



Informatica® PowerExchange for Microsoft
Azure Data Lake Storage Gen1
10.2.1

User Guide

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.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, the Informatica logo, PowerExchange, and Big Data Management 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.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

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 products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA 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.

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrixes.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to PowerExchange for Microsoft Azure Data Lake Store.....	 7
PowerExchange for Microsoft Azure Data Lake Store Overview.	7
Introduction to Microsoft Azure Data Lake Store.	8
Administration of Microsoft Azure Data Lake Store Connector.	8
 Chapter 2: PowerExchange for Microsoft Azure Data Lake Store Configuration.....	 9
PowerExchange for Microsoft Azure Data Lake Store Configuration Overview.	9
Prerequisites.	9
 Chapter 3: Microsoft Azure Data Lake Store Connections.....	 10
Microsoft Azure Data Lake Store Connections Overview.	10
Microsoft Azure Data Lake Store Connection Properties.	11
Creating a Microsoft Azure Data Lake Store Connection.	12
 Chapter 4: PowerExchange for Microsoft Azure Data Lake Store Data Objects.....	 13
Microsoft Azure Data Lake Store Data Object Overview.	13
Microsoft Azure Data Lake Store Data Object Properties.	14
Microsoft Azure Data Lake Store Data Object Read Operation.	14
Directory Source in Microsoft Azure Data Lake Store Sources.	14
Microsoft Azure Data Lake Store Object Read Operation Properties.	15
Source Properties of the Data Object Read Operation.	15
General Properties - Source.	15
Ports Properties - Source.	15
Advanced Properties - Source.	16
Column Projection Properties.	16
Microsoft Azure Data Lake Store Data Object Write Operation Properties.	17
Input Properties of the Data Object Write Operation.	17
Ports Properties - Input Write.	18

Run-time Properties.	18
Advanced Properties - Target.	18
Column Projection Properties.	19
Importing a Microsoft Azure Data Lake Store Data Object.	19
Creating a Microsoft Azure Data Lake Store Object Read or Write Operation.	20
Chapter 5: Microsoft Azure Data Lake Store Mappings.	22
Microsoft Azure Data Lake Store Mapping Overview.	22
Mapping Validation and Run-time Environments.	22
Microsoft Azure Data Lake Store Mapping Example.	23
Appendix A: Microsoft Azure Data Lake Store Datatype Reference.	24
Datatype Reference Overview.	24
Microsoft Azure Data Lake Store and Transformation Data Types.	25
Avro File Data Types and Transformation Data Types.	25
Intelligent Structure Model Data Types and Transformation Data Types.	26
JSON File Data Types and Transformation Data Types.	27
Parquet File Data Types and Transformation Data Types.	27
Index.	29

Preface

The *Informatica PowerExchange® for Microsoft Azure Data Lake Store User Guide* provides information about reading data from and writing data to Microsoft Azure Data Lake Store. It is written for administrators and developers who create mappings to read from or write data to a Microsoft Azure Data Lake Store resource.

This guide assumes that you have knowledge of Microsoft Azure Data Lake Store 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.

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 Data Lake Store

This chapter includes the following topics:

- [PowerExchange for Microsoft Azure Data Lake Store Overview, 7](#)
- [Introduction to Microsoft Azure Data Lake Store, 8](#)
- [Administration of Microsoft Azure Data Lake Store Connector, 8](#)

PowerExchange for Microsoft Azure Data Lake Store Overview

All the occurrences of *PowerExchange for Microsoft Azure Data Lake Store* in this document denote *PowerExchange for Microsoft Azure Data Lake Storage Gen1*.

You can use PowerExchange for Microsoft Azure Data Lake Store to connect to Microsoft Azure Data Lake Store from Informatica.

Use PowerExchange for Microsoft Azure Data Lake Store to read data from and write data to Microsoft Azure Data Lake Store. You can collate and organize the details from multiple input sources and use PowerExchange for Microsoft Azure Data Lake Store to write data to Microsoft Azure Data Lake Store. You can use Microsoft Azure Data Lake Store objects as sources and targets in mappings. When you use Microsoft Azure Data Lake Store objects in mappings, you must configure properties specific to Microsoft Azure Data Lake Store. You can validate and run mappings in native or Hadoop environments.

Use PowerExchange for Microsoft Azure Data Lake Store to read flat files and the industry-standard file formats, such as Avro, Parquet, and JSON files. You can read and write hierarchical data present in the Avro, Parquet, and JSON files. You can read data from and write data to only uncompressed Avro, Parquet, and JSON files. In addition to the industry-standard file formats, you can also read data from intelligent structure sources.

Introduction to Microsoft Azure Data Lake Store

You can use Microsoft Azure Data Lake Store to store data irrespective of size, structure, and format. Use Microsoft Azure Data Lake Store to process large volumes of data to achieve faster business outcomes.

The Microsoft Azure Data Lake Store is a combination of following primary services:

Microsoft Azure Data Lake Store: Microsoft Azure Data Lake is an Hadoop Distributed File System (HDFS) compatible data repository that stores structured, semi structured, and unstructured data of your organization in native format. Data scientists and data analysts can use data in the Data Lake to find out specific patterns before you move the analyzed data to a data warehouse.

Azure HDInsight: Azure HDInsight is an Apache Hadoop distributed system that is configured with cloud. HDInsight supports the Hadoop components such as Apache HBase, Apache Spark, and Apache Storm and other technologies under the Hadoop system.

Data Lake Analytics: Data Lake Analytics manages the distributed infrastructure. You can dynamically allocate and de-allocate resources with Microsoft Azure Data Lake Store Analytics so that you can pay for services that you use.

Administration of Microsoft Azure Data Lake Store Connector

Before you use PowerExchange for Microsoft Azure Data Lake Store, perform the following tasks:

- Create a Microsoft Azure Data Lake Store.
- Authorize users to access the Microsoft Azure Data Lake Store and the directory in the Data Lake Store.
- Create an Azure AD web application for service-to-service authentication with Azure Data Lake Store. For information on service-to-service authentication of your web application, see, <https://kb.informatica.com/howto/6/Pages/20/512374.aspx>.

Note: Ensure that you have superuser privileges to access the folders or files created in the application using PowerExchange for Microsoft Azure Data Lake Store.

CHAPTER 2

PowerExchange for Microsoft Azure Data Lake Store Configuration

This chapter includes the following topics:

- [PowerExchange for Microsoft Azure Data Lake Store Configuration Overview, 9](#)
- [Prerequisites, 9](#)

PowerExchange for Microsoft Azure Data Lake Store Configuration Overview

PowerExchange for Microsoft Azure Data Lake Store installs with the Informatica services and clients.

To configure PowerExchange for Microsoft Azure Data Lake Store, complete the prerequisites.

Prerequisites

Before you use PowerExchange for Microsoft Azure Data Lake Store, you must complete the following prerequisites:

- Install and configure the Informatica services.
- Install and configure the Developer tool. You can install the Developer tool when you install Informatica clients.
- Create a Data Integration Service and a Model Repository Service in the Informatica domain.

CHAPTER 3

Microsoft Azure Data Lake Store Connections

This chapter includes the following topics:

- [Microsoft Azure Data Lake Store Connections Overview, 10](#)
- [Microsoft Azure Data Lake Store Connection Properties, 11](#)
- [Creating a Microsoft Azure Data Lake Store Connection, 12](#)

Microsoft Azure Data Lake Store Connections Overview

Microsoft Azure Data Lake Store connection enables you to read data from or write data to Microsoft Azure Data Lake Store.

You can use Microsoft Azure Data Lake Store 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 a Microsoft Azure Data Lake Store connection from the Developer tool or the Administrator tool. Create and manage connections in the Preferences dialog box or the Connection Explorer view.

Microsoft Azure Data Lake Store Connection Properties

When you set up a Microsoft Azure Data Lake Store connection, you must configure the connection properties.

The following table describes the Microsoft Azure Data Lake Store 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	The description of the connection. The description cannot exceed 4,000 characters.
Location	The domain where you want to create the connection.
Type	The connection type. Select Microsoft Azure Data Lake Store.

The following table describes the properties for metadata access:

Property	Description
ADLS Account Name	The name of the Microsoft Azure Data Lake Store.
ClientID	The ID of your application to complete the OAuth Authentication in the Active Directory.
Client Secret	The client secret key to complete the OAuth Authentication in the Active Directory.
Directory	The Microsoft Azure Data Lake Store directory that you use to read data or write data. The default is root directory.
AuthEndpoint	The OAuth 2.0 token endpoint from where access code is generated based on based on the Client ID and Client secret is completed.

For more information about creating a client ID, client secret, and auth end point, contact the Azure administrator or see Microsoft Azure Data Lake Store documentation.

Creating a Microsoft Azure Data Lake Store Connection

Create a Microsoft Azure Data Lake Store connection before you create a Microsoft Azure Data Lake Store 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 > ADLS**, and click **Add**.
5. Enter a connection name and an optional description.
6. Select Microsoft Azure Data Lake Store as the connection type.
7. Click **Next**.
8. Configure the connection properties.
9. Click **Test Connection** to verify the connection to Microsoft Azure Data Lake Store.
10. Click **Finish**.

CHAPTER 4

PowerExchange for Microsoft Azure Data Lake Store Data Objects

This chapter includes the following topics:

- [Microsoft Azure Data Lake Store Data Object Overview, 13](#)
- [Microsoft Azure Data Lake Store Data Object Properties, 14](#)
- [Microsoft Azure Data Lake Store Data Object Read Operation, 14](#)
- [Microsoft Azure Data Lake Store Data Object Write Operation Properties, 17](#)
- [Importing a Microsoft Azure Data Lake Store Data Object, 19](#)
- [Creating a Microsoft Azure Data Lake Store Object Read or Write Operation, 20](#)

Microsoft Azure Data Lake Store Data Object Overview

A Microsoft Azure Data Lake Store data object is a physical data object that represents Microsoft Azure Data Lake Store table as a source or target. A Microsoft Azure Data Lake Store data object is the representation of data that is based on a Microsoft Azure Data Lake Store object. You can configure the data object read and write operation properties that determine how data can be read from Microsoft Azure Data Lake Store or loaded to Microsoft Azure Data Lake Store.

To read data from the Microsoft Azure Data Lake Store, create a data object read operation based on the Microsoft Azure Data Lake Store data object. Configure the read operation properties to determine how the Data Integration Service must read data from the Microsoft Azure Data Lake Store table. Add the read operation as a Read transformation in a mapping.

To write data to the Microsoft Azure Data Lake Store, create a data object write operation based on the Microsoft Azure Data Lake Store data object. Configure the write operation properties to determine how the Data Integration Service must write data to the Microsoft Azure Data Lake Store. Add the write operation as a Write transformation in a mapping.

Microsoft Azure Data Lake Store Data Object Properties

The Microsoft Azure Data Lake Store Overview view displays general information about the Microsoft Azure Data Lake Store data object and the object properties that apply to the Microsoft Azure Data Lake Store table you import.

General Properties

You can configure the following properties for a Microsoft Azure Data Lake Store data object:

- Name. Name of the Microsoft Azure Data Lake Store data object.
- Description. Description of the Microsoft Azure Data Lake Store data object.
- Connection. Name of the Microsoft Azure Data Lake Store connection.

Microsoft Azure Data Lake Store Data Object Read Operation

The Data Integration Service reads data from a Microsoft Azure Data Lake Store table based on the data object read operation properties that you specify.

When you create a data object read operation, the Developer tool creates a Source transformation and an Output transformation.

The Source transformation represents the data that the Data Integration Service reads from the Microsoft Azure Data Lake Store table.

The Output transformation represents the data that the Data Integration Service passes into the mapping pipeline.

Directory Source in Microsoft Azure Data Lake Store 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 Data Lake Store data object read operation:

- File
- Directory

Note: To read compressed `.csv` files, the content type of the Microsoft Azure Data Lake Store source must be `application/x-gzip` 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.
- The directory read is not applicable to the binary file type.
- 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 file formats in the directory you select and might result into errors.

Microsoft Azure Data Lake Store Object Read Operation Properties

The Data Integration Service reads data from a Microsoft Azure Data Lake Store object based on the data object read operation. The Developer tool displays the data object read operation properties of the Microsoft Azure Data Lake Store data object in the Data Object Operation view.

You can view or configure the data object read operation from the source and output properties.

Source properties

Represents data that the Data Integration Service reads from the Microsoft Azure Data Lake Store object. Select the source properties to view data, such as the name and description of the Microsoft Azure Data Lake Store object, the ports, and advanced properties.

Output properties

Represents data that the Data Integration Service passes into the mapping pipeline. Select the output properties to edit the port properties of the data object read operation. You can also use the Tracing Level advanced property to set the amount of detail that the Data Integration Service writes in the log.

Source Properties of the Data Object Read Operation

When you create a data object, the source properties populate based on the Microsoft Azure Data Lake Store object that you add. The source properties of the data object read operation include general, column, and advanced properties that apply to the Microsoft Azure Data Lake Store object.

You can view the source properties of the data object read operation from the **General**, **Column**, and **Advanced** tabs.

General Properties - Source

The following table describes the source general properties of the data object read operation:

Property	Description
Name	Name of the Microsoft Azure Data Lake Store source object.
Description	Description of the data object read operation.

Ports Properties - Source

The column properties display the data types, precision, and scale of the source property in the data object read operation.

The following table describes the source column properties of the data object read operation:

Property	Description
Name	Name of the column.
Type	Native data type of the column.

Property	Description
Precision	Maximum number of significant digits for Numeric data types, or maximum number of characters for String data types. For numeric data types, precision includes scale.
Scale	Maximum number of digits after the decimal point for numeric values.
Description	Description of the column.

Advanced Properties - Source

You can use the advanced properties to specify data object read operation properties to read data from a Microsoft Azure Data Lake Store server.

The following table describes the Advanced source column properties of the data object read operation:

Property	Description
Directory Path Override	Overrides the default directory path.
File Name Override	Overrides the file name.
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. The directory read is not applicable to the binary file type. You can read compressed <code>.csv</code> files in the native environment.

Column Projection Properties

The Developer tool displays the column projection properties for Intelligent Structure Model, Avro, JSON, and Parquet complex file sources in the Properties view of the **Read** operation.

The following table describes the column projection properties that you configure for the complex file sources:

Property	Description
Enable Column Projection	Displays the column details of the complex files sources.
Schema Format	Displays the schema format that you selected while creating the complex file data object. You can change the schema format and provide respective schema.
Schema	Displays the schema associated with the complex file. You can select a different schema. Note: If you disable the column projection, the schema associated with the complex file is removed. If you want to associate schema again with the complex file, enable the column projection and click Select Schema.

Property	Description
Model	Displays the intelligent structure model associated with the complex file. You can select a different model. Note: If you disable the column projection, the intelligent structure model associated with the data object is removed. If you want to associate an intelligent structure model again with the data object, enable the column projection and click Select Model.
Column Mapping	Displays the mapping between input and output ports. Note: If you disable the column projection, the mapping between input and output ports is removed. If you want to map the input and output ports, enable the column projection and click Select Schema to associate a schema to the complex file.
Project Column as Complex Data Type	Displays columns with hierarchical data as a complex data type, such as, array, map, or struct. Select this property when you want to process hierarchical data on the Spark engine. Note: If you disable the column projection, the data type of the column is displayed as binary type.

Microsoft Azure Data Lake Store Data Object Write Operation Properties

The Data Integration Service writes data to a Microsoft Azure Data Lake Store object based on the data object write operation. The Developer tool displays the data object write operation properties for the Microsoft Azure Data Lake Store data object in the Data Object Operation section.

You can view the data object write operation from the Input and Target properties.

Input properties

Represent data that the Data Integration Service reads from a Microsoft Azure Data Lake Store directory server. Select the input properties to edit the port properties and specify the advanced properties of the data object write operation.

Target properties

Represent data that the Data Integration Service writes to Microsoft Azure Data Lake Store. Select the target properties to view data, such as the name and description of the Microsoft Azure Data Lake Store object.

Input Properties of the Data Object Write Operation

Input properties represent data that the Data Integration Service writes to a Microsoft Azure Data Lake Store directory server. Select the input properties to edit the port properties of the data object write operation. You can also specify advanced data object write operation properties to write data to Microsoft Azure Data Lake Store objects.

The input properties of the data object write operation include general properties that apply to the data object write operation. Input properties also include port, source, and advanced properties that apply to the data object write operation.

You can view and change the input properties of the data object write operation from the **General**, **Ports**, **Targets**, **run-time**, and **Advanced** tabs.

Ports Properties - Input Write

The input ports properties list the data types, precision, and scale of the data object write operation.

The following table describes the input ports properties that you must configure in the data object write operation:

Property	Description
Name	Name of the port.
Type	Data type of the port.
Precision	Maximum number of significant digits for numeric data types, or maximum number of characters for string data types. For numeric data types, precision includes scale.
Scale	Maximum number of digits after the decimal point for numeric values.
Description	Description of the port.

Run-time Properties

The run-time properties display the name of the connection used for write transformation.

The following table describes the run-time properties that you configure for a Microsoft Azure Data Lake Store write operation:

Property	Description
Connection	Name of the Microsoft Azure Data Lake Store connection.

Advanced Properties - Target

You can use the advanced properties to specify data object write operation properties to write data to a Microsoft Azure Data Lake Store server.

The following table describes the advanced properties that you configure for a Microsoft Azure Data Lake Store write operation:

Property	Description
Tracing Level	By default, the tracing level for every transformation is Normal. Change the tracing level to a Verbose setting when you need to troubleshoot a transformation that is not behaving as expected. Set the tracing level to Terse when you want the minimum amount of detail to appear in the log.
Maintain row order	Not applicable
Directory Path Override	Overrides the default directory path.
File Name Override	Overrides the file name.

Property	Description
Compress Newly Created File	Enables compression. Applicable to the .csv files in the native environment.
If file exists	If the target exists, overwrites or appends the data to the existing file. You can also configure the mapping to fail if the target already exists. Applicable to .csv files in the native environment.

Column Projection Properties

The Developer tool displays the column projection properties for Avro, JSON, and Parquet complex file targets in the Properties view of the **Write** operation.

The following table describes the advanced properties that you configure for Avro, JSON, and Parquet complex file targets:

Property	Description
Enable Column Projection	Displays the column details of the complex files sources.
Schema Format	Displays the schema format that you selected while creating the complex file data object. You can change the schema format and provide respective schema.
Schema	Displays the schema associated with the complex file. You can select a different schema. Note: If you disable the column projection, the schema associated with the complex file is removed. If you want to associate schema again with the complex file, enable the column projection and click Select Schema.
Column Mapping	Displays the mapping between input and output ports. Note: If you disable the column projection, the mapping between input and output ports is removed. If you want to map the input and output ports, enable the column projection and click Select Schema to associate a schema to the complex file.
Project Column as Complex Data Type	Displays columns with hierarchical data as a complex data type, such as, array, map, or struct. Select this property when you want to process hierarchical data on the Spark engine. Note: If you disable the column projection, the data type of the column is displayed as binary type.

Importing a Microsoft Azure Data Lake Store Data Object

Import a Microsoft Azure Data Lake Store 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 **ADLS Data Object** and click **Next**.

The **ADLS 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:
 - Intelligent Structure Model: to read any format that an intelligent structure parses.
 - Binary: to read and write any resource format.
 - 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.

Note: Intelligent structure model and JSON formats are supported only on the Spark engine. For a data object with an intelligent structure model, create a read operation. You cannot use a write transformation for a data object with an intelligent structure model in a mapping. Avro and Parquet are supported in the native environment and on the Spark engine.

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 Microsoft Azure Data Lake Store connection from which you want to import the Microsoft Azure Data Lake Store resource metadata.
8. To add a resource, click **Add** next to the **Selected Resources** option.

The **Add Resource** dialog box appears.

9. From the Package Explorer, select a naming context from which you want to import the schema.
10. You can Perform one of the following tasks to import an Microsoft Azure Data Lake Store table, and then click **OK**:
 - Navigate to the Microsoft Azure Data Lake Store table that you want to import.
 - Search for the Microsoft Azure Data Lake Store table, enter the name of the Microsoft Azure Data Lake Store table entity that you want to add and click **OK**.
11. Click **Finish**.

The data object appears under Data Objects in the project or folder in the **Object Explorer** view.

Creating a Microsoft Azure Data Lake Store Object Read or Write Operation

You can add a Microsoft Azure Data Lake Store data object read or write operation to a mapping or mapplet as a source. You can create the data object read or write operation for one or more Microsoft Azure Data Lake Store data objects.

Before you create a Microsoft Azure Data Lake Store data object read or write operation, you must create at least one Microsoft Azure Data Lake Store data object.

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 read or write operation.
4. Select **Read** or **Write** as the type of data object operation.

5. Click **Add**.

The **Select Resources** dialog box appears.

6. Select the Microsoft Azure Data Lake Store object for which you want to create the data object read or write operation and click **OK**.

7. Click **Finish**.

The Developer tool creates the data object read or write operation for the selected data object.

CHAPTER 5

Microsoft Azure Data Lake Store Mappings

This chapter includes the following topics:

- [Microsoft Azure Data Lake Store Mapping Overview, 22](#)
- [Mapping Validation and Run-time Environments, 22](#)
- [Microsoft Azure Data Lake Store Mapping Example, 23](#)

Microsoft Azure Data Lake Store Mapping Overview

After you create the Microsoft Azure Data Lake Store data object with a Microsoft Azure Data Lake Store connection, you can develop a mapping. You can define the following types of objects in the mapping:

- A Read transformation of the Microsoft Azure Data Lake Store data object to read data from Microsoft Azure Data Lake Store in native or Hadoop run-time environment.
- A Write transformation of the Microsoft Azure Data Lake Store data object to write data to Microsoft Azure Data Lake Store in native or Hadoop run-time environment.

Note: If you use multiple connections in a mapping, verify that all connections point to the same Microsoft Azure Data Lake Store. If you run a mapping on the HDInsight cluster that uses Microsoft Azure Data Lake Store file system, the Microsoft Azure Data Lake Store connector must point to the same Microsoft Azure Data Lake Store.

Validate and run the mapping. You can deploy the mapping and run it or add the mapping to a Mapping task in a workflow. When you read data from Microsoft Azure Data Lake Store and terminate the mapping, the Data Integration Service downloads the entire data set and deletes the data set after long time.

Mapping Validation and Run-time Environments

You can validate and run mappings in the native environment or a 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 in a Hadoop environment, you can select the Spark engine. The Data Integration Service pushes the mapping to the Spark engine for processing.

Microsoft Azure Data Lake Store Mapping Example

You work as a data analyst for a large financial enterprise. The enterprise performs risk management, fraud detection, and other analysis with Microsoft Azure Data Lake Analytics. You need to write the data to Microsoft Azure Data Lake Store to perform the analytics. Create a mapping to read data from a flat file source and write data to Microsoft Azure Data Lake Store. After the data is available in the Microsoft Azure Data Lake Store, you can perform the data analytics.

You can use the following objects in a Microsoft Azure Data Lake Store mapping:

Flat file input

The input file is a flat file that contains the customer names and other details about customers.

Create a flat file data object. Configure the flat file connection and specify the flat file that contains the customer data as a resource for the data object. Drag the data object into a mapping as a read data object.

Transformations

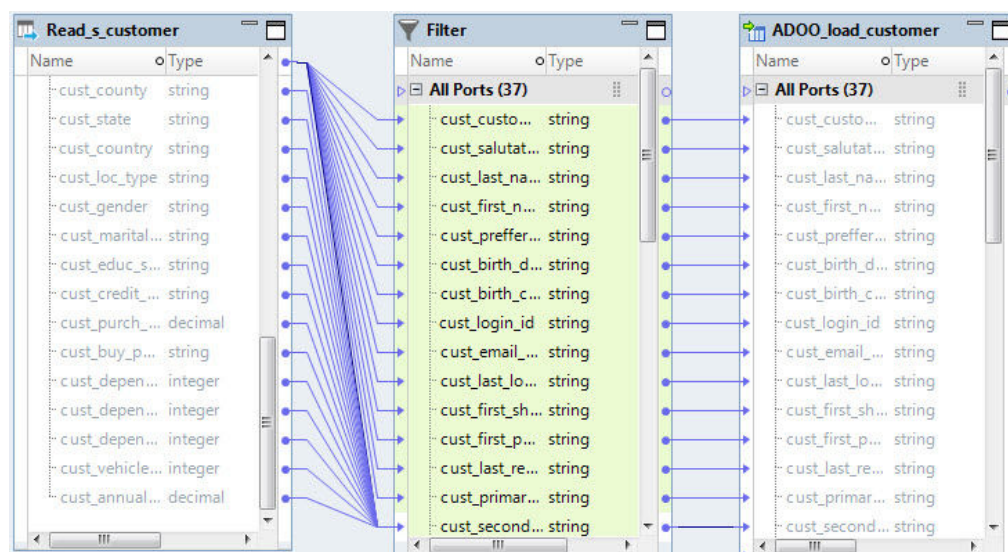
Add Filter transformation to get customer data in a particular region.

The Filter transformation filters the source data based on the value you specify for the region ID column. The Data Integration Service returns the rows that meet the filter condition.

Microsoft Azure Data Lake Store output

Create a Microsoft Azure Data Lake Store data object write operation. Configure the Microsoft Azure Data Lake Store connection and specify the Microsoft Azure Data Lake Store object as a target for the data object. Drag the data object into a mapping as a target data object.

The following image shows the Microsoft Azure Data Lake Store mapping example:



When you run the mapping, the customer records are read from the flat file and written to the Microsoft Azure Data Lake Store table.

APPENDIX A

Microsoft Azure Data Lake Store Datatype Reference

This appendix includes the following topics:

- [Datatype Reference Overview, 24](#)
- [Microsoft Azure Data Lake Store and Transformation Data Types, 25](#)
- [Avro File Data Types and Transformation Data Types, 25](#)
- [Intelligent Structure Model Data Types and Transformation Data Types, 26](#)
- [JSON File Data Types and Transformation Data Types, 27](#)
- [Parquet File Data Types and Transformation Data Types, 27](#)

Datatype Reference Overview

Informatica Developer uses the following data types in Microsoft Azure Data Lake Store mappings:

- Microsoft Azure Data Lake Store native data types. Microsoft Azure Data Lake Store 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 PowerExchange for Microsoft Azure Data Lake Store reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When PowerExchange for Microsoft Azure Data Lake Store writes to a target, it converts the transformation data types to the comparable native data types.

Microsoft Azure Data Lake Store and Transformation Data Types

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

Microsoft Azure Data Lake Store Native Data Type	Transformation Data Type	Range
String	string	1 to 104,857,600 characters
Number	Decimal	Precision 1 to 28 digits, scale 0 to 28
Binary	Binary	1 to 104,857,600 bytes The binary data type is applicable when you run a mapping in the native environment.

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
Array	Array	Unlimited number of characters
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
Map	Map	Unlimited number of characters

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
Record	Struct	Unlimited number of characters
String	String	1 to 104,857,600 characters

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

Avro Union Data Type

A union indicates that a field might have more than one data type. For example, a union might indicate that a field can be a string or a null. A union is represented as a JSON array containing the data types.

The Developer tool only interprets a union of ["primitive_type|complex_type", "null"] or ["null", "primitive_type|complex_type"]. The Avro data type converts to the corresponding transformation data type. The Developer tool ignores the null.

Unsupported Avro Data Types

The Developer tool does not support the following Avro data types:

- enum
- null

Intelligent Structure Model Data Types and Transformation Data Types

Intelligent structure model complex file data types map to transformation data types that the Data Integration Service uses to move data across platforms.

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

Intelligent Structure Model Complex File Data Type	Transformation Data Type	Range and Description
String	String	-1 to 104,857,600 characters

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
Array	Array	Unlimited number of characters.
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
Object	Struct	Unlimited number of characters.
String	String	1 to 104,857,600 characters

Unsupported JSON Data Types

The Developer tool does not support the following JSON data types:

- date/timestamp
- enum
- union

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)
Byte_Array	Binary	Arbitrarily long byte array

Parquet File Data Type	Transformation Data Type	Range and Description
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
Map	Map	Unlimited number of characters.
Struct	Struct	Unlimited number of characters.
Union	Corresponding primitive data type in a union of ["primitive_type", "null"] or ["null", "primitive_type"].	Dependent on primitive data type.

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.

Parquet Union Data Type

A union indicates that a field might have more than one data type. For example, a union might indicate that a field can be a string or a null. A union is represented as a JSON array containing the data types. The Developer tool only interprets a union of ["primitive_type", "null"] or ["null", "primitive_type"]. The Parquet data type converts to the corresponding transformation data type. The Developer tool ignores the null.

Unsupported Parquet Data Types

The Developer tool does not support the following Parquet data types:

- int96 (TIMESTAMP_MILLIS)

INDEX

A

advanced properties
 source [16](#)
 target [18](#)

C

create
 Microsoft Azure Data Lake Store connection [12](#)
creating
 Microsoft Azure Data Lake Store data object read operation [20](#)

D

data object read operation
 creating [20](#)
Data types
 Avro file [25](#)
 JSON file [27](#)
 transformation [25](#)
directory source
 Microsoft Azure Data Lake Store sources [14](#)

G

General properties [15](#)

I

importing
 Microsoft Azure Data Lake Store data object [19](#)
input properties [17](#)
intelligent structure model complex file data types
 transformation data types [26](#)

M

mapping example
 Microsoft Azure Data Lake Store [23](#)
Microsoft Azure Data Lake
 introduction [8](#)
Microsoft Azure Data Lake Store
 create [12](#)
 importing a data object [19](#)
Microsoft Azure Data Lake Store connection
 overview [10](#)
 properties [11](#)

Microsoft Azure Data Lake Store Connector
 administration [8](#)
Microsoft Azure Data Lake Store data object
 overview [13](#)
Microsoft Azure Data Lake Store data object read operation
 creating [20](#)
Microsoft Azure Data Lake Store data types
 comparing with transformation data types [25](#)
 overview [24](#)
Microsoft Azure Data Lake Store mappings
 overview [22](#)
Microsoft Azure Data Lake Store run-time environment
 description [22](#)
Microsoft Azure Data Lake Store validation environment
 description [22](#)
Microsoft Azure Data Lake Store write operation
 properties [17](#)

O

Output properties [15](#)
overview
 Microsoft Azure Data Lake Store connection [10](#)

P

Parquet file data types
 transformation data types [27](#)
Ports properties [15](#)
PowerExchange for Microsoft Azure Data Lake Store configuration
 overview [9](#)
 prerequisites [9](#)

R

read operation properties
 column projection properties [16](#)

S

source properties [15](#)

W

write operation properties
 column projection properties [19](#)