



Informatica® PowerExchange for Amazon S3
10.1.1

User Guide for PowerCenter

© Copyright Informatica LLC 2015, 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 © 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 © TMat 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/hsqldbLicense.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.schneider.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/ssl/>, <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: 2018-09-27

Table of Contents

Preface	6
Informatica Resources.	6
Informatica Network.	6
Informatica Knowledge Base.	6
Informatica Documentation.	6
Informatica Product Availability Matrixes.	7
Informatica Velocity.	7
Informatica Marketplace.	7
Informatica Global Customer Support.	7
 Chapter 1: Introduction to PowerExchange for Amazon S3.....	 8
PowerExchange for Amazon S3 Overview.	8
PowerCenter Integration Service and Amazon S3 Integration.	8
Introduction to Amazon S3.	9
Amazon S3 Objects.	9
Amazon S3 Object Format.	9
 Chapter 2: PowerExchange for Amazon S3 Configuration.....	 10
PowerExchange for Amazon S3 Configuration Overview.	10
Prerequisites.	10
Create an Access Key ID and Secret Access Key.	10
Enable Client-side Encryption.	11
Create Minimal Amazon S3 Bucket Policy.	11
Registering the Plug-in.	12
 Chapter 3: Amazon S3 Sources and Targets.....	 13
Amazon S3 Sources and Targets Overview.	13
Import Amazon S3 Objects.	13
 Chapter 4: Amazon S3 Sessions.....	 15
Amazon S3 Sessions Overview.	15
Amazon S3 Connections.	15
Amazon S3 Connection Properties.	15
Configuring an Amazon S3 Connection.	17
Configuring the Source Qualifier.	17
Amazon S3 Source Sessions.	17
Amazon S3 Target Sessions.	17
Data Encryption in Amazon S3 Targets.	18
Success Files.	18
Error Files.	18

Amazon S3 Target Session Configuration.	19
Appendix A: Data Type Reference.....	20
Data Type Reference Overview.	20
Index.	21

Preface

The *Informatica PowerExchange® for Amazon S3 User Guide for PowerCenter®* describes how to read data from and write data to Amazon S3. The guide is written for database administrators and developers who are responsible for moving delimited file data from a source to an Amazon S3 target, and from an Amazon S3 source to a target. This guide assumes that you have knowledge of database engines, Amazon S3, 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.

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 PowerExchange for Amazon S3

This chapter includes the following topics:

- [PowerExchange for Amazon S3 Overview, 8](#)
- [PowerCenter Integration Service and Amazon S3 Integration, 8](#)
- [Introduction to Amazon S3, 9](#)

PowerExchange for Amazon S3 Overview

You can use PowerExchange for Amazon S3 to connect PowerCenter and Amazon S3.

Amazon S3 is a cloud-based store that stores many objects in one or more buckets.

You can also connect to Amazon S3 buckets available in Virtual Private Cloud (VPC) through VPC endpoints. Use PowerExchange for Amazon S3 to read delimited file data from and write delimited file data to Amazon S3. You can use Amazon S3 objects as sources and targets in mappings. When you use Amazon S3 objects in mappings, you must configure properties specific to Amazon S3. When you write to Amazon S3, you can enable data encryption to protect data.

Example

You are a medical data analyst in a medical and pharmaceutical organization who maintains patient records. A patient record can contain patient details, doctor details, treatment history, and insurance from multiple data sources.

You use PowerExchange for Amazon S3 to collate and organize the patient details from multiple input sources and write the data to Amazon S3.

PowerCenter Integration Service and Amazon S3 Integration

The PowerCenter Integration Service uses the Amazon S3 connection to connect to Amazon S3.

When you run an Amazon S3 session, the PowerCenter Integration Service reads data from Amazon S3 based on the session and Amazon S3 connection configuration. The PowerCenter Integration Service connects and

reads data from Amazon S3 through a TCP/IP network. The PowerCenter Integration Service then stores data in a staging directory on the PowerCenter Integration Service machine and writes to any target.

When you run the Amazon S3 session, the PowerCenter Integration Service writes data to Amazon S3 based on the session and Amazon S3 connection configuration. The PowerCenter Integration Service reads from any source and stores data in a staging directory on the PowerCenter Integration Service machine. The PowerCenter Integration Service then connects and writes data to Amazon S3 through a TCP/IP network.

Introduction to Amazon S3

Amazon Simple Storage Service (Amazon S3) is storage service in which you can copy data from source and simultaneously move data to any target. You can use Amazon S3 to store and retrieve any amount of data at any time, from anywhere on the web. You can accomplish these tasks using the AWS Management Console web interface.

Amazon S3 stores data as objects within buckets. An object consists of a file and optionally any metadata that describes that file. To store an object in Amazon S3, you upload the file you want to store to a bucket. Buckets are the containers for objects. You can have one or more buckets. When using the AWS Management Console, you can create folders to group objects, and you can nest folders.

Amazon S3 Objects

PowerExchange for Amazon S3 sources and targets represent delimited file data objects that are read from or written to Amazon S3 buckets as CSV files.

Use PowerExchange for Amazon S3 to read delimited files from Amazon S3 and to insert data to delimited files in Amazon S3 buckets.

Amazon S3 Object Format

Amazon S3 objects are CSV or delimited text files. All fields in an Amazon S3 file are of string data type with a data format that you cannot change and with a defined precision of 256. Data in Amazon S3 text files is written in String 256 format.

PowerExchange for Amazon S3 accepts target data with a precision greater than 256. You do not need to change the precision in the Target transformation.

To read source data with a precision greater than 256, increase the precision in the Source transformation to view the complete data.

To write Amazon S3 source data to any relational target data source, you can specify field expressions in the Fields page. The PowerCenter Integration Service converts the Amazon S3 string data to the target data format.

An Amazon S3 file uses the following data format:

- The delimiter is a colon.
- The qualifier is a double-quote.
- The escape character is a backslash.

CHAPTER 2

PowerExchange for Amazon S3 Configuration

This chapter includes the following topics:

- [PowerExchange for Amazon S3 Configuration Overview, 10](#)
- [Prerequisites, 10](#)
- [Create Minimal Amazon S3 Bucket Policy, 11](#)
- [Registering the Plug-in, 12](#)

PowerExchange for Amazon S3 Configuration Overview

You can use PowerExchange for Amazon S3 on Windows or Linux. You must configure PowerExchange for Amazon S3 before you can extract data from or load data to Amazon S3.

Prerequisites

Before you can use PowerExchange for Amazon S3, perform the following tasks:

1. Install or upgrade to PowerCenter 10.1.1.
2. Create an Access Key ID and Secret Access Key in AWS. You can provide these key values when you create an Amazon S3 connection.
3. To encrypt data inserted in Amazon S3 target objects, enable client-side encryption.
4. Verify that you have read, write, and execute permissions on the following directory: `<Informatica installation directory>/server/bin`.

Create an Access Key ID and Secret Access Key

When you configure an Amazon S3 connection, you need to specify the access key ID and secret key to access the Amazon resources.

1. Log in to Amazon Web Services with your AWS credentials.

2. Navigate to the Security Credentials page.
3. Expand the **Access Keys** section, and click **Create New Access Key**.
4. Click the **Show Access Key** link.
5. Click **Download Key File** and save the file on the machine that hosts the PowerCenter Integration Service.

Enable Client-side Encryption

To enable client-side encryption, you must specify the master symmetric key when you create an Amazon S3 connection. The PowerCenter Integration Service uses the master symmetric key to encrypt the data while uploading the files to Amazon S3 targets.

1. Create a master symmetric key, which is a 256-bit AES encryption key in Base64 format.
2. Update the security policy .jar files on the machine that hosts the PowerCenter Integration Service. Update the `local_policy.jar` and the `US_export_policy.jar` files in the following directory:
`<Informatica installation directory>\java\jre\lib\security`. From the Oracle website, download the .jar files that are supported by the JAVA environment on the machine that hosts the PowerCenter Integration Service.

Create Minimal Amazon S3 Bucket Policy

The minimal Amazon S3 bucket policy ensures PowerExchange for Amazon S3QuickSight Connector performs read and write operations successfully.

You can restrict user operations and user access to particular Amazon S3 buckets by assigning an AWS Identity and Access Management (IAM) policy to users. Configure the IAM policy through the AWS console. Following are the minimum required permissions for users to successfully read data from and write data to Amazon S3 bucket.

- PutObject
- GetObject
- GetObjectVersion
- DeleteObject
- DeleteObjectVersion
- ListBucket
- GetBucketPolicy

Sample Policy:

```
{
  "Version": "2012-10-17", "Statement": [
    {
      "Effect": "Allow", "Action": [ "s3:PutObject", "s3:GetObject", "s3:GetObjectVersion",
        "s3:DeleteObject", "s3:DeleteObjectVersion", "s3:ListBucket", "s3:GetBucketPolicy" ],
      "Resource": [ "arn:aws:s3:::<specify_bucket_name>/*", "arn:aws:s3:::<specify_bucket_name>/
        *" ] }
  ]
}
```

Registering the Plug-in

After you install or upgrade to PowerCenter 10.1.1, you must register the plug-in for PowerExchange for Amazon S3 with the PowerCenter repository.

A plug-in is an XML file that defines the functionality of PowerExchange for Amazon S3. To register the plug-in, the repository must be running in exclusive mode. Use the Administrator tool or the pmrep RegisterPlugin command to register the plug-in.

The plug-in file for PowerExchange for Amazon S3 is `AmazonS3Plugin.xml`. When you install PowerExchange for Amazon S3, the installer copies the `AmazonS3Plugin.xml` file to the following directory: `<Informatica Installation Directory>\server\bin\Plugin`.

Note: If you do not have the correct privileges to register the plug-in, contact the user who manages the PowerCenter Repository Service.

CHAPTER 3

Amazon S3 Sources and Targets

This chapter includes the following topics:

- [Amazon S3 Sources and Targets Overview, 13](#)
- [Import Amazon S3 Objects, 13](#)

Amazon S3 Sources and Targets Overview

Create a mapping with an Amazon S3 source to read delimited file data from Amazon S3 and write to a target. Create a mapping with any source and an Amazon S3 target to write delimited file data to Amazon S3.

Import Amazon S3 Objects

You can import Amazon S3 source and target objects before you create a mapping.

Ensure that you have valid AWS credentials before you create a connection.

1. Start the PowerCenter Designer and connect to a PowerCenter repository.
2. Open a source or target folder.
3. Select **Source Analyzer** or **Target Designer**.
4. Click **Sources** or **Targets**, and then click **Import from AmazonS3Plugin**.
The **Establish Connection** dialog box appears.
5. Specify the following information and click **Connect**.

Connection Property	Description
Access Key	The access key ID used to access the Amazon account resources. Required if you do not use AWS Identity and Access Management (IAM) authentication.
Secret Key	The secret access key used to access the Amazon account resources. This value is associated with the access key and uniquely identifies the account. You must specify this value if you specify the access key ID.

Connection Property	Description
Folder Path	The complete path to the Amazon S3 objects and must include the bucket name and any folder name. Ensure that you do not use a forward slash at the end of the folder path. For example, <bucket name>/<my folder name>
Master Symmetric Key	Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool. If you specify a value, ensure that you specify the encryption type as client side encryption in the target session properties.
Code Page	The code page compatible with the Amazon S3 source. Select one of the following code pages: <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European data. - UTF-8. Select for Unicode and non-Unicode data. - Shift-JIS. Select for double-byte character data. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1.
Formatting Options	Select a delimiter, text qualifier, or an escape character. Choose Other for the delimiter, if you want to specify a delimiter other than comma, tab, colon, and semi-colon.
Region Name	The name of the region where the Amazon S3 bucket is available. Select one of the following regions: <ul style="list-style-type: none"> - US East (N. Virginia) - US West (N. California) - US West (Oregon) - EU (Ireland) - EU (Frankfurt) - Asia Pacific (Tokyo) - Asia Pacific (Seoul) - Asia Pacific (Singapore) - Asia Pacific (Sudney) - South America (Sao Paulo) Default is US East (N. Virginia).

6. Click **Connect**.
7. Click **Next**.
8. Select the Amazon S3 object that you want to import.
9. Optionally, click **Data Preview** to view the resource metadata.
10. Click **Finish**.

CHAPTER 4

Amazon S3 Sessions

This chapter includes the following topics:

- [Amazon S3 Sessions Overview, 15](#)
- [Amazon S3 Connections, 15](#)
- [Amazon S3 Source Sessions, 17](#)
- [Amazon S3 Target Sessions, 17](#)

Amazon S3 Sessions Overview

You can configure an Amazon S3 connection in the Workflow Manager to read delimited file data from or write delimited file data to an Amazon S3. Ensure that you have write access to the Amazon S3 bucket you want to access.

When you write to Amazon S3 targets, you can only insert data to Amazon S3 targets. You cannot update or delete data. Any data in the target is overwritten when you select an existing Amazon S3 target. To protect data, you can also enable data encryption before writing data to Amazon S3 targets.

Amazon S3 Connections

Amazon S3 connections enable you to read data from or write data to Amazon S3. The PowerCenter Integration Service uses the connection when you run an Amazon S3 session.

Amazon S3 Connection Properties

When you configure an Amazon S3 connection, you define the connection attributes that the PowerCenter Integration Service uses to connect to Amazon S3.

The following table describes the Amazon S3 connection properties:

Connection Property	Description
Name	The name of the Amazon S3 connection.
Type	The AmazonS3 connection type.
Access Key	The access key ID used to access the Amazon account resources. Required if you do not use AWS Identity and Access Management (IAM) authentication. Note: Ensure that you have valid AWS credentials before you create a connection.
Secret Key	The secret access key used to access the Amazon account resources. This value is associated with the access key and uniquely identifies the account. You must specify this value if you specify the access key ID.
Folder Path	The complete path to the Amazon S3 objects and must include the bucket name and any folder name. Ensure that you do not use a forward slash at the end of the folder path. For example, <bucket name>/<my folder name>
Master Symmetric Key	Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool. If you specify a value, ensure that you specify the encryption type as client side encryption in the target session properties.
Code Page	The code page compatible with the Amazon S3 source. Select one of the following code pages: <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European data. - UTF-8. Select for Unicode and non-Unicode data. - Shift-JIS. Select for double-byte character data. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1.
Region Name	The name of the region where the Amazon S3 bucket is available. Select one of the following regions: <ul style="list-style-type: none"> - US East (N. Virginia) - US West (N. California) - US West (Oregon) - EU (Ireland) - EU (Frankfurt) - Asia Pacific (Tokyo) - Asia Pacific (Seoul) - Asia Pacific (Singapore) - Asia Pacific (Sudney) - South America (Sao Paulo) Default is US East (N. Virginia).
Formatting Options	Select a delimiter, text qualifier, or an escape character. Choose Other for the delimiter, if you want to specify a delimiter other than comma, tab, colon, and semi-colon.

Configuring an Amazon S3 Connection

Configure an Amazon S3 connection in the Workflow Manager to define the connection attributes that the PowerCenter Integration Services uses to connect to Amazon S3.

1. In the Workflow Manager, click **Connections > Application**.
The **Application Connection Browser** dialog box appears.
2. Click **New**.
The **Select Subtype** dialog box appears.
3. Select **AmazonS3** and click **OK**.
The **Connection Object Definition** dialog box appears.
4. Enter a name for the Amazon S3 connection.
5. Enter the Amazon S3 connection properties.
6. Click **OK**.

Configuring the Source Qualifier

After you import a source to create a Mapping for Amazon S3 source, you must configure the source qualifier.

1. In a mapping, double-click the Source Qualifier.
2. Select the **Configure** tab and click **Configure**.
The Establish Connection dialog box appears.
3. Specify the Amazon S3 connection properties and click **Connect**.
4. Click **Finish**.
5. Save the mapping.

Amazon S3 Source Sessions

Create a mapping with an Amazon S3 source and a target to read data from Amazon S3. If the file size of an Amazon S3 object is greater than 8 MB, you can enable the **Enable Downloading S3 Files in Multiple Parts** option to download the object in multiple parts in parallel.

For Amazon S3 sources, you can set the tracing level session property, which sets the amount of detail that appears in the log file. You can choose terse, normal, verbose initialization, or verbose data. Default is normal.

Amazon S3 Target Sessions

Create a session and associate it with a mapping that you created to write data to Amazon S3. Define the session properties to write data to Amazon S3.

Data Encryption in Amazon S3 Targets

To protect data, you can enable server-side encryption or client-side encryption to encrypt data inserted in Amazon S3 buckets.

Server-side Encryption

Enable server-side encryption if you want Amazon S3 to encrypt the data while uploading the CSV files to the buckets. To enable server-side encryption, select Server Side Encryption as the encryption type in the target session properties.

Client-side Encryption

Enable client-side encryption if you want the PowerCenter Integration Service to encrypt the data while uploading the CSV files to the buckets. To enable client-side encryption, perform the following tasks:

1. Provide a master symmetric key when you create an Amazon S3 connection. Ensure that you provide a 256-bit AES encryption key in Base64 format.
2. Select Client Side Encryption as the encryption type in the target session properties.
3. Ensure that an organization administrator updates the security policy .jar files on the machines that host the PowerCenter Integration Service.

Success Files

The PowerCenter Integration Service generates a success file after you run a session.

The success file contains an entry for each record that is successfully written to Amazon S3. Each entry contains the values that are written for all the fields of the record. Use the success file to understand the data that the PowerCenter Integration Service writes to the Amazon S3.

If you want the PowerCenter Integration Service to generate a success file, specify a directory for the success file in the Amazon S3 target session properties. Specify a directory on the machine that hosts the PowerCenter Integration Service. If you do not specify a directory for the success file, the PowerCenter Integration Service creates the success file in the following directory: `$PMTargetFileDir`.

The PowerCenter Integration Service generates the success file with the following naming convention:

```
<sessionName>_<timestamp>_success.csv.
```

Error Files

If errors occur when you run a session, the PowerCenter Integration Service generates an error file.

The error file contains an entry for each data error that occurs when you write to Amazon S3 target. Each entry in the file contains the values for all fields of the record and the error message. Use the error file to understand why the PowerCenter Integration Service did not write data to the Amazon S3 target.

If you want the PowerCenter Integration Service to generate an error file, specify a directory for the error file in the Amazon S3 target session properties. Specify a directory on the machine that hosts the PowerCenter Integration Service. If you do not specify a directory for the error file as a session property, the PowerCenter Integration Service creates the error file in the following directory: `$PMBadFileDir`.

The PowerCenter Integration Service generates an errors file with the following naming convention:

```
<sessionName>_<timestamp>_error.csv.
```

Amazon S3 Target Session Configuration

You can configure a session to write data to Amazon S3. Define the properties for each target instance in the session.

The following table describes the session properties:

Property	Description
Encryption Type	Method you want to use to encrypt data. Select one of the following values: <ul style="list-style-type: none">- None. The data is not encrypted.- Client Side Encryption. The PowerCenter Integration Service encrypts data while uploading the CSV files to Amazon buckets. You must select client-side encryption if you specify a master symmetric key in the Amazon S3 connection properties.- Server Side Encryption. Amazon S3 encrypts data while uploading the CSV files to Amazon buckets.
INSERT	Inserts the source data in the Amazon S3 target. Overwrites any existing data in the target object. Note: You can only insert data to Amazon S3 objects. You cannot perform delete or update operations on Amazon S3 targets.
Success File Directory	Directory for the Amazon S3 success file. Specify a directory on the machine that hosts the PowerCenter Integration Service.
Error File Directory	Directory for the Amazon S3 error file. Specify a directory on the machine that hosts the PowerCenter Integration Service.

APPENDIX A

Data Type Reference

This appendix includes the following topic:

- [Data Type Reference Overview, 20](#)

Data Type Reference Overview

PowerExchange for Amazon S3 uses only CSV files in PowerCenter sessions.

PowerExchange for Amazon S3 uses the following data types in PowerCenter sessions with Amazon S3 objects:

Amazon S3 native data types

Amazon S3 data types appear on the Datatype tab for source qualifiers and target definitions when you edit metadata for the fields.

Transformation data types

Set of data types that appear in the remaining transformations. They are internal data types based on ANSI SQL-92 generic data types, which PowerCenter uses to move data across platforms.

Transformation data types appear in all remaining transformations in a PowerCenter sessions.

When PowerExchange for Amazon S3 reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When PowerExchange for Amazon S3 writes to a target, it converts the transformation data types to the comparable native data types.

The following table lists the Amazon S3 data types that PowerExchange for Amazon S3 supports and the corresponding transformation data types:

Amazon S3 Native Data Type	Transformation Data Type	Description
String	String	1 to 104,857,600 characters

INDEX

A

- access key
 - creating [10](#)
- administration
 - minimal Amazon S3 bucket policy [11](#)
- Amazon S3
 - connections [15](#)
 - integration with PowerCenter [8](#)
 - object format [9](#)
 - objects [9](#)
 - overview [9](#)
- Amazon S3 connection
 - configuring [17](#)
 - properties [15](#)
- Amazon S3 objects
 - importing [13](#)
- Amazon S3 target
 - error file directory [18](#)
 - session configuration [19](#)
 - success file directory [18](#)

C

- client-side encryption
 - enabling [11](#)

D

- data encryption
 - client-side [18](#)
 - server-side [18](#)
- data type reference
 - overview [20](#)

P

- PowerExchange for Amazon S3
 - overview [8](#)

S

- secret access key
 - creating [10](#)
- source qualifier
 - configuring [17](#)