



Informatica® PowerExchange for Microsoft
Azure Blob Storage

10.2

User Guide

© Copyright Informatica LLC 2016, 2018

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. 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.

Informatica, the Informatica logo, and PowerExchange are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

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.opendap.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/license.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/license.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://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.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, please report them to us in writing at Informatica LLC 2100 Seaport Blvd. Redwood City, CA 94063.

INFORMATICA LLC PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2018-09-27

Table of Contents

Preface	6
Informatica Resources.	6
Informatica Network.	6
Informatica Knowledge Base.	6
Informatica Documentation.	7
Informatica Product Availability Matrixes.	7
Informatica Velocity.	7
Informatica Marketplace.	7
Informatica Global Customer Support.	7
Chapter 1: Introduction to PowerExchange for Microsoft Azure Blob Storage	8
PowerExchange for Microsoft Azure Blob Storage Overview.	8
PowerExchange for Microsoft Azure Blob Storage Example.	8
Introduction to Microsoft Azure Blob Storage.	9
Chapter 2: PowerExchange for Microsoft Azure Blob Storage Installation and Configuration	10
Installation and Configuration Overview.	10
Prerequisites.	10
Installing the Server Component.	11
Installing the Server Component on Linux Master Gateway.	11
Installing the Server Component on Other Linux Nodes (Optional).	11
Installing the Client Component.	12
Java Heap Memory Configuration (Optional).	12
Configure Temporary Directory Location (Optional).	12
Chapter 3: Microsoft Azure Blob Storage Connections	14
Microsoft Azure Blob Storage Connection Overview.	14
Microsoft Azure Blob Storage Connection Properties.	14
Creating a Microsoft Azure Blob Storage Connection.	15
Chapter 4: Microsoft Azure Blob Storage Data Objects	16
Microsoft Azure Blob Storage Data Objects Overview.	16
Microsoft Azure Blob Storage Data Object Properties.	16
Microsoft Azure Blob Storage Data Object Read Operation Properties.	17
Microsoft Azure Blob Storage Data Object Write Operation Properties.	18
Creating a Microsoft Azure Blob Storage Data Object.	18
Creating a Data Object Operation.	20

Chapter 5: Microsoft Azure Blob Storage Mappings.....	21
Microsoft Azure Blob Storage Mappings Overview.	21
Mapping Validation and Run-time Environments.	21
Chapter 6: Data Type Reference.....	22
Data Type Reference Overview.	22
Microsoft Azure Blob Storage and Transformation Data Types.	23
Avro File Data Types and Transformation Data Types.	23
JSON File Data Types and Transformation Data Types.	24
Parquet File Data Types and Transformation Data Types.	24
Index.....	26

Preface

The Informatica PowerExchange® for Microsoft Azure Blob Storage User Guide provides information about reading data from and writing data to Microsoft Azure Blob Storage. The guide is written for database administrators and developers who are responsible for developing mappings that read data from and write data to Microsoft Azure Blob Storage.

The guide assumes you have knowledge of Microsoft Azure Blob Storage and Informatica.

Informatica Resources

Informatica Network

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

As a member, you can:

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

As a member, you can:

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

Informatica Knowledge Base

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

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

Informatica Documentation

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

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

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at

<https://network.informatica.com/community/informatica-network/product-availability-matrixes>.

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 Microsoft Azure Blob Storage

This chapter includes the following topics:

- [PowerExchange for Microsoft Azure Blob Storage Overview, 8](#)
- [PowerExchange for Microsoft Azure Blob Storage Example, 8](#)
- [Introduction to Microsoft Azure Blob Storage, 9](#)

PowerExchange for Microsoft Azure Blob Storage Overview

You can use PowerExchange for Microsoft Azure Blob Storage to connect to Microsoft Azure Blob Storage from Informatica.

Use PowerExchange for Microsoft Azure Blob Storage to read data from or write data to Microsoft Azure Blob Storage. Use PowerExchange for Microsoft Azure Blob Storage to read delimited files, Avro, Parquet, and JSON files from or write to the Microsoft Azure Blob Storage. You can read or write delimited files, Avro, Parquet, and JSON files when you run a mapping on the Spark engine.

You can read and write the gzip compressed `.csv` files both in the native environment and on the Spark engine. Create a Microsoft Azure Blob Storage connection to read or write Microsoft Azure Blob Storage data into a Microsoft Azure Blob Storage data object. When you use Microsoft Azure Blob Storage objects in mappings, you must configure properties specific to Microsoft Azure Blob Storage. You can validate and run mappings in the native environment or in the Spark mode.

PowerExchange for Microsoft Azure Blob Storage Example

You work in sales operations and want to score leads to drive higher sales for your organization. You need to bring in leads from Salesforce to Microsoft Azure Blob Storage. You can score leads for sales readiness in Microsoft Azure Machine Learning, and then load the lead scores back into Salesforce. You can keep data up to date with the latest leads and lead scores by scheduling a workflow to run on a regular basis.

You have leads in Salesforce with data such as the contact information, industry, company size, and marketing information.

You configure a mapping to insert leads from Salesforce to Microsoft Azure Blob Storage. Use Microsoft Azure Machine Learning to score the leads, and then create another mapping to load the lead scores into Salesforce.

You create a workflow so that the tasks run serially promising leads and increase efficiency.

Introduction to Microsoft Azure Blob Storage

Microsoft Azure Blob Storage is a cloud-storage solution that stores unstructured data in the cloud as objects or blobs. Microsoft Azure Blob Storage can store text or binary data of any type, such as a document, media files, or application installer. Microsoft Azure Blob Storage is referred to as object storage.

Blobs are files of any type and size, and are organized into containers in Microsoft Azure Storage. You can access delimited files that are append blobs or block blobs with Microsoft Azure Blob Storage connections.

CHAPTER 2

PowerExchange for Microsoft Azure Blob Storage Installation and Configuration

This chapter includes the following topics:

- [Installation and Configuration Overview, 10](#)
- [Prerequisites, 10](#)
- [Installing the Server Component, 11](#)
- [Installing the Client Component, 12](#)
- [Java Heap Memory Configuration \(Optional\), 12](#)
- [Configure Temporary Directory Location \(Optional\), 12](#)

Installation and Configuration Overview

You can install PowerExchange for Microsoft Azure Blob Storage on Red Hat Enterprise Linux 64-bit machine.

Prerequisites

You must perform the following prerequisites before you can use PowerExchange for Microsoft Azure Blob Storage:

- Install or upgrade to Informatica 10.2.
- Back up the Domain and the Model repository.
- Ensure that the Hadoop cluster configuration for HDInsight is done correctly. Run a FF->FF mapping on the Spark engine to validate the settings.

For more information about product requirements and supported platforms, see the Product Availability Matrix on Informatica

Network:<https://network.informatica.com/community/informatica-network/product-availability-matrices>

Installing the Server Component

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.

Installing the Server Component on Linux Master Gateway

The PowerExchange for Microsoft Azure Blob Storage server component installs the Data Integration Service and Model Repository Service components.

1. Log on to the master gateway node.
2. Shutdown the domain.
3. Install the EBFs in the following order:
 - a. EBF-11099
 - b. EBF-11098

Note: Follow the installation instructions provided with the EBFs.
4. Run the following command from `<INFA_HOME>\isp\bin`.

```
./infasetup.sh validateAndRegisterFeature -ff <INFA_HOME>\plugins\conf\azureblob.xml -up true
```
5. Delete the contents, including the hidden files (.metadata folder) and directories, from the following directories:
 - `$INFA_HOME/services/work_dir`
 - `$INFA_HOME/tomcat/bin/workspace`
6. Restart the domain.

Installing the Server Component on Other Linux Nodes (Optional)

1. Log on to the Linux node.
2. Shutdown the node.
3. Install the EBFs in the following order:
 - a. EBF-11099
 - b. EBF-11098

Note: Follow the installation instructions provided with the EBFs.
4. Run the following command from `<INFA_HOME>\isp\bin`. For example:

```
./infacmd.sh validateFeature -dn <domain_name> -un <username> -pd <password> -ff <INFA_HOME>\plugins\conf\azureblob.xml
```
5. Delete the contents, including the hidden files (.metadata folder) and directories, from the following directories:
 - `$INFA_HOME/services/work_dir`
 - `$INFA_HOME/tomcat/bin/workspace`
6. Restart the node.

Installing the Client Component

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

Ensure that the Informatica client machine has the client version 10.2 running before you install the Microsoft Azure Blob Storage client component.

1. Close the Developer tool.
2. Install the EBFs in the following order:
 - a. EBF-11099
 - b. EBF-11098

Note: Follow the installation instructions provided with the EBFs.

3. Delete the contents from the following directory:
`$INFA_HOME\clients\DeveloperClient\workspace`
4. Delete the configuration files and retain the `config.ini` file from the following directory:
`$INFA_HOME\clients\DeveloperClient\configuration`
5. Edit the `developerCore.ini` file available at `$INFA_HOME\clients\DeveloperClient`. Change the `-DINFA_HADOOP_DIST_DIR` value to `hadoop\HDInsight_3.6` and save the `developerCore.ini` file.
6. Restart the Developer tool.

Java Heap Memory Configuration (Optional)

Configure the memory for the Java heap size in the node that runs the Data Integration Service.

1. In the Administrator tool, navigate to the Data Integration Service for which you want to change the Java heap size.
2. Click the **Processes** tab.
3. Edit **Advance Properties** section.
4. Specify the maximum heap size based on the data you want to process.
5. Click Ok.
6. Restart the Data Integration Service.

Configure Temporary Directory Location (Optional)

Follow below steps to configure the temporary directory location in the node that runs the Data Integration Service.

1. In the Administrator tool, navigate to the Data Integration Service for which you want to change the temporary directory location.
2. Click the **Processes** tab.
3. Click **Custom Properties**. The **Edit Custom Properties** dialog box appears.

4. Click **New** to add a new custom property.
5. Add the JVMOption custom property for the Data Integration Service and specify the value in the following format:
`-Djava.io.tmpdir=<required tmp directory location>`
For example,
`Property Name: JVMOption1`
`Value: -Djava.io.tmpdir=/opt/Informatica/tmp/ZUDAP/`
6. Click **Ok**.
7. Restart the Data Integration Service.

CHAPTER 3

Microsoft Azure Blob Storage Connections

This chapter includes the following topics:

- [Microsoft Azure Blob Storage Connection Overview, 14](#)
- [Microsoft Azure Blob Storage Connection Properties, 14](#)
- [Creating a Microsoft Azure Blob Storage Connection, 15](#)

Microsoft Azure Blob Storage Connection Overview

Microsoft Azure Blob Storage connection enables you to read data from or write data to Microsoft Azure Blob Storage.

You can use Microsoft Azure Blob Storage connections to create data objects and run mappings. The Developer tool uses the connection when you create a data object. The Data Integration Service uses the connection when you run mappings.

You can create an Microsoft Azure Blob Storage connection from 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.

Microsoft Azure Blob Storage Connection Properties

When you set up a Microsoft Azure Blob Storage connection, you must configure the connection properties.

The following table describes the Microsoft Azure Blob Storage connection properties:

Property	Description
Name	Name of the Microsoft Azure Blob Storage connection.
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.

Property	Description
Description	Description of the connection.
Location	The domain where you want to create the connection.
Type	Type of connection. Select AzureBlob.

The **Connection Details** tab contains the connection attributes of the Microsoft Azure Blob Storage connection. The following table describes the connection attributes:

Property	Description
Account Name	Name of the Microsoft Azure Storage account.
Account Key	Microsoft Azure Storage access key.
Container Name	The root container or sub-folders with the absolute path.

Creating a Microsoft Azure Blob Storage Connection

Before you create a Microsoft Azure Blob Storage data object, create a connection in the Developer tool.

1. Click **Window > Preferences**.
2. Select **Informatica > Connections**.
3. Expand the domain in the **Available Connections**.
4. Select the connection type **File System > AzureBlob**, and click **Add**.
5. Enter a connection name and an optional description.
6. Enter an ID for the connection.
7. Select **AzureBlob** as the connection type.
8. Click **Next**.
9. Configure the connection properties.
10. Click **Test Connection** to verify the connection to Microsoft Azure Blob Storage.
11. Click **Finish**.

CHAPTER 4

Microsoft Azure Blob Storage Data Objects

This chapter includes the following topics:

- [Microsoft Azure Blob Storage Data Objects Overview, 16](#)
- [Microsoft Azure Blob Storage Data Object Properties, 16](#)
- [Creating a Microsoft Azure Blob Storage Data Object, 18](#)
- [Creating a Data Object Operation, 20](#)

Microsoft Azure Blob Storage Data Objects Overview

A Microsoft Azure Blob Storage data object is a physical data object that uses Microsoft Azure Blob Storage as a source or target. A Microsoft Azure Blob Storage data object represents the data in a Microsoft Azure Blob Storage file.

You can configure the data object read and write operation properties that determine how data can be read from Microsoft Azure Blob Storage sources and loaded to Microsoft Azure Blob Storage targets. You first create a connection to create a Microsoft Azure Blob Storage data object. When you create a data object, the read and write operations are created by default. You can modify the default read and write operations or create additional operations.

Microsoft Azure Blob Storage 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 Microsoft Azure Blob Storage data objects:

Property	Description
Name	Name of the Microsoft Azure Blob Storage data object.
Location	The project or folder in the Model Repository where you want to store the Microsoft Azure Blob Storage data object.
Connection	Name of the Microsoft Azure Blob Storage connection.

Microsoft Azure Blob Storage Data Object Read Operation Properties

Microsoft Azure Blob Storage data object read operation properties include advanced properties that apply to the Microsoft Azure Blob Storage data object.

The Developer tool displays advanced properties for the Microsoft Azure Blob Storage data object operation in the **Advanced** view.

The following table describes the advanced properties that you can configure for a Microsoft Azure Blob Storage data object read operation:

Property	Description
Number of concurrent connections to Blob Store	The number of concurrent connections to Blob Store to download files. Default is 4.
Source Type	Select the type of source from which you want to read data. You can select the following source types: <ul style="list-style-type: none"> - File - Directory Default is File .
Blob Name Override	Overrides the file name.
Blob Container Override	Overrides the default container name. You can specify the container name or sub-folders in the root container with the absolute path.

Directory Source in Microsoft Azure Blob Storage Sources

You can select the type of source from which you want to read data.

You can select the following type of sources from the **Source Type** option under the advanced properties for a Microsoft Azure Blob Storage data object read operation:

- File
- Directory

Note: The content type of the blob must be `application/x-gzip` to read compressed files for both File and Directory source types.

Use the following rules and guidelines to select **Directory** as the source type:

- All the source files in the directory must contain the same metadata.

- All the files must have data in the same format. For example, delimiters, header fields, and escape characters must be same.
- All the files under a specified directory are parsed. The files under subdirectories are not parsed.
- The connector does not perform any validation if there are multiple blob formats in the directory you select and might result into errors.

Microsoft Azure Blob Storage Data Object Write Operation Properties

Microsoft Azure Blob Storage data object write operation properties include advanced properties that apply to the Microsoft Azure Blob Storage data object.

The Developer tool displays advanced properties for the Microsoft Azure Blob Storage data object operation in the **Advanced** view.

The following table describes the advanced properties that you can configure for a Microsoft Azure Blob Storage data object write operation:

Property	Description
Number of concurrent connections to Blob Store	The number of concurrent connections to Blob Store to upload files. Default is 4.
Blob Name Override	Overrides the file name. Use .gz extension to write compressed blob files to Microsoft Azure Blob Storage.
Blob Container Override	Overrides the default container name. You can specify the container name or sub-folders in the root container with the absolute path.
Compress newly created Blob	Compresses the newly created blob when set to True. Applicable only to delimited files.
Write Strategy	Override: overrides an existing blob. Append: appends to an existing blob. Append operation is applicable only to the append blob type.
Blob Type	Writes data to a block blob or an append blob. You can write data to an append blob only in the native environment.

Creating a Microsoft Azure Blob Storage Data Object

Create a Microsoft Azure Blob Storage 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 **AzureBlob Data Object** and click **Next**.
The AzureBlob Data Object dialog box appears.
4. Enter a name for the data object.

5. In the **Resource Format** list, select any of the following formats:
 - Flat: to read and write delimited resources.
 - Avro: to read and write Avro resources.
 - Json: to read and write JSON resources.
 - Parquet: to read and write Parquet resources.
6. Click **Browse** next to the **Location** option and select the target project or folder.
7. Click **Browse** next to the **Connection** option and select the AzureBlob connection from which you want to import the Microsoft Azure Blob Storage object.
8. To add a resource, click **Add** next to the **Selected Resources** option.
The Add Resource dialog box appears.
9. Select the checkbox next to the Microsoft Azure Blob Storage object you want to add and click **OK**.
You must select the object according to the selected resource format.
10. Click **Finish** if you selected Avro, Json, or Parquet file and skip the remaining steps. Click **Next** if you selected a flat file.
11. Applicable only to the delimited files. Choose **Sample Metadata File** .
You can click Browse and navigate to the directory that contains the file.
12. Click **Next**.
13. Configure the format properties.

Property	Description
Delimiters	Character used to separate columns of data. Make sure that the delimiter you select is not a part of the data. If you enter a delimiter that is the same as the escape character or the text qualifier, you might receive unexpected results. Microsoft Azure Blob Storage reader and writer support Delimiters.
Text Qualifier	Quote character that defines the boundaries of text strings. Make sure that the text qualifier you select is not a part of the data unless it is escaped. If you select a quote character, the Developer tool ignores delimiters within pairs of quotes. Microsoft Azure Blob Storage 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. Do not select this option to read header-less files. Applicable only to the native environment.

14. Click **Next** to configure the column properties and edit the column attributes.
15. Click **Finish**.
The data object appears under Data Objects in the project or folder in the Object Explorer view. The data object read and write operations are created by default when a data object is created.

Creating a Data Object Operation

You can create the data object read or write operation for a Microsoft Azure Blob Storage data objects. You can then add the Microsoft Azure Blob Storage 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 read or write operation.
5. Click **Add**.
The Select Resources dialog box appears.
6. Select the Microsoft Azure Blob Storage 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.

CHAPTER 5

Microsoft Azure Blob Storage Mappings

This chapter includes the following topics:

- [Microsoft Azure Blob Storage Mappings Overview, 21](#)
- [Mapping Validation and Run-time Environments, 21](#)

Microsoft Azure Blob Storage Mappings Overview

After you create a Microsoft Azure Blob Storage data object operation, you can develop a mapping.

You can define the following objects in the mapping:

- Microsoft Azure Blob Storage data object read operation as the input to read data from Microsoft Azure Blob Storage.
- Microsoft Azure Blob Storage data object write operation as the output to write data to Microsoft Azure Blob Storage.

Note: You do not require any additional steps to run PowerExchange for Microsoft Azure Blob Storage in Hadoop Environment.

Mapping Validation and Run-time Environments

You can validate and run mappings in the native environment or Hadoop Environment.

You can validate a mapping in the native environment, Hadoop Environment, or both. 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.

When you run a mapping in the native environment, the Data Integration Service runs the mapping.

When you run a mapping on a Hadoop cluster, you can select the Spark engine. The Data Integration Service pushes the mappings to the selected engine for processing.

CHAPTER 6

Data Type Reference

This chapter includes the following topics:

- [Data Type Reference Overview, 22](#)
- [Microsoft Azure Blob Storage and Transformation Data Types, 23](#)
- [Avro File Data Types and Transformation Data Types, 23](#)
- [JSON File Data Types and Transformation Data Types, 24](#)
- [Parquet File Data Types and Transformation Data Types, 24](#)

Data Type Reference Overview

Informatica developer uses the following data types in Microsoft Azure Blob Storage mappings:

- Microsoft Azure Blob Storage native data types. Microsoft Azure Blob Storage data types appear in the physical data object column properties.
- Transformation data types. Set of data types that appear in the transformations. They are internal data types based on ANSI SQL-92 generic data types, which the Data Integration Service uses to move data across platforms. Transformation data types appear in all transformations in a mapping.

When the Data Integration Service reads a source, it converts the native data types to the comparable transformation data types before transforming the data. When the Data Integration Service writes to a target, it converts the transformation data types to the comparable native data types.

Microsoft Azure Blob Storage and Transformation Data Types

The following table lists the Microsoft Azure Blob Storage data types that the Data Integration Service supports and the corresponding transformation data types:

Microsoft Azure Blob Storage Native Data Type	Transformation Data Type	Range and Description
Bigint	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Decimal	Decimal	Precision 15
Integer	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
String	String	1 to 104,857,600 characters

Avro File Data Types and Transformation Data Types

Avro file data types map to transformation data types that the Data Integration Service uses to move data across platforms.

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

Avro File Data Type	Transformation Data Type	Range and Description
Boolean	Integer	TRUE (1) or FALSE (0)
Bytes	Binary	Precision 4000
Double	Double	Precision 15
Float	Double	Precision 15
Int	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
Long	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0

Avro File Data Type	Transformation Data Type	Range and Description
Null	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
String	String	1 to 104,857,600 characters

Note: The Null data type is supported only in the native environment.

JSON File Data Types and Transformation Data Types

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

JSON File Data Type	Transformation Data Type	Range and Description
Bigint	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Double	Double	Precision 15
Int	Integer	-2,147,483,648 to 2,147,483,647 Precision 10, scale 0
String	String	1 to 104,857,600 characters

Parquet File Data Types and Transformation Data Types

Parquet file data types map to transformation data types that the Data Integration Service uses to move data across platforms.

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

Parquet File Data Type	Transformation Data Type	Range and Description
Binary (UTF-8)	String	1 to 104,857,600 characters
Boolean	Integer	TRUE (1) or FALSE (0)

Parquet File Data Type	Transformation Data Type	Range and Description
Byte_Array	Binary	Arbitrarily long byte array
Double	Double	Precision 15
Float	Double	Precision 15
Int32	Integer	-2,147,483,648 to +2,147,483,647 Precision of 10, scale of 0
Int64	Bigint	-9,223,372,036,854,775,808 to +9,223,372,036,854,775,807 Precision of 19, scale of 0

The Parquet schema that you specify to read or write a Parquet file must be in smaller case. Parquet does not support case-sensitive schema.

INDEX

C

- client component
 - installation [12](#)
- Configuration [10](#)
- Configure Temporary Directory Location [12](#)
- Connection
 - details [14](#)
 - properties [14](#)

D

- Data
 - objects [16](#)
- Data object
 - storage [18](#)
- Data types
 - Avro file [23](#)
 - description [23](#)
 - JSON file [24](#)
 - range [23](#)
 - transformation [23](#)
- directory source
 - Microsoft Azure Blob Storage sources [17](#)

I

- installation
 - client component [12](#)
- Installation [10](#)
- installation on Linux master gateway
 - server component [11](#)
- installation on Linux nodes
 - server component [11](#)

J

- java heap size [12](#)

M

- mapping
 - validation [21](#)
- Microsoft Azure
 - Blob Storage [8](#)

P

- Parquet file data types
 - transformation data types [24](#)
- PowerExchange [8](#)
- Prerequisites [10](#)

S

- server component
 - installation on Linux master gateway [11](#)
 - installation on Linux Nodes [11](#)
- storage solution
 - cloud [9](#)

U

- unstructured data
 - blobs [9](#)
 - objects [9](#)