Informatica

Informatica® Test Data Management 9.7.1 HotFix1

User Guide

Informatica Test Data Management User Guide
9.7.1 HotFix1
August 2016

© Copyright Informatica LLC 2003, 2018

This software and documentation contain proprietary information of Informatica LLC and are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright law. Reverse engineering of the software is prohibited. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC. This Software may be protected by U.S. and/or international Patents and other Patents Pending.

Use, duplication, or disclosure of the Software by the U.S. Government is subject to the restrictions set forth in the applicable software license agreement and as provided in DFARS 227.7202-1(a) and 227.7702-3(a) (1995), DFARS 252.227-7013©(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19, or FAR 52.227-14 (ALT III), as applicable.

The information in this product or documentation is subject to change without notice. If you find any problems in this product or documentation, please report them to us in writing.

Informatica, Informatica Platform, Informatica Data Services, PowerCenter, PowerCenterRT, PowerCenter Connect, PowerCenter Data Analyzer, PowerExchange, PowerMart, Metadata Manager, Informatica Data Quality, Informatica Data Explorer, Informatica B2B Data Transformation, Informatica B2B Data Exchange Informatica On Demand, Informatica Identity Resolution, Informatica Application Information Lifecycle Management, Informatica Complex Event Processing, Ultra Messaging, Informatica Master Data Management, and Live Data Map are trademarks or registered trademarks of Informatica LLC in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties, including without limitation: Copyright DataDirect Technologies. All rights reserved. Copyright © Sun Microsystems. All rights reserved. Copyright © RSA Security Inc. All Rights Reserved. Copyright © Ordinal Technology Corp. All rights reserved. Copyright © Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright © Meta Integration Technology, Inc. All rights reserved. Copyright © Intalio. All rights reserved. Copyright © Oracle. All rights reserved. Copyright © Adobe Systems Incorporated. All rights reserved. Copyright © DataArt, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Microsoft Corporation. All rights reserved. Copyright © Rogue Wave Software, Inc. All rights reserved. Copyright © Teradata Corporation. All rights reserved. Copyright © Yahoo! Inc. All rights reserved. Copyright © Glyph & Cog, LLC. All rights reserved. Copyright © Thinkmap, Inc. All rights reserved. Copyright © Clearpace Software Limited. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright © International Organization for Standardization 1986. All rights reserved. Copyright © ej-technologies GmbH. All rights reserved. Copyright © Jaspersoft Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © yWorks GmbH. All rights reserved. Copyright © Lucent Technologies. All rights reserved. Copyright (c) University of Toronto. All rights reserved. Copyright © Daniel Veillard. All rights reserved. Copyright © Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright © MicroQuill Software Publishing, Inc. All rights reserved. Copyright © PassMark Software Pty Ltd. All rights reserved. Copyright © LogiXML, Inc. All rights reserved. Copyright © 2003-2010 Lorenzi Davide, All rights reserved. Copyright © Red Hat, Inc. All rights reserved. Copyright © The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright © EMC Corporation. All rights reserved. Copyright © Flexera Software. All rights reserved. Copyright © Jinfonet Software. All rights reserved. Copyright © Apple Inc. All rights reserved. Copyright © Telerik Inc. All rights reserved. Copyright © BEA Systems. All rights reserved. Copyright © PDFlib GmbH. All rights reserved. Copyright © Orientation in Objects GmbH. All rights reserved. Copyright © Tanuki Software, Ltd. All rights reserved. Copyright © Ricebridge. All rights reserved. Copyright © Sencha, Inc. All rights reserved. Copyright © Scalable Systems, Inc. All rights reserved. Copyright © jQWidgets. All rights reserved. Copyright © Tableau Software, Inc. All rights reserved. Copyright© MaxMind, Inc. All Rights Reserved. Copyright © TMate Software s.r.o. All rights reserved. Copyright © MapR Technologies Inc. All rights reserved. Copyright © Amazon Corporate LLC. All rights reserved. Copyright © Highsoft. All rights reserved. Copyright © Python Software Foundation. All rights reserved. Copyright © BeOpen.com. All rights reserved. Copyright © CNRI. All rights reserved.

This product includes software developed by the Apache Software Foundation (http://www.apache.org/), and/or other software which is licensed under various versions of the Apache License (the "License"). You may obtain a copy of these Licenses at http://www.apache.org/licenses/. Unless required by applicable law or agreed to in writing, software distributed under these Licenses is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the Licenses for the specific language governing permissions and limitations under the Licenses.

This product includes software which was developed by Mozilla (http://www.mozilla.org/), software copyright The JBoss Group, LLC, all rights reserved; software copyright © 1999-2006 by Bruno Lowagie and Paulo Soares and other software which is licensed under various versions of the GNU Lesser General Public License Agreement, which may be found at http:// www.gnu.org/licenses/lgpl.html. The materials are provided free of charge by Informatica, "as-is", without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (copyright The OpenSSL Project. All Rights Reserved) and redistribution of this software is subject to terms available at http://www.openssl.org and http://www.openssl.org/source/license.html.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://curl.haxx.se/docs/copyright.html. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

The product includes software copyright 2001-2005 (©) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://www.dom4j.org/ license.html.

The product includes software copyright © 2004-2007, The Dojo Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://dojotoolkit.org/license.

This product includes ICU software which is copyright International Business Machines Corporation and others. All rights reserved. Permissions and limitations regarding this software are subject to terms available at http://source.icu-project.org/repos/icu/icu/trunk/license.html.

This product includes software copyright © 1996-2006 Per Bothner. All rights reserved. Your right to use such materials is set forth in the license which may be found at http:// www.gnu.org/software/ kawa/Software-License.html.

This product includes OSSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at http://www.opensource.org/licenses/mit-license.php.

This product includes software developed by Boost (http://www.boost.org/) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at http:/ /www.boost.org/LICENSE_1_0.txt.

This product includes software copyright © 1997-2007 University of Cambridge. Permissions and limitations regarding this software are subject to terms available at http:// www.pcre.org/license.txt.

This product includes software copyright © 2007 The Eclipse Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http:// www.eclipse.org/org/documents/epl-v10.php and at http://www.eclipse.org/org/documents/edl-v10.php.

This product includes software licensed under the terms at http://www.tcl.tk/software/tcltk/license.html, http://www.bosrup.com/web/overlib/?License, http://www.stlport.org/doc/ license.html, http://asm.ow2.org/license.html, http://www.cryptix.org/LICENSE.TXT, http://hsqldb.org/web/hsqlLicense.html, http://httpunit.sourceforge.net/doc/ license.html, http://jung.sourceforge.net/license.txt , http://www.gzip.org/zlib/zlib_license.html, http://www.openldap.org/software/release/license.html, http://www.libssh2.org, http://slf4j.org/license.html, http://www.sente.ch/software/OpenSourceLicense.html, http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3- license-agreement; http://antlr.org/license.html; http://aopalliance.sourceforge.net/; http://www.bouncycastle.org/licence.html; http://www.jgraph.com/jgraphdownload.html; http://www.jcraft.com/jsch/LICENSE.txt; http://jotm.objectweb.org/bsd_license.html; . http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231; http://www.slf4j.org/license.html; http://nanoxml.sourceforge.net/orig/copyright.html; http://www.json.org/license.html; http://forge.ow2.org/projects/javaservice/, http://www.postgresql.org/about/licence.html, http://www.sqlite.org/copyright.html, http://www.tcl.tk/software/tcltk/license.html, http://www.jaxen.org/faq.html, http://www.jdom.org/docs/faq.html, http://www.slf4j.org/license.html; http://www.iodbc.org/dataspace/iodbc/wiki/iODBC/License; http://www.keplerproject.org/md5/license.html; http://www.toedter.com/en/jcalendar/license.html; http://www.edankert.com/bounce/index.html; http://www.net-snmp.org/about/license.html; http://www.openmdx.org/#FAQ; http://www.php.net/license/3_01.txt; http://srp.stanford.edu/license.txt; http://www.schneier.com/blowfish.html; http://www.jmock.org/license.html; http://xsom.java.net; http://benalman.com/about/license/; https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js; http://www.h2database.com/html/license.html#summary; http://jsoncpp.sourceforge.net/LICENSE; http://jdbc.postgresql.org/license.html; http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto; https://github.com/rantav/hector/blob/master/LICENSE; http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html; http://jibx.sourceforge.net/jibx-license.html; https://github.com/lyokato/libgeohash/blob/master/LICENSE; https://github.com/hjiang/jsonxx/blob/master/LICENSE; https://code.google.com/p/lz4/; https://github.com/jedisct1/libsodium/blob/master/LICENSE; http://one-jar.sourceforge.net/index.php?page=documents&file=license; https://github.com/EsotericSoftware/kryo/blob/master/license.txt; http://www.scala-lang.org/license.html; https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt; http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html; https://aws.amazon.com/asl/; https://github.com/twbs/bootstrap/blob/master/LICENSE; https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt; https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE, and https://github.com/apache/hbase/blob/master/LICENSE.txt.

This product includes software licensed under the Academic Free License (http://www.opensource.org/licenses/afl-3.0.php), the Common Development and Distribution License (http://www.opensource.org/licenses/cddl1.php) the Common Public License (http://www.opensource.org/licenses/cpl1.0.php), the Sun Binary Code License Agreement Supplemental License Terms, the BSD License (http:// www.opensource.org/licenses/bsd-license.php), the new BSD License (http://opensource.org/licenses/BSD-3-Clause), the MIT License (http://www.opensource.org/licenses/mit-license.php), the Artistic License (http://www.opensource.org/licenses/artistic-license-1.0) and the Initial Developer's Public License Version 1.0 (http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at http://xstream.codehaus.org/license.html. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit http://www.extreme.indiana.edu/.

This product includes software Copyright (c) 2013 Frank Balluffi and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.

See patents at https://www.informatica.com/legal/patents.html.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

Publication Date: 2018-06-26

Table of Contents

Preface 16

Informatica Resources. 16

Informatica Network. 16

Informatica Knowledge Base. 16

Informatica Documentation. 17

Informatica Product Availability Matrixes. 17

Informatica Velocity. 17

Informatica Marketplace. 17

Informatica Global Customer Support. 17

Chapter 1: Introduction to Test Data Management. 18

Test Data Management Overview. 18

Test Data Management Use Cases. 19

Using TDM to Improve Performance in Application Testing Environments. 20

Using TDM for Security and Compliance in Application Testing Environments. 20

Using TDM With Data Validation Option to Create Verified Test Data. 21

TDM Architecture. 21

TDM Tools. 22

TDM Server. 22

Application Services. 22

TDM Databases. 23

TDM Connections. 23

TDM Process. 25

Create a Data Masking Policy. 25

Create a Project and Import Metadata. 26

Discover Source Information. 26

Define Data Masking and Data Subset Operations. 26

Define a Data Generation Operation. 27

Create a Plan for Data Masking and Data Subset. 27

Create a Plan for Data Generation. 27

Monitor the Workflow. 28

TDM Example. 28

Chapter 2: Test Data Manager. 30

Test Data Manager Overview. 30

Test Data Manager User Interface. 31

Views. 31

Search Field. 32

Quick Links Menu. 32

User Preferences Menu. 32

Actions Menu. . . . 32
Data Filter. . . . 32
Overview View. . . . 33
Dashboards. . . . 33
Project Risk Analysis. . . . 33
Project Assignments and Project Objects. . . . 34
Recent Project Sensitive Data Distribution. . . . 34
Recent Activities. . . . 34
Plan Execution. . . . 34
Global Components. . . . 34
Recent Projects. . . . 34
Policies View. . . . 35
Data Sets View. . . . 35
Projects View. . . . 36
Monitor View. . . . 37
Administrator View. . . . 37
Expression Builder. . . . 38
Logging In to the Test Data Manager. . . . 39

Chapter 3: Projects. . . . 41

Projects Overview. . . . 41
Project Components. . . . 42
Project Logs. . . . 43
Project Management. . . . 43
Creating a Project. . . . 44
Editing a Project. . . . 44
Copying a Project. . . . 45
Deleting a Project. . . . 45
Configuring Project Log Locations. . . . 45
Exporting a Project. . . . 46
Importing a Project. . . . 46
Data Sources. . . . 47
TDM Source Definitions. . . . 47
XSD Data Sources. . . . 47
Editing the Properties of an Element in an XSD File. . . . 49
XSD Import Guidelines. . . . 50
PowerExchange Nonrelational Sources. . . . 50
Importing Data Sources. . . . 50
Identity Columns in TDM. . . . 51
Project Permission and Security. . . . 51
Project Permissions. . . . 52
Updating User and Group Security. . . . 52

Chapter 4: Policies 54

Policies Overview 54
Policies View 55
Policies Task Flow 55
Data Masking Rules 55
- Standard Masking Rules 56
- Advanced Masking Rules 57
- Mapplet Rules 58

Data Domains 59
- Apply Masking Rules to a Data Domain 60
- Apply Generation Rules to a Data Domain 60
- Metadata and Data Patterns for Data Domains 60
- Regular Expression Syntax 60
- Data Patterns 61
- Metadata Patterns 61
- Data Domain Options 62
- Creating a Data Domain 62
- Copying a Data Domain 63
- Editing a Data Domain 63
- Deleting a Data Domain 63

Policy Packs 64
- PII Policy Pack 64
- PHI Policy Pack 65
- PCI Policy Pack 66

Import and Export 67
- Exporting Policy Components 67
- Importing Policy Components 67
- Data Domain Import 67

Linking Business Glossary Terms to Global Objects 68
- Linking a Business Term to an Object 69
- Deleting a Business Term Link to an Object 69

Policy Management 69
- Creating a Policy 70
- Copying a Policy 70
- Editing a Policy 70
- Deleting a Policy 71

Chapter 5: Data Discovery 72

Data Discovery Overview 72
Data Discovery Sources 73
- Rules and Guidelines for Data Discovery Sources 73

Discover View 74

Column Properties. . . . 75
Sensitive Columns. . . . 75
Value Cascades. . . . 75
Auto Cascades. . . . 77
Data Discovery Task Flow. . . . 77
Primary Key Discovery 77
Primary Key Profile Options. . . . 78
Entity Discovery. . . . 78
Entity Profile Options. . . . 78
Data Domain Discovery. . . . 79
Data Domain Profile Sampling Options. . . . 80
Assigning a Data Domain to Multiple Columns. . . . 80
Manually Updating the Column Data Domain. . . . 80
Column Profile. . . . 81
Inferred Rules. . . . 81
Column Profile Options. . . . 82
Profile Management. . . . 82
Creating a Primary Key Profile. . . . 82
Creating an Entity Profile. . . . 83
Creating a Data Domain Profile. . . . 84
Creating a Column Profile. . . . 84
Editing a Profile. . . . 85
Deleting a Profile. . . . 85
Profile Import. . . . 86
Importing a Profile. . . . 86
Apply the Results. . . . 87
Primary Key Results. . . . 87
Entity Discovery Results. . . . 88
Data Domain Discovery Results. . . . 88
Column Profile Results. . . . 90
Project Tables. . . . 91
Table Classification. . . . 91
Constraints. . . . 91
Manually Add Keys to Tables in a Project. . . . 93
Creating a Primary Key Constraint. . . . 93
Creating a Logical Relationship Between Tables. . . . 93
Creating a Unique Key Constraint. . . . 94
Creating a Conditional Constraint. . . . 95

Chapter 6: Data Subset. . . . 96

Data Subset Overview. . . . 96
Data Subset Task Flow. . . . 97
Entities. . . . 97

Entity Example. . . . 97
Entities View. . . . 98
Entity Map View. . . . 98
Tables View. . . . 99
Relationships View. . . . 99
Review Changes. . . . 101
Entity Management. . . . 102
Creating an Entity. . . . 102
Parameter Values in Entity Criteria. . . . 102
Database Functions in Entity Criteria 103
Editing an Entity. . . . 104
Optimizing Relations. . . . 104
Exporting an Entity. . . . 105
Importing an Entity. . . . 105
Deleting an Entity. . . . 105
Groups. . . . 106
Creating a Group. . . . 106
Editing a Group. . . . 106
Copying a Group. . . . 106
Exporting a Group. . . . 107
Importing a Group. . . . 107
Deleting a Group. . . . 107
Templates. . . . 108
Template Example. . . . 108
Creating a Template. . . . 108
Editing a Template. . . . 109
Deleting a Template. . . . 109
Data Subset for XSD Sources. . . . 109
Data Subset for XSD Sources Task Flow. . . . 110
Applying Criteria to an Element or Attribute. . . . 110
Example - Data Subset for XSD Data Sources. . . . 111
Data Subset Plans and Workflows. . . . 118

Chapter 7: Data Masking. . . . 119

Data Masking Overview. . . . 119
Data Masking Components. . . . 120
Data Masking Task Flow. . . . 120
Masking Rules Management. . . . 121
Modify Masking Rules. . . . 121
Editing a Masking Rule. . . . 121
Copying a Masking Rule. . . . 122
Deleting a Masking Rule. . . . 122
Masking Rule Assignments. . . . 122

Available Data Masking Rules . . . 123
Assigning Masking Rules to Columns. . . . 124
Overriding a Masking Rule. . . . 124
Assigning Multiple Masking Rules to a Column. . . . 124
Assigning Default Values to Multiple Columns . . . 125
Assigning Custom Masking Rules to Columns. . . . 125
Assigning Advanced Masking Rules to Columns. . . . 126
Deleting Masking Rule Assignments. . . . 126
Data Masking for XSD Sources. . . . 127
Data Masking Task Flow for XSD Sources. . . . 127
Applying Data Masking Rules to XML Elements and Attributes. . . . 127
Data Masking Plans and Workflows. . . . 128
Masking Components in PowerCenter. . . . 129
Mapplet Export. . . . 129

Chapter 8: Data Masking Techniques and Parameters. . . . 130

Data Masking Techniques and Parameters Overview. . . . 131
Data Masking Techniques. . . . 131
Data Masking Parameters. . . . 132
Repeatable Output. . . . 133
Exception Handling. . . . 134
Custom Masking. . . . 135
Custom Masking Parameters. . . . 135
Advanced Masking. . . . 135
Advanced Masking Parameters. . . . 136
Advanced Masking Example. . . . 137
Credit Card Masking. . . . 137
Credit Card Masking Parameters. . . . 138
Email Masking. . . . 138
Email Masking Parameters. . . . 138
Encryption Masking. . . . 140
Expression Masking. . . . 140
Expression Masking Parameters. . . . 141
Rules and Guidelines for Expression Masking. . . . 141
IP Address Masking. . . . 141
Key Masking. . . . 142
Mask Format. . . . 142
Source String Characters. . . . 143
Result String Replacement Characters. . . . 143
Date Key Masking . . . 144
Numeric Key Masking Parameters. . . . 144
String Key Masking Parameters. . . . 144
Nullification Masking. . . . 145

Phone Masking. 145
Random Masking. 145
Range Masking. 146
Blurring. 146
Mask Format. 147
Source String Characters. 147
Result String Replacement Characters. 148
Date Random Masking Parameters. 148
Numeric Random Masking Parameters. 149
String Random Masking Parameters. 149
Shuffle Masking. 150
Shuffle Masking Parameters. 150
SIN Masking. 151
SSN Masking. 151
Substitution Masking. 152
Substitution Masking Parameters. 152
URL Masking. 153
Name Substitution Example 153
Add a Dictionary in Test Data Manager. 154
Creating the Substitution Rule. 155
Creating the Advanced Masking Rule. 155
Shuffle Address Example. 159
Creating the Shuffle Rule. 160
Create the Advanced Masking Rule. 160

Chapter 9: Data Generation 166

Data Generation Overview. 166
Data Generation Components. 167
Data Generation Task Flow. 167
Data Generation Rule Types. 168
Default Settings. 169
Define Default Generation Rules. 170
Standard Generation Rules. 170
Creating a Standard Generation Rule. 170
Editing a Generation Rule. 171
Overriding a Generation Rule. 171
Deleting a Generation Rule. 171
Custom Generation Rules. 172
Creating a Custom Generation Rule. 172
Ad Hoc Generation Rules. 173
Creating an Ad Hoc Generation Rule. 173
Editing an Ad Hoc Generation Rule. 174
Conditional Constraints. 175

Conditional Constraints and Data Conflicts. 175
Data Generation Rule Assignments. 176
Available Data Generation Rules. 177
Assigning Standard Generation Rules to Columns. 177
Assigning Custom Generation Rules to Columns. 177
Assigning Auto Rules to Columns. 178
Deleting Generation Rule Assignments. 178
Data Generation for XSD Sources. 179
Data Generation Task Flow for XSD Sources. 179
Applying Data Generation Rules to XML Elements and Attributes. 179
Data Generation Plans and Workflows. 180

Chapter 10: Data Generation Techniques and Parameters. 181

Data Generation Techniques and Parameters Overview. 181
Data Generation Techniques. 182
Data Generation Parameters. 183
Exception Test Data. 183
Custom Generation. 184
Custom Generation Parameters. 185
Dictionary Generation. 185
Dictionary Generation Parameters. 186
Effective Dates Generation. 187
Effective Dates Generation Parameters. 188
Effective Dates Generation Example. 189
Expression Generation. 189
Expression Generation Parameters. 190
Random Generation. 191
Date Random Generation Parameters. 191
String Random Generation Parameters. 192
Numeric Random Generation Parameters. 193
Data Patterns for Random Generation. 194
Reference Lookup Generation. 197
Reference Lookup Generation Parameters. 198
Sequence Generation. 199
Date Sequence Generation Parameters. 200
Numeric Sequence Generation Parameters. 201
Set of Values Generation. 201
Set of Values Generation Parameters. 202

Chapter 11: Working with Test Data Warehouse. 203

Test Data Warehouse Overview. 203
Test Data Warehouse Process. 204
Data Sets. 204

Data Set Tags. . . . 205
Creating a Data Set. . . . 206
Reset a Data Set. . . . 206
Classification of Data Set Tables. . . . 207
Classifying a Data Set Table. . . . 208
Reset With and Without Downtime. . . . 208
Resetting a Data Set. . . . 209
Reset Process Flow. . . . 210
Reset with Truncate Tables Process Flow. . . . 211
Reset Without Truncate Tables Process Flow. . . . 212
Restoring Deleted Data After a Failed Reset. . . . 213
Editing the Metadata of a Data Set. . . . 213
Deleting a Data Set. . . . 213
Data Set Permissions. . . . 214
Editing Data Set Version Permission. . . . 214
Locking and Unlocking a Data Set Version. . . . 215
Monitor a Data Set Job. . . . 215

Chapter 12: Plans and Workflows. . . . 217

Plans and Workflows Overview. . . . 217
Plans and Workflows Task List. . . . 218
Workflow Connections. . . . 218
Plan Components. . . . 219
Plan Settings. . . . 220
Connection Options. . . . 220
Data Set Properties. . . . 222
Test Tool Integration Options. . . . 223
Target Options. . . . 223
Update Strategy Options. . . . 224
Data Generation Properties. . . . 224
Error and Recovery Options. . . . 225
Advanced Options. . . . 225
Log Settings. . . . 228
Masking Components. . . . 229
Subset Components. . . . 229
Generation Components. . . . 230
Component Criteria. . . . 230
Filtering Data Subset Components. . . . 230
Disabling Masking for a Column. . . . 231
Source Settings. . . . 231
Connection Properties. . . . 231
Test Tool Integration Properties. . . . 234
Target Properties. . . . 234

Update Strategy Properties. . . . 234
Error and Recovery Properties. . . . 235
Source and Target Properties. . . . 235
Advanced Properties. . . . 235
Partitioning Properties. . . . 236
XSD Data Source Connection Settings. . . . 237
XSD Data Source Advanced Settings. . . . 238
Plan Management. . . . 238
Creating a Data Masking and Data Subset Plan. . . . 238
Creating a Data Generation Plan. . . . 239
Creating an Ad Hoc Generation Plan. . . . 240
Copying a Plan. . . . 240
Exporting a Plan. . . . 241
Importing a Plan. . . . 241
Deleting a Plan. . . . 241
Workflow Generation. . . . 241
Generating a Workflow. . . . 242
Export to HP ALM. . . . 242
Workflow View. . . . 242
Executing a Workflow. . . . 243
Workflow Executions View. . . . 243
Workflow Tasks. . . . 244
Workflow Properties Panel. . . . 245

Chapter 13: Monitor. . . . 247

Monitor Overview. . . . 247
Jobs. . . . 248
Job Details. . . . 248
Monitor Tasks. . . . 249
Logs. . . . 250
Severity Levels. . . . 250
Viewing the Log Messages. . . . 250
Sessions. . . . 251

Chapter 14: Reports. . . . 252

Reports Overview. . . . 252
Data Masking Report. . . . 253
Running the Data Masking Report. . . . 253
Plan Audit Report. . . . 253
Running a Plan Audit Report. . . . 253
Plan Detail Report. . . . 254
Running the Plan Detail Report. . . . 254
Row Count Report. . . . 255

Running the Row Count Report. 255

Chapter 15: ilmcmd 256

ilmcmd Overview. 256
Configuring ilmcmd. 257
Running ilmcmd. 257
Entering Options and Arguments. 257
Syntax Notation. 258
Delete. 258
Delete Examples. 259
Export. 260
Export Examples. 261
Import. 261
Import Examples. 262
Search. 262
Search Examples. 263
Workflow. 264
Workflow Examples. 265
Reset. 265
Register. 267

Chapter 16: tdwcmd. 269

tdwcmd Overview. 269
Running tdwcmd. 269
Entering Options and Arguments. 270
Syntax Notation. 270
Version. 271
List. 271
List Examples. 273

Chapter 17: tdwquery. 274

tdwquery Overview. 274
Configuring tdwquery. 274
Running tdwquery. 275
Select Clause. 276
Joins. 276
Aggregate Functions. 277
Where Clause. 278
Additional Clauses. 280

Appendix A: Data Type Reference. 281

Data Type Reference Overview. 281
Oracle. 282

Microsoft SQL Server. 286
DB2 for Linux, UNIX, and Windows. 290
Sybase ASE. 292
MySQL. 295
Flat File. 297
VSAM. 298
DB2 for z/OS. 299
DB2 for IOS. 301
IMS Flat. 303
Sybase IQ. 304
Netezza. 306
Teradata. 308
Cassandra. 310
MongoDB. 311

Appendix B: Data Type Reference for Test Data Warehouse. 312

Data Type Reference for Test Data Warehouse Overview. 312
Oracle. 313
Microsoft SQL Server. 315
DB2 for Linux, UNIX, and Windows. 317
Sybase ASE. 318
Teradata. 320

Appendix C: Glossary. 322

Index. 328

Preface

The Informatica *Test Data Management User Guide* describes how to protect sensitive data and create lean nonproduction systems for test and development. It shows you how to implement data subset, data masking, data generation, and data discovery operations. This guide is written for users who use Test Data Manager with PowerCenter. It assumes knowledge of operating systems, database engines, flat files, and PowerCenter.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit https://network.informatica.com.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit https://kb.informatica.com. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com.

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at https://network.informatica.com/community/informatica-network/product-availability-matrices.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services. Developed from the real-world experience of hundreds of data management projects, Informatica Velocity represents the collective knowledge of our consultants who have worked with organizations from around the world to plan, develop, deploy, and maintain successful data management solutions.

If you are an Informatica Network member, you can access Informatica Velocity resources at http://velocity.informatica.com.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that augment, extend, or enhance your Informatica implementations. By leveraging any of the hundreds of solutions from Informatica developers and partners, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at https://marketplace.informatica.com.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through Online Support on Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link: http://www.informatica.com/us/services-and-training/support-services/global-support-centers.

If you are an Informatica Network member, you can use Online Support at http://network.informatica.com.

CHAPTER 1

Introduction to Test Data Management

This chapter includes the following topics:

- Test Data Management Overview, 18
- Test Data Management Use Cases, 19
- TDM Architecture, 21
- TDM Process, 25
- TDM Example, 28

Test Data Management Overview

Test Data Management (TDM) integrates with PowerCenter, PowerExchange, and Informatica applications to manage nonproduction data in an organization.

With TDM, an organization can create a smaller copy of the production data and mask the sensitive data. An organization can discover the sensitive columns in the test data, and ensure that the sensitive columns are masked in the test data. An organization can also create test data that does not contain sensitive data from the production database.

Organizations create multiple copies of application data to use for testing and development. Organizations often maintain strict controls on production systems, but data security in nonproduction systems is not as secure. An organization must maintain knowledge of the sensitive columns in the production data and ensure that sensitive data does not appear in the test environment. Development must not have to rewrite code to create test data.

Manage data discovery, data subset, data masking, and data generation in Test Data Manager.

Data discovery

Use data discovery to run sensitive field profiles to identify the columns that contain sensitive data. Use the profile results to determine which columns to mask and which data masking techniques to apply. Define data domains to identify sensitive data columns by patterns in the data or the column metadata. When you apply data masking, you can apply the same rule to multiple columns in the same data domain. You can run primary and foreign key profiles to discover potential primary key-foreign key constraints to define relationships between parent and child tables.

Data subset

Use data subset to create a small environment for testing and development. You can define the type of data that you want to include in the subset database. You might create a subset database with data based on time, function, or geographic location. For example, a time-based subset database might include recent payment transactions from all invoice data in a production system.

Data masking

Create data masking rules to apply to source columns and data domains. You can apply different masking techniques such as substitution masking, shuffle masking, key masking, and encryption. You can configure repeatable results in the masked data. You can assign multiple rules to the same column.

Data generation

Use data generation to create a testing environment that does not use data from the production database. Create data generation rules to define the type of data you want to generate. TDM generates data in a schema that you can use for testing.

To perform data subset and masking operations, you can generate and run workflows from data subset and data masking plans in Test Data Manager. To perform data generation operations, you can generate and run data generation plans in Test Data Manager.

You can export test data to an HP ALM server from TDM. You can copy results of subset, masking, and generation operations that have flat file targets. Integrate the HP-ALM test tool with TDM to directly copy and maintain flat file results in an HP ALM server. You can then use the data to create and run test cases in HP ALM.

You can configure a test data warehouse for TDM. Store multiple versions of test data that you can retrieve, enhance, and reset, in the test data warehouse. For example, consider a situation where you run multiple test cases, or multiple test teams work on an application. Store the non production data in the test data warehouse. When one test team completes testing, save the modified test data as another version of the original data set in the test data warehouse. Restore the required version from the test data warehouse to the test environment to run other test cases or for a different team to work with.

TDM includes the ilmcmd command line program. Run ilmcmd commands to perform a subset of the Test Data Manager tasks from the command line.

TDM users have roles and privileges that determine the tasks that they can perform through Test Data Manager or the ilmcmd command line program. The administrator manages roles and privileges for users from the Informatica Administrator.

Test Data Management Use Cases

Improving performance, security, and compliance in application testing environments are all use cases for Test Data Management.

Using TDM to Improve Performance in Application Testing Environments

Testing teams need to be efficient and agile to create and maintain test data while maintaining a collaborative testing environment. You can use the features available in TDM to help you meet this requirement.

You can perform the following tasks to improve performance in a testing environment:

- You can create subsets of production data to use as test data. The data subset retains primary and foreign key relations and the data remains referentially intact. Use data subset to create manageable and scalable data for testing.
- You can directly copy flat file test data from TDM to an integrated test tool and use the data to run test cases.
- You can generate data based on rules that define the type of data required. Use data generation when you need test data but do not want to use production data. For example, consider test teams that work on applications that handle sensitive data. The test teams do not have access to production data. They can use data generation to create test data.
- You can store multiple versions of test data in a common location. Use the test data warehouse as a single and secure location to maintain required versions of test data. You can store data from multiple machines that host a Test Data Manager Service in a domain in a single test data warehouse. Multiple testing teams can then share the stored data and collaborate efficiently.
- You can restore a data set version to return the test environment to a particular state. For example, you might need to return the test environment to an older state or to the original state after you run some test cases. Use the test data warehouse to reduce the time and effort to create and manually manipulate test data.

Using TDM for Security and Compliance in Application Testing Environments

Testing teams need to be compliant and maintain strict controls on the data used in testing environments.

You must identify sensitive information in huge volumes of production data and then ensure that you do not expose the information in the test data. This is a challenge and a compliance and security risk.

You can perform the following tasks to help with security and compliance in a test environment:

- To analyze the data and identify sensitive information in the data, run profiles in TDM. You can compare the data against multiple compliance regulation standards to ensure that the data meets standard compliance regulations. For example, run a profile on the data to identify columns that contain sensitive data according to PII, PCI, or PHI compliance regulations.
- To remove sensitive data that you identify, run a data masking operation in TDM. You can use standard masking formats such as credit card numbers or Social Security numbers to mask data or create customized formats to mask data.
- To reduce the risk associated with creating multiple copies of test data, store the data in the test data warehouse. Multiple teams that use different machines that host a Test Data Manager Service in a domain can access a single test data warehouse.

Using TDM With Data Validation Option to Create Verified Test Data

Data Validation Option (DVO) is a solution that you use with PowerCenter to validate data. You can validate target data to verify that it is accurate and that the transformation process did not introduce errors or inconsistencies.

Perform data validation to compare production and development environments when you use TDM to create test data.

For example, consider test teams that need to submit compliance reports for test data created from production data. You can use TDM with DVO to deliver test data with verifiable proof that the test data does not contain sensitive data values.

TDM Architecture

The TDM architecture consists of tools, the TDM server, the Test Data Manager Service and other application services, and databases.

The following image shows the components of TDM:

The TDM architecture consists of the tools that you use to modify the data.

The application services include Model Repository Service, Data Integration Service, PowerCenter Repository Service, PowerCenter Integration Service, and Test Data Manager Service. The Model Repository Service and Data Integration Service access the Model repository and profiling warehouse.

Test Data Manager accesses the Test Data Manager Service and uses the database content from the TDM repository and the test data warehouse repository. PowerCenter services access the PowerCenter repository. The application services access the domain configuration repository and Model repository that maintain connections and metadata for the domain.

TDM Tools

The TDM tools consist of Test Data Manager, Informatica Developer, Informatica Administrator, and the PowerCenter Client.

Test Data Manager

A web-based client application that you can use to configure data masking, data subset, data generation, and profiles for data discovery. You can also configure connections, and manage project permissions for users and user groups.

Informatica Developer

A client application that you use to create and export profiles for data discovery.

Informatica Administrator

A web-based client that a domain administrator uses to manage application services and create users and user groups.

PowerCenter Client

A client application that you use to configure permissions on folders and on connection objects for the PowerCenter repository.

TDM Server

The TDM server is the interface between Test Data Manager and the application services.

Application Services

TDM uses Informatica services. Create the services in the Administrator tool.

TDM uses the following application services:

PowerCenter Repository Service

An application service that manages the PowerCenter repository. The PowerCenter Repository Service accepts requests from the PowerCenter Integration Service when a workflow runs.

PowerCenter Integration Service

An application service that runs data subset, data generation, and data masking workflows. When you create the Test Data Manager Service in the Informatica Administrator, you select the PowerCenter Integration Service that runs the workflows.

Model Repository Service

An application service that manages the Model repository for data discovery operations.

Data Integration Service

An application service that performs the data discovery operations. The Data Integration Service connects to the Model Repository Service to store metadata from data discovery profiles in the Model repository. When you create a Data Integration Service in the Administrator tool, you select the data profiling warehouse to store data from data discovery profiles.

Analyst Service

Manages the Analyst tool that TDM uses to link TDM global objects to objects in the Business Glossary. The license for the Analyst Service to use database content from the TDM repository.

Test Data Manager Service

An application service that creates and manages the TDM repository. The Test Data Manager accesses the Test Data Manager Service to use database content from the TDM repository.

TDM Databases

TDM connects to databases for PowerCenter metadata, profiling, TDM configuration, domain configuration, and the test data warehouse.

TDM needs connections to the following databases:

TDM repository

A relational database that contains tables that TDM requires to run and the tables that store metadata.

Model repository

A relational database that stores table metadata for data discovery profiles and the connections that you create in Test Data Manager.

PowerCenter repository

A relational database that stores metadata for PowerCenter sources and targets. The PowerCenter repository also stores metadata for the workflows that you generate from plans in Test Data Manager.

Profiling warehouse

A relational database that stores profile results for data discovery.

Domain configuration repository

A relational database that stores connections and metadata for the Informatica domain.

Test data warehouse repository

A relational database that stores the project metadata and source table metadata for data sets that you store in the test data warehouse.

Test data warehouse

A relational database that stores the source data of data sets that you create in Test Data Manager.

TDM Connections

TDM connects to databases, repositories, and services to perform data subset, masking, generation, and profiles for discovery operations. The connection requirements are based on the operations that you need to perform.

To perform data discovery operations, TDM requires connections to a database source and a Data Integration Service.

Workflows for data subset and masking operations require connections to services, the TDM repository, external database sources, and PowerCenter sources and targets.

Workflows for operations to store data in the test data warehouse require connections to external database sources, the test data warehouse repository, and the test data warehouse.

To run workflows to copy flat file results to a test tool, you must first integrate the test tool server with TDM. TDM connects to the integrated server to copy the flat file to the configured location.

Service Connections

TDM requires connections to the following services:

Data Integration Service

TDM requires a connection to a Data Integration Service for data discovery operations. The Data Integration Service is the service in the Informatica domain that performs the data discovery operation.

PowerCenter Integration Service

A data masking, data generation, or data subset workflow requires a connection to a PowerCenter Integration Service. When you generate a workflow from a plan, select the name of the PowerCenter Integration Service to use.

PowerCenter Repository Service

When you import metadata into Test Data Manager from a PowerCenter source, the TDM Server sends a request to the PowerCenter Repository Service to extract source metadata from the PowerCenter repository and load it to the TDM repository.

You create a Test Data Manager Service in the Informatica domain. When you create the Test Data Manager Service, you select the PowerCenter Repository Service. The TDM Server connects to the PowerCenter repository and creates the TDM repository tables.

Repository Connections

TDM requires connections to repositories.

TDM accesses the following repositories:

TDM repository

The PowerCenter Integration Service stores TDM components in the TDM repository.

A data masking, data generation, or data subset workflow requires a connection to the TDM repository.

Model repository

When you run profiles to discover data, the TDM Server sends a request to the Data Integration Service to extract data for the source tables. The Data Integration Service sends a request to its associated Model Repository Service to load the metadata for the tables to the Model repository.

Profiling warehouse

The Data Integration Service loads the profile results to the profiling warehouse. When you create a Data Integration Service in the Administrator tool, you configure a profiling warehouse.

PowerCenter repository

When you import metadata into Test Data Manager from a PowerCenter source, the TDM Server sends a request to the PowerCenter Repository Service to copy source metadata from the PowerCenter repository into the TDM repository.

Test data warehouse repository

When you run a TDM operation to create a data set in the test data warehouse, the Test Data Manager Service stores the project metadata and source table metadata in the test data warehouse repository.

Database Connections

TDM requires a connection to a database source to perform data discovery operations and a database target to perform data generation operations. It also requires connections for relational sources and targets for data subset and data masking. To store data sets in the test data warehouse, TDM requires a connection to a test data warehouse.

Configure connections in the Administrator view of Test Data Manager. To connect TDM to databases, you do not need to install a separate driver. To create a test data warehouse, add the database connection and

then configure the connection as the test data warehouse from the **Administrator | Test Data Warehouse** view in Test Data Manager.

When you generate workflows from data subset or masking plans, you must select connections for relational sources and targets. When you generate workflows from data generation plans, you must select connections for targets. Select these connections in Test Data Manager when you create a plan. TDM tests the connections when it generates workflows from plans, and the PowerCenter Integration Service uses the connections when it runs the workflows.

TDM Process

Run a profile against source data, create a subset of the data, and mask the subset data.

The TDM process includes the following high-level steps:

1. Create policies that define the types of data you want to mask and the rules that you might use to mask the data.
2. Create a project and import data sources.
3. Optionally, discover information about the source data. Run profiles for data and metadata discovery to discover primary keys, entities, and data domains.
4. Define data subset operations and data masking operations. Define the tables that you want to include in the subset database and the relationships between the tables. Assign data masking rules to columns in the source data.
5. Define data generation operations. Define the tables that you want to include and assign data generation rules to columns in the target table.
6. In addition to the target connection, to directly copy flat file results to an HP ALM server, enter the test tool integration properties in the plan.
7. To store data in the test data warehouse, select the test data warehouse as the target in the plan.
8. Generate and run the workflow for data masking, data subset, or data generation.
9. Monitor the workflow.

Create a Data Masking Policy

Design policies to mask specific types of data. A policy includes the data domains that describe the data that you want to mask. A policy does not contain any data source. You can apply a policy to more than one project in Test Data Manager.

Define data domains to group sensitive fields by column name or by the column data. Define patterns in the column name or the column data using regular expressions. A data domain also contains masking rules that describe how to mask the data.

To design a data masking rule, select a built-in data masking technique in Test Data Manager. A rule is a data masking technique with specific parameters. You can create data masking rules with mapplets imported into TDM from PowerCenter.

Create a Project and Import Metadata

Create a project to organize the components for data discovery, masking, subset and generation operations.

Import data sources in the project. Create a target schema. TDM overwrites any data that already exists in the target schema. Import metadata for the sources on which you want to perform data subset or masking operations. Import target metadata to perform data generation operations. You can import metadata from a PowerCenter folder or an external database source.

When you import PowerCenter source metadata, the TDM Server sends a request to the PowerCenter Repository Service to extract source metadata from the PowerCenter repository. The PowerCenter Repository Service loads the source metadata to the TDM repository. When you import external database metadata, the TDM Server extracts metadata from the source tables and loads it into the TDM repository.

Discover Source Information

You can run profiles to discover primary and foreign key data, entity relationships, and data domains in source tables.

When a data source has no keys, you can run a primary key profile to identify possible primary keys. When the project contains multiple sources, you can run an entity profile to discover possible relationships between tables. Select the primary keys and the entities from the profile results to define the subset data structure.

You can run a data domain profile to search for columns in the source data to add to each data domain. Use data domain profile results to determine which columns to mask with the same masking rules.

When you run profiles for data discovery, the TDM Server sends a request to the Data Integration Service to extract data from the source tables. The Data Integration Service loads the profile results to the profiling warehouse. When you add constraints to tables, the TDM Server stores the constraints in the TDM repository. The TDM server does not update the data sources.

Define Data Masking and Data Subset Operations

To define data subset operations, define the tables that you want to include in the subset database and the relationships between the tables. To perform data masking operations, create a plan to run the masking operations. Add policies for masking the data. You can also add rules that are not in policies.

Perform the following tasks in Test Data Manager to define the data masking and data subset operations:

1. Create entities, groups, and templates to define the tables that you want to copy to the subset database. An entity defines a set of tables that are related based on physical or logical constraints. A group defines a set of unrelated tables. A template is an optional component that contains the entities and groups.
2. Assign data masking rules to columns in the data source.
3. Create a data subset plan, and add the entities, groups, and templates to it. For each column in a parent table, you can define criteria to filter the data.
4. Create a data masking plan and assign the policies and rules to the plan that you want to apply.

The TDM Server stores projects, entities, groups, templates, and plans in the TDM repository. When you generate and run workflows from plans, the PowerCenter Integration Service runs the workflows and loads the data into the target database.

Define a Data Generation Operation

To perform a data generation operation, create a data generation plan. Add tables and entities to the plan.

Perform the following tasks in Test Data Manager to define the data generation operation:

1. Create entities that you want to add to the generation plan.
2. Create data generation rules and assign the rules to the columns in the target table.
3. Create a data generation plan, and add the entities and tables to the plan. Assign default data generation rules to the columns that do not have rule assignments.

Create a Plan for Data Masking and Data Subset

Create a plan for the data masking and data subset operations. A plan includes the components that you need to generate a workflow. You can combine a data masking and a data subset operation in the same plan, or you can create separate plans. To save the results to an integrated HP ALM server or to store the results in the test data warehouse, select the appropriate properties in the plan.

1. Create a data subset plan and add the entities, groups, and templates to it. You can define additional criteria to filter the data.
2. Create a data masking plan and assign the policies and rules to the plan that you want to apply.
3. To store the results in the test data warehouse, select the test data warehouse from the list of target connections in the plan.
4. To also copy flat file target results to an integrated HP ALM server, enter the test tool integration properties in the plan.
5. Generate a workflow from the plan.
6. Run the workflow.

When you generate and run workflows from plans, the PowerCenter Integration Service runs the workflows and loads the data into the target database.

Create a Plan for Data Generation

Create a plan to perform data generation operations. You cannot combine a data generation operation with a data masking or a data subset operation in the same plan. You must create a separate plan for data generation. To save the results to an integrated HP ALM server or to store the results in the test data warehouse, select the appropriate properties in the plan.

Perform the following tasks when you want to create a data generation plan:

1. Create a data generation plan and add the tables and entities. You can define additional criteria to filter the data.
2. Enter the number of records that you want to generate.
3. To store the results in the test data warehouse, select the test data warehouse from the list of target connections in the plan.
4. To also copy flat file target results to an integrated HP ALM server, enter the test tool integration properties in the plan.
5. Generate a workflow from the plan.
6. Run the workflow.

Monitor the Workflow

Monitor workflow progress and monitor progress and logs of other jobs such as importing metadata and profiling in the Monitor view. Each workflow appears as a job in the Monitor view.

Access the Monitor view to determine the status of the workflow jobs. You can run the Row Count Report on a successfully run workflow to view the number of rows that a plan affects. View the workflow job status. Access the TDM job log to troubleshoot problems.

TDM Example

An organization wants to enforce a policy to mask sensitive employee stock data in a large data processing environment.

The IT department needs test data for a new employee stock plan in an organization. The organization must ensure that the sensitive data is not compromised in the test data. The test database must contain representative data from the various application environments, including employee personal data, salary data, stock purchases, and job information. Multiple test teams must be able to access the test data and replace modified test data with the original test data when required. The organization uses TDM to establish and enforce a policy for creating the data in the test environment and to store and reuse the test data in the test data warehouse.

The organization completes the following steps:

1. Create a policy. The compliance officer determines the type of employee data that should be masked. The compliance officer creates an Employee_Stock policy.
2. Define data domains. The compliance officer defines data domains to group similar fields for data masking. For example, the data contains columns called Employee_Salary, Yearly_Salary, and Salary_History. All columns that contain "Salary" in the name belong to the same data domain. All columns in the same data domain can receive the same data masking rules.
3. Define data masking rules. The compliance officer creates data masking rules to mask the employee data. For example, the compliance officer masks employee names with substitution masking from a dictionary. The compliance officer applies random masking to the salary columns. He applies Social Security masking to Social Security numbers.
4. Define a project. A project developer defines an Employee_Stock project and imports the data sources to the project. The project developer performs all the data subset, data profiling, and data masking configuration in the project.
5. Run a profile for data discovery. The project developer runs a profile for data discovery. The profile identifies sensitive columns in the source tables and it populates the data domains that the compliance officer defined in the policy.
6. Create table relationships. The database does not contain primary and foreign keys. The project developer runs a profile for primary keys and entities to find relationships between tables. The project developer examines the primary key profile results and the entity profile results to create relationships. The project developer creates logical primary and foreign keys in the tables. In some cases, the project developer selects an entity to use from the profile results.
7. Create entities and groups for data subset. With the constraints in place, the project developer can create entities in an Employee_Stock project. An entity defines a set of related source tables based on constraints. The project includes the Employee, JobHistory, Salary, and Employee_Stock tables. The project developer also creates a group in the project. A group defines unrelated tables to include in the test database. The group includes a table called Stock_History.

8. Approve or reject profile job results. The compliance officer reviews the results and approves or rejects the column assignments to the data domains.
9. Verify all sensitive fields are masked. The compliance officer reviews reports that describe what source data is masked in the project.
10. Create a plan to run data subset and data masking. The project developer creates one plan to run the data masking and subset operations in a workflow. The project developer adds the entities and groups to the plan to define which data to copy to the subset database. The project developer adds the Employee_Stock policy to the plan to define how to mask the data. When the project developer runs a workflow from the plan, the PowerCenter Integration Service runs the workflow and loads the masked data into the subset database.
11. The compliance officer validates the results in the subset database.
12. Create a plan to move the masked data subset to the test data warehouse. The project developer creates a plan with the subset database as the source connection and the test data warehouse as the target connection. When the project developer runs a workflow from the plan, the PowerCenter Integration Service runs the workflow and loads the masked data as a data set in the test data warehouse.
13. Reset a data set from the test data warehouse. The project developer runs a reset operation on the data set to restore the original test data to the required connection. When the reset operation runs, the PowerCenter Integration Service runs the workflow and loads the data set from the test data warehouse to the target connection.

CHAPTER 2

Test Data Manager

This chapter includes the following topics:

- Test Data Manager Overview, 30
- Test Data Manager User Interface, 31
- Overview View, 33
- Policies View, 35
- Data Sets View, 35
- Projects View, 36
- Monitor View, 37
- Administrator View, 37
- Expression Builder, 38
- Logging In to the Test Data Manager, 39

Test Data Manager Overview

Test Data Manager is a web-based user interface that you use to perform data discovery, data subset, data masking, and data generation operations. Open a view in Test Data Manager based on the task you need to perform.

A compliance officer uses Test Data Manager **Policies** view to create policies, data masking rules, data generation rules, and data domains. The compliance officer assigns the data masking rules to data domains.

A developer uses the **Projects** view in Test Data Manager to define a project and import the data sources to the project. The developer runs profiles to discover primary keys, entities, and data domain assignments. The developer creates entities for data subset operations and assigns rules to columns for data masking and data generation operations. The developer creates plans, generates workflows, and runs the workflows.

The developer opens the **Monitor** view to check the status of jobs that perform data masking, data subset, data generation, and other operations.

An administrator uses the **Administrator** view to create connections, configure workflow options, perform server management, and restrict user access to TDM components. An administrator can also integrate a test tool with TDM and configure the test data warehouse from the **Administrator** view.

Test Data Manager User Interface

Test Data Manager contains options to view and edit TDM components.

The following image shows a view in Test Data Manager:

The Contents panel shows an overview of the items in a view. The Details panel shows additional details for a single item in the contents panel.

Views

Access Test Data Manager views to perform tasks such as defining data masking policies or configuring projects.

Test Data Manager contains the following views:

Overview

View dashboard reports about projects in the TDM repository.

Policies

Define policies, masking rules, and generation rules that you can add to projects.

Projects

Define a project that contains source data and the data subset, data masking, data profiling, or data generation operations for the data.

Data Sets

Manage data sets that you store in the test data warehouse.

Monitor

View the status of jobs that import sources or perform data subset, data masking, data profiling, or data generation operations. Stop or abort jobs.

Administrator

Manage connections, dictionaries, test tool integration, test data warehouse, and workflow options.

Note: By default, an administrator can access the **Administrator** view of Test Data Manager. A user must have privileges to access the other views in Test Data Manager.

Search Field

Use the Search field to search for objects within the TDM repository and the test data warehouse repository. Search for objects such as projects, plans, assignments, and data sets. You cannot search for connections.

Enter the name or part of the name of the object you want to find. A search tab opens where you can filter the types of objects to include in the search. A search count displays the number of objects that match the search criteria. The search results return all objects and assignments contained within it and objects that contain it. For example, when you include projects within the search objects, the results also list objects such as policies, rules, and entities that the project contains. When you search for a rule, the results include the rule, assignments, objects that contain the rule, and dictionaries in the rule.

You can search for objects other than connections globally from the Search field. You cannot search for objects when indexing tables.

Quick Links Menu

You can quickly navigate to content tabs in Test Data Manager from the **Quick Links** menu.

From the **Quick Links** menu, you can open tabs to manage data subset, masking, and generation. You can use the **Quick Links** menu to access shortcuts to create a generation rule, masking rule, policy, data domain, or project.

The items that you can access from the **Quick Links** menu change based on user permissions.

User Preferences Menu

The **User Preferences** menu contains options that you can use to modify Test Data Manager.

Click **User Preferences** to view the **Edit Preferences** dialog box. The **Edit Preferences** dialog box contains the following tabs:

General

Displays general Test Data Manager preferences. Select **Show Start Screen** to display the **Informatica Test Data Manager** start screen when you log in to Test Data Manager.

Projects

Displays a list of projects that are visible in the dashboards in the **Overview** view. Disable projects in the **Projects** tab that you do not want to appear in the **Overview** view.

Actions Menu

Click the **Actions** menu to modify data that you select in the Content panel or to perform tasks such as importing and exporting objects.

You can choose different options in the Actions menu based on the Test Data Manager view.

Data Filter

You can filter data that appears in the **Content** and **Details** panel of Test Data Manager views.

Test Data Manager **Content** and **Details** panels might contain multiple rows of data. For example, the Policies view might show 25 data domain names. You might want to limit the data domain names to names that include "Emp." To apply the filter, type "Emp" in the filter field for the data domain name. Click the **Filter** icon. The **Policies** view shows data domains such as Employee_Domain or Emp_Number.

To reset the filter results, click the **Reset Filter** icon.

You can apply filters to different columns of data in the **Contents** panel based on Test Data Manager view.

Overview View

View dashboard reports about projects in the **Overview** view. The dashboard reports are graphs that show the distribution of data masking, data subset, and data generation objects in projects. You can use links in the **Overview** view to access project and the policy objects.

The following image shows the **Overview** view:

Dashboards

The **Overview** view shows the TDM dashboards. The dashboards are summary reports about projects, data masking objects, data generation objects, and administration objects.

You can view the distribution of data by sensitivity level in projects. You can view the distribution of data masking data domains, policies, and rules in projects. View the distribution of entities, groups, and templates in data subset projects.

You can view the number of data masking policies, the number of data domains, and the number of rules in the TDM repository.

Project Risk Analysis

The **Project Risk Analysis** dashboard shows the percentage of columns that are in each domain sensitivity level. It also shows the percentage of columns that do not belong to domains.

When you create a data domain, you select the sensitivity level for all columns in the domain. You can select a sensitivity level from the levels configured by the administrator.

The **Project Risk Analysis** dashboard shows the percentage of columns that belong to domains at each sensitivity level. You can view the projects that contain the columns. Move the pointer over the section of the pie chart that you want to view. The dashboard shows the number of domains in that sensitivity level, and the number of projects that contains columns in the domains. Click the **Number of Projects** link to view a list of the projects.

Project Assignments and Project Objects

The Project Assignments and Project Objects dashboard shows graphs of the number of objects and assignments in up to 10 recently accessed projects.

The Project Assignment tab displays by default. This graph displays the number of assignments, including data domains, policies, masking rules, and generation rules, in the last 10 accessed projects. The Project Objects tab displays the number of subset objects, including entities, groups, and templates, in up to 10 recently accessed projects.

Recent Project Sensitive Data Distribution

The **Recent Project Sensitive Data Distribution** dashboard shows a graph of the sensitive domain information in the last 10 accessed projects.

You must have set data domain sensitivity levels to view the distribution information. If you do not set data domain sensitivity levels, no graph appears.

Recent Activities

The **Recent Activities** panel shows the last nine Test Data Manager components that changed.

The **Recent Activities** panel shows the name of the TDM component that changed, the type of component, and who made the change. The change actions are created, updated, and deleted. The panel shows the date of the change.

Plan Execution

The **Plan Execution** dashboard displays plan execution data for a particular date.

You can select the plan component, masking, subset, or generation, for which to view data. You can view data for plans in progress, completed, or both. The default display is for the current date, all plan components, and both completed and in progress status. You can browse the calendar to choose a different date, or enter the specific date. Use the arrows to move the date forward or backward.

Global Components

The **Global Components** dashboard shows the proportion of policies, masking rules, generation rules, and data domains, out of the total number of components available in the repository.

Use the **Global Components** dashboard to view the distribution of policies, data domains, masking, and generation rule assignments in the repository in a pie chart. Masking rules and generation rules are displayed separately.

Recent Projects

The **Recent Projects** dashboard lists up to the last 10 projects you modified. You cannot view information on projects created or modified by other users.

The **Recent Projects** dashboard shows the name of the modified project, the project description, the user name, and the date of modification. Click on the project name to open the project.

Policies View

Maintain policies, the data domains, and the data masking and generation rules in the **Policies** view. Create a policy and add the data domains to the policy. You can apply the policy to multiple data masking projects in Test Data Manager.

The **Policies** view lists the policies, the data domains, and the rules in the TDM repository. Select a policy, a data domain, or a rule to change it. Or, choose to add a new policy, data domain, or rule, from the **Actions** menu.

The following image shows the **Policies** view:

Data Sets View

Manage multiple versions of data sets from the **Data Sets** view.

All machines that host a Test Data Manager Service in the domain can connect to a single test data warehouse. The **Data Sets** view lists all data sets stored in the test data warehouse. Select a version of a data set from the list to view the general properties in the **Overview** tab. The **Tables** tab lists all tables in the data set that you select. The **Components** tab lists the components in the data set.

Use the **Data Sets** list filter to filter and search for a data set. You can reset a specific version of a data set to a target, delete a data set, or lock and unlock a data set from the **Actions** menu.

The following image shows the **Data Sets** view:

Projects View

Maintain projects in the **Projects** view. A project is the top-level container that you use to organize the components for data discovery, data masking, data subset, and data generation operations.

From the **Projects** view, you can create a project and you can import source data for the project. Assign policies and rules to columns in the source data. Run profiles for data discovery against the data. Run data masking and data subset operations on the source data. Run data generation operations to generate data.

The **Projects** view shows a list of projects. You can select a project to view the data sources and recent activities.

When you select a project, you can update it. You can define profiles for data discovery, data masking, and data subset operations. You can generate a PowerCenter workflow from within a project.

The following image shows the **Projects** view:

Monitor View

In the **Monitor** view, you can review the status of jobs that perform tasks such as import data, run profiling, or run data masking, data subset, and data generation operations. You can review the TDM Server job log to investigate problems. You can also stop jobs and recover workflows in the **Monitor** view.

The following image shows the **Monitor** view:

Select a job in the **Monitor** view and choose options from the **Actions** menu.

Administrator View

An administrator can set default settings, create connections, add and manage dictionaries, manage the test data warehouse settings, and perform TDM server tasks in the **Administrator** view.

An administrator can perform the following tasks in the **Administrator** view:

- Configure default settings including generation rule settings and child and parent record settings.
- Configure data domain sensitivity levels for tracking the sensitive data that users need to mask.
- Add and modify default project configuration fields.
- Configure workflow properties and set default staging and storage connections and create storage tables.
- Create connections to source or target databases.
- Add and manage dictionaries.
- Configure test data warehouse settings.
- View and search application level logs.
- Set or modify the log levels to debug the logs.
- Integrate a test tool with TDM.

The following image shows the **Administrator** view:

Expression Builder

Use the **Expression Builder** to build an expression using PowerCenter functions.

Click **Expression** to open the **Expression Builder** when you perform any of the following tasks:

- Create masking rules with Expression masking type.
- Add pre-processing or post-processing expressions to a standard masking rule.
- Create advanced masking rules.
- Assign rules to a column.
- Configure expressions in an ad hoc generation rule.
- Configure advanced expressions in plans.

The **Expression Builder** shows a list of PowerCenter functions by category. To view the complete list, expand the **All Functions** list.

You can choose from a list of available columns to include in an expression. Select a function and click the **Add** arrow to add the function to an expression. You can choose columns and operators to enter join conditions.

For example, you might create an advanced masking rule that contains a first name and a last name input port. The masking rule has two variable ports that receive masked values for the first name and the last name. You might create an expression for the output port that concatenates the masked names in the variable ports and includes a space between them:

```
CONCAT( CONCAT( VAR1, ' ' ), VAR2 )
```

The following image shows the **Expression Builder** where you configure expressions:

For more information about the function syntax, see the *Informatica Transformation Language Reference*.

Logging In to the Test Data Manager

To access the Test Data Manager, enter the host name and port number of the TDM Server in a web browser.

To log in, enter a user name and password defined in Informatica Administrator.

1. In the address bar of a web browser, enter the Test Data Manager URL.
   - Use the following format if Transport Layer Security is enabled:

     `https://hostname:portnumber/tdm/`
   - Use the following format if Transport Layer Security is not enabled:

     `http://hostname:portnumber/tdm/`

   Where:
   - *hostname* is the host name or IP address of the machine where you installed the TDM Server.
   - *portnumber* is the port number. The default is 6643 if Transport Layer Security is enabled. The default is 6605 if Transport Layer Security is not enabled.

   For example, you might enter the following URL:

   `http://TXW1779:6643/tdm/`

   The **Login** dialog box of the Test Data Manager appears.
2. Enter the user name and password.

Select the security domain. If the Informatica domain is configured to use LDAP authentication, the default security domain **Native**.

3. Click **Login**.

   The Test Data Manager appears.

To log out of the Test Data Manager, click **Logout**.

CHAPTER 3

Projects

This chapter includes the following topics:

- Projects Overview, 41
- Project Components, 42
- Project Logs, 43
- Project Management, 43
- Data Sources, 47
- Project Permission and Security, 51

Projects Overview

A project is the top-level container that you use to organize the components for data discovery, masking, subset, and generation operations.

The following image shows a project in Test Data Manager:

A project contains the following views:

Overview

Edit the project general properties in the **Properties** view. Add policies and rules to the project in the **Policies** view. In the **Data Sources** view, import data sources to a project for data discovery, data subset, data masking, and data generation operations. You can import a source from the PowerCenter

repository, or you can import a source from a database. You can import multiple types of sources to the same project and define relationships between them.

Discover

Discover the primary keys, the table relationships, and the data domains in the source data.

Define

Define data masking, data subset, and data generation operations. When you define a data masking operation, assign the rules and policies to columns in the project source. When you need to create a data subset, define entities, groups, and templates. When you need to perform a data generation, assign the generation rules to the columns in the tables of the project.

Execute

Define a data subset, data masking, or data generation plan. Generate and run a PowerCenter workflow from the plan.

Monitor

View the status of data source import jobs, profiling jobs, workflow generation jobs, data subset jobs, data masking jobs, and data generation jobs.

Permissions

Apply the user group and the user permission to projects. When you create a project, you are the project owner and have access to the project. If other users need access to the project, you can add the users in the **Permissions** view. Apart from permission on a specific project, users need the minimum required privileges to access or perform any tasks on the project.

Project Components

A project contains one or more data sources. Other components that you add to a project depend on the operations that you need to perform on the data.

The following table describes the components that you can create in a project:

Component	Operation	Description
Data Source	Data subset Data discovery Data masking Data generation	Defines the input data that you want to transform.
Entity	Data subset Data generation	Defines a set of related tables based on physical or logical constraints. Define an entity for a data subset or a data generation operation. Create an entity in the **Define** view of a project.
Group	Data subset	Defines a set of unrelated tables to copy to a target subset database with data subset. Create a Group in the **Define** view of a project.
Plan	Data masking Data subset Data generation	Defines data subset, data masking, or data generation operations. Generate a PowerCenter workflow from a plan. Define a plan in the **Execute** view of a project.

Component	Operation	Description
Policy assignments	Data masking	Assigns policies to the project. When you assign a policy to a project, the project receives the data domains and the data masking rules in the policy.
Profile	Data discovery	Suggests the primary keys, relationships, and data domains in a source based on source data and metadata. Define a profile in the **Discover** view of a project.
Rule assignments	Data masking Data generation	Defines which data masking or generation rules apply to a source column.
Template	Data subset	A reusable component that contains the entities and groups that you want to subset based on common parameters. Create a template in the **Define** view of a project.

Project Logs

PowerCenter generates workflow and session logs when you run a plan or workflow from within a project.

You can configure the location where you want to save the log files for a project. If you do not specify a location, PowerCenter saves the log files in the default location set in the PowerCenter Integration Service properties.

The PowerCenter Integration Service stores the logs and reject files in the following default locations:

- Workflow log directory: `$PMWorkflowLogDir\`
- Session log directory: `$PMSessionLogDir\`
- Reject file directory: `$PMBadFileDir\`

You can configure the location where you want to save workflow logs, session logs, and reject files from the **Edit Project** dialog box. You can specify the location in an existing project. You cannot specify the location when you create the project.

You can configure workflow log properties and session log properties when you create a plan. For example, you can choose to save session logs or workflow logs by the time stamp, the number of runs, or the maximum size of the log files.

You can specify the type of error log files to generate error logs and the location to store error logs for a specific plan when you create the plan.

Project Management

When you create a project, you add one or more sources to the project. When you edit a project, you can update the name, description, folder location, and associated sources. You can also edit the location where you want to store the log files of the project.

When you create a project, you are the project owner. You have access to the project folder by default.

You can export a project and project components to an XML file. You can import the data from the XML file into Test Data Manager. You might export and import a project to back it up, or you might need to move a project from a development to a production system.

Creating a Project

Create a project folder to contain the data discovery, subset, masking, and generation components to apply to a data source.

1. In Test Data Manager, click **Projects** to access the projects.

   A list of projects appears.
2. Click **Actions** > **New**.
3. In the **Create Project** dialog box, enter project properties. The following table describes the project properties:

Option	Description
Name	The name of the project.
Description	The description of the project.
PowerCenter Repository	The name of the PowerCenter repository to store the repository folder.
Folder	The name of the project folder in the repository. Default is the project name. You can choose another folder in the repository.
Owner	The name of the user that owns the folder. The folder owner has all permissions on the folder. The default is the name of the user that created the folder. You can select another user as the folder owner.

4. Click **OK**.

   The properties for the project appear in Test Data Manager.

Editing a Project

You can edit a project to change its name, description, or associated sources.

1. To access the projects, click **Projects**.

   A list of projects appears.
2. Click on the project that you want to edit.

   The project opens in a separate tab with project properties and data sources details.
3. To change the project properties, click **Edit**.

   You cannot remove a source that is used by entities or groups in the project.

4. In the **Edit** dialog box, change project options. The following table describes the project options that you can change:

Option	Description
Name	The name of the project.
Description	The description of the project.
Folder	The name of the project folder in the repository. Default is the project name. You can choose another folder in the repository.
Owner	The name of the user that owns the folder. The folder owner has all permissions on the folder. The default is the name of the user that created the folder. You can select another user as the folder owner.

5. Click **Save**.

Copying a Project

You can create a project by copying a project. When you copy a project, Test Data Manager creates a copy of the project but does not import the metadata or objects of the project.

1. To access the **Projects** view, click **Projects**.
2. Click a project description to select the project.

   Do not open the project.
3. Click **Actions** > **Duplicate**.

   The **Copy <Project Name>** dialog box appears.
4. Change the name and description of the project. Click **Save**.

Deleting a Project

When you delete a project, all sources that are associated with the project are deleted.

1. To access the projects, click **Projects**.

   The list of projects appears.
2. To select a project without opening it, click the Description or Created On column of the project.
3. Click **Actions** > **Delete**.
4. In the **Confirm Delete** dialog box, click **Yes**.

Configuring Project Log Locations

You can configure the location where you want to save the log files that the PowerCenter Integration Service generates for project workflows and sessions.

1. Open the project and click **Actions** > **Edit**.

   The **Edit Project** dialog box opens.
2. Click the **Default Log File** tab.

3. Perform the following tasks to change the location of the log files:
   - In the **Workflow Log Directory** field, enter the location where you want to store workflow logs.
     If you enter a wrong location, a workflow fails to run without generating an error log.
   - In the **Session Log Directory** field, enter the location where you want to store session logs.
   - In the **Reject File Directory** field, enter the location where you want to store reject files.

   You can use parameters or enter a fully defined path. You must add a delimiter with a slash or backslash at the end of the file path.
   For example, you can enter the workflow log directory in the following ways:
   - `$PMWorkflowLogDir\`
   - `$PMWorkflowLogDir\projectlogs\`
   - `\TDM\Logs\Project\`

   Note: The directory path must exist. TDM does not create the directory.
4. Click **Save**.

Exporting a Project

You can export a project to an XML file and import the XML file to another TDM repository. When you export a project, the XML file contains all the project components, such as the source information, the connections, the constraints, the data domains, and the assignments.

1. To access the projects, click **Projects**.

   A list of projects appears.
2. Click a project to view the project properties and the data sources.
3. Click **Actions** > **Export**.
4. Choose the name and path of the XML file to create.
5. Click **OK**.

Importing a Project

You can import a project from an XML file that was exported from another TDM repository.

Before you import a project into Test Data Manager, you must import all the global components that are assigned in the project.

1. To access the projects, click **Projects**.
2. Click **Actions** > **Import**.
3. Browse for the XML file that contains the project to import. Click **Finish**.

   Test Data Manager imports the project.

Data Sources

To perform data subset, masking, and discovery operations, you must import source metadata into the TDM repository. You can import sources from the PowerCenter repository or from a source database. To perform data generation operations, you must import target metadata into the TDM repository.

When you create a project, add one or more sources to the project. You can add more than one type of source to the project. For example, you can add a flat file source and a relational source to the project. You can create constraints to create relationships between the sources and apply filter criteria for data subset and data masking.

Note: If the source data has a numeric column with precision greater than 28, you cannot apply a data masking or data subset condition to that column. The PowerCenter Integration Service passes the row and the column to the target without applying the condition.

TDM Source Definitions

Import source metadata from PowerCenter or directly from the source database to the TDM repository. The TDM project and PowerCenter folders share the same name to ensure that the PowerCenter Integration Service stores workflow information in the correct PowerCenter folder.

You can import source metadata directly from the source if the source is a relational database. If the source contains many tables, you can increase the performance if you import the metadata directly from the source instead of from PowerCenter.

When you import source definitions from the PowerCenter repository into the TDM repository, you can import all the definitions supported by PowerCenter. When you import source metadata from PowerCenter, you import the PowerCenter folder that contains source metadata. Test Data Manager creates a copy of the folder in the TDM repository.

When you import the MongoDB sources to PowerCenter, you must ensure that the column names are unique and does not contain more than 80 characters in a column name.

To use flat file sources, you must import flat files from the PowerCenter repository. You cannot create flat file connections in Test Data Manager.

Note: Rowid is a reserved keyword. You cannot import Oracle sources that have rowid as a column name or a table name.

XSD Data Sources

You can perform data subset, data masking, and data generation operations on XML sources.

Import XML source metadata from the PowerCenter repository. To import an XML source, you import the source definition from the PowerCenter repository. A workflow that you generate and run fails if you import the XML instead of the XML source definition (XSD).

You can view the list of XSD files in a project from the **Files** tab in the **Discover** view of the project. Highlight a file to view the properties in the tab below. You can open and view the properties of each file in a separate page.

The following image shows the file page in Test Data Manager:

The file page contains the following tabs:

Properties

Displays the name of the XSD file, project information, and a count of the objects with data masking, data generation, or data subset criteria assignments.

View

You can view the elements in a file as a map or in a grid. View the elements in the file and assign a data domain. To apply a masking rule to multiple elements, you can mark elements as similar value objects. You cannot perform tasks on all elements in the map or grid view. You can choose to view select objects from a list of options in the **Map View**. You can filter and search for elements and attributes based on general properties in the **Grid View**.

Details

Apply subset criteria, data masking rules, or data generation rules from the **Details** tab on the file page.

The **Details** tab has the following tabs:

General

View the properties of an element or attribute. You can add or edit a data domain assignment, and you can mark an attribute or element as restricted or sensitive. When you run a workflow, the PowerCenter Integration Service ignores elements and attributes that you mark as restricted.

Data Masking

Assign a policy and data masking rule or directly assign a data masking rule to an element or attribute.

Data Generation

Assign a data generation rule to an element or attribute, or create and assign a data generation rule.

Criteria

Specify the data subset criteria to apply to an element or attribute.

Editing the Properties of an Element in an XSD File

You can view the general properties of an element or attribute that you select in the map view or grid view. You can edit the properties that are not read-only.

1. Open the project and click the **Files** tab in the **Discover** view.

   A list of XSD files in the project appears.
2. Select the required file to open the file in a separate page.
3. Select the element or attribute that you want to edit.

   The properties of the element or attribute appear on the **Details** tab.
4. To open the **Edit Properties** window, click **Edit** on the **General Properties** tab.

   The **Edit Properties** window lists the following properties:

Property	Description
Name	Read-only. Name of the element or attribute.
XPath	Read-only. The path expression of the element or attribute in the XSD file.
Data Type	Read-only. The data type of the element or attribute.
Precision	Read-only. The precision for the element or attribute. The maximum number of digits or the maximum number of characters that the element or attribute can accommodate.
Data Domain	The data domain assigned to the element or attribute. Select a data domain from the list.
Restricted	The PowerCenter Integration Service does not write the value of a restricted element or attribute to a target. Select this option to mark the element or attribute as restricted.
Sensitive	Indicates whether an element or attribute contains sensitive data. Select this option to mark the element or attribute as sensitive.

5. You can edit the following properties as required:
   - To assign a data domain to the element or attribute, select a data domain from the list of data domains.
   - To mark the attribute or element as restricted, select the **Restricted** check box. During TDM operations the PowerCenter Integration Service ignores elements and attributes that you mark as restricted.
   - To mark the element or attribute as sensitive, select the **Sensitive** check box.
6. Click **OK** to save the changes.

XSD Import Guidelines

Read the following information before you import an XSD file into the PowerCenter repository from PowerCenter Designer:

- To open the **Change XMLViews Creation and Naming Options** window, click **Advanced Options** in the **Import XML Definition** dialog box. Set the following options for the import:
  - Choose to override all infinite lengths with a low value. The value is the default length if the XSD file does not specify a length. You do not require a high value if the elements are of a small length. A high length value requires more memory and affects the performance. Select the option and enter an override value.
  - If the data contains elements that require a higher precision, manually set the precision of the element to the required value after you import the file. Open the file in PowerCenter Designer and select the element in the **Schema Components** panel. Edit the length on the **Actions** tab.
  - Choose to analyze elements and attributes in standalone XML as global declarations. Required in TDM to ensure correct mapping of the namespaces of the different elements and attributes. Import of the XSD in TDM fails if the namespaces of elements and attributes are not mapped correctly.
  - Use the default settings for the other options.
  - Enter a value for the **Default length for anyType element mapped to string** option.
- In the **XML Definition Creation Options** page of the XML Wizard, select **Hierarchy Relationships** and the **Normalized XML Views** option. Use this option to create a root and expand the XML components under the root. When you choose normalized XML views, every element or attribute appears once. One-to-many relationships become separate XML views with keys to relate the views. If you create denormalized XML views, all elements and attributes display in one hierarchical group.

PowerExchange Nonrelational Sources

You can perform TDM operations on nonrelational sources that you define in PowerExchange.

You first need to define datamaps in PowerExchange and import them into PowerCenter. Create the policies, projects, and operations in TDM. Import the PowerCenter sources into a project in TDM.

You can perform TDM operations on the following PowerExchange nonrelational sources:

- Adabas
- IMS unload
- Sequential
- Hierarchical and nonhierarchical VSAM

Note: You cannot import a PowerExchange source directly into the TDM repository.

Importing Data Sources

Import data sources from PowerCenter or from an external database to a project in Test Data Manager.

1. To view a list of the projects in Test Data Manager, click **Projects**.
2. Click a project in the list to open it.
3. Click **Actions** > **Import Metadata**.

   The **Import Metadata** window appears.
4. Choose one of the following options:

- PowerCenter Repository. Import metadata from PowerCenter.
- Datasource Connection. Import metadata from a database connection.

5. Choose whether you want to review the metadata changes before you import the data sources. Test Data Manager displays the information before the import runs. You can choose to skip the import option.

   Test Data Manager shows you the metadata you are importing and the impact on rules and domain assignments or entities and groups in Test Data Manager.
6. Click **Next**.
7. Choose one of the following options:
   - If you selected **PowerCenter Repository**, select the PowerCenter folder that contains the data source you want to import. You can filter folders by the folder name or the description.
   - If you selected **Datasource Connection**, select the schema to import. You can filter schemas by schema name.
8. Click **Next**.
9. Select the files or tables that you want to import. You can filter the tables by data source, table name, or table description.
10. Click **Next**.
11. Choose when to import the sources. Choose one of the following options:
    - Import Now. Import the data source immediately. To run the import in the background, select **Run import in the background**.
    - Schedule Later. Schedule the import to occur at a specific date and time. Click the calendar to select a date. Use the hour and minute sliders to set the time.
12. Click **Finish**.

    If the import is running, view the progress of the import job in the **Monitor** view. After the job finishes, access the imported metadata through the **Data Sources** details view.

Identity Columns in TDM

Read the following information before you perform TDM operations on identity columns:

- You cannot move data from a source DBO schema to a target non-DBO schema on a Sybase database if the data contains an identity column and indexes.
- If you enable the auto identity option in the database and there are no primary keys or unique keys in a table, Sybase adds the SYB_IDENTITY_COL column. You must mark the column as restricted. A workflow fails if you do not mark the SYB_IDENTITY_COL column as restricted.
- To disable an identity column on a Microsoft SQL Server database, the database user must have execute permission on sp_help.

Project Permission and Security

Use project permissions to control access to projects.

When you create a project, you become the owner of the project. As the project owner, you can add users and user groups and assign the required levels of permission. The domain administrator can also add and edit project permissions.

You can access projects based on the permissions that you have. For example, if you do not have any permissions for a project, it does not appear in the list of projects. You cannot view projects that appear in search results if you do not have the required permissions.

Projects have the following levels of permission:

- Read
- Write
- Execute

Project Permissions

You must have the required project permissions to access and perform tasks in a project.

The following table lists the project permission levels, the tasks that you can perform with each level, and the minimum required privileges for each task:

Permission	Description	Minimum Required Privilege
Read	- Open and view the project. - Monitor logs for the project workflows.	- View project - Monitor project - Audit project
Write	- Open and view the project. - Monitor logs for the project workflows. - Import metadata. - Create entities, groups, and templates. - Assign rules. - Generate workflows. - Run profiles. - Copy the project. - Delete the project.	- View project - Monitor project - Audit project - Import metadata - Generate project - Manage project - Discover project
Execute	- Open and view the project. - Monitor logs for the project workflows. - Run workflows.	- View project - Monitor project - Audit project - Execute project

Updating User and Group Security

When you create a project, you can assign read, write, and execute permissions to users and user groups. Edit the project permissions assigned to users and user groups from the **Permissions** tab of the project. Changes to permissions take effect from the subsequent login.

1. Open a project and click **Permissions**.

   A list of the users and user groups with permissions for the project appears.

2. Click **Edit** on the **Users** or **User Groups** tab.

   The **Edit Project Permissions** dialog box opens.

3. To edit the permission of a user or user group, select the user or user group from the list and edit the permission as required. You must save the changes for each user or user group.
4. To delete a user or user group, select the user or user group from the list and click **Delete**.

5. To add a user or a user group:
   a. Click **Add Users** or **Add User Groups**.
   b. Select one or more users or user groups.
   c. Optional. From the list of permissions, select the required permissions if either of the following statements is true:
      - You selected a single user or user group.
      - You want to assign the same levels of permission to all selected users or user groups.
   d. Click **OK**. TDM adds the users or user groups to the list.
   e. Select each user or user group and assign the required permission levels. You must save the changes for each user or user group. Skip this step if you performed step c.
6. Click **OK**.

CHAPTER 4

Policies

This chapter includes the following topics:

- Policies Overview, 54
- Policies View, 55
- Policies Task Flow, 55
- Data Masking Rules, 55
- Data Domains, 59
- Policy Packs, 64
- Import and Export, 67
- Linking Business Glossary Terms to Global Objects , 68
- Policy Management, 69

Policies Overview

A policy is a data masking component that describes the methods to maintain the privacy of specific types of source data.

A policy contains data domains. A data domain describes the functional meaning of a column based on the column data or the column name. For example, a Social_Security data domain contains all database columns with numbers in the following format: 999-99-9999. A Salary data domain might include the Salary, Bonus, and Commission columns in a database.

A data domain contains data masking rules. A data masking rule is a data masking technique to mask a specific type of data. For example, you might configure the Substitution data masking technique for First Names and Last Name columns. You configure two Substitution masking rules because each rule contains different parameters.

You can configure data domains, rules, and policies separately. Apply the rules to data domains and add the data domains to a policy. After you define the policy, you can assign the policy to a data source in a project. You can apply a policy to multiple projects.

Policies View

The **Policies** view shows the policies, data domains, and rules in the TDM repository.

The following image shows the **Policies** view in Test Data Manager:

Policies Task Flow

You can create policies, rules, and data domains in any order.

Complete the following high-level steps to define policies:

- Create data masking rules.
- Define data domains to describe which columns can receive the same masking rules.
- Assign the data masking rules to the data domains.
- Create a policy.
- Assign the data domains to the policy.

Data Masking Rules

A data masking rule is a data masking technique to mask a specific type of data. You can create a standard rule, advanced rule, or a rule that you import as a mapplet.

A data masking technique defines the logic that masks the data. Masking parameters are options that you configure for a masking technique. For example, you can define different dictionary files for substitution masking rules. Dictionary files contain the sample data for substitution. You might blur output results by different percentages for different columns. Most masking techniques have associated masking parameters.

You can enable users to override masking parameters for a rule. For example, you create a rule with the substitution masking technique to mask column data based on a flat file substitution source. You set the

override option for the rule. When a developer assigns this rule to columns in a source, the developer can select a relational database as a substitution source rather than a flat file.

You can assign rules to source columns, data domains, policies, and plans.

Standard Masking Rules

A standard masking rule is a data masking rule that applies a built-in masking technique. A standard masking rule has one input column and one output column.

When you create a standard masking rule, you select the masking technique from a list. You can define one masking technique in a standard masking rule and you can apply the rule to one column.

Test Data Manager has masking techniques that you can select to create masking rules. You can use standard masking techniques based on the source datatype and masking type that you configure for a column. You can restrict the characters in a string to replace and the characters to apply in the mask. When you mask numbers and dates, you can provide a range of numbers for the masked data. You can configure a range that is a fixed or percentage variance from the original number.

Creating a Standard Masking Rule

Create a rule to define a masking technique, the datatype to mask, and masking parameters that define how to apply the technique.

1. To access the **Policies** view, click **Policies**.
2. Click **Actions** > **New** > **Masking Rule**.

   The **Rule Wizard** appears.
3. Enter a name and optional description for the rule.
4. Select the datatype of the column to apply the masking rule to.
5. Select the Standard masking rule.
6. To enable users to override masking parameters for a rule, select the **Override Allowed** option.
7. Click **Next**.

   Note: The **Masking Parameters** dialog box changes based on the **Masking Technique** you select.
8. Enter the masking parameters.
9. Enter the exception handling options. Configure how to handle null or empty spaces. Configure whether to continue processing on error.
10. Click **Finish**.

Rule Simulation

You can simulate the output of a standard rule to preview the output before you assign the rule to a column.

Use the Rule Simulator to view the output of a standard rule before you assign it to a column or add it to a plan. View the output of the rule and change the rule properties if required before assigning it to a column. You can choose to include data from a connection in the simulation. Alternatively, you can use default sample data or enter up to 100 rows of sample data on which to simulate the rule output. View the original data values and the masked values in the Rule Simulator tab.

The latest simulation configuration details are stored in the browser cache. You can edit the properties of a rule after viewing the simulation results, and run the simulation again on the same data with the updated rule. Clearing the cache deletes the configuration information.

Previewing Rule Output

Use the Rule Simulator to view the output of a standard masking rule on selected data. You can use data from a connection, use default sample data, or enter sample data on which to view the rule output.

1. To access the **Policies** view, click **Policies**.
2. Click a masking rule name to open the Rule Properties page of the masking rule.
3. Click **Rule Simulator** to open the Rule Simulator configuration tab.
4. Select the type of configuration from the Source Details list and click **Edit**. Select Default to use default sample data, My Test Data to enter sample data, or Connection to use data from a connection. The **Configure Source Details** page opens.
5. On the **Configure Source Details** page, to use source data from a connection:
   a. Select the connection and owner from the lists.
   b. To select a table, type the table name, or click **Browse** to select the table from a list.
   c. Enter the number of rows to include in the simulation. The default value is 20. The maximum number of rows you can include is 100.
   d. Click **OK** to return to the Rule Simulator tab.
6. To enter sample data:
   a. Enter the data in the data fields in the **Configure Source Details** window. Use the buttons to add or delete rows. You can enter a maximum of 100 rows.
   b. Click **OK** to return to the Rule Simulator tab.
7. To use default sample data:
   a. Click **Copy default data**.
   b. Click **OK** to return to the Rule Simulator tab.
8. On the **Rule Simulator** tab, click **Go** to start the simulation.

   View the original source values and the masked values in the **Rule Simulator** tab.

Advanced Masking Rules

An advanced masking rule is a combination of masking techniques that mask multiple source columns or a target column based on values of more than one input column.

For example, you can create a full masked name by masking the first name and last name input columns. Define variable columns to contain the masked names. Add an output column that contains a result of an expression that combines the first name and last name variable columns.

Create the following types of columns in an advanced rule:

Input

The source column that you want to mask.

Variable

A column that contains intermediate values in a calculation. The variable column receives a value from an expression or a masking technique. You can configure multiple variable columns in order to combine multiple masking techniques.

Output

The target column that receives the masked value. The output column type contains a masking technique and masking parameters.

Related Topics:

- "Advanced Masking" on page 135

Creating an Advanced Masking Rule

Create an advanced masking rule to combine more than one masking technique or to mask multiple columns.

1. To access the **Policies** view, click **Policies**.
2. Click **Actions** > **New** > **Masking Rule**.

   The **Rule Wizard** appears.
3. Enter a name and optional description for the rule.
4. Select the Advanced masking rule.
5. Click **Next**.
6. In the **Input Columns** section, click **Add New**.

   The **Add Column** dialog box appears.
7. Enter the column properties, such as the name, datatype, precision, and scale. Select whether the column is mandatory to assign to a data source in all projects.

   You cannot enter masking properties for input columns.
8. Click **OK**.
9. To enter more input columns, click **Create Input Column** in the **New Masking Rule** dialog box.
10. In the **Variable Columns** section, click **Add New**.
11. Enter an expression, a masking rule, or a dependent column for each variable column that you define. If you apply a masking rule, configure the input column to create the variable column from.
12. Click **OK**.
13. In the **Output Columns** section, click **Add New**.
14. Enter an expression, a masking rule, or a dependent column for the output column.
15. Click **OK**.

Mapplet Rules

You can create rules from a mapplet. The mapplet contains the logic to mask the input columns and return data to the target columns. When you create a rule from a mapplet, you assign the mapplet column names to input and output columns when you assign the rule to a column in the data source.

Import a mapplet from an XML file that you exported from the Model repository or the PowerCenter repository. The mapplet can contain any passive transformations.

To perform TDM masking operations, you can import PowerCenter mapplets.

A mapplet can contain multiple input and multiple output columns. All columns might not be available in all projects. You must configure one input column and one output column as required columns. The mandatory columns must have source and target assignments when you assign the rule to a column in the data source. Test Data Manager has an interface to assign multiple columns to a rule from a mapplet.

The TDM repository stores the mapplet logic when you import the mapplet. You cannot change the mapplet in Test Data Manager.

Creating a Mapplet Masking Rule

You can create a data masking rule from a mapplet. The mapplet contains the logic to mask the source fields.

Export the mapplet to an XML file from the PowerCenter repository or the Model repository before you import it to Test Data Manager.

1. To access the **Policies** view, click **Policies**.
2. Click **Actions** > **New** > **Masking Rule**.

   The **Rule Wizard** appears.
3. Enter a name and optional description for the rule.
4. Choose the data type to mask.
5. Select the Mapplet masking type.
6. Browse for the XML file that contains the PowerCenter mapplet to import.
7. Click **Next**.
8. Select at least one input column and output column as mandatory columns.

   Select the column and click the mandatory column to change the value from `No` to `Yes`.
9. Click **Finish**.

Data Domains

A data domain is an object that represents the functional meaning of a column based on the column data or the column name. Configure data domains to group data source columns for data masking. You can assign a masking rule to a data domain and all the columns in the data domain are masked with the same rule. You can add generation rules to a data domain so that TDM generates data with the same generation rule.

Create data domains to describe the columns you need to mask with the same masking rules. Assign at least one masking rule to each data domain.

For example, you might need to mask all the instances of Social Security number with the same masking rule. You can create a data domain that describes the Social Security data that occurs in the different columns. A database might have a Social Security number in a column called SSN. The database also has a column called SOCIAL_SECURITY in a different table. A Social Security number might also appear in a COMMENTS column.

When you create the data domain, you create a data expression that describes the data format for Social Security numbers. A Social Security number has this format: `999-99-9999`. You can also create multiple metadata expressions that describe possible column names for Social Security numbers. Social Security column names might include `SSN` or `Social`.

You can add data generation rules to a data domain. TDM lists the preferred data generation rules for a data domain. You can edit the list or add another generation rule.

After you define a data domain, you can add the data domain to a policy. You can run profiles for data discovery against data sources in a project. Run profiles to find the columns for data domains. For example, the profile job can find all the Social Security numbers in the source data based on how you defined the data domain. The profile assigns data domains to columns.

Note: If you do not have Data Discovery, you can still use data domains to aggregate data. However, you must manually associate source columns with the data domains.

Apply Masking Rules to a Data Domain

You can assign one or more data masking rules to the data domain. When you assign a masking rule to a data domain, the columns in the domain receive the data masking rule when you configure data masking.

When you assign data masking rules to the data domain, the rules are called preferred rules. If you assign multiple rules to the data domain, you enable one of the rules to be the default rule. The default rule is applied to all columns in the data domain. You can manually change the masking rule for a column to a different preferred rule. You can also apply more than one masking rule to a column.

For example, an organization has a data domain called Last_Name. The Last_Name data domain describes columns that contain last names in company databases. The company can use a shuffle masking rule to mask the last names of customers in a database. The shuffle masking rule is the default rule. The organization applies a substitution masking technique to mask the last names of customers in a different table. The substitution masking rule is a different preferred masking rule in the data domain.

Apply Generation Rules to a Data Domain

You can assign one or more data generation rules to a data domain. When you assign a generation rule to a data domain, the columns in the data domain receive the data generation rule when you configure data generation.

TDM lists the preferred data generation rules for a data domain. You can edit the list or add another generation rule.

When you add data generation rules to a data domain, you can enable one of the rules to be the default rule. The default rule is applicable to all columns in the data domain. To change a default rule, you must edit the data domain and enable the default rule that you want. You can also apply more than one generation rule to a data domain.

For example, an organization has a data domain called Date_of_Birth. The Date_of_Birth data domain describes the columns that contain birth dates of the customers. The organization can use the date random generation rule to generate birth dates. The date random generation rule is the default rule. The organization applies a date sequence generation technique to generate the birth dates of customers in a different table. The date sequence generation rule is a different preferred generation rule in the data domain.

Metadata and Data Patterns for Data Domains

A data pattern and a metadata pattern are regular expressions that you configure to group columns into a data domain. Use regular expressions to find sensitive data such as IDs, telephone numbers, postal codes, and Social Security numbers in the source data.

A regular expression is a text string that describes a search pattern. A regular expression provides a way to match strings of text or patterns of characters in the source data.

A data domain expression can contain data expressions and metadata expressions. A data expression identifies data values in a source. A metadata expression identifies column names in a source. When a data domain contains multiple expressions, any column name or column value that matches an expression in the pattern appear in the search results.

Regular Expression Syntax

A regular expression contains characters that represent source character types, source character sets, and string or word boundaries in the source columns. A regular expression can also contain quantifiers that determine how many times characters can occur in the source data. Regular expressions are case sensitive.

The following special characters are examples of characters that you can include in a regular expression:

Any character except [\^$.|?*+()

All characters except the listed special characters match a single instance of themselves. For example, `abc` always matches `abc`.

\ (backslash) followed by any of the following special characters: `[\^$.|?*+(){}`

A backslash escapes any special character in a regular expression, so the character loses the special meaning.

*** (asterisk)**

Matches the preceding token zero or more times.

[(left bracket)

Marks the beginning of specifications for one character that you want to match.

- (hyphen)

Specifies a range of characters. For example, [a-zA-Z0-9] matches any letter or digit.

] (right bracket)

Marks the end of the specifications for one character.

? (question mark)

Makes the preceding item optional.

{n} where n is an integer > = 1

Repeats the previous item n times.

For information about creating regular expressions, see tutorials and documentation for regular expressions on the internet such as http://www.regular-expressions.info/tutorial.html.

Data Patterns

Data patterns are regular expressions that describe the format of source data in a data domain.

A data pattern can contain multiple data expressions. If any of the expressions match patterns of the data for a column, then the column belongs in the data domain. You can configure detailed regular expressions to identify data in columns.

For example, a Social Security number contains numbers in the following pattern:

```
999-99-9999
```

The following regular expression shows a data pattern that describes the format of a Social Security number:

```
[0-9]{3}-[0-9]{2}-[0-9]{4}
```

Metadata Patterns

A metadata pattern is a regular expression that identifies column names in a source. A metadata pattern can contain multiple metadata expressions.

A metadata expression can be a column name or part of a column name. For example, if you configure `.*Name*` as a metadata expression, column names such as Name, Employee_Name, and Organization_Name in the source appear in the search result.

A column name that matches any metadata expression in the pattern appears in the search results.

A Social Security number might have different column names. The following regular expressions are metadata expression to find Social Security numbers by column name:

```
.*SSN*
.*SOCIAL*
.*SECURITY*
```

Data Domain Options

When you create a data domain you configure options that describe the data domain.

Configure the following options to describe a data domain:

Name

Data domain name.

Sensitivity level

The sensitivity level for all columns in the data domain. The Administrator defines the sensitivity levels that you can choose from when you apply the sensitivity level option.

Description

Description of the data domain.

Status

Data domain status is enabled or disabled. When the data domain is enabled, a profile for data discovery includes the data domain. Default is enabled.

Creating a Data Domain

When you create a data domain, you can enter regular expressions that describe the data that you want to include in the data domain. You can also enter regular expressions that describe the names of database columns to include.

1. To access the policies, click **Policies**.

   The **Policies** view shows a list of the policies, data domains, and rules in the TDM repository.
2. Click **Actions** > **New** > **Data Domain**.
3. Enter the name, sensitivity level, and description for the data domain. Click **Next**.
4. Click **Next**.
5. Optionally, enter a regular expression to filter columns by data pattern.
6. To add more expressions for data patterns, click the **+** icon.
7. To add regular expressions that filter columns by column name, click **Next**. Or, click **Finish** to skip entering any more data domain information.

   You can add multiple expressions.
8. Enter regular expressions to filter columns by column name.
9. Click **Next** if you want to apply preferred masking or generation rules to the data domain. Or, click **Finish** to finish configuring the data domain.
10. To add preferred masking and generation rules to the data domain, click **Add Rules**.

    The **Add Rules** dialog box appears.
11. Select the data masking rules and data generation rules that you want to add.
12. Click **OK**.

13. Enable a default masking rule and a default generation rule.
14. Click **Finish**.

Copying a Data Domain

You can create a data domain by copying a data domain.

1. To access the policies, click **Policies**.
2. Click a data domain description to select the data domain.

   Do not open the data domain.
3. Click **Actions** > **Duplicate**.

   The **Copy <Data Domain Name>** dialog box appears.
4. Change the name and description of the data domain. Click **Save**.

Editing a Data Domain

You can edit a data domain to update the rules, data patterns, and metadata patterns.

1. To access the policies, click **Policies**.

   The **Policies** view shows a list of the policies, data domains, and rules in the TDM repository.
2. Click the name of the data domain that you want to edit.

   The data domain opens in a tab.
3. Click **Actions** > **Edit**.

   The **Edit** dialog box appears.
4. To add or edit expressions for data patterns, click the **Data Patterns** tab.
5. To add or edit expressions for metadata patterns, click the **Metadata Patterns** tab.
6. To add or edit masking or generation rules, click the **Preferred Rules** tab. Click **Save**.

   If you delete a rule that contains assignments, the **Impacted Objects** dialog box appears with the list of affected columns and plans.
7. To download the list of affected columns and plans, click **Export**, and save the .csv file.
8. To save the changes, click **Continue**.

   To update the changes in a plan, you must generate and run the plan again.

Deleting a Data Domain

When you delete a data domain, you delete the data domain assignments. When you delete a data domain, you do not delete the rules that you add to the data domain.

1. To access the policies, click **Policies**.

   The **Policies** view shows a list of the policies, data domains, and rules in the TDM repository.
2. Select the name of the data domain that you want to delete.
3. Click **Actions** > **Delete**.

   The **Delete Data Domain** dialog box appears. If you delete a data domain that contains assignments, the **Impacted Objects** dialog box appears with the list of affected columns and plans.

4. To delete the data domain that has no assignments, click **OK**.
5. To delete the data domain that contains assignments, click **Continue**. To download the list of affected objects, click **Export**, and save the .csv file.

   To update the changes in a plan, you must generate and run the plan again.

Policy Packs

A policy pack contains rules that mask common types of sensitive data in the business applications.

TDM includes the following three policy packs:

- PII
- PHI
- PCI

After you install TDM, the policy packs are available in the `TDM/Content/Policies` folder. Import the TDM policy packs in Test Data Manager. The policy packs include the data and metadata patterns of the data domains. You can view the regular expression strings that define the search patterns in TDM policy packs.

PII Policy Pack

The Personally Identifiable Information (PII) pack contains data masking rules and policies specific to masking personal information.

The following table describes the data domains and the corresponding default rules available in PII policy pack:

Data Domain Name	Default Rule
Age Domain	Age Rule
Birth Date Domain	Birth Date Rule
Birth Place Domain	Birth Place Rule
Country Domain	Country Rule
Credit Card Domain	Credit Card Numbers Rule
Drivers License Number Domain	Drivers License Rule
Email Address Domain	Email Address Rule
First Name Domain	First Name Rule
Full Address UPPER Domain	Full Address UPPER Rule
Full Name Domain	Full Name Rule
Gender Domain	Gender Rule

Data Domain Name	Default Rule
Grades Domain	Grades Rule
IP Address Domain	IP Address Rule
Job Position Domain	Job Position Rule
Last Name Domain	Last Name Rule
Organization Name Domain	Organization Name Rule
Passport Domain	Passport Rule
Phone Domain	Phone Number Rule
Salary Domain	Salary Rule
School Name Domain	School Name Rule
Social Security Number Domain	Social Security Number Rule
State Domain	State Rule
Street Name Domain	Address Street Name Rule
UK National Identifier Domain	National Identifier Rule
Vehicle Registration Number Domain	Vehicle Registration Rule
ZIP Domain	ZIP Rule

PHI Policy Pack

The Protected Health Information (PHI) pack contains data masking rules and policies specific to the healthcare and the pharmaceutical industries.

The following table describes the data domains and the corresponding default rules available in PHI policy pack:

Policy Pack	Description
Account Number Domain	Account Number Rule
Birth Date Domain	Birth Date Rule
Certificate License Number Domain	Certificate License Number Rule
Device Identifier Serial Number Domain	Device Identifier Serial Number Rule
Email Address Domain	Email Address Rule
First Name Domain	First Name Rule

Policy Pack	Description
Health Plan Beneficiary Number Domain	Health Plan Beneficiary Number Rule
IP Address Domain	IP Address Rule
Last Name Domain	Last Name Rule
Medical Record Number Domain	Medical Record Number Rule
Phone Domain	Phone Number Rule
Primary Account Number Domain	Primary Account Number Rule
Social Security Number Domain	Social Security Number Rule
State Domain	State Rule
Unique Identifier Domain	Unique Identifying Number Rule
Web URL Domain	Web URL Rule
ZIP Domain	ZIP Rule

PCI Policy Pack

The Payment Card Industry (PCI) pack contains data masking rules and policies specific to the banking and the finance industries.

The following table describes the data domains and the corresponding default rules available in PCI policy pack:

Policy Pack	Description
Account Number Domain	Account Number Rule
Birth Date Domain	Birth Date Rule
Credit Card Domain	Credit Card Rule
Expiration Date Domain	Expiration Date Rule
First Name Domain	First Name Rule
Last Name Domain	Last Name Rule

Import and Export

You can export policies, domains, and rules to XML files and import the XML files to another TDM repository.

When you export a policy, Test Data Manager exports references to the data domains in the policy. When you export data domains, Test Data Manager exports references to the rules in the repository. However, Test Data Manager does not export the data domains or rules unless you choose to export them.

Test Data Manager exports the policies, data domains, and rules to separate XML files. Each file has a default name and number that includes the date. For example: `Policy_121112141309.xml` or `Domain_121112141309.xml`. You can rename the files before you save them.

Exporting Policy Components

Export policies, domains, and rules to separate XML files.

1. In the **Policies** view, choose the policies, data domains, and rules to export.

   To choose an object, select the check box next to the object name.
2. Click **Actions** > **Export**.
3. Click **Open** to open the XML file in a text editor or click **Save** to save the XML file on the system.

   Test Data Manager prompts you for each type of XML file you create. It downloads separate files file policies, data domains, and rules.

Importing Policy Components

You can import policies, data domains, and rules from XML files that have been exported from a TDM repository.

1. In the **Policies** view, click **Actions** > **Import**.
2. Browse for the XML file to import.
3. Click **Finish** to import the file.

   If the file policy, data domain, or rule already exists in the TDM repository, Test Data Manager overwrites the original file.

Data Domain Import

Before you import a profile, verify that TDM contains the objects in the profile. So before you import a data domain profile, you must import the data domains in the profile. You can import data domains from the Model repository into TDM. You can edit the data domain in Test Data Manager after you import it.

If you edit a data domain in the Developer tool after you import it, you can import it again to update the data domain in TDM.

You can view data domains that you import in the list of data domains in the **Policies** view.

When you import a data domain from the Model repository, TDM imports the following information in the data domain:

- Name
- Description
- Links to business terms created in the Analyst tool

TDM does not import the following information in the data domain:

- Data pattern
- Metadata pattern
- Preferred rule

If a data domain with the same name exists in TDM, you can choose to overwrite it. If you continue, TDM imports and updates the description and linked business terms of the data domain. It does not overwrite the data patterns, metadata patterns, and preferred rules that you configured within Test Data Manager.

When you import a data domain, TDM merges the linked business terms and retains existing linked business terms. If you re-import a data domain after you delete a linked term from it in the Developer tool, TDM retains the linked term that you deleted. You can manually delete the linked business term from Test Data Manager.

If you re-import a data domain after you change the name of the data domain in the Developer tool or in Test Data Manager, TDM updates the name. You cannot re-import the data domain if a data domain in TDM has the same name.

You can edit data domains that you import in Test Data Manager. You can edit the description, or add a data pattern, metadata pattern, or preferred rule. You can add or delete links to business terms.

Importing a Data Domain

You can import a data domain from the Model repository into the TDM repository. Import data domains before you import a data domain profile.

TDM does not import the data patterns, metadata patterns, or the preferred rules of the data domain.

1. Click the **Policies** view.
2. Click **Actions** > **Import Data Domains**.

   The **Import Data Domains** window opens with a list of data domains in the Model repository.
3. Select the required data domains.
4. Assign a sensitivity level to each data domain that you want to import. Click the **Sensitivity Level** column and choose the required level from the list. Click **Save**.
5. Click **Import**.

   If a data domain with the same name exists, you are prompted to overwrite the data domain or cancel the import. TDM imports the data domain with the description and any linked business terms. You can view the data domain and its properties in the list of data domains.

Linking Business Glossary Terms to Global Objects

You can link business terms in the Informatica Business Glossary to TDM global objects such as policies, data domains, masking rules, and generation rules. A business term includes the business definition and examples of how to use the business term. Use this to store information that users can look up from one central location. When you link a business term to a TDM object, you can view the TDM object when you select the business term in the Analyst tool.

View business terms linked to objects from the **Policies** view. You must configure an Analyst Service in the Test Data Manager Service properties to view or link business terms with TDM objects.

When you select an object from the **Policies** view, a list of business terms linked to the object appears in the Linked Business Terms tab. The list displays the term, the name of the glossary that the term belongs to, and a description.

You can add and remove business term links to objects from the **Policies** view. Click the business term link in the Linked Business Terms tab to view the business term in the Analyst tool.

If you delete a TDM object that has a business term linked to it, you delete the link to the business term. If you edit the name of a linked TDM object, the edit reflects in the glossary. If you delete a business term from a glossary, you delete the object link to the business term.

Linking a Business Term to an Object

You can link business terms to an object on which you have Manage privileges.

Perform the following steps to link a business term to a TDM global object.

1. To access the **Policies** view, click **Policies**.
2. Select the policy, data domain, masking rule, or generation rule to which you want to link a business term.
3. Click **Add** on the **Linked Business Terms** tab. The **Link Asset to Business Term** page opens.
4. Select the glossary from which you want to select a business term. A list of business terms in the glossary appears.
5. Select the terms that you want to link to the object and then click **OK**. The business terms appear in the list of linked business terms.

Deleting a Business Term Link to an Object

You can delete a business term link to objects on which you have Manage privileges. You can delete the link from Test Data Manager or from the Business Glossary UI.

Perform the following steps to delete a business term link to a TDM global object from Test Data Manager.

1. To access the **Policies** view, click **Policies**.
2. Select the policy, data domain, masking rule, or generation rule that you want to edit. The list of business terms linked to the object appears in the **Linked Business Terms** tab.
3. Select the business term that you want to delete.
4. Click **Delete** on the **Linked Business Terms** tab. A confirmation page opens.
5. Click **Yes** to confirm the deletion. TDM deletes the link to the object and updates the **Linked Business Terms** tab.

Policy Management

The **Policies** view shows the policies in the TDM repository. You can create and edit a policy from the **Policies** view.

You can create a policy before or after you define data masking rules and data domains. After you create a policy, you can edit the policy to add data domains and rules. You can copy a policy if you need to create a policy with similar rules or data domains. The duplicate policy contains the rules and data domains from the original policy.

You can export a policy from a TDM repository to an XML file. Import the XML file in another TDM repository.

Creating a Policy

Create a policy in the **Policies** view. You can create the policy before you create data domains and rules. You can add the data domains and rules to the policy at any time.

1. In the **Policies** view, click **Actions** > **New** > **Policy**.

   The **New Policy** dialog box appears.
2. Enter a name and optional description for the policy, and click **Next**.
3. To add data domains to the policy, click **Add Data Domains**.
4. Select the data domains from the list.
5. Click **Finish**.

   The policy appears in the **Policies** view.

Copying a Policy

You can create a policy by copying a policy. When you copy a policy, Test Data Manager copies the data domains and rules in the original policy to the new policy.

1. To access the **Policies** view, click **Policies**.
2. Click a policy description to select the policy.

   Do not open the policy.
3. Click **Actions** > **Duplicate**.

   The **Copy <Policy Name>** dialog box appears.
4. Change the name and description of the policy. Click **Save**.

Editing a Policy

You can edit the policy properties and update the data domains in a policy.

1. To access the **Policies** view, click **Policies**.
2. Click the name of the policy that you want to edit.

   The policy opens in a tab.
3. Click **Actions** > **Edit**.

   The **Edit Policy** dialog box appears.
4. You can change the policy name, description, or status.
5. Click the **Data Domains** tab to edit data domains in the policy.
6. Click **Add** to add data domains to the policy.

   A list of data domains appears.
7. Select the data domains that you want to add to the policy.
8. Click **OK** to select the data domains.
9. Click **Save**.

   If you delete a data domain from the policy and if the data domain contains rule assignments, the **Impacted Objects** dialog box appears with the list of affected columns and plans.
10. To download the list of affected columns and plans, click **Export**, and save the .csv file.

11. To save the changes, click **Continue**.

    To update the changes in a plan, you must generate and run the plan again.

Deleting a Policy

When you delete a policy, you do not delete the data domains or rules that you added to the policy. When you delete a policy, you remove all the assignments that you create with the policy.

1. To access the **Policies** view, click **Policies**.
2. Click to open the policy you want to delete.
3. Click **Actions** > **Delete**.

   The **Delete Policies** dialog box appears.
4. Confirm that you want to delete the policy.

   If the policy contains assignments, the **Impacted Objects** dialog box appears with the list of affected columns and plans.
5. To download the list of affected objects, click **Export**, and save the .csv file.
6. To delete the policy, click **Continue**.

   Test Data Manager deletes the policy. To update the changes in a plan, generate and run the plan again.

CHAPTER 5

Data Discovery

This chapter includes the following topics:

- Data Discovery Overview, 72
- Data Discovery Sources, 73
- Discover View, 74
- Column Properties, 75
- Data Discovery Task Flow, 77
- Primary Key Discovery , 77
- Entity Discovery, 78
- Data Domain Discovery, 79
- Column Profile, 81
- Profile Management, 82
- Profile Import, 86
- Apply the Results, 87
- Project Tables, 91

Data Discovery Overview

Run profiles to discover source data for data masking and data subset operations.

Before you run data masking, you can discover which columns to mask with the same masking rules. Before you create a data subset, you can discover relationships between tables and the entities for the data subset database. You can apply profiling results instead of manually configuring data subset entities or manually applying a data masking rule to one column at a time.

You can run the following types of profiles:

Foreign key

Determines relationships between tables. When you run the profile, the Data Integration Service evaluates the structure of the source tables and identifies potential table relationships. Select entities from the profile results. TDM adds the logical constraints in the TDM repository, but does not apply these constraints in the source tables. When you generate a workflow from a data subset plan, TDM uses the constraints to create subset data with referential integrity.

Primary key

When you run the profile, the Data Integration Service evaluates the structure of the source tables and identifies potential primary keys. The primary key profile includes the percentage of unique values in a column or combination of columns. Select primary keys from the profile results. TDM adds the primary keys in the TDM repository, but does not add them to the source tables.

Data domain

Identifies the columns that belong in a data domain based on the data value of the column or the column name. Use the results when you apply data masking rules. You can apply a rule to all columns that belong to the data domain instead of manually applying the rule to one column at a time.

Column

Analyzes data quality in selected columns in a table. A column profile determines the characteristics of columns in a data source, such as range of values, list of values, and data patterns. When you create a column profile, select the tables on which you want to run a profile and configure data sampling options.

You can run data domain profiles and primary key profiles at the same time. You can also import and run profiles that you create in Informatica Developer.

Data Discovery Sources

You can run profiles for data discovery on relational sources.

You can run profiles for data discovery on the following sources:

- Oracle
- Microsoft SQL Server
- IBM DB2
- Sybase
- Teradata

Use ODBC to connect to Sybase and Teradata sources.

Rules and Guidelines for Data Discovery Sources

Use the following rules and guidelines when you configure the data discovery sources:

- You cannot run a profile for data discovery on nonrelational sources. However, you can use profiling in PowerCenter or Informatica Data Quality to create and run profiles and then export the results. You can import the results to the TDM repository with Test Data Manager. For information on how to run profiles in Informatica Data Quality and export them, see the Informatica *Profile Guide*.
- You can import the results of profiles that use specific options in Informatica Data Quality.
  - You can view the results of profiles that use the Enterprise Discovery Profile and Profile options.
  - You can import the results of data discovery performed with mapplets created using simple regular expressions. You cannot import results if the mapplets use labeler, tokenizer, content set, or reference tables.
  - You can import data domains from the Model repository into TDM. TDM does not import the data pattern, metadata pattern, and preferred rules.

- You can import and view domain discovery results of profiles run in the project. You can import the results of profiles created in folders within the project, but you cannot view the results in TDM.

- A table that you import to the repository must have the same connection when you use it in a profile. If you use a different connection for the profile than you used to import the data source, you might receive unexpected results.
- You cannot run a profile that contains two tables with the same name. For example, a project might have more than one EMPLOYEE table. Each EMPLOYEE table might have a different owner. You must create a separate profile for each EMPLOYEE table.

Discover View

View tables, manage primary key-foreign key relationships, configure profiles, and update columns for masking in the **Discover** view.

The following image shows the **Discover** view in a Project:

The **Discover** view contains the following views:

Tables

View the tables in the data source. Select a table and view the relationship between the table and other tables in the project. View the columns in a table. Define primary key and foreign keys in tables.

You can disable physical primary key columns when you do not need the composite primary keys or entity relationships. In the **Tables | Columns** view, you select a row with a physical primary key and enable or disable the primary key from the **Actions** menu. You can also edit the precision of a source table column based on the values that you need in the target column.

Profile

Define a profile for data discovery. Configure the profile to find columns for data domains and possible primary key-foreign key relationships between tables.

Columns

View the results of a data domain profile. Maintain data domain assignments and define similar column values for value cascades.

Files

View a list of XSD files in the project. Select a file in the list to open the file information in a separate tab. You can perform data subset, data masking, and data generation tasks on XSD sources.

Column Properties

You can configure column properties for a project source.

The following table describes the properties listed on the **Discover** | **Columns** view:

Property	Description
Table Name	Name of the table.
Column	Name of the column.
Datatype	Column datatype.
Domain	Name of the data domain that the column belongs to.
Restricted	The column in the database is a restricted column. The PowerCenter Integration Service does not write the value of a restricted column to a target.
Sensitive	Indicates whether a column contains sensitive data.
Similar Value Columns	Shows the number of columns that are related to this column in a value cascade.

Sensitive Columns

When you mark a column as sensitive, you indicate that the column contains sensitive data. Sensitive columns are for reference only and they do not affect processing.

When you assign a data domain to a column, Test Data Manager marks the column as sensitive.

Value Cascades

Use a value cascade to mask similar columns in different tables with the same masking rules. Use a value cascade when the columns that you want to mask are not related by key constraints. Data masking returns the same values into each column. Configure a group of similar value columns to define a value cascade.

Configure value cascades as similar value columns and define one of the columns as the driving column. If you assign a rule to the driving column, Test Data Manager assigns the rule to all the columns in the group. For example, you configure employee_num column in one table and EmpID in another table as similar value columns. Define employee_num as the driving table. When you assign a data masking rule to employee_num, EmpID receives the same masking rule. The columns in a value cascade must be the same type.

When you view the columns in a project, a driving table shows a number in the **Similar Value Columns** column. The number indicates how many other columns are in the value cascade. The similar value columns do not show a number in the **Similar Value Columns** column.

You can view a list of the columns in a value cascade. Select the driving column in the content panel. Click the **Similar Value Columns** view in the details panel. A list of the similar value columns appears.

To configure value cascades for an XSD source, open the XSD file from the **Discover | Files** view within a project. Select the elements and attributes, and configure similar value columns.

Creating a Value Cascade

Configure a value cascade as a group of similar value columns and define one of the columns as the driving column.

1. In the **Discover | Columns** view, select multiple columns that you need to define as similar columns.
2. Click **Actions** > **Set as similar value columns**.

   A list of the columns with similar values appears.
3. Choose one column as the driving column.

   To select the column, click the driving table column checkbox to change the value from No to Yes.

Changing a Value Cascade

You can change a value cascade configuration. You can change the driving column, delete a column, or add similar value columns.

1. In the **Discover | Columns** view, select the driving table of the value cascade you want to change.

   A driving table has a numeric value in the **Similar Value Columns** column.
2. Click the **Similar Value Columns** view in the **Details** panel.

   A list of the columns in the value cascade appears.
3. Click **Edit Similar Value Columns**.

   The **Edit Similar Value Columns** dialog box appears.
4. To change the driving column to a different column, change the **Driving Column** value to Yes for the column that you want.

   You do not have to disable the previous driving table column. The value changes to No by default.
5. To add similar value columns to the group, click **Add**.

   Select the columns from the **Add Similar Value Columns** dialog box.
6. To remove columns from the group, select the columns on the **Edit Similar Values** dialog box and then click **Delete**.

Deleting a Value Cascade Group

You can delete a value cascade.

1. In the **Discover | Columns** view, select the driving table for the value cascade.

   A driving table has a numeric value in the **Similar Value Columns** column.
2. Click the **Similar Value Columns** view in the **Details** panel.

   A list of the similar value columns in the value cascade appears.
3. Select any column in the group and click **Delete.**
4. Confirm that you want to delete the similar values group.

   Test Data Manager removes the value cascade.

Auto Cascades

An auto cascade masks the primary key in a parent table and related foreign keys in child tables with the same value.

Tables with key constraints have auto cascades applied by default. You can manually disable or enable auto cascades. For example, you need to mask the employee ID in an employee master table and two child tables. The employee ID is a primary key in the master table. When you mask the employee ID in the employee table, the employee ID in the child tables receives the same masked value.

To use auto cascades, key constraints must exist between the parent and the child tables.

You cannot use auto cascades for XSD, MongoDB, and Cassandra data sources.

Data Discovery Task Flow

You can run profiles for data discovery to find primary keys and entity relationships between tables in the source data. You can run a data domain profile to search for columns to assign to data domains for data masking.

Before you can run profiles, the administrator must configure a connection to the source database for data discovery. The administrator must also configure connections to the Data Integration Service and the Model Repository Service.

Complete the following high-level steps to perform data discovery:

1. Create a profile.
2. Select the type of profiling you want to perform. You can choose to run a primary key profile, an entity profile, or a data domain discovery profile.
3. If you choose to run a data domain discovery profile, choose the data domains to search for.
4. Choose the sampling size for the profile.
5. Run the profile and monitor the job.
6. After the job completes, open the profile again.
7. Review the primary key profile results, the entity profile results, and data domain profile results.
8. Select and approve the results that you want to use for data masking and data subset operations.

Primary Key Discovery

To discover possible keys in source tables, create a primary key profile.

To identify potential data quality issues, create a primary key profile that discovers non-unique rows. A primary key discovery profile is also useful when the primary key is a combination of columns that uniquely defines a row. Non-conforming rows might contain duplicate information. After you discover potential keys in a table, you can add the keys to the source table in a project.

For example, an Employee table contains an EmpID column. The value of the column is 100% unique. A primary key profile identifies the column as a possible primary key that you can add to the table in the project.

Primary Key Profile Options

When you run a primary key profile, the Data Integration Service infers primary key candidates from the source tables. To limit the candidate results, configure inference options.

A primary key profile identifies the columns and combinations of columns that meet a specific conformance level. You can edit the maximum number of columns to combine for unique key identification.

The following table describes the options for primary key profiles:

Option	Description
Primary Key Rows	The maximum number of columns that the profile can combine to form a primary key. Default is 3.
Max Rows	The maximum number of rows to profile. Default is 1,000.
Conformance Criteria	The minimum percentage or maximum number of key violations that the profile allows. Default is a minimum percentage of 99.

Entity Discovery

To discover data relationships between source tables, create an entity profile. Entity discovery shows the tables that are potential entities.

The entity profile discovers entity relationships in the source data. An entity consists of related parent and child tables. The entity profile returns a list of possible entities by comparing column names in the source tables. You can select entities from the profile results to add to the project. When you run a data subset operation, the entities that you define for the subset determine the subset database structure and the data in the subset.

For example, the Employee table, the Employee_Detail table, and an Employee_Address table contain an EmpID column. The entity profile suggests that the Employee table could be an entity with two direct relationships. The Employee table has two child tables. When you view the profile results, you select the Employee table as an entity. When you run a data subset operation, you can create a filter on EmpID to retrieve the data for employees with an EmpID less than 2000. The data subset operation filters data for each table in the entity relationship.

If you select Entity as the profile type, and the physical database does not contain primary keys, the profile does not return any results. If you run the profile as a primary key and entity profile, the profile discovers possible primary keys and foreign keys. The profile returns results even if the physical database does not contain primary keys.

Entity Profile Options

TDM infers foreign key candidates from the source tables. To limit the candidate results, specify inference options.

Define options to configure how the entity profile matches the datatypes of the primary keys and the foreign keys of the entity table.

The following table describes the entity profile options:

Option	Description
Datatypes used in comparisons	The datatype to use for comparisons. Select one of the following options: - Inferred Datatypes. Compares the primary key and the foreign key based on the datatypes inferred by TDM. - Documented Datatypes. Compares the primary key and the foreign key based on the datatypes defined for the source column.
Comparison case sensitivity	Performs case-sensitive comparisons.
Trim spaces before comparison	Trims spaces before comparing values. Select one of the following options: - None. Does not trim leading or trailing spaces. - Both. Trims both leading and trailing spaces. - Leading. Trims leading spaces. - Trailing. Trims trailing spaces.
Maximum foreign keys returned	Maximum number of matched foreign key results that you want to return.
Minimum confidence percent	Minimum percentage of foreign key values that match the inferred datatype.
Regenerate signature	Determines whether the data profiling warehouse should clear the existing object information before the Data Integration Service runs the profile.

Data Domain Discovery

Data domain discovery finds the source columns that contain similar data. The profile assigns the same data domain name to each column that contains similar data. You can assign the same data masking rules to all the columns in a data domain at the same time.

Create a data domain to describe the columns you need to mask with the same data masking rules. When you create a data domain, you configure regular expressions that define patterns in the data or patterns in the column names.

Run the data domain discovery profile to find the columns that match the criteria in the data domain regular expressions. When you configure a profile for data domain discovery, select the tables to search in the data domain discovery operation. Select which data domains to search for in the tables. You can select policies that contain data domains instead of selecting each data domain to search for.

After you run the profile for data discovery you can view the profile results. The profile results assign source columns to data domains. You can choose which profile results to use for data masking.

Data Domain Profile Sampling Options

When you run a profile for data domain discovery, configure sampling options to limit the number of rows to search or limit the regular expressions to search with.

The following table describes the sampling options that you can select in a profile for data discovery:

Option	Description
Data	Search for patterns in the data only.
Column Name	Search for patterns in the column name only.
Data and Column Name	Search for patterns in the data and in the column name.
Maximum Rows to Profile	Limit the number of rows to profile. Default is 1000.
Minimum Conformance Percent	Minimum percentage of rows where the column data or metadata matches the data domain.

Assigning a Data Domain to Multiple Columns

You can manually assign a data domain to multiple columns at a time. You can also remove the data domain assignment from multiple columns at a time.

1. Open a project.
2. Navigate to the **Discover | Columns** view.

   A list of all the columns in the project appears.
3. Select the columns that you want to assign the data domain to.
4. Click **Actions** > **Edit Assignments**.

   The **Edit Data Domain Assignment** dialog box appears.
5. Choose the data domain to assign to the columns.

   You can choose a blank data domain to remove the previous data domain assignment.
6. Click **Save**.

   The data domain assignments appear in the **Columns** view.

Manually Updating the Column Data Domain

You can manually update a data domain for a column. When you add a data domain to a column, Test Data Manager marks the column as a sensitive column.

1. Open the project and click **Discover | Columns**.
2. Click the **Domain** column for the column you want to update.
   A list of data domains appears.
3. Select the data domain to add to the column.

Note: You can run a profile for data domain discovery to update the data domain for columns.

Column Profile

Create and run a column profile to determine the data source column characteristics, such as a list of values, data patterns, and value ranges. Create a column profile, and extrapolate the profile results to generate test data.

Select the columns on which you want to run a profile and configure data sampling options when you create a column profile.

After you configure and run a column profile, TDM analyzes column data and infers rules based on the data. You can view the profile results for the columns. You can manually assign the inferred rules to the columns in the profile results or apply the inferred rules that TDM provides.

You can run column profiles on Oracle and Microsoft SQL Server databases.

Inferred Rules

After you configure and run a column profile, TDM infers analyzes column data and infers rules based on the data. Column data types include all the inferred data types for each column in the profile results.

Column profile results contain the following types of inferred rules:

Range

The range of column values, such as the maximum and minimum lengths of values and the first and last values in each column. The range of values includes the percentage of null values in the columns. You can use rang of values for numeric and date data types.

Patterns

The patterns of data in each column and the percentage of distribution for each pattern. The sum of the distribution percentage must be equal to 100%. You can use data patterns for string and numeric data types.

List of Values

The list of unique and nonunique values in the columns and the percentage of distribution. When you assign the inferred rule type to generate test data, the sum of the distribution percentage must be equal to 100%. You can use list of values for string, numeric, and date data types.

You can assign the inferred rule type to the target columns from the **Define | Data Generation** page. Auto rules are the inferred rules that you assign after you run a column profile. When the status of inferred rule is Yes, you can assign Auto Rule to target columns to generate test data. You can edit the properties of the inferred rule types from the **Define | Data Generation** page.

Column Profile Options

You can configure the sampling options when you create or edit a column profile.

The following table describes the sampling options for a column profile:

Option	Description
Minimum conformance percentage for patterns	Required. The minimum percentage of patterns that you want in the profile results. Default is 5. If the patterns are less than the minimum conformance percentage, TDM does not display the patterns in the profile results.
Maximum number of rows to be profiled	Required. The maximum number of rows that you want to run the profile on. You can choose all rows in the data object or enter the number of rows that you want to profile on. Default is All.
Minimum data coverage to show List of Values	Required. The minimum percent of data that you want in the list of values. Default is 80. If the maximum number of values that you want in the list of values is 20, then 80% of the data must contain these 20 values.
Maximum number of values in List of Values	Required. The maximum number of values that you want in the list of values. Default is 20.

Profile Management

You can create and run profiles, add primary and unique keys to the source, and verify and add constraints to the source. Alternatively, you can import profiles that you create and run in Informatica Developer.

You can run a profile multiple times, and you can edit a profile between runs.

Creating a Primary Key Profile

Create a primary key profile to identify a list of possible primary keys that you can add to tables without keys. Create a primary key profile if the source data does not contain primary and foreign key relationships.

1. Open the project, and click the **Discover** view.
2. Click the **Profile** view.
3. Click **Actions** > **New Profile** to create a profile.
4. In the **New Profile** dialog box, enter the profile name and description. Choose to create a primary key profile.
5. Click **Next.**
6. Select the tables to profile, and click **OK**.
7. Click **Next.**
8. In the **Sampling Options** page, enter the maximum number of columns that you want to include in the primary key.
9. Enter the maximum number of rows that you want to run the profile on.
10. Enter the minimum conformance required, or enter the maximum number of rows that can violate the criteria.

11. Click **Finish**.
12. Click **Actions** > **Execute**.

Creating an Entity Profile

The entity profile creates a list of possible parent tables that you can define for data subset operations. Create and run entity profiles in the **Discover** view.

1. Open the project, and click the **Discover** view.
2. Click the **Profile** view.

   The **Profile** view shows a list of the profiles in the project.
3. Click **Actions** > **New Profile** to create a profile.
4. In the **New Profile** dialog box, enter the profile name and description. Choose to create an entity profile.

   Select primary key and entity profile if the physical database does not contain any primary keys.
5. Click **Next**.
6. Select the tables to profile, and click **OK**.
7. Click **Next**.
8. In the **Sampling Options** page, enter the maximum number of columns that you want to include in the primary key.
9. Enter the maximum number of rows that you want to run the profile on.
10. Enter the minimum conformance required, or enter the maximum number of rows that can violate the criteria.

    The following table describes the entity options that you can update:

Entity Option	Description
Data types used in comparison	Determines how Test Data Manager derives data types to compare in the tables. Choose one of the following options: - Inferred data types. Compares the primary key and the foreign key based on the data types inferred by TDM. - Documented data types. Compares the primary key and the foreign key based on the data types defined for the source column.
Comparison case sensitivity	Indicates if the foreign key comparison is case-sensitive or not. If you choose case-sensitivity, profile performance might be affected.
Trim values before comparison	Trims leading values, trailing values, or leading and trailing values.
Minimum conformance percent	Minimum percentage of rows without key violations that the profile allows for foreign key discovery.
Regenerate Signature	Determines whether the data profiling warehouse should clear the existing object information before the Data Integration Service runs the profile.

11. Click **Save**.
12. Click **Actions** > **Execute** to run the profile.

Creating a Data Domain Profile

Create and run data domain profiles in the **Discover** view.

A project must contain policies before you create a data domain profile. The policies contain the data domains that you can use in a profile for data discovery.

1. Open the project and click the **Discover** view.
2. Click the **Profile** view.

   The **Profile** view shows a list of the profiles in the project.
3. Click **Actions** > **New Profile** to create a new profile.
4. In the **New Profile** dialog box, enter the profile name and description. Choose to create a data domain profile.
5. Select the tables to profile and click **OK**.
6. Click **Next.**
7. In the **Select Sampling Options** pane, choose whether to add policies or data domains to the profile. When you select a policy, Test Data Manager includes all data domains in the policy.

   Test Data Manager returns a list of policies or data domains in the pane.
8. Select the policies or the data domains to profile.
9. In the **Sampling** panel, select whether to run data discovery on the source data, the column name, or the data and the column name.

   You can run a profile for column metadata and then run it again for the source data.
10. Enter the maximum number of rows to profile.
11. Enter the minimum conformance percent.

    All rows might not conform to the data domain expression pattern. You can enter a minimum percentage of the profiled rows that must conform.
12. Click **Save.**
13. Click **Actions** > **Execute**.

Creating a Column Profile

To perform a data generation operation, you can create and run a column profile.

1. Open a project and click the **Discover** view.
2. Click the **Profile** view.

   The **Profile** view shows a list of the profiles in the project.
3. Click **Actions** > **New Profile**.

   The **New Profile** dialog box appears.
4. Enter the profile name and description.
5. Select **Column**.
6. Click **Next**.
7. To add the tables that you want to run a profile on, click **Add**.

   The **Select Tables** dialog box appears.
8. Select the tables, and click **OK**.

   The list of selected tables appears.

9. Click **Next**.
10. To create the column profile, configure the required sampling options.
11. Click **Finish**.

    The profile opens in a tab.
12. To run the profile, click **Actions** > **Execute**.

    The **Execute Profile** dialog box appears.
13. Select a connection to run the column profile.

    Note: A column profile run fails if you select a native Oracle ODBC connection. Create an Oracle ODBC connection with the DataDirect driver.

    A column profile run fails if you do not specify a schema name in the Microsoft SQL Server database connection.
14. Click **Execute**, and close the profile.
15. In the project, click **Monitor** to view the progress of the profile run.

Editing a Profile

You can edit a profile and run it multiple times. When you run the profile, the Data Integration Service overwrites the original profile results.

When you edit a profile, you can change the profile name and description. You can add or remove tables. When you run data domain profiles, you can change the policies and data domains in the profile. When you run primary key profiles and entity profiles, you can change the options.

1. Open the project and click the **Discover** | **Profile** view.
2. In the **Discover** view, click a profile name to open it.
3. Click **Edit**.

   The **Edit Profile** dialog box appears. The dialog box has the **General** tab, the **Tables** tab, and the **Relationship Sampling** tab or the **Sampling** tab depending on the profile type.
4. On the **General** tab, you can edit the name and description of the profile.

   You cannot change the profile type.
5. On the **Tables** tab, add or remove tables that you want to profile.

   To add a table, click **Add** and select the tables you want to add. To delete a table, select the check box for the table you want to delete. Click **Delete**.
6. On the **Relationship Sampling** tab, select to override default inference options. Edit the options for finding primary key and foreign keys in the source data.
7. On the **Sampling** tab, you can change the data domains or policies in the profile. You can choose to run the profile against the source data, the column name, or both data and column name.
8. Click **Actions** > **Execute** to run the profile.

Deleting a Profile

You can delete profiles.

1. Open the project and click the **Discover** view.
2. In the **Discover** pane, select the profile row, but do not open the profile.
3. Click **Actions** > **Delete** to delete the profile.

4. Click **Yes** to confirm the delete.

Profile Import

You can import a profile from the Model repository and view the profile in Test Data Manager.

For example, you can run a profile on a flat file source in the Developer tool. You can import the profile into TDM. You can then use the profile information to apply masking rules to mask source data or to create a data subset in TDM.

Import the profile from within a project. You can run a single profile to identify multiple types of data. So a profile might contain data domain, primary key, foreign key, and column profile information. Select the types of profile information that you want to import.

Before you import a profile, import the metadata and the objects in the profile into the TDM repository. For example, before you import a data domain profile, the TDM repository must contain all the data domains that appear in the profile. You can import the data domains from the Model repository. You cannot view the profile after import if data domains with the same name do not exist in the TDM repository.

You can import profiles with table names that contain alphanumeric characters or the special character _.

Importing a Profile

You can import profiles from the Model repository into TDM. You can then view the profile in Test Data Manager.

Verify that the source metadata and the data domains are in the TDM repository before you import the profile. Import the metadata into the project from a connection by the same name as the connection that you used to run the profile.

1. Open the project and click the **Discover** | **Profile** view.
2. Click **Actions** > **Import Profile**.

   The **Import Profiles** dialog box appears.
3. Select the profile that you want to import from the Model repository. Click **Next**.

   You cannot import a profile that contains a table name with special characters other than _.
4. Optional. TDM imports the profile with the same name and description. Edit the profile name and description if required.
5. Optional. Choose to import the curation information with the profile.

   When you import the curation information, TDM imports the saved review information and performs tasks based on the review information. For example, TDM assigns a data domain to a column if the inferred data domain is approved in a data domain profile.
6. Select the required profile types.

   By default, TDM selects all the profile types that the profile contains. You can choose the profile types that you want to import.
7. If you select the **Column** profile type, enter the column profile properties.
8. Click **Finish**.

   The imported profile appears in the **Discover** view.

Apply the Results

After the profile run, you must close the profile in Test Data Manager and open it again to view the profile results. Click separate views for the primary key results, the entity profile results, the data domain profile results, and the column profile results.

Primary Key Results

The primary key results is a list of possible primary keys that you can add to tables that do not have keys. Use the unique percentage value in the profile results to determine which column or combination of columns that you might add as a primary key for a table.

The following image shows sample primary key profile results:

To view primary key results, close the profile and open it again to display the results. Click the **Profile | Primary Key** view. Examine the primary key profile results. The primary key candidates include the percentage of unique values in a column or combination of columns.

The following table describes the fields in the primary key results:

Property	Description
Table Name	The name of the table that contains the column with the primary key candidate.
Column Name	The name of the column or combination of columns that the Data Integration Service infers to be a primary or unique key.
Relationship	The function of the column in the table is primary key.
Uniqueness	The percentage of unique values in the column. Use the unique percentage value in the profile results to determine which column or combination of columns to add as a primary key. A unique percentage of 100 indicates that all values in the column are unique.
Verified	Indicates whether the primary key candidate has been verified as not containing another constraint.
Existing constraints	Shows any existing primary key or foreign key constraint for the column.
Status	Indicates whether a user verified, approved, or rejected the primary key candidate.

Apply the Primary Key Profile Results

You can choose primary key candidates from the profile results and add the keys to tables in the project.

You can configure a status that indicates whether you want to accept the primary key candidate or reject the primary key candidate. When you approve a primary key, the Test Data Manager adds the key to the table in the TDM repository.

Select a column and change the status from the **Actions** menu. Select **Verify**, **Approve**, or **Reject**.

When you approve the primary key candidate, the Test Data Manager adds the primary key to the table. When you reject the primary key candidate, it does not affect the subset operation.

Use the **Verified** column to track results as you finish the review. The **Verified** column does not affect data subset operations.

Entity Discovery Results

The **Entity** view shows the list of possible entity relationships in the source data. You can select a column in the profile results and see other columns that might contain the same data. You can select entities from the profile results to add to a data subset operation.

To view entity discovery results after you run the profile, close the profile and open it again. Click the **Profile | Entity** view.

Select an entity to view the constraints in the **Constraints** view. Test Data Manager shows the parent columns, the child columns, and the relationships. Approve the constraints to add them to an entity.

You can approve each entity that you want to add to a project.

The following image shows the results of an entity discovery profile:

Data Domain Discovery Results

The data domain profile results show a list of source columns and possible data domains to assign to the columns. You can select which data domain candidates to use for data masking from the profile results.

To view data domain discovery results, close the profile and open it again. Click the **Profile | Data Domain** view.

Select a column and click on the **Data Preview** tab to view the source data of the selected column. The data viewer displays the first 200 records of columns returned in the data domain profile.

You can select rows and approve the data domain for each column. When you approve the suggested data domain for the column, you can assign the rules in the domain to each column in the data domain. You can assign the rules in the **Define | Data Masking** view.

After you are finished working on the data domain for a column, you can verify the data domain for each column in the **Data Domain** view. The **Verify** column is for tracking. It does not affect the data domain profile operation.

When you finish approving the data domains, you can mark the data domain classification as completed. Use this method to verify that you reviewed all the results. Completing the data domain classification does not affect any process.

The following image shows the data domain discovery results:

The following table describes the columns in the data discovery results:

Column Name	Description
Table	Name of the table.
Source	Name of the column to mask.
% Data Conformance	The percentage of rows that contain metadata or data patterns that match the data domain.
% Null	The number of rows that contain NULL values in the source column.
Patterns	The number of data domain data or metadata patterns that match the column.
Column Name Match	Indicates if the column name is the matching data domain pattern.
Current Data Domain	Shows the previous data domain when you run the data domain profile more than once.
Profiled data domain	The data domain name returned from the latest profile run.
Status	Shows whether the data domain is verified or approved.

Column Profile Results

After you run a column profile, you can review the results and either accept or reject the results.

To view column profile results, click the **Profile | Column Discovery** view. Review inferred pattern for all the columns. Accept or reject profiled patterns for each column. Edit the profiled pattern for columns and reject unwanted patterns or specify a pattern. Select Yes in the Inferred Rule column, select the inferred type, and save the profiled pattern.

The following image shows the column profile results:

The following table describes the columns in the column profile results:

Column Name	Description
Table	Name of the table.
Column	Name of the target column to apply the profile result and generate test data.
Data Type	The data type of the target column.
% Null	The number of rows that contain NULL values in the source column.
Inferred Rule	Indicates that you can apply inferred rule types to the column. Default is a hyphen. Select Yes to apply the inferred rule.
Select Inferred Type	The type of inferred rule that is applicable for the column. You can select pattern, list of values, and range based on the data type of the column. Select a list of values or a data pattern for a string data type. Select a list of values or a range of values for a date data type. You can select a list of values, a data pattern, or a range of values for a numeric data type.

Project Tables

You can view the source tables in a project. You can view the primary keys and foreign keys in each table.

Click the **Discover | Tables** view to see the data source tables in the project. You can filter the list of tables by the table name or by the description.

Click the table name to view the table properties in the **General Properties** pane. View the columns in the table in the **Columns** view. Test Data Manager also shows the child tables in the **Relationship Overview** pane.

Table Classification

You can assign a classification to source tables to identify what kind of data the table contains. The classification does not affect profile operations. You can filter tables by this classification when you create entities or plans.

In the **Discover | Tables** view, select the tables that you want to assign a classification to. Choose one of the following table classifications:

- Temporary
- Master
- Configuration
- Transactional
- Log
- Seed

For example, you might want to know which tables are master tables when you determine which tables are driving tables for entities. You might want to configure more filter criteria for tables that contain transactions or logs.

Constraints

Foreign key constraints define parent-child relationships between the source tables. Use constraints to determine the tables to include in a data subset. You can also limit the values that you want to store in the data subset table columns.

When you create a data entity, you select one table for the entity. The table that you select is the entity driving table. Test Data Manager adds other tables to the entity based on the constraints for the driving table.

Use data discovery to find relationships between tables. When you identify the relationships that you want to add to the TDM repository, create and edit constraints in Test Data Manager.

The following types of constraints define relationships between tables in an entity:

Primary Key

A column or combination of columns that uniquely identifies a row in a data source. A table can have one primary key.

Logical Manual

A parent-child relationship between tables based on columns that are not keys. You can create the following logical constraints:

- Logical constraints that you accept from data discovery profiles. You can delete these logical constraints.

- Logical constraints that you define in Test Data Manager. You can delete these logical constraints.

Related Topics:

- "Entities" on page 97

Major and Minor Constraint Severity

When you define a foreign key constraint, configure the constraint severity as major or minor. The severity determines the scope of the data that a data subset receives based on the constraints. You can change the severity to change the final output of a subset operation. The constraint severity level does not affect the structure of an entity. You can change the severity level for a constraint after you create an entity. You can also change the severity level at run time in a subset plan. Doing this does not change the severity levels in the entity.

Major constraints are constraints that move in both directions, child to parent and parent to child. When you want the parent table to select additional child records that are related to the parent, assign a major constraint between two tables.

When you define a constraint, you define it on the child table. When you define a constraint with a major severity level, all of the children of the parent are included in the data subset. For example, you define a constraint between Customer and Leads. Customer is the parent table. You create a filter on Leads to return only high-quality leads. The subset receives the customer for a high-quality lead, but it also receives all the leads for the customer. This major constraint honors referential and transactional integrity. You can apply only such a major constraint from within an entity.

The child record selected for the subset operation pulls in related parent records from both parent tables. The parent records select all the related child records from the child table. As more records are selected in the child table, the related parent records are selected again. This is because the records selected by one parent in the child table might refer to parent records in another parent that is not a part of the subset. This process continues in a loop until there are no more related records to select.

To configure a subset operation that aims to exclude additional data, you can select the Referential Integrity Only option in a plan at run time. When you select this option, the data included remains referentially intact, but this option does not guarantee that the subset will include all child records pertaining to a parent. Therefore, the subset operation might not maintain transactional integrity.

When you change the constraint definition at run time, the relations set in the entity do not change.

Minor constraints are unidirectional constraints, from child to parent. Minor constraints do not move in the reverse direction from parent to child. When you do not want the parent table to select additional child table records that are related to the parent, assign a minor constraint between two tables.

When you define a constraint with a minor severity level, the parent row is included in the data subset, but the data subset does not receive all of the children of the parent row. From the example above, the data subset receives the high-quality leads and the customers for the high-quality leads.

If the schema has constraints that form cycles, or if the schema has objects that have more than one major parent constraint, TDM performs recursive evaluation to calculate the subset. TDM needs a staging connection to perform recursive evaluation. You must therefore create a staging connection to perform subset operations on tables with cyclic or tending to cyclic constraints. It is recommended that you use the source connection as the staging connection.

Manually Add Keys to Tables in a Project

You can manually add primary keys and foreign keys to tables to establish relationships between tables for data subset operations. When you manually add keys, you do not need to run a primary key discovery profile.

When you add keys, you define constraints for data subset, data masking, and data generation operations in the project. You do not update the source database.

You can add the following types of keys constraints:

Primary Key

You can add one column or column combination as the primary key for a table. A primary key column cannot contain null or duplicate values. You cannot add more than one primary key constraint.

Foreign Key

Add a column or column combination as a foreign key in a table. When you define a constraint, you define a foreign key in a table and relate it to a column in a parent table.

Unique Key

Add a column or set of columns as a constraint to define a unique key in a table. A unique key column can contain null values. You can create more than one unique key in table. You can perform multiple data generation assignments when you filter the columns based on unique keys.

Conditional

Add a conditional constraint to perform a data generation operation. Create a conditional constraint to restrict the data that you want to generate in a column based on the data in another column. The conditional constraint that you enter becomes a part of the metadata of the target table. Configure expressions to apply constraints at the table level.

Creating a Primary Key Constraint

You can add a logical primary key to a table to create table relationships for data subset.

1. In a project, click the **Discover | Tables** view.
2. Click a table name to select the table.
3. Click **Constraints**.
4. Click **Create New Constraint**.

   The **New Constraint** dialog box appears.
5. Select **Primary Key**.
6. Click **Next**.
7. To add the columns, click **Add**. Select a column for which you want to add a primary key constraint.
8. Click **Finish**.

Creating a Logical Relationship Between Tables

You can add a logical foreign key to a table to create a table relationship for data subset. Choose a column from a parent table to establish a key relationship between the tables.

1. In a project, click **Discover** > **Tables**.
2. Click the table in which you want to create the foreign key.
3. Click the **Constraints** tab.

4. Click **Create New Constraint**.
5. Enter the constraint properties.

   The following table describes the constraint properties:

Property	Description
Name	Constraint identifier.
Constraint Type	Select the constraint type as Foreign Key. When you want to perform a data generation operation, you select the constraint type as Conditional.
Severity Level	Choose one of the following values: - Major. The relationship includes the parent table and it includes the other children of the parent. - Minor. The relationship includes the parent table but it does not include the other children of the parent.
Parent Table	Choose the parent table to establish the foreign key with.
Enable Constraint	Enable the foreign key relationship.

6. Click **Next**.

   A list of the columns in the table appears in the left panel. A list of the columns in the parent table appears in the right pane.
7. Click a child column from the left pane. Click a parent column from the right pane. Click the **Link** icon to map the parent-child relationship.
8. Click **Finish**.

Creating a Unique Key Constraint

You can add a unique key to a table to create table relationships for data subset and to perform bulk generation rule assignments at a time.

1. In a project, click the **Discover | Tables** view.
2. Click a table name to select the table.
3. Click **Constraints**.
4. Click **Create New Constraint**.

   The **New Constraint** dialog box appears.
5. Select **Unique Key**.
6. Click **Next**.
7. To add the columns, click **Add**. Select the column for which you want to add a unique key constraint.
8. Click **Finish**.

Creating a Conditional Constraint

Create a conditional constraint to restrict the data generated in a column based on the data in another column. Configure an expression to create a conditional constraint.

1. In a project, click **Discover** > **Tables**.
2. Click the table in which you want to create the conditional constraint.
3. Click the **Constraints** tab.
4. Click **Create New Constraint**.

   The **New Constraint** dialog box appears.
5. Enter a name for the conditional constraint.
6. Select the constraint type as Conditional.
7. Click **Next**.
8. Create an expression in the **Expression Builder**. Select the required columns, functions, variables, and operators to build expressions.
9. Click **Finish**.

CHAPTER 6

Data Subset

This chapter includes the following topics:

- Data Subset Overview, 96
- Data Subset Task Flow, 97
- Entities, 97
- Entity Management, 102
- Groups, 106
- Templates, 108
- Data Subset for XSD Sources, 109
- Data Subset Plans and Workflows, 118

Data Subset Overview

You can create a subset of production data if you need a small, targeted, and referentially intact copy of the production data to use in a nonproduction environment. A nonproduction environment might include development, test, or training environments.

For example, you might create a subset of financial data for a specific region or time period.

To create a data subset from relational and flat file sources, you define one or more of the following components:

Entity

Defines a set of tables that are related based on physical or logical constraints. Create an entity when you need to copy data from related tables to a subset database. When you create an entity, select parameters to filter data in columns that you want to include in the subset database. You can manually create an entity or use an entity from profile results.

Group

Defines a set of unrelated tables. Create a group when you need to copy data from one or more unrelated tables to a subset database.

Template

Contains entities and groups. A template must contain at least one entity. Create a template when you want to reuse a set of entities and groups.

After you define the data subset components, add them to a plan. You can generate a workflow from the plan. The workflow loads a subset of the source data into a target database.

Data Subset Task Flow

To define a data subset, add entities, groups, and templates to an project. Create a data subset plan and add the entities, groups, and templates from the project to the plan. To load data into the subset database, generate and start a workflow from the plan.

Complete the following high-level tasks to create the components for the data subset plan:

1. Create a project that contains one or more sources.
2. Create an entity to select related parent and child tables to include in the subset database.
3. Create filter criteria that defines what data to include in the subset database.
4. Optionally, create a group to select unrelated tables to include in the subset database.
5. Optionally, create a template that contains the entities and groups.
6. Create a data subset plan and add the entities, groups, and templates to it. Define parameters to filter the data in the columns. For example, you can define a parameter to filter a month column to contain only October.
7. Generate a workflow from the plan.
8. Start the workflow.
9. Monitor the progress of the workflow.
10. In the target database, verify that the correct data was copied.

Entities

An entity defines a structure for copying related data to the subset database.

An entity consists of a driving table and related tables. A driving table is the starting point for defining relationships between tables in the entity. Test Data Manager defines the tables that relate to the driving table based on physical and logical constraints. You can add tables that have cyclic or circular relationships within the table or between tables. You must add a staging connection before you can add these tables to an entity.

When you create an entity, you select a table as the driving table. Test Data Manager retrieves all related tables to the driving table based on the constraints in the driving table. Test Data Manager adds the related tables to the entity by default.

Define filter parameters for columns of tables in the entity. The filter parameters define the criteria for the subset of the data. To define the subset criteria, you can enter a specific value or a variable that is present in the parameter file.

Related Topics:

- "Constraints" on page 91

Entity Example

An employee has an ID number of 10 and you want information about EmpID 10 from multiple tables. A table called Employee contains an EmpID column. The Employee_History and Employee_Order tables have constraints that relate them to the Employee table that contain the EmpID column.

Create an entity and define Employee as the driving table. Test Data Manager adds the Employee_History and Employee_Order tables to the entity. Define a criteria in which the EmpID value must be equal to 10.

Create a plan, generate the workflow, and start the workflow. The target database contains all the data for EmpID 10. The Employee table has one row with EmpID 10, the Employee_History table has multiple rows of history for EmpID 10, and Employee_Order has all the orders for EmpID 10.

Entities View

Create, edit, and delete entities in the **Define | Data Subset** view of a project.

When you create an entity, you can configure it with the following views:

Entity Map

The **Entity Map** view is a graphical representation of the tables in the entity and the relationships between the tables. You can disable and enable relationships, and edit filter criteria for each table in the **Entity Map** view. To select multiple rows in the **Relationships** view, you can click the required links in the graphical representation of the entity.

Tables

The **Tables** view contains a list of the tables in the entity. You can edit filter criteria for each table in the **Tables** view.

Relationships

The **Relationships** view shows a list of relationships in the entity. For each constraint, the view shows the parent table and the child table in the relationship. You can disable and enable relationships between tables in the **Relationships** view. To select specific links in the **Entity Map** view, you can select the corresponding rows in the **Relationships** view.

Entity Map View

The entity map is a graphical representation of the tables in an entity and the relationships between the tables.

When you create an entity, you define a driving table for the entity. The driving table is a starting point for determining the structure of the entity. Test Data Manager determines the rest of the tables in the entity based on the constraints that include the driving table.

If you drag the entity map panel to the left, a **Table Properties** panel appears that includes information about a selected table or constraint.

You can select any table in the map to edit the filter criteria. You can click a link to edit the constraint between the tables. You can disable relationships between tables from the map. When you select links in the **Entity Map** view, the corresponding rows are selected in the **Relationships** view.

Entity Map Example

You might define the driving table as Employee. Employee has a primary key, EmpID. Employee has three child tables that have the EmpID foreign key. The EmployeeBankDetails child table has an AccountNum constraint that relates EmployeeBankDetails to a BankAccounts table. When you create an entity with Employee as the driving table, Test Data Manager might create an entity map similar to the following map:

The driving table icon contains a star in the entity map.

To add a filter criteria for Employee, click the Employee table in the entity map. Edit the filter criteria from the Table Properties panel.

To edit the constraint between the EmployeeDetails and Employee table, click the link between the tables.

When you click the links between the Employee, EmployeeBankDetails, and BankAccounts table, the corresponding rows are selected in the **Relationships** view.

Choose Tables to View from the Entity Map

You can change which tables you view in the entity map. You can view all the tables or you can view the tables by their relationships with a selected table. By default the entity map shows enabled relationships.

You can choose the **Show excluded tables** option with any of the views to show disabled relationships.

Select one of the following views:

All

Shows all tables in the entity.

All parents

Shows all the parent tables of a selected table.

All immediate relationships

Shows the tables that have constraint relationships with a selected table. You can view one level of parent and child tables from the selected table.

All children

Shows all the child tables of a selected table.

All cyclic relationships

Shows all constraint relationships that create cyclic relationships. The maximum number of cyclic relationships that TDM displays is 1000. You can choose to hide nodes and paths that are not a part of the cycles.

Tables View

The **Tables** view contains a list of the tables in the entity. You can edit filter criteria for each table from the **Tables** view.

The **Tables** view shows the table name, description, and the name of the database that contains the table. The status field indicates whether the entity includes or excludes the table. When you disable a relationship in the entity and the table is not related to any table in the entity map, the table is excluded from the data subset database.

Relationships View

The **Relationships** view shows a list of the relationships between tables in the entity. You can disable and enable relationships from the **Relationships** view.

The **Relationships** view shows the parent table, the child table, and the database names for each table. The view shows the name of the constraint that defines the relationship between the tables. When a table has more than one relationship in an entity, the table appears more than once in the view.

The **Status** field indicates whether the relationship is enabled or disabled.

When you select a row in the **Relationships** view, the corresponding link is selected in the **Entity Map** view.

Disabling and Enabling Relationships in the Relationships View

You can enable or disable a relationship in the **Relationships** view.

1. In Test Data Manager, open a project.
2. Navigate to the **Define | Data Subset** view.
3. Select the entity you need to change.
4. Click the **Relationships** view.

5. Select the relationship to change.
6. Click **Enable Relationship** or **Disable Relationship**.

   The status changes to disabled or enabled.

Enabling a Relationship in the Entity Map

You can enable a disabled relationship from the entity map.

1. In the **Entity Map** view, select **Show Excluded Tables**.

   Test Data Manager includes disabled relationships in the entity map. The entity map shows a disabled relationship with broken lines between the objects in the relationship.
2. Click the broken line between the tables in the relationship you want to enable.
3. Click **Enable Relationship**.

   The broken line becomes a solid line when Test Data Manager enables the relationship.

Disabling a Relationship in the Entity Map

You can disable relationships between entity objects.

1. In Test Data Manager, open a project.
2. Navigate to the **Define | Data Subset** view.
3. Click the entity to open the **Entity Map** view.
4. Click the link between the parent and child icons in the map to select the link.

   Test Data Manager highlights the link.
5. Click **Disable Relationship**.

   The link appears as a dotted line.

6. Click the **Relationships** view.

   The relationship status is disabled.

Review Changes

When you modify a constraint that affects a table in an entity, you must update the entity to reflect the change.

For example, you create an entity that contains a table called EMP. After you create the entity, you add a constraint relationship between EMP and another table. When you add the constraint, the additional table does not appear in the entity unless you update the entity relationships.

When you view the list of entities in a project, Test Data Manager flags the entity as having a change impact. Open the entity and click **Review Changes** to see the relationship changes that impacted the entity. You can select a relationship and accept the change to update the entity.

Note: After you review and update the entity, you must indicate that you are finished reviewing the changes. In the Entity view, click **Actions** > **Mark Changes Review as Complete**. The Review Changes notification no longer appears for the entity.

The following image shows the **Review Changes** dialog box:

Entity Management

Manage entities to perform data subset operations.

You can create, edit, and delete entities in the **Define | Data Subset** view. You can export and import entities.

Creating an Entity

When you create an entity, you select the driving table. Test Data Manager retrieves the related tables based on the constraints. You can add filter criteria to the driving table to filter source data when you create the data subset database. You can disable relationships between tables in the entity.

Before you create an entity, perform data discovery to identify constraints. Add constraints to the source to define the child tables.

1. Open a project.
2. Click **Define** > **Data Subset**.
3. Click **Actions** > **New** > **Entities**.
4. In the **New Entity** dialog box, enter a name and optional description for the entity.
5. Click **Select Table** to select a driving table from the list.
6. Click **OK**.

   Test Data Manager displays a diagram of the relationship between the tables in the entity. You can view a list of the tables or a list that shows the relationships between the tables.
7. To create a subset of the data based on filter criteria, click **Edit** in the **Properties** pane.

   The **Edit Criteria** dialog box appears.
8. To add criteria for a column, click the **Entity Criteria** tab.
9. Click **Add Criteria**.
10. Select the columns on which you want to filter data and click **OK**.

    The columns that you selected appears in the columns list in the **Entity Criteria** tab.
11. If you want to define the filter expression in the entity, select an operator from the list to filter the data.

    You can indicate a filter condition in the entity, but define the expression in the plan.
12. If you enter a filter condition, enter a specific value or a variable to complete the filter expression.

    Note: If you want to provide a variable in the entity criteria, you must start the parameter value with "$$" for nonrelational sources. For relational sources, you must start the parameter value with "$".
13. Click **Save** for each filter criteria that you create. If you define multiple filters in an entity, the filter conditions act as "AND."
14. Select the rows and click **Save** to exit the **Edit Criteria** dialog box.

Parameter Values in Entity Criteria

A parameter file is an XML file that lists user-defined parameters and their assigned values. You can use parameters when you specify the criteria to create a data subset. Parameter files provide the flexibility to change parameter values each time you run a workflow.

When you use parameters to specify the criteria, you enter a parameter name instead of a specific value.

For example, you want to create a data subset that includes all rows with the EMPID greater than 100. Instead of entering the value as 100 in the value field, you enter the name of the parameter. Enter the parameter value as 100 in the parameter file.

Enter the criteria EMPID > $PARAMETERVALUE. The parameter file must contain an entry $PARAMETERVALUE=100.

Enter the parameter file name and the file path when you configure the plan.

When you use parameters for a string field or a date field on a relational source, you must enclose the value in single quotes in the parameter file.

Database Functions in Entity Criteria

You can use database functions in entity filter criteria. To ensure that TDM processes the filter query correctly, you must use the IN operator and a Select clause.

For example, consider filter criteria that uses the add_months database function. Workflow generation fails if the plan contains the query entered directly, as TDM does not recognize this as a database function. TDM is unable to pass the criteria with the database function as is to the database.

To use the filter criteria with a database function, use the IN operator and enter the query with a Select clause.

To add the required filter criteria, in the **Edit Criteria** dialog box, select the operator IN from the list of operators. Enter the database function as a Select clause in the **Value** field.

```
select HIREDATE from <table> where HIREDATE > add_months(current_date, -3)
```

TDM processes the filter criteria correctly and passes the query to the database. The workflow runs successfully with this entity criteria.

Editing an Entity

You can edit an entity to change the general properties, relationships, and filter criteria for tables.

1. In a project, click **Define | Data Subset**.

   The **Entities** panel shows a list of the entities in the project.
2. Click the entity name to open the entity.
3. Click **Actions** > **Edit**.

   The **Edit** dialog box appears.
4. On the **General** tab, edit the entity name, description, and status.
5. On the **Entity Criteria** tab, select columns and enter filter criteria.
6. Click **Save**.

   A list of any plans that the entity is included in appears. To update the changes in the plans, generate and run the plans again.
7. Optional. To download the list of plans in a .csv file, click **Export**.

   The file contains a list of the plans with the description and the project information.
8. Click **Continue**.

Optimizing Relations

By default, all constraint relationships within an entity are major. This can sometimes result in including excess data in the subset, when you run a subset operation. Optimize the relations such that the entity has a minimum number of major relations to ensure the subset operation fetches minimal data while considering all tables in the entity. You can also optimize relations at run time in a plan. When you optimize relations at run time, you do not save the changes to the entity.

The entity must have a criteria before you can optimize relations in the entity. TDM generates an error if you optimize relations without assigning criteria at the entity level.

Perform the following steps to optimize relations in an entity:

1. Open the entity.
2. Click **Actions** > **Optimize Relations**.

   TDM analyzes the constraints in the entity and proposes major relationships or minor relationships based on analysis. The **Optimize Relationships** page displays the proposed relationship between each parent table and child table along with the current relationship.
3. Optional. Click **Download Optimizations** to download and save the optimization information in a CSV file.
4. Click **Optimize** to accept all the proposed relationships. You cannot apply results selectively .

   A list of any plans that the entity is included in appears. To update the changes in the plans, generate and run the plans again.
5. Optional. To download the list of plans in a .csv file, click **Export**.

   The file contains a list of the plans with the description and the project information.
6. Click **Continue**.

   The entity is updated with the optimized relations. You can manually edit any required constraints even after you save the changes.

Exporting an Entity

You can export an entity to an XML file and import the XML file to another TDM repository.

1. Click **Projects** to open the **Projects** view.

   A list of projects appears.
2. Open the project that contains the entity to export.
3. Click the **Define** view.

   The list of entities, groups, and templates in the project appear.
4. Select an entity to export.

   Use the checkbox to select the entity.
5. Click **Actions** > **Export**.
6. Choose the name and path of the XML file to create.

   The default name is a string that contains "Entity_" and the current date and the time.
7. Click **OK**.

Importing an Entity

You can import an entity from an XML file that was exported from another TDM repository.

1. To open the **Projects** view, click **Projects**.
2. Click the project description to select a project to import the entity into.

   Do not open the project.
3. Click **Actions** > **Import**.
4. Browse for the XML file that contains the entity to import.

   The XML file has a default name similar to `Entity_130315081854.xml`.
5. To import the entity, click **Finish**.

Deleting an Entity

If you do not use an entity, you can delete the entity. If you delete an entity that is assigned to a plan, the plan is not valid.

1. In the **Define | Data Subset** view in the project, click an entity name to select it.

   Do not open the entity.
2. Click the **Actions** > **Delete**.
3. In the **Delete Entity** message box, click **OK**.

   A list of any plans that the entity is included in appears. To update the changes in the plans, generate and run the plans again.
4. Optional. To download the list of plans in a .csv file, click **Export**.

   The file contains a list of the plans with the description and the project information.
5. Click **Continue**.

Groups

A group defines one or more unrelated tables that you want to copy to a subset database. Create a group to add unrelated tables to a plan, or to copy unfiltered data to a target.

When you add tables to a group, you can include residual tables in the group. A residual table is a table that has not been added to a group or entity. Select residual tables when you want to include all the tables in a source in a data subset plan.

When you edit a group, you can add or remove tables from the group.

You create, edit, and delete groups in the **Define** view of an application.

Creating a Group

To create a group, select the tables that you want to add to a data subset plan.

1. On the **Define | Data Subset** view in the project, click **Actions** > **New** > **Groups**.
2. In the **Create Group** dialog box, enter a name for the group and optional description for the group. Click **Next**.
3. To select one or more tables for the group, click **Add Tables**.
4. Optionally, you can filter the list of tables to search for.
5. Click **Save**.

Editing a Group

You can edit a group to change its description, tables, and tables.

1. In a project, click **Define | Data Subset**.

   The **Groups** panel shows a list of the groups in the project.
2. Click the group name to open the group.
3. Click **Actions** > **Edit**.

   The **Edit** dialog box appears.
4. On the group **General** tab, edit the group name or the description.
5. On the **Tables** tab, change the tables in the group.
6. Click **Save**.

   A list of any plans that the group is included in appears. To update the changes in the plans, generate and run the plans again.
7. Optional. To download the list of plans in a .csv file, click **Export**.

   The file contains a list of the plans with the description and the project information.
8. Click **Continue**.

Copying a Group

You can create a group by copying a group. When you copy a group, Test Data Manager copies the tables in the original group to the new group.

1. Open a project and click **Define** > **Data Subset** to view a list of objects in the project.

2. Click a group description to select the group.

   Do not open the group.
3. Click **Actions** > **Duplicate**.

   The **Copy <Group Name>** dialog box appears.
4. Change the name and description of the group. Click **Save**.

Exporting a Group

You can export a group to an XML file and import the XML file to another TDM repository.

1. Click **Projects** to open the **Projects** view.

   A list of projects appears.
2. Open the project that contains the group to export.
3. Click the **Define** view.

   The list of entities, groups, and templates in the project appear.
4. Select group to export.

   Use the checkbox to select the entity.
5. Click **Actions** > **Export**.
6. Choose the name and path of the XML file to create.

   The default name is a string that contains "Group_" and the current date and the time.
7. Click **OK**.

Importing a Group

You can import a group from an XML file that was exported from another TDM repository.

1. To open the **Projects** view, click **Projects**.
2. Click the project description to select a project to import the group into.

   Do not open the project.
3. Click **Actions** > **Import**.
4. Browse for the XML file that contains the group to import.

   The XML file has a default name similar to `Group_130315081854.xml`.
5. To import the group, click **Finish**.

Deleting a Group

If you do not use a group in a project, you can delete it.

1. In a project, click **Define** | **Data Subset**.

   The **Groups** panel shows a list of the groups in the project.
2. Click a group name to select a group.

   Do not open the group.
3. Click **Actions** > **Delete**.

4. In the **Delete Group** message box, click **OK**.

   A list of any plans that the group is included appears. To update the changes in the plans, generate and run the plans again.

5. Optional. To download the list of plans in a .csv file, click **Export**.

   The file contains a list of the plans with the description and the project information.

6. Click **Continue**.

Templates

A template is a container for data subset groups and entities. A template must contain at least one entity.

Use a template to add multiple groups and entities to a plan or to add groups and entities to multiple plans. You can add groups and entities to more than one template.

You create, edit, and delete templates in the **Define | Data Subset** view in an project.

Template Example

You want information from multiple tables about different employees. You create a template that gets information about an employee. The template contains an entity that uses multiple tables to gather information about an employee with ID number 10. The template also contains a group that contains unrelated configuration tables.

After you create the template, you can change it to get information about a different employee. For example, you can change the employee to ID number 25 and get the same information about employee ID 25.

Creating a Template

Create a template to contain the entities and groups to add to a data subset plan. You can map template parameters to the entity parameters.

1. On the **Define | Data Subset** view in the project, click **Actions** > **New** > **Templates**.
2. In the **Create Template** dialog box, enter a name for the group and optional description for the group.
3. In the **Create Template** dialog box, enter a name for the template and optional description for the template. Click **Next**.
4. Click **Add Entities** to add entities to the template.
5. On the **Add Entities** dialog box select the checkbox of the entities you want to add. Click **Next**.
6. Click **Next**.

   The template parameter dialog box appears. The right panel of the dialog box shows the parameters for entities.

7. To add a template parameter, click **Add**.

   The template parameters dialog box appears.

8. Enter the template parameter name, description, and datatype.
   a. Add a parameter for each of the column datatypes in the main table. For each parameter that you want to add, enter a name, description, and the datatype.

      Each parameter that you add appears in the list.
   b. To assign parameters to columns in the main table, link the template parameter to the entity parameter. Select the template parameter, select an entity parameter, and click the link button.
9. Click **Next** to add Groups to the template or click **Finish** to exit the template wizard.
10. If you chose to enter groups, select one or more groups, and click **Finish**.

Editing a Template

Edit a template to add or remove entities, groups, and parameters.

1. On the **Define** | **Data Subset** view in the project, click a template to open it.
2. Click **Edit**.

   The Edit template dialog box displays the **General**, **Entities**, **Parameter Map**, or **Group**.
3. Click a tab, update the template and click **Save**.

Deleting a Template

You can delete template if it is not assigned to a plan.

To delete a template, first remove it from the assigned plan.

1. On the **Define** | **Data Subset** view in the project, select a template.
2. Click **Delete**.
3. In the **Delete Confirmation** message box, click **Yes**.

Data Subset for XSD Sources

You can create a data subset from XML sources. TDM creates a data subset that contains complete XML objects. The subset operation includes all child elements of any parent element that the subset includes based on the filter criteria.

To create a data subset, apply the subset criteria to one or more elements or attributes in the XSD file and define the scope of the criteria. Criteria determines the filter that you apply to an element or attribute. You can apply one filter for each hierarchy.

The scope determines the level at which you apply the filter. The level at which you apply the criteria affects the outcome of the data subset operation.

After you apply the criteria and define the scope of the criteria, add the XSD file as the subset component to a plan. Generate and run a workflow from the plan. The workflow creates and saves a subset of the source data into the target directory.

Data Subset for XSD Sources Task Flow

To define a data subset, import the XSD files into the project and then apply the criteria and the required scope of the criteria. Create a data subset plan and add the XSD files to the plan. To create the data subset, generate and start a workflow from the plan.

Perform the following high-level tasks to create the data subset:

1. Create a project.
2. Import the XML source definition from the PowerCenter repository.
3. Open the XSD file from within the project and apply the criteria and scope of the criteria to an element or attribute.
4. Create a data subset plan and add the XSD file to the plan. You can add multiple XSD files to a plan. Define parameters to create the data subset. For example, you can choose to format the data in the file, or to create the file with unformatted data.
5. Generate a workflow from the plan.
6. Start the workflow.
7. Monitor the progress of the workflow.
8. In the target file directory, verify that the data subset contains the correct data.

Applying Criteria to an Element or Attribute

Apply subset criteria and specify the scope of the criteria to create a data subset from an XSD source. The scope and criteria determine the output of a data subset operation. You can apply criteria to elements and attributes in an XSD file.

Import the XML source definition from the PowerCenter repository into the project before you apply the criteria.

1. Open a project and click **Discover** > **Files**.

   A list of XSD files in the project appears.
2. To open the required file in a separate page, select the file.
3. Select the element or attribute on which you want to apply the subset criteria.
4. Click **Edit** on the **Criteria** tab of the **Details** tab.

   The **Edit Criteria** window opens.
5. Select the required operator from the list of operators.
6. Enter the value or set of values for the criteria.

   You cannot enter a parameter.
7. Select the scope from the list of options and click **OK**.

   The criteria appears in the **Criteria** tab. To clear criteria, select **Criteria** from the list and click the **Clear** button.

Add the XSD file with the scope and criteria assignment to a plan. Run the plan and workflow to create a data subset with the required data.

Example - Data Subset for XSD Data Sources

A product test team maintains its test data in a master XML file. To test specific features of the product, different test groups require specific test data. Test group A requires data on a specific skill set Java. Test group B requires data on a specific location BLR.

To create different test data files for each test group, the test team uses TDM to create data subsets with the required information.

The master data file contains records of employee and contractor information including name, department, address, office location, and skill set.

The XML source definition file (XSD) defines the structure of the master file and data subset files that you create from the master data file. The XSD file defines the Street, Zip, City, and State elements as child elements of the element Address. The elements Name, Skill, and Address are child elements of the elements Employee and Contractor. Address, Employee, and Contractor are child elements of the element Department.

Ancestor and child relations affect the outcome of a data subset operation.

The following image shows the **Criteria** tab on the file page in Test Data Manager:

To create a data subset that contains employee information for the skill set Java, apply the criteria to the element Skill on the Employee branch. Select the operator as equals. Enter the value on which to filter the data as JAVA. Define the scope as Employee.

Create and run a plan that includes the XSD file and the criteria and scope of the data subset.

The subset operation applies the filter on the element Employee and includes all employees with skills that match the criterion. The data subset file that the workflow creates contains data on all employees with Java skills. Department is an ancestor element of Employee. The subset therefore includes all departments that have employees with Java skills and all child elements of Departments.

Edit the data subset criteria and then run the plan again to create a data subset with the data that test group B requires. Delete the criteria on the element Skill.

To create a data subset of departments located in BLR, apply the filter criteria to the element City. Select the operator as equals. Enter the value on which to filter the data as BLR. Define the scope as Department.

The subset operation applies the filter on the element City and includes all departments with addresses that have cities that match the criterion. The data subset file that the workflow creates contains data on all departments with the city as BLR. The subset includes all child elements of the element Department.

Example XML Structure Definition

Consider the following XSD structure of the master test data file:

```
<?xml version="1.0" encoding="utf-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
    <xsd:element name="Skill" type="xsd:string"/>
    <xsd:element name="Name" type="xsd:string"/>
    <xsd:element name="Street" type="xsd:string"/>
    <xsd:element name="State" type="xsd:string"/>
    <xsd:element name="City" type="xsd:string"/>
    <xsd:element name="Zip" type="xsd:string"/>
    <xsd:attribute name="id" type="xsd:integer"/>
    <xsd:attribute name="name" type="xsd:string"/>
    <xsd:attribute name="location" type="xsd:string"/>

    <xsd:element name="Address">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Street" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Zip" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="City" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="State" minOccurs="1" maxOccurs="1"/>
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="Employee">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Name" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Address" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Skill" minOccurs="1" maxOccurs="unbounded"/>
            </xsd:sequence>
        <xsd:attribute ref="id"/>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="Department">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Address" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Employee" minOccurs="1" maxOccurs="unbounded"/>
                <xsd:element ref="Contractor" minOccurs="1" maxOccurs="unbounded"/>
            </xsd:sequence>
            <xsd:attribute ref="name"/>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="Departments">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Department" minOccurs="1" maxOccurs="unbounded"/>
            </xsd:sequence>
            <xsd:attribute ref="location"/>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="Contractor">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Name" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Address" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Skill" minOccurs="1" maxOccurs="unbounded"/>
            </xsd:sequence>
            <xsd:attribute ref="id"/>
        </xsd:complexType>
    </xsd:element>
</xsd:schema>
```

Example Master XML Data

Consider the following master test data file:

```
?xml version="1.0" encoding="US-ASCII" standalone="no"?>
<Departments location="INDIA" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <Department name="QA">
    <Address>
      <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
    </Address>
    <Employee id="101">
      <Name>RAMA</Name>
      <Address>
        <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>SELENIUM</Skill>
    </Employee>
    <Employee id="102">
      <Name>KRISHNA</Name>
      <Address>
        <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>JAVA</Skill>
    </Employee>
    <Contractor id="901">
      <Name>RAMESH</Name>
      <Address>
        <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>TESTING</Skill>
    </Contractor>
    <Contractor id="902">
      <Name>PAVAN</Name>
      <Address>
        <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>JAVA</Skill>
    </Contractor>
  </Department>
  <Department name="DEV">
    <Address>
      <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>
    <Employee id="201">
      <Name>ANAY</Name>
      <Address>
        <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
      </Address>
      <Skill>MAPGEN</Skill>
    </Employee>
```

```
    <Contractor id="903">
      <Name>AJEET</Name>
      <Address>
        <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
      </Address>
      <Skill>JAVA</Skill>
    </Contractor>
  </Department>
  <Department name="DEVOPS">
    <Address>
       <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
    </Address>
    <Employee id="301">
      <Name>NARAYAN</Name>
      <Address>
       <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>BUILD</Skill>
    </Employee>
    <Contractor id="904">
      <Name>SHILPA</Name>
      <Address>
         <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>BUILD</Skill>
    </Contractor>
    <Contractor id="905">
      <Name>AJEET</Name>
      <Address>
        <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
      </Address>
      <Skill>JAVA</Skill>
    </Contractor>
  </Department>
  <Department name="DBA">
    <Address>
      <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>
    <Employee id="402">
      <Name>REKHA</Name>
      <Address>
        <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
      </Address>
      <Skill>SQL_SYBASE</Skill>
    </Employee>
    <Employee id="402">
      <Name>RAMESH</Name>
      <Address>
        <Street>VBIT</Street>
      <Zip>500081</Zip>
```

```
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>ORACLE</Skill>
    </Employee>
    <Contractor id="907">
      <Name>RAMANA</Name>
      <Address>
        <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
      </Address>
      <Skill>JAVA</Skill>
    </Contractor>
    <Contractor id="906">
      <Name>VASU</Name>
      <Address>
        <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
      </Address>
      <Skill>TERADATA</Skill>
    </Contractor>
  </Department>
  <Department name="HR">
    <Address>
       <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>
    <Employee id="501">
      <Name>SURENDRA</Name>
      <Address>
         <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
      </Address>
      <Skill>RECRUIT</Skill>
    </Employee>
    <Contractor id="907">
      <Name>LAVANYA</Name>
      <Address>
         <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
      </Address>
      <Skill>RESOURCE</Skill>
    </Contractor>
  </Department>
</Departments>
```

Data Subset for Group A

Run the plan the first time with the criteria to create a data subset that meets the requirements of Group A.

The output of the subset operation includes the following data:

```
<?xml version="1.0" encoding="US-ASCII" standalone="no"?>
<Departments location="INDIA" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <Department name="QA">
    <Address>
      <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
```

```
      </Address>

      <Employee id="102">
        <Name>KRISHNA</Name>
        <Address>
          <Street>VBIT</Street>
          <Zip>500081</Zip>
          <City>HYD</City>
          <State>TELANGANA</State>
        </Address>

        <Skill>JAVA</Skill>
      </Employee>

      <Contractor id="901">
        <Name>RAMESH</Name>
        <Address>
          <Street>VBIT</Street>
          <Zip>500081</Zip>
          <City>HYD</City>
          <State>TELANGANA</State>
        </Address>

        <Skill>TESTING</Skill>
      </Contractor>
      <Contractor id="902">
        <Name>PAVAN</Name>
        <Address>
          <Street>VBIT</Street>
          <Zip>500081</Zip>
          <City>HYD</City>
          <State>TELANGANA</State>
        </Address>

        <Skill>JAVA</Skill>
      </Contractor>
    </Department>
  </Departments>
```

Data Subset for Group B

Run the plan again with the criteria to create a data subset that meets the requirements of Group B.

The output of the subset operation includes the following data:

```
<?xml version="1.0" encoding="US-ASCII" standalone="no"?>
<Departments location="INDIA" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <Department name="DEV">
    <Address>
      <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>

    <Employee id="201">
      <Name>ANAY</Name>
      <Address>
        <Street>INFA</Street>
        <Zip>500030</Zip>
        <City>BLR</City>
        <State>KARNATAKA</State>
      </Address>

      <Skill>MAPGEN</Skill>
    </Employee>

    <Contractor id="903">
      <Name>AJEET</Name>
      <Address>
```

```
      <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>

    <Skill>JAVA</Skill>
  </Contractor>
</Department>
<Department name="DBA">
  <Address>
    <Street>INFA</Street>
    <Zip>500030</Zip>
    <City>BLR</City>
    <State>KARNATAKA</State>
  </Address>

  <Employee id="402">
    <Name>REKHA</Name>
    <Address>
      <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>

    <Skill>SQL_SYBASE</Skill>
  </Employee>
  <Employee id="402">
    <Name>RAMESH</Name>
    <Address>
      <Street>VBIT</Street>
      <Zip>500081</Zip>
      <City>HYD</City>
      <State>TELANGANA</State>
    </Address>

    <Skill>ORACLE</Skill>
  </Employee>

  <Contractor id="907">
    <Name>RAMANA</Name>
    <Address>
      <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>

    <Skill>JAVA</Skill>
  </Contractor>
  <Contractor id="906">
    <Name>VASU</Name>
    <Address>
      <Street>INFA</Street>
      <Zip>500030</Zip>
      <City>BLR</City>
      <State>KARNATAKA</State>
    </Address>

    <Skill>TERADATA</Skill>
  </Contractor>
</Department>
<Department name="HR">
  <Address>
    <Street>INFA</Street>
    <Zip>500030</Zip>
    <City>BLR</City>
    <State>KARNATAKA</State>
  </Address>
```

```
    <Employee id="501">
      <Name>SURENDRA</Name>
      <Address>
        <Street>INFA</Street>
        <Zip>500030</Zip>
        <City>BLR</City>
        <State>KARNATAKA</State>
      </Address>

      <Skill>RECRUIT</Skill>
    </Employee>

    <Contractor id="907">
      <Name>LAVANYA</Name>
      <Address>
        <Street>VBIT</Street>
        <Zip>500081</Zip>
        <City>HYD</City>
        <State>TELANGANA</State>
      </Address>

      <Skill>RESOURCE</Skill>
    </Contractor>
  </Department>
</Departments>
```

Data Subset Plans and Workflows

A data subset plan contains the XSD files, or the entities, groups, and templates that represent the tables and the data that you want to load to the subset database. Generate a workflow from the plan to create PowerCenter mappings to run the subset operation.

When you start a workflow for the plan, the data subset operation starts.

CHAPTER 7

Data Masking

This chapter includes the following topics:

- Data Masking Overview, 119
- Data Masking Components, 120
- Data Masking Task Flow, 120
- Masking Rules Management, 121
- Masking Rule Assignments, 122
- Data Masking for XSD Sources, 127
- Data Masking Plans and Workflows, 128
- Masking Components in PowerCenter, 129

Data Masking Overview

Use data masking to replace source data in sensitive columns with realistic test data for nonproduction environments. When you create data masking rules, you define the logic to replace sensitive data. To configure the sensitive columns that you want to mask, assign data masking rules to source columns, data domains, and policies.

A policy defines the data masking rules, the data to mask, and the masking parameters for a source. When you assign data masking rules to policies, you can assign multiple source columns to the data masking rules. You can also assign a rule directly to a source column. You can assign data masking rules based on the data type of the source columns.

To implement data masking, create a data masking plan and generate a workflow from the plan. A data masking plan can contain policies and rules. In a data masking plan, select a rule that is assigned to a column. Policies and rules define how to mask sensitive and confidential data in a target database. A data masking plan contains at least one rule or one policy.

When you start the workflow, the PowerCenter Integration Service performs the masking operation.

Data Masking Components

To perform data masking operations, assign rules to data domains, policies, and columns. Use data domains and data discovery to find columns that you want to mask. Create cascades to mask similar columns.

The following table describes the components that you create to implement data masking operations:

Component	Description
Assignments	The allocation of rules to a column to mask the column data. You assign a rule to a column through either a column assignment or a data domain assignment. A column assignment assigns a rule directly to a column in a source. A data domain assignment assigns one or more rules in a data domain to columns in a source.
Column sensitivity	A sensitive column contains sensitive data. Configure column sensitivity to mark columns that you want to mask.
Data domain	An object that represents the functional meaning of a column based on the column data or the column name. Use a data domain to filter the ports that you want to mask when you assign a rule to columns. Define patterns in the data or patterns in the column names when you configure a data domain.
Plan	Defines data masking operations. A data masking plan indicates whether to mask data in place in the source database or in stream in a target database.
Policy	Defines the data masking rules, the data to mask, and the masking parameters for a source.
Rule	Defines the data masking technique, an optional rule qualifier, and masking parameters. A masking technique defines the logic that is used to mask the data. Masking parameters define how a masking technique in a rule masks source data. You can set an override option in a rule that defines whether users can modify the masking parameters for the rule when they assign the rule to columns in a source.
Value cascade	Masks similar columns across tables. You can identify similar columns in a project and configure them to cascade masking rules. Use cascades when some fields are denormalized across multiple tables.

Data Masking Task Flow

To implement data masking operations, assign masking rules to columns in a source. Create a plan and add policies and rules to the plan. Generate a workflow from the plan to mask data in a target database.

Complete the following high-level steps to create the components that you need in a data masking plan:

1. Create rules, data domains, and policies to define the masking techniques and masking parameters.
2. Create a project that contains one or more data sources.
3. Add policies to the project. When you add a policy to a project, the project receives the data domains and the rules that the policy contains.
4. Add additional rules to the project. When you add additional rules to a project, the project receives the masking rules.

5. Assign rules to columns in the source. You can assign rules from a policy or data domain. You can assign the default rules from data domains to multiple columns at a time. Manually assign advanced rules and mapplet rules to columns.
6. Create a plan and add data masking components to the plan. Generate a workflow from the plan. Monitor the progress of the workflow in the **Monitor** view or in the **Plan** view.

Masking Rules Management

Manage data masking rules to perform data masking operations. update the masking rule properties, and change the masking rule assignments.

The **Policies** view shows the masking rules in the TDM repository. After you create a masking rule, you can edit and delete a masking rule from the **Policies** view.

Modify Masking Rules

You can update or delete masking rules after you assign the rules to the target columns.

When you update or delete a masking rule, a warning message appears that contains the list of columns, plans, and data domains that contain the rule assignment. You can export the list of affected objects and save them in a .csv file.

If you update a masking rule, you change the properties within the masking rule. The changes do not take effect in the plans that contain the rules. You must fix the changes in the plan, and generate and run the plan again. If you make changes in the name or type of the field of a custom masking rule, the assignments do not contain the changes. You must import the mapplet again.

If you delete masking rules, TDM does not delete the masking rule from the columns, data domains, and plans that contain the rule assignment. To update the changes in a plan, you must generate and run the plan again.

Editing a Masking Rule

You can edit a masking rule and update the masking rule properties that you want.

1. To access the **Policies** view, click **Policies**.
2. Click the name of the masking rule that you want to edit.

   The rule opens in a tab.
3. Click **Actions** > **Edit**.
4. Edit the parameters that you want to change and click **OK** to save the rule.

   If the masking rule has assignments, the **Impacted Objects** dialog box appears with the list of affected columns, plans, and data domains.
5. To download the list of affected columns and plans, click **Export**, and save the .csv file.
6. To save the changes, click **Continue** .

   To update the changes in a plan, you must generate and run the plan again.

Copying a Masking Rule

You can create a masking rule by copying a masking rule. When you copy a masking rule, Test Data Manager copies the masking rule properties from the original masking rule to the new masking rule.

1. To access the **Policies** view, click **Policies**.
2. Click a masking rule description to select the masking rule.

   Do not open the masking rule.
3. Click **Actions** > **Duplicate**.

   The **Copy <Masking Rule Name>** dialog box appears.
4. Change the name and description of the masking rule. Click **Save**.

Deleting a Masking Rule

You can delete a masking rule before or after you assign the masking rule to a column.

1. To access the Policies view, click **Policies**.
2. Click the box to the left of the generation rule that you want to delete.
3. Click **Actions** > **Delete**.

   The **Delete Masking Rule** dialog box appears with a warning message. If the masking rule has assignments, the **Impacted Objects** dialog box appears with the list of affected columns, plans, and data domains.
4. To delete the masking rule that do not have assignments, click **OK**.
5. To delete the masking rule that have assignments, click **Continue**. To download the list of affected objects, click **Export**, and save the .csv file.

   To update the changes in a plan, you must generate and run the plan again.

Masking Rule Assignments

You can assign a data masking rule to a column in the **Define** | **Data Masking** view. Choose a rule from a list in the **Masking Rule** column in the view. The data domain default rule appears at the top of the list when you click the **Masking Rule** column.

The following image shows the **Define** | **Data Masking** view in a project:

The following table describes the fields in the **Data Masking** view:

Column	Description
Name	Name of the table.
Owner	Name of the database schema.
Columns	Name of the column to mask.
Data Type	Data type of the column to mask.
Domain	Name of the domain that you assigned to the column either from a data domain discovery or a manual assignment.
Sensitive	Indicates if the column is a sensitive column. Value is Yes or No.
Similar Value Columns	Indicates that the column is configured with other columns in a cascade. The column shows the number of other columns in the cascade.
Policy	Policy name that the column is assigned to.
Masking Rule	The rules to apply to the column. When you click inside the Masking Rule column, you can choose which rule to apply to the column. A preferred rule from a data domain has an asterisk (*) before the name.
Override	Shows the override property status for a rule. If the property is Yes, you can override the rule properties when you assign the rule to a column. When you override the rule parameters for a column the **Override** column value is Yes-Overridden.

Available Data Masking Rules

The **Data Masking** view lists the rules that you can assign to each column of the data source. You can assign one or more rules for a column if the masking rules are available for the column.

If no rules appear for a column when you click **Masking Rule**, check for the following situations:

- The project has no rules for the column datatype.
- The project does not contain policies with masking rules.
- A data domain that has a preferred rule that you need to use is not in a policy that you assigned to the project.

The PowerCenter Integration Service does not mask integers greater than 28 characters.

If a column has a data domain assignment, the preferred rules for the data domain appear at the top of the list in the **Masking Rule** column. If you forget to add a rule to a data domain, you can assign rules from policy preferred rules. The policy preferred rules appear below the data domain rules in the rules list.

You can apply the default data domain rules to multiple columns at a time. Select multiple columns and click **Rule Assignment**. You can choose to assign the default rules in the **Rule Assignment** dialog box.

Note: In the **Data Masking** view, you must click the **Save** icon after each time you assign a rule. If you do not save the rule assignment before you assign another rule, the Test Data Manager discards the rule assignment.

When you select mapplet rules, advanced rules, or any rule that uses multiple columns, you must configure a mapping between the source ports and the rule ports. The Test Data Manager notifies you when you need to configure the source to rule port mapping.

Assigning Masking Rules to Columns

Assign masking rules from data domains or policies to one or more columns in the project source that you want to mask.

1. In the project, click **Define | Data Masking** to access the **Data Masking** view.
2. Select a column to assign a masking rule to.
3. If the **Domain** is blank for the column, click the **Policy** column and choose a policy that contains the data masking rule that you want to assign.
4. Click inside the **Rule** column to view the list of available rules.

   The data domain preferred rules appear at the top of the list. The other rules in the policy appear at the bottom of the list.
5. Select a masking rule. If you choose substitution masking rule, you can specify the rule assignment parameters. If you choose mapplet or advanced rule, you can assign the rule columns to the columns in the project.
6. Click **Save** for each column that you update.
7. If you want to update a masking rule assignment, select another masking rule.

   If the masking rule contains assignments, the **Impacted Objects** dialog box appears with the list of affected plans, columns, and data domains.
8. To download the list of affected columns and plans, click **Export**, and save the .csv file.
9. To save the changes, click **Continue**.

   To update the changes in a plan, you must generate and run the plan again.

Overriding a Masking Rule

You can override rule parameters when you assign the masking rule to a column. The rule must have the override property enabled.

1. In the **Define | Data Masking** view, select a column that has a masking rule with the override property enabled.

   The **Override** column value is Yes.
2. Click **Actions** > **Override**.

   The **Edit <Rule Type>** dialog box appears and shows the rule parameters.
3. Change the rule parameters and the exception parameters as required.
4. Click **Save**.

   Test Data Manager adds an Overridden flag to the **Override** column when you override the rule.

Assigning Multiple Masking Rules to a Column

You can assign multiple masking rules to a column. To assign more than one rule, create a duplicate row in the **Define | Data Masking** view and update the rule.

1. Open the **Define | Data Masking** view for a project.
2. Select a column that you want to assign multiple rules to.

   Use the checkbox to select the column.
3. Click **Rule Assignment** .

The **Rule Assignments** dialog box appears.

4. Select the column that you want to assign multiple policies to.

   Use the checkbox to select the column.

5. Click **Copy Rule Assignment**.

   The Test Data Manager creates a row that is a duplicate of the row you selected.

6. Change the policy to view more preferred masking rules.
7. In the **Masking Rule** column, select a rule from the list.
8. Click **Save**.

   The **Define | Data Masking** view shows two rows for the same column. Each column has a different rule assignment.

Assigning Default Values to Multiple Columns

You can assign the default data domain rules to multiple columns at the same time.

1. Open the **Define | Data Masking** view for a project.
2. Select the columns that you want to assign default values to.
3. Click **Rule Assignment**.

   The **Rule Assignments** dialog box appears.

4. Select which columns you want to update with the default values. You can select the checkbox at the top of the dialog box to select all rows.
5. Click **Default Assignments**.

   The Test Data Manager updates each column with the default rule.

6. Click **Save**.

Assigning Custom Masking Rules to Columns

Mapplet rules can require values from multiple ports. You must map each source input column to a rule input and output port that you configure in the mapplet rule.

1. In the **Define | Data Masking** view, click the **Masking Rule** column for the column that requires a mapplet rule.
2. Select the mapplet rule from the list.

   The **Custom Rule Assignment** dialog box appears with the list of columns in the table.

3. Select a source input column in the left panel and a rule input port in the right panel.

   When you select a source input column, the rule ports with same data type become available for mapping.

4. Click the **Link** icon to create a mapping.

   An arrow appears that links the source port to the rule port.

   Note: To create a mapping, you can also click **Show Simple** and map the ports.

5. If you need to remove the link, select the source input port and the rule input port. Click the **Unlink** icon.

   The arrow between the ports disappears.

6. After you map the rule input ports, click **Next**.
7. Select a rule output port in the left panel and a source input column in the right panel.

8. Click the **Link** icon to create a mapping. To remove a mapping, click the **Unlink** icon.
9. Click **Save**.

Assigning Advanced Masking Rules to Columns

Assign an advanced masking rule to a column and map source input columns to the rule input and output ports. To perform expression cascade, you can add another table, select columns from the table, and map the input columns. An expression cascade joins the two tables to generate a combined masked output.

1. In the **Define | Data Masking** view, click the **Masking Rule** column for the column that requires an advanced rule.
2. Select the advanced rule from the list.

   The **Advanced Rule Assignment** dialog box appears with the list of columns in the table.
3. To select columns from another table, click the **Add Associated Table** icon.

   The **Add Associated Table** dialog box appears.
4. Select an associated table and click **OK**.

   The table name appears with a list of column.
5. Select a source input column in the left panel and a rule input port in the right panel.

   Note: When you select a source input column, the rule ports with same data type become available for mapping.
6. Click the **Link** icon to create a mapping.

   An arrow appears that links the source port to the rule port.

   Note: To create a mapping, you can also click **Show Simple** and map the ports.
7. If you need to remove the link, select the source input port and the rule input port. Click the **Unlink** icon.

   The arrow between the ports disappears.
8. After you map the rule input ports, click **Next**.
9. Provide a join condition for the two source input tables to perform expression cascade.
10. Click **Next**.
11. Select a rule output port in the left panel and a source input column in the right panel.
12. Click the **Link** icon to create a mapping. To remove a mapping, click the **Unlink** icon.
13. Click **Save**.

Deleting Masking Rule Assignments

You can delete a masking rule assignment to a column.

1. In the **Define | Data Masking** view, select the column that you need to delete the rule assignments for.
2. Click **Clear Assignment**.

   Test Data Manager prompts you to confirm deleting the assignment.
3. To delete the rule assignment from the column, click **OK**.

   If a plan contains the masking rule assignment, the **Impacted Objects** dialog box appears with the list of affected columns, plans, and data domains.

4. To delete the masking rule assignment, click **Continue**. To download the list of affected objects, click **Export**, and save the .csv file.

   To update the changes in a plan, you must generate and run the plan again.

Data Masking for XSD Sources

You can apply data masking rules and policies to XML elements and attributes to mask XML data.

View and open the XSD file from the **Discover | Files** view of the project. You can view the details of the XML elements and attributes in the XSD map view and the grid view.

Select an element, and apply a data masking rule to the element. You can select multiple elements and assign a default masking rule to the selected elements. You can remove all the default masking rule assignments, and assign another masking rule.

You can mask similar objects in the XSD file. Select multiple objects and set them as similar value objects.

You can assign standard, mapplet, and advanced masking rules to the elements in XSD sources. You cannot use shuffle masking technique for XSD sources.

Data Masking Task Flow for XSD Sources

To mask data from an XSD source, you apply the masking rules to the XML elements and attributes.

Perform the following high-level tasks to mask the elements in an XSD data source:

1. Create a project.
2. Import the XML source definition from the PowerCenter repository.
3. Add policies and additional rules to the project.
4. Open the XSD file from the project.
5. Select an element or an attribute. Specify a masking rule for a data domain or apply masking rules individually.
6. Create a plan and add the masking components that you assigned in the XSD file. Define the data masking parameters, and specify the target directory.
7. Generate a workflow from the plan.
8. Run the workflow, and monitor the progress of the workflow.
9. Verify that the data is masked in the target.

Applying Data Masking Rules to XML Elements and Attributes

Apply data masking rules to XML elements and attributes to mask data from an XML source.

1. In a project, click **Discover** > **Files**.

   A list of XSD files appears in the project.
2. Click the XSD file.

   The XSD file opens in a separate tab.
3. Select the elements and attributes on which you want to apply the masking rule.
4. In the **Details** pane, click **Edit** in the **Data Masking** section.

5. Assign a policy and a data masking rule to the selected element or attribute.

   The following image shows a sample data masking rule assignment to an XML element:

6. To save the assignment, click **OK**.
7. To configure value cascades, click **Set as Similar Value Objects**.

   The **Set as Similar Value Elements and Attributes** dialog box appears.
8. Select the driving object and click **OK**.
9. To assign default rules to multiple objects, click **Default Assignment**.
10. To clear rule assignments, click **Clear Assignment**.

Data Masking Plans and Workflows

A data masking plan contains columns and policies that represent the tables and data that you want to mask. The plan also contains the rules to use to mask the data. A data masking plan contains at least one port with an assigned rule or one policy. You generate a workflow from a plan to perform the masking operations.

When you start a workflow for the plan, the data masking operation starts.

Related Topics:

- "Creating a Data Masking and Data Subset Plan" on page 238
- "Generating a Workflow" on page 242

Masking Components in PowerCenter

When you generate and run a workflow from a data masking plan, the PowerCenter Integration Service creates the masking components that are required to run a data masking workflow. The PowerCenter Integration Service stores these components in the PowerCenter repository.

The PowerCenter Integration Service creates the following components to perform the data masking operations:

- Mappings
- Transformations
- Mapplets
- Sessions and workflows

Important: If you edit the components that are generated from Test Data Manager in the PowerCenter Client, the session might generate unexpected results.

Mapplet Export

To implement PowerCenter masking techniques that are not available in Test Data Manager, you can create and export masking mapplets from PowerCenter to an XML file. You can import mapplets as rules into Test Data Manager.

Include one or more Data Masking transformations in the mapplet. The mapplet can be a single or multiple port mapplet. The transformations must be passive transformations. For a single input and output port mapplet, you can configure the same port as both the input and output port, and you can complete this action for multiple ports at a time when you assign a mapplet that you imported as a rule to columns in a project.

When you assign a mapplet to a columns, assign the source and target columns to the mapplet. The TDM Server applies the PowerCenter transformation within the mapplet to the columns. You can also export rules as PowerCenter mapplets.

CHAPTER 8

Data Masking Techniques and Parameters

This chapter includes the following topics:

- Data Masking Techniques and Parameters Overview, 131
- Data Masking Techniques, 131
- Data Masking Parameters, 132
- Custom Masking, 135
- Advanced Masking, 135
- Credit Card Masking, 137
- Email Masking, 138
- Encryption Masking, 140
- Expression Masking, 140
- IP Address Masking, 141
- Key Masking, 142
- Nullification Masking, 145
- Phone Masking, 145
- Random Masking, 145
- Shuffle Masking, 150
- SIN Masking, 151
- SSN Masking, 151
- Substitution Masking, 152
- URL Masking, 153
- Name Substitution Example , 153
- Shuffle Address Example, 159

Data Masking Techniques and Parameters Overview

A data masking technique is the type of data masking to apply to a selected column. The masking parameters are that options that you configure for the technique.

The type of masking technique that you can apply depends on the datatype of the column that you need to mask. When you choose a masking technique, Test Data Manager displays parameters for the masking technique.

You can restrict the characters in a string to replace and the characters to apply in the mask. You can provide a range of numbers to mask numbers and dates. You can configure a range that is a fixed or percentage variance from the original number.

You can configure different masking parameters for a masking technique and save each configuration as a masking rule. The PowerCenter Integration Service modifies source data based on masking rules that you assign to each column. You can maintain data relationships in the masked data and maintain referential integrity between database tables.

Data Masking Techniques

You can apply masking techniques based on the source datatype that you configure for a column. For example, if the column datatype is numeric, you can define a masked value that is within a fixed or percent variance from the original value.

The PowerCenter Integration Service does not mask integers greater than 28 characters.

The following table describes the masking techniques that you can choose when you define a rule:

Masking Technique	Description
Advanced	Applies masking techniques to multiple input and output columns. You can create an expression to combine multiple columns. You can mask all datatypes.
Credit card	Applies a built-in mask format to disguise credit card numbers. You can mask the string datatype.
Email address	Applies a built-in mask format to disguise email addresses. You can mask the string datatype.
Encryption	Applies encryption to the source data. You can assign an encryption method to mask the data with. You can encrypt string data types.
Expression	Applies an expression to a column to create or disguise data. Use expression masking if the expression references one column. If the expression includes multiple columns, use the calculated masking technique with advanced masking. You can mask all datatypes.
IP address	Applies a built-in mask format to disguise IP addresses. You can mask the string datatype.
Key	Produces deterministic results for the same source data, masking rules, and seed value. You can mask date, numeric, and string datatypes.
Mapplet	Applies masking rules from a PowerCenter mapplet. The mapplet contains the logic to mask the input columns and return data to the target. A mapplet can have multiple input and output columns.

Masking Technique	Description
Nullification	Replaces a column of data with a null value. You can mask all datatypes.
Phone	Applies a built-in mask format to disguise phone numbers. You can mask the string datatype.
Random	Produces random, non-repeatable results for the same source data and masking rules. You can mask date, numeric, and string datatypes.
Shuffle	Applies sensitive column values from one row to another row in the same table. You can restrict which rows to shuffle data from. You can mask date, numeric, and string datatypes.
SIN	Applies a mask to Social Insurance numbers. You can mask a string datatype.
SSN	Applies a built-in mask format to disguise Social Security numbers. You can mask the string datatype.
Substitution	Replaces a column of data with similar but unrelated data from a dictionary. You can mask date, numeric, and string datatypes.
URL	Applies a built-in mask format to disguise URL data. You can mask a string datatype.

Data Masking Parameters

Configure data masking parameters to define how to apply a data masking technique. The parameters that you configure depend on the data type that you need to mask. Some masking techniques are not available for all data types.

You can configure the data masking properties and the exception handling parameters in the **New Masking Rule** wizard.

The following image shows data masking parameters that appear when you configure a Credit Card data masking rule:

Repeatable Output

Repeatable data masking output returns deterministic values. Use repeatable masking when you generate a data masking workflow more than once and you need to return the same masked values each time it runs.

Configure repeatable output if you have the same value in multiple source tables and you want to return the masked value in all of the target tables. The tables in the target database receive consistent masked values.

For example, customer John Smith has two account numbers, 1234 and 5678. The account numbers are in multiple tables. The PowerCenter Integration Service masks John Smith as Frank Martinez in all the tables. It always masks account number 1234 as 6549 and account number 5678 as 3214.

You can enter a seed value when you configure repeatable output. You can configure a dictionary file with replacement data values for substitution masking. When you configure repeatable output, TDM returns the same value from the dictionary whenever a specific value appears in the source data.

Seed

Apply a seed value to create repeatable output for data masking output. The seed value is a starting point for generating masked values.

You can define a seed value from 1 through 999. The default seed value is 1. Apply the same seed value to a column to return the same masked data values in different source data. For example, you have the same `Cust_ID` column in four tables. If you want all of them to output the same masked values, you can apply the same seed value when you mask each column.

You can create a variable for the seed value in a parameter file. When you configure repeatable output in a masking rule, enter the name of the variable instead of a seed value. You can then change the seed value in the parameter file if you need to change the seed value, instead of editing the masking rule.

The TDM administrator can also set a central seed value for all data masking components with repeatable output. The central seed value overrides any other seed value.

Exception Handling

Data masking exception handling defines options for handling nulls, blanks, empty strings, and errors in source data. You can configure exception handling for each data masking rule that you create. You can specify preprocessing and post processing expression parameters to apply changes before and after masking the data.

The following table describes the exception handling options:

Option	Description
Preprocessing Expression	Optional. Expression to define changes to make to the data before masking. Click **Edit** to configure the preprocessing expression.
Post processing Expression	Optional. Expression to define changes to make to the masked data before saving the data to the target. Click **Edit** to configure the post processing expression.
Null and Empty Spaces	The default behavior for handling null values or empty columns in the source data. Choose one of the following options: - Constant. Mask the data with a constant value. Enter the value to use. - Default server value. Mask the source column with the default value from thedefaultValue.xml file. - Log error and continue. Log an error in the PowerCenter session log and continue processing. - Treat as value. Treat null values or spaces as a valid source value. Mask the space or null value with a valid value. - Ignore. Do not mask the null or empty space. **Note:** For columns with dependent masking, there are no separate rules to handle null and empty spaces. TDM masks the dependent column based on the rules and values of the column that it depends on.
Error Handling	The default behavior for handling errors in the source data. Choose one of the following options: - Constant. Mask the data with a constant value. Enter the value to use. - Default server value. Mask the source column with the default value from thedefaultValue.xml file - Log exception and continue. Log an exception in the PowerCenter session log and continue processing. - Ignore and continue. Do not mask the null or empty space. - Error. Log an error in the PowerCenter session log and stop processing.
Trim Leading or Trailing Spaces	Trims the leading and trailing spaces from source data. When you enable this option the following source fields are the same: " Jones", "Jones", "Jones " .

Custom Masking

Use custom masking when you want to apply multiple masking techniques to a column or when you want to configure multiple input and output columns in the same masking rule. You can mask all data types.

To create a custom masking rule, you must import the PowerCenter mapplets. Set the mapplet input columns and output columns as required. The mapplet must contain at least one input and one output column.

When you assign the custom masking rule to columns, you must map the source columns to the columns in the custom masking rule.

Custom Masking Parameters

Configure input and output columns that you import from a mapplet when you create a custom masking rule.

The following table describes the general properties that you can configure for input and output columns:

Parameter	Description
Column Name	The name of an input or an output column within a mapplet. When you assign the rule to a column, you map the column names in the rule to column names in the database.
Column Type	The column type. The mapplet can contain the following types of columns: - Input. Receives the input data. - Output. Returns the output data.
Data Type	The data type of the column.
Precision	The precision for the column. The maximum number of digits or the maximum number of characters that the column can accommodate. For example, 874.560 has a precision of 6.
Mandatory	Indicates if you must assign the column to a table column in the project. Applies to input and output columns. You must set at least one input and one output column as required.
Group Name	Specifies the group to which the columns belong. The group name can be input, output, or any other name that you provide when you create a mapplet.

Advanced Masking

Use advanced masking when you want to apply multiple masking techniques to a column or when you want to configure multiple input and output columns in the same masking rule. You can mask all datatypes.

When you configure advanced masking, you configure the input columns, output columns, and variable columns. Variable columns are work fields that you can define to temporarily store data.

When you create the columns in the masking rule, the column names do not need to be the same as the column names in the source. When you assign the masking rule to columns, you must map the source columns to columns in the advanced masking rule.

Related Topics:

- "Advanced Masking Rules" on page 57

Advanced Masking Parameters

Configure parameters for each column that you create in an advanced masking rule.

The following table describes the general properties that you can configure for input, output, and variable columns:

Parameter	Description
Column Name	The name of an input, output, or variable column. Enter any name. The name does not have to match the name of a column in the source. When you assign the rule to source data in a project, you map the column names in the rule to column names in the database.
Column Type	The column type. You can configure the following types of columns: - Input. Receives the source data. - Variable. A temporary column that contains intermediate values. You can apply masking rules to variable column values in order to mask data before returning data to output columns. - Output. Returns the output data. You can apply an expression or a masking rule to variable column data and return the data in the output column.
Datatype	The datatype of the column.
Precision	The precision for the column. The maximum number of digits or the maximum number of characters that the column can accommodate. For example, 798.650 has a precision of 6.
Scale	Number of digits to the right of the decimal point in a number.
Mandatory	Indicates if you must assign the column to a column in the source. Applies to input and output columns.

The following table describes the masking properties that you can configure for variable columns and output columns:

Parameter	Description
Expression	An expression to apply to the variable column. You can create the expression in the Expression Builder.
Masking Rule	Applies a masking rule to the input column and writes the results in the variable column. You can enter the following parameters. - Condition. Defines whether an input column should be masked or not. If the condition is true, the PowerCenter Integration Service masks the column. - Rule. The data masking rule to apply to the input column. - Override properties. You can change the data masking rule parameters if the rule owner enabled the rule to be overridden. - Input column. The name of the input column to apply the masking rule to. Select an input column from the columns that you added to the rule.

Parameter	Description
Condition Input	Applies an expression to the Output column only. Select a condition from the list. If you applied conditions on the input column, you can use it for the output column.
Dependent	Applies dependent masking. Dependent masking replaces the value of a column based on the values returned from a dictionary row for another column. You must define substitution masking for another column before configuring a column for dependent masking. Enter the following parameter: - Input column. The name of the input column that you configured for substitution masking. - Dictionary column. Choose the dictionary column to replace the dependent column with.

Advanced Masking Example

You can create an expression to combine multiple columns in an advanced masking rule.

Create an expression in the **Expression Builder**. Select columns, functions, variables, and operators to build expressions. The expression can reference input columns and variable columns.

For example, you have a CUSTOMERS table that contains first and last names. You want to mask the names and combine the masked values into a masked full name.

The following table shows the columns that you create in an the advanced masking rule to combine the masked names:

Column Name	Column Type	Masking Technique	Expression
FIRST_NAME	INPUT	–	–
LAST_NAME	INPUT	–	–
FIRST_MASKED	VARIABLE	Substitution on FIRST_NAME	_
LAST_MASKED	VARIABLE	Substitution on LAST_NAME	_
FULL_NAME	OUTPUT	Expression	FIRST_MASKED || ' ' || LAST_MASKED

Mask the FIRST_NAME and LAST_NAME with Substitution. Return FULL_NAME using an expression that combines the FIRST_MASKED and LAST_MASKED columns.

When you create expressions in the Expression Builder, use the point-and-click interface to minimize errors. Verify that the expression returns a value that matches the output column data type.

Credit Card Masking

Credit card masking applies a built-in mask format to disguise credit card numbers. The PowerCenter Integration Service creates a masked number that has a valid checksum. You can choose from multiple credit card number formats. Mask the string datatype with credit card masking.

The PowerCenter Integration Service generates a logically valid credit card number when it masks a credit card number. The length of the source credit card number must be between 13 to 19 digits. The input credit card number must have a valid checksum based on credit card industry rules.

You can apply repeatable masking to credit card numbers. Each credit card type has a different format. You can choose to keep the same credit card type in the masked credit card number, or you can change the credit card type. For example, if the source credit card number is a Visa credit card number, you can configure the rule to return a Visa credit card number. Or, you can configure the rule to return a credit card number that is a different credit card type.

The source credit card number can contain numbers, spaces, and hyphens. If the credit card has incorrect characters or is the wrong length, the Integration Service writes an error to the session log. The PowerCenter Integration Service applies a default credit card number mask when the source data is not valid.

The PowerCenter Integration Service does not mask the six-digit Bank Identification Number (BIN) at the beginning of the number. For example, the PowerCenter Integration Service might mask the credit card number `4539 1596 8210 2773` as `4539 1516 0556 7067`.

Credit Card Masking Parameters

The following table describes the parameters that you can configure for credit card masking:

Parameter	Description
Repeatable	Returns the same masked value when you generate a workflow multiple times, or when you generate masked values for a column that is in multiple tables.
Seed	Starting point for creating repeatable output. Enter a number between 1 and 999. Default seed value is 1.
Keep card issuer	The PowerCenter Integration Service returns the same credit card type for the masked credit card. For example, if the source credit card is a Visa card, generate a masked credit card number that is the Visa format.
Replace card issuer	Replaces the source credit card type with another credit card type. When you enable replace card issuer, select which type of credit card to replace it with. You can choose credit cards such as AMEX, VISA, JCB, and MASTERCARD. Default is ANY.

Email Masking

Email masking generates realistic email addresses. You can apply a mask from the masked values of a first and last name. You can apply a constant domain name or create an expression to define the domain. Mask the string datatype with email masking.

You can configure email masking to mask email addresses the same way each time the same user appears in the source data. You can configure an email address from the masked first and last name columns. You can configure email masking as repeatable between workflows, or you can configure email masking to be repeatable in one workflow.

Note: The PowerCenter Integration Service always returns ASCII characters for an email address.

Email Masking Parameters

When you configure email masking, you can configure parameters to mask the user name and the domain name in the email address. You can specify dictionary files that contain the user names and the domain names. You can define an expression that combines names to create realistic email user names.

Email Masking General Parameters

The following table describes the parameters that define how to mask an email address:

Parameter	Description
Repeatable output	Returns the same masked value when you generate a workflow multiple times, or when you generate masked values for a column that is in multiple tables.
Seed	Starting point for creating repeatable output. Enter a number between 1 and 999. Default is 1.
Standard	Replaces an email string with characters. The standard email masking technique does not create realistic domain or user names.
Advanced	Replaces an email string with name and domain values from a dictionary or mapping.
Columns to form email	Indicates where to retrieve the columns to form the email address. Choose one of the following options: - From mapping. Do not use a dictionary for user names. Assign the column names when you assign the rule to source columns. - From dictionary. Mask user names from a dictionary file.

User Name Masking Parameters

The following table describes the parameters that define how to mask the user name in the email address:

Parameter	Description
Dictionary for columns	The dictionary to use when you choose to form an email from a dictionary column. Choose a dictionary that contains first and last names.
Expression	Defines an expression when you select to create an email from a mapping. You can create an expression to define the user name from parts of the first and last names. The column names in the expression do not have to match the source column names. Assign columns to the expression when you assign rules to columns.
First name column	The dictionary column that contains the first name of the email address.
First name length	Length for the first name. For example, you might choose the first letter of the name, or the first 4 letters of the name. Default is 10.
Delimiter	The delimiter between the first and last names of the email address. Choose a period, underscore, or hyphen.
Last name column	Dictionary column that contains the last name of the email address.
Last name length	The length for the last name. For example, you might choose the first 6 characters of the last name. Default is 10.

Domain Name Masking Parameters

The following table describes the parameters that define how to mask the domain name:

Parameter	Description
Constant	A constant value to mask the email address with. Each email address receives the same domain name.
Random	Indicates whether to use a flat file or relational dictionary.
Domain lookup dictionary	The dictionary file to use from the imported sources.
Domain name column	The column that contains a dictionary name in the domain lookup dictionary.

Encryption Masking

Encryption masking applies encryption algorithms to mask source data. You can choose the algorithm to encrypt the data. Mask string datatypes with encryption masking.

Select from the following encryption algorithms:

AES

Advanced Encryption Standard with 128-bit encoding.

CRC

Cyclic Redundancy Check. Finds data transmission errors or verifies that data is not modified. Computes a checksum.

MD5

MD5 Message-Digest Algorithm. One-way cryptographic hash function with a 128-bit hash value.

To configure encryption masking, enter an encryption key that is 16 characters or less.

Expression Masking

Expression masking applies an expression to a column to mask or change data. Mask all data types with expression masking.

To configure expression masking, you create an expression in the **Expression Builder**.

Select column, functions, variables, and operators to build expressions. The expression can reference input columns and output columns.

In expression masking, you can append additional data to the source column data. For example, you want to mask the first names. The source has the FirstName column and you want to concatenate the values in the FirstName column with the string ABC. When you configure the data masking rule, select the FirstName column and enter the following expression in the expression editor.

```
CONCAT(FirstName,'ABC')
```

When you configure expression masking for a column, the column name appears as the expression by default.

Select functions, columns, variables, and operators from the point-and-click interface to minimize errors when you build expressions.

When you create an expression, verify that the expression returns a value that matches the column data type. The PowerCenter Integration Service returns zero if the return value does not match a numeric column. It returns NULL if the return value does not match a string column.

For information about expression syntax, see the *Informatica Transformation Language Reference*.

Expression Masking Parameters

You can configure parameters for the expression masking technique.

The following table describes the parameters that you can configure for expression masking:

Parameter	Description
Repeatable Output	Determines if the masked value should persist for a given column value. You must enter a seed value when you enable repeatable output.
Alphanumeric Seed	The alphanumeric seed is a key that allows multiple data masking rules to generate the same masked values from the same source values. Define the same seed in each data masking rule that requires the same results for a column. The seed can be any text.
Expression	Accepts PowerCenter expressions and performs a calculation based on values within a single row.

Rules and Guidelines for Expression Masking

Use the following rules and guidelines for expression masking:

- You cannot use the output from an expression as input to another expression. If you manually add the output column name to the expression, you might get unexpected results.
- Select functions, columns, variables, and operators from the point-and-click interface to minimize errors when you build expressions.
- If the data masking rule is configured for repeatable masking, specify the storage table in the data masking plan. If the storage table does not exist, the data masking operation fails.

IP Address Masking

IP address masking applies a built-in mask format to change IP addresses. Mask string datatypes with IP address masking.

Use IP address masking to mask data with the string datatype.

IP masking splits an IP address into four numbers, separated by periods. The first number is the network. The PowerCenter Integration Service masks the network number within the network range.

The PowerCenter Integration Service does not mask the class and private network address. The PowerCenter Integration Service masks a Class A IP address as a Class A IP Address and a 10.x.x.x address as a 10.x.x.x address.

For example, the PowerCenter Integration Service can mask `11.12.23.34` as `75.32.42.52`, and `10.23.24.32` as `10.61.74.84`.

Key Masking

Key masking produces deterministic results for the same source data, masking rules, and seed value. Mask date, numeric, and string datatypes with key masking.

The following table describes the parameters that you can configure for key masking:

Parameter	Description
Seed	A start number that enables the PowerCenter Integration Service to return deterministic data. You can mask the date, numeric, and string datatypes.
Mask Format	The type of character to substitute for each character in the input data. You can limit each character to an alphabetic, numeric, or alphanumeric character type. You can mask the string datatype.
Source String Characters	The characters in the source string that you want to mask. You can mask the string datatype.
Result String Replacement Characters	Substitutes the characters in the target string. You can mask the string datatype.

Mask Format

Configure a mask format to limit each character in the output column to an alphabetic, numeric, or alphanumeric character.

If you do not define a mask format, the PowerCenter Integration Service replaces each source character with any character. If the mask format is longer than the input string, the PowerCenter Integration Service ignores the extra characters in the mask format. If the mask format is shorter than the source string, the PowerCenter Integration Service does not mask the characters at the end of the source string.

Note: The mask format contains uppercase characters. When you enter a lowercase mask character, Test Data Manager converts the character to uppercase.

The following table describes mask format characters:

Character	Description
A	Alphabetical characters. For example, ASCII characters a to z and A to Z.
D	Digits. From 0 through 9.
N	Alphanumeric characters. For example, ASCII characters a to z, A to Z, and 0-9.

Character	Description
X	Any character. For example, alphanumeric or symbol.
+	No masking.
R	Remaining characters. R specifies that the remaining characters in the string can be any character type. R must appear as the last character of the mask.

Source String Characters

Configure source string characters to choose the characters that you want to mask.

For example, if you set the number sign (#) as a source string character, it is masked every time it occurs in the input data. The position of the characters in the source string does not matter, and you can configure any number of characters. If you do not configure source string characters, the masking replaces all the source characters in the column.

The source characters are case sensitive. The PowerCenter Integration Service does not always return unique data if the number of source string characters is fewer than the number of result string characters.

The following table describes the options that you can configure for source string characters:

Option	Description
Mask Only	Masks characters in the source that you configure as source string characters. For example, if you enter A and b as source string characters, every instance of A and b in the source data will change. A source character that is not an A or b will not change.
Mask all except	Masks all characters in the source except for source string characters. For example, if you enter "-" as the source string character, every character except for "-" will change.

Result String Replacement Characters

Configure result string replacement characters to specify masking output.

The PowerCenter Integration Service replaces characters in the source string with the result string replacement characters. For example, enter the following characters to configure each mask to contain uppercase alphabetic characters A through F:

```
ABCDEF
```

To avoid generating the same output for different input values, configure a wide range of substitute characters, or mask only a few source characters. The position of each character in the string does not matter.

The following table describes the options for result string replacement characters:

Option	Description
Use only	Masks the source with only the characters you define as result string replacement characters. For example, if you enter the characters A, B, and c, the masking replaces every character in the source column with an A, B, or c. The word "horse" might be replaced with `BAcBA`.
Use all except	Masks the source with any characters except the characters you define as result string replacement characters. For example, if you enter A, B, and c result string replacement characters, the masked data never has the characters A, B, or c.

Date Key Masking

You can configure key masking with dates to generate deterministic output.

You can change the seed to match the seed value for another column to return repeatable datetime values between the columns.

The PowerCenter Integration Service can mask dates between 1753 and 2400 with key masking. The PowerCenter Integration Service always generates valid dates. If the source year is in a leap year, the PowerCenter Integration Service returns a year that is also a leap year. If the source month contains 31 days, the PowerCenter Integration Service returns a month that has 31 days. If the source month is February, the PowerCenter Integration Service returns "February."

Numeric Key Masking Parameters

You can configure key masking for numeric values and generate deterministic output.

When you configure a column for numeric key masking, you can select a seed value for the column. When the PowerCenter Integration Service masks the source data, it applies a masking algorithm that requires the seed.

You can change the seed value for a column to produce repeatable results if the same source value occurs in a different column. Configure repeatable results when you want to maintain a primary key-foreign key relationship between two tables. In each rule, enter the same seed value for the primary-key column as the seed value for the foreign-key column. The PowerCenter Integration Service produces deterministic results for the same numeric values. The referential integrity is maintained between the tables.

String Key Masking Parameters

Configure string key masking to mask all or part of a string. To limit the masking output to certain characters, specify a mask format and result string replacement characters. If you need repeatable output, specify a seed value.

The following table describes the masking parameters that you can configure for key masking string values:

Parameter	Description
Seed	A start number that enables the PowerCenter Integration Service to return deterministic data. Select a seed value between 1 and 1,000. Apply the same seed value to a column to return the same masked data values in different source data.
Mask Format	The type of character to substitute for each character in the input data. You can limit each character to an alphabetic, numeric, or alphanumeric character type.

Parameter	Description
Source String Characters	The characters in the source string that you want to mask. For example, mask the number sign (#) character whenever it occurs in the input data. Leave this field blank to mask all the input characters. The PowerCenter Integration Service does not always return unique data if the number of source string characters is less than the number of result string characters.
Result String Replacement Characters	Substitutes the characters in the target string. For example, enter the following characters to configure each mask to contain uppercase alphabetic characters A through F: `ABCDEF`

Nullification Masking

Nullification masking replaces a column of data with a null value. Use nullification masking to mask binary, date, numeric, or string data.

Nullification masking has no parameters.

Phone Masking

Phone masking applies a built-in masking format to change phone number data. Mask string datatypes with phone masking.

Phone masking does not change the format of the original phone number. For example, phone masking can mask the phone number `(408)382 0658` as `(408)256 3106`.

The source data can contain numbers, spaces, hyphens, and parentheses. Phone masking does not mask alphabetic or special characters.

Random Masking

Random masking produces random, non-repeatable results for the same source data and masking rules. Random masking does not require a seed value. Mask date, numeric, and string datatypes with random masking.

The results of random masking are non-deterministic. Use random masking to mask date, numeric, or string datatypes.

The following table describes the options that you can configure for random masking:

Option	Description
Range	A range of output values. The PowerCenter Integration Service returns data between the minimum and maximum values. You can configure a range for date, numeric and string datatypes.
Blurring	A range of output values with a fixed or percent variance from the source data. Returns data that is close to the value of the source data. You can configure blurring for date and numeric datatypes.
Mask Format	The type of character to substitute for each character in the input data. You can limit each character to an alphabetic, numeric, or alphanumeric character type. You can configure a mask format for the string datatype.
Source String Characters	The characters in the source string that you want to mask. You can configure source string characters for the string datatype.
Result String Replacement Characters	Substitutes the characters in the target string. You can configure replacement characters for the string datatype.

Range Masking

Configure a range to define an output range for numeric, date, or string data.

When you define a range for numeric or date values, the PowerCenter Integration Service masks the source data with a value between the minimum and maximum values. When you configure a range for a string, you configure a range of string lengths.

Note: When you configure date random masking, the maximum datetime must be later than the minimum datetime.

Blurring

Configure blurring to return a random value that is close to the original value. For random masking of datetime or numeric data, blurring creates an output value within a fixed or percent variance from the source data value.

Date Blurring

To blur a datetime source value, select a unit of time to blur, a high bound, and a low bound. You can select year, month, day, or hour as the unit of time. By default, the blur unit is year.

For example, to restrict the masked date to a date within two years of the source date, select year as the unit. Enter two as the low and high bound. If a source date is 02 February, 2006, the PowerCenter Integration Service returns a date between 02 February, 2004 and 02 February, 2008.

Numeric Blurring

To blur a numeric source value, select a fixed or percent variance, a high bound, and a low bound. The high and low bounds must be greater than or equal to zero.

The following table lists the masking results for blurring range values when the input source value is 66:

Blurring Type	Low	High	Result
Fixed	0	10	Between 66 and 76
Fixed	10	0	Between 56 and 66
Fixed	10	10	Between 56 and 76
Percent	0	50	Between 66 and 99
Percent	50	0	Between 33 and 66
Percent	50	50	Between 33 and 99

Mask Format

Configure a mask format to limit each character in the output column to an alphabetic, numeric, or alphanumeric character.

Note: The mask format contains uppercase characters. When you enter a lowercase mask character, Test Data Manager converts the character to uppercase.

The following table describes mask format characters:

Character	Description
A	Alphabetical characters. For example, ASCII characters a to z and A to Z.
D	Digits. From 0 through 9.
N	Alphanumeric characters. For example, ASCII characters a to z, A to Z, and 0-9.
X	Any character. For example, alphanumeric or symbol.
+	No masking.
R	Remaining characters. R specifies that the remaining characters in the string can be any character type. R must appear as the last character of the mask.

If you do not define a mask format, the PowerCenter Integration Service replaces each source character with any character. If the mask format is longer than the input string, the PowerCenter Integration Service ignores the extra characters in the mask format. If the mask format is shorter than the source string, the PowerCenter Integration Service does not mask the characters at the end of the source string.

Source String Characters

Source string characters are characters that you want to mask in the source. Configure source string characters if you want to mask a few of the characters in the input string.

For example, if you set the number sign (#) as a source string character, it is masked every time it occurs in the input data. The position of the characters in the source string does not matter, and you can configure any number of characters. If you do not configure source string characters, the masking replaces all the source characters in the column.

The source characters are case sensitive. The PowerCenter Integration Service does not always return unique data if the number of source string characters is fewer than the number of result string characters.

The following table describes the options that you can configure for source string characters:

Option	Description
Mask Only	Masks characters in the source that you configure as source string characters. For example, if you enter A and b as source string characters, every instance of A and b in the source data will change. A source character that is not an A or b will not change.
Mask all except	Masks all characters in the source except for source string characters. For example, if you enter "-" as the source string character, every character except for "-" will change.

Result String Replacement Characters

Result string replacement characters are a set of characters that the PowerCenter Integration service can use to mask to the source data. You can configure the masking rule to mask the source only from the set of characters, or you can configure the masking rule to mask the source with any character except the result string replacement characters.

The PowerCenter Integration Service replaces characters in the source string with the result string replacement characters. For example, enter the following characters to configure each mask to contain uppercase alphabetic characters A through F:

```
ABCDEF
```

To avoid generating the same output for different input values, configure a wide range of substitute characters, or mask only a few source characters. The position of each character in the string does not matter.

The following table describes the options for result string replacement characters:

Option	Description
Use only	Masks the source with only the characters you define as result string replacement characters. For example, if you enter the characters A, B, and c, the masking replaces every character in the source column with an A, B, or c. The word "horse" might be replaced with `BAcBA`.
Use all except	Masks the source with any characters except the characters you define as result string replacement characters. For example, if you enter A, B, and c result string replacement characters, the masked data never has the characters A, B, or c.

Date Random Masking Parameters

To mask datetime values with random masking, either configure a range of output dates or choose a variance.

When you configure a variance, choose a part of the date to blur. Choose the year, month, day, hour, minute, or second. The PowerCenter Integration Service returns a date that is within the range you configure.

The following table describes the parameters that you can configure for random masking of datetime values:

Parameter	Description
Range	The minimum and maximum values to return for the selected datetime value. The date range is a fixed variance.
Blurring	Masks a date based on a variance that you apply to a unit of the date. The PowerCenter Integration Service returns a date that is within the variance. You can blur the year, month, day, or hour. Choose a low and high variance to apply.

Numeric Random Masking Parameters

When you mask numeric data, you can configure a range of output values for a column.

The PowerCenter Integration Service returns a value between the minimum and maximum values of the range depending on column precision. To define the range, configure the minimum and maximum ranges or a blurring range based on a variance from the original source value.

The following table describes the parameters that you can configure for random masking of numeric data:

Parameter	Description
Range	A range of output values. The PowerCenter Integration Service returns numeric data between the minimum and maximum values.
Blurring Range	A range of output values that are within a fixed variance or a percent variance of the source data. The PowerCenter Integration Service returns numeric data that is close to the value of the source data. You can configure a range and a blurring range.

String Random Masking Parameters

Configure random masking to generate random output for string columns.

To configure limitations for each character in the output string, configure a mask format. Configure filter characters to define which source characters to mask and the characters to mask them with.

The following table describes the parameters that you can configure for random masking of string columns:

Parameter	Description
Range	The minimum and maximum string length. The PowerCenter Integration Service returns a string of random characters between the minimum and maximum string length.
Mask Format	The type of character to substitute for each character in the input data. You can limit each character to an alphabetic, numeric, or alphanumeric character type.
Source String Characters	The characters in the source string that you want to mask.
Result String Replacement Characters	Substitutes the characters in the target string.

Shuffle Masking

Shuffle masking masks the data in a column with data from the same column in another row of the table. Shuffle masking switches all the values for a column in a file or database table. You can restrict which values to shuffle based on a lookup condition or a constraint. Mask date, numeric, and string datatypes with shuffle masking.

For example, you might want to switch the first name values from one customer to another customer in a table. The table includes the following rows:

```
100 Tom Bender
101 Sue Slade
102 Bob Bold
103 Eli Jones
```

When you apply shuffle masking, the rows contain the following data:

```
100 Bob Bender
101 Eli Slade
102 Tom Bold
103 Sue Jones
```

You can configure shuffle masking to shuffle data randomly or you can configure shuffle masking to return repeatable results.

Note: If the source file might have empty strings in the shuffle column, set the **Null and Empty Spaces** option to Treat as Value in the rule exception handling. When you set the option to Treat as Value, the PowerCenter Integration Service masks the space or the null value with a valid value. The default is to skip masking the empty column.

Shuffle Masking Parameters

You can configure masking parameters to determine if shuffle masking is repeatable, the masking is repeatable for one workflow run, or the masking is random. You can also configure a lookup to ensure that replacement values originate from rows that contain specific values.

The following image shows Data Masking parameters that appear when you configure a Shuffle data masking rule:

The following table describes the parameters that you can configure for shuffle masking:

Parameter	Description
Shuffle Type	Select random or representative shuffling: - Random. Shuffle values from one row to another without checking if the target values are unique for each source value. For example, the Integration Service masks 12345 with 65432 in a row. The Integration Service can also replace 33333 with 12345 in another row. - Representative. All source rows with the same value receive the same shuffle value. When the Integration Service replaces 12345 with 65432, then it can use 65432 as a mask value for any row with a 12345 source value. Representative masking does not save values between workflow runs. Use repeatable masking to return the same values between workflow runs.
Seed	Starting point for creating repeatable output. Enter a number between 1 and 999. Default is 1. Enabled when Representative Shuffle Type is selected.
Constrained	Restricts applying shuffle masking to rows that are constrained by another column. For example, shuffle employee names based on gender. Or, shuffle addresses within the same city. Choose the constraint column when you assign the rule to columns in a project.

SIN Masking

You can mask a Social Insurance number (SIN) that is nine digits. The digits can be delimited by any set of characters. Mask string datatypes with social insurance masking.

If the number contains no delimiters, the masked number contains no delimiters. Otherwise the masked number has the following format: `xxx-xxx-xxx`.

You can define the first digit of the masked SIN.

Enable **Start Digit** and enter the digit. The PowerCenter Integration Service creates masked Social Insurance numbers that start with the number that you enter.

You can configure repeatable masking for Social Insurance numbers. To configure repeatable masking for SIN numbers, click Repeatable Output and enter a Seed Value.

SSN Masking

SSN masking applies a built-in mask format to change Social Security numbers. Mask string datatypes with SSN masking.

The PowerCenter Integration Service generates valid Social Security numbers. To avoid generating numbers that the Social Security Administration has already issued, you can download the latest version of the High Group List at the following location:

```
http://www.ssa.gov/employer/highgroup.txt
```

The PowerCenter Integration Service accesses the latest High Group List from the following location:

```
<PowerCenter Installation Directory>\infa_shared\SrcFiles\highgroup.txt
```

The PowerCenter Integration Service generates Social Security numbers that are not on the High Group List.

The SSN masking accepts any SSN format that contains nine digits. You can delimit the digits with any set of characters. For example, the SSN masking accepts the following format:

```
+=54-*9944$#789-,*()”
```

You can configure repeatable masking for Social Security numbers. Select the **Repeatable Output** option, select the **Seed** option, and enter a value.

The PowerCenter Integration Service returns deterministic Social Security numbers with repeatable masking. The PowerCenter Integration Service cannot return all unique Social Security numbers because it cannot return valid Social Security numbers that the Social Security Administration has issued.

Substitution Masking

Substitution masking replaces a column of data with similar but unrelated data from a dictionary. Mask date, numeric, and string datatypes with substitution masking.

Use substitution masking to mask string data with realistic output. For example, if you want to mask address data, you specify a dictionary file that contains addresses.

Substitution is an effective way to replace production data with realistic test data. When you configure substitution masking, select the relational or flat file dictionary that contains the substitute values. The PowerCenter Integration Service performs a lookup on the dictionary and replaces source data with data from the dictionary.

When you assign a substitution masking rule to a column, you can specify the rule assignment parameters.

The following table describes the rule assignment parameters that you can configure:

Parameter	Description
Lookup Condition	The column name in the source table you can refer to match with the column in the dictionary. This field is optional.
Unique Substitution Column	The column name in the source table to substitute with unique data. This field is optional.

You can substitute data with repeatable or non-repeatable values. When you choose repeatable values, the PowerCenter Integration Service produces deterministic results for the same source data and seed value. You must configure a seed value to substitute data with deterministic results. The PowerCenter Integration Service maintains a storage table of source and masked values for repeatable masking. You can specify the storage table you want to use when you generate a workflow.

Substitution Masking Parameters

You can substitute data with repeatable or non-repeatable values.

When you choose repeatable values, the PowerCenter Integration Service produces deterministic results for the same source data and seed value. You must configure a seed value to substitute data with deterministic results.

You can configure the following substitution masking parameters:

Parameter	Description
Repeatable Output	Returns deterministic results between sessions. The PowerCenter Integration Service saves masked values in the storage table.
Seed	A start number that the PowerCenter Integration Service uses to return deterministic data.
Unique Substitution Data	Replaces the target column with unique masked values for every unique source column value. If there are more values in the source than in the dictionary file, the data masking operation fails because the dictionary file does not contain sufficient unique values to substitute the data. For security, the default is nonunique substitution.
Dictionary Information	Required. Configuration of the flat file or relational table that contains the substitute data values. Configure the following parameters: - Dictionary. Displays the flat file or relational table name that you select. - Masked Value. The column returned to the masking rule. - Lookup Column. The source data column to use in the lookup. - Serial Number Column. The column in the dictionary that contains the serial number.

URL Masking

URL masking applies a built-in mask format to change URL data. Mask string datatypes with URL masking.

The PowerCenter Integration Service parses a URL by searching for the `://` string and parsing the substring to the right of it. The source URL must contain the `://` string. The source URL can contain numbers and alphabetic characters.

The PowerCenter Integration Service does not mask the protocol of the URL. For example, if the URL is `http://www.yahoo.com`, the PowerCenter Integration Service can return `http://MgL.aHjCa.VsD/`. The PowerCenter Integration Service might generate a URL that is not valid.

Note: The PowerCenter Integration Service always returns ASCII characters for a URL.

Name Substitution Example

You want to mask employee names and you want to preserve the gender and nationality of the names in the masked data.

You create substitution masking rules to mask first names based on gender and nationality. Substitution masking replaces a column of data with similar but unrelated data from a dictionary. You use substitution masking to mask string data with realistic output. You use advanced masking to apply multiple masking techniques to a column.

Complete the following steps to create rules to mask names based on gender and nationality:

1. Add a dictionary in Test Data Manager.
2. Create a substitution rule that substitutes first names based on gender.
3. Create an advanced masking rule to substitute first names based on nationality.

Add a Dictionary in Test Data Manager

Add a dictionary in Test Data Manager to use for substitution masking.

The dictionary must contain country, gender, and first name columns. You can use a flat file or a relational dictionary. Test Data Manager uses the dictionary to substitute data.

The following text is a sample from the flat file dictionary that you use to mask the employee names:

```
SNO, COUNTRY, GENDER, FIRSTNAME
1, US, M, Adam
2, US, F, Sarah
3, JP, M, Mahito
4, JP, F, Kyoko
```

Adding a Relational Dictionary

When you add a relational dictionary, you define the connection to the dictionary.

1. In the **Administrator | Dictionaries** view, click **Actions** > **New Dictionary**.

   The **New Dictionary** tab appears.
2. Enter the name of the dictionary, an optional description of the dictionary, and the type of the dictionary.
3. Click **Select** to define a connection.

   The **Select Relational Dictionary** dialog box appears.
4. Select a datasource connection from the menu, and click **Next**.
5. Select a datasource, and click **Next**.
6. Select a table from the list of tables in the datasource, and click **Finish**.
7. Review the **Connection**, **Schema**, and **Table** properties you selected.
8. Click **Save**.

   A tab with the dictionary properties opens and the dictionary appears in the **Administrator | Dictionaries** view.

Adding a Flat File Dictionary

When you add a flat file dictionary, you can select a dictionary from the local system.

1. In the **Administrator | Dictionaries** view, click **Actions** > **New Dictionary**.

   The **New Dictionary** tab appears.
2. Enter a name and optional description for the dictionary. Select **Flat File** as the dictionary type.
3. Click **Choose File** to browse to the location of the dictionary. Select the dictionary file, and click **Open**.
4. Optionally, select to overwrite the file if a dictionary exists in the `<TDM installation directory>/TDM/infa_shared/LkpFiles` directory.
5. Select the file encoding type.
6. Click **Save**.

   A tab with the dictionary properties opens and the dictionary appears in the **Administrator | Dictionaries** view.

Creating the Substitution Rule

Create a substitution rule that substitutes first name based on gender.

1. In the **Policies** view, click **Actions** > **New** > **Masking Rule**.

   The **New Rule** window appears.
2. Enter a name and optional description for the rule.
3. Select the string datatype and standard substitution masking type. Select override allowed.
4. Click **Next**.
5. Enter the following substitution masking parameters:
   - Dictionary. Select the dictionary to use for the masking rule.
   - Masked Value. Select FIRSTNAME.
   - Lookup column. Select GENDER.
   - Serial number column. Select the serial number column.

   The following image shows the substitution masking parameters:

6. Click **Finish**.

   The rule appears in the **Policies** view.

Creating the Advanced Masking Rule

Create an advanced rule that substitutes first names based on nationality with a lookup condition on gender.

1. In the **Policies** view, click **Actions** > **New** > **Masking Rule**.

   The **New Masking Rule** window appears.
2. Enter a name and optional description for the rule.
3. Select the advanced masking type.
4. Click **Next**.
5. Click **Create Input Column** in the **Specify Masking Properties** window.

   The **Add Column** window appears.

6. Enter the following general properties:
   - Column name. Enter in_Country.
   - Column type. Select input.
   - Datatype. Select string.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.

   The following image shows the in_Country column properties:

7. Click **OK**.

   The in_Country input column appears in the list of input columns.
8. Click **Create Input Column**.

   The **Add Column** window appears.
9. Enter the following general properties:
   - Column name. Enter in_FirstName.
   - Column type. Select input.
   - Datatype. Select string.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.
10. Click **OK**.

    The in_FirstName input column appears in the list of input columns.
11. Click **Create Input Column**.

    The **Add Column** window appears.
12. Enter the following general properties:
    - Column name. Enter in_Gender.
    - Column type. Select input.
    - Datatype. Select string.
    - Precision. Select 10.
    - Scale. Select 10.
    - Mandatory. Select the check box.
13. Click **OK**.

The in_Gender input column appears in the list of input columns.

14. Click **Create Input Column**.

    The **Add Column** window appears.

15. Enter the following general properties:
    - Column name. Enter var_FirstName_us.
    - Column type. Select variable.
    - Datatype. Select string.
    - Precision. Select 10.
    - Scale. Select 10.
    - Mandatory. Select the check box.

16. Select the masking rule masking property. Configure the following masking properties:
    - Condition. Enter in_Country='us'.
    - Rule. Select the substitution rule you created in Step 2.
    - Override Properties. Click **Edit**. Enable **Repeatable Output** and **Unique Substitution Data** options, and click **Save**. The property appears as Yes (Overridden).
    - Lookup column. Select in_Gender.
    - Unique column. Selct in_Country.
    - Input column. Select in_FirstName.

    The following image shows the var_FirstName_us column properties:

17. Click **OK**.

    The var_FirstName_us variable column appears in the list of variable columns.

18. Click **Create Input Column**.

    The **Add Column** window appears.

19. Enter the following general properties:
    - Column name. Enter var_FirstName_jp.
    - Column type. Select variable.
    - Datatype. Select string.
    - Precision. Select 10.

- Scale. Select 10.
- Mandatory. Select the check box.

20. Select the masking rule masking property. Configure the following masking properties:
    - Condition. Enter in_Country='jp'.
    - Rule. Select the substitution rule you created in Step 2.
    - Override Properties. Click **Edit**. Enable **Repeatable Output** and **Unique Substitution Data** options, and click **Save**. The property appears as Yes (Overridden).
    - Lookup column. Select in_Gender.
    - Unique column. Selct in_Country.
    - Input column. Select in_FirstName.
21. Click **OK**.

    The var_FirstName_jp variable column appears in the list of variable columns.
22. Click **Create Input Column**.

    The **Add Column** window appears.
23. Enter the following general properties:
    - Column name. Enter o_FirstName.
    - Column type. Select output.
    - Datatype. Select string.
    - Precision. Select 10.
    - Scale. Select 10.
    - Mandatory. Select the check box.
24. Select the conditional inputs masking property. Configure the following masking property:
    - Conditional input. Select in_FirstName.

    The following image shows the o_FirstNames column properties:

25. Click **OK**.

    The o_FirstName variable column appears in the list of output columns.

The following image shows the rule columns:

Columns				
Column Name	Dat...	Pr...	S...	Ma...
Input Columns (3)				
in_Country	String	10	10	true
in_FirstName	String	10	10	true
in_Gender	String	10	10	true
Variable Columns (2)				
var_FirstName_us	String	10	10	true
var_FirstName_jp	String	10	10	true
Output Columns (1)				
o_FirstName	String	10	10	true

26. Click **Next**.

    Review the rule map that appears.

27. Click **Finish**.

    The rule appears in the **Policies** view.

Shuffle Address Example

You want to mask employee addresses and keep the ZIP code unmasked.

Create a shuffle rule and an advanced masking rule to shuffle addresses and leave the ZIP code unmasked. Shuffle masking masks the data in a column with data from the same column in another row of the table. Shuffle masking switches all the values for a column in a file or database table. You can restrict which values to shuffle based on a lookup condition or a constraint.

Create shuffle masking rules with a dictionary that contains three address columns and a ZIP code column. The ZIP code remains unmasked. The three address columns shuffle, but remain consistent.

The following text shows a sample flat file with the required columns:

```
SNO,AddressLine1,AddressLine2,AddressLine3,ZIP
1,3290 Apple Lane,Chillicothe,IL,61523
2,7760 Ash Street,Dallas,TX,75240
3,2229 Ash Street,Moscow,TN,38057
4,6698 Caldwell Road,Rochester,NY,14620
```

Complete the following steps to create rules to shuffle addresses with the ZIP code unmasked in Test Data Manager:

1. Create a shuffle rule.
2. Create an advanced masking rule that keeps the ZIP code column unmasked and shuffles the address columns together.

Creating the Shuffle Rule

Create a shuffle rule in Test Data Manager.

1. In the **Policies** view, click **Actions** > **New** > **Masking Rule**.

   The **New Rule** window appears.
2. Enter a name and optional description for the rule.
3. Select the string datatype and standard shuffle masking type. Select override allowed.
4. Click **Next**.
5. Enter the following shuffle masking parameters:
   - Shuffle type. Select Random.
   - Constrained. Select Constrained.

   The following image shows the shuffle masking parameters:

6. Click **Finish**.

   The rule appears in the **Policies** view.

Create the Advanced Masking Rule

Create an advanced masking rule that shuffles address lines.

The advanced rule shuffles three address line columns together and keeps the ZIP code column unmasked. The rule uses shuffle masking with a lookup on AddressLine3 and dependent masking on AddressLine1 and AddressLine2.

In the rule editor, you create input columns, variable columns, and output columns.

Setting Up the Advanced Masking Rule

Set up the advanced masking rule.

1. In the **Policies** view, click **Actions** > **New** > **Masking Rule**.

   The **New Masking Rule** window appears.

2. Enter a name and optional description for the rule.
3. Select the advanced masking type.
4. Click **Next**.

Creating the Input Columns

Create input columns in the advanced masking rule.

1. Create an input column. Click **Create Input Column** in the **Specify Masking Properties** window.

   The **Add Column** window appears.
2. Enter the following general properties:
   - Column name. Enter i_AddressLine1.
   - Column type. Select Input.
   - Datatype. Select String.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.

   The following image shows the i_AddressLine1 column properties:

3. Click **OK**.

   The i_AddressLine1 input column appears in the list of input columns.
4. Create an input column with the following general properties:
   - Column name. Enter i_AddressLine2.
   - Column type. Select Input.
   - Datatype. Select String.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.
5. Create an input column with the following general properties:
   - Column name. Enter i_AddressLine3.
   - Column type. Select Input.
   - Datatype. Select String.
   - Precision. Select 10.

- Scale. Select 10.
- Mandatory. Select the check box.

6. Create an input column with the following general properties:
   - Column name. Enter i_ZIP.
   - Column type. Select Input.
   - Datatype. Select String.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.

Creating the Variable Columns

Create variable columns in the advanced masking rule.

1. Create a variable column to shuffle AddressLine3 based on ZIP. Click **Create Input Column**.

   The **Add Column** window appears.
2. Enter the following general properties:
   - Column name. Enter v_AddressLine3.
   - Column type. Select Variable.
   - Datatype. Select String.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.
3. Select the masking rule masking property. Configure the following masking properties:
   - Condition. Leave blank.
   - Rule. Select the shuffle rule that you created in Step 2.
   - Override Properties. Select None.
   - Lookup column. Select i_ZIP.
   - Input column. Select i_AddressLine3.

   The following image shows the v_AddressLine3 column properties:

4. Click **OK**.

   The v_AddressLine3 variable column appears in the list of variable columns.

5. Create a variable column to mask AddressLine2 based on the masked value of AddressLine3. Click **Create Input Column**.

   The **Add Column** window appears.

6. Enter the following general properties:
   - Column name. Enter v_AddressLine2.
   - Column type. Select Variable.
   - Datatype. Select String.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.

7. Select the dependent masking property. Configure the following masking properties:
   - Input column. Select v_AddressLine3.
   - Dictionary column. Select i_AddressLine2.

   The following image shows the v_AddressLine2 column properties:

8. Click **OK**.

   The v_AddressLine2 variable column appears in the list of variable columns.

9. Create a variable column to mask AddressLine1 based on the masked value of AddressLine3. Click **Create Input Column**.

   The **Add Column** window appears.

10. Enter the following general properties:
    - Column name. Enter v_AddressLine1.
    - Column type. Select Variable.
    - Datatype. Select String.
    - Precision. Select 10.
    - Scale. Select 10.
    - Mandatory. Select the check box.

11. Select the dependent masking property. Configure the following masking properties:
    - Input column. Select v_AddressLine3.
    - Dictionary column. Select i_AddressLine1.
12. Click **OK**.

    The v_AddressLine1 variable column appears in the list of variable columns.

Creating the Output Columns

Create output columns in the advanced masking rule.

1. Click **Create Input Column**.

   The **Add Column** window appears.
2. Enter the following general properties:
   - Column name. Enter o_AddressLine1.
   - Column type. Select Output.
   - Datatype. Select String.
   - Precision. Select 10.
   - Scale. Select 10.
   - Mandatory. Select the check box.
3. Select the dependent masking property. Configure the following dependent properties:
   - Input column. Select v_AddressLine3.
   - Dictionary column. Select i_AddressLine1.

   The following image shows the o_AddressLine1 column properties:

4. Click **OK**.

   The o_AddressLine1 variable column appears in the list of variable columns.
5. Click **Create Input Column**.

   The **Add Column** window appears.
6. Enter the following general properties:
   - Column name. Enter o_AddressLine2.
   - Column type. Select Output.
   - Datatype. Select String.

- Precision. Select 10.
- Scale. Select 10.
- Mandatory. Select the check box.

7. Select the dependent masking property. Configure the following dependent properties:
   - Input column. Select v_AddressLine3.
   - Dictionary column. Select i_AddressLine2.
8. Click **OK**.

   The o_AddressLine2 variable column appears in the list of variable columns.
9. Click **Create Input Column**.

   The **Add Column** window appears.
10. Enter the following general properties:
    - Column name. Enter o_AddressLine3.
    - Column type. Select Output.
    - Datatype. Select String.
    - Precision. Select 10.
    - Scale. Select 10.
    - Mandatory. Select the check box.
11. Select the dependent masking property. Configure the following dependent properties:
    - Input column. Select v_AddressLine3.
    - Dictionary column. Select i_AddressLine3.
12. Click **OK**.

    The o_AddressLine3 variable column appears in the list of variable columns.

Saving the Advanced Masking Rule

Save the advanced masking rule.

1. Verify that all the rule columns that you created are visible in the columns list.

   The following image shows the rule columns:

2. Click **Finish**.

   The rule appears in the **Policies** view.

CHAPTER 9

Data Generation

This chapter includes the following topics:

- Data Generation Overview, 166
- Data Generation Components, 167
- Data Generation Task Flow, 167
- Data Generation Rule Types, 168
- Default Settings, 169
- Define Default Generation Rules, 170
- Standard Generation Rules, 170
- Custom Generation Rules, 172
- Ad Hoc Generation Rules, 173
- Conditional Constraints, 175
- Data Generation Rule Assignments, 176
- Data Generation for XSD Sources, 179
- Data Generation Plans and Workflows, 180

Data Generation Overview

Use data generation to create realistic test data for test environments. You can define generation rules that define the logic to generate data.

Import metadata into a project to define the type of data that you want to generate. You assign generation rules that you create or default rules to the target columns to generate data based on the data type of the column. When you create a rule, you can choose a generation technique and configure parameters to create random test data.

To implement data generation, you create a data generation plan and a workflow from the plan. If the target is a flat file, you can configure test tool integration properties in the plan. Configure test tool integration to copy the results to a location in an integrated HP ALM server. To store the test data along with the metadata in the test data warehouse, select the test data warehouse as the target connection.

Data Generation Example

You work for an organization that sells plane tickets. You want to generate data in tables that contain customer information such as identification number, status of membership, and address. You want

additional tables to store ticket details such as ticket number and flight number. To generate the data, you can perform the following tasks:

- Create data generation rules that load dictionary values such as names into the tables.
- Create random number strings for ticket numbers.
- Create a numeric sequence for identification numbers.
- Use a reference lookup for values such as airport codes.
- Create projects to import metadata, enable relationships, and create entities.
- Make rule assignments, create a plan, and run the plan to generate the data.

Data Generation Components

To perform data generation operations, assign rules to columns. Import metadata to define the columns in the target database.

The following table describes the components that you create to implement data generation:

Component	Description
Assignments	The allocation of rules to a column to generate the column data.
Plan	Defines data generation operations. You import the target schema metadata into the repository in the plan and configure the settings.
Rule	Defines the data generation technique and parameters. A generation technique defines the logic to generate the data. Generation parameters define how a generation technique in a rule generates data. You can set an override option in a rule that defines whether users can modify the generation parameters for the rule when they assign the rule to the columns in a target.
Table	The target table in which you assign a generation rule.
Entity	Defines a set of tables that are related based on physical or logical constraints.

Data Generation Task Flow

To implement data generation operations, assign generation rules to target columns. Create a plan and add rules to the plan. Generate a workflow from the plan and run the workflow to generate data in the target database.

To create the components that you need in a data generation plan, perform the following high-level steps:

1. Create generation rules that define the generation techniques and configure the parameters. Define default rules that generate data for columns that you do not assign a generation rule to.
2. Create a project and import metadata into the project.
3. If you want to restrict the generated values in a column based on values in another column, create a conditional constraint.

4. Add generation rules to the project and assign the rules to columns in the project.
5. Create a plan and add data generation components to the plan.
6. Generate and run the plan. View the progress of the workflow in the **Monitor** view.

Data Generation Rule Types

A data generation rule is a rule that defines how you want to generate data in the target table. When you create a data generation rule, you choose a generation technique and configure generation parameters.

You can create the following data generation rule types:

Standard

A data generation rule that uses predefined generation techniques.

Custom

A data generation rule that generates data based on PowerCenter mapplets logic.

Ad Hoc

A data generation rule that you create within a project for a column in the target table.

Standard rules and custom rules are global rules. Global rules are rules that you create in the **Policies** view of Test Data Manager. When you create a global generation rule in the **Policies** view, you specify a rule name. This rule is available for projects that you create. You assign the rule to one or more columns in the project that you want to generate data into.

Ad hoc generation rules are rules that you create in the **Define | Data Generation** view of a project. Ad hoc rules do not have a name, and you cannot move an ad hoc rule. You can create ad hoc generation rules within the projects. You select the target columns for an ad hoc generation rule when you create the rule. You can select a standard generation technique for an ad hoc generation rule or you can select an ad hoc generation technique. You can use reference lookup, effective dates, and expression generation techniques only in an ad hoc generation rule. In an ad hoc rule, if you want to generate numbers in a column that uses string data type, you can change the data type to numeric.

After you install TDM and log in to Test Data Manager, TDM configures the default generation rules that comes with the installation package.

A data generation technique defines the logic that generates the data. When you create a generation rule, the techniques vary based on datatype you select. Generation parameters are options that you configure for a generation technique. For example, you can define different start dates and end dates for random date generation.

You can enable users to override generation parameters for a rule. For example, you create a rule that generates random dates from 2011 to 2012. You set the override option for the rule. When a user assign this rule to a target column, the user can select Saturday and Sunday as exclude dates so that the rule generates random weekday dates from 2011 to 2012.

Default Settings

Use default settings to generate test data for the columns that you have not assigned specific generation rules to.

When you create a project and import target metadata, the project contains the default generation rules for all data types that a TDM administrator configures. When you open a project, you can view the default data generation rules, default primary key generation rules, and default unique key generation rules for all data types from the **Overview | Properties** tab. Edit the default settings if you want to change the default generation rules.

The following table describes the properties to configure the default settings:

Property	Description
Default Rule for Numeric Datatype	The default rule that you can select to generate data for numeric data types.
Default Rule for String Datatype	The default rule that you can select to generate data for string data types.
Default Rule for Date Datatype	The default rule that you can select to generate data for date data types.
Default Rule for Binary Datatype	The default rule that you can select to generate data for binary data types.
Default Primary Key Rule for Numeric Datatype	The default primary key rule that you can select to generate data for numeric data types.
Default Primary Key Rule for String Datatype	The default primary key rule that you can select to generate data for string data types.
Default Primary Key Rule for Date Datatype	The default primary key rule that you can select to generate data for date data types.
Default Primary Key Rule for Binary Datatype	The default primary key rule that you can select to generate data for binary data types.
Default Unique Key Rule for Numeric Datatype	The default unique key rule that you can select to generate data for numeric data types.
Default Unique Key Rule for String Datatype	The default unique key rule that you can select to generate data for string data types.
Default Unique Key Rule for Date Datatype	The default unique key rule that you can select to generate data for date data types.
Default Unique Key Rule for Binary Datatype	The default unique key rule that you can select to generate data for binary data types.
Apply Default Rule to	The default rule that you can apply to all the columns without generation rules. You can also apply the default rule to the columns that do not have null values.
Default Character Set	The character set that you can select. Default is ASCII.
Minimum	The minimum number of child records for a parent record. Default is 1.
Maximum	The maximum number of child record for a parent record. Default is 1.

Property	Description
Parent Records Without Children	The number of parent records that have no child.
Child Records Without a Parent	The number of child records that have no parent.

Define Default Generation Rules

To generate test data, you can apply auto rule, default data domain rule, default global rule, default primary key rule, or default unique key rule to a target column.

View the default generation rules from the **Overview** | **Properties** tab. Edit the default settings if you want to change the default generation rules that you want to assign to the target columns.

You can apply the default generation rules from the **Define** | **Data Generation** page. The default data domain rule is the default rule that you configure when you create and run a data domain profile.

After you run a column profile, you can review and finalize the profile results. You can assign auto rule to columns that contain Yes in the Inferred Rule column. When you apply auto rule to a target column, TDM populates the target column with the approved profile results. You must not apply an auto rule to a primary key column because a primary key column cannot contain null values.

Standard Generation Rules

A standard generation rule is a global rule that applies a system-defined generation technique. A standard generation rule has one output column.

When you create a standard generation rule, you select the generation technique. You can define one generation technique in a standard rule and you can apply the rule to one column at a time. You can apply the rule multiple times to different columns.

You can use standard generation techniques to generate string, numeric, date, and binary datatypes. The techniques vary based on the datatype.

For example, you create a standard generation rule with the random date technique to generate birth dates. You can assign the rule multiple times in a project to generate birth dates in more than one column or you can use the rule in multiple projects.

Creating a Standard Generation Rule

To create a standard generation rule, you choose a generation technique based on the data type and configure the generation parameters to apply the technique.

1. Click **Policies** to access the **Policies** view.
2. Click **Actions** > **New** > **Generation Rule**.

   The **Rule Wizard** appears.
3. Enter a name and optional description for the rule.

4. Select the data type of the column to apply the generation rule to.
5. Select the standard generation rule.
6. To enable users to override generation parameters for a rule, select **Override Allowed**.
7. Click **Next**.
8. Enter the generation parameters.

   Note: The parameters that appear depend on the **Generation Technique** that you select.
9. Enter the exception test data options. Configure whether to allow null or empty spaces. Configure whether to generate data that is not valid.
10. Click **Finish**.

Editing a Generation Rule

You can edit a global generation rule to change the rule parameters.

1. Click **Policies**.
2. Click the name of the generation rule that you want to edit.

   The rule opens in a tab.
3. Click **Edit**.
4. Edit the parameters that you want to change.
5. To save the rule, click **OK**.

   If the generation rule has assignments, the **Impacted Objects** dialog box appears with the list of affected columns, plans, and data domains.
6. To download the list of affected objects, click **Export**, and save the .csv file.
7. To save the changes, click **Continue**.

   To update the changes in a plan, generate and run the plan again.

Overriding a Generation Rule

After you assign a generation rule to a column, you can override the rule properties for the column. You can override rule properties for global generation rules.

1. Open a project and click **Define** > **Data Generation**.
2. Select the column that you need to override the rule properties for.
3. Click **Actions** > **Override**.

   The **Generation Rule Override** window appears.
4. Update the values in the generation rule properties.
5. Click **Save** for each column that you update.

   The **Override** column value shows Yes-Overridden for the column that you updated.

Deleting a Generation Rule

You can delete a global generation rule. If the rule has assignments, you must make changes to the plan, and generate and run the plan again.

1. Click **Policies**.

2. Click the box to the left of the generation rule that you want to delete.
3. Click **Actions** > **Delete**.

   If the generation rule has no assignments, the **Delete Generation Rule** dialog box appears. If the generation rule has assignments, the **Impacted Objects** dialog box appears with the list of affected plans, columns, and data domains.
4. To delete the generation rule that has no assignments, click **OK**.
5. To delete the generation rule that contains assignments, click **Continue**. To download the list of affected objects, click **Export**, and save the .csv file.

   To update the changes in a plan, generate and run the plan again.

Custom Generation Rules

A custom generation rule is a global rule that generates data based on PowerCenter mapplet logic.

You can import a mapplet to create a custom generation rule. You must mark at least one input and one output column as required. Use custom rules to generate data for string, numeric, and date data types.

For example, an EMPLOYEES table contains columns for first name and last name. You create a custom generation rule to combine the first and last name of an employee to generate the full name.

You can also create custom rule with mapplets when you want to use PowerCenter transformations, such as an SQL transformation or a Stored Procedure transformation. You can also use custom mapplets when you need to use the logic to generate records or generate output with the user defined procedure.

Creating a Custom Generation Rule

Create a custom rule using PowerCenter mapplet to generate data.

1. To access the **Policies** view, click **Policies**.
2. Click **Actions** > **New** > **Generation Rule**.

   The **Rule Wizard** appears.
3. Enter a name and optional description for the rule.
4. Select the **Custom Rule** generation type.
5. To select a mapplet from the local folder, click **Browse**.
6. Click **Next**.

   The **Configure Mandatory Columns** dialog box appears with a list of input columns and output columns that are available in the mapplet.
7. To mark the columns as required fields, select the columns, and click **Set Mandatory**.

   You must at least select one input column and one output column as required fields.
8. Click **Finish**.

   The custom rule appears as another tab. Review and close the tab.

Ad Hoc Generation Rules

An ad hoc generation rule is a data generation rule that you create within a project. You can create an ad hoc generation rule for a column in the target table.

An ad hoc rule does not have a name. You cannot move an ad hoc rule because you use target columns to define the rule.

An ad hoc generation rule includes the standard data generation techniques and the techniques such as reference lookup generation, effective dates generation, and expression generation. An ad hoc rule can have more than one target column, based on the generation technique that you select.

When you create an ad hoc rule, the generation techniques that you can select depend on the data type of the column. To create an ad hoc rule, you can use a standard or an ad hoc generation technique. The ad hoc generation techniques include expression generation, reference lookup generation, and effective dates generation.

For example, an Airport table contains airport codes. You create a reference lookup rule on a From_City column that generates data from the Airport table and inserts it into the From_City column in the target table.

In an ad hoc rule, if you want to generate numbers in a column that uses string data type, you can change the data type to numeric. For example, Employee_ID is a sequence of numbers and the target table contains the Employee_ID column with string data type. To generate numbers in the column, select **Override Data Type to Numeric** when you create an ad hoc generation rule.

Creating an Ad Hoc Generation Rule

You can create an ad hoc generation rule in a project. The generation techniques change based on the data type of the column that you selected. The generation parameters change based on the generation technique that you select.

1. In a project, click **Define** > **Data Generation** to access the **Data Generation** view.
2. Select a column to assign the ad hoc generation rule to.
3. Click inside the **Generation Rule** column to view the list of rules available for the data type of the column.
4. Click **New Generation Rule**.

The **New Generation Rule** window appears.

5. If you want to generate numbers in a column with string data type, select **Override Data Type To Numeric**.
6. Select the generation technique from the list.
7. Enter the generation parameters. You can choose to generate unique output values.

   Note: If a Java heap space exception occurs, contact the Administrator to increase the heap size.
8. Enter the exception test data options. Configure whether to allow null or empty spaces. Configure whether to generate data that is not valid.

Editing an Ad Hoc Generation Rule

You can edit an ad hoc rule assignment to a column.

1. In a project, click **Define** > **Data Generation** to access the Data Generation view.
2. Select a column to edit the ad hoc rule assignment.
3. Click **Actions** > **Edit Generation Rule**.

   The **Edit Generation Rule** window appears.
4. Edit the parameters that you want to change. Click **OK**.

   If there are plans that contain the generation rule assignment, the **Impacted Objects** dialog box appears with the list of affected plans.
5. To download the list of affected plans, click **Export** and save the .csv file.
6. To save the changes, click **Continue**.

   To update the changes in a plan, generate and run the plan again.

Conditional Constraints

Create conditional constraints to restrict the data that you want to generate in a column based on the data in another column. The conditional constraint that you enter becomes a part of the metadata of the target table.

Configure expressions to apply constraints at the table level. Use the **Expression Builder** to enter the expressions. You can create multiple conditional constraints for a table.

You cannot use conditional constraints for the following data generation techniques:

- Numeric sequence
- Date sequence
- Random data patterns

Conditional Constraints Example

You have a table that contains airline ticket information and you want to generate airport codes in From_City and To_City columns in the table. You create and assign a dictionary generation rule to the columns. You do not want the From_City column and the To_City column to have the same value in a row. You create a conditional constraint and enter the following expression in the **Expression Builder**:

```
To_City != From_City
```

When you perform the data generation operation, TDM adheres to the conditional constraint and generates data in the From_City and To_City columns.

The following table lists sample flight details with generated data:

Flight Code	From_City	To_City
11A1	BOS	SFO
22B2	AUS	EWR
33C3	SEA	JFK

Conditional Constraints and Data Conflicts

When you perform a data generation operation, there might be conflicts between the conditional constraint and the values specified in the data generation rule parameters. The data generation operation prioritizes the data generation rule over the conditional constraint and TDM generates data with warning messages.

Conditional Constraints and Data Conflicts Example

You want to generate data in the Created_Date and Modified_Date columns in an Employee table. You create a random generation rule with a date range from January 1, 1970 to July 31, 2014, and assign the rule to both Created_Date and Modified_Date columns. You want the value in the Modified_Date column to be greater than the value in the Created_Date column. You create a conditional constraint and enter the following expression in the **Expression Builder**:

```
Modified_Date > Created_Date
```

When the Created_Date column contains July 31, 2014, TDM cannot generate data that adheres to the conditional constraint because the data generation rule parameters restrict the end dates to July 31, 2014. TDM generates a date within the specified range in the Modified_Date column and generates warning messages.

The following table lists sample employee details with generated data:

Employee ID	Created_Date	Modified_Date
1111	7/31/2014	1/13/2002
2111	6/5/2000	5/6/2010
8112	7/31/2014	5/16/1990

Data Generation Rule Assignments

You can assign a global or an ad hoc generation rule to a column within a project.

To assign a rule, choose a rule from the list in the **Generation Rule** column or select **New Generation Rule**. If you do not select a rule, the column receives the default rule that you configure in default settings.

The following image shows the **Define** | **Data Generation** view in a project:

The following table describes the fields in the **Data Generation** view:

Column	Description
Table	Name of the table.
Column	Name of the column of the generated data.
Data Type	Data type of the column of the generated data.
Is Null	Indicates whether the column can contain null values.

Column	Description
Generation Rule	The generation rule that you want to apply to the column. When you click inside the column, you can choose which rule to apply to the column. The rules change based on the data type of the column. Select **New Generation Rule** to create an ad hoc generation rule for the column.
Override	Shows the override property status for a rule. If the property is **Yes**, you can override the rule properties when you assign the rule to a column. When you override the rule parameters for a column the **Override** column value is **Yes-Overridden**.

Available Data Generation Rules

The **Data Generation** view lists the rules that you can assign to each column of the target table.

If no global rules appear for a column when you click **Generation Rule**, make sure that the project contains global generation rules for the column datatype.

Ad hoc generation rules are available for columns of all datatypes. Click **New Generation Rule** to create an ad hoc generation rule.

Note: In the **Data Generation** view, you must click the **Save** icon each time you assign a rule. If you do not save the rule assignment before you assign the next rule, the Test Data Manager discards the rule assignment.

Assigning Standard Generation Rules to Columns

Assign standard generation rules to one or more target columns to generate data in the columns.

1. In a project, click **Define** > **Data Generation** to access the **Data Generation** view.
2. Select a column to assign a generation rule to.
3. Click inside the **Generation Rule** column to view the list of rules available for the data type of the column.
4. Select a rule.
5. Click **Save** for each column that you update.

   If there are plans that contain the generation rule assignment, the **Impacted Objects** dialog box appears with the list of affected columns, plans, and data domains.
6. To download the list of affected objects, click **Export**, and save the .csv file.
7. To save the updates, click **Continue**.

   To update the changes in a plan, you must generate and run the plan again.

Assigning Custom Generation Rules to Columns

Assign custom generation rules to one or more target columns to generate data in the columns.

1. In the **Define | Data Generation** view, click the **Generation Rule** column for the column that requires a mapplet rule.
2. Select the mapplet rule from the list.

   The **Custom Rule Assignment** dialog box appears with the list of columns in the table.
3. Select a source input column in the left panel and a rule input port in the right panel.

Note: When you select a source input column, the rule ports with same datatype become available for mapping.

4. Click the **Link** icon to create a mapping.

   An arrow appears that links the source port to the rule port.

   Note: To create a mapping, you can also click **Show Simple** and map the ports.

5. If you need to remove the link, select the source input port and the rule input port. Click the **Unlink** icon.

   The arrow between the ports disappears.

6. After you map the rule input ports, click **Next**.
7. Select a rule output port in the left panel and a source input column in the right panel.
8. Click the **Link** icon to create a mapping. To remove a mapping, click the **Unlink** icon.
9. Click **Save**.

Assigning Auto Rules to Columns

You can assign an auto rule to target columns if the Inferred Rule status is Yes. You can edit the inferred rule properties after you assign auto rule to the target.

1. In a project, click **Define** > **Data Generation**.
2. Select a table column on which you want to assign the rule.
3. Click inside the **Generation Rule** column and select **Auto Rule**.
4. To save the assignment, click **Save**.
5. To view the assigned inferred rule properties, select the column. To update the rule properties, click **Edit**.
6. To assign Auto Rule to multiple columns, select multiple rows and click **Generation Rule Assignment**.

   The **Apply Generation Rules** dialog box appears.

7. Select **Auto Rule**, and click **Apply** to save the assignment.

   Note: You must not apply an auto rule to a primary key column because a primary key column cannot contain null values.

Deleting Generation Rule Assignments

You can delete a generation rule assignment to a column.

1. In a project, click **Define** > **Data Generation**.
2. Select the column that you want to delete the rule assignment for.
3. Click **Actions** > **Clear Assignment**.

   If the rule assignment does not exist in a plan, TDM deletes the generation rule assignment. If there are plans that contain the generation rule assignment, the **Impacted Objects** dialog box appears with the list of affected columns, plans, and data domains.

4. To delete the generation rule assignment, click **Continue**. To download the list of affected objects, click **Export**, and save the .csv file.

   To update the changes in a plan, you must generate and run the plan again.

Data Generation for XSD Sources

You can apply data generation rules to XML elements and attributes to generate test data.

View and open the XSD file from the **Discover | Files** view of the project. You can view the details of the XML elements and attributes in the XSD map view or the grid view.

Select an element, and apply a data generation rule to the element. You can select multiple elements and assign a default generation rule to the selected elements. You can remove all the default generation rule assignments, and assign another generation rule.

You can assign standard and custom generation rules to generate test data in XSD sources. You cannot use reference lookup and effective dates generation rules for an XSD source.

Data Generation Task Flow for XSD Sources

To generate data from XSD sources, you can assign the generation rules to the XML elements and attributes.

Perform the following high-level tasks to generate test data:

1. Create a project.
2. Import the XML source definition from the PowerCenter repository.
3. Add data generation rules to the project.
4. Open the XSD file from the project.
5. Select an element or an attribute. Specify a generation rule for a data domain or apply generation rules individually.
6. Create a data generation plan and add the XSD file. You can add multiple XSD files to a plan. Define the data generation parameters, and specify the target directory.
7. Generate a workflow from the plan.
8. Run the workflow, and monitor the progress of the workflow.
9. Verify that the data is generated in the target.

Applying Data Generation Rules to XML Elements and Attributes

Apply data generation rules to XML elements and attributes to generate test data in an XML file.

1. In a project, click **Discover** > **Files**.

   A list of XSD files in the project appears.
2. Click the XSD file.

   The XSD file opens in a separate tab.
3. Select the elements and attributes on which you want to apply the generation rule.
4. In the **Details** pane, click **Edit** in the **Data Generation** section.
5. Assign a data generation rule to the selected element or attribute. You can also create and assign an ad hoc generation rule.

The following image shows a sample data generation rule assignment to an XML element:

6. To save the assignment, click **OK**.
7. To assign default rules to multiple objects, click **Default Assignment**.
8. To clear rule assignments, click **Clear Assignment**.

Data Generation Plans and Workflows

A data generation plan contains XSD files, tables, and entities that represent the data that you want to generate.

The plan includes the components that have generation rule assignments. You generate a workflow from a plan to perform the generation operations. When you start a workflow for the plan, the data generation operation starts.

CHAPTER 10

Data Generation Techniques and Parameters

This chapter includes the following topics:

- Data Generation Techniques and Parameters Overview, 181
- Data Generation Techniques, 182
- Data Generation Parameters, 183
- Custom Generation, 184
- Dictionary Generation, 185
- Effective Dates Generation, 187
- Expression Generation, 189
- Random Generation, 191
- Reference Lookup Generation, 197
- Sequence Generation, 199
- Set of Values Generation, 201

Data Generation Techniques and Parameters Overview

A data generation technique defines how you want to generate data in a column. A generation parameter is an option that you configure for the technique.

The type of generation technique that you can apply depends on the datatype of the column that you need to generate. When you choose a generation technique, Test Data Manager displays parameters for the generation technique.

Some generation techniques are available only for ad hoc generation rules. The generation techniques that you can select to create a rule can change when you create an ad hoc generation rule.

You can create generation rules that create random strings, use data from a dictionary or set of values, or that create data based on a sequence that you define.

Data Generation Techniques

You can apply generation techniques based on the target datatype that you configure for a column.

For example, if the column datatype is numeric, you can define generation values that are within a fixed range or values in a sequence. You can select an additional rule to create values that are not valid values within the data set.

The following table describes the generation techniques that you can choose when you define a rule:

Generation Technique	Description
Custom	Standard technique that applies generation rules from a mapplet. The custom mapplet accepts input and uses the logic in the mapplet to generate the output.
Dictionary	Standard technique that imports dictionary values to generate data.
Effective Dates	Ad hoc technique that generates related dates for two columns in the target database.
Expression	Ad hoc technique that accepts expression as the input to generate the output.
Random	Standard technique that generates random strings, values, and dates.
Reference Lookup	Ad hoc technique that generates data from a reference lookup table. You can have multiple column assignments within a reference lookup rule.
Sequence	Standard technique that generates numeric and date values in a sequence.
Set of Values	Standard technique that uses a finite set of values to use in data generation.

Data Generation Parameters

Configure data generation parameters to define how to apply a data generation technique. The parameters that you configure depend on the data type that you need to generate. Some generation techniques are not available for all data types.

The following image shows data generation parameters that appear when you configure a random string generation rule:

Exception Test Data

Configure exception test data parameters to generate null and values that are not valid.

Select a global rule that you previously saved or select default to generate random null values and values that are not valid. Invalid values are data values that do not conform to the generation rule.

For example, you want to create a random date generation rule that generates 85% of the dates between 2012, and 15% of the dates between 2005 and 2010. You create a rule called Dates_2005 that generates dates from 2005 to 2010. You save the Dates_2005 rule. You create a rule called Recent_Dates that generates dates from 2012 to 2013. When you create the Recent_Dates rule, you select Dates_2005 in the invalid values parameter and enter 15 in the percent parameter.

Exception Test Data Parameters

You can configure exception test data parameters based on the data generation rules that you create. You can use exception test data parameters to generate null values and values that are not valid.

The following table describes the exception test data parameters:

Parameter	Description
Null Values	A list of global rules that you previously created or the default rule to generate null values. The rules are available based on the data type you select. Percentage defines the percentage of rows that can contain the null values. You can enter a whole number or a decimal number up to two places.
Invalid Values	A list of global rules that you previously created or the default rule to generate values that are not valid. The rules are available based on the data type you select. Percentage defines the percentage of rows that can contain the values that are not valid. You can enter a whole number or a decimal number up to two places.

Custom Generation

Create custom generation rules to generate data for string, numeric, and date data types.

Custom generation rule is a global rule. To create custom generation rule, import a PowerCenter mapplet. Mark at least one input and one output column as required.

The following image shows the custom generation parameters:

Column Name	Column Type	Data Type	Precision	Mandatory	Group Name
NEWFIELD	Input	String	10	☑	Input
NEWFIELD1	Input	String	10	No	Input
NEWFIELD2	Input	String	10	No	Input
NEWFIELD3	Input	String	10	No	Input
out_NEWFIELD	Output	String	10	Yes	Output
out_NEWFIELD1	Output	String	10	No	Output
out_NEWFIELD2	Output	String	10	No	Output
out_NEWFIELD3	Output	String	10	No	Output

Custom Generation Example

An EMPLOYEES table contains columns for first name and last name. Create and assign standard generation rules to generate first names and last names. Create a custom generation rule to combine the first and last name of an employee to generate the full name in the target table.

Custom Generation Parameters

Configure input and output columns that you import from a mapplet when you create a custom generation rule.

The following table describes the general properties that you can configure for input and output columns:

Parameter	Description
Column Name	The name of an input or an output column within a mapplet. When you assign the rule to a column, you map the column names in the rule to column names in the database.
Column Type	The column type. The mapplet can contain the following types of columns: - Input. Receives the input data. - Output. Returns the output data.
Data Type	The data type of the column.
Precision	The precision for the column. The maximum number of digits or the maximum number of characters that the column can accommodate. For example, 874.560 has a precision of 6.
Mandatory	Indicates if you must assign the column to a table column in the project. Applies to input and output columns. You must set at least one input and one output column as required.
Group Name	Specifies the group to which the columns belong. The group name can be input, output, or any other name that you provide when you create a mapplet.

Dictionary Generation

A dictionary generation rule is a global or ad hoc generation rule that loads dictionary values into the target test database.

Before you create a dictionary generation rule, you must add a dictionary to TDM in the **Administrator | Dictionaries** view. Users that have permissions to create generation rules have permissions to add dictionaries.

Dictionary generation can use the string, numeric, date, and binary data types. You can create a rule based on a dictionary column. The binary data type contains a serial number column that increases performance. Generate dictionary values for columns that have a finite number of values, such as states or countries.

You can create dictionary generation rule either from a flat file or relational database. You can generate data for a binary data type from relational databases but not from flat files.

Dictionary generation rules can be global rules or ad hoc rules.

The following image shows the dictionary generation parameters:

Dictionary Generation Example

You want to load customer names into the target database. You import a dictionary that has first and last name columns. You create two dictionary rules with the string data type. When you create the first name rule, you select the dictionary that contains the names and choose the first name dictionary column. In the last name rule, you select the same dictionary and choose the last name dictionary column.

Dictionary Generation Parameters

Create a dictionary generation rule to load dictionary values into target databases. Configure dictionary generation parameters for string, date, numeric, and binary data types.

The following table describes the parameters that you can configure to generate dictionary data:

Parameter	Description
Dictionary Name	The name of the dictionary. You need to import a dictionary in TDM.
Dictionary Column	The column that you want to get the data from. If the values in the dictionary are larger than the size of the column then it truncates the values.
Dictionary Lookup Column	The lookup column in the dictionary that you can match with the table column to generate data. When you assign a standard generation rule to a column or create an ad hoc generation rule, you can specify the lookup column. This field is optional. If you enter double data type value in the lookup column and table column, the process generates null values in the target table.
Table Column	The column in the target table based on which you can match the lookup column in the dictionary. When you assign a standard generation rule to a column or create an ad hoc generation rule, you can specify the table column. This field is optional. If you enter double data type value in the lookup column and table column, the process generates null values in the target table.

Parameter	Description
Serial Number Columns	A serial number to find the location of the data. You can configure serial number for relational databases and not for flat files. You enter the serial number column for binary data types.
Generate Unique Values	Generates unique data values. You cannot generate unique values for flat file dictionaries and binary data types.
Null Values	A list of global rules or a default rule that is available for the data type you select. A null value is a value that a default rule or a global rule generates in the target for the columns based on the data type you select.
Invalid Values	A list of global rules or a default rule that is available for the data type you select. Any value that does not exist in the dictionary is not valid. You cannot generate values that are not valid for binary data types.

Effective Dates Generation

An effective dates generation rule is an ad hoc rule that generates related date data for two columns. The effective dates are sequences of dates that consist of start dates and end dates. You can create effective dates rules with the date data type.

An effective dates generation rule generates a random start date within a range. The rule creates an end date that is after the start date and another start date that immediately follows the end date. The number of rows in the grouping column that have the same value determines the number of dates in the effective dates sequence.

The column in which you create the rule must be either the effective start date or the effective end date. If you create the rule on the effective start date column, you must select a different column as the effective end date column. If you create the rule on the effective end date column, you must select a different column as the effective start date column. The rule generates date values for both columns. You select the grouping column to generate sequences of dates in multiple rows.

The following image shows the effective dates generation parameters:

Effective Dates Generation Parameters

Create effective dates generation rule for the date data type.

The following table describes the parameters that you can configure to generate data for effective dates:

Parameter	Description
Start Date/Time	The date and time at which you can start to generate dates.
End Date/Time	The date and time at which you can stop to generate dates.
Exclude	The days of the week that you do not want to include in the generated dates. You can enter up to two days, and both the days cannot be same.
Effective Start Date	The column name in the target table where you want to put the generated effective start date. Default is the column name in which you create the effective dates generation rule. You can change it to any other available column name.
Effective End Date	The column name in the target table where you want to put the generated effective end date. You can select any column other than the effective start date.

Parameter	Description
Grouping Column	The column that determines how you group the data to generate effective dates in the target columns.
Invalid Values	Any value with effective end date greater than the effective start date. If effective start date is less than effective end date of the previous record for the grouping column, the values are not valid. This parameter has default rule selected and you cannot change the rule. You enter the percent for values that are not valid.

Effective Dates Generation Example

You want to generate start dates and end dates for customer addresses.

The target table has two customers. Each customer has one identification number. The table has three address rows for each customer. You choose a date range to generate a random effective start date for each customer. You select an effective start date column to put the start date into. You select an effective end date column to put an end date into. You select the customer identification number as the grouping column.

The rule generates a random start date for the first address row for each customer. The rule generates a random end date that occurs after the start date. The rule generates a start date for the next row that is after the end date of the previous row.

The following table shows a sample table with the generated data:

Customer ID	Address	Start Date	End Date
1111	55 Crestwood Dr	11/20/2001	06/17/2003
1111	200 Larkin St	06/18/2003	04/16/2004
1111	1000 Cooke Ln	04/17/2004	08/16/2012
2222	12708 Danbrooke Ct	03/06/2008	05/19/2010
2222	1200 S Sixth Ave	05/20/2010	07/06/2011
2222	2243 Creeksie Ct	07/07/2011	11/24/2013

Expression Generation

An expression generation rule is an ad hoc rule that generates data from an expression. You can create expression rules with the string, numeric, and date datatypes.

When you create an expression generation rule, you use the **Expression Builder** to create an expression. In the **Expression Builder**, you can choose from a list of columns and functions to include in the expression. To enter a join condition, you can choose columns and operators. Click the **Add** arrow to include the column in the expression.

If you select **Override Data Type to Numeric** for a column with string data type, you cannot enter a numeric expression.

The following image shows the expression generation parameters:

Expression Generation Example

You want to create a Name column that combines the data from the FirstName and LastName columns. You use the following expression to create the rule:

```
CONCAT (FirstName+'/'+LastName)
```

Expression Generation Parameters

You can configure expressions to generate data for string, date, and numeric data types.

The following table describes the parameters that you can configure to generate expressions:

Parameter	Description
Expression	An expression that you can configure to generate test data. If you select to override a string data type column with numeric data type, you cannot enter expression to generation numbers.
Null Values	A list of global rules or default rule that are available for the data type you select. A null value is value that a default rule or a global rule generates in the target column based on the datatype you select.
Invalid Values	A list of global rules or default rule that are available for the data type you select. Any value that does not follow the specified expression is not valid.

Random Generation

A random generation rule is a global or ad hoc rule that creates random string, numeric, and date values. You can use random generation to generate data such as salary and birth dates.

When you configure random generation, you can specify a minimum and maximum length for data strings and a minimum and maximum value for numeric data. You can include a regular expression to create a fixed percentage of the generated data. The percentages of the pattern distributions must total 100. When you generate dates, you can select a start date and end date to create random dates within a range.

Random generation rules can be global rules or ad hoc rules.

The following image shows the random generation parameters:

Random Generation Example

You want to generate addresses for customers. You create a rule that uses the random technique and the string data type. You define a regular expression in the rule. The regular expression creates rows with a four digit number, a random string of 12 characters, and "St." or "Ave." You set the pattern percentage to 100 to generate all the rows with the regular expression.

Date Random Generation Parameters

Create a random data generation rule for date data types.

The following table describes the parameters that you can configure to generate random data for date data types:

Parameter	Description
Start Date/Time	The date and time at which you can start to generate random dates.
End Date/Time	The date and time at which you can stop to generate random dates.

Parameter	Description
Exclude	The days of the week that you do not want to include in the generated dates. You can enter up to two days, and both the days cannot be same.
Null Values	A list of global rules or default rule that are available for the date data type. A null value is a value that a default rule or a global rule generates in the target column for the date data type.
Invalid Values	A list of global rules or default rule that you can select for the date data type. Any value that is less than the minimum or more than the maximum for a specified date range is not valid. Any date that falls on excluded dates is not valid.

String Random Generation Parameters

Create a random data generation rule for string datatypes.

The following table describes the parameters that you can configure to generate random data for string datatype:

Parameter	Description
Random	A range provided to generate random data. You can select Random to specify the range of the string for data generation.
Minimum Length	The minimum length for the string datatype. Default is 1.
Maximum Length	The maximum length for the string datatype. Default is the maximum value that the column permits.
From Regular Expression	A regular expression to generation random data. You can select From Regular Expression to create fixed percentage of data for generation.
Pattern	A regular expression that describes the format of generated data.
Distribution	The percentage of rows with the regular expression data. If you want to use more expressions, you can add more rows.
Generate Unique Values	Generates unique data values. You cannot generate unique values when you select From Regular Expression.

Parameter	Description
Null Values	A list of global rules or default rule that are available for the string datatype. A null value is a value that a default rule or a global rule generates in the target columns for the string datatype.
Invalid Values	A list of global rules or default rule that you can select for the string datatype. Any value that is less than the minimum length or more than maximum length for a specified range is not valid. A string that does not match the specified pattern is not valid. You cannot generate values that are not valid when you select From Regular Expression.

Numeric Random Generation Parameters

Create random data generation rule for numeric datatypes.

The following table describes the parameters that you can configure to generate random data for numeric datatype:

Parameter	Description
Random	A rule to generate random data. You can select Random to specify the range of numeric values for data generation.
Minimum Value	The minimum value for the numeric datatype. Default is 0.
Maximum Value	The maximum value for the numeric datatype. Default is the maximum value that the column permits.
From Regular Expression	A regular expression to generation random data. You can select From Regular Expression to create fixed percentage of data for generation.
Pattern	A regular expression that describes the format of generated data.
Distribution	The percentage of rows with the regular expression data. If you want to use more expressions, you can add more rows.
Generate Unique Values	Generates unique data values. You cannot generate unique values when you select From Regular Expression.

Parameter	Description
Null Values	A list of global rules or default rule that are available for the numeric datatype. A null value is a value that a default rule or a global rule generates in the target for the numeric datatype.
Invalid Values	A list of global rules or default rule that you can select for the numeric datatype. Any value that is less than the minimum or more than the maximum for a specified range is not valid. A value that does not match the specified pattern is not valid. You cannot generate values that are not valid when you select From Regular Expression.

Data Patterns for Random Generation

You can enter data patterns from regular expressions to generate string and numeric data.

To generate numbers that contain special characters or any other operators, you use random string data generation technique. You can use the following operators to generate string data patterns: `. , \d , \w, (opt1| opt2|…..), {} , []`.

To generate numbers that do not contain special characters or any other operators, you use random numeric data generation technique. To generate numeric data, you can combine the following patterns: `\d, alternates (1|2|3|…), and [0-9]`. You cannot nest the alternates.

When you enter data patterns to generate the credit card number, Social Security number, and Social Insurance numbers, the generated data might not be valid. These numbers follow certain algorithms and you cannot use data patterns to generate valid numbers.

String Data Pattern Examples

You want to generate email addresses, credit card numbers, phone numbers, Social Insurance numbers, Social Security numbers, dates, IP addresses, and employee IDs from regular expressions.

Email Address

To generate email addresses, you must enter the pattern in the following order:

1. Enter the first eight characters as lowercase alphabet characters [a-z].
2. Enter the sign (@).
3. Enter hotmail, yahoo, gmail, aol, or Comcast.
4. Enter .com, co.uk, or .net.

To generate an email address, use the following pattern: `[a-z]{8}@(hotmail|yahoo|gmail|aol|Comcast)\.(com|co\.uk|net)`. To generate a data pattern for uppercase and lower alphabet characters, you can enter `([a-z]|[A-Z]){8}@(gmail|hotmail)\\.(com|co\\.uk)`.

Credit Card Number

To generate a credit card number, you enter the first character and the pattern.

The following table shows the output for sample patterns that generate credit card numbers:

Pattern	Sample Output
4\d{3}-\d{4}-\d{4}-\d{4}	4312-0036-5436-0876
3\d{4} \d{4} \d{6}	3852 0000 023237

Phone Number

The following table shows the output for sample patterns that generate phone numbers:

Pattern	Sample Output
\d{3}\.\d{3}\.\d{4}	617.555.1212
\d{3}-\d{3}-\d{4}	617-555-1212
\(\d{3}\) \d{3}-\d{4}	(617) 555-1212
\+\d\.\d{3}\.\d{3}\.\d{4}	+1.617.555.1212
\d\.\d{3}\.\d{3}\.\d{4}	1.617.555.1212

Social Insurance Number

The following table shows the output for sample patterns that generate Social Insurance numbers:

Pattern	Sample Output
\d{3}-\d{3}-\d{3}	289-645-236
\d{3}\.\d{3}\.\d{3}	289.645.236
\d{3} \d{3} \d{3}	289 645 236

Social Security Number

The following table shows the output for a sample pattern that generates a sample Social Security number:

Pattern	Sample Output
\d{3}-\d{2}-\d{4}	999-99-9999

Date

The following table shows the output for sample patterns that generate dates:

Pattern	Sample Output
(1|3|5|7|8|10|12)-([1-9]|[12][0-9]|30|31)-(200[0-9]| 201[0-3])	5-30-2013
(4|6|9|11)-([1-9]|[12][0-9]|30)-(200[0-9]|201[0-3])	11-22-2012
(0[1-9]|1[0-2])\/((0[1-9]|2\d)|3[0-1])\/(19\d\d|200[0-3])	12/31/2003

IP Address

The following table shows the output for sample patterns that generate IP addresses:

Pattern	Sample Output
([0-9]|[1-9][0-9]|1[0-9]{2}|2[0-4][0-9]|25[0-5])\.([0-9]|[1-9][0-9]|1[0-9]{2}|2[0-4][0-9]|25[0-5])\.([0-9]|[1-9][0-9]|1[0-9]{2}|2[0-4][0-9]|25[0-5])\.([0-9]|[1-9][0-9]|1[0-9]{2}|2[0-4][0-9]|25[0-5])	255.255.0.0
(25[0-5]|2[0-4][0-9]|[0-1]{1}[0-9]{2}|[1-9]{1}[0-9]{1}|[1-9])\.(25[0-5]|2[0-4][0-9]|[0-1]{1}[0-9]{2}|[1-9]{1}[0-9]{1}|[1-9]|0)\.(25[0-5]|2[0-4][0-9]|[0-1]{1}[0-9]{2}|[1-9]{1}[0-9]{1}|[1-9]|0)\.(25[0-5]|2[0-4][0-9]|[0-1]{1}[0-9]{2}|[1-9]{1}[0-9]{1}|[0-9])	192.168.0.1

Employee ID

The following table shows the output for sample patterns that generate employee IDs:

Pattern	Sample Output
([A-Z][a-z]{2}_\d+)	Idd_1111
([A-Z][A-Z]-\d\d\d\d)	ID-4321

Numeric Data Pattern Examples

You want to generate credit card numbers, phone numbers, Social Insurance numbers, and Social Security numbers from regular expressions.

To generate a 10-digit phone number, consider the following scenarios:

- A phone number that starts with 3 or 4, you enter the data pattern as (3|4)\d{9}.
- A phone number that starts with 6 and consists of digits such as 2,4,6, and 8, you enter the data pattern as 6(2|4|6|8){9}.
- A phone number that starts with 6 and consists of digits between 3-7, you enter the data pattern 6 [3-7]{9}.

The following table shows the sample output for the patterns that you can enter:

Example	Pattern
Credit card number	You enter the pattern as 5\d{15} to generate a sample output as 5312003654360876.
Phone number	You enter the pattern as \d{10} to generate a sample output as 6175551212.
Social Insurance number	You enter the pattern as \d{9} to generate a sample output as 289645236.
Social Security number	You enter the pattern as \d{9} to generate a sample output as 999999999.

Reference Lookup Generation

A reference lookup generation rule is an ad hoc rule that generates data from a reference table. You can create reference lookup rules with the string, numeric, and date data types.

Create a reference lookup generation rule to generate column data using data from an existing table. Specify the parent table and assign the parent columns to child columns in the target database.

You must assign a parent table and parent column to the target column that you create the rule for. When you create the rule you can also assign parent columns from the same parent table to other target child columns.

The following image shows the reference lookup generation parameters:

Reference Lookup Generation Example

You want to generate customer identification numbers in a column. You want to use the same customer identification numbers that exist in a different table. You create a reference lookup generation rule that references the column in the other table and generates the values from that table.

Reference Lookup Generation Parameters

Create reference lookup rules for string, numeric, and date data types.

The following table describes the parameters that you can configure to generate data from a reference lookup table:

Parameter	Description
Minimum Value	The minimum number of child records for a parent record. Default is 1.
Maximum Value	The maximum number of child records for a parent record. Default is 3.
Data Source	The connection to the parent table.
Parent Table	The parent table that you can get the data from.
Child Column	The column name on which you create the rule. You cannot remove the first child column. To enter additional child columns, you can add more rows.

Parameter	Description
Parent Column	The parent column for the corresponding child column. You add more rows to enter additional child and parent columns. You can also assign parent columns from the same parent table to the other target child columns.
Parent Table Filter	A conditional expression that you enter to extract the data. The data you extract from the parent table must conform to this condition.
Parent records without children	The percent of parent records without a child.
Child records without a parent	The percent of child records without a parent. You can enter a value greater than 0. If you enter 0, you must provide a default value instead of null values.
Default Value	The default value that you want to generate in the target column instead of null values. The default value can be one of the values present in the parent table if the tables are in a parent-child relationship. Required if the percent of child records without a parent is set to 0. Optional if the percent of child records without a parent field is blank. Disabled if you enter a non-zero value in the percent of child records without parent field to generate null values.

Sequence Generation

A sequence generation rule is a global or ad hoc rule that generates numeric and date values in a sequence.

You select a beginning value and an increment value when you create a sequence generation rule.

When you create a sequence generation rule for a numeric data type, you can choose to start the sequence from the last output value. When you run the plan again, the sequence continues to generate numbers from the previous output value

If you try to generate numbers in a column with string data type and if you choose to start the sequence from the last value, TDM generates values from the beginning when you run the plan again.

The following image shows the sequence generation parameters:

Sequence Generation Example

You want to generate ticket numbers in ascending order. The target table has a TICKET_NUMBER column. You create a sequence generation rule with the numeric data type. You enter a start value of 1000 and an increment value of one (1).

Date Sequence Generation Parameters

Create sequence generation rule for date data types.

The following table describes the parameters that you can configure to generate a sequence of data for date data type:

Parameter	Description
Start Date	The date at which you can start generation of sequence data for date data types.
Increment by	The value to increment after the start date. You can enter a number and select year, month, day, hour, minutes, or seconds.
Null Values	A list of global rules or default rule that are available for the date data type. A null value is a value that a default rule or a global rule generates in the target column for the date data types.
Invalid Values	A list of global rules or default rule that are available for the date data type. Any value that is less than the sequence start value for a positive increment or more than the sequence start value for a negative increment is not valid.

Numeric Sequence Generation Parameters

Create sequence generation rule for numeric data types.

The following table describes the parameters that you can configure to generate a sequence of data for numeric data type:

Parameter	Description
Start Value	The minimum value to start generation of a sequence. You can enter a number to generate the sequence data.
Increment Value	The value to increment after the start value during generation of the sequence data. You can enter a number to increment the value.
Start Sequence from Last Value	Starts the sequence from the last output value when you run the plan again. **Note:** If you try to generate numbers in a column with string data type and if you choose to start the sequence from the last value, TDM generates values from the beginning when you run the plan again.
Null Values	A list of global rules or default rule that are available for the numeric data type. A null value is value that a default rule or a global rule generates in the target column for the numeric data type.
Invalid Values	A list of global rules or default rule that are available for the numeric data type. Any value that is less than the sequence start value for a positive increment or more than the sequence start value for a negative increment is not valid.

Set of Values Generation

A set of values generation rule is a global or ad hoc rule that writes a small data set to the target table. You can list the data values that you want to write to the table when you create a set of values generation rule.

You can use the string, numeric, and date data types in a set of values generation rule. You add data values in the rule and specify what percent of the rows in the target table you want to have the value. You can enter the percent in decimal values. You can use set of values generation for columns that have a small number of values, such as membership status.

Set of values rules can be global rules or ad hoc rules.

The following image shows the set of values parameters:

Set of Values Generation Example

You want to generate payment information. The target database has a PAYMENT_TYPE column with Visa, MasterCard, and Paypal values. You want 60% of the rows to have the Visa value, 30% of the rows to have the MasterCard value, and 10% of the rows to have the Paypal value.
You create a set of values rule with the string data type. You add Visa in the value field and enter 60 in the distribution percent field. You click the add button twice to add two additional values to the rule. In the second row, you enter MasterCard in the value field and 30 in the distribution percent field. In the third row, you enter Paypal in the value field and 10 in the distribution field.

Set of Values Generation Parameters

Configure set of values parameters to generate data for string, date, and numeric data types.

The following table describes the parameters that you can configure to generate a set of values:

Parameter	Description
Value	The exact data value that you want to write to a target table. You can enter exact values for the string and numeric data types. You enter a date for date data type. If you want to use more than one value, you can add more rows.
Distribution%	The percent of rows for the specified value. The sum of the distribution of the values have to be 100%. You can enter a decimal value in the distribution percent field.
Null Values	A list of global rules or default rule that are available for the data type you select. A null value is a value that a default rule or a global rule generates in the target columns based on the data type you select.
Invalid Values	A list of global rules or default rule that are available for the data type you select. Any number, date, or string that is not a part of the set of values is not valid.

CHAPTER 11

Working with Test Data Warehouse

This chapter includes the following topics:

- Test Data Warehouse Overview, 203
- Test Data Warehouse Process, 204
- Data Sets, 204
- Creating a Data Set, 206
- Reset a Data Set, 206
- Editing the Metadata of a Data Set, 213
- Deleting a Data Set, 213
- Data Set Permissions, 214
- Locking and Unlocking a Data Set Version, 215
- Monitor a Data Set Job, 215

Test Data Warehouse Overview

Create a test data warehouse to create a central location that stores multiple versions of test data. Multiple Test Data Manager services can access the same test data warehouse to create a central storage location across teams.

Create and manage the test data warehouse from Test Data Manager. To create a test data warehouse, you configure a test data warehouse repository and a test data warehouse.

Data is stored as data sets in the test data warehouse. The test data warehouse repository stores the project metadata and the source table metadata associated with the data set. Project metadata includes the plan information, version, and tag information. The test data warehouse stores the source data that you included in the data set.

You can create multiple versions of a data set to maintain different versions of the test data. To return a test environment to a previous state, you can choose to restore a specific version of the data from the test data warehouse.

For example, before you run any test cases, you create a data set to store the original test data in the test data warehouse. After each team runs its tests and the data is modified, each team creates another version of the data set to store the modified test data. To return the test environment to its original state, you can

restore the version of the data set that contains the original data. You can retrieve any of the other data set versions to return a test environment to a required state.

Test Data Warehouse Process

Store data in the test data warehouse to manage multiple versions of test data in a central location.

The following image shows the tasks that you perform to store data in the test data warehouse and the tasks that you can perform on the data in the test data warehouse:

You can perform the following tasks on data that you store in the test data warehouse:

Edit metadata

You can edit the tags and the description that you add to a data set.

Change permission settings

You can change user permission levels for a data set to control the level of access to the data set.

Lock and unlock a data set

The administrator and the data set owner can lock and unlock a data set. Lock a data set to prevent users from editing or resetting a data set to a target test system. For example, if you want to check the data before users reset a system with the data. A data set owner can perform a reset on a data set that is locked.

Reset

You can reset a data set to a target system to return the test data on the system to a specific state.

Delete

You can delete a data set from the test data warehouse.

Data Sets

A data set is a collection of data that consists of one or more tables from a data source. Manage and store multiple versions of data sets in the test data warehouse.

Tables in a data set can be from a single data source or from multiple data sources. A data set generally corresponds to one or more test cases. You can create data sets based on the test cases that you want to

run, and store them in the test data warehouse. You can create a data set from TDM when you perform one or more of the following tasks:

- Create a subset of production data.
- Create a masked copy of production data.
- Generate data.

To create a data set, select the test data warehouse as the target when you run a plan. You can add tags and other metadata information to identify the data set. Metadata includes the source table metadata and plan metadata. You can search the test data warehouse based on the metadata properties.

Test data might change when you use the data to run test cases. You can store the changed data as another version of a data set in the test data warehouse. You can create multiple versions of a data set.

You can create another version of a data set in the following ways:

- Rerun the plan that you used to create the data set without changing the name of the data set.
- Create and run a different plan and enter the same data set name.

Different versions of a data set can have similar metadata or different metadata. For example, you create a data set version DA_APP when you test version 1.0 of an application for a specific customer. You test specific features of the application in this release. You use source data that contains tables relevant to the tests you run. You might add tags to identify the customer. When you test the next version of the application for the customer, you test different features that require different test data. You create and run a different plan with a different data source. Because this is test data for a single application, you enter the same data set name to create another version of the data set. The version retains the tags that identify the customer. This version contains different data and metadata.

Test data on a test environment can get corrupted. You might require the original test data to run additional tests. You can return a test environment to a specific previous state when you reset a specific data set from the test data warehouse to the required connection.

For example, consider a situation where test teams run multiple test cases, or when multiple test teams work on an application. Store the original test data as a data set in the test data warehouse. When one test team completes testing, create another version of the data set with the modified test data. You can then reset the original test data from the test data warehouse to the test environment. You can reset any version of a data set to return the test environment to the required state.

View the list of data sets in the **Data Sets** view.

Data Set Tags

You can add tags to a data set to classify and identify each version of the data set.

A data set contains test data that you require to run specific test cases. You can add the name of the test case, an identification number, or any relevant information as a tag to identify the data set.

You can perform a keyword search for data sets based on tags. When multiple test teams create data sets, the teams can add tags to identify the data sets. Share tag information to help users search for a data set that you want to share.

You can also use tags to filter and search for data sets in the **Data Sets** view.

You can add tags when you create a data set. When you create another version of a data set, the new version inherits the tags added to the previous version. You can edit or delete the inherited tags and add tags when you create another version of the data set. You can add multiple tags to a data set. Use a comma to separate the tags.

You can add tags to an existing data set from the **Overview** tab on the data set page.

Creating a Data Set

Run a plan to create a data set. Create and run different plans for data generation operations.

- You cannot create a data set if the tables contain columns with user-defined data types. If a table contains columns with user-defined data types, mark the columns as restricted before you include the tables in a plan to create a data set. To mark a column as restricted, open the project and click **Discover** > **Columns** and select the option **Restricted** for the required column.
- Verify whether the tables in the data source contain primary keys. Create logical keys in all transactional tables that do not have any primary keys. A reset operation might fail if a transactional table with null values does not have primary keys.
- If a table contains more than two character columns with a precision of 4000 each, edit the precision to 3,999 or less. A reset operation might fail if a data set contains more than two character columns with a precision of 4,000 each and you do not disable constraints during the reset operation.

1. Open a project and add the required data sources and the project components.
2. To create a plan, click **Execute** and then click **Actions** > **New**.
3. Enter a name and optional description for the plan. Click **Next**.
4. Add and configure the required components.
5. Configure the plan settings.
6. To create a data set, select **Test Data Warehouse** from the list of target connections.

   The **Data Set Properties** section appears.
7. Enter a data set name. TDM enters the plan name by default. You can edit the name.

   When you enter a data set name that does not exist in the test data warehouse, you create the first version of a data set. TDM enters the data set version number.
8. Enter an application name and application version to identify the data set.
9. Add tags to the data set. Use a comma to separate tags.
10. Configure the error and recovery settings and advanced settings as required.
11. Optionally, configure an override strategy for a data source or a table.
12. To save the plan, click **Finish**.
13. To generate and run the workflow, click **Actions** > **Generate and Execute** in the **Plan Properties** page.

You can view the job log messages in the **Monitor** page. When the job completes successfully, you can view the data set in the **Data Sets** view.

Related Topics:

- "Project Components" on page 42
- "Creating a Project" on page 44

Reset a Data Set

Reset is the process of moving data from a data set in the test data warehouse to a target.

During the course of testing, the test data might change. You might want to reset the data in a test environment with a version of the data set from the test data warehouse.

For example, you are responsible for testing certain features of an application. You create test data to use in the test cases that you run. Before you use the data, you save the data as a data set in the test data warehouse. When you complete tests on some features, the data might change or get corrupted. You might require the original test data to run the rest of the test cases. You can replace the data in the test system with the data set in the test data warehouse. Perform a reset operation from the data set to a target. Use the test system as the target.

You can configure a reset operation to delete the target data and replace it with the data in the data set. In this case, the reset operation restores the target data to the exact same state as the data set. Perform this reset operation if testers do not share the tables. You do not need records that exist in the target and not in the data set.

You might want to retain data in the target that is not in the data set. Consider a case where multiple testers use the test data. You need the first 100 records in a Customer table. You create a data set that contains the first 100 records of the Customer table. After running a few tests, you want to replace the test data with the data set. You must configure the reset operation to ensure that the additional data in the target remains untouched when you reset the data in the data set.

In reset operations that selectively delete and load data, TDM runs an SQL script to create a backup of the deleted data in the target.

You can select the required configuration when you configure the reset operation.

Classification of Data Set Tables

You can classify the tables in a data set as master tables or transactional tables. The results of the reset operation depend on how you classify the tables.

A reset operation does not delete records in master tables. It performs an upsert on master tables. When testers share all records in a table, classify the table as a master table. Because the reset does not delete records in master tables, the reset operation does not delete records that might impact the work of other testers.

Classify tables that are descendants of master tables as transactional tables. Records in transactional tables are generally not shared records. When you delete a record, you do not impact the work of another tester. By default, during a reset operation TDM performs an insert on transactional tables. You can choose to change the default behavior when you configure the reset operation.

If you do not classify a table, by default a reset operation considers the table as a transactional table.

TDM performs the following tasks during an upsert:

- Updates records that exist in both the data set and the target.
- Inserts records that are present in the data set and not in the target.

During an insert, TDM inserts records from the data set that do not exist in the target. If the record exists, the insert fails.

Based on how you classify tables as master and transactional, TDM identifies top-level transactional tables and begins the cascade delete at this level.

A top-level transactional table is a direct child of a master table. All parents of a top-level transactional table, if any, are master tables.

Verify the table classification before you perform a reset operation.

Classifying a Data Set Table

Classify tables in a data set as master tables or transactional tables to control the results of a reset operation.

1. Open the required data set.
2. Click the **Plan Components** tab.
3. Click the **Tables** tab in the panel below the list of plan components to display a list of tables and the properties.
4. Click **Edit Table Type** to open the **Edit Table** dialog box.
5. Select the required table type from the list of table types.
6. If you classify a table as a master table, you must choose one of the following options:
   - Mark all descendant tables as transactional tables. TDM marks all related descendant tables as transactional tables.
   - Mark all other tables as transactional. TDM marks all related descendant tables and all unrelated tables as transactional tables.

   A list of the tables in the data set with the impact of the classification on the tables appears. The following image shows the **Edit Table** dialog box:

7. Click **OK**.

Reset With and Without Downtime

You can choose to disable constraints and indexes when you perform a reset operation.

Disabling constraints during the reset might increase the performance of the reset operation because the constraints do not exist. TDM performs delete and load in multiple tables in parallel and the reset operation might take less time to complete.

When you do not disable constraints between tables during a reset operation, TDM must check the order in which deletes and inserts happen. TDM cannot perform load and delete for multiple tables in parallel. TDM must consider constraints before deleting records. The reset operation might take longer to complete than when you disable constraints and indexes.

You apply constraints between tables. When you disable constraints, you disable constraints for the entire table. During the time the constraints are disabled, tests that you run using the tables might fail as the constraints do not exist. Consider a case where multiple testers use the test data. You create a data set that contains the first 100 records of the Customer table. After running a few tests, you want to reset the test data with the data set. If you disable constraints for the table during the reset operation, tests run by other testers that include the Customer table might fail. The reset operation causes a downtime for running tests that use the target connection.

While disabling constraints can increase performance, there might be a tradeoff with a downtime in testing if the target tables are shared among testers.

If multiple testers use the tables, a slower reset operation might be better than a downtime in testing for multiple testers. If the target tables are not used by other testers, and you are sure that no tests that include the tables are run during the reset, you can increase the performance of the reset by disabling constraints.

Analyze the requirement and consider this tradeoff when you configure a reset operation.

Resetting a Data Set

You can reset a specific version of a data set to a target. Reset a data set to a target to reuse a specific version of test data or to reset test data to an earlier state. You can configure the reset operation based on how you want to replace the test data.

Ensure that you classify the tables in the data set before you perform a reset operation. TDM considers unclassified tables as transactional tables during a reset operation.

1. Open the required version of the data set.
2. Click **Actions** > **Reset**.

   The **Reset** dialog box opens.
3. Select the target connection from the list of connections.

   The source is the test data warehouse. You cannot edit the source. Select the database where you want to replace the data as the target connection.
4. Configure the settings for the reset operation:
   - Truncate Tables. Select this option if you want to completely replace the target connection with data from the data set. TDM deletes all data in the target, and copies the data in the data set to the target. TDM disables constraints and indexes when it performs this operation.
   - Disable Constraints and Indexes. Select this option if you want to disable constraints and indexes during the reset operation.
   - Upsert in Transactional Tables. Select this option if you want to perform an upsert on transactional tables instead of an insert. If the parent records of records in the data set have changed to point to a different parent in the target, the insert might fail. The reset operation fails if the insert fails. Select this option and perform an upsert on transactional tables.
5. Enter the reset recovery properties.
   - Backup Recovery Prefix. Enter a prefix for the backup schema. TDM creates a backup schema in the format *<prefix>_t<ablename>*.
   - Backup Recovery User Name. The name of the schema where you want to store the backup tables.

You cannot enter the reset recovery properties if you choose to truncate tables. TDM does not create a backup when it truncates tables.

6. Edit the **Error and Recovery** and **Advanced Settings** as required.
7. Click **Next**.
8. Optional. You can override the reset settings at the data source level or at the table level.
9. Select the PowerCenter Integration Service to use for the reset operation. You must select the service that you configured in the Test Data Manager Service.
10. Select to run the operation or schedule it to run later.
11. Click **OK**.

You can view the job log messages in the **Monitor** tab.

Reset Process Flow

You can perform a reset operation that truncates tables or a reset operation that does not truncate tables.

Perform a reset with the truncate tables option when you own the target schema or when you do not share the data. By default, TDM disables the constraints and indexes during the reset. The downtime that this causes does not impact the work of other users because you do not share the tables.

Do not truncate tables if you share the table data and you need to reset a part of the data. You might delete data that other users need.

The following image indicates the steps that a reset operation goes through:

Note: The image does not include tasks performed during backup.

Reset with Truncate Tables Process Flow

You can choose to truncate the tables when you perform a reset operation.

TDM performs the following tasks when you perform a reset operation and select the **Truncate Tables** option:

1. Disables the constraints and indexes.
2. Deletes all tables in the target.
3. Inserts all tables from the data set to the target.

4. Enables the constraints and indexes.

Reset Without Truncate Tables Process Flow

You can perform a reset operation without truncating the tables.

TDM performs the following tasks when you perform a reset operation and do not select the truncate tables option:

1. Identifies records for deletion from transactional tables that have child tables.
2. If there are no records to delete, TDM skips this step. If records for deletion exist, TDM performs the following tasks.
   a. Creates staging tables for the identified records and stages the keys.
   b. Identifies top-most transactional tables in the data set.
   c. For each record in a top-most transactional table in the target, if a record exists in the data set, TDM marks the record in the staging table.
   d. Runs an SQL script to cascade mark records in the staged tables for deletion. If a parent record is marked for deletion, TDM marks the child record for deletion.
   e. Checks if recovery is enabled for the reset operation and that the user name and prefix information are available.
   f. If the recovery option is enabled and the prefix information provided, TDM performs the following tasks. If the recovery option is not enabled, TDM skips this step.
      a. Creates temporary tables in the backup schema and then moves records from the staging area to the backup schema.
      b. Stores a backup of all records marked for deletion in the backup tables in the backup schema. TDM runs an SQL script to back up the tables and uses the naming convention *<prefix>_<targetname>*.
3. Checks if the **Disable Constraints and Indexes** option is enabled. If the option is enabled, the reset operation performs the following steps:
   a. Disables the constraints and indexes.
   b. Deletes marked records from the target. TDM reads the marked records from the staging tables.
   c. Moves data from the data set to the target. TDM processes multiple tables in parallel because the constraints and indexes are disabled.
   d. Enables the constraints and indexes.
   If the option is not enabled, the reset operation performs the following steps:
   a. Deletes marked records from the target. TDM reads the marked records from the staging tables.
   b. Moves data from the data set to the target. TDM processes the tables according to constraints because the constraints exist.
4. Deletes the temporary tables, if created.
5. Deletes the staging tables, if created.
6. If the reset operation is successful, TDM deletes the backup tables. If the reset operation is not successful, TDM retains the backup tables and displays an error message.

Restoring Deleted Data After a Failed Reset

If you need to restore data that was deleted during a failed reset operation, you can restore the data from the backup created during the reset. Run a separate plan to restore the data from the backup. Ensure that you have the backup prefix used during the failed reset operation.

Perform the following high-level steps to restore data from the backup schema:

1. Create a group that contains all the tables that were backed up during the reset operation.

   The backed up tables include all transactional tables.
2. Open a project and create a plan.
3. Add the group as a subset component to the plan.
4. Add the connection that contains the backup schema as the source connection.
5. Add the database where you want to restore the data as the target connection.
6. In the target settings in the plan, disable the constraints and indexes.

   Select the following values for the **Update Strategy** properties:

   - **Treat Source Row As**: Update.
   - **Update As**: Update else Insert.
7. In the **Advanced Settings**, enter the backup prefix in the **Reset Recovery Prefix** field.
8. Run the plan.

Editing the Metadata of a Data Set

You can edit the description and tags of a data set.

1. To open the data set, click the required version of the data set in the **Data Sets** view.

   The data set page opens.
2. To edit the description or tags, click **Edit** on the **General** or **Tags** tab.
3. Add or edit the description as required. You can enter a maximum of 4,000 characters.
4. To add a tag, enter the text in the **Tags** field. You can use special characters and spaces in a tag. Use a comma to separate tags.
5. Click **Save**.

Deleting a Data Set

You can delete a data set or specific versions of a data set from the test data warehouse. Delete a data set from the **Data Sets** view. When you delete a data set, you delete the data and the metadata related to the data set.

1. Open the **Data Sets** view.
2. You can delete a data set or specific versions of data sets.

- To delete a data set, select the data set. TDM selects all versions of the data set. You can select multiple data sets.
- To delete a specific version of a data set, select the version. You can select multiple versions of different data sets.

3. Click **Actions** > **Delete**.

Data Set Permissions

Assign permissions to a data set to control the level of access to the data set.

When you run the plan that creates the data set, you become the owner of the data set. A data set owner and the domain administrator can add users and user groups and assign the required level of access. You can access data set based on the permissions that you have. If you do not have the required permissions, you cannot open data sets that appear in search results.

Data sets have the following levels of permission:

- Read
- Write
- Execute

To perform a task that uses a data set, you must also have the minimum required level of privileges as a TDM user.

The following table lists the permission levels, the tasks that you can perform with each level, and the minimum required privileges for each task:

Permission	Description	Minimum Required Privilege
Read	- Open and view the data set.	- View project
Write	- Open and view the data set. - Edit the data set description and tags. - Delete an unlocked data set.	- View project - Manage project
Execute	- Open and view the data set. - Reset a data set.	- View project - Execute project

The data set owner and the administrator can lock and unlock a data set. You cannot edit or reset a locked data set.

Editing Data Set Version Permission

When you create a data set version, you can provide read, write, and execute permissions to other users and user groups. Edit the data set version permissions assigned to users and user groups from the **Permissions** tab of the data set version.

1. Open a data set version and click **Permissions**.

   A list of the users and user groups with permissions for the data set version appears.

2. Click **Edit** on the **Users** or **User Groups** tab.

The **Edit Dataset Version Permissions** dialog box opens.

3. To edit the permission of a user or user group, select the user or user group from the list and edit the permission as required. You must save the changes for each user or user group.
4. To delete a user or user group, select the user or user group from the list and click **Delete**.
5. To add a user or a user group:
   a. Click **Add Users** or **Add User Groups**.
   b. Select one or more users or user groups.
   c. Optional. From the list of permissions, select the required permissions if either of the following statements is true:
      - You selected a single user or user group.
      - You want to assign the same levels of permission to all selected users or user groups.
   d. Click **OK**. TDM adds the users or user groups to the list.
   e. Select each user or user group and assign the required permission levels. You must save the changes for each user or user group. Skip this step if you performed step c.
6. Click **OK**.

Locking and Unlocking a Data Set Version

To prevent users from editing or resetting a data set version, you can lock the data set version. You cannot make changes to a locked data set version. The owner and the administrator can reset a locked data set version to a target, but cannot edit or delete the data set version when it is locked. You can reset a locked data set version if you need to check the data before you allow other users to edit or reset it. You can unlock the data set version when you want to allow edits to the data set version.

1. Open a data set version and click the **Overview** or **Permissions** tab.
2. To lock the data set version, click **Actions** > **Lock**.

   The data set version appears as locked in the **Data Sets** view. The **Edit** options do not appear in a locked data set version.
3. To unlock the data set version, click **Actions** > **Unlock**.

   The status of the data set version is updated in the **Data Sets** view. The **Edit** options appear in the data set version.

Monitor a Data Set Job

You can monitor all data set jobs in the **Monitor** view from the Test Data Manager home page. You can monitor jobs for a specific data set from the **Monitor** tab in the data set page.

You can view jobs created by any Test Data Manager Service in the domain. Identify the node based on the Test Data Manager Service listed in the properties.

You can monitor the following types of jobs for a data set:

- Created. A job that creates a data set in the test data warehouse.

- Reset. A job that resets a data set to a target.

CHAPTER 12

Plans and Workflows

This chapter includes the following topics:

- Plans and Workflows Overview, 217
- Workflow Connections, 218
- Plan Components, 219
- Plan Settings, 220
- Masking Components, 229
- Subset Components, 229
- Generation Components, 230
- Component Criteria, 230
- Source Settings, 231
- Plan Management, 238
- Workflow Generation, 241
- Executing a Workflow, 243
- Workflow Executions View, 243

Plans and Workflows Overview

A plan defines a data subset, data masking, or data generation operation. It includes the components that you need to generate a workflow.

When you create a plan, you add data masking, data subset components, or data generation components based on the operation that you need to perform. You cannot add data masking or data subset components to a plan when you want to add data generation components. You can add policies and rules to run data masking operations. Add entities, groups, and templates to run data subset operations on relational or flat file sources.

Add XSD files if you want to perform data subset or data masking on XSD sources. Add entities and tables or XSD files to run data generation operations.

If you add both subset and masking components to a plan, the subset component gets priority. TDM performs the subset operation first. TDM then applies masking rules to columns in the subset data that have masking assignments. TDM does not apply masking assignments to source data that is not included in the subset, even if the data is part of a cascade.

Define the workflow connections in the plan. Define the source and target connections. If required for the plan, edit staging and storage connections and define lookup and dictionary connections.

When you update or delete a plan component, you must generate and run the plan again to update the changes.

To copy flat file results to a test tool that you have integrated with TDM, configure the test tool properties in the plan.

To store data in the test data warehouse, you select the test data warehouse as the target in the plan.

You can create multiple workflows from one plan. Define workflow properties in the plan, such as commit properties, update strategies, and recovery strategies.

View the sessions that run from each workflow. View statistics that describe the number of rows that the target receives. The **Session Details** tab shows the affected rows, the rejected rows, and the number of rows that the session wrote to the target.

When you start a workflow, the PowerCenter Integration Service completes the plan operations.

Plans and Workflows Task List

Complete the following steps to create the plan and run the workflow:

1. Create a plan. Enter a plan name and description.
2. Add data masking rules and policies to the plan.
3. Add data subset entities, groups, and templates.
4. Add data generation entities and tables to the plan. When you enter data generation components, you cannot enter data masking or data subset components.
5. Switch masking rules off or on.
6. Optionally, change the criteria of data subset components.
7. Configure plan settings including the plan properties, error and recovery strategies, and advanced settings.
8. To store data in the test data warehouse, select the test data warehouse as the target connection.
9. To copy flat file results to a test tool that you have integrated with TDM, configure the test tool properties in the plan.
10. Optionally, configure an override strategy for a data source or a table.
11. Generate a workflow.
12. Run the workflow.
13. Monitor the workflow.

Workflow Connections

Workflows contain connections to the PowerCenter Integration Service, the TDM repository, and one or more connection objects.

An offline job requires three connections to the TDM repository. When you run a plan with multiple workflows, each workflow requires one connection. You can run a maximum of eight workflows at a time. When you run eight workflows, you need 11 connections. Additional workflows move into the queued state.

Each workflow can use different connections to relational sources and paths to flat files. Choose the connections in the plan.

When you create a plan, you can select the following connections:

- Source. The connection that is used to connect to the source.
- Target. The connection that is used to connect to the target.
- Staging connection. A connection to a database for staging data. The default connection entered in the Preferences view appears. You can change the connection for a plan.
- Lookup connection. A connection to a database that contains lookup tables.
- Storage connection. A connection to a database that contains storage tables. The default connection entered in the Preferences view appears. You can change the connection for a plan.
- Teradata connection. An ODBC connection to a Teradata database.
- Dictionary connection. A connection to a database that contains a dictionary table for substitution, email masking, and data generation.

Plan Components

When you create a plan, you add components to the plan based on the operations that you want to perform.

You can perform data masking operations, data subset operations, or both. You can also perform data generation operations. When you perform data generation operations, you cannot perform data subset or data masking operations.

The following table describes the components that you can add to a plan for each type of operation:

Component	Description
Rule	Data masking rules. For a data masking operation, you can add rules assigned to a policy and add rules that are not assigned to a policy. You can add rules to a data masking plan.
Policy	A Policy is a set of data domains. Each data domain can have multiple rules. You can add policies to a data masking plan.
Entity	Defines a set of tables that are related based on physical or logical constraints. An entity can contain parent tables and child tables. An entity maintains relational hierarchy in the tables. Define filter parameters for ports in the entity to extract the columns from the tables in an entity. When you run a workflow generated from a data subset plan, the PowerCenter Integration Service extracts source data based on the filter parameters defined in the entity and loads the data to the subset tables. If an entity is part of a template that is in the plan, you cannot add the same entity to the plan. Remove the template from the plan, or remove the entity from the template to continue. You can add entities to a data subset or a data generation plan.
Group	Defines a set of tables to copy to a target subset database. You can add groups to a data subset plan.
Template	A reusable component that contains the entities and groups of a specific module that you want to subset based on common parameters. You can add templates to a data subset plan.

Component	Description
Table	Defines the target table columns that contain generation rule assignments. You can add tables to a data generation plan.
XSD File	The XML source definition that you import from the PowerCenter repository. Add to a plan if you want to perform data subset or data masking on an XML source or if you want to generate XML data. You can add XSD files to a data subset, data masking, or data generation plan.

Plan Settings

Configure plan settings that apply to all data sources in the plan.

The plan settings include PowerCenter options, connection information, test data warehouse settings, test tool integration information, recovery strategies, and source and target properties.

Connection Options

Enter source and target connections.

The following table describes connection options:

Connection Options	Description
Treat the subset criteria as a variable	Appears if you select at least one subset component in a plan. Use the subset criteria as a variable only when there is an entity with a parameter value instead of a specific value in the entity criteria. When you select this option, you must enter the parameter file name and parameter file path.
Use Source Connection as Target Connection	Use the same connection for source and targets. Use with inplace masking.
Source Connection	A connection to the source database. Select a source connection from the list.
Source Connection Directory Path	Displays if flat file sources are present. Enter the path to the flat file source directory.
Target Connection	Choose variable, relational, test data warehouse, or flat file connection from the list. Select a target connection from the list. If you enter the target connection as a variable, ensure that you define the scope of the variable as **Global** or **Integration Service** in the parameter file. In a multinode setup, if you use parameters to enter the connection information, the parameter file must be present on the node on which the Test Data Manager Service runs. The file must be in the same directory path on both nodes.

Connection Options	Description
Target JDBC Connection	Appears if you select an ODBC target connection type. The connection that contains the JDBC connection string of the ODBC target database. Required if the target connection type is ODBC and either of the following statements is true: - You want to disable and enable constraints in the plan. - The source connection type is ODBC and you want to generate mappings with the source data type. Optional if the target is ODBC and you want to perform simple data movement operations.
Source File Directory	Appears if the project includes XML sources. Location of the source XML file. Enter the path to the XML source file directory. The source file must reside on the machine on which the PowerCenter Integration Service runs. Default is the PowerCenter Integration Service process variable directory $PMSourceFileDir.
Target File Directory	Appears if the project includes XML sources. Enter the path to the XML target file directory. The path must be on the machine on which the PowerCenter Integration Service runs. Default is the PowerCenter Integration Service process variable directory $PMTargetFileDir.
Parameter File Name	Appears if one or more of the following conditions are true: - You select variable source and target connections. - You choose to treat the subset criteria as a variable. - You enter the seed value as a parameter or variable in a masking rule that you use in the plan. The name of the parameter file that contains the entity parameter values and the variable names for source and target connections. Enter the name of the parameter file.
Parameter File Path	Appears if one or more of the following conditions are true: - You select variable source and target connections. - You choose to treat the subset criteria as a variable. - You enter the seed value as a parameter or variable in a masking rule that you use in the plan. The path to the file that contains the entity parameter values and the variable names for source and target connections. Enter the path to the parameter file.
Source File Name	The name of the IMS source file or VSAM source file. If the source is a Hierarchical IMS file, enter the name of the source unload file. If the source is a Hierarchical VSAM/SEQ file, enter the name of the source file. Required if the source connection is PXW_NRDB Batch.
Target File Name	The name of the IMS target file or VSAM target file. If the target is a Hierarchical IMS file, enter the name of the target unload file. If the target is a Hierarchical VSAM/SEQ file, enter the name of the target file. The target file name must be different from the source file name. Required if the target connection is PXW_NRDB Batch.
Teradata ODBC Connection	Name of the Teradata ODBC connection. Required if the target connection is Teradata PT. Select a connection from the list.
Dictionary Lookup Connection	Connection to a database that contains lookup tables. Required if the plan has a dictionary generation rule assignment or a mapplet rule that includes a transformation. Choose relational or flat file from the list. Select a dictionary lookup connection from the list.
Dictionary Connection	A connection to the database that contains the dictionary table. The dictionary connection option does not appear unless the project contains a rule that requires a relational dictionary.
Email Dictionary Connection	A connection to the database that contains the email dictionary table. The email dictionary connection option does not appear unless the project contains a masking rule that requires a relational email dictionary.

Connection Options	Description
Number of Records to Generate	Required. The number of records that you want to generate in the target table when you create a generation plan.
Reference Lookup Connection	A connection to the database that contains parent tables. Required if the generation plan has reference lookup rule assignment. Choose relational or flat file from the list. Select a dictionary lookup connection from the list.

Dictionary Connections

A dictionary is a flat file or relational table that contains substitute data. When you define a relational dictionary, you can define the connection to the dictionary in the project plan and workflow.

The **Dictionary Connection** field appears in the **Plan Settings** if the plan requires a connection.

When you define a substitution masking rule with a flat file dictionary, you can browse to a dictionary from Test Data Manager. Test Data Manager imports the structure of the dictionary file to enable you to define the serial number column and the lookup column.

By default, when you run a workflow, the PowerCenter Integration Service finds flat file dictionaries in the following location: `<Informatica installation directory>\server\infa_shared\lkpfiles`

You cannot change this location. The TDM installation installs the built-in dictionaries in the lkpfiles location. When you create a custom dictionary file, Test Data Manager copies the file to the `lkpfiles` directory.

Note: To access a dictionary on Microsoft SQL Server, create an ODBC connection using the DataDirect SQL Server Wire Protocol driver.

Data Set Properties

Configure the data set properties to use the test data warehouse as the target connection in a plan.

The data set properties appear when you select the test data warehouse from the list of target connections.

The following table describes the test data warehouse settings:

Property	Description
Name	Required. Name of the data set.
Description	Enter a description for the data set.
Data Set Version	The data set version. You cannot edit the version.
Application	The application name.
Application Version	The application version.
Tags	Tags for the data set. Enter the required tags separated by commas.

Test Tool Integration Options

Configure test tool integration in a plan to copy the flat file target results to a specific location in an HP ALM server. The administrator must first enable test tool integration for the TDM setup. You must run the workflow from Test Data Manager. The test tool integration job does not run when you run the workflow from the PowerCenter client.

The following table describes the test tool integration properties that you can configure:

Test Tool Integration Options	Description
Attach to Test Tool	Select this option to copy the flat file target results to a location in the HP ALM server integrated with TDM.
Project	Required. Browse to select the project in the HP ALM server where you want to store the flat file results.

Target Options

Configure properties for committing data and error recovery in the **Target** fields.

The following table describes target options:

Other Property Options	Description
Truncate Table	Truncates the table before loading it. By default, this is not selected.
Disable Indexes	Disables indexes for faster loading.
Disable Constraints	Disables physical constraints in the database before loading the data to the target and enables the constraints after the data loads. If TDM fails to enable the constraints after loading the data, the workflow fails. You must then manually enable the constraints in the database and run the workflow again. **Note:** When the target connection is DB2 for zOS and if you select disable constraints, TDM drops and recreates the constraints and moves the data. But the database tablespace goes into Check Pending state and you cannot perform operations on any of the tablespace. You must run the Check Data utility in the current database tablespace.
Optimize Relationship Type for Entities	Optimizes relations in an entity at the plan level. Optimizes relations to ensure a minimal number of major relations such that the operation includes all tables in the entity. **Note:** You must have provided criteria at the plan level when you select this option.

Update Strategy Options

Configure how the PowerCenter Integration Service updates the target in the plan update strategy options. Configure whether the source rows are insert or update rows. Determine how the PowerCenter Integration Service updates the target based on whether the rows exist in the target.

The following table describes the options for update strategy:

Update Strategy Option	Description
Treat source rows as	Indicates how the PowerCenter Integration Service treats source rows. Choose Insert or Update. If an imported PowerCenter mapplet contains a transformation that is configured to set the update strategy, the default option is Insert.
Update as	Determines whether the PowerCenter Integration Service updates or inserts rows. Choose one of the following options: - Update. The PowerCenter Integration Service updates all rows flagged for update. - Insert. The PowerCenter Integration Service inserts all rows flagged for update. - Update else Insert. The PowerCenter Integration Service updates rows flagged for update if they exist in the target, and inserts remaining rows marked for insert.

Data Generation Properties

Before you create a plan to generate data, you can define default settings in the **Administrator | Preferences** view. You can update the default settings when you create a data generation plan.

The default rules are not populated if the rules are not part of a particular project. If the default rules are blank, then you need to select the default rules that are applicable to the plan.
The following table describes the properties to update default settings for data generation:

Property	Description
Default Numeric Generation Rule	The default rule you have set to generate data for numeric datatype. You can select a separate default rule for numeric datatype.
Default String Generation Rule	The default rule you have set to generate data for string datatype. You can select a separate default rule for string datatype.
Default Date Generation Rule	The default rule you have set to generate data for date datatype. You can select a separate default rule for date datatype.
Default Binary Generation Rule	The default rule you have set to generate data for binary datatype. You can select a separate default rule for date datatype.
Apply Default Rule to	The default rule you have set for all the columns without generation rules. You can also apply the default rule to the columns that do not have null values.
Minimum	The minimum number of child records for a parent record. Default is 1. You can enter a minimum numeric value.
Maximum	The maximum number of child record for a parent record. Default is 1. You can enter a maximum numeric value.

Property	Description
Parent Records Without Children	The number of parent records that have no child. Enter the percent value for self-cyclic tables or reference lookup generation rules.
Child Records Without Parent	The number of child records that have no parent. Enter the percent value for self-cyclic tables or reference lookup generation rules.

Error and Recovery Options

Configure properties for committing data and error recovery in the **Error and Recovery** fields. The section is minimized by default. The options are populated with default data if available.

The following table describes the error and recovery options:

Error and Recovery Options	Description
Stop on Error	Indicates how many non fatal errors the PowerCenter Integration Service encounters before it stops the session. If you enter zero, the session does not stop for non fatal errors. Default is zero.
Suspend on Error	Suspend workflow processing when an error occurs. You can fix errors and then restart the workflow. The default is No.
Rollback Transactions on Error	Rolls back the transaction at the next commit point when it encounters a non fatal write error. The default is Yes.
Recover Strategy	Strategy for recovering a workflow when errors occur. Choose one of the following recovery strategies: - Resume from last checkpoint. The PowerCenter Integration Service saves the session state of operation and maintains target recovery tables. - Restart task. The PowerCenter Integration Service runs the task again when it recovers the workflow. - Fail task and continue workflow. Default. The PowerCenter Integration Service cannot recover the task, but it continues the workflow.
Enable High Availability Recovery	Select Yes to enable workflow recovery. High availability recovery is not available for web service workflows. The default is No.
Commit Type	Choose a source-based or target-based commit. By default, the PowerCenter Data Integration Service performs a target-based commit.
Commit Interval	The number of rows to process before a commit. Default is 10,000.
Error Log Type	Specifies the type of error log to create. You can specify flat file or none. Default is None.
Error Log File Directory	Required if you enable flat file logging. Specifies the directory where the log file is stored. Default is $PMBadFileDir\.

Advanced Options

The Advanced Options include options to set the locale, change the staging connection and storage connection, and set the number of tables to process in a workflow. You can also configure an email address

to send notifications to. This section is minimized by default. The options are populated with default data if available.

The following table describes advanced options:

Advanced Options	Description
Staging Connection	A connection to a database for creating temporary staging tables. The default set in the Workflow Generation section appears. You can change the connection used in a plan. Use this connection for scenarios that involve Cyclic Subset, Tending to Cyclic subset, Inplace Batch Updates, Non Relational Shuffle, and Hierarchical PWX. It is recommended that you use the source connection as the staging connection.
Storage Connection	A connection to a database that contains storage tables. The Data Masking transformation maintains the storage tables for repeatable substitution and expression masking between sessions. A storage table row contains the source column and a masked value pair. The storage tables for substitution and expression masking are separate tables, IDM_SUBSTITUTION_STORAGE and IDM_EXPRESSION_STORAGE.
XML date-time format	Appears if the source is an XML file. Date-time format defined in the session properties. Choose one of the following values: - Local Time. The time zone on the machine on which the PowerCenter Integration Service runs. - Local Time with Time Zone. The difference in hours between the PowerCenter Integration Service time zone and Greenwich Mean Time. - UTC. Greenwich Mean Time. Default is Local Time.
Datetime Format String	Date-time format defined in the session properties. You can enter seconds, milliseconds, microseconds, or nanoseconds. - Seconds. MM/DD/YYYY HH24:MI:SS - Milliseconds. MM/DD/YYYY HH24:MI:SS.MS - Microseconds. MM/DD/YYYY HH24:MI:SS.US - Nanoseconds. MM/DD/YYYY HH24:MI:SS.NS Default is microseconds.
Enable High Precision	Processes the Decimal datatype to a precision of 28. Default is No.
Tables Per Workflow	Sets the number of tables to process in a workflow. Default is 10.
Mapplet	Use to Reuse or Replace the mapplets in PowerCenter.
Max Parallel Sessions	The maximum number of sessions that can run at the same time. Applicable if you disable indexes and constraints in the target by selecting the options under Target options.
Locale	Sets the locale. A locale properties file corresponding to the locale you select must exist in the `<Informatica installation directory>/TDM/lang` location.

Advanced Options	Description
Data Integrity for Subset	Sets the integrity level for major constraints in the source data. Choose one of the following options: - Referential Integrity only. Ensures that the subset operation maintains referential integrity. This might fetch minimal data but might break transactional integrity. - Transactional Integrity for Major Relation Type. Ensures that the subset operation maintains referential and transactional integrity. This might fetch additional data, but honors referential and transactional integrity. Default is Transactional Integrity for Major Relation Type.
Email	An email address to send notifications to when the plan execution completes. Click settings to configure the email address, subject, and text of the email.
Target Load Type	Choose Normal or Bulk. If you select Normal, the PowerCenter Integration Service loads targets normally. Choose Normal mode if the mapping contains an Update Strategy transformation. Choose Bulk to improve session performance. Bulk load limits the ability to recover because no database logging occurs. Default is Normal.
Target Schema Name	Table name prefix. Specify the target schema name if the target is in a different schema. If the target is Cassandra, you must specify the target schema name.
Source Schema Name	Table owner name. Specify the source schema name if the source is in a different schema. Enter the schema name in uppercase unless the schema name is case sensitive in the database. If the schema name is case sensitive in the database, enter the name as in the database. If the source is Cassandra, you must specify the source schema name.
Batch Update	Enables batch update to update target rows in a batch mode. Improves performance. Use with inplace masking. When you use connection variables, ensure that the staging connection is same as the source connection.
Batch Size	The number of target rows that you want to move at a time. This option is available when you select the Batch Update as Yes. Use with inplace masking. **Note:** The batch size must be greater than the total number of records.
Use Oracle ROWID in Source	Determines whether to use the ROWID from Oracle sources as a column. The ROWID indicates the physical location of a row in the database. Enabled when the Oracle source and target connections are the same.
Join parent and child tables using	Choose how to join the parent and the child tables in the SQL. Choose one of the following options: - IN clause. At least one value in the parent table must match a value in the child table. - EXISTS clause. The join must return at least one row.
Enable special characters in metadata	Select whether to allow special characters in plan metadata. Choose one of the following options: - Yes. Select yes to allow special characters in metadata. - No. Select no to disable special characters in metadata. Default is No.
Reset Recovery prefix	Required if you are running the plan to recover records deleted during a failed reset operation. The backup recovery prefix that you entered during the reset operation.

Advanced Options	Description
Generate Mappings With Source Data type	Select Yes if the data source was created with the ODBC connection type and if the source table or column names contain special characters. Select Yes when you want to generate a mapping with the actual source data type if ODBC does not support a data type. Applicable for databases that do not have native drivers support on Linux.
Pre 85 Timestamp Compatibility	Trims subseconds to maintain compatibility with versions prior to 8.5. The Integration Service converts the Oracle Timestamp data type to the Oracle Date data type. The Integration Service trims subsecond data for the following sources and targets: - Relational sources and targets - XML sources and targets

Log Settings

Configure the log properties in the **Log Settings** fields. The section is minimized by default. The options are populated with default data if available.

The following table describes the log settings:

Log Settings	Description
Save Workflow Logs	You can create workflow logs according to the following options: - By Runs. The PowerCenter Integration Service creates a designated number of workflow logs. Configure the number of workflow logs in the **Save Workflow Log for These Runs** option. The PowerCenter Integration Service does not archive binary logs. - By Time Stamp. The PowerCenter Integration Service creates a log for all workflows, appending a time stamp to each log. When you save workflow logs by time stamp, the PowerCenter Integration Service archives binary logs and workflow log files.
Save Workflow Logs For These Runs	Number of historical workflow logs you want the PowerCenter Integration Service to create. The PowerCenter Integration Service creates the number of historical logs you specify, plus the most recent workflow log.
Write Backward Compatible Workflow Log	Writes workflow logs to a text log file. Select this option if you want to create a log file in addition to the binary log for the log events.
Save Session Log By	You can create session logs according to the following options: - Session Runs. The PowerCenter Integration Service creates a designated number of session logs. Configure the number of session logs in the **Save Session Logs for These Runs** option. The PowerCenter Integration Service does not archive binary logs. - Session Time Stamp. The PowerCenter Integration Service creates a log for all sessions, appending a time stamp to each log. When you save session logs by time stamp, the PowerCenter Integration Service archives binary logs and workflow log files.
Save Session Log For These Runs	Number of historical session logs you want the PowerCenter Integration Service to save. The PowerCenter Integration Service creates the number of historical logs you specify, plus the most recent session log.

Log Settings	Description
Session Log file Maximum Size	The maximum number of megabytes for a log file. Configure a maximum size to enable log file rollover by file size. When the log file reaches the maximum size, the PowerCenter Integration Service creates a new log file. If you set the size to zero the session log file size has no limit. Configure this option for real-time sessions that generate large session logs. The PowerCenter Integration Service writes the session logs to multiple files. Each file is a partial log file. Default is zero.
Session Log file Maximum Time Period	The maximum number of hours that the PowerCenter Integration Service writes to a session log. Configure the maximum time period to enable log file rollover by time. When the period is over, the PowerCenter Integration service creates another log file. Configure this option for real-time sessions that generate large session logs. The PowerCenter Integration Service writes the session logs to multiple files. Each file is a partial log file. Default is zero.
Maximum Partial Session Log Files	Maximum number of partial session log files to save. The PowerCenter Integration Service overwrites the oldest partial log file if the number of log files has reached the limit. Configure this option in conjunction with the maximum time period or maximum file size option. You must configure one of these options to enable session log rollover. If you configure a maximum of zero, the number of session log files is unlimited. Default is one.
Write Backward Compatible Session Log	Writes session logs to a text log file. Select this option if you want to create a log file in addition to the binary log for the log events.

Masking Components

Masking components are the policies and masking rules in a data masking operation. When you create a plan for a data masking operation, select the data masking components that you want to include in the workflow.

When you run a data masking operation, you can restrict the masking operation to some of the masking rules instead of applying all of the masking rules. You can choose the policies and masking rules that you want to test.

The **Masking Components** dialog box shows the policies and rules in the project. Select the policies and rules that you want to apply in the masking operation. The PowerCenter Integration Service performs masking with the components that you select. You can disable and enable masking for specific columns in a workflow run when you configure the plan criteria.

Subset Components

Subset components are the entities, groups, and templates or XSD files in a data subset operation. When you create a plan for a data subset operation, select the subset components that you want to include in the workflow.

When you run a data subset operation on a relational or flat file source, you can restrict the subset operation to specific entities in the project. You can choose the entities, groups, and templates that you want to include.

When you run a data subset operation on an XML data source, you can select the XSD files that you want to include in the operation.

The **Add Subset Components** dialog box shows the entities, groups, templates, or XSD files in the project. Select the components that you want to apply in the data subset operation. The PowerCenter Integration Service performs the data subset operation with the components that you select.

Generation Components

Generation components are the entities and tables in a data generation operation. If you select a target XML metadata, you must select the XSD file component in the data generation plan.

When you create a plan for a data generation operation, select the data generation components that you want to include in the workflow. To run a data generation operation, you can restrict the generation operation to some of the generation rules instead of applying all the generation rules. You can choose the generation rules that you want to test. When you configure the plan criteria, you can enter the number of records that you want to generate.

The **Generation Components** dialog box shows the entities and tables in the project. Select the components that you want to apply in the data generation operation.

Component Criteria

You can define filter criteria for data subset components and for data masking components in a plan. Define the filter criteria for each column you need to filter.

You can define expressions to limit data values for specific columns in a data subset operation. You can also limit the value to a percentage of rows, an interval of rows, or to an absolute value. You can disable and enable data masking for a column.

Define filter criteria for a column in the **Plan Components** panel. Choose the data masking or data subset component to filter.

Filtering Data Subset Components

You can define filter criteria for data subset components at the column level.

1. To create a data subset filter for a column, click the data subset component in the **Plan Components** panel.
2. Select a column from the **Criteria** panel to apply the filter to.

   Test Data Manager shows existing filter criteria. You can change the filter criteria or add new criteria.
3. Choose a method to filter the column data:

Option	Description
Custom	Create a basic expression or an advanced expression to filter the data.

Option	Description
Limit	Limit the subset results by percentage, by absolute value, or by an interval of rows to write to the target.

4. If you choose the **Custom** method, select one of the following options to define a filter:

Option	Description
Basic	Define a simple expression that includes just an operator and operand.
Advanced	Define an expression in the **Expression Editor**. You can include multiple columns and functions in the expression.

Disabling Masking for a Column

You can disable or enable masking for a column in the plan.

1. Click a masking component in the **Plan Components** panel.
2. Select a column from the **Criteria** panel.
3. To disable masking for the column, click **OFF**.
4. To enable masking for the column, click **ON**.

Source Settings

The source level properties are a subset of plan settings. You can change the settings for all the tables in the source or you can change the settings for each table in the source.

If a property is not available at the data source level, the property value at the plan level takes precedence. A lookup connection appears in the settings if there is a mapplet rule that is associated with a column in the data source.

To change settings at the source or table level, choose the **Source Settings** tab when you edit the plan.

Connection Properties

The following table describes connection options that you can configure at the data source level and the table level:

Connection Options	Description
Source Connection	A connection to the source database. Select a source connection from the list.
Source Connection Directory Path	The path to the flat file source directory. Appears if flat file sources are present.

Connection Options	Description
Source Filename	The name of the source file. Required if the source is a flat file. Default is <name of selected source>.dat.
Target Connection	A connection to the target database or the target file. Choose variable, relational, or flat file from the list. Select a target connection from the list. If you enter the target connection as a variable, ensure that you define the scope of the variable as **Global** or **Integration Service** in the parameter file. In a multinode setup, if you use parameters to enter the connection information, the parameter file must be present on the node on which the Test Data Manager Service runs. The file must be in the same directory path on both nodes.
Target JDBC Connection	Appears if you select an ODBC target connection type. The connection that contains the JDBC connection string of the ODBC target database. Required if the target connection type is ODBC and either of the following statements is true: - You want to disable and enable constraints in the plan. - The source connection type is ODBC and you want to generate mappings with the source data type. Optional if the target is ODBC and you want to perform simple data movement operations.
Output Filename	The name of the target file. Required if the target is a flat file. Default is <name of selected source>.out.
File Encoding	The file encoding type. Required if the target is a flat file. If the source is a flat file, the default type is what the flat file contains. If the source is relational, default is MS Windows Latin 1 (ANSI), superset of Latin1. You can select the type of flat file encoding that you want in the target.
File Format	The format of a flat file. Required if the target is a flat file. You can choose one of the following options: - Fixed Width. The width of the columns is fixed. You cannot specify a column delimiter. - Delimited. You can limit the width of the columns. Specify a column delimiter. Default is Delimited.
Column Delimiter	A character that separates columns of data. Required if the source is a flat file and if the file format is delimited. Default is a comma (,).
Row Separator	A character that separates rows of data. Required if the source is a flat file. Default is \012 LF (\n).

Connection Options	Description

Optional Quotes	Select No Quotes, Single Quote, or Double Quotes. If you select a quote character, the Integration Service ignores delimiter characters within the quote characters. Therefore, the Integration Service uses quote characters to escape the delimiter.
	For example, a source file uses a comma as a delimiter and contains the following row:
	`342-3849, ‘Smith, Jenna’, ‘Rockville, MD’, 6.`
	If you select the optional single quote character, the Integration Service ignores the commas within the quotes and reads the row as four fields.
	If you do not select the optional single quote, the Integration Service reads six separate fields.
	When the Integration Service reads two optional quote characters within a quoted string, it treats them as one quote character. For example, the Integration Service reads the following quoted string as
	`I’m going tomorrow:`
	`2353, ‘I’’m going tomorrow’, MD`
	Additionally, if you select an optional quote character, the Integration Service reads a string as a quoted string if the quote character is the first character of the field.
	Note: You can improve session performance if the source file does not contain quotes or escape characters.

Include Headers	Optional. You can choose to include the headers in the target flat file. Default is Yes.
Source File Name	The name of the IMS source file or VSAM source file. If the source is a Hierarchical IMS file, enter the name of the source unload file. If the source is a Hierarchical VSAM/SEQ file, enter the name of the source file. Required if the source connection is PXW_NRDB Batch.
Target File Name	The name of the IMS target file or VSAM target file. If the target is a Hierarchical IMS file, enter the name of the target unload file. If the target is a Hierarchical VSAM/SEQ file, enter the name of the target file. The target file name must be different from the source file name. Required if the target connection is PXW_NRDB Batch.
Teradata ODBC Connection	Name of the Teradata ODBC connection. Required if the target connection is Teradata PT. Select a connection from the list.
Lookup Connection	Connection to a database that contains lookup tables. Required if the plan has a mapplet rule which includes a transformation. Choose relational or flat file from the list. Select a lookup connection from the list.
Dictionary Connection	A connection to the database that contains the dictionary table. The dictionary connection option does not appear unless the project contains a rule that requires a relational dictionary.
Email Dictionary Connection	A connection to the database that contains the email dictionary table. The email dictionary connection option does not appear unless the project contains a masking rule that requires a relational email dictionary.

Test Tool Integration Properties

The following table describes the test tool integration properties that you can configure at the data source level:

Test Tool Integration Options	Description
Attach to Test Tool	Select this option to copy the flat file target results to a location in the HP ALM server integrated with TDM. You must run the workflow from Test Data Manager. The test tool integration job does not run when you run the workflow from the PowerCenter client.
Project	Required. Browse to select the project in the HP ALM server where you want to store the flat file results.

Target Properties

The following table describes target options that you can configure at the data source level and at the source level:

Other Property Options	Description
Truncate Table	Truncates the table before loading it. Default is disabled.

Update Strategy Properties

The following table describes strategy options that you can configure at the data source level:

Update Strategy Option	Description
Treat source rows as	Indicates how the PowerCenter Integration Service treats source rows. Choose Insert or Update. If an imported PowerCenter mapplet contains a transformation that is configured to set the update strategy, the default option is Insert.
Update as	Determines whether the PowerCenter Integration Service updates or inserts rows. Choose one of the following options: - Update. The PowerCenter Integration Service updates all rows flagged for update. - Insert. The PowerCenter Integration Service inserts all rows flagged for update. - Update else Insert. The PowerCenter Integration Service updates rows flagged for update if they exist in the target, and inserts remaining rows marked for insert.

Error and Recovery Properties

The following table describes connection options that you can configure at the data source level:

Error and Recovery Options	Description
Commit Type	Choose a source-based or target-based commit. By default, the PowerCenter Data Integration Service performs a target-based commit.
Commit Interval	The number of rows to process before a commit. Default is 10,000.

Source and Target Properties

You can override the name of the table owner if you want to select a table from a different schema.

The following table describes source and target options that you can configure at the data source level:

Source and Target Options	Description
Imported Source Owner Name	The default name of the source owner.
Runtime Source Owner Name	The name of the table owner from which you want the source data. Specify the owner name of the source table if you need the table from a different schema. Enter the schema name in uppercase unless the schema name is case sensitive in the database. If the schema name is case sensitive in the database, enter the name as in the database.
Target Prefix Name	The name prefix of the target table. Specify the target schema name if the target is in a different schema.

Advanced Properties

The following table describes advanced options that you can override at the data source level and at the table level:

Advanced Options	Description
Datetime Format String	Date-time format defined in the session properties. You can enter seconds, milliseconds, microseconds, or nanoseconds. - Seconds. MM/DD/YYYY HH24:MI:SS - Milliseconds. MM/DD/YYYY HH24:MI:SS.MS - Microseconds. MM/DD/YYYY HH24:MI:SS.US - Nanoseconds. MM/DD/YYYY HH24:MI:SS.NS Default is microseconds.
Target Load Type	Choose Normal or Bulk. If you select Normal, the PowerCenter Integration Service loads targets normally. Choose Normal mode if the mapping contains an Update Strategy transformation. Choose Bulk to improve session performance. Bulk load limits the ability to recover because no database logging occurs. Default is Normal.

Advanced Options	Description
Target Schema Name	Table name prefix. Specify the target schema name if the target is in a different schema. If the target is Cassandra, you must specify the target schema name.
Source Schema Name	Table owner name. Specify the source schema name if the source is in a different schema. If the source is Cassandra, you must specify the source schema name.
Source Pre SQL	SQL query executed before the plan is run. Required if you need to perform an operation before you run the plan. Enter the complete SQL statement.
Source Post SQL	SQL query executed after the plan is run. Required if you need to perform an operation after you run the plan. Enter the complete SQL statement.
Batch Size	The number of target rows that you want to move at a time. This option is available when you select the Batch Update as Yes. Use with inplace masking. **Note:** The batch size must be greater than the total number of records.
Use Oracle ROWID in Source	Determines whether to use the ROWID from Oracle sources as a column. The ROWID indicates the physical location of a row in the database. Enabled when the Oracle source and target connections are the same.
Performance Hint	Optimizes the SQL query after the plan is run. Enter a performance hint value only for Oracle databases.

.

Partitioning Properties

You must specify partitioning details if the tables in the source database contain partitions.

The following table describes partitioning options that you can configure at the table level:

Partitioning Property	Description
Enable Partitioning	Select to enable partitioning of output tables.
Partition Type	Select the type of partitioning: - Pass Through. Use pass through partitioning to increase data throughput without increasing the number of partitions. - Key Range. Use key range when the source data tables are partitioned by key range. - Database Partitioning. Use database partitioning for Oracle or IBM DB2 sources and targets.
Number of partitions	The number of partitions in the source data. Enter the number of partitions and click **Select** to enter partition information for the target data.

XSD Data Source Connection Settings

You can override the plan settings and enter different source and target file directories in a plan that contains XSD data sources. The settings apply to all files in the directory. You can override the data source connection settings for individual files in the directory.

The following table describes the connection options that you can configure at the data source level and at the file level:

Connection Options	Description
Source File Directory	The path to the XML source file directory. The source file must reside on the server on which the PowerCenter Integration Service runs.
Target File Directory	The path to the XML target file directory.
Source Type	Indicates whether the source file contains the source data or whether the source file contains a list of files with the same file properties. Use to configure multiple file sources with a file list. - Direct. The source file contains the source data. - Indirect. The source file is an index file. The index file contains a list of file names. The PowerCenter Integration Service finds the index file and reads each listed file when it runs the session. The index file must have an empty line at the end of the file. Default is Direct.
Source File Name	The file name or file name and path. Optionally, use the $InputFileName session parameter for the file name. If you specify both the directory and file name in the **Source File Directory** field, clear this field. The PowerCenter Integration Service concatenates this field with the **Source File Directory** field when it runs the session. For example, if you have "C:\XMLdata\" in the **Source File Directory** field, then enter "filename.xml" in the Source File Name field. When the PowerCenter Integration Service begins the session, it looks for "C:\XMLdata\filename.xml".
Output File Name	The file name, or file name and path. Optionally, use the $OutputFileName session parameter for the file name. If you specify both the directory and file name in the **Output File Directory** field, clear this field. The PowerCenter Integration Service concatenates this field with the **Output File Directory** field when it runs the session.

XSD Data Source Advanced Settings

You can override the plan settings for XSD sources in a plan that contains XSD data sources. The settings apply to all files in the directory. You can override the data source connection settings for individual files in the directory.

The following table describes the advanced options that you can configure at the file level:

Advanced Options	Description
XML date-time format	Date-time format defined in the session properties. Choose one of the following values: - Local Time. The time according to the PowerCenter Integration Service server time zone. - Local Time with Time Zone. The difference in hours between the PowerCenter Integration Service time zone and Greenwich Mean Time. - UTC. Greenwich Mean Time. Default is Local Time.
Treat empty content as NULL	Treat empty XML components as null. By default, the PowerCenter Integration Service does not output element tags for null values. The PowerCenter Integration Service outputs tags for empty content. Default is No.
Format Output	Format the XML target file so the XML elements and attributes indent. If you do not select Format Output, each line of the XML file starts in the same position. Default is No.
NULL content representation	Choose how to represent null content in the target. Default is No tag.
Empty String content representation	Choose how to represent empty string content in the target. Default is Tag with empty content.
NULL attribute representation	Choose how to represent null attributes. Default is No attribute.
Empty String attribute representation	Choose how to represent empty string attributes in the target. Default is Attribute with empty string.

Plan Management

After you create a plan you can edit it. You can copy the plan, export it to an XML file, or delete the plan. You can import a plan that you created in another TDM repository and exported.

Creating a Data Masking and Data Subset Plan

When you create a plan, add components to it to define its operations. You can combine a data subset and a data masking operation in the same plan, or create separate plans. To perform a data generation operation, you must create a separate plan. Add entities, groups, and templates to complete data subset operations. Add rules and policies to plans to perform data masking operations.

1. Open a project and click **Execute** to view the project plans.
2. Click **Actions** > **New**.

3. In the **New Plan** dialog box, enter a name and optional description for the plan.
4. To add a data masking operation to the plan, click **Add Masking Components**.
5. Select the policies and rules to add to the plan. Click **Next**.
6. To add a data subset operation to the plan, click **Add Subset Components**.
7. Select the entities, groups, and templates, or XSD files to add to the plan. Click **Next**.

   You can assign criteria for masking columns and generating subsets.
8. To skip adding a data generation component, click **Next**.
9. To skip masking a rule, select the check box for the rule and click **Off**.
10. To filter subset components, select the component and choose to enter a basic or advanced expression.

    Test Data Manager shows any existing filter criteria. You can change the criteria. You cannot override subset criteria for an XSD subset component.
11. To limit the subset results, click **Limit** and choose to limit by percentage, by absolute value, or by defining an interval of rows to create. Click **Next**.
12. Configure the connections and other properties. If the target is a flat file, you can configure test tool integration properties. If you select the test data warehouse as the target connection, you can enter the data set properties. If the source is an XSD source, you cannot add other types of sources to the plan. Click **Next**.
13. To override plan settings for a data source, select the data source and click **Override Plan Settings** and enter the properties.
14. To override data source settings for a table, select the table and click **Override Data Source Settings** and enter the properties.
15. Click **Finish**.

    The plan appears in the project.

Related Topics:

- "Generating a Workflow" on page 242
- "Data Masking Plans and Workflows" on page 128

Creating a Data Generation Plan

To perform a data generation operation, you create a plan and add data generation components to the plan. When you add data generation components, you cannot add data subset or data masking components to the same plan.

1. Open a project and click **Execute**.
2. Click **New Plan**.
3. In the **New Plan** dialog box, enter a name and optional description for the plan.
4. Click **Next**.
5. To skip adding the masking components, click **Next**.
6. To skip adding the subset components, click **Next**.
7. To add the generation components to the plan, click **Add Generation Components**.
8. Select the entities and the tables to add to the plan. If you want to generate test data in an XML file, select the XSD file to add to the plan.
9. Click **OK**.

The entities and tables appear in the list.

10. Click **Next**.
11. To view or filter the table names, select the table.
12. To specify generation criteria for an entity, select the entity and enter the values.
13. Click **Next**.
14. To configure generation plan settings, enter the number of records that you want to generate.
15. Configure connection and target properties. If the target is a flat file, you can configure test tool integration properties. If you select the test data warehouse as the target connection, you can enter the data set properties.
16. To make changes to data generation, error recovery, and advanced settings, you can edit the settings.
17. Click **Next**.
18. Click **Next**.
19. To override plan settings, click **Override Plan Settings** and enter the properties.
20. To override table settings, click **Override Data Source Settings** and enter the properties.
21. Click **Finish**.

    The generation plan appears in the project.

Creating an Ad Hoc Generation Plan

You can create an ad hoc generation plan for a column.

1. Open a project and click **Define** > **Data Generation**.
2. Select a column from the table for which you want to create an ad hoc generation plan.
3. Click **Actions** > **Generate and Execute**.

   The **Generate and Execute Plan** dialog box appears.
4. Enter a name for the plan or keep the default plan name that appears on the screen.
5. Select the Integration Service.
6. Enter the number of records that you want to generate.
7. Configure connections and target properties.
8. To make changes to data generation, error and recovery, and advanced settings, edit the settings.
9. Click **OK**.

   A **Create and Execute** dialog box appears.
10. To run the plan, click **Yes**.

Copying a Plan

Copy a plan to create another plan with similar components. Create a copy of a plan and edit the copy.

1. Open a project and click **Execute** to view the project plans.
2. Click the plan **Description** or **Status** field to select a plan.

   Do not open the plan.
3. Click **Actions** > **Duplicate**.

4. Optionally, enter a project name and a description.

   The default name is `Copy of <original name>`.

Exporting a Plan

You can export a group to an XML file and import the XML file to another TDM repository.

1. Open a project and click **Execute** to view the project plans.
2. Click the plan **Description** or **Status** field to select a plan.

   Do not open the plan.
3. Click **Actions** > **Export**.
4. Choose to save the file.
5. Enter the XML file name and the path of the file.

   The default name is a string that contains "Plan_" and the current date and the time.

Importing a Plan

You can import a plan from an XML file that was exported from another TDM repository.

1. To open the **Projects** view, click **Projects**.
2. Click the project description to select a project to import the plan into.

   Do not open the project.
3. Click **Actions** > **Import**.
4. Browse for the XML file that contains the group to import.

   The XML file has a default name similar to `Plan_130315081854.xml`.
5. Click **Finish** to import the plan.

Deleting a Plan

You can delete plans. When you delete the plan, you delete the workflow for the plan.

1. Open a project and click **Execute** to view the project plans.
2. Click the plan **Description** or **Status** field to select a plan.

   Do not open the plan.
3. Click **Actions** > **Delete**.
4. Click **Yes** to confirm the delete.

Workflow Generation

Generate a workflow from a plan, and then run the workflow to perform data subset, data masking, and data generation operations.

You can generate and start a workflow in a single step.

When you generate a workflow, the PowerCenter Repository Service generates mappings for the workflow. When you start a workflow, the PowerCenter Integration Service performs the data masking, data subset, and data generation tasks defined in the plan. You can select the PowerCenter Integration Service to run the workflow.

You can view the status of generated workflows or the load configuration on the **Monitor** view.

Generating a Workflow

After you create a plan, generate a workflow. The PowerCenter Repository Service generates the mappings to include in the workflow.

1. In a project click **Execute** to access the plans in the project.
2. Select a plan from the list.
3. Click **Actions** > **Generate Workflow**.
4. Choose whether to run it immediately or to generate the workflow later.

Related Topics:

- "Creating a Data Masking and Data Subset Plan" on page 238
- "Data Masking Plans and Workflows" on page 128

Export to HP ALM

You can export flat file results of a plan in TDM to an HP ALM server. You must enable and configure test tool integration before you can export the results.

Enter the test tool integration settings when you define the plan settings. TDM runs a separate job in the same workflow to copy the flat file to the HP ALM server in addition to the TDM server. The test tool integration job creates a test resource folder in the HP ALM server and appends the plan name and the job ID to the folder name.

You must run the workflow from Test Data Manager. The test tool integration job does not run if you run the workflow from the PowerCenter client.

Note: You cannot export a flat file from TDM if you enable version control on the HP ALM server. You must disable version control and then run the TDM workflow.

Workflow View

The Workflow view contains a list of the workflows that you generate from a plan. You can view details about the workflows and sessions that you generated from the plan.

The Workflow Details panel contains the list of workflows in the plan and the properties panel contains the workflow properties and session names.

The **Workflow** view has a **Workflow Details** panel and a **Properties** panel. The **Workflow Details** panel contains the list of workflows in the plan. The Properties panel contains properties for a workflow and the session names in the workflow.

Properties Panel

The **Properties** panel in the **Workflow** view contains a list of the workflows generated from a plan.

You can view the following information about generated workflows:

Property	Description
Workflow Name	A string that includes the project name, plan name, connection, and table name.
User	The name of the user that generated the workflow.
Generation	Date and time that the workflow was generated.

Details Panel

The **Details** panel shows the **Workflow** details and the **Session** details for a workflow that you select in the **Workflows** view.

The **Workflow** details shows the summary information about a workflow. When you select a workflow in the **Workflow** view, the workflow details appear. You can view the workflow name, the user, and the project name.

The **Session** details panel appears next to the **Workflow** details panel. The panel shows each session name and the number of tables that the session processes for a workflow.

Executing a Workflow

After you generate a workflow, you can execute the workflow to run sessions from the PowerCenter mappings. If you generated multiple workflows for the same plan, you can execute each workflow separately.

You can start a workflow from the **Plan | Properties** view or the **Plan | Workflow** view.

1. If the plan you selected contains one workflow, click **Actions** > **Execute Workflow** in either view.
2. If the plan you selected contains multiple workflows, click the **Workflow** view.
   a. Select the workflow that you want to start.
   b. Click **Actions** > **Execute Workflow**.
3. View workflow status in the **Workflow Execution** view.

Note: If you run the workflow with the **Disable Constraints** option selected, and TDM fails to enable the constraints after loading the data, the workflow fails. You must then manually enable the constraints in the database and run the workflow again.

Workflow Executions View

The **Workflow Executions** view shows the current workflow that runs for the plan. You can also view the status of all the previous workflows. The list can contain workflow runs for workflows that no longer exist in

the plan. The order of the workflow is based on the start date and time. To copy flat files to an integrated test tool, the workflow runs a separate job.

To view the workflow log, you can click **Job ID**. To view the session log, you can click **Session ID** in the **Sessions** tab.

The following table describes the fields in the **Workflow Executions** view for each workflow run:

Field	Description
Job ID	The job number that identifies the job. If you click the Job ID, you can view the workflow log.
Name	The workflow string that includes the project name, plan name, connection, and table name.
Description	Describes the type of job that the workflow performs. The job type can be profiling, import, and workflow operations.
Status	The current status of the workflow. - In Queue. The PowerCenter Integration Service is waiting for resources before it starts the workflow. - Running. The workflow is running. - Success. The workflow finished successfully. - Error. The workflow did not complete due to errors.
Start Date/Time	The date and time that the workflow started.
End Date/Time	The date and time that the workflow ended.
User	The name of the user that started the workflow.

Workflow Tasks

You can stop, abort, and recover the workflow in the Workflow Executions view.

You can perform the following tasks on the Workflow Executions view:

Auto Refresh

Causes the view to refresh every ten seconds automatically. You can turn auto refresh on and off.

Abort

Stops a job immediately without waiting for a commit. You can abort all jobs, such as profiling and import. You cannot abort workflows.

Workflow Stop

Stops a workflow after the next commit occurs.

Workflow Abort

Stops a workflow immediately without waiting for a commit.

Workflow Recover

Recover a stopped workflow. If high availability recovery is enabled, the PowerCenter Integration Service restarts the workflow from the last saved checkpoint. Otherwise, the PowerCenter Integration Service recovers the workflow according to the recovery strategy that you set in the plan.

Workflow Properties Panel

The **Properties** panel shows summary information about a workflow that you select in the **Workflow Executions** view.

The **Properties** panel shows the same information as the workflow you select in the list. The properties also include generation date, the elapsed time, and the project name.

Workflow Sessions Tab

The **Sessions** tab lists the sessions that ran in the workflow that you select in the **Workflow Executions** view.

The following table describes the fields for each session in the **Sessions** tab:

Field	Description
Job ID	The job number that identifies the session. If you click the Job ID, you can view the session log.
Session Name	The session name is the same value as the workflow name, except the session name starts with "S" and ends with a sequential number. If you click the session name, you can view the session details.
Status	The current status of the session. - In Queue. The PowerCenter Integration Service is waiting for resources before it starts the session. - Running. The session is running. - Success. The session finished successfully. - Error. The session did not complete due to errors.
Source Rows Succeeded	The number of valid source rows read.
Source Rows Failed	The number of source rows in error.
Target Rows Succeeded	The number of rows written to the target.
Target Rows Failed	The number of rows that were not written to the target.
Start Date/Time	The date and time that the session started.
End Date/Time	The date and time that the session ended.
User	The name of the user that started the workflow.

Session Details

The **Session Details** dialog box contains a list of tables that a session processed. The **Session Details** dialog box appears when you click the session name in the **Workflow Execution Sessions** tab.

If the **Workflow Executions** view has refresh enabled, the session details refresh when the workflows refresh.

The following table describes the fields in the **Session Details** dialog box:

Field	Description
Tables	The names of the tables that the session processed.
Affected Rows	The number of rows that the session processed from the table.
Rejected Rows	The number of rows that the PowerCenter Integration Service did not process due to errors.
Throughput	The number of affected rows that the PowerCenter Integration Service did not process due to errors.
Last Error Message	The text of the last error message in the session. Click the link to view the complete message.

CHAPTER 13

Monitor

This chapter includes the following topics:

- Monitor Overview, 247
- Jobs, 248
- Monitor Tasks, 249
- Logs, 250
- Sessions, 251

Monitor Overview

In the **Monitor** view, you can monitor the status of jobs that you start in Test Data Manager. You can stop jobs from running and view job and session logs.

You can monitor jobs for all projects in the **Monitor** view. To monitor jobs for a single project, open the project and click **Monitor**.

You can monitor jobs for all data set versions in the **Monitor** view.

To monitor jobs for a specific data set version, open the data set version and click **Monitor**. You can view jobs run on the data set version from any machine that hosts a Test Data Manager Service in the domain.

You can sort, filter, and perform tasks on jobs in the **Monitor** view. Select a job in the **Monitor** view to view the job details in the **Properties** pane. You can also view the workflow and session logs for a job.

The following image shows the **Monitor** view:

Jobs

Check the status of a job and view the job details in the **Monitor** view.

You can view the following types of jobs in the **Monitor** view:

Import from Source

Imports source data from a source file.

Import from PowerCenter

Imports a source from the PowerCenter repository.

Profiling

Performs data discovery for primary keys, entities, and data domains.

Generate Workflow

Generates a workflow from a plan.

Execute Workflow

Runs a workflow for data subset, data masking, or data generation operations after you generate the workflow.

Workflow

Runs the data subset, data masking, or data generation operation. The execute workflow job might run multiple workflows.

Upload to test tool

Uploads all flat files created by a successful workflow to the integrated HP ALM server location.

Reset Data Set Version from Test Data Warehouse

Resets target data with a data set from the test data warehouse.

Session

Performs a task within the workflow. A workflow might have multiple sessions. Select a workflow to view the session details in the **Session** pane.

Job Details

You can sort and filter jobs by job details.

The **Monitor** view contains the following job details:

Job ID

The job ID number. TDM creates consecutive job ID numbers for each job.

Name

The name or the type of job. You can view import, profiling, workflow, and session jobs.

Description

The name of the plan, profile, PowerCenter folder, or connection associated with the job.

Status

The status of the job. A job can have the following statuses:

- Error. The job did not run successfully. Click the job ID to view the job log file.

- In Queue. The job is in the queue to run.
- Running. The job is running.
- Success. The job ran successfully.
- Terminated. The job was terminated.

Project

The name of the project that contains the job. Project details are not visible in the **Monitor** view within a project.

Start Date / Time

The date and time the job started.

End Date / Time

The date and time the job ended.

User

The user that ran the job.

Monitor Tasks

You can perform tasks in the Monitor view based on the job that you view.

You can perform the following tasks in the Monitor view:

Auto Refresh

Refreshes the view every ten seconds. If you disable the auto refresh, click **Actions** > **Refresh** to manually refresh.

Abort

Stops a job immediately. You can abort all jobs except workflows.

Workflow Stop

Stops a workflow after the next commit occurs.

Workflow Abort

Stops a workflow immediately.

Workflow Recover

Recovers an aborted or stopped workflow. If high availability recovery is enabled, the PowerCenter Integration Service restarts the workflow from the last saved checkpoint. Otherwise, the PowerCenter Integration Service clears the state of operation and the recovery information before it restarts the workflow.

Logs

You can view logs to troubleshoot jobs. To view the logs, select a job and click the **Logs** tab. You can view the session, workflow, and console logs.

When a TDM job runs, the TDM server generates logs. The PowerCenter Integration Service generates the session and workflow logs. When a TDM job fails, you can view the logs to debug problems.

When a job is triggered, TDM always generates console logs. If the job logs are not available, you can view the console logs to check the messages.

When you click a job ID, you can view the following log details:

Date/ Time

The date and time the job ended.

Severity

The severity level of the log messages.

Description

The detailed description of the error message, the cause of the problem, and the solution.

Severity Levels

You can view the severity level of a log message and get a clear understanding of the problem level.

The log messages include the following severity levels:

Error

Indicates that the TDM server failed to perform an operation or respond to a request from a client application.

Warning

Indicates that the TDM server is performing an operation that might cause an error.

Info

Indicates that the TDM server is performing an operation that does not cause errors or problems.

Debug

Indicates TDM Server operations at a detailed level. The debug messages generally record the success or failure of server operations.

Trace

Indicates TDM Server operations at a more specific level than the debug logs. Trace messages are generally trace code paths.

Viewing the Log Messages

View the log messages to troubleshoot the problems when a TDM job fails. You can search and filter the log messages based on the dates, severity levels, and keywords in the error description.

1. Click **Monitor**.
2. Click a job ID and click the **Logs** tab.

   A list of log messages appears.
3. Search and filter the logs from the list of log messages.

4. Select the log message and download the log file.
5. To download the log file, click **Download**.
6. To view console logs for each job, select a job ID and click **Actions** > **View Console Logs**.

Sessions

When you select a job ID within the execution workflow, you can view the **Sessions** tab. The **Sessions** tab lists the sessions that ran for the workflow.

The **Sessions** tab contains the following fields:

Job ID

The job number that identifies the session. If you click **Job ID**, you can view the processed tables, affected and rejected rows of the tables, error messages, and session logs.

Session Name

The name of the session that ran in the workflow.

Status

The current status of the session. The session can have the following statuses:

- In Queue. The PowerCenter Integration Service is waiting for resources before it starts the workflow.
- Running. The workflow is running.
- Success. The workflow completed successfully.
- Error. The workflow did not complete due to errors.

Source Rows Succeeded

The number of valid source rows read during the session.

Source Rows Failed

The number of source rows that failed due to errors.

Target Rows Succeeded

The number of rows written to the target.

Target Rows Failed

The number of rows that were not written to the target.

Start Date/Time

The date and time that the session started.

End Date/Time

The date and time that the session ended.

User

The name of the user that started the workflow.

CHAPTER 14

Reports

This chapter includes the following topics:

- Reports Overview, 252
- Data Masking Report, 253
- Plan Audit Report, 253
- Plan Detail Report, 254
- Row Count Report, 255

Reports Overview

You can run a report from Test Data Manager to see detailed information about a plan or a policy. Run the type of report that contains the information you want to see.

Test Data Manager contains data masking, plan audit, plan detail, and row count reports. You can run a data masking report on a project. You can run a plan audit, plan detail, or a row count report on a plan.

When you run a report in Test Data Manager, the report opens in a new window. Make sure you do not have a pop-up blocker enabled so that the report window can open.

The following table describes the reports that you can run in Test Data Manager:

Report	Description
Data Masking Report	A report that you run on a project that lists all the table columns and the rules and policies that are assigned to the columns.
Plan Audit Report	A report that you run on a plan that contains detailed information about the policy assignments in the plan.
Plan Detail Report	A report that you run on a plan that lists policy assignments, entity details, groups, templates, and template components that are part of the plan.
Row Count	A report that you run on a plan that lists the tables in the plan and the number of rows in each table that the plan affects.

Data Masking Report

A data masking report lists all the columns within a project and the rules and policies that you assign to the project.

Running the Data Masking Report

1. Click **Projects** to view a list of the projects in Test Data Manager.
2. Click a project in the list to open it.
3. Click **Define** > **Data Masking** > **Print Masking Report** to view the project plans.

   The data masking report opens in a new window.

Plan Audit Report

Run a plan audit report to view detailed information about the data masking components that took part in a plan execution.

You can generate a plan audit report to present to auditors or administrators that need comprehensive information about the policies and rules that took part in a plan execution.

You can create a plan audit report for a plan that has successfully completed at least one execution.

The following table lists the properties in a plan audit report:

Component	Properties
Policy Assignment	You can view the following policy assignment details: - Data Source - Table - Column - Data Type - Data Domain - Policy - Rule - Rule Type - Additional Details - Masking Properties - Exception Handling
Plan Execution	You can view the following plan execution details: - Source Connection - Target Connection - Created Date - Plan Start Date - Plan End Date - Total Rows Processed - Status of the Plan

Running a Plan Audit Report

1. Click **Projects** to view a list of the projects in Test Data Manager.

2. Click a project in the list to open it.
3. Click **Execute** to view the project jobs.
4. Click a plan.
5. Click **Actions** > **Plan Audit Report**.

   The row count report opens in a new window.

Plan Detail Report

A plan detail report lists the data subset and data masking components within a plan.

The following table lists the properties in a plan detail report:

Component	Detail
Masking Assignment	You can view the following policy assignment details: - Data Source - Table - Column - Data Type - Data Domain - Policy - Rule - Masking Properties
Entity	You can view the following entity details: - Name - Description - Filter Criteria
Group	You can view the following group details: - Name - Description
Template	You can view the following template details: - Name - Description - Filter Criteria
Template Component	You can view the following template component details: - Name - Type - Description - Component Name

Running the Plan Detail Report

1. To view a list of the projects, click **Projects**.
2. Click a project in the list to open it.
3. To view the project plans, click **Execute**.
4. Click a plan.

5. Click **Actions** > **Plan Detail Report**.

   The plan detail report opens in a new window.

Row Count Report

A row count report contains a list of the tables that took part in a plan execution and the number of rows that the plan affected.

Running the Row Count Report

1. To view a list of the projects, click **Projects**.
2. Click a project in the list to open it.
3. To view the project jobs, click **Monitor**.
4. Click a job that has an Execute Workflow job type.
5. Click **Actions** > **Row Count Report**.

   The row count report opens in a new window.

CHAPTER 15

ilmcmd

This chapter includes the following topics:

- ilmcmd Overview, 256
- Configuring ilmcmd, 257
- Running ilmcmd, 257
- Entering Options and Arguments, 257
- Syntax Notation, 258
- Delete, 258
- Export, 260
- Import, 261
- Search, 262
- Workflow, 264
- Reset, 265
- Register, 267

ilmcmd Overview

ilmcmd is a command line program that you use to perform TDM tasks. You can use ilmcmd to complete a subset of the tasks that you can perform with Test Data Manager.

Use ilmcmd to perform the following tasks:

- Import XML files into the TDM repository.
- Export data subset, data masking, and data generation objects to XML files.
- Search for data subset and data masking objects.
- Validate policies, plans, and data subset objects.
- Delete data subset, data masking, and data generation objects.
- Generate and run workflows.
- Display the status of workflows.

You must have the required privileges to perform these tasks. For tasks related to projects, you must also have the required permission on the project.

You cannot use ilmcmd to search for data generation objects.

Configuring ilmcmd

When Test Data Manager runs on HTTPS, you must configure the ilmcmd before you can run ilmcmd commands.

Before you configure ilmcmd, you must import the browser certificate to the following location:

```
<Install directory>/TDM/utilities/ilmcli/conf
```

To generate a keystore, you must import the browser certificate. The keystore is necessary to run ilmcmd commands.

1. To set the password and generate a keystore, run the following command:

   ```
   keytool -import -file <Imported certificate> -keystore client.ks
   ```

2. Edit the userConfig.ilm file and add the keystore location for the javax.net.ssl.trustStore parameter. You can find the userConfig.ilm file in the following location:

   ```
   <Installation directory>/TDM/utilities/ilmcli/conf/userConfig.ilm
   ```

3. Add the password that you used to create the keystore for the javax.net.ssl.trustStorePassword parameter.

Running ilmcmd

Invoke ilmcmd from the command line. You can issue commands directly or from a script, batch file, or other program. On Windows, ilmcmd is a batch file with a .bat extension. On UNIX, ilmcmd is a shell script with no extension.

1. At the command line, switch to the directory where the ilmcmd executable is located.

   By default, ilmcmd is installed in the following directory:

   ```
   <installation directory>\utilities\ilmcli\bin
   ```

2. Enter `ilmcmd` followed by the command name and its required options and arguments.

   For example:

   ```
   ilmcmd -command_name [-option1] argument_1 [-option2] argument_2...
   ```

To view the command line syntax, enter the following command:

```
ilmcmd -h
```

Entering Options and Arguments

The ilmcmd command line program uses a set of options and arguments.

Use the following rules when you enter command options and arguments:

- To enter options, type a hyphen followed by the program syntax for the command.
- Enter options in any order.
- If an argument contains spaces, enclose the argument in double quotes.
- The first word after the option is the argument.
- Most options require arguments. You must separate options from arguments with a single space.

- Commands, options, and object names are case sensitive.

Syntax Notation

To use the ilmcmd command line program, review the syntax notation.

The following table describes the syntax notation for the ilmcmd command line program:

Convention	Description
-x	Option placed before an argument. This designates the parameter you enter. For example, to enter the user name for ilmcmd, type -un or -UserName followed by the user name.
<x>	Required option. If you omit a required option, the command line program returns an error message.
<x | y>	Select between required options. If you omit a required option, the command line program returns an error message. If a pipe symbol (|) separates options, specify exactly one option. If options are not separated by pipe symbols, specify all the options.
[x]	Optional option. The command runs whether or not you enter these options.
[x | y]	Select between optional options. For example, you can display help for all ilmcmd commands by using the -h or -help option. `[-h|-help]` The command runs whether or not you enter the optional parameter.
< <x | y> | <a | b> **>**	When a set contains subsets, the superset is indicated with bold brackets (**< >**). Subsets are separated with the bold pipe symbol (|).

Delete

Deletes objects from the ILM repository. To delete an object, specify the name of the object or the name of an XML file that lists the object details.

When you delete an object by name, you specify the object type and the location of the object in the repository. You can delete the following object types:

- Data domain
- Masking rule
- Generation rule
- Policy
- Project
- Plan

The ilmcmd delete command uses the following syntax:

```
ilmcmd
<-Delete | -d>
<<-File | -f> XML_file_location | <-Object | -o>>
<-ObjectType | -ot> object_type
<-Project> project_name
<-Name | -n> object_name
[-Help | -h]
```

The following table describes ilmcmd delete options and arguments:

Option	Argument	Description
-ObjectType -ot	object_type	Required to delete an object by name. The type of object to export. For example, you can enter "Group."
-Project	project_name	Required to delete an object by name. The name of the project that contains the object to delete.
-Name -n	object_name	Required to delete an object by name. The name of the object that you want to delete.
-Help -h	n/a	Optional. Displays help for the command.

Delete Examples

Deleting a Masking Rule

The following sample command deletes the `Credit_Card_Mask` rule on UNIX:

```
./ilmcmd -d -ot MaskingRule -n Credit_Card_Mask
```

Deleting a Policy

The following sample command deletes the `Personal` policy on UNIX:

```
./ilmcmd -d -ot Policy -n Personal
```

Deleting a Generation Rule

The following sample command deletes the `SIN_RandomNumeric` rule on Windows:

```
ilmcmd -d -ot GenerationRule -n SIN_RandomNumeric
```

Deleting a Project

The following sample command deletes the `Customer_DataGeneration` project on Windows:

```
ilmcmd -d -ot Project -n Customer_DataGeneration
```

Export

Exports objects from the ILM Test Data Manager to an XML file.

When you export an object from the ILM repository, you specify the object type and the location of the object in the repository. You can export the following object types:

- Masking rule
- Generation rule
- Policy
- Data domain
- Entity
- Group
- Template
- Profile
- Project
- Data source
- Source definition
- Connections
- Plan

The ilmcmd export command uses the following syntax:

```
ilmcmd
<-Export | -e>
<-ObjectType | -ot> object_type
<-Name | -n> object_name
<-Project> project_name
<-File | -f> XML_file_location
<-DataSourceName | -dsn> schema_name
[-Help | -h]
```

The following table describes the ilmcmd export options and arguments:

Option	Argument	Description
-ObjectType -ot	object_type	Required. The type of object to export. For example, you can enter "Group" or "RuleAssignment." If the object type contains a space in it, remove the space when you enter the argument.
-Name -n	object_name	Required. The name of the object you want to export.
-Project -p	project_name	Required to export a plan, entity, group, or template. The name of the project that contains the plan or port assignment to export.
-File -f	XML_file_location	Required. The path and file name of the XML file to which you export the object.

Option	Argument	Description
-DataSourceName -dsn	schema_name	Required to export a port assignment. The name of the schema that contains the port assignment to export.
-Help -h	n/a	Optional. Displays help for the command.

Export Examples

Exporting a Policy

The following sample command exports the `Policy_Customer` policy to the `Policy_Customer.xml` file:

```
ilmcmd -e -ot Policy -n Policy_Customer -f C:\Informatica\ILMServer\CLI
\Policy_Customer.xml
```

Exporting a Masking Rule

The following sample command exports the Credit_Card_Mask rule to the `CCR.xml` file:

```
ilmcmd -e -ot MaskingRule -n Credit_Card_Mask -f /home/infa1/Desktop/CCR.xml
```

Exporting a Generation Rule

The following sample command exports the SIN_RandomNumeric rule to the `SINRN.xml` file:

```
ilmcmd -e -ot GenerationRule -n SIN_RandomNumeric -f E:\rules\SINRN.xml
```

Exporting a Project

The following sample command exports the CustRecord_Data project to the `CustRecord.xml` file:

```
ilmcmd -e -ot Project -n CustRecord_Data -f E:\project\CustRecord.xml
```

Import

Imports objects from an XML file to the ILM Test Data Manager.

You can import the following object types:

- Masking rule
- Generation rule
- Policy
- Data domain
- Entity
- Group
- Template
- Profile
- Project
- Data source
- Source definition

- Plan

The ilmcmd import command uses the following syntax:

```
ilmcmd
<-Import | -i>
<-File | -f> XML_file_location
[-Help | -h]
```

The following table describes the ilmcmd import options and arguments:

Option	Argument	Description
-File -f	XML_file_location	Required. The path and file name of the import file.
-Help -h	n/a	Optional. Displays help for the command.

Import Examples

Importing a Plan

The following sample command imports the plan listed in the `Plan_Customer.xml` file to the repository:

```
ilmcmd -i -f C:\Informatica\ILMServer\CLI\Plan_Customer.xml -un Administrator -pd
Administrator -hn ilmserver -port 6002
```

Importing a Source Definition

The following sample command imports the source definition listed in the `SrcDef_CustomerDB.xml` file to the repository:

```
ilmcmd -i -f C:\Informatica\ILMServer\CLI\SrcDef_CustomerDB.xml -un Administrator -pd
Administrator -hn ilmserver -port 6002
```

Importing a Generation Rule

The following sample command imports the generation rule listed in the `SIN_RandomNumeric.xml` file to the repository:

```
ilmcmd -i -f E:\Rules\SIN_RandomNumeric.xml
```

Search

Searches for source definitions, masking rules, policies, and plans.

You can search for the following object types in the repository:

- Masking rule
- Policy
- Data domain
- Entity
- Group
- Template

- Profile
- Project
- Data source
- Source definition
- Connection
- Plan

You cannot use ilmcmd to search data generation objects.

The ilmcmd search command uses the following syntax:

```
ilmcmd
<-Search | -s>
<-ObjectType | -ot> object_type
<-Project | -p> project_name
<-DataSourceName | -dsn> schema_name
<-NamePattern | -np> name_pattern
[-Help | -h]
```

The following table describes the ilmcmd search options and arguments:

Option	Argument	Description
-ObjectType -ot	object_type	Required. The type of object to export. For example, you can enter "Group" or "Project." If the object type contains a space in it, remove the space when you enter the argument.
-Project -p	folder_name	Required to search for a source definition, plan, entity, group, or template. The name of the project that contains the object to search for.
-DataSourceName -dsn	schema_name	Required to search for a source definition. The name of the schema that contains the source definition.
-NamePattern -np	name_pattern	Required. The name pattern. Use the asterisk (*) character as a wild card. For example, the name pattern `Rule_C*` returns the following rules: - `Rule_Customer` - `Rule_CustID`
-Help -h	n/a	Optional. Displays help for the command.

Search Examples

Searching for a Masking Rule

The following sample command searches for rules that match the name pattern `Rule*`:

```
ilmcmd -s -ot Rule -np Rule.*
```

The sample command returns the following output:

```
Rule_Customer
Rule_Ticket
```

Searching for a Policy

The following sample command searches for policies that match the name pattern `CUST*`:

```
ilmcmd -s -ot Policy -np CUST.*
```

The sample command returns the following output:

```
CUSTOMER
CUST_COUPONS
```

Workflow

Generates a workflow, displays the status of a workflow run, or runs a workflow.

The ilmcmd workflow command uses the following syntax:

```
ilmcmd
<-WorkFlow | -wf>
<<-Generate | -g> |  <-Execute | -ex>  | <-GetStatus | -gs>>
<-Project> folder_name
<-PlanName | -pn> plan_name
<-WorkFlowName | -wfn> Workflow_name
<-IntegrationService | -is> integration_service_name
[-Help | -h]
```

The following table describes the ilmcmd workflow options and arguments:

Option	Argument	Description
-Generate -g	n/a	Generates a workflow.
-Execute -ex	n/a	Runs the workflow.
-GetStatus -gs	n/a	Gets the status of a workflow.
-GetWorkFlows -gwf	n/a	Gets workflows for a plan.
-Project	project_name	Required. The name of the folder that contains the plan.
-PlanName -pn	plan_name	Required. The plan name.
-WorkFlowName -wfn	workflow_name	Required to generate a workflow. The location of the workflow properties file.
-IntegrationService -is	integration_service_name	Required to run a workflow. The name of the PowerCenter Integration Service. Required when you run the workflow.
-JobId	job_id	Required. The job ID of the workflow run.
-Help -h	n/a	Optional. Displays help for the command.

Workflow Examples

Generating a Workflow

The following sample command generates a workflow for the `Plan_NameMasking` plan:

```
ilmcmd -wf -g -Project Infa_Project -pn Plan_NameMasking
```

Monitoring a Workflow

The following sample command monitors the status of a workflow for a plan:

```
ilmcmd -wf -gs -JobId 360
```

Running a Workflow

The following sample command runs a workflow for the `Plan_NameMasking` plan:

```
ilmcmd -wf -ex -Project TDM -pn Plan_NameMasking -is PCInteg
```

Reset

You can reset a data set version to a target and override the source properties.

The reset command uses the following syntax:

```
ilmcmd
[options]
<-Reset | -r>
<-DatasetVersionName | -dvn> Data set version name
<-ResetFile | -rf> File location
<-ResetIntegrationService | -ris> Integration Service
[-Verbose | -v]
[-Help | -h]
```

The following table describes the reset command options and arguments:

Option	Argument	Description
-Reset -r	n/a	Required. The reset job that you create to reset a data set version.
-DatasetVersionName -dvn	Data set version name	Required. The name of the data set version that you want to reset.
-ResetFile -rf	File location	Required. The location of the file that contains the reset objects.
-ResetIntegrationService -ris	Integration Service	Required. The name of the Integration Service that runs the reset workflow operation.
-Verbose -v	n/a	Optional. Verbose output. Displays more than the minimum information about the object.
-Help -h	n/a	Optional. Displays or prints the help options.

Reset Example

The following sample command resets a data set version:

```
ilmcmd -r -dvn "RESET Version1" -rf "C:\Documents\ResetObject.json" -ris PIS
```

Sample Reset File Properties

Use a sample reset file to configure the reset properties.

After you install TDM, you can find the `ResetObject.xml` file or the `ResetObject.json` file in the following location:

```
<Installation directory>/TDM/utilities/ilmcli/conf/
```

The following table describes the reset file properties and the description:

Property	Description
SOURCE_CONN_NAME	The name of the source connection.
TRUNCATE_TRGT_TABLES	The target tables that TDM truncates if you want to completely replace the target connection with data from the data set version. TDM deletes all data in the target, and copies the data set version to the target. TDM disables constraints and indexes when it performs this operation.
AUTOMATIC_DISABLE_CONSTRAINT	Disable the constraints automatically during the reset operation.
MANUAL_DISABLE_CONSTRAINT	Disable the constraints manually during the reset operation.
DISABLE_INDEXES	Disable indexes during the reset operation.
UPSERT_IN_TRANSACTIONAL_TABLES	Performs an upsert on transactional tables instead of an insert. If the parent records of records in the data set version have changed to point to a different parent in the target, the insert might fail. The reset operation fails if the insert fails.
ENABLE_RECOVERY	Enables the reset recovery options.
BACKUP_RECOVERY_PREFIX	A prefix for the backup schema. TDM creates a backup schema in the format prefix_tablename.
BACKUP_RECOVERY_USER_NAME	The name of the schema where you want to store the backup tables.
STOP_ON_ERROR	Indicates how many non fatal errors the PowerCenter Integration Service encounters before it stops the session. If you enter zero, the session does not stop for nonfatal errors. Default is zero.
ENABLE_HIGH_PRECISION	Processes the decimal data type to a precision of 28. Default is No.
MAX_TABLES_PER_MAPPING	Sets the number of tables to process in a workflow. Default is 10.
MAX_PARALLEL_SESSIONS	The maximum number of sessions that can run at the same time. Applicable if you disable indexes and constraints in the target.
LOCALE_CODE	Sets the locale code for data reset operations.

Property	Description
EMAIL_ADDRESS	An email address to send notifications to.
EMAIL_SUBJECT	The subject of the email that you enter.
EMAIL_TEXT	The content of the email that you enter.
TARGET_LOAD_TYPE	The target load type. You can choose Normal or Bulk. If you select Normal, the PowerCenter Integration Service loads targets normally. Choose Normal mode if the mapping contains an Update Strategy transformation. Choose Bulk to improve session performance. Bulk load limits the ability to recover because no database logging occurs. Default is Normal.
ENABLE_SPECIAL_CHARACTERS_IN_METADATA	Enables the support for special characters in metadata.
TARGET_CONN_TYPE	The type of the target connection. You can select the relational connection types such as Oracle, Microsoft SQL Server, DB2, and Sybase.
TARGET_CONN_NAME	The name of the target connection.
TARGET_PREFIX_NAME	The name prefix of the target table. Specify the target schema name if the target is in a different schema.

Register

You can register a test data warehouse to perform the test data warehouse operations. If you did not select Enable Test Data Warehouse from the Administrator tool, you can provide the URL to register test data warehouse from the command line.

The register command uses the following syntax:

```
ilmcmd
[options]
<-RegisterTestDataWarehouse | -rtdw>
[-tdwUrl | -tu] <URL>
[-Verbose | -v]
[-Help | -h]
```

The following table describes the register command options and arguments:

Option	Argument	Description
-RegisterTestDataWarehouse -rtdw	n/a	The test data warehouse server that you register to perform the test data warehouse operations.
-tdwUrl -tu	URL	The URL for the test data warehouse.

Option	Argument	Description
-Verbose -v	n/a	Optional. Verbose output. Displays more than the minimum information about the object.
-Help -h	n/a	Optional. Displays or prints the help options.

Register Example

The following sample command registers the test data warehouse server URL:

```
ilmcmd -rtdw -tu http://10.65.146.136:6605/tdm
```

CHAPTER 16

tdwcmd

This chapter includes the following topics:

- tdwcmd Overview, 269
- Running tdwcmd, 269
- Entering Options and Arguments, 270
- Syntax Notation, 270
- Version, 271
- List, 271

tdwcmd Overview

tdwcmd is a command line program that you can use to perform test data warehouse tasks.

Use the tdwcmd command to perform the following tasks:

- View the version details of the client on which the Test Data Manager Service is running.
- List the test data warehouse repository, test data warehouse, connection, and data set details.

You must have the required privileges to perform these tasks. For tasks related to projects, you must also have the required permission on the project.

Running tdwcmd

Invoke tdwcmd from the command line. You can issue commands directly or from a script, batch file, or other program.

On Windows, tdwcmd is a batch file with a .bat extension. On UNIX, tdwcmd is a shell script with no extension.

1. At the command line, switch to the directory where the tdwcmd executable file is located.

   By default, tdwcmd is installed in the following directory:

   ```
   <TDM installation directory>\utilities\tdwcli\bin
   ```

2. Enter tdwcmd followed by the command name and the required options and arguments.

   For example:

   ```
   tdwcmd -command_name [-option1] argument_1 [-option2] argument_2...
   ```

To view the command line syntax, enter the following command:

```
tdwcmd -h
```

Entering Options and Arguments

The tdwcmd command line program uses a set of options and arguments.

Use the following rules when you enter the tdwcmd command options and arguments:

- To enter options, type a hyphen followed by the program syntax for the command. Use two consecutive hyphens for the long name of the option, and use a single hyphen for the short name of the option.
- You can enter options in any order.
- If an argument contains spaces, enclose the argument in double quotes.
- The first word after the option is the argument.
- Most options require arguments. You must separate options from arguments with a single space.
- Commands, options, and object names are case sensitive.

Syntax Notation

To use the tdwcmd command line program, review the syntax notation.

The following table describes the syntax notation for the tdwcmd command line program:

Convention	Description
--x or -x	An option placed before an argument. The option designates the parameter that you enter. For example, to enter the user name for tdwcmd, type --UserName or -un followed by the user name.
<x>	Required option. If you omit a required option, the command line program returns an error message.
<x | y>	Select between required options. If you omit a required option, the command line program returns an error message. If a pipe symbol (|) separates options, specify one option. If there are no pipe symbols to separate the options, specify all the options.
[x]	Optional option. The command runs whether or not you enter these options.
[x | y]	Select between optional options. For example, you can display help for all tdwcmd commands by using the --Help or -h option. `[--Help|-h]` The command runs whether or not you enter the optional parameter.
< <x | y> **|** <a | b> **>**	When a set contains subsets, indicate the supersets with bold brackets (**< >**). Separate the subsets with the bold pipe symbol (|).

Version

You can print the version information for the client and the version on which the Test Data Manager Service is running.

With the tdwcmd version command, you can perform the following tasks:

- Find the version of the client on which the Test Data Manager Service is running.
- Print the version information of the client.

The tdwcmd version command uses the following syntax:

```
tdwcmd
<--version>
```

The following table describes tdwcmd CreateTDWREPO options and arguments:

Option	Argument	Description
--version	n/a	View and print the version details of the client on which the Test Data Manager Service is running.

List

You can search and list test data warehouse repository, connection, test data warehouse, and data set details.

You can search for the following object types in the repository:

- Test data warehouse repository
- Test data warehouse
- Connection
- Data set
- Data set version
- Data set version tables
- Test Data Manager Service details
- User authentication parameters

The list command uses the following syntax:

```
tdwcmd
<--list | -list>
[--Repository | -Repository]
[--TestDataWarehouse | -TestDataWarehouse]
[--Connection | -Connection]
<--Dataset | -Dataset>
[--TestDataWarehouseName | -twn] TestDataWarehouseName
<--DatasetVersion | -DatasetVersion>
[--TestDataWarehouseName | -twn] TestDataWarehouseName
[--DatasetName | -dn] DatasetName
<--DatasetVersionTable | -DatasetVersionTable>
[--TestDataWarehouseName | -twn] TestDataWarehouseName
[--TableDatasetName | -tdn] TableDatasetName
[--DatasetVersionNumber | -dvn] DatasetVersionNumber
<--HostName | -host_name] HostName
```

```
[--HTTPPort | -http_port] HTTPPort
[--HTTPSPort | -https_port] HTTPSPort
<--UserName | -un> UserName
<--Password | -pd> Password
<--SecurityDomain | -sdn> SecurityDomain
<--DomainName | -dn> DomainName
<--TDMServiceName | -tsn> TDMServiceName
[--Verbose | -v]
[--Help | -?, -h]
```

The following table describes the List command options and arguments:

Option	Argument	Description
--Repository -Repository	n/a	The test data warehouse repository.
--TestDataWarehouse -TestDataWarehouse	n/a	The relational database that stores the data related to data sets that you store in the test data warehouse.
--Connection -Connection	n/a	The connection to the data warehouse repository.
--Dataset -Dataset	n/a	Required. The data set that you create in the test data warehouse.
--TestDataWarehouseName -twn	TestDataWarehouseName	The name of the test data warehouse.
--DatasetVersion -DatasetVersion	n/a	Required. The version of the data set.
--DatasetName -dn	DatasetName	The name of the data set.
--DatasetVersionTable -DatasetVersionTable	n/a	Required. The version of the data set table.
--TableDatasetName -tdn	TableDatasetName	The name of the data set.
--DatasetVersionNumber -dvn	DatasetVersionNumber	The version number of the data set.
--HostName -host_name	HostName	Required. The TDM host name.
--HTTPPort -http_port	HTTPPort	Required if you do not specify an HTTPS port. Port number to connect to the TDM service.
--HTTPSPort -https_port	HTTPSPort	Required if you do not specify an HTTP port. Port number to connect to the Test Data Manager Service.
--UserName -un	UserName	Required. The Informatica domain user name.

Option	Argument	Description
--Password -pd	Password	Required. The password for the Informatica domain user.
--SecurityDomain -sdn	SecurityDomain	Required. The Informatica security domain.
--DomainName -dn	DomainName	Required. The name of the TDM domain.
--TDMServiceName -tsn	TDMServiceName	Required. The name of the Test Data Manager Service.
--Verbose -v	n/a	Optional. Verbose output. Displays more than the minimum information about the object.
--Help -?, -h	n/a	Optional. Displays or prints the help options.

List Examples

Listing a Test Data Warehouse Repository

The following sample command lists the test data warehouse repository details:

```
tdwcmd --list --Repository --HostName localhost --HTTPPort 6605 -un Administrator -pd
Administrator -sdn Native -dn  Domain_SandboxVMLin_TDM961HF14 --TDMServiceName TDMS
```

Listing a Test Data Warehouse

The following sample command lists the test data warehouse details:

```
tdwcmd --list --TestDataWarehouse --HostName localhost --HTTPPort 6605 -un Administrator
-pd Administrator -sdn Native -dn  Domain_SandboxVMLin_TDM961HF14 --TDMServiceName TDMS
```

Listing a Connection

The following sample command lists the connection details:

```
tdwcmd --list --Connection --HostName localhost --HTTPPort 6605 -un Administrator -pd
Administrator -sdn Native -dn  Domain_SandboxVMLin_TDM961HF14 --TDMServiceName TDMS
```

CHAPTER 17

tdwquery

This chapter includes the following topics:

- tdwquery Overview, 274
- Configuring tdwquery, 274
- Running tdwquery, 275
- Select Clause, 276

tdwquery Overview

tdwquery is a command line program that you can use to analyze data in a data set that you store in the test data warehouse.

You can run queries to view data in a data set. You can run queries on tables in a single data set at a time. You can use multiple joins, functions, and clauses to filter and search for data when you run queries.

Before you run a query, you must configure the properties file.

Configuring tdwquery

You must configure the tdwquery properties file before you can run a query.

1. Browse to the following location and find the properties file `userConfig.ilm`:

   `<Informatica installation directory>/TDM/utilities/tdwquery/conf/`

2. Use a text editor to open the file and enter the values of the properties in the file.

You must configure the following properties:

Property	Description
TDW_REPO_URL	The URL of the test data warehouse repository that you want to connect to. The URL appears in the **Repository** tab of the **Administrator	Test Data Warehouse** view in Test Data Manager.
USERNAME	Name of the TDM user that runs the query. The TDM user must have the Test Data Manager Service View Project privilege and at least read permission on the data set version.
PASSWORD	Password of the TDM user.
DOMAINNAME	Name of the domain to which the TDM user belongs.
NAMESPACE	Namespace of the TDM user.
SERVICENAME	Name of the Test Data Manager Service
TESTDATAWAREHOUSENAME	Name of the test data warehouse where you stored the data set.
DATASETVERSIONNAME	Name of the data set version that you want to run a query on.
PAGENUMBER	Required if you enter a value for PAGESIZE. The number of the page that you want to view.
PAGESIZE	Required if you enter a value for PAGENUMBER. The number of records that you want to appear on each page.

Running tdwquery

Run tdwquery from the command line. On Windows, tdwquery is a batch file with a .bat extension. On UNIX, tdwquery is a shell script with no extension.

You must configure the tdwquery properties file before you run a query.

1. At the command line, switch to the directory where the tdwquery executable is located.

   By default, tdwquery is installed in the following directory:
   `<Informatica installation directory>/TDM/utilities/tdwquery`.

2. To run a query:

   - On Windows, enter tdwquery followed by the clause and the required syntax.
     For example:

     ```
     tdwquery "SELECT * FROM td_empdetails"
     ```

     You can also prefix the schema name to the clause. For example:

     ```
     tdwquery "SELECT * FROM AA.td_empdetails"
     ```

   - On UNIX, enter ./ tdwquery followed by the clause and the required syntax.
     For example:

     ```
     ./tdwquery "SELECT * FROM td_empdetails"
     ```

You can also prefix the schema name to the clause. For example:

```
./tdwquery "SELECT * FROM AA.td_empdetails"
```

Select Clause

You can use the Select and Select Distinct clauses to run queries to analyze data in a table.

For example, the following query returns all data in table1:

```
SELECT * from table 1
```

The following query returns distinct data values from the columns col 1 and col 2 in table 2.

```
SELECT DISTINCT col 1, col 2 from table 2
```

You can use joins, aggregate functions, the Where clause with options, and other clauses with the Select and Select Distinct clauses.

You can run queries with table names and column names that contain spaces or dashes. You must enter the table name or column name within '\"'. Use the following syntax:

```
SELECT \"<column-name>\" from \"<table name>\"
```

Note:

- You cannot use other special characters in a query.
- You can use only English characters, the underscore ('_'), and numeric characters in a query.

Alias

You can use aliases when you run a query with the Select and Select Distinct clauses.

For example:

```
SELECT <alias_name>.<column_name> FROM <table_name> <alias_name>

SELECT <column_name> as <alias_name> from <table_name>

SELECT e.dept_id FROM Employee e

SELECT emp_name as name from TDWEMP

SELECT e.dept_id, e.name FROM Employee e WHERE e.name like 'J%'
```

Joins

You can use joins when you run a query with the Select and Select Distinct clauses.

You can use the following types of joins in a query:

Inner Join

Use the following syntax to perform an inner join:

```
SELECT <columns>
FROM <table 1>
INNER JOIN <table 2>
ON <table 1>.<column_name> = <table 2>.<column_name>
```

For example:

```
SELECT suppliers.supplier_id, suppliers.supplier_name, orders.order_date
FROM suppliers,
INNER JOIN orders
ON suppliers.supplier_id = orders.supplier_id
```

Left Outer Join

Use the following syntax to perform a left outer join:

```
SELECT <columns>
FROM <table 1>
LEFT OUTER JOIN <table 2>
ON <table 1>.<column_name> = <table 2>.<column_name>
```

For example:

```
SELECT suppliers.supplier_id, suppliers.supplier_name, orders.order_date
FROM suppliers
LEFT OUTER JOIN orders
ON suppliers.supplier_id = orders.supplier_id
```

Right Outer Join

Use the following syntax to perform a right outer join:

```
SELECT <columns>
FROM <table 1>
RIGHT OUTER JOIN <table 2>
ON <table 1>.<column_name> = <table 2>.<column_name>
```

For example:

```
SELECT orders.order_id, orders.order_date, suppliers.supplier_name
FROM suppliers
RIGHT OUTER JOIN orders
ON suppliers.supplier_id = orders.supplier_id
```

Full Outer Join

Use the following syntax to perform a full outer join:

```
SELECT <columns>
FROM <table 1>
FULL OUTER JOIN <table 2>
ON <table 1>.<column_name> = <table 2>.<column_name>
```

For example:

```
SELECT suppliers.supplier_id, suppliers.supplier_name, orders.order_date
FROM suppliers
FULL OUTER JOIN orders
ON suppliers.supplier_id = orders.supplier_id
```

Aggregate Functions

You can use aggregate functions when you run a query with the Select and Select Distinct clauses.

You can use the following types of aggregate functions in a query:

AVG

Use the following syntax:

```
SELECT AVG(<column_name>) FROM <table_name>
```

For example:

```
SELECT AVG(SAL) FROM TDWEMP
```

COUNT

Use the following syntax:

```
SELECT COUNT(<column_name>) FROM <table_name>
```

For example:

```
SELECT COUNT(*) FROM TDWEMP
```

MAX

Use the following syntax:

```
SELECT MAX(<column_name>) FROM <table_name>
```

For example:

```
SELECT MAX(SAL) FROM TDWEMP
```

MIN

Use the following syntax:

```
SELECT MIN(<column_name>) FROM <table_name>
```

For example:

```
SELECT MIN(SAL) FROM TDWEMP
```

SUM

Use the following syntax:

```
SELECT SUM(<column_name> FROM <table_name>
```

For example:

```
SELECT SUM(SAL) FROM TDWEMP
```

Where Clause

You can use the Where clause to specify filter criteria in a query with the Select and Select Distinct clauses.

You can run queries that use parentheses with the Where clause. For example:

```
SELECT * from Employee where NAME ='Mary' OR (NAME='Jessica' AND DEPT_ID=1)
```

You can use the Where clause with the following conditions in a query:

Single condition

Use the following syntax:

```
SELECT * FROM <table_name> WHERE <column_name> = <column_value>
```

For example:

```
SELECT * FROM customers WHERE state = 'California'
```

And

Use the following syntax:

```
SELECT * FROM <table_name> WHERE <column_name> = <column_value> AND <column_value>
<operator> <column_value>
```

For example:

```
SELECT * FROM customers WHERE state = 'California' AND company_name = 'Informatica'
```

OR

Use the following syntax:

```
SELECT * FROM <table_name> WHERE <column_name> = <column_value> OR <column_name>
<operator> <column_value>
```

For example:

```
SELECT * FROM customers WHERE state = 'California' OR company_name = 'Informatica'
```

Combining OR and AND

Use the following syntax:

```
SELECT * FROM <table_name> WHERE <column_name> = <column_value> OR <column_name>
<operator> <column_value> AND <column_name> <operator> <column_value>
```

For example:

```
SELECT * FROM customers WHERE state = 'California' OR available_credit > 500 and
revenue < 90
```

Joining and where

Use the following syntax:

```
SELECT columns FROM <table 1> LEFT OUTER JOIN <table 2> ON <table 1>.<col 1> =
<table 2>.<col 1> WHERE <table 1>.<col 1> <operator> <column_value>
```

For example:

```
SELECT suppliers.supplier_id, suppliers.supplier_name, orders.order_date FROM
suppliers LEFT OUTER JOIN orders ON suppliers.supplier_id = orders.supplier_id WHERE
supplier.supplier_id > 10
```

Exists

Use the following syntax:

```
SELECT DISTINCT <col 1> FROM <table 2> WHERE EXISTS (SELECT * FROM <table 2> WHERE
<table 1>.<col 1> <operator> <table 1>.<col 1>)
```

For example:

```
SELECT DISTINCT store_type FROM stores WHERE EXISTS (SELECT * FROM cities_stores
WHERE cities_stores.store_type = stores.store_type)
```

In

Use the following syntax:

```
SELECT * FROM <table 1> WHERE <col 1> IN (SELECT <col 1> FROM <table 2>)
```

For example:

```
SELECT * FROM Employee WHERE dept_id IN (SELECT DEPT_ID FROM TDWDEPT)
```

Like

Use the following syntax:

```
SELECT * FROM <table 1> WHERE <col 1> LIKE <value>
```

You can use characters or wildcards '%' and '_'. For example:

```
SELECT * FROM Employee WHERE name LIKE 'J%'
```

Is Null or Is Not Null

Use the following syntax:

```
SELECT * FROM <table 1> WHERE <col 1> IS NULL
```

For example:

```
SELECT * FROM Employee WHERE ssn IS NULL
```

Between

Use the following syntax:

```
SELECT * FROM <table 1> WHERE <col 1> BETWEEN <value> AND <value>
```

For example:

```
SELECT * FROM Employee WHERE delt_id BETWEEN 2 AND 3
```

Not

Use the following syntax:

```
SELECT * FROM <table 1> WHERE <col 1> NOT <CONDITION> <value>
```

For example:

```
SELECT DISTINCT * from STATE s WHERE NOT EXISTS (SELECT ct.state_id from city ct
where ct.state_id = s.state_id)
```

```
SELECT * FROM Employee WHERE dept_id NOT IN (SELECT DEPT_ID FROM TDWDEPT)
```

```
SELECT * FROM Employee WHERE name NOT LIKE 'J%'
```

```
SELECT * FROM Employee WHERE delt_id NOT BETWEEN 2 AND 3
```

Additional Clauses

You can use the following clauses when you run a query with the Select clause:

Group by

Use the following syntax:

```
SELECT <col 1> FROM <table 1> GROUP BY <col 1>
```

For example:

```
SELECT name FROM Employee GROUP BY NAME
```

Having

Use the following syntax:

```
SELECT <col 1> FROM <table 1> GROUP BY <col 1> HAVING <col 1> LIKE <value>
```

For example:

```
SELECT name FROM Employee GROUP BY NAME HAVING name like 'J%'
```

Order by

Use the following syntax. You can use the optional modifiers ASC and DESC:

```
SELECT <col 1> FROM <table 1> ORDER BY <col 1> ASC
```

For example:

```
SELECT dept_id, name FROM Employee ORDER BY dept_id ASC
```

APPENDIX A

Data Type Reference

This appendix includes the following topics:

- Data Type Reference Overview, 281
- Oracle, 282
- Microsoft SQL Server, 286
- DB2 for Linux, UNIX, and Windows, 290
- Sybase ASE, 292
- MySQL, 295
- Flat File, 297
- VSAM, 298
- DB2 for z/OS, 299
- DB2 for IOS, 301
- IMS Flat, 303
- Sybase IQ, 304
- Netezza, 306
- Teradata, 308
- Cassandra, 310
- MongoDB, 311

Data Type Reference Overview

You can pass different data types from sources to targets and perform TDM operations on the data types.

In TDM, you can perform data masking, data subset, and data discovery operations on the source databases for supported data types. You cannot perform the operations on data types that TDM does not support. For some data types, you might need to configure the precision and scale to a specific range of values.

Oracle

The following table describes the operations that you can perform on data types in Oracle database:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Binary_Double	Yes	Yes	Yes	Yes	Yes	
Binary_Float	Yes	Yes	Yes	Yes	Yes	
Blob	Yes	Yes	Yes	Yes	No	You can perform nullification masking.
Char	Yes	Yes	Yes	Yes	Yes	
Clob	Yes	Yes	Yes	Yes	No	
Date	Yes	Yes	Yes	Yes	Yes	You cannot run data domain profiles.
Dburitype	No	No	No	Yes	No	Use dictionary generation rule to generate data.
Decimal(P,S)	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Float	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles. To generate test data, you enter a precision value less than 15.
Httpuritype	No	No	No	Yes	No	Use dictionary generation rule to generate data.
Integer	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Interval day to second	Yes	Yes	No	No	No	You can enter the following criteria: Is Null, Is Not Null, Like, Not Like, Is Empty, and Is Not Empty. For the Interval day(9) to second(9) data type, you must set the precision greater than 18.
Interval year to month	Yes	Yes	No	Yes	No	You can enter the following entity criteria: Is Null, Is Not Null, Like, Not Like, Is Empty, and Is Not Empty. For the Interval year(9) to month data type, you must set the precision greater than 11. Use dictionary generation rule to generate data.
Long	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Longraw	Yes	Yes	Yes	Yes	No	You can perform nullification masking. Use dictionary generation rule to generate data.

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Longvarchar	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Nchar	Yes	Yes	Yes	Yes	Yes	
Nclob	Yes	Yes	Yes	Yes	No	Use string generation rule to generate data.
Number	Yes	Yes	Yes	Yes	Yes	
Nvarchar	Yes	Yes	Yes	Yes	Yes	
Nvarchar2(N)	Yes	Yes	Yes	Yes	Yes	
Raw	Yes	Yes	Yes	Yes	No	You can perform nullification masking. Use dictionary generation rule to generate data.
Real	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Rowid	Yes	Yes	No	No	No	
Smallint	Yes	Yes	Yes	Yes	Yes	
Timestamp	Yes	Yes	Yes	Yes	Yes	You cannot run data domain profiles.
Timestamp with time zone	Yes	Yes	No	No	No	You can enter the following entity criteria: Is Null, Is Not Null, Like, Not Like, Is Empty, and Is Not Empty.
Urowid	Yes	Yes	No	No	No	
Varchar	Yes	Yes	Yes	Yes	Yes	
Varchar2(N)	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
XML	Yes	Yes	No	Yes	No	You can enter the following entity criteria: Is Null, Is Not Null, Is Empty, and Is Not Empty. If you assign a dictionary generation rule that contains specified invalid values for the XML data type in an Oracle database, the data generation task fails.
Xdburitype	No	No	No	Yes	No	Use dictionary generation rule to generate data.

You cannot perform data masking, data subset, profiling, and data generation operations on the following Oracle data types:

- Anydata
- Anydataset
- Anytype
- Bfile
- Ordaudio
- Orddicom
- Orddoc
- Ordimage
- Ordimagesignature
- Ordvideo
- Sdo_Geometry
- Sdo_Georaster
- Sdo_Topo_Geometry
- SI_Averagecolor
- SI_Color
- SI_Colorhistogram
- SI_Featurelist
- SI_Positionalcolor
- SI_Stillimage
- SI_Texture

- Timestamp with local time zone
- Uritype

Microsoft SQL Server

The following table describes the operations that you can perform on data types in Microsoft SQL Server database:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bigint	Yes	Yes	Yes	Yes	Yes	
Binary	Yes	Yes	Yes	Yes	No	Use dictionary generation rule to generate test data. Use nullification data masking rule to mask the data.
Bit	Yes	Yes	Yes	Yes	Yes	If you enter data patterns, you must provide T and F values instead of 1 and 0.
Char	Yes	Yes	Yes	Yes	Yes	
Date	Yes	Yes	Yes	Yes	Yes	You can primary profiles.
Datetime	Yes	Yes	No	Yes	Yes	You cannot run data domain profiles.
Datetime2	Yes	Yes	Yes	Yes	No	
Datetimeoff set	Yes	Yes	No	Yes	No	
Decimal	Yes	Yes	Yes	Yes	Yes	You can enter a precision value up to 28.

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Float	Yes	Yes	Yes	Yes	Yes	To perform a data masking operation, you must set the precision to less than 28. When you run a test data generation operation, TDM truncates data if the precision value is greater than 17. You cannot run entity profiles.
Geography	Yes	Yes	No	No	No	You can enter the following entity criteria: Is Null and Is Not Null.
Geometry	Yes	Yes	No	No	No	You can enter the following entity criteria: Is Null and Is Not Null.
Image	Yes	Yes	Yes	Yes	No	
Int	Yes	Yes	Yes	Yes	Yes	
Money	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Nchar	Yes	Yes	Yes	Yes	Yes	
Ntext	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Numeric	Yes	Yes	Yes	Yes	Yes	You can enter a precision value up to 28.
Nvarchar	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Real	Yes	Yes	Yes	Yes	Yes	When you run a test data generation operation, TDM truncates data if the precision is greater than 15. You cannot run entity profiles.
Smalldatetime	Yes	Yes	No	Yes	Yes	You cannot run data domain profiles.
Smallint	Yes	Yes	Yes	Yes	Yes	
Smallmoney	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Sql_variant	Yes	Yes	No	Yes	No	You can enter the following entity criteria: Is Null, Is Not Null, Is Empty, and Is Not Empty. In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Text	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Time	Yes	Yes	Yes	Yes	Yes	When you enter values in milliseconds, TDM truncates the data. You can run primary key profiles.
Timestamp	n/a	n/a	n/a	Yes	No	
Tinyint	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Uniqueidentifier	Yes	Yes	No	Yes	Yes	You can enter the following entity criteria: Is Null, Is Not Null, Like and Not Like. You must not assign a data generation rule. You cannot run entity and data domain profiles.
Varbinary	Yes	Yes	Yes	Yes	No	Use dictionary generation rule to generate test data. Use nullification data masking rule to mask the data.
Varchar	Yes	Yes	Yes	Yes	Yes	
XML	Yes	Yes	No	Yes	No	You can enter the following entity criteria: Is Null and Is Not Null. Use dictionary generation rule to generate data.

You cannot perform data masking, data subset, profiling, and data generation operations on the following Microsoft SQL Server data types:

- Hierarchyid

DB2 for Linux, UNIX, and Windows

The following table describes the operations that you can perform on data types in DB2 database for Linux, UNIX, and Windows operating systems:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bigint	Yes	Yes	Yes	Yes	Yes	
Blob (N)	Yes	Yes	Yes	Yes	No	You can perform nullification masking. Use dictionary generation rule to generate data.
Char for bit data	Yes	Yes	Yes	Yes	No	
Char varying (Length)	Yes	Yes	Yes	Yes	Yes	To run a profile, you must keep the field length less than or equal to 255.
Character (N)	Yes	Yes	Yes	Yes	Yes	
Clob (N)	Yes	Yes	Yes	Yes	No	You can perform nullification masking.
Date	Yes	Yes	Yes	Yes	Yes	
Dbclob (N)	Yes	Yes	Yes	Yes	No	You can perform nullification masking.
Dec	Yes	Yes	Yes	Yes	Yes	
Decfloat (16)	Yes	Yes	Yes	Yes	No	
Decfloat (34)	Yes	Yes	Yes	Yes	No	
Decimal	Yes	Yes	Yes	Yes	Yes	
Double	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Double Precision	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Float	Yes	Yes	Yes	Yes	Yes	You can enter a precision value up to 28. You cannot run entity profiles.
Graphic (N)	Yes	Yes	Yes	Yes	Yes	
Integer	Yes	Yes	Yes	Yes	Yes	
Long Varchar	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Long Vargraphic	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Num	Yes	Yes	Yes	Yes	Yes	
Numeric	Yes	Yes	Yes	Yes	Yes	
Real	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Smallint	Yes	Yes	Yes	Yes	Yes	
Time	Yes	Yes	Yes	No	Yes	
Timestamp	Yes	Yes	Yes	Yes	Yes	
Varchar for Bit Data	Yes	Yes	Yes	<TBD>	No	
Varchar (N)	Yes	Yes	Yes	Yes	Yes	To run a profile, you must keep the field length less than or equal to 255.

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Vargraphic (N)	Yes	Yes	Yes	Yes	Yes	
XML	Yes	Yes	No	Yes	No	You can enter the following entity criteria: Is Null and Is Not Null. Use dictionary generation rule to generate data.

Sybase ASE

The following table describes the operations that you can perform on data types in Sybase ASE database:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bigdatetime	Yes	Yes	Yes	Yes	Yes	
Bigint	Yes	Yes	Yes	Yes	Yes	
Bigtime	Yes	Yes	Yes	Yes	Yes	
Bit	Yes	Yes	Yes	Yes	Yes	You can enter the following entity criteria: Is Null, Is Not Null, Like, and Not Like.
Binary (n)	Yes	Yes	Yes	Yes	No	You can perform nullification masking.
Char (n)	Yes	Yes	Yes	Yes	Yes	You can enter characters up to 2000. To run a profile, you must enter the field length less than or equal to 255.
Date	Yes	Yes	Yes	Yes	Yes	
Datetime	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Decimal (P,S)	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles. You can enter a precision value up to 28.
Double Precision	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles. You can perform data masking for precision values up to 28.
Float (P)	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles. You can perform data masking for precision values up to 28.
Image	Yes	Yes	Yes	Yes	No	Use dictionary generation rule to generate date. You can perform nullification masking.
Int	Yes	Yes	Yes	Yes	Yes	
Longsysnam e	Yes	Yes	Yes	Yes	Yes	
Money	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Nchar (N)	Yes	Yes	Yes	Yes	No	
Numeric (P,S)	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles. You can enter a precision value up to 28.
Nvarchar (N)	Yes	Yes	Yes	Yes	No	
Real	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Smalldateti me	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Smallint	Yes	Yes	Yes	Yes	Yes	
Smallmoney	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Sysname	Yes	Yes	Yes	Yes	Yes	
Text	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Time	Yes	Yes	Yes	Yes	Yes	
Timestamp	Yes	Yes	Yes	Yes	No	You can perform nullification masking. Timestamp values are system generated in Sybase database. You can restrict the data type from Test Data Manager to generate data.
Tinyint	Yes	Yes	Yes	Yes	Yes	When you assign a data generation rule, ensure that you provide a range between 0 and 255.
Unichar (N)	Yes	Yes	Yes	Yes	Yes	
Unitext	Yes	Yes	Yes	Yes	No	In a sequence generation rule, you cannot choose to start the sequence from the last output value.
Univarchar (N)	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Unsigned bigint	Yes	Yes	Yes	Yes	Yes	You cannot provide negative values.
Unsigned int	Yes	Yes	Yes	Yes	Yes	You cannot provide negative values.
Unsigned smallint	Yes	Yes	Yes	Yes	Yes	You cannot provide negative values.
Unsigned tinyint	Yes	Yes	Yes	Yes	Yes	When you assign a data generation rule, ensure that you provide a range between 0 and 255.
Varbinary (N)	Yes	Yes	Yes	Yes	No	You can perform nullification masking.
Varchar (N)	Yes	Yes	Yes	Yes	Yes	To run a profile, you must enter the field length less than or equal to 255.

MySQL

The following table describes the operations that you can perform on data types in MySQL database:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bigint	Yes	Yes	Yes	Yes	Yes	
Binary	Yes	Yes	Yes	Yes	No	
Bit	Yes	Yes	Yes	Yes	Yes	You cannot run data domain profiles.
Blob	Yes	Yes	Yes	Yes	No	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Char	Yes	Yes	Yes	Yes	Yes	
Date	Yes	Yes	Yes	Yes	Yes	
Datetime	Yes	Yes	Yes	Yes	Yes	
Decimal	Yes	Yes	Yes	Yes	Yes	
Double	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Enum	Yes	Yes	Yes	Yes	Yes	
Float	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Int	Yes	Yes	Yes	Yes	Yes	
Mediumblob	Yes	Yes	Yes	Yes	No	
Mediumint	Yes	Yes	Yes	Yes	Yes	
Mediumtext	Yes	Yes	Yes	Yes	No	
Set	Yes	Yes	Yes	Yes	Yes	
Smallint	Yes	Yes	Yes	Yes	Yes	
Text	Yes	Yes	Yes	Yes	No	
Time	Yes	Yes	Yes	Yes	Yes	
Timestamp	Yes	Yes	Yes	Yes	Yes	
Tinyblob	Yes	Yes	Yes	Yes	No	
Tinyint	Yes	Yes	Yes	Yes	Yes	
Tinytext	Yes	Yes	Yes	Yes	No	
Varbinary	Yes	Yes	Yes	Yes	No	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Varchar	Yes	Yes	Yes	Yes	Yes	
Year	Yes	Yes	Yes	Yes	Yes	TDM generates data for the year range from 1901 to 2155. For any other date range, TDM generates zero in the target database.

You cannot perform a data masking, data subset, profiling, and data generation operation on the following MySQL data types:

- Longblob
- Longtext

Flat File

The following table describes the operations that you can perform on data types in flat files:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bigint	Yes	Yes	Yes	Yes	Yes	
Datetime	Yes	Yes	Yes	Yes	Yes	
Double	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Int	Yes	Yes	Yes	Yes	Yes	
Nstring	Yes	Yes	Yes	Yes	Yes	
Number	Yes	Yes	Yes	Yes	Yes	
String	Yes	Yes	Yes	Yes	Yes	

VSAM

The following table describes the operations that you can perform on data types in VSAM:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bin	Yes	Yes	No	Yes	No	You can enter the following entity criteria: Is Null and Is Not Null.
Char	Yes	Yes	Yes	Yes	Yes	
Date	Yes	Yes	Yes	Yes	Yes	
Double	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Float	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num16	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num16U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num32	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num32U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num64	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num64U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num8	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num8U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Numchar	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Packed	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Spacked	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Time	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Timestamp	Yes	Yes	Yes	Yes	Yes	
Upacked	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Uzoned	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Varbin	Yes	Yes	No	Yes	No	You can enter the following entity criteria: Is Null and Is Not Null.
Varchar	Yes	Yes	Yes	Yes	Yes	
Zoned	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.

DB2 for z/OS

The following table describes the operations that you can perform on data types in DB2 for z/OS database:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bigint	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Binary	Yes	Yes	Yes	Yes	No	
Blob	n/a	n/a	n/a	Yes	n/a	You can generate data through the native client.
Char	Yes	Yes	Yes	Yes	Yes	
Char for bit data	Yes	Yes	Yes	Yes	No	
Clob	n/a	n/a	n/a	Yes	n/a	You can generate data through the native client.
Date	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Dbclob	n/a	n/a	n/a	Yes	n/a	You can generate data through the native client.
Decfloat	n/a	n/a	n/a	Yes	n/a	You can generate data through the native client.
Decimal	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Double	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles. You can enter a precision value up to 28.
Float	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Graphic	No	No	No	Yes	Yes	
Integer	Yes	Yes	Yes	Yes	Yes	
Long Varchar	Yes	Yes	Yes	Yes	No	
Long Varchar for bit data	Yes	Yes	Yes	Yes	No	
Long Vargraphic	No	No	No	Yes	No	
Real	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Smallint	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Time	Yes	Yes	Yes	Yes	Yes	
Timestamp	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Varbinary	Yes	Yes	Yes	Yes	No	
Varchar	Yes	Yes	Yes	Yes	Yes	
Varchar for Bit Data	Yes	Yes	Yes	Yes	No	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Vargraphic	No	No	No	Yes	Yes	
XML	n/a	n/a	n/a	Yes	n/a	You can generate data through the native client.

DB2 for IOS

You can perform data masking, data movement, data subset, and data generation operations on the DB2 for IOS data types. You cannot apply constraints on DB2 for IOS databases. To perform data discovery on DB2 for IOS data types, you must import profiles that you create in Informatica Developer.

The following table describes the operations that you can perform on data types in DB2 for IOS database:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Comments
Bigint	Yes	Yes	Yes	Yes	You must enter a precision value up to 19.
Binary	Yes	Yes	Yes	Yes	
Char	Yes	Yes	Yes	Yes	
Char for bit data	Yes	Yes	Yes	Yes	
Date	Yes	Yes	Yes	Yes	
Decimal	Yes	Yes	Yes	Yes	
Double	Yes	Yes	Yes	Yes	
Graphic	Yes	Yes	Yes	Yes	
Integer	Yes	Yes	Yes	Yes	You must enter a precision value up to 10.
Long Varchar	Yes	Yes	Yes	Yes	
Long Varchar for bit data	Yes	Yes	Yes	Yes	
Long Vargraphic	Yes	No	Yes	Yes	
Numeric	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Comments
Real	Yes	Yes	Yes	Yes	
Smallint	Yes	Yes	Yes	Yes	You must enter a precision value up to 19.
Time	Yes	Yes	Yes	No	To generate test data in the supported data types, you must restrict the Time data type.
Timestamp	Yes	Yes	Yes	Yes	
Varbinary	Yes	Yes	Yes	Yes	
Varchar	Yes	Yes	Yes	Yes	
Varchar for Bit Data	Yes	Yes	Yes	Yes	
Vargraphic	Yes	Yes	Yes	Yes	

When you want to generate test data, you must restrict all the above mentioned data types that are not supported. You cannot perform TDM operations on the following DB2 for IOS data types:

- Blob
- Clob
- Datalink
- Dbclob
- Clob
- Datalink
- Decfloat
- Rowid
- User_defined

IMS Flat

The following table describes the operations that you can perform on data types for IMS Flat file system:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bin	Yes	Yes	No	Yes	Yes	You cannot run entity profiles.
Char	Yes	Yes	Yes	Yes	Yes	
Date	Yes	Yes	Yes	Yes	Yes	
Double	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Float	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num16	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num16U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num32	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num32U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num64	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num64U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num8	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Num8U	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Numchar	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Packed	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Spacked	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Time	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Timestamp	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Upacked	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Uzoned	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Varbin	Yes	Yes	No	Yes	Yes	You cannot run entity profiles.
Varchar	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Zoned	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.

Sybase IQ

The following table describes the operations that you can perform on data types for Sybase IQ database:

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Bigint	Yes	Yes	Yes	Yes	Yes	
Binary	Yes	Yes	Yes	Yes	No	You can assign the nullification data masking rule.
Bit	Yes	Yes	Yes	Yes	Yes	You can enter the following entity criteria: Is Null and Is Not Null.
Blob	Yes	Yes	Yes	Yes	No	You can assign the nullification data masking rule.
Char	Yes	Yes	Yes	Yes	Yes	
Clob	Yes	Yes	Yes	Yes	No	
Date	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Datetime	Yes	Yes	Yes	Yes	Yes	
Decimal	Yes	Yes	Yes	Yes	Yes	You can enter precision value up to 28. You cannot run entity profiles.
Double	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Float	Yes	Yes	Yes	Yes	Yes	You can enter precision value up to 28. You cannot run entity profiles.
Image	Yes	Yes	Yes	Yes	No	You can assign the nullification data masking rule.
Int	Yes	Yes	Yes	Yes	Yes	
Long Binary	Yes	Yes	Yes	Yes	No	You can assign the nullification data masking rule.
Long Varchar	Yes	Yes	Yes	Yes	No	
Money	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Numeric	Yes	Yes	Yes	Yes	Yes	You can enter precision value up to 28. You cannot run entity profiles.
Real	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Smalldatetime	Yes	Yes	Yes	Yes	Yes	
Smallint	Yes	Yes	Yes	Yes	Yes	
Smallmoney	Yes	Yes	Yes	Yes	Yes	You cannot run entity profiles.
Sysname	Yes	Yes	Yes	Yes	Yes	

Data Types	Data Movement	Entity Criteria	Data Masking	Data Generation	Profiling	Comments
Text	Yes	Yes	Yes	Yes	No	
Time	Yes	Yes	Yes	Yes	Yes	
Timestamp	Yes	Yes	Yes	Yes	Yes	
Tinyint	Yes	Yes	Yes	Yes	Yes	When you assign a data generation rule, ensure that you provide a range between 0 and 255.
Unique Identifier	No	Yes	No	Yes	Yes	You can enter the Is Null entity criteria.
Unsigned bigint	Yes	Yes	Yes	Yes	Yes	
Unsigned int	Yes	Yes	Yes	Yes	Yes	
Unsigned tinyint	Yes	Yes	Yes	Yes	Yes	
Varbinary	Yes	Yes	Yes	Yes	No	You can assign the nullification data masking rule.
Varchar	Yes	Yes	Yes	Yes	Yes	

Netezza

The following table describes the operations that you can perform on data types in Netezza database:

Data Types	Data Generation
Bigint	Yes
Boolean	Yes
Bpchar	Yes
Byteint	Yes
Char	Yes

Data Types	Data Generation
Date	Yes
Double	Yes
Float	Yes
Int1	Yes
Int2	Yes
Int4	Yes
Int8	Yes
Integer	Yes
Nchar	Yes
Nvarchar	Yes
Real	Yes
Smallint	Yes
Time	Yes
Timestamp	Yes
Varchar	Yes

You cannot perform data masking, data subset, profiling, and data generation operations on the following Netezza data types:

- Decimal
- Decimal (38,0)
- Decimal (38,38)
- Interval
- Interval day
- Interval day to hour
- Interval day to minute
- Interval hour to second
- Interval hour
- Interval hour to minute
- Interval hour to second
- Interval minute
- Interval minute to second
- Interval month
- Interval second

- Interval year
- Interval year to month
- Numeric
- Timetz

Teradata

The following table describes the operations that you can perform on data types in Teradata database:

Data Types	Data Generation	Comments
Bigint	Yes	
Blob	Yes	You must set the precision in the range of 1 to 64000.
Byte	Yes	
Byteint	Yes	
Char	Yes	
Clob	Yes	You must set the precision in the range of 1 to 64000.
Date	Yes	
Decimal	Yes	You must set the precision in the range of 1 to 37.
Double precision	Yes	
Float	Yes	
Integer	Yes	
Interval day	Yes	
Interval day to hour	Yes	
Interval day to minute	Yes	
Interval day to second	Yes	
Interval hour	Yes	
Interval hour to minute	Yes	
Interval hour to second	Yes	
Interval minute	Yes	

Data Types	Data Generation	Comments
Interval minute to second	Yes	
Interval month	Yes	
Interval second	Yes	
Interval year	Yes	
Interval year to month	Yes	
Long varchar	Yes	
Real	Yes	
Smallint	Yes	
Time	Yes	
Time with time zone	Yes	You must set the precision in the range of 1 to 29.
Timestamp	Yes	
Timestamp with time zone	Yes	
Varbyte	Yes	
Varchar	Yes	

You cannot perform data masking, data subset, profiling, and data generation operations on the following Teradata data types:

- Graphic
- Long vargraphic
- Mbr
- Period (date)
- Period (time with time zone)
- Period (time)
- Period (timestamp with time zone)
- Period (timestamp)
- St_geometry
- Vargraphic

Cassandra

The following table describes the data masking and data domain discovery operations that you can perform on the Cassandra data types:

Data Types	Data Masking	Data Domain Discovery	Comments
ASCII	Yes	Yes	
Bigint	Yes	Yes	
Blob	Yes	Yes	
Boolean	Yes	Yes	
Decimal	Yes	Yes	
Double	Yes	Yes	
Float	Yes	Yes	
Inet	Yes	Yes	
Int	Yes	Yes	
Text	Yes	Yes	
Timestamp	Yes	Yes	
UUID	Yes	Yes	
TimeUUID	Yes	Yes	
Varchar	Yes	Yes	
Varint	Yes	Yes	

You cannot perform data masking and data domain discovery operations on the following Cassandra data types:

- Counter
- Custom

MongoDB

The following table describes the data masking and data domain discovery operations that you can perform on the MongoDB data types:

Data Types	Data Masking	Data Domain Discovery	Comments
Binary	Yes	No	You cannot perform data domain discovery and conditional data masking operations.
Boolean	Yes	Yes	
Date	Yes	Yes	
jstOID	Yes	Yes	
NumberDouble	Yes	Yes	You can enter a precision value up to 15.
NumberInt	Yes	Yes	You can enter a precision value up to 10.
NumberLong	Yes	Yes	You can enter a precision value from 1 to 28. You can a scale value from 0 to 28.
String	Yes	Yes	

You cannot perform data masking and data domain discovery operations on the following MongoDB data types:

- Timestamp

APPENDIX B

Data Type Reference for Test Data Warehouse

This appendix includes the following topics:

- Data Type Reference for Test Data Warehouse Overview, 312
- Oracle, 313
- Microsoft SQL Server, 315
- DB2 for Linux, UNIX, and Windows, 317
- Sybase ASE, 318
- Teradata, 320

Data Type Reference for Test Data Warehouse Overview

You can perform data subset, data masking, data discovery, and data generation operations and store the results in a test data warehouse.

When the target is test data warehouse, you can perform the TDM operations on the supported data types in the following databases:

- Oracle
- Microsoft SQL Server
- DB2 for Linux, UNIX, and Windows
- Sybase ASE
- Teradata

Oracle

The following table describes the data masking, data subset, data discovery, and data generation operations that you can perform on the Oracle data types when the target is test data warehouse:

Data Types	Test Data Warehouse	Comments
Binary_Double	Yes	You can enter a precision value up to 28.
Binary_Float	Yes	
Blob	Yes	
Char	Yes	
Clob	Yes	
Date	Yes	
Decimal(P,S)	Yes	You can enter a precision value up to 28. You cannot enter the data type with value Decimal(38,38).
Float	Yes	
Integer	Yes	
Long	Yes	
Longraw	Yes	
Longvarchar	Yes	
Nchar	Yes	
Nclob	Yes	
Number	Yes	You can enter a precision value up to 28. You cannot enter the data type with value Number(38,38).
Nvarchar2(N)	Yes	
Raw	Yes	
Real	Yes	
Smallint	Yes	
Timestamp	Yes	

Data Types	Test Data Warehouse	Comments
Timestamp with local time zone	Yes	You must set the precision to 40. You cannot perform the test generation operation.
Timestamp with time zone	Yes	You must set the precision to 40. You cannot perform the test generation operation.
Urowid	Yes	You cannot assign a data generation rule. TDM generates null values in the target.
Varchar2(N)	Yes	
XML	Yes	If you assign a dictionary generation rule that contains specified invalid values for the XML data type in an Oracle database, the data generation task fails.

When the target is test data warehouse, you cannot perform data masking, data subset, data discovery, and data generation operations on the following Oracle data types:

- Anydata
- Anydataset
- Anytype
- Bfile
- Dburitype
- Httpuritype
- Interval day to second
- Interval year to month
- Ordaudio
- Orddicom
- Orddoc
- Ordimage
- Ordimagesignature
- Ordvideo
- Rowid
- Sdo_Geometry
- Sdo_Georaster
- Sdo_Topo_Geometry
- Sl_Averagecolor
- Sl_Color

- SI_Colorhistogram
- SI_Featurelist
- SI_Positionalcolor
- SI_Stillimage
- SI_Texture
- Uritype
- Xdburitype

Microsoft SQL Server

The following table describes the data masking, data subset, data discovery, and data generation operations that you can perform on the Microsoft SQL Server data types when the target is test data warehouse:

Data Types	Test Data Warehouse	Comments
Bigint	Yes	
Binary	Yes	Use dictionary generation rule to generate test data. Use nullification data masking rule to mask the data.
Bit	Yes	
Char	Yes	
Date	Yes	
Datetime	Yes	
Datetime2	Yes	
Datetimeoffset	Yes	
Decimal	Yes	You can enter a precision value up to 28.
Float	Yes	To perform a data masking operation, you must set the precision to less than 28. When you run a data generation operation, TDM truncates data if the precision is greater than 17.
Image	Yes	
Int	Yes	
Money	Yes	
Nchar	Yes	

Data Types	Test Data Warehouse	Comments
Ntext	Yes	To perform a reset operation on tables with Ntext data type, create a logical primary key in Test Data Manager if a physical primary key does not exist.
Numeric	Yes	You can enter a precision value up to 28.
Nvarchar	Yes	
Real	Yes	When you run a data generation operation, TDM truncates data if the precision is greater than 15. To perform a reset operation on tables with Real data type, create a logical primary key in Test Data Manager if a physical primary key does not exist.
Smalldatetime	Yes	
Smallint	Yes	
Smallmoney	Yes	
Sql_variant	Yes	You cannot perform a data masking operation.
Text	Yes	To perform a reset operation on tables with Text data type, create a logical primary key in Test Data Manager if a physical primary key does not exist.
Time	Yes	When you enter values in milliseconds, TDM truncates the data.
Tinyint	Yes	
Uniqueidentifier	Yes	You cannot assign string masking rule even if the data type is present in the `dataMapping.xml` file. You cannot perform data generation operations.
Varbinary	Yes	Use dictionary generation rule to generate test data. Use nullification data masking rule to mask the data.
Varchar	Yes	
XML	Yes	You cannot perform a data masking operation.

When the target is test data warehouse, you cannot perform data masking, data subset, data discovery, and data generation operations on the following Microsoft SQL Server data types:

- Geography
- Geometry
- HierarchyID
- Timestamp

DB2 for Linux, UNIX, and Windows

The following table describes the data masking, data subset, data discovery, and data generation operations that you can perform on the DB2 data types when the target is test data warehouse:

Data Types	Test Data Warehouse	Comments
Bigint	Yes	
Blob (N)	Yes	
Char for bit data	Yes	You can perform the data generation operation.
Char varying (Length)	Yes	
Character (N)	Yes	
Clob (N)	Yes	
Date	Yes	
Dbclob (N)	Yes	
Dec	Yes	You can perform the data generation operation.
Decfloat (16)	Yes	
Decfloat (34)	Yes	
Decimal	Yes	
Double	Yes	
Double Precision	Yes	
Float	Yes	
Graphic (N)	Yes	
Integer	Yes	
Long Varchar	Yes	
Long Vargraphic	Yes	
Num	Yes	
Numeric	Yes	
Real	Yes	
Smallint	Yes	
Time	Yes	You cannot assign a data generation rule. TDM generates null values in the target.

Data Types	Test Data Warehouse	Comments
Timestamp	Yes	
Varchar for Bit Data	Yes	
Varchar (N)	Yes	
Vargraphic (N)	Yes	
XML	Yes	Use dictionary generation rule to generate data. You cannot perform data masking operations.

Sybase ASE

The following table describes the data masking, data subset, data discovery, and data generation operations that you can perform on the Sybase ASE data types when the target is test data warehouse:

Data Types	Test Data Warehouse	Comments
Bigdatetime	Yes	
Bigint	Yes	
Bigtime	Yes	
Bit	Yes	
Binary (n)	Yes	
Char (n)	Yes	
Date	Yes	
Datetime	Yes	
Decimal (P,S)	Yes	You can enter a precision value up to 28. You cannot enter the data type with value Decimal(38,38).
Double Precision	Yes	
Float (P)	Yes	
Image	Yes	
Int	Yes	
Longsysname	Yes	

Data Types	Test Data Warehouse	Comments
Money	Yes	
Nchar (N)	Yes	
Numeric (P,S)	Yes	You can enter a precision value up to 28. You cannot enter the data type with value Numeric(38,38).
Nvarchar (N)	Yes	
Real	Yes	
Smalldatetime	Yes	
Smallint	Yes	
Smallmoney	Yes	
Sysname	Yes	
Text	Yes	
Time	Yes	
Tinyint	Yes	When you assign a data generation rule, ensure that you provide a range between 0 and 255.
Unitext	Yes	
Unsigned bigint	Yes	You cannot provide negative values.
Unsigned int	Yes	You cannot provide negative values.
Unsigned smallint	Yes	You cannot provide negative values.
Unsigned tinyint	Yes	When you assign a data generation rule, ensure that you provide a range between 0 and 255.
Varbinary (N)	Yes	
Varchar (N)	Yes	

When the target is test data warehouse, you cannot perform data masking, data subset, data discovery, and data generation operations on the following Sybase ASE data types:

- Timestamp. You cannot perform a data masking operation on Timestamp data type because the value is system generated in Sybase database.
- Unichar
- Univarchar

Teradata

The following table describes the operations that you can perform on data types in a Teradata database:

Data Types	Test Data Warehouse	Comments
Bigint	Yes	
Blob	Yes	You must set the precision in the range of 1 to 4000.
Byte	Yes	
Byteint	Yes	
Char	Yes	
Char varying	Yes	
Clob	Yes	You must set the precision in the range of 1 to 4000.
Date	Yes	
Decimal	Yes	You must set the precision in the range of 1 to 38.
Double precision	Yes	
Float	Yes	
Integer	Yes	
Interval day	Yes	
Interval day to hour	Yes	
Interval day to minute	Yes	
Interval day to second	Yes	
Interval hour	Yes	
Interval hour to minute	Yes	
Interval hour to second	Yes	
Interval minute	Yes	
Interval minute to second	Yes	
Interval month	Yes	
Interval second	Yes	
Interval year	Yes	

Data Types	Test Data Warehouse	Comments
Interval year to month	Yes	
Long varchar	Yes	
Number	Yes	
Real	Yes	
Smallint	Yes	
Time	Yes	
Time with time zone	Yes	You must set the precision in the range of 1 to 29.
Timestamp	Yes	
Timestamp with time zone	Yes	
Varbyte	Yes	
Varchar	Yes	

You cannot perform test data warehouse operations on the following Teradata data types:

- Graphic
- Long vargraphic
- Mbr
- Period (date)
- Period (time with time zone)
- Period (time)
- Period (timestamp with time zone)
- Period (timestamp)
- St_geometry
- Vargraphic

APPENDIX C

Glossary

accelerator

An Informatica solution that contains predefined data subset and data masking components for major business applications.

activity

In the empty shell creation process, defines a set of related operations in the source or target system. Use the Informatica Oracle Shell Build tool to create an empty shell.

ad hoc generation rule

A data generation rule that you can create within a project. You can use ad hoc generation techniques within an ad hoc generation rule.

advanced masking

A data masking technique that can include other masking techniques and multiple input and output ports.

auto cascade

A cascade type that masks the primary key in the parent table and the related foreign keys in the child tables with the same value.

blurring

A masking parameter for numeric and date values that creates an output value within a fixed or percent variance from the source data value. Configure blurring to return a random value that is close to the original value.

calculated masking

A masking technique that uses an expression to combine the values of multiple ports.

cascade

A data masking component that masks columns across tables based on column value. Define similar columns in Test Data Manager in order to mask them with the same rules.

column profile

A type of data analysis that determines the characteristics of columns in a data set, such as value frequency, percentages, and patterns.

custom generation rule

A data generation rule that uses PowerCenter mapplets to generate data.

data discovery

The process of discovering the metadata of source systems that includes content, such as data values and frequencies, and structure such as primary keys, foreign keys, and functional dependencies.

data generation

The process to generate realistic test data for the testing environment without using the production data.

data masking

The process of replacing sensitive columns of source data with realistic test data.

data subset

A small, targeted, and referentially intact copy of production data.

dependent masking

A masking technique that replaces the values of one source column based on the values of another source column.

dictionary

A flat file or relational table of data that substitution masking uses to replace a column of data with similar but unrelated data.

dictionary generation

A standard data generation technique that imports dictionary values into the target column.

effective dates generation

An ad hoc generation technique that generates related dates for two columns in the target database.

empty shell

A database that contains no application data but contains metadata that is identical to the metadata in a source database. You can use an empty shell as the target database in a data subset plan to create a non-production database. Use the Informatica Oracle Shell Build tool to create an empty shell for an Oracle database.

encryption masking

A data masking technique that applies encryption algorithms to mask source data. You can choose the algorithm to encrypt the data.

entity

Defines a set of tables that are related based on physical or logical constraints. An entity can contain parent tables and child tables. An entity maintains relational hierarchy in the tables. Define filter parameters for ports in the entity to extract the columns from the tables in an entity. When you run a workflow generated from a data subset plan, the PowerCenter Integration Service extracts source data based on the filter parameters defined in the entity and loads the data to the subset tables.

Expression Builder

A wizard in Test Data Manager that includes Informatica functions, variables, and operators to build expressions from the input and the output column values.

expression generation

An ad hoc generation technique that generates data based on the expression.

expression masking

A data masking technique that applies an expression to a port to change or create data.

foreign key profile

A type of data analysis that finds column values in a data source that match the primary key column values in another data source.

generation rule

A rule that you create to define how you want to generate data in the target table.

group

Defines a set of tables to copy to a target subset database.

ilmcmd

A command line program that you can use to complete TDM tasks from any machine. ilmcmd allows you to complete tasks that you can complete in Test Data Manager.

key masking

A data masking technique that produces deterministic results for the same source data, masking rules, and seed value.

mask format

A masking parameter that limits each character in the output column to an alphabetic, numeric, or alphanumeric character.

nullification masking

A data masking technique that replaces column data with null values.

pattern

A data masking component that uses a regular expression to filter the ports that you want to mask when you assign a rule to columns. Use a pattern to mask sensitive data such as IDs, telephone numbers, postal codes, and state names.

plan

Defines data subset, data masking, or data generation operations. You can add entities, groups, templates, policies, rules, and tables to a plan. When you generate and run workflows from a plan, the PowerCenter Integration Service generates and runs the workflows.

policy

Defines the data masking rules, the data to mask, and masking parameters for a source.

PowerCenter

An Informatica solution that can extract data from multiple sources, transform the data according to business logic you build in the client application, and load the transformed data into file and relational targets for data warehousing.

profile

Used in data discovery.

Defines the tables, fields, and keys to discover. You can create profiles to discover the following items in tables:

- Sensitive fields
- Foreign keys
- Primary keys

project

A container component for entities, groups, templates, and one or more sources that you want to use in data discovery, subset, and masking operations. When you create a project, you add one or more sources to the project. Any source that you add to a project is not available to other projects.

random generation

A standard data generation technique that generates random data for string, numeric, and date datatypes.

random masking

A masking technique that produces random, non-repeatable results for the source data and masking rules.

range

A data masking parameter that defines the minimum and maximum values to mask the numeric and date datatypes, and the minimum and maximum values for string lengths of string datatype.

reference lookup generation

An ad hoc generation technique that generates data from a reference table.

residual table

In a data subset plan, a table that is missing from entities and groups in the plan but which is part of the project. You can add residual tables to a data subset plan.

result string replacement characters

A data masking parameter that defines the substitute characters for masked data.

rule

Defines the data masking technique, an optional rule qualifier, and masking parameters.

rule qualifier

A rule component that filters the records that you want to mask.

seed

A start number that key masking uses to return deterministic data.

sequence generation

A standard data generation technique that generates numeric and date values in a sequence.

set of values generation

A standard data generation technique that defines a finite set of values to write to the target table.

shuffle masking

A data masking technique that masks the data in a column with data from the same column from another row of the table.

simulation masking

A masking type that generates a simulation report without masking any data. The report contains policies, rules, masking parameters, qualified ports, and column values that should be masked and their corresponding masked values.

source string characters

A data masking technique that masks the selected characters. You can choose to mask the selected source string characters or mask all characters except the source string characters.

standard generation rule

A data generation rule that applies a predefined data generation technique.

substitution masking

A data masking technique that replaces a column of data with similar but unrelated data.

TDM repository

A relational database that stores the components that you define in Test Data Manager, such as policies, projects, entities, and data masking rules. The TDM repository stores metadata that you import into Test Data Manager from a source database or from the PowerCenter repository. The TDM repository stores constraints that define relationships between the source tables in a project.

TDM Server

Server that runs Test Data Manager and integrates with Informatica application services to perform data subset, data masking, and data discovery operations.

tdwcmd

A command line program that you use to perform test data warehouse tasks. You can create a test data warehouse repository, create a test data warehouse, and list all the test data warehouse details.

tdwquery

A command line program that you use to analyze data in a data set version. You can run queries with multiple joins and clauses to filter and search for data.

template

Used in data subset.

A reusable component that contains the entities and groups of a specific module that you want to subset based on common parameters.

Test Data Management (TDM)

The Informatica solution that bundles Data Subset, Data Generation, and Data Masking to protect sensitive data and create lean nonproduction systems for test and development purposes.

Test Data Manager

The web-based user interface that you use to configure and run data subset, data masking, and discovery operations.

test data repository

A relational database that stores the project metadata and the source table metadata related to data sets that you store in the test data warehouse.

test data warehouse

A central and location to store and manage test data in TDM. You can store data from data subset, data masking, and data generation plans. Restore or reset required data from the test data warehouse to a target connection in TDM.

value cascade

A cascade type that masks the column values in different tables with the same value.

INDEX

A

abort
  workflows 244
Actions menu
  Test Data Manager 31
ad hoc
  generation plan 240
ad hoc generation rules
  creating 173
  description 173
Administrator view
  description 37
advanced masking
  description 135
  parameters 136
advanced options
  plan settings 226
advanced rules
  assigning rule to columns 124
  creating 58
  description 57
application services
  TDM 22
architecture
  TDM 21
asset linking
  create link 69
  delete link 69
assign
  advanced 126
    advanced rule
      assignment 126
  custom generation 177
assigning
  rules in policies to columns in an project source 124
assignments
  deleting 126
  multiple rules to a column 124
  overriding a rule 124
  rules to columns 122, 176
auto cascades
  overview 77

B

blurring
  random masking 146

C

calculated masking
  description 137
cascades
  auto 77
  value 75
CLI
  ilmcmd description 256
column profile
  profiled patterns 90
  results 90
  inferred rules 81
  list of values 81
  patterns 81
  range 81
  sampling options 82
columns
  cascades 75
  sensitivity 75
commit interval
  workflow properties 225
commit type
  workflow properties 223, 225
components
  data generation 167
  data masking 120
conditional
  constraints 175
conditional constraint
  create 95
connections
  Administrator view 37
  configuring workflows 218
  plan settings 220
  TDM 23
  to databases 24
  to repositories 24
  to services 23
constraint
  major and minor severity 92
    relationships
      optimize 104
constraint type
  conditional 95
constraint-based loading
  workflow properties 223, 225
constraints
  adding 93
  reviewing changes 101
content panel
  Test Data Manager 31
copying
  plans 240
creating
  conditional constraints 95
  entities 102
credit card masking
  description 137

criteria
parameters 102
subset and masking 230
custom generation
assign 177
parameters 185
custom generation rules
description 172
custom mapplet
masking rule 135
custom masking
parameters 135

D

dashboards
Test Data Manager 33
data discovery
column profile 81
nonrelational sources 73
overview 72
rules and guidelines 73
task flow 77
TDM process 26
data discovery profiles
creating 84
deleting 85
editing 85
data domain
default masking rules 60
delete 63
edit metadata 63
data domain definitions
data patterns 61
metadata patterns 61
data domain profile
sampling options in 80
data domains
assigning to multiple columns 80
copying 63
creating 62
data generation rules 60
discovering 79
exportingand importing 67
manually adding to columns 80
options 62
overview 59
status 62
data expressions
patterns 61
data generation
expression generation 190
random generation 193
sequence generation 201
set of values 202
ad hoc rules 173
assign auto rule 178
components 167, 230
conditional constraints 175
creating ad hoc rules 173
creating standard rules 170
custom rules 172
deleting rules 171
dictionary generation 186
editing rules 171
effective dates generation
parameters 188
data generation *(continued)*
exception test data 183
overview 166
parameters 183
plans 180
rules 181
standard rules 170
task flow 167
techniques 181, 182
data generation parameters
random generation 191, 192
reference lookup 198
data generation rules
preferred rules 60
data masking
assigning multiple rules to a column 124
assigning rules in policies to columns 124
components 120
creating advanced rules 58
creating rules from a mapplet 59
creating standard rules 56
delete rules 122
disabling at the plan level 231
editing rules 121
masking techniques list 131
metadata and data patterns 60
overview 119
plans 217
repeatable output 133
rules 131
rules overview 55
standard rules 56
task flow 120
techniques 131
data masking plans
creating 238
data patterns
data domain definitions 61
TDM data domains 60
data set
edit 213
classify tables 208
create 206
delete 213
lock 204
metadata 213
tags 213
unlock 204
data set version
delete 213
data subset
creating groups 106
editing groups 106
editing templates 109
entities 102, 104
overview 96
plans 118, 128, 217
process 26, 27
task flow 97
templates 108
data subset plans
creating 238
databases
connections to 24
TDM 23
date format string
workflow properties 225

date values
key masking 144
random masking 148
default rule
data domain 170
TDM data domains 60
Define view
Test Data Manager 41
definitions
TDM repository 47
delete
ilmcmd command 258
deleting 241
details panel
Test Data Manager 31
dictionary connections
plan settings 220
relational and flat file 222
dictionary generation
description 185
parameters 186
dictionary location
flat files 222
discover
discover
discover
columns 74
profiles 74
tables 74
Discover view)
Test Data Manager 41
discovering
data domains 79
primary key 77
domains
assigning to multiple columns 80
exporting and importing 67
duplicate
data domains 63
groups 106
masking rules 122
policies 70
projects 45

E

editing
entities 104
effective dates generation
description 187
email masking
description 138
parameters 138
encryption masking
description 140
entities
circular relationships 97
creating 102
deleting 105
editing 104
exporting TDM 105
importing TDM 105
overview 97
entity
TDM projects 42
entity discovery
description 72
entity profiles
inference options 78
example
data subset with XSD source 111
examples
TDM 28
export
ilmcmd command 260
exporting
entities 105
groups 107
policy 67
TDM projects 46
expression
cascade 126
Expression Builder
overview 38
expression generation
parameters 190
description 189
expression masking
description 140
parameters 141
rules and guidelines for 141

F

filter criteria
disabling masking by column 231
filtering data
Test Data Manager 32
foreign key
constraints 93

G

generate staging DDL
workflow properties 225
generating
PowerCenter mappings 128
workflows 128, 180
generation rules
assigning 177
deleting 171
description 168
editing 171
generation techniques
dictionary 185
effective dates 187
expression 189
random 191
reference lookup 197
sequence 199
set of values 201
groups
copying 106
creating 106
deleting 107
editing 106
exporting TDM 107
importing TDM 107
overview 106

H

high availability recovery
  workflow properties 225
high precision
  workflow properties 225

I

ilmcmd
  command line help 257
  register 267
  reset 265
ilmcmd commands
  delete 258
  entering options and arguments 257
  export 260
  import 261
  overview 256
  search 262
  syntax for 258
  workflow 264
import
  ilmcmd command 261
importing
  entities 105
  groups 107
  metadata from PowerCenter into TDM 50
  plans 241
  profiles 86
  TDM projects 46
importing metadata
  from PowerCenter into Test Data Manager 50
index
  disabling in plan 226
inference options
  entity profiles 78
inferred rule
  types 81
integers
  masking large integers 123
IP address masking
  description 141

K

key masking
  date values 144
  description 142
  masking numeric values 144
  masking string values 144
  result string replacement characters 143
  source string characters 143

L

limit
  data subset plan 230
locale
  plan settings 226
logging in
  Test Data Manager 39
logs
  project 43

M

major severity
  configuring constraints 92
manage logs
  project 45
mandatory columns
  description 58
mapplet
  parameters 135, 185
mapplets
  assigning rule to columns 124
  creating rules from 58
  TDM 129
mark changes review complete
  entity changes 101
mask format
  description 142
masking
  disabling by column 230
  large integers 123
masking options
  blurring 146
  mask format 142
  range 146
  result string replacement characters 143
  seed 133
  source string characters 143
masking parameters
  advanced masking 136
  date key masking 144
  date random masking 148
  email masking 138
  expression masking 141
  numeric key masking 144
  numeric random masking 149
  random strings 149
  string key masking 144
  substitution 152
masking rule
  copying 122
  custom mapplet 135
masking rules
  delete 121, 122
  editing 121
  update 121
masking techniques
  advanced 135
  calculated 137
  credit card 137
  email 138
  encryption 140
  expression 140
  IP address 141
  key masking 142
  lists 131
  nullification 145
  phone 145
  random 145
  Social Security number 151
  substitution 152
  URL 153
maximum confidence percent
  inference options 78
maximum foreign keys returned
  inference options 78
metadata patterns
  data domain definitions 61

metadata patterns *(continued)*
TDM data domains 60
minimum confidence percent
inference options 78
minor severity
configuring constraints 92
Monitor
details 248
jobs 248, 250
logs 250
tasks 249
Monitor view
description 247
Test Data Manager 37

N

nullification masking
description 145
numeric random generation
parameters 193
numeric sequence generation
parameters 201
numeric values
key masking 144
random masking 149

O

overriding a rule
data masking 124
overview
data discovery 72
data generation 166
data masking 119
groups 106
TDM 18
Overview view
Test Data Manager 33

P

parameters
data generation 181, 183
data masking 131
date sequence generation 200
partitioning
properties 236
patterns
data domain definitions 61
TDM data domains 60
permission view
data set version 214
projects 52
Persistent Data Masking
masking Social Insurance numbers 151
phone masking
description 145
plan criteria
configuring 230
plan settings
advanced options 226
connections 220
description 220
source settings 231
plan settings *(continued)*
Workflow Executions view 244
Workflow view 242
plans
creating 238
data generation 180
data subset 118, 128
deleting 241
disabling masking for columns 231
generating PowerCenter workflows from 242
importing TDM 241
properties 219
TDM projects 42
TDM workflows 217
policies
copying 70
creating 70
editing 70
exporting and importing 67
managing 69
overview 54
task flow 55
Policies view
Test Data Manager 35
policy
delete 71
export 67
PowerCenter
objects 129
PowerCenter mappings
generating 128
PowerCenter repository
importing metadata from 50
PowerCenter workflows
generating from plans 242
PowerExchange sources
importing 50
preferred masking rules
overview 60
primary key
create constraint 93
primary key discovery
description 72
primary key options
primary key profiles 78
primary key profiles
description 77
inference options 78
results 87
processes
data subset 26, 27
profile
TDM projects 42
profiles
data domains 79
deleting 85
editing 85
importing 86
primary key and entity discovery 77
project
list of tables 91
project risk analysis
description 33
project security
maintaining in TDM 52
projects
copying 45
creating in TDM 44

projects *(continued)*
deleting in TDM 45
editing in TDM 44
exporting TDM 46
importing TDM 46
TDM 41
TDM components 42
Projects view
Test Data Manager 36
properties
plans 219
properties panel
workflow executions 245

Q

quick links
Overview view 33

R

random generation
numeric random 193
description 191
random masking
blurring 146
description 145
masking date values 148
masking string values 149
numeric values 149
range 146
result string replacement characters 143
source string characters 143
range
random masking 146
reference lookup generation
description 197
parameters 198
regenerate signature
inference options 78
regular expressions
syntax 60
repeatable output
data masking 133
reports
Test Data Manager 252
repositories
connections to 24
result string replacement characters
key masking 143
random masking 143
results
primary key profiles 87
resume from last checkpoint
workflow properties 225
review changes
entity constraints 101
rollback transactions on error
workflow properties 225
rule assignment
multiple rules to a column 124
rule assignments
description 122, 176
overriding the rule 124
rule simulation
preview rule output 57
rule simulation *(continued)*
simulator 56
standard rule 56
rules
ad hoc generation rules 173
adding to policies 70
advanced 57
assigning multiple rules to a column 124
assigning to columns 122, 124, 176
creating from a mapplet 59
custom generation rules 172
deleting column assignments 126
exporting and importing 67
from PowerCenter mapplets 58
not available for assignment 123, 177
overview 55
standard generation rules 170
standard masking rules 56
rules and guidelines
expression masking 141

S

sampling options
data domain profile 80
search
ilmcmd command 262
seed 133
sensitivity
for ports 75
sensitivity levels
data domains 62
sequence generation
description 199
numeric 201
services
connections to 23
session
details 245
sessions
tab 245
set of values generation
parameters 202
description 201
severity
configuring constraints 92
similar value columns
deleting the group 76
description 75
editing 76
setting 76
SIN numbers
masking Social Insurance numbers 151
source settings
plan settings 231
source string characters
key masking 143
random masking 143
sources
TDM 47
special format masking
Social Insurance numbers 151
special masking techniques
credit card 137
email 138
IP address 141
phone 145

special masking techniques *(continued)*
Social Security number 151
URL 153
SSN masking
description 151
standard generation rules
creating 170
description 170
standard masking rules
description 56
standard rules
creating 56
start digit
Social Insurance numbers 151
status
data domains 62
stop on error
workflow properties 225
string values
key masking 144
random data masking 149
subset
relations
optimize 104
substitution masking
description 152
parameters 152
syntax
ilmcmd commands 258

T

table classification
description 91
tables
classification 91
tables per mapping
description 226
targets
TDM 47
task flow
data subset 97
TDM
application services 22
architecture 21
connections 23
database connections 24
databases 23
example 28
overview 18
process overview 25
repository connections 24
service connections 23
sources 47
targets 47
tools 22
TDM processes
data discovery 26
TDM projects
components 42
managing 43
owner 43
TDM repository
definitions 47
tdwcmd
arguments 270
list 271
tdwcmd *(continued)*
options 270
overview 269
run command 269
syntax notation 270
version 271
techniques
data generation 181, 182
data masking 131
templates
creating 108
deleting 109
description 108
editing 109
TDM projects 42
Test Data Manager
Administrator view 37
filtering data 32
logging in to 39
Monitor view 37
Overview view 33
Policies view 35
Projects view 36
reports 252
user interface overview 30
views 31
test data warehouse
reset 206
reset a data set 209
security
data set version 214
tags 214
thinkmaps
for policy assignments 252
for rule assignments 252
tools
TDM 22
trim spaces before comparison
inference options 78

U

unique key
create constraint 94
URL masking
description 153

V

value cascades
deleting 76
description 75
editing 76
setting 76

W

workflow
execute 243
workflow executions
properties panel 245
Workflow Executions view
plan settings 244
Sessions tab 245
tasks 244

Workflow view
plan settings 242
workflows
aborting 244
connections for 218
generating 128, 180
generating data subset 118
generating from plans 242
ilmcmd command 264
properties in plan settings 223, 225
recovering 244

X

XSD
data generation 179
data masking 127
data subset 109
XSD source
criteria 110
example 111
scope 110