Informatica® PowerExchange for Amazon S3 10.1

User Guide

Informatica PowerExchange for Amazon S3 User Guide
10.1
September 2016

© Copyright Informatica LLC 2016, 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: 2018-09-27

Table of Contents

Preface

The *PowerExchange for Amazon S3 Guide* contains information about how to set up and use PowerExchange for Amazon S3. The guide explains how organization administrators and business users can use PowerExchange for Amazon S3 to read from and write data to Amazon S3.

This guide assumes that you have knowledge of Amazon S3 and Informatica Data Services.

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

You can use PowerExchange for Amazon S3 to read and write delimited flat file data and binary files as pass-through data from and to Amazon S3 buckets.

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

Create an Amazon S3 connection to specify the location of Amazon S3 sources and targets you want to include in a data object. You can use the Amazon S3 connection in data object read and write operations.

You can validate and run mappings in the native environment or Blaze mode.

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 in Amazon S3 buckets.

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.

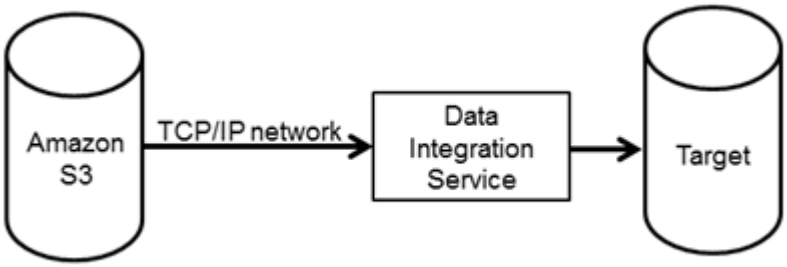
Buckets are the containers for objects. You can have one or more buckets. When using the AWS Management Console, you can create folders to group objects, and you can nest folders.

Data Integration Service and Amazon S3 Integration

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

Reading Amazon S3 Data

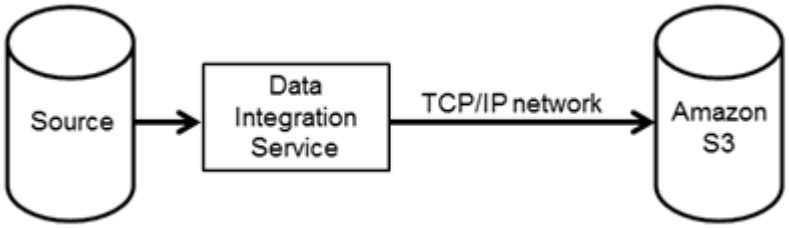
The following image shows how Informatica connects to Amazon S3 to read data:



When you run the Amazon S3 session, the Data Integration Service reads data from Amazon S3 based on the workflow and Amazon S3 connection configuration. The Data Integration Service connects and reads data from Amazon Simple Storage Service (Amazon S3) through a TCP/IP network. The Data Integration Service then stores data in a staging directory on the Data Integration Service host. Amazon S3 is a storage service in which you can copy data from source and simultaneously move data to any target. The Data Integration Service issues a copy command that copies data from Amazon S3 to the target.

Writing Amazon S3 Data

The following image shows how Informatica connects to Amazon S3 to write data:



When you run the Amazon S3 session, the Data Integration Service writes data to Amazon S3 based on the workflow and Amazon S3 connection configuration. The Data Integration Service stores data in a staging directory on the Data Integration Service host. The Data Integration Service then connects and writes data to Amazon Simple Storage Service (Amazon S3) through a TCP/IP network. Amazon S3 is a storage service in which you can copy data from source and simultaneously move data to Amazon S3 clusters. The Data Integration Service issues a copy command that copies data from Amazon S3 to the Amazon S3 target table.

CHAPTER 2

PowerExchange for Amazon S3 Installation and Configuration

This chapter includes the following topics:

PowerExchange for Amazon S3 Installation and Configuration Overview

The PowerExchange for Amazon S3 installation consists of a server installation and a client installation. You can install PowerExchange for Amazon S3 on Windows or UNIX.

To install and configure PowerExchange for Amazon S3, perform the following steps:

1. Install or upgrade Informatica services. Create and configure a Model Repository Service and a Data Integration Service.
2. Install the PowerExchange for Amazon S3 server component after you install the Informatica services. The server binaries are copied to the Informatica server installation directory.
3. Install the PowerExchange for Amazon S3 client component after you install the Informatica clients. The client binaries are copied to the Informatica client installation directory.

To run Amazon S3 mappings in Blaze mode, follow these additional steps:

1. Install Informatica Big Data Management on every node of the Hadoop cluster and verify that you have completed the prerequisite tasks. For information, see the *Informatica Big Data Management Installation and Configuration Guide.*.
2. Install EBF 17588 on top of Informatica Big Data Management. For information, see the *Informatica Big Data Management EBF 17588 Release Notes*.
3. Install the Amazon S3 component on every node of the Hadoop cluster.

Prerequisites

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

1. Complete the installation and configuration process for PowerExchange for Amazon S3.
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. Verify that you have write permissions on all the directories within the `<INFA_HOME>`directory.

Installing the Server Component

You can install the server component on Windows or Linux machines.

Install the PowerExchange for Amazon S3 server component on all the nodes in the domain.

Installing the Server Component on Windows

If multiple nodes exist in your environment, you must first install the server component on the master gateway node. You can then install the server component on the other nodes in the domain.

Before you install, shut down the Informatica domain.

1. Delete the contents from the following directories:
   - `<INFA_HOME>\services\work_dir`
   - `<INFA_HOME>\tomcat\bin\workspace`
2. Navigate to the root directory of the extracted installer files.
3. Run the `install.bat` script file.

   The **Welcome** page appears.
4. Click **Next**.

   The **Installation Directory** page appears.
5. Enter the absolute path to the Informatica installation directory. Click **Browse** to find the directory or use the default directory.

   By default, the server components are installed in the following location:

   `C:\Informatica\<version folder>\`

   If you did not shut down the domain, a message appears asking you to shut down the domain.
6. Click **Next**.

   The **Pre-Installation Summary** page appears.
7. Verify that all installation requirements are met and click **Install**.

   The **Domain Information Panel** page appears.

8. View or enter the domain information.

Property	Description
Domain Name	Name of the domain where Informatica services are installed. This field is read-only.
Node Name	Name of the node on which you are installing the PowerExchange for Amazon S3 server component. This field is read-only.
Domain User Name	User name of the administrator for the domain.
Domain Password	Password for the domain administrator.
Master Gateway Node	Indicates whether the node on which you are installing the server component is the master gateway node. Select the option for the master gateway node. Clear the option for all other nodes on which you install the server component.

9. Click **Next**.

   The installer shows the progress of the installation. When the installation is complete, the **Post-Installation Summary** page displays the status of the installation.

10. Click **Done** to close the installer.

    For more information about the tasks performed by the installer, view the installation log files.

Installing the Server Component on Linux

If multiple nodes exist in your environment, you must first install the server component on the master gateway node. You can then install the server component on the other nodes in the domain.

1. Delete the contents from the following directories:
   - `$INFA_HOME/services/work_dir`
   - `$INFA_HOME/tomcat/bin/workspace`
2. Navigate to the root directory of the extracted installer files.
3. Enter `./install.sh` at the command prompt.

   Note: The `install.sh` file must have executable permissions.
4. Enter the path to the Informatica installation directory.

   By default, the server components are installed in the following location:

   `<User Home Directory>/Informatica/<version folder>`

   If you did not shut down the domain, a message appears asking you to shut down the domain.
5. Review the installation information and press **Enter** to begin the installation.

6. View or enter the domain information.

Property	Description
Domain Name	Name of the domain where Informatica services are installed. This field is read-only.
Node Name	Name of the node on which you are installing the PowerExchange for Amazon S3 server component. This field is read-only.
Domain User Name	User name of the administrator for the domain.
Domain Password	Password for the domain administrator.
Master Gateway Node	Indicates whether the node on which you are installing the server component is the master gateway node. Select from the following options: 1. Yes. Select Yes if the node is the master gateway node. 2. No. Select No for all other nodes on which you install the server component.

For more information about the tasks performed by the installer, view the installation log files.

Installing the Client Component

Install the client component on every Informatica Developer client machine that connects to the domain.

1. Delete the contents from the following directory:

   `<INFA_HOME>\clients\DeveloperClient\workspace`
2. Delete the configuration files and retain the `config.ini` file from the following directory:

   `<INFA_HOME>\clients\DeveloperClient\configuration`
3. Unzip the client installation archive and navigate to the root directory of the extracted installer files.
4. Run the `install.bat` script file.

   The **Welcome** page appears.
5. Click **Next**.

   The **Installation Directory** page appears.
6. Enter the absolute path to the Informatica installation directory. Click **Browse** to find the directory or use the default directory.
7. Click **Next**.

   The **Pre-Installation Summary** page appears.
8. Verify that all installation requirements are met and click **Install**.

   The installer shows the progress of the installation. When the installation is complete, the **Post-Installation Summary** page displays the status of the installation.
9. Click **Done** to close the installer.

   For more information about the tasks performed by the installer, view the installation log files.

Configuring the Amazon S3 Component on Hadoop

Install the Amazon S3 component on every node of the Hadoop cluster.

1. Download the Amazon S3 component file, `EBFAmazon-S3_HadoopRPM_EBFInstaller.tar.z`.
2. Log in as an user with sudo privileges on the Hadoop cluster node.
3. Use a zip utility to extract the Amazon S3 component file.
4. Use a text editor to open the `input.properties` file and change the path to the Informatica installation directory.
5. Use a text editor to open the `HadoopDataNodes` file and list the host names of all the nodes in the Hadoop cluster on which an user can run the Amazon S3 mappings.
6. Run the `InformaticaHadoopEBFInstall` script file.
7. If you agree to the terms of the end user license agreement, specify Yes. If you specify No, you cannot continue with the installation process.
8. Based on your requirement, choose one of the following options:
   - Enter 1 to install the component on a single node on the local machine.
   - Enter 2 to install the component on multiple remote nodes of the Hadoop cluster.
   - Enter 3 to exit from the installation process.

PowerExchange for Amazon S3 Administration

As a user, you can use PowerExchange for Amazon S3 after an organization administrator performs the following tasks:

- Create an Access Key ID and Secret Access Key in AWS. Provide these key values when you create an Amazon S3 connection.
- Enable client-side encryption if you want to encrypt the data while uploading files to Amazon S3 buckets.
- Optionally, create a minimal Amazon S3 bucket policy for PowerExchange for Amazon S3.

Create an Access Key ID and Secret Access Key

1. Log in to Amazon Web Services and navigate to the Security Credentials page.
2. Expand the **Access Keys** section, and click **Create New Access Key**.
3. Click the **Show Access Key** link.
4. Click **Download Key File** and save the file on the Secure Agent machine.

Enable Client-side Encryption

An organization administrator must perform the following tasks to enable client-side encryption:

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 Data 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`.

   You can download the `.jar` files supported by the Java environment on the machine that hosts the Data Integration Service from the Oracle website.

Create Minimal Amazon S3 Bucket Policy

You can create a minimal Amazon S3 bucket policy to ensure that PowerExchange for Amazon S3 successfully reads and writes data from and to Amazon S3.

To restrict the user operations and user access to specific Amazon S3 buckets, assign an AWS Identity and Access Management (IAM) policy to users. Configure the IAM policy through the AWS console. To successfully read data from and write data to Amazon S3, users need the following permissions:

- 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>/*" ] }
]
}
```

CHAPTER 3

Amazon S3 Connections

This chapter includes the following topics:

Amazon S3 Connections Overview

Amazon S3 connections enable you to read data from or write data to Amazon S3.

When you create an Amazon S3 connection, you define connection attributes. You can create an Amazon S3 connection in the Developer tool or the Administrator tool. The Developer tool stores connections in the domain configuration repository. Create and manage connections in the connection preferences.

The Developer tool uses the connection when you create data objects. The Data Integration Service uses the connection when you run mappings.

Amazon S3 Connection Properties

When you set up an Amazon S3 connection, you must configure the connection properties.

The following table describes the Amazon S3 connection properties:

Property	Description
Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:~ ` ! $ % ^ & * ( ) - + = { [}] | \ : ; " ' < , > . ? /
ID	String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name.
Description	Optional. The description of the connection. The description cannot exceed 4,000 characters.

Property	Description
Type	The Amazon S3 connection type.
Access Key	The access key ID for access to Amazon account resources.
Secret Key	The secret access key for access to Amazon account resources. The secret key is associated with the access key and uniquely identifies the account.
Folder Path	The complete path to Amazon S3 objects. The path must include the bucket name and any folder name. Do not use a 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. **Note:** You can enable Client Side Encryption as the encryption type in the advanced properties of the data object write operation.
Region Name	Select the AWS region in which the bucket you want to access resides.

Creating an Amazon S3 Connection

Create an Amazon S3 connection before you create an Amazon S3 data object.

1. In the Developer tool, click **Window** > **Preferences**.
2. Select **Informatica** > **Connections**.
3. Expand the domain in the **Available Connections**.
4. Select the connection type **Enterprise Application** > **Amazon S3**, and click **Add**.
5. Enter a connection name and an optional description.
6. Select Amazon S3 as the connection type.
7. Click **Next**.
8. Configure the connection properties.
9. Click **Test Connection** to verify the connection to Amazon S3.
10. Click **Finish**.

CHAPTER 4

PowerExchange for Amazon S3 Data Objects

This chapter includes the following topics:

Amazon S3 Data Object Overview

An Amazon S3 data object is a physical data object that uses Amazon S3 as a source or target. An Amazon S3 data object is the physical data object that represents data based on an Amazon S3 resource.

You can configure the data object read and write operation properties that determine how data can be read from Amazon S3 sources and loaded to Amazon S3 targets.

Create an Amazon S3 data object from the Developer tool. Create a data object read operation or data object write operation for the Amazon S3 data object. You can then add the data object read or write operation to a mapping.

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 files to the buckets. To enable server-side encryption, select **Server Side Encryption** as the encryption type in the advanced properties of the data object write operation.

Client-side Encryption

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

1. Ensure that an organization administrator creates a master symmetric key, which is a 256-bit AES encryption key in Base64 format.
2. Provide the master symmetric key when you create an Amazon S3 connection.
3. Select **Client Side Encryption** as the encryption type in the advanced properties of the data object write operation.
4. Ensure that an organization administrator updates the security JAR files, required by the Amazon S3 client encryption policy, on the machine that hosts the Data Integration Service.
   For information about the Amazon S3 client encryption policy, see the Amazon S3 documentation.

Amazon S3 Data Object Properties

Specify the data object properties when you create the data object.
The following table describes the properties that you configure for the Amazon S3 data objects:

Property	Description
Name	Name of the Amazon S3 data object.
Location	The project or folder in the Model Repository Service where you want to store the Amazon S3 data object.
Connection	Name of the Amazon S3 connection.

Amazon S3 Data Object Read Operation Properties

Amazon S3 data object read operation properties include run-time properties that apply to the Amazon S3 data object.

The Developer tool displays advanced properties for the Amazon S3 data object operation in the Advanced view. The following table describes the Advanced properties for an Amazon S3 data object read operation:

Property	Description
File Name	Name of the Amazon S3 resource file from which you want to read data.
S3 Folder Path Name	Bucket name that contains the Amazon S3 source file. If applicable, include the folder name that contains the source file in the `bucket_name>/<folder_name>` format.
Enable Download S3 Files in Multiple Parts	Download large Amazon S3 objects in multiple parts. When the file size of an Amazon S3 object is greater than 8 MB, you can choose to download the object in multiple parts in parallel.
Local Temp Folder Path	Amazon S3 staging directory. Ensure that the user has write permissions on the directory. In addition, ensure that there is sufficient space to enable staging of the entire file. Default staging directory is the `/temp` directory on the machine that hosts the Data Integration Service.

Amazon S3 Data Object Write Operation Properties

Amazon S3 data object write operation properties include run-time properties that apply to the Amazon S3 data object.

The Developer tool displays advanced properties for the Amazon S3 data object operation in the Advanced view. The following table describes the Advanced properties for an Amazon S3 data object write operation:

Property	Description
File Name	Name of the Amazon S3 resource file to which you want to write the source data.
S3 Folder Path Name	Bucket name that contains the target file. If applicable, include the folder name that contains the target file in the `bucket_name>/<folder_name>` format.
Encryption Type	Method you want to use to encrypt data. Select one of the following values:- - None. The data is not encrypted. - Client Side Encryption. The Data Integration Service uses the master symmetric key you specify in the Amazon S3 connection properties to encrypt data. - 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 files to Amazon buckets.
Local temp folder path	Amazon S3 staging directory. Ensure that the user has write permissions on the directory. In addition, ensure that there is sufficient space to enable staging of the entire file. Default staging directory is the `/temp` directory on the machine that hosts the Data Integration Service.

Creating an Amazon S3 Data Object

Creata an Amazon S3 data object to add to a mapping.

1. Select a project or folder in the **Object Explorer** view.
2. Click **File** > **New** > **Data Object**.
3. Select **Amazon S3 Data Object** and click **Next**.

   The **Amazon S3 Data Object** dialog box appears.
4. Enter a name for the data object.
5. Click **Browse** next to the **Location** option and select the target project or folder.
6. Click **Browse** next to the **Connection** option and select the Amazon S3 connection from which you want to import the Amazon S3 object.
7. To add a resource, click **Add** next to the **Selected Resources** option.

   The **Add Resource** dialog box appears.
8. Select the check box next to the Amazon S3 object you want to add and click **OK**.
9. Click **Finish**.
   The data object appears under Data Objects in the project or folder in the **Object Explorer** view.

Creating a Data Object Operation

You can create the data object read or write operation for an Amaazon S3 data object. You can then add the Amazon S3 data object operation to a mapping.

1. Select the data object in the **Object Explorer** view.
2. Right-click and select **New** > **Data Object Operation**.

   The **Data Object Operation** dialog box appears.
3. Enter a name for the data object operation.
4. Select the type of data object operation. You can choose to create a read or write operation.
5. Click **Add**.

   The **Select Resources** dialog box appears.
6. Select the Amazon S3 data object for which you want to create the data object operation and click **OK**.
7. Click **Finish**.

The Developer tool creates the data object operation for the selected data object.

Configuring Column Projection

After you create a data object operation, you can project the columns as a binary data type, sample the metadata of an Amazon S3 file and project the columns, or manually project the columns.

Projecting Binary Columns

Perform the following steps to project columns as a binary data type:

1. Go to **Column Projection** tab.
2. Clear the **Enable Column Projection** field.

   The columns appear as a binary data type.

Sampling Metadata

Perform the following steps to sample metadata file and project columns:

1. Go to **Column Projection** tab.
2. Click **Edit Column Projection**.
3. Select **Reconfigure**.

   The **Column Projection** page appears.
4. Choose **Sample Metadata File**.

   You can click **Browse** and navigate to the directory that contains the file.
5. Select a code page in **Code page** field.

   The page matches the code page of the data that you want to process.

   Note: The **Delimited** and **Fixed-width** format properties are not applicable for PowerExchange for Amazon S3.
6. Click **Next**.
7. Configure the format properties.

Property	Description
Delimiters	Character used to separate columns of data. If you enter a delimiter that is the same as the escape character or the text qualifier, you might receive unexpected results. Amazon S3 reader and writer support Delimiters.
Text Qualifier	Quote character that defines the boundaries of text strings. If you select a quote character, the Developer tool ignores delimiters within pairs of quotes. Amazon S3 reader supports Text Qualifier.
Import Column Names From First Line	If selected, the Developer tool uses data in the first row for column names. Select this option if column names appear in the first row. The Developer tool prefixes"FIELD_" to field names that are not valid. Amazon S3 reader and writer support Import Column Names From First Line.
Escape Character	Character immediately preceding a column delimiter character embedded in an unquoted string, or immediately preceding the quote character in a quoted string. When you specify an escape character, the Data Integration Service reads the delimiter character as a regular character.

   Note: The **Start import at line**, **Treat consecutive delimters as one**, and **Retain escape character in data** properties in the **Column Projection** dialog box are not applicable for PowerExchange for Amazon S3.
8. Click **Next** to preview the flat file data object.

   You must change the data types to string manually.
9. Click **Finish**.

Projecting Columns Manually

After sampling the metadata, you can manually edit the projected columns.

Perform the following steps to project columns manually:

1. Go to **Column Projection** tab.
2. Click **Edit Column Projection**.
3. Click **New** icon and add fields manually.

CHAPTER 5

PowerExchange for Amazon S3 Mappings

This chapter includes the following topics:

PowerExchange for Amazon S3 Mappings Overview

After you create an Amazon S3 data object read or write operation, you can create a mapping.

You can create an Informatica mapping containing an Amazon S3 data object read operation as the input, and a relational or flat file data object operation as the target. You can create an Informatica mapping containing objects such as a relational or flat file data object operation as the input, transformations, and an Amazon S3 data object write operation as the output to load data to Amazon S3 buckets.

Validate and run the mapping. You can deploy the mapping and run it or add the mapping to a Mapping task in a workflow.

Mapping Validation and Run-time Environments

You can validate and run mappings in the native environment or in Blaze mode.

The Data Integration Service validates whether the mapping can run in the selected environment. You must validate the mapping for an environment before you run the mapping in that environment.

Native environment

You can configure the mappings to run in the native or Hadoop run-time environment. When you run mappings in the native environment, the Data Integration Service processes the mapping and runs the mapping from the Developer tool.

Blaze Mode

When you run mappings in the Blaze mode, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on a Blaze engine. The Data Integration Service generates an execution plan to run mappings on the Blaze engine.

The Blaze engine execution plan simplifies the mapping into segments. The plan contains tasks to start the mapping, run the mapping, and create and cleanup the temporary tables and file required to run the mapping. The plan contains multiple tasklets and the task recovery strategy. The plan also contains pre and post grid task preparation commands for each mapping before running the main mapping on a Hadoop cluster. A pre-grid task can include a task such as copying data to HDFS. A post-grid task can include tasks such as cleaning up temporary files or copying data from HDFS.
You can view the plan in the Developer tool before you run the mapping and in the Administrator tool after you run the mapping. In the Developer tool, the Blaze engine execution plan appears as a workflow. You can click on each component in the workflow to get the details. In the Administrator tool, the Blaze engine execution plan appears as a script.

For more information about the Hadoop environment and Blaze mode, see the *Informatica Big Data Management User Guide*.

APPENDIX A

Amazon S3 Datatype Reference

This appendix includes the following topics:

Datatype Reference Overview

When you run the session to read data from or write data to Amazon S3, the Data Integration Service converts the transformation data types to comparable native Amazon S3 data types.

Amazon S3 and Transformation Datatypes

The following table lists the Amazon S3 data types that the Data Integration Service supports and the corresponding transformation data types:

Amazon S3 Data Type	Transformation Data Type	Description
INT	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
BIGINT	Bigint	Precision of 19 digits, scale of 0
DOUBLE	Double	Precision 15
NUMBER	Decimal	For transformations that support precision up to 28 digits, the precision is 1 to 28 digits, and the scale is 0 to 28. If you specify the precision greater than the maximum number of digits, the Data Integration Service converts decimal values to double in high precision mode.
STRING	String	1 to 104,857,600 characters
NSTRING	String	1 to 104,857,600 characters

INDEX