Informatica

Informatica® PowerExchange for Microsoft Azure Blob Storage

10.1.1

User Guide for PowerCenter

Informatica PowerExchange for Microsoft Azure Blob Storage User Guide for PowerCenter
10.1.1
November 2016

Publication Date: 2018-09-27

Table of Contents

Preface

The Informatica PowerExchange® for Microsoft Azure Blob Storage User Guide for PowerCenter® describes how to read data from and write data to Microsoft Azure Blob Storage. The guide is written for database administrators and developers who are responsible for moving data from a source to a Microsoft Azure Blob Storage target and from a Microsoft Azure Blob Storage source to a target. This guide assumes that you have knowledge of database engines, Microsoft Azure Blob Storage, and PowerCenter.

Informatica Resources

Informatica Network

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

As a member, you can:

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

Informatica Knowledge Base

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

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

Informatica Documentation

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

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

Informatica Product Availability Matrixes

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

Informatica Velocity

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

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

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

Informatica Marketplace

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

Informatica Global Customer Support

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

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

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

CHAPTER 1

Introduction to PowerExchange for Microsoft Azure Blob Storage

This chapter includes the following topics:

PowerExchange for Microsoft Azure Blob Storage Overview

You can use PowerExchange for Microsoft Azure Blob Storage to connect PowerCenter and Microsoft Azure Blob Storage.

Use PowerExchange for Microsoft Azure Blob Storage to read delimited files from or write delimited files to Microsoft Azure Blob Storage. You can read compressed files from or write compressed files to Microsoft Azure Blob Storage. You can use Microsoft Azure Blob Storage objects as sources and targets in mappings. When you use Microsoft Azure Blob Storage objects in mappings, you must configure properties specific to Microsoft Azure Blob Storage.

Note: You cannot configure a Microsoft Azure Blob Storage session to partition data.

Microsoft Azure Blob Storage stores unstructured data. Blobs are files of any type and size and are organized into containers in Microsoft Azure Storage. You can access delimited files that are page blobs or block blobs with PowerExchange for Microsoft Azure Blob Storage.

PowerExchange for Azure Blob Example

You work in sales operations and want to score leads to drive 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 that contain 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 mappings run serially on a schedule. The sales organization can see the most current lead scores for prospects and focus on the most promising leads.

PowerCenter Integration Service and Microsoft Azure Blob Storage Integration

The PowerCenter Integration Service uses the Microsoft Azure Blob Storage connection to connect to Microsoft Azure Blob Storage.

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

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

CHAPTER 2

PowerExchange for Microsoft Azure Blob Storage Installation and Configuration

This chapter includes the following topics:

PowerExchange for Microsoft Azure Blob Storage Installation and Configuration Overview

You can install PowerExchange for Microsoft Azure Blob Storage on Windows 64-bit, or Red Hat Enterprise Linux 64-bit machines.

When you install the PowerExchange for Microsoft Azure Blob Storage server component, you enable the PowerCenter Integration Service to read data from or write data to Microsoft Azure Blob Storage.

Prerequisites

Before you can use PowerExchange for Microsoft Azure Blob Storage, perform the following tasks:

1. Install or upgrade to PowerCenter 10.1.1.
2. Get the Account Name and Account Key.

   You can create AzureBlob service on the Microsoft Azure portal. Once the AzureBlob service is up and running, you can get the Account Name and Account Key from the properties of the AzureBlob service.
3. Verify that you have read, write, and execute permissions on the following directory:

   `<Informatica installation directory>/server/bin`.

Installing the Server Component

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

If you configure the PowerCenter Integration Service or PowerCenter Repository Service to run on primary and backup nodes, install the PowerExchange for Microsoft Azure Blob Storage server component on each node configured to run the PowerCenter Integration Service or PowerCenter Repository Service.

If you configure the PowerCenter Integration Service to run on a grid, install the PowerCenter for Microsoft Azure Blob Storage server component on each node configured to run on the grid. If you cannot install the PowerCenter for Microsoft Azure Blob Storage server component on each node on the grid, create a resource in the domain and assign it to each node where you installed the PowerCenter for Microsoft Azure Blob Storage server component. When you create a session, configure the session to use the resource.

Installing the Server Component on Windows

You can install the PowerExchange for Microsoft Azure Blob Storage server component on Windows 64-bit machine. The PowerExchange for Microsoft Azure Blob Storage server component installs the PowerCenter Integration Service and PowerCenter Repository Service components.

1. Run `install.bat` from the installation package.
2. Click **Next**.
3. Select the Informatica installation directory.

   By default, the server components are installed in the following location: `C:\Informatica installation directory\<version folder>`.
4. Click **Next**.
5. Click **Install** to begin the installation.
6. Click **Done** when the installation is complete.

The PowerCenter Integration Service and PowerCenter Repository Service components are installed.

Installing the Server Component on Linux

Install the PowerExchange for Microsoft Azure Blob Storage server component on a Red Hat Enterprise Linux 64-bit machine when the PowerCenter Integration Service or PowerCenter Repository Service runs on Linux.

1. Enter `sh install.sh` at the prompt.
2. 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>`.

The PowerCenter Integration Service and PowerCenter Repository Service components are installed.

Installing the Client Component

Install the client component on every PowerCenter client machine that connects to the domain where the PowerExchange for Microsoft Azure Blob Storage server is installed.

1. Unzip the installation archive and navigate to the root directory of the extracted installer files.

2. Run the `install.bat` script file.

   The Welcome page appears.

3. Click **Next**.

   The Installation Directory page appears.

4. Enter the absolute path to the Informatica client installation directory.

   You can click **Browse** to find the directory or use the default directory.
   By default, the PowerCenter client is installed in the following location: `C:\Informatica\<version folder>`.

5. Click **Next**.

   The Pre-Installation Summary page appears.

6. 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.

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

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

Registering the Plug-in

After you install or upgrade PowerExchange for Microsoft Azure Blob Storage, you must register the plug-in with the PowerCenter repository.

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

The plug-in file for PowerExchange for Microsoft Azure Blob Storage is `AzureBlobPlugin.xml`. When you install PowerExchange for Microsoft Azure Blob Storage, the installer copies the `AzureBlobPlugin.xml` file to the following directory: `<Informatica Installation Directory>\server\bin\Plugin`.

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

CHAPTER 3

Microsoft Azure Blob Storage Sources and Targets

This chapter includes the following topics:

Microsoft Azure Blob Storage Sources and Targets Overview

You can create a mapping with a Microsoft Azure Blob Storage source to read delimited files from Microsoft Azure Blob Storage and write to a target. You can create a mapping with any source and a Microsoft Azure Blob Storage target to write delimited files to Microsoft Azure Blob Storage.

Importing a Microsoft Azure Blob Storage Source or Target

Import Microsoft Azure Blob Storage source and target objects before you create a mapping.

1. Start the PowerCenter Designer and connect to a PowerCenter repository.
2. Open a source or target folder.
3. Select **Source Analyzer** or **Target Designer**.
4. Click **Sources** or **Targets**, and then click **Import from AzureBlob**.

   The Establish Connection dialog box appears.

5. Specify the following information and click **Connect**.

Connection Property	Description
Account Name	Name of the Microsoft Azure Storage account.
Account Key	Microsoft Azure Storage access key.
Container Name	Microsoft Azure Blob Storage container name.
File Delimiter	Character used to separate fields in the file. Default is a comma (,).

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

CHAPTER 4

Microsoft Azure Blob Storage Sessions

This chapter includes the following topics:

Microsoft Azure Blob Storage Sessions Overview

After you create mappings, you can create a session to extract, transform, and load data.

You can configure a Microsoft Azure Blob Storage connection in the Workflow Manager to read delimited and compressed files from or write delimited and compressed files to Microsoft Azure Blob Storage. Ensure that you have write access to the Microsoft Azure Blob Storage bucket you want to access. You can define properties in a session to determine how the PowerCenter Integration Service reads delimited or compressed files from a Microsoft Azure Blob Storage source or writes delimited or compressed files to a Microsoft Azure Blob Storage target.

When you write to Microsoft Azure Blob Storage targets, you can only insert data to Microsoft Azure Blob Storage targets. You cannot perform update or delete operation on Microsoft Azure Blob Storage. Any data that exists in the Microsoft Azure Blob Storage target is overwritten when you run a session to write to Microsoft Azure Blob Storage.

Note: If you get Java Heap Size error when you read from or write to Microsoft Azure Blob Storage, increase the Java heap size.

Microsoft Azure Blob Storage Connections

Microsoft Azure Blob Storage connections enable you to read data from or write data to Microsoft Azure Blob Storage. The PowerCenter Integration Service uses the connection when you run a Microsoft Azure Blob Storage session.

Microsoft Azure Blob Storage Connection Properties

When you configure a Microsoft Azure Blob Storage connection, you define the connection properties that the PowerCenter Integration Service uses to connect to Microsoft Azure Blob Storage.

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

Connection Property	Description
Account Name	Name of the Microsoft Azure Blob Storage account.
Account Key	Microsoft Azure Storage access key.
Container Name	Microsoft Azure Blob Storage container name.
File Delimiter	Character used to separate fields in the file. Default is a comma (,).

Configuring a Microsoft Azure Blob Storage Connection

Configure a Microsoft Azure Blob Storage connection in the Workflow Manager to define the connection properties that the PowerCenter Integration Services uses to connect to Microsoft Azure Blob Storage.

1. In the Workflow Manager, select **Connections** > **Application**.

   The Application Connection Browser dialog box appears.
2. Click **New**.

   The Select Subtype dialog box appears.
3. Select **AzureBlob** and click **OK**.

   The Connection Object Definition dialog box appears.
4. Enter a name for the Microsoft Azure Blob Storage connection.
5. Enter the Microsoft Azure Blob Storage connection properties.
6. Click **OK**.

Configuring the Source Qualifier

After you import a source to create a mapping for Microsoft Azure Blob Storage source, you must configure the source qualifier.

1. In a mapping, double-click the **Source Qualifier**.
2. Select the **Configure** tab and click **Configure**.

   The Establish Connection dialog box appears.
3. Specify the Microsoft Azure Blob Storage connection properties and click **Connect**.
4. Click **Finish**.
5. Save the mapping.

Microsoft Azure Blob Storage Source Session Properties

Create a mapping with a Microsoft Azure Blob Storage source and a target to read data from Microsoft Azure Blob Storage. If the file size of a Microsoft Azure Blob Storage object is greater than 8 MB, you can enable the Enable Downloading Blobs in Multiple Parts option to download the object in multiple parts in parallel.

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

The following table describes the session properties, which you can configure for a Microsoft Azure Blob Storage source session:

Session Property	Description
Blob Name Override	Overrides the default folder name. Use this property to read compressed blob files from Microsoft Azure Blob Storage.
Blob Container Override	Overrides the default container name.
Header in the first row of Blob	Indicates that the first row of the Blob is a header and the PowerCenter Integration Service does not read the first row.
Number of concurrent connections to Blob Store	Number of concurrent connections to extract data from the Microsoft Azure Blob Storage. Default is 4.
INSERT	This property is not applicable.
DELETE	This property is not applicable.
UPDATE	This property is not applicable.
Success File Directory	This property is not applicable.
Error File Directory	This property is not applicable.

Microsoft Azure Blob Storage Target Session Properties

Create a session and associate it with a mapping that you created to write data to Microsoft Azure Blob Storage. Define the session properties to write data to Microsoft Azure Blob Storage.

The following table describes the session properties, which you can configure for a Microsoft Azure Blob Storage target session:

Session Property	Description
Number of concurrent connections to Blob Store	Number of concurrent connections to load data to the Microsoft Azure Blob Storage. Default is 4.
Blob Name Override	Overrides the default folder name. Use this property to write compressed blob files to Microsoft Azure Blob Storage.
Blob Container Override	Overrides the default container name.
Header in the first row of Blob	Adds the header to the target file when you select this field.
Compress newly created Blob	Compresses the newly created blob when set to True.
Write Strategy	Appends block to a blob, when you select append blob.
Blob Type	Writes data to a block blob or an append blob.
INSERT	Required. Inserts the source data to the Microsoft Azure Blob Storage target. Overwrites any existing data in the target object. **Note:** You can only insert data to Microsoft Azure Blob Storage objects. You cannot perform delete or update operations on Microsoft Azure Blob Storage targets.
DELETE	This property is not applicable.
UPDATE	This property is not applicable.
Success File Directory	This property is not applicable.
Error File Directory	This property is not applicable.

APPENDIX A

Data Type Reference

This appendix includes the following topic:

Data Type Reference Overview

PowerExchange for Microsoft Azure Blob Storage uses only CSV files in PowerCenter sessions.

PowerExchange for Microsoft Azure Blob Storage uses the following data types in PowerCenter sessions with Microsoft Azure Blob Storage objects:

Microsoft Azure Blob Storage native data types

Microsoft Azure Blob Storage data types appear on the Datatype tab for source qualifiers and target definitions when you edit metadata for the fields.

Transformation data types

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

When PowerExchange for Microsoft Azure Blob Storage reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When PowerExchange for Microsoft Azure Blob Storage writes to a target, it converts the transformation data types to the comparable native data types.

The following table lists the Microsoft Azure Blob Storage data types that PowerExchange for Microsoft Azure Blob Storage supports and the corresponding transformation data types:

Microsoft Azure Blob Storage Native Data Type	Transformation Data Type	Description
String	String	1 to 104,857,600 characters