



Informatica® Data Integration - Free & PayGo

# Microsoft Azure Synapse SQL Connector

© Copyright Informatica LLC 2018, 2023

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, Informatica Cloud, and PowerCenter 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.

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, report them to us at [infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com).

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.

Publication Date: 2023-04-04

# Table of Contents

<b>Preface .....</b>	<b>5</b>
Informatica Resources. ....	5
Informatica Documentation. ....	5
Informatica Intelligent Cloud Services web site. ....	5
Informatica Intelligent Cloud Services Communities. ....	5
Informatica Intelligent Cloud Services Marketplace. ....	6
Simple Data Integration connector documentation. ....	6
Informatica Knowledge Base. ....	6
Informatica Intelligent Cloud Services Trust Center. ....	6
Informatica Global Customer Support. ....	6
 <b>Chapter 1: Introduction to Microsoft Azure Synapse SQL Connector.....</b>	<b>7</b>
Microsoft Azure Synapse SQL Connector assets. ....	7
Administration of Microsoft Azure Synapse SQL Connector. ....	8
Azure Active Directory authentication. ....	9
Managed identity authentication for Microsoft Azure Synapse SQL. ....	10
Managed identity authentication for Microsoft Azure Data Lake Storage Gen2. ....	10
Optimize the staging performance for a mapping. ....	11
 <b>Chapter 2: Connections for Microsoft Azure Synapse SQL.....</b>	<b>13</b>
Microsoft Azure Synapse SQL connection properties. ....	13
Proxy server configuration. ....	15
Configure proxy server settings on Windows. ....	15
Configure proxy server settings on Linux. ....	16
Bypassing the proxy server. ....	17
 <b>Chapter 3: Mappings for Microsoft Azure Synapse SQL.....</b>	<b>18</b>
Microsoft Azure Synapse SQL sources in mappings. ....	18
Adding multiple source objects. ....	21
Microsoft Azure Synapse SQL targets in mappings. ....	24
Copy command. ....	26
Rejected rows. ....	28
Mapping tasks with CDC sources. ....	28
Changed data capture limitations. ....	29
Bulk processing for write operations. ....	29
Rules and guidelines for configuring update override. ....	30
Microsoft Azure Synapse SQL lookups in mappings. ....	30
Enabling lookup caching. ....	33
Process SQL queries and stored procedures using an SQL transformation. ....	34
Configuring an SQL transformation. ....	34

Rules and guidelines for SQL transformations. . . . .	35
Dynamic schema handling . . . . .	36
Rules and guidelines for dynamic schema handling. . . . .	36
IDENTITY column. . . . .	36
Parameterization. . . . .	37
Rules and guidelines for mappings and mapping tasks. . . . .	37
<b>Chapter 4: Pushdown optimization. . . . .</b>	<b>41</b>
Pushdown optimization types. . . . .	41
Data Integration behavior with source and full pushdown optimization. . . . .	41
Pushdown optimization scenarios. . . . .	42
Pushdown optimization using a Microsoft Azure Synapse SQL connection. . . . .	42
Previewing pushdown optimization . . . . .	43
Configuring optimization for a Microsoft Azure Synapse SQL mapping task. . . . .	43
Configuring a custom query for the Microsoft Azure Synapse SQL source object. . . . .	54
Rules and guidelines for pushdown optimization. . . . .	54
<b>Chapter 5: Data type reference. . . . .</b>	<b>57</b>
Microsoft Azure Synapse SQL and transformation data types. . . . .	57
<b>Chapter 6: Troubleshooting. . . . .</b>	<b>59</b>
Troubleshooting a Microsoft Azure Synapse SQL connection. . . . .	59
Troubleshooting a mapping or mapping task. . . . .	59
<b>Index. . . . .</b>	<b>61</b>

# Preface

Use *Microsoft Azure Synapse SQL Connector* to learn how to read from or write to Microsoft Azure Synapse SQL. Learn to create a connection, develop and run mappings, mapping tasks, and data transfer tasks in Data Integration. Learn how to push down the transformation logic for processing to the Microsoft Azure Synapse SQL database.

## Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

### Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at [infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com).

### Informatica Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at <http://www.informatica.com/cloud>. This site contains information about Informatica Cloud integration services.

### Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:

<https://network.informatica.com/community/informatica-network/products/cloud-integration>

Developers can learn more and share tips at the Cloud Developer community:

<https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers>

## Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and maplets:

<https://marketplace.informatica.com/>

## Simple Data Integration connector documentation

You can access documentation for Simple Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit <https://docs.informatica.com>.

## Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at [KB\\_Feedback@informatica.com](mailto:KB_Feedback@informatica.com).

## Informatica Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at <https://www.informatica.com/trust-center.html>.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The [Informatica Intelligent Cloud Services Status](#) page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this page, and during an outage, it will have the most current information. To ensure you are notified of updates and outages, you can subscribe to receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, go to <https://status.informatica.com/> and click **SUBSCRIBE TO UPDATES**. You can then choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

## Informatica Global Customer Support

You can contact a Customer Support Center by telephone or online.

For online support, click **Submit Support Request** in Informatica Intelligent Cloud Services. You can also use Online Support to log a case. Online Support requires a login. You can request a login at <https://network.informatica.com/welcome>.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <https://www.informatica.com/services-and-training/support-services/contact-us.html>.

## CHAPTER 1

# Introduction to Microsoft Azure Synapse SQL Connector

You can use Microsoft Azure Synapse SQL Connector to read data from or write data to Microsoft Azure Synapse SQL.

When you use Microsoft Azure Synapse SQL Connector, you can create a Microsoft Azure Synapse SQL connection and use the connection in Data Integration mappings and tasks. You can use Microsoft SQL Server or Azure Active Directory authentication to connect to Microsoft Azure Synapse SQL.

You can use external tables as Microsoft Azure Synapse SQL sources in mappings. You can read data from or write data to Microsoft Azure Synapse SQL case-sensitive databases as well. You can create mappings and mapping tasks to read data from or write data to a Microsoft Azure Synapse SQL endpoint that resides in a virtual network (VNet).

Microsoft Azure Synapse SQL Connector is optimized for large data sets and can perform better than traditional data integration methods, such as ODBC or JDBC. When you write data to Microsoft Azure Synapse SQL, the connector stages data files to Microsoft Azure Storage and uses T-SQL commands with Microsoft Polybase to extract or load relational and non-relational data in parallel.

## Microsoft Azure Synapse SQL Connector assets

Create assets in Data Integration to integrate data using Microsoft Azure Synapse SQL Connector.

When you use Microsoft Azure Synapse SQL Connector, you can include the following Data Integration assets:

- Data transfer task
- Mapping
- Mapping task

For more information about configuring assets and transformations, see *Mappings, Transformations, and Tasks* in the Data Integration documentation.

# Administration of Microsoft Azure Synapse SQL Connector

Before you use a Microsoft Azure Synapse SQL connection, you must complete the following prerequisites:

- Obtain the JDBC URL from Microsoft Azure Synapse SQL.
- If you have an existing Microsoft Azure Synapse SQL connection, apply the latest SDKPatch. To get the latest SDKPatch, contact Informatica Global Customer Support.
- Ensure that a default schema is present at the account level or user or group level in Microsoft Azure Synapse SQL.
- Verify that either the `db_owner` privilege or the following more granular privileges are granted to the user to connect to Microsoft Azure Synapse SQL and perform read and write operations successfully:
  - `EXEC sp_addrolemember 'db_datareader', '<user>';` // Alternately assign permission to individual table
  - `EXEC sp_addrolemember 'db_datawriter', '<user>';` // Alternately assign permission to individual table
  - `GRANT ALTER ANY EXTERNAL DATA SOURCE TO <user>;`
  - `GRANT ALTER ANY EXTERNAL FILE FORMAT TO <user>;`
  - `GRANT CONTROL TO <user>;` // To grant all permissions on the database.
  - or
  - `GRANT ALTER ANY SCHEMA TO <user>;` // To grant permissions only on the schema.
  - `GRANT CREATE TABLE TO <user>;`
  - Assign required privileges for tasks performed through Pre-SQL and Post-SQL commands.
- When you use managed identity authentication to connect to Microsoft Azure Synapse SQL, the user for the system assigned identity is the virtual machine for which you enable the identity.  
For example, `GRANT CONTROL TO <virtual machine name>;`  
The user for the user assigned identity is the user identity that you create in the Azure portal.  
For example, `GRANT CONTROL TO <user identity name>;`
- If you have the ALTER ANY SCHEMA permissions, you must create the Master Key, Database Scoped Credential, and External Data Source in Microsoft Azure Synapse SQL that require the CONTROL permission on the database and specify the external data source when you create a connection. Also, Microsoft Azure Synapse SQL Connector does not delete the Database Scoped Credential and External Data Source. You must manually delete the Database Scoped Credential and External Data Source.
- When you read from or write to Microsoft Azure Synapse SQL, an interim directory in the agent machine is used to stage the files.  
Ensure that the interim directory is created in the agent machine and you have the read and write permissions to the interim directory.
- If you receive a Java heap space error when you create or run a task, you must set the JVM options for type Tomcat JRE to increase the -Xms and -Xmx values in **System Configuration Details > Data Integration Server** for the Secure Agent. You must then restart the Secure Agent.

For more information about configuring a Microsoft Azure Synapse SQL connection, see the Informatica How-To Library article, [Prerequisites to create a Microsoft Azure Synapse SQL connection](#).



## Azure Active Directory authentication

You can configure Azure Active Directory (AAD) authentication to connect to Microsoft Azure Synapse SQL. Default is Microsoft SQL Server authentication.

To configure the AAD authentication, perform the following tasks:

### Import server certificate

Applicable if a trust store file is not configured for your organization and you want to use AAD authentication with Active Directory Federation Services in Azure. Import the server certificate to the following location:

```
<Secure Agent installation directory>\jdk\jre\lib\security\cacerts
```

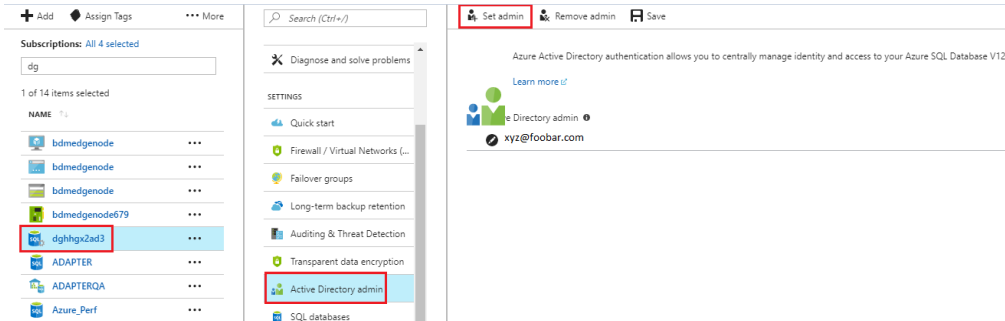
Use the following command to import the certificate:

```
keytool -import -trustcacerts -alias <alias name of the certificate> -file <certificate file path> -keystore <Secure Agent installation directory>\jdk\jre\lib\security\cacerts -storepass <password for the truststore>
```

### Set admin

Perform the following steps to set admin between Microsoft SQL Server that has the Microsoft Azure Synapse SQL hosted and the Azure Active Directory:

1. Log on to the Microsoft Azure portal using your credentials. The Dashboard page appears.
2. From the All Resources page, select the Microsoft SQL Server that has the Microsoft Azure Synapse SQL hosted.
3. Select the **Active Directory admin** option under Settings displayed for the Microsoft SQL Server. The image shows the Active Directory admin settings:



4. Click **Set admin**. The Add admin page appears.
5. Enter the email ID that you want to use as admin and click **Select**.
6. Click **Save**.

### Create a user

Perform the following steps to create a user:

1. Connect to Microsoft Azure Synapse SQL using the Azure Active Directory admin created in the previous steps. You can use Microsoft SQL Server Management Studio to connect to the Microsoft Azure Synapse SQL.
2. Type and run the following command to create a user: `create user [user@foobar.com] from external provider;`  
For more information, see Microsoft Azure documentation.
3. Assign the required privileges to the user.

Configure the JDBC URL and the user you created in connection properties to enable AAD authentication.

## Managed identity authentication for Microsoft Azure Synapse SQL

Before you use managed identity authentication to connect to Microsoft Azure Synapse SQL, complete the following prerequisites:

1. Create an Azure virtual machine.
2. Install the Secure Agent on the Azure virtual machine.
3. Enable system assigned identity or user assigned identity for the Azure virtual machine.  
If you enable both and do not specify the client ID, the system assigned identity is used for authentication.
4. After you add or remove a managed identity, restart the Azure virtual machine.

### Rules and guidelines for managed identity authentication

Consider the following rules and guidelines for managed identity authentication:

- You cannot use a proxy server with managed identity authentication.
- If the object ID for the system assigned identity is reset in the Azure portal, you must delete and recreate the virtual machine.

## Managed identity authentication for Microsoft Azure Data Lake Storage Gen2

Before you use managed identity authentication to connect to Microsoft Azure Data Lake Storage Gen2, complete the following prerequisites:

1. Create an Azure virtual machine.
2. Install the Secure Agent on the Azure virtual machine.
3. Enable system assigned identity or user assigned identity for the Azure virtual machine.  
If you enable both and do not specify the client ID, the system assigned identity is used for authentication.
4. After you add or remove a managed identity, restart the Azure virtual machine.

### Rules and guidelines for managed identity authentication

Consider the following rules and guidelines for managed identity authentication:

- You can use managed identity authentication for Microsoft Azure Data Lake Storage Gen2 to stage files only when you run a mapping to read data from a Microsoft Azure Synapse SQL source or Microsoft Azure Data Lake Storage Gen2 source.
- You cannot use a proxy server with managed identity authentication.

### Rules and guidelines for role-based access control and access control lists

You can use role-based access control or access control lists to authenticate access to the storage account for Microsoft Azure Data Lake Storage Gen2.

Consider the following rules and guidelines when you use managed identity authentication with role-based access control or access control lists:

- You cannot use both role-based access control and access control lists for the same storage account.
- When you use access control lists, assign the read, write, and execute permissions to Microsoft Azure Synapse SQL to access the file system.
- Assign read, write, and execute permissions to Microsoft SQL Server to access the file system.

- If you enable system assigned identity, assign the required role or permissions to the Azure virtual machine to run the mappings and tasks.  
If you enable user assigned identity, assign the required role or permissions to the Azure virtual machine and the user assigned identity to run the mappings and tasks.

## Optimize the staging performance for a mapping

Data Integration, by default, creates a flat file locally in a temporary folder to stage the data after reading from or before writing to Microsoft Azure Synapse SQL. You can set Data Integration to optimize the staging performance.

If you do not set the staging property, Data Integration performs staging without the optimized settings, which might impact the performance of the task.

Consider the following rules when you enable the staging property:

- When you read multiple objects, you cannot optimize the staging performance.
- If you run a mapping enabled for pushdown optimization, the mapping runs without pushdown optimization.
- If the data contains timestamp data types with time zone, the job runs without staging the data in the local flat file.
- When you use a custom query or SQL override with the Order By clause and the `AzureSynapseDisableStagingForSort` property is set to `true`, the mapping runs without staging the data in the local flat file.
- When you read source data with decimal values and write to a Microsoft Azure Synapse SQL target and enable the staging property to optimize the staging performance, the decimal values in the target get truncated and rounded off to the nearest integer.  
To avoid this issue, in the system configuration details section, select the Type as DTM for the Data Integration service, and set the value for `DisableInfraDoubleHandlingForStaging` as `yes`.

### Enabling the Microsoft Azure Synapse SQL Connector to optimize the staging performance

Perform the following tasks to enable the Microsoft Azure Synapse SQL Connector to optimize the staging performance:

1. In Administrator, click **Runtime Environments**.
2. Edit the Secure Agent for which you want to set the property.
3. In the **System Configuration Details** section, select the **Service** as **Data Integration Server** and the type as **Tomcat**.

4. To stage the data after reading from Microsoft Azure Synapse SQL, set the value of the Tomcat property *INFA\_DTM\_RDR\_STAGING\_ENABLED\_CONNECTORS* to the plugin ID of the Microsoft Azure Synapse SQL connector.

To stage the data before writing to Microsoft Azure Synapse SQL, set the value of the Tomcat property *INFA\_DTM\_STAGING\_ENABLED\_CONNECTORS* to the plugin ID of the Microsoft Azure Synapse SQL connector.

You can find the plugin ID in the manifest file located in the following directory: <Secure Agent installation directory>/downloads/<Microsoft Azure Synapse SQL package>/CCIManifest

The following image shows the property set for the Secure Agent:

Tomcat	INFA_DTM_STAGING_ENABLED_CONNECTORS	<input type="text" value="453100"/>
Tomcat	INFA_DTM_RDR_STAGING_ENABLED_CONNECTORS	<input type="text" value="453100"/>

To set the property for multiple connectors, specify the plugin ID for each connector separated by a comma.

When you run the mapping, the flat file is created in the following directory in your machine: C:\Windows\Temp\AzureSynapse\stage\<AzureSynapse\_Target.txt>

You can check the session logs. If the flat file is created successfully, Data Integration logs the following message in the session log:

The INFA\_DTM\_STAGING is successfully enabled to use the flat file to create local staging files.

## CHAPTER 2

# Connections for Microsoft Azure Synapse SQL

Create a Microsoft Azure Synapse SQL connection to securely read data from or write data to Microsoft Azure Synapse SQL. You can use a Microsoft Azure Synapse SQL connection to specify sources, targets, and lookups in mappings and mapping tasks. You can also use the Microsoft Azure Synapse SQL connection in an SQL transformation.

## Microsoft Azure Synapse SQL connection properties

When you set up a Microsoft Azure Synapse SQL connection, configure the connection properties.

The following table describes the Microsoft Azure Synapse SQL connection properties:

Property	Description
Connection Name	Name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + , Maximum length is 255 characters.
Description	Description of the connection. Maximum length is 4000 characters.
Type	The Microsoft Azure Synapse SQL connection type.
Runtime Environment	The name of the runtime environment where you want to run the tasks. You can specify a Secure Agent or a Hosted Agent.
Azure DW JDBC URL	The Microsoft Azure Synapse SQL JDBC connection string. Enter the connection string in the following format for Microsoft SQL Server authentication: <code>jdbc:sqlserver://&lt;Server&gt;.database.windows.net:1433;database=&lt;Database&gt;</code> Enter the connection string in the following format for Azure Active Directory (AAD) authentication: <code>jdbc:sqlserver://&lt;Server&gt;.database.windows.net:1433; database=&lt;Database&gt;;encrypt=true;trustServerCertificate=false; hostNameInCertificate=*.database.windows.net;loginTimeout=30; Authentication=ActiveDirectoryPassword;</code> Default is Microsoft SQL Server authentication.

Property	Description
Azure DW JDBC Username	User name to connect to the Microsoft Azure Synapse SQL account. Provide AAD user name for AAD authentication.
Azure DW JDBC Password	Password to connect to the Microsoft Azure Synapse SQL account.
Azure DW Schema Name	Name of the schema in Microsoft Azure Synapse SQL.
Azure DW Client ID	Required if you want to use the user-assigned managed identity for Managed Identity Authentication to connect to Microsoft Azure Synapse SQL. The client ID of the user-assigned managed identity. If the managed identity is system-assigned, leave the field empty.
Azure Storage Type	Type of Azure storage to stage the files. Select one of the following storage types: <ul style="list-style-type: none"> <li>- Azure Blob. Uses Microsoft Azure Blob Storage to stage the files.</li> <li>- ADLS Gen2. Uses Microsoft Azure Data Lake Storage Gen2 to stage the files.</li> </ul> Default is Azure Blob.
Authentication Type	Authentication type to connect to Azure storage to stage the files. Select one of the following options: <ul style="list-style-type: none"> <li>- Shared Key Authentication. Uses the account name and account key to connect to Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2.</li> <li>- Service Principal Authentication. Applies to Microsoft Azure Data Lake Storage Gen2. Uses the client ID, client secret, and tenant ID to connect to Microsoft Azure Data Lake Storage Gen2. To use Service Principal authentication, register an application in the Azure Active Directory, generate a client secret, and then assign the Storage Blob Contributor role to the application.</li> <li>- Managed Identity Authentication. Applies to Microsoft Azure Data Lake Storage Gen2. Select to authenticate using identities that are assigned to applications in Azure to access Azure resources in Microsoft Azure Data Lake Storage Gen2.</li> </ul>
Azure Blob Account Name	Applies to Shared Key Authentication for Microsoft Azure Blob Storage. Name of the Microsoft Azure Blob Storage account to stage the files.
Azure Blob Account Key	Applies to Shared Key Authentication for Microsoft Azure Blob Storage. The Microsoft Azure Blob Storage access key to stage the files.
Container Name	Applies to Microsoft Azure Blob Storage. The name of the container in the Azure Blob Storage account.
ADLS Gen2 Storage Account Name	Applies to Shared Key Authentication and Service Principal Authentication for Microsoft Azure Data Lake Storage Gen2. Name of the Microsoft Azure Data Lake Storage Gen2 account to stage the files.
ADLS Gen2 Account Key	Applies to Shared Key Authentication for Microsoft Azure Data Lake Storage Gen2. The Microsoft Azure Data Lake Storage Gen2 access key to stage the files.

Property	Description
Client ID	Applies to Service Principal Authentication and Managed Identity Authentication for Microsoft Azure Data Lake Storage Gen2. The client ID of your application. To use service principal authentication, enter the application ID or client ID for your application registered in the Azure Active Directory. To use managed identity authentication, enter the client ID for the user-assigned managed identity. If the managed identity is system-assigned, leave the field empty.
Client Secret	Applies to Service Principal Authentication for Microsoft Azure Data Lake Storage Gen2. The client secret for your application.
Tenant ID	Applies to Service Principal Authentication for Microsoft Azure Data Lake Storage Gen2. The directory ID or tenant ID for your application.
File System Name	Applies to Microsoft Azure Data Lake Storage Gen2. The name of the file system in the Microsoft Azure Data Lake Storage Gen2 account.
Blob End-point	Type of Microsoft Azure endpoints. Select one of the following endpoints: - core.windows.net. Connects to Azure endpoints. - core.usgovcloudapi.net. Connects to US Government Microsoft Azure Synapse SQL endpoints. - core.chinacloudapi.cn. Connects to Microsoft Azure Synapse SQL endpoints in the China region. Default is core.windows.net.
VNet Rule	Enable to connect to a Microsoft Azure Synapse SQL endpoint residing in a virtual network (VNet).

## Proxy server configuration

If your organization uses an outgoing proxy server to connect to the internet, the agent connects to Informatica Intelligent Cloud Services and the Microsoft Azure Synapse SQL endpoints through the proxy server.

You can configure the Secure Agent to use the proxy server on Windows and Linux. You can use the unauthenticated or authenticated proxy server.

You can use one of the following types of proxy servers:

- Unauthenticated proxy - Requires only the host and port address for configuration.
- Authenticated proxy - Requires the host address, port address, user name, and password for configuration.

You cannot configure the HTTPS authenticated proxy using the JVM options.

## Configure proxy server settings on Windows

You can configure the proxy server on a Windows machine either through the Secure Agent Manager or the `proxy.ini` file.

## Configure proxy server settings through the Secure Agent Manager

Use the Secure Agent Manager to configure the Secure Agent to use the proxy server on Windows.

1. In the Secure Agent Manager, click **Proxy**.
2. To enter proxy server settings, click **Use a proxy server**.
3. Configure the following fields:

Field	Description
Proxy Host	Required. Host name of the outgoing proxy server that the Secure Agent uses.
Proxy Port	Required. Port number of the outgoing proxy server.
User Name	User name to connect to the outgoing proxy server.
Password	Password to connect to the outgoing proxy server.

4. Click **OK**.  
The Secure Agent Manager restarts the Secure Agent to apply the settings.

## Configure proxy server settings through the proxy.ini file

You can set the proxy server details for the Secure Agent in the `proxy.ini` file.

1. Navigate to the following directory on the Secure Agent machine:  
`<Secure Agent installation directory>\Informatica Cloud Secure Agent\apps\agentcore\conf`
2. Open the `proxy.ini` file and specify the following parameters:

Parameter	Description
InfAgent.ProxyHost=	Required. Host name of the outgoing proxy server that the Secure Agent uses.
InfAgent.ProxyPort=	Required. Port number of the outgoing proxy server.
InfAgent.ProxyUser=	User name to connect to the outgoing proxy server.
InfAgent.ProxyPassword=	Password to connect to the outgoing proxy server.

3. Save the `proxy.ini` file.
4. Restart the Secure Agent.

## Configure proxy server settings on Linux

Use the shell command to configure proxy server settings for the Secure Agent on a Linux machine.

1. Navigate to the following directory:  
`<Secure Agent installation directory>/apps/agentcore/`
2. Run the following command:  
`./consoleAgentManager.sh configureProxy <proxy host> <proxy port> <user name> <password>`
3. Restart the Secure Agent.



## Bypassing the proxy server

When you connect to Microsoft Azure Data Lake Storage Gen2 or Microsoft Azure Blob Storage to stage files, you can bypass the proxy server settings configured for the Secure Agent.

Perform the following steps to bypass the proxy server:

1. Navigate to the following directory:

```
<Secure Agent installation directory>/apps/agentcore
```

2. Specify the following entry in the `proxy.ini` file:

```
InfraAgent.NonProxyHost=localhost|127.*|[\:1]|<accountname>.blob.core.windows.net |  
<accountname>.dfs.core.windows.net | login.microsoftonline.com| 169.254.169.254
```

To bypass the proxy server for service principal authentication, append `login.microsoftonline.com` to the entry.

To bypass the proxy server for managed identity authentication, append `169.254.169.254` to the entry.

3. Restart the Secure Agent.

## CHAPTER 3

# Mappings for Microsoft Azure Synapse SQL

When you configure a mapping, you describe the flow of data from the source to the target.

A mapping defines reusable data flow logic that you can use in mapping tasks.

When you create a mapping, you define the Source, Target, and Lookup transformations to represent a Microsoft Azure Synapse SQL object. Use the Mapping Designer in Data Integration to add the Source, Target, or Lookup transformations in the mapping canvas and configure the Microsoft Azure Synapse SQL source, target, and lookup properties.

You can use Monitor to monitor the jobs.

## Microsoft Azure Synapse SQL sources in mappings

In a mapping, you can configure a Source transformation to represent a single Microsoft Azure Synapse SQL object.

The following table describes the Microsoft Azure Synapse SQL properties that you can configure in a Source transformation:

Property	Description
Connection	<p>Name of the source connection.</p> <p>You can select an existing connection, create a new connection, or define parameter values for the source connection property.</p> <p>If you want to overwrite the parameter at runtime, select the <b>Allow parameter to be overridden at run time</b> option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.</p> <p>You can switch between a non-parameterized and a parameterized Microsoft Azure Synapse SQL connection. When you switch between the connections, the advanced property values are retained.</p>
Source Type	<p>Type of the source object.</p> <p>Select Single Object, Multiple Objects, Query, or Parameter.</p> <p>When you specify a custom SQL query as source type, the Secure Agent evaluates the properties in the following order to run a mapping:</p> <ol style="list-style-type: none"> <li>1. SQL Override</li> <li>2. Table Name Override</li> <li>3. Custom SQL</li> </ol> <p>You cannot parameterize a source query object at runtime in a mapping.</p>
Object	<p>The source object for the task.</p> <p>You can use external tables as Microsoft Azure Synapse SQL sources in mappings.</p>
Parameter	<p>Select an existing parameter for the source object or click <b>New Parameter</b> to define a new parameter for the source object. The <b>Parameter</b> property appears only if you select Parameter as the source type.</p> <p>If you want to overwrite the parameter at runtime, select the <b>Allow parameter to be overridden at run time</b> option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties. Ensure that the parameter file is in the correct format.</p>
Sort	Not applicable.

The following table describes the Microsoft Azure Synapse SQL source advanced properties that you can configure in a Source transformation:

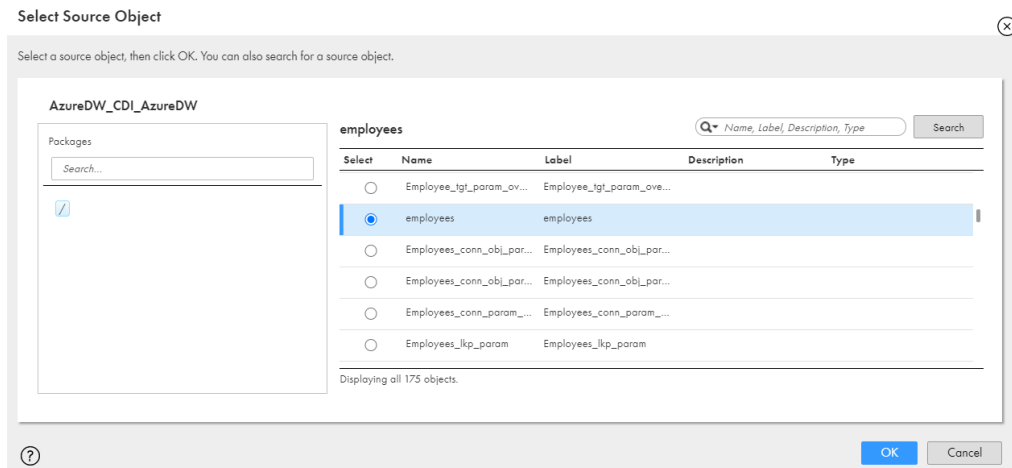
Property	Description
Azure Blob Container Name	Required if you select Azure Blob storage in the connection properties. The name of the container in Microsoft Azure Blob Storage. The container name must not contain special characters.
ADLS FileSystem Name	Required if you select ADLS Gen2 storage in the connection properties. The name of the file system in Microsoft Azure Data Lake Storage Gen2. The file system name must not contain special characters. You can also specify the path of the directory under the file system. Use only a forward slash to specify the directory path.
Schema Name Override	Overrides the schema specified in the connection.
Table Name Override	Overrides the table name of the imported Microsoft Azure Synapse SQL source table.
Field Delimiter	Character used to separate fields in the file. Default is 0x1e. You can specify 'TAB' or 0-256 single-char printable and non-printable ASCII characters. Non-printable characters must be specified in hexadecimal. <b>Note:</b> Multi-char ASCII characters except TAB are not applicable. You cannot use the following non-printable characters: 00x0, 0x0, 0x0A , 0x1B, 0x0D, and 0x1F
Number of Concurrent Connections to Blob Store	Number of concurrent connections to extract data from the Microsoft Azure Blob Storage. When reading a large-size blob, you can spawn multiple threads to process data. Configure <b>Blob Part Size</b> to partition a large-size blob into smaller parts. Default is 4. Maximum is 10.
Blob Part Size	Partitions a blob into smaller parts each of specified part size. When reading a large-size blob, consider partitioning the blob into smaller parts and configure concurrent connections to spawn required number of threads to process data in parallel. Default is 8 MB.
Pre-SQL	Pre-SQL command that must be run before reading data from the source.
Post-SQL	Post-SQL command that must be run after reading data from the source.
SQL Override	When you read data from a Microsoft Azure Synapse SQL object, you can configure SQL overrides and define constraints.

Property	Description
On Pre-Post SQL Error	<p>Determines the behavior when a task that includes pre-SQL or post-SQL commands encounters errors.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- Continue. The task continues regardless of errors.</li> <li>- Stop. The task stops when errors occur while executing pre-SQL or post-SQL commands.</li> </ul>
Quote Character	<p>Specifies the quote character to skip when you read data from Microsoft Azure Synapse SQL.</p> <p>The quote character that you specify must not exist in the source table. If it exists, enter a different quote character value.</p> <p>Default is <code>0x1f</code>.</p>
Interim Directory	<p>Optional. Path to the staging directory in the Secure Agent machine. Specify the staging directory where you want to stage the files when you read data from Microsoft Azure Synapse SQL. Ensure that the directory has sufficient space and you have write permissions to the directory.</p> <p>Default staging directory is <code>/tmp</code>.</p> <p>You cannot specify an interim directory when you use the Hosted Agent.</p>
Tracing Level	<p>Sets the amount of detail that appears in the log file. You can choose terse, normal, verbose initialization, or verbose data. Default is normal.</p>

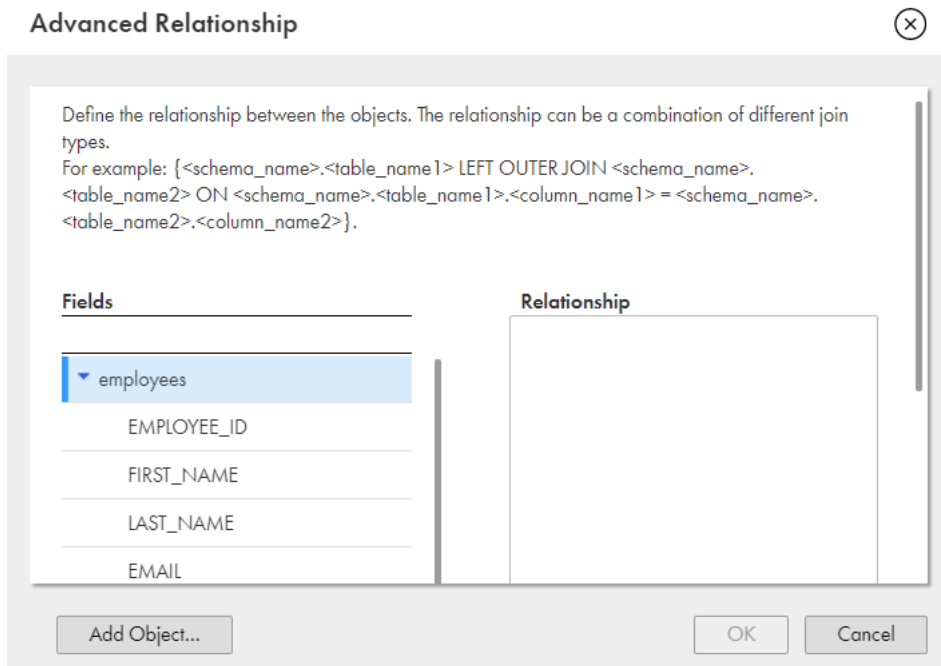
## Adding multiple source objects

Perform the following steps to add multiple source objects in a mapping:

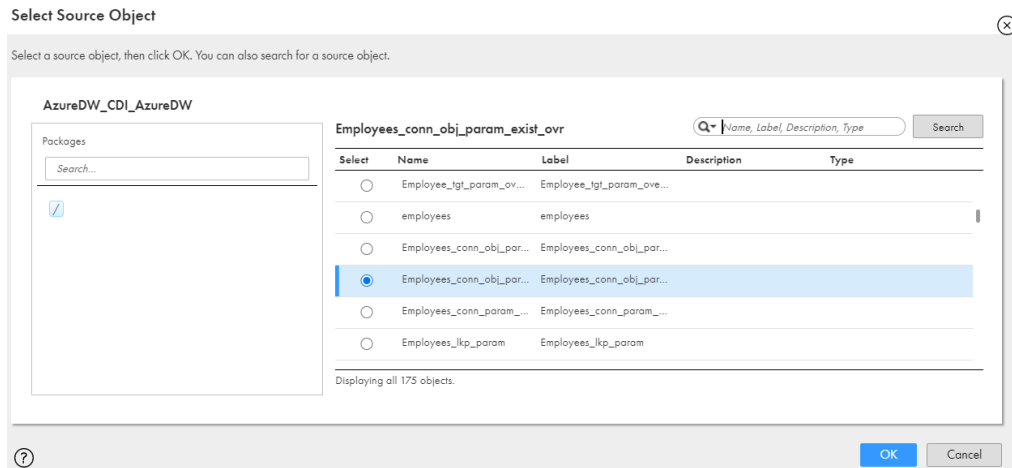
1. Click **New > Mappings**.
2. Select **Mapping** and click **Create**.
3. In the Source Properties page, specify the name and provide a description in the **General** tab.
4. Click the **Source** tab.
5. Select the source connection and source type as **Multiple Objects** to be used for the task.
6. In the **Objects and Relationships** section, click the arrow to open the **Action** menu and then select **Add Source Object**.
7. Select a source object from the list and click **OK**.



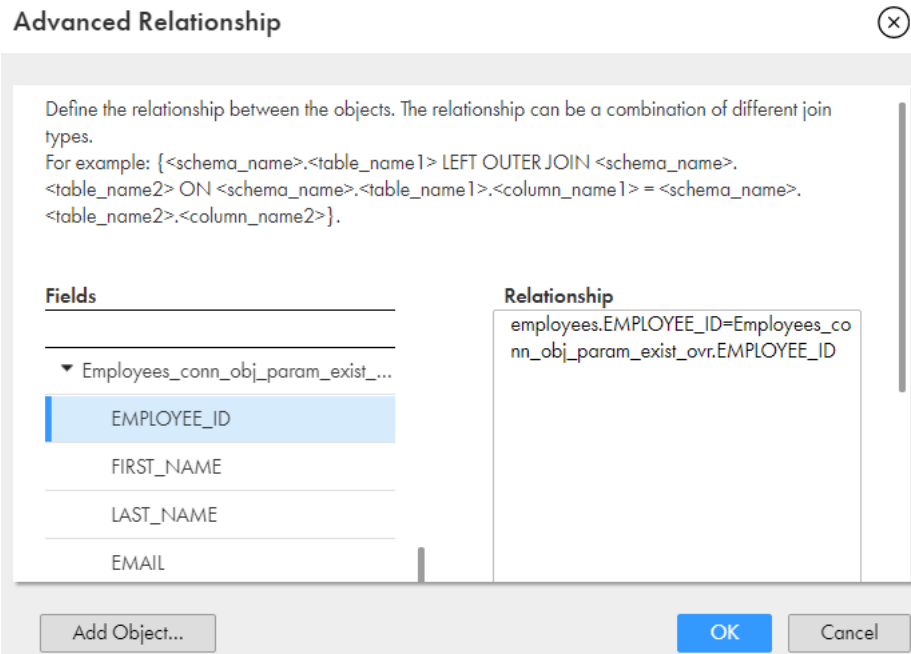
8. In the **Objects and Relationships** section, click the arrow next to the source object and then click **Advanced Relationship** to add related objects.
9. In the Advanced Relationship page, click **Add Object**.



10. In the Select Source Object page, select an object from the list and click **OK**.



11. In the Advanced Relationship page, select the required fields for the objects and create a join condition to define the relationship for all objects.



**Note:** You can join only two source tables. You must create an inner join to define the relationship. You must define only a join condition, not the complete join query. If the table name or column name is a reserved keyword or the table name contains special characters, enclose the name in square brackets. For example, [AUTHORIZATION].[BROWSE]=[int].[BROWSE].

12. Click **OK**.

# Microsoft Azure Synapse SQL targets in mappings

In a mapping, you can configure a Target transformation to represent a single Microsoft Azure Synapse SQL.

The following table describes the Microsoft Azure Synapse SQL properties that you can configure in a Target transformation:

Property	Description
Connection	<p>Name of the target connection. Select a target connection or click <b>New Parameter</b> to define a new parameter for the target connection.</p> <p>You can switch between a non-parameterized and a parameterized Microsoft Azure Synapse SQL connection. When you switch between the connections, the advanced property values are retained.</p>
Target Type	<p>Target type. Select one of the following types:</p> <ul style="list-style-type: none"><li>- Single Object</li><li>- Parameter. Select <b>Parameter</b> to define the target type when you configure the task.</li></ul>
Object	<p>Name of the target object.</p>
Parameter	<p>Select an existing parameter for the target object or click <b>New Parameter</b> to define a new parameter for the target object. The <b>Parameter</b> property appears only if you select Parameter as the target type.</p>
Create New at Runtime	<p>Creates a Microsoft Azure Synapse SQL target at runtime.</p> <p>Enter a name for the target object and select the source fields that you want to use. By default, all source fields are used.</p> <p>Select <b>Use Exact Source Field Names in Target</b> to write special characters from the source column names to the target.</p> <p>You cannot parameterize the target at runtime.</p>
Operation	<p>The target operation.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"><li>- Insert</li><li>- Update</li><li>- Upsert</li><li>- Delete</li><li>- Data Driven. Select to honor flagged rows for an insert, update, delete, or reject operation from the Update Strategy transformation or a CDC source.</li></ul> <p>You can also use the <b>Treat Source Rows As</b> advance property to perform insert, delete, update, or upsert operations.</p> <p>However, if you want to define a key for the delete, update, or upsert operation, you must use the Operation property.</p> <p>For more information about the operations, see <a href="#">“Rules and guidelines for mappings and mapping tasks” on page 37</a>.</p>



Property	Description
Data Driven Condition	Enables you to define expressions that flag rows for an insert, update, delete, or reject operation. <b>Note:</b> Appears only when you select Data Driven as the operation type. However, you may leave the field empty as the rows in the Update Strategy transformation and CDC source tables are already marked with the operation types.
Update Column	The key columns to upsert or update data to or delete data from Microsoft Azure Synapse SQL. This property is honored only if you select delete, update, or upsert operation in the <b>Treat Source Rows As</b> advance property. <b>Note:</b> This property appears only if you select delete, update, upsert, or data driven operation in the <b>Operation</b> property.

The following table describes the Microsoft Azure Synapse SQL properties that you can configure in a Target transformation:

Advanced Property	Description
Azure Blob Container Name	Required if you select Azure Blob storage in the connection properties. The name of the container in Microsoft Azure Blob Storage. The container name cannot contain special characters.
ADLS FileSystem Name	Required if you select ADLS Gen2 storage in the connection properties. The name of the file system in Microsoft Azure Data Lake Storage Gen2. The file system name cannot contain special characters. You can also specify the path of the directory under the file system. Use only a forward slash to specify the directory path.
Copy Method	The method to load data from the staging location to Microsoft Azure Synapse SQL. Select one of the following options: - Polybase - Copy Command Default is Polybase.
Copy Command Options	Options for the copy command in key=value format. Specify each option on a new line. For more information on copy command options, see <a href="#">"Copy command" on page 26</a> .
Schema Name Override	Overrides the schema specified in the connection.
Table Name Override	Overrides the table name of the imported Microsoft Azure Synapse SQL Data Warehouse target table.
Field Delimiter	Character used to separate fields in the file. Default is 0x1e. You can specify 'TAB' or 0-256 single-char printable and non-printable ASCII characters. Non-printable characters must be specified in hexadecimal. <b>Note:</b> Multi-char ASCII characters except TAB are not applicable. You cannot use the following non-printable characters: 00x0, 0x0, 0x0A, 0x1B, 0x0D, and 0x1F
Number of Concurrent Connections to Blob Storage	Number of concurrent connections to extract data from the Microsoft Azure Blob Storage. When reading a large-size blob, you can spawn multiple threads to process data. Default is 4. Maximum is 10.

Advanced Property	Description
Truncate Table	Truncates the target before inserting data to the target.
Pre-SQL	Pre-SQL command that must be run before reading data from the source.
Post-SQL	Post-SQL command that must be run after writing data to the target.
On Pre-Post SQL Error	Determines the behavior when a task that includes pre-SQL or post-SQL commands encounters errors. You can select any of the following options: <ul style="list-style-type: none"> <li>- Continue. The task continues regardless of errors.</li> <li>- Stop. The task stops when errors occur while executing pre-SQL or post-SQL commands.</li> </ul>
Treat Source Rows As	Select one of the following options: <ul style="list-style-type: none"> <li>- NONE</li> <li>- INSERT</li> <li>- DELETE</li> <li>- UPDATE</li> <li>- UPSERT</li> <li>- DATA DRIVEN. Select to honor the flagged rows from the update strategy or any other custom transformation, or a CDC source.</li> </ul> Default is None.
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 2000000.
Reject Threshold	Number of errors within a batch that causes a batch to fail. Enter a positive integer. If the number of errors is equal to or greater than the property value, the Secure Agent rejects the entire batch to the error file and marks the session failed. <b>Note:</b> When you do not set the reject threshold, the mapping fails when an error is encountered.
Quote Character	Specifies the quote character to skip when you write data to Microsoft Azure Synapse SQL. The quote character that you specify must not exist in the source table. If it exists, enter a different quote character value.
Compression Format	Compresses the staging files in the .Gzip format. Default is None.
Update Override	Overrides the default update SQL statement that the Secure Agent generates.
Interim Directory	Optional. Path to the staging directory in the Secure Agent machine. Specify the staging directory where you want to stage the files when you write data to Microsoft Azure Synapse SQL. Ensure that the directory has sufficient space and you have write permissions to the directory. Default staging directory is /tmp. You cannot specify an interim directory when you use the Hosted Agent.
Forward Rejected Rows	Determines whether the transformation passes rejected rows to the next transformation or drops rejected rows. By default, the mapping task forwards rejected rows to the next transformation.

## Copy command

You can use the copy command to load data from Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2 to Microsoft Azure Synapse SQL.

When you stage files in Microsoft Azure Data Lake Storage Gen2, you can use the copy command only with Service Principal Authentication.

You can specify the options for the copy command in key=value format in the Copy Command Options field. Specify each option in a new line.

The following image shows an example of the copy command options:

The screenshot shows a configuration window with a sidebar on the left containing tabs: General, Incoming Fields, Target (selected), Target Fields, and Field Mapping. The main area is titled 'Copy Command Options:' and contains several input fields:

- Copy Command Options:** A text area containing three lines: `FIRSTROW=11`, `IDENTITY_INSERT=OFF`, and `ERRORFILE=test/test2/test8/nir`.
- Schema Name Override:** An empty text box.
- Table Name Override:** An empty text box.
- Field Delimiter:** A text box containing `0x40`.
- Number of concurrent connections to Blob Store:** A dropdown menu with the value `4` selected.

You can configure only the following copy command options in the advanced target properties. The default value is considered for other copy command options supported by Microsoft Azure.

#### ERRORFILE

Specifies the directory where you want to write rejected rows and the corresponding error file. The ERRORFILE option is equivalent to the Reject Directory advanced target property.

Use the following format: `ERRORFILE=<Directory Location>`.

You can specify the absolute path or relative path for the directory. The error file path must not contain special characters.

If you specify an absolute path, for example, `Dir1/Dir2`, the agent creates the reject directory in the following path:

`Dir1/Dir2`

If you specify a relative path, for example, `/Dir1/Dir2`, the agent creates the reject directory in the following path:

`<staging path>/Dir1/Dir2`

If you do not specify the directory, the agent creates the reject directory in the following path:

`<staging path>/Reject_<UUID Randomly Generated>`

#### MAXERRORS

Specifies the maximum number of reject rows allowed in the load before the copy command is canceled. Each row that cannot be imported by the copy command is ignored and counted as one error. The MAXERRORS option is equivalent to the Reject Threshold advanced target property.

Use the following format: `MAXERRORS=max_errors`. For example, `MAXERRORS=20`.

## COMPRESSION

Specifies the data compression method for the data. You can use only Gzip compression for CSV files.

The COMPRESSION option is equivalent to the Compression Format advanced target property.

Use the following format: `COMPRESSION='method'`. For example, `COMPRESSION='GZIP'`.

## FIELDQUOTE

Specifies a single character that is used as the quote character in the CSV file.

The FIELDQUOTE option is equivalent to the Quote Character advanced target property.

Use the following format: `FIELDQUOTE='field_quote'`.

## FIELDTERMINATOR

Specifies the field terminator that is used in the CSV file.

The FIELDTERMINATOR option is equivalent to the Field Delimiter advanced target property.

Use the following format: `FIELDTERMINATOR='field_terminator'`. For example,  
`FIELDTERMINATOR='oxlf'`.

For more details about the copy command options, see the Microsoft Azure documentation.

## Rejected rows

The Secure Agent generates two error files in the Microsoft Azure Blob container specified in the target properties.

The error files are generated with the keyword *Reject* as a prefix in the error file names. One error file contains an entry for each rejected row and the other error file lists the cause for the rejected rows. To generate the error files, specify **Reject Threshold** in advanced target properties.

## Mapping tasks with CDC sources

Your organization needs to replicate real-time changed data from a mission-critical Oracle production system to minimize intrusive, non-critical work, such as offline reporting or analytical operations system. You can use Microsoft Azure Synapse SQL Connector to capture changed data from the Oracle CDC source and write the changed data to a Microsoft Azure Synapse SQL. Add the Oracle CDC sources in mappings, and then run the associated mapping tasks to write the changed data to the target.

1. In Data Integration, click **New > Mapping > Create**.  
The **New Mapping** dialog box appears.
2. Enter a name and description for the mapping.
3. On the Source transformation, specify a name and description in the general properties.
4. On the **Source** tab, select the configured Oracle CDC connection and specify the required source properties.
5. On the Target transformation, specify a name and description in the general properties.
6. On the **Target** tab, perform the following steps to configure the target properties:
  - a. In the **Connection** field, select the Microsoft Azure Synapse SQL connection.
  - b. In the **Target Type** field, select the type of the target object.
  - c. In the **Object** field, select the required target object.

- d. In the **Operation** field, select **Data Driven** to properly handle insert, update, and delete records from the source.
  - e. In the **Data Driven Condition** field, leave the field empty.
  - f. In the **Update Column** field, select the key columns to upsert or update data to or delete data from Microsoft Azure Synapse SQL. This property is honored only if you select delete, update, or upsert operation in the **Treat Source Rows As** advance property.
  - g. In the **Advanced Properties** section, you must select Data Driven in the **Treat Source Rows As** property.
7. On the **Field Mapping** tab, map the incoming fields to the target fields. You can manually map an incoming field to a target field or automatically map fields based on the field names.
  8. In the **Actions** menu, click **New Mapping Task**.  
The **New Mapping Task** page appears.
  9. In the **Definition** tab, enter the task name and select the configured mapping.
  10. In the **CDC Runtime** tab, specify the required properties.  
For more information about the **CDC Runtime** properties, see the *Informatica Cloud Data Integration Oracle CDC Connector User Guide*.
  11. In the **Schedule** tab, specify the following properties in the **Advanced Session Properties** section:
    - a. In the **Commit on End of File** field, select the value of the property as **No**.
    - b. In the **Commit Type** field, select the value of the property as **Source**.
    - c. In the **Recovery Strategy** field, select the value of the property as **Resume from last checkpoint**.
  12. Click **Save > Run** the mapping.  
Alternatively, you can create a schedule that runs the mapping task on a recurring basis without manual intervention. You can define the schedule to minimize the time between mapping task runs.
- In **Monitor**, you can monitor the status of the logs after you run the task.

## Changed data capture limitations

Consider the following limitations when working with a Microsoft Azure Synapse SQL change data capture (CDC) target:

- The staging temporary tables are not dropped if a mapping task is terminated. You must clean up the temporary tables manually.
- The following keywords are not allowed for the CDC target:
  - Keywords not allowed in the table name: "(", " SET ", " WHERE ", "."
  - Keywords not allowed in column name: ") VALUES", ",", " SET ", " WHERE ", " = ?", " AND "

**Note:** Consider the leading and trailing spaces with certain keywords.
- You cannot capture changed data for `datetimeoffset` and `time` data type.
- Date conversion from other data types to `smalldatetime` is not applicable.

## Bulk processing for write operations

You can enable bulk processing to write large amounts of data to Microsoft Azure Synapse SQL. Bulk processing utilizes minimal number of API calls and the performance of the write operation is optimized.

To enable bulk processing, specify the property `-DENABLE_WRITER_BULK_PROCESSING=true` in the Secure Agent properties:

Perform the following steps to configure bulk processing before you run a mapping:

1. In Administrator, select the Secure Agent listed on the **Runtime Environments** tab.
2. Click **Edit**.
3. In the **System Configuration Details** section, select **Data Integration Server** as the service and **DTM** as the type.
4. Edit the JVM option, and enter `-DENABLE_WRITER_BULK_PROCESSING=true`.
5. Click **Save**.

You cannot use bulk processing in mapping tasks configured with pushdown optimization using the Microsoft Azure Synapse SQL ODBC connection or the Microsoft Azure Synapse SQL connection.

## Rules and guidelines for configuring update override

Consider the following rules and guidelines when you use the update override property for a Microsoft Azure Synapse SQL target:

- Select **Update** for the **Treat source rows as** property in the advanced target properties.
- Specify the update query in the following format:

```
UPDATE <Target table name> SET <Target table name>.<Column1> = :TU.<Column1>, <Target table name>.<Column2> = :TU.<Column2>, ... <Target table name>.<ColumnN> = :TU.<ColumnN> FROM :TU WHERE <Target table name>.<Update Column1> = :TU.<Update Column1> AND <Target table name>.<Update Column2> = :TU.<Update Column2> AND ... <Target table name>.<Update ColumnN> = :TU.<Update ColumnN>
```
- Column names for :TU must match the target table column names.
- You must specify the update query with a valid SQL syntax because Microsoft Azure Synapse SQL Connector replaces :TU with a temporary table name and does not validate the update query.
- You cannot change the order of the column mappings using the update override.
- You cannot use unconnected fields to the target in the update query.

## Microsoft Azure Synapse SQL lookups in mappings

Use a Lookup transformation to retrieve data based on a specified lookup condition.

Use a Microsoft Azure Synapse SQL Lookup transformation to look up data in a Microsoft Azure Synapse SQL object. For example, the source table includes the customer code, but you want to include the customer name in the target table to make summary data easy to read.

You can use the Microsoft Azure Synapse SQL Lookup transformation to look up the customer name in another Microsoft Azure Synapse SQL object.

You can create the following lookups when you configure field mappings in a mapping task:

- Connected with cached or uncached
- Unconnected lookup
- Connected lookup with dynamic lookup cache

The following table describes the Microsoft Azure Synapse SQL lookup object properties that you can configure in a Lookup transformation:

Property	Description
Connection	<p>Name of the lookup connection.</p> <p>You can select an existing connection, create a new connection, or define parameter values for the lookup connection property.</p> <p>If you want to overwrite the parameter at runtime, select the <b>Allow parameter to be overridden at run time</b> option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.</p> <p>You can switch between a non-parameterized and a parameterized Microsoft Azure Synapse SQL connection. When you switch between the connections, the advanced property values are retained.</p>
Source Type	<p>Type of the source object.</p> <p>Select Single Object, Query, or Parameter.</p> <p>You cannot use custom query as the source type in cached and dynamic cache lookups.</p>
Lookup Object	Name of the lookup object for the mapping.
Parameter	<p>A parameter file where you define values that you want to update without having to edit the task. Select an existing parameter for the lookup object or click <b>New Parameter</b> to define a new parameter for the lookup object.</p> <p>The <b>Parameter</b> property appears only if you select parameter as the source type.</p> <p>If you want to overwrite the parameter at runtime, select the <b>Allow parameter to be overridden at run time</b> option when you create a parameter. When the task runs, the agent uses the parameters from the file that you specify in the task advanced session properties.</p>
Define Query	If the source type is a query, displays the <b>Edit Custom Query</b> dialog box. Enter a valid custom query and click <b>OK</b> .
Multiple Matches	<p>The behavior when the lookup condition returns multiple matches.</p> <p>Select one of the following options:</p> <ul style="list-style-type: none"> <li>- Return first row</li> <li>- Return last row</li> <li>- Return any row</li> <li>- Return all rows</li> <li>- Report error</li> </ul> <p><b>Note:</b> If the lookup table has an IDENTITY column, the data for the first row and last row is returned based on the sorting of the IDENTITY column. If the matched rows have duplicates then the values returned are random.</p>

You cannot use Binary, Varbinary, and Datetimeoffset datatypes columns as lookup conditions.

When you use a cached lookup with Microsoft Azure Synapse SQL connection and if the lookup condition contains a NULL value, the lookup condition is ignored.

The following table describes the Microsoft Azure Synapse SQL lookup advanced properties that you can configure in a Lookup transformation:

Property	Description
Azure Blob Container Name	Required if you select Azure Blob storage in the connection properties. The name of the container in Microsoft Azure Blob Storage. The container name must not contain special characters.
ADLS FileSystem Name	Required if you select ADLS Gen2 storage in the connection properties. The name of the file system in Microsoft Azure Data Lake Storage Gen2. The file system name must not contain special characters. You can also specify the path of the directory under the file system. Use only a forward slash to specify the directory path.
Schema Name Override	Overrides the schema specified in the connection.
Table Name Override	Overrides the table name of the imported Microsoft Azure Synapse SQL source table.
Field Delimiter	Character used to separate fields in the file. Default is 0x1e. You can specify 'TAB' or 0-256 single-char printable and non-printable ASCII characters. Non-printable characters must be specified in hexadecimal. <b>Note:</b> Multi-char ASCII characters except TAB are not applicable. You cannot use the following non-printable characters: 0x00, 0x01, 0x0A , 0x1B, 0x0D, and 0x1F
Number of Concurrent Connections to Blob Store	Number of concurrent connections to extract data from the Microsoft Azure Blob Storage. When reading a large-size blob, you can spawn multiple threads to process data. Configure <b>Blob Part Size</b> to partition a large-size blob into smaller parts. Default is 4. Maximum is 10.
Blob Part Size	Partitions a blob into smaller parts each of specified part size. When reading a large-size blob, consider partitioning the blob into smaller parts and configure concurrent connections to spawn required number of threads to process data in parallel. Default is 8 MB.
Pre-SQL	Pre-SQL command that must be run before reading data from the source.
Post-SQL	Post-SQL command that must be run after reading data from the source.
SQL Override	When you read data from a Microsoft Azure Synapse SQL object, you can configure SQL overrides and define constraints.



Property	Description
On Pre-Post SQL Error	Determines the behavior when a task that includes pre-SQL or post-SQL commands encounters errors. You can select any of the following options: <ul style="list-style-type: none"> <li>- Continue. The task continues regardless of errors.</li> <li>- Stop. The task stops when errors occur while executing pre-SQL or post-SQL commands.</li> </ul>
Quote Character	The Secure Agent skips the specified character when you read data from Microsoft Azure Synapse SQL. Default is <code>0x1f</code> .
Interim Directory	Optional. Path to the staging directory in the Secure Agent machine. Specify the staging directory where you want to stage the files when you read data from Microsoft Azure Synapse SQL. Ensure that the directory has sufficient space and you have write permissions to the directory. Default staging directory is <code>/tmp</code> . You cannot specify an interim directory when you use the Hosted Agent.
Tracing Level	Sets the amount of detail that appears in the log file. You can choose terse, normal, verbose initialization, or verbose data. Default is normal.

## Enabling lookup caching

When you configure a Lookup transformation in a mapping, you can cache the lookup data during the runtime session.

When you select **Lookup Caching Enabled**, Data Integration queries the lookup source once and caches the values for use during the session, which can improve performance. You can specify the directory to store the cached lookup.

### Lookup Cache Persistent

Use lookup cache persistent to save the lookup cache file to reuse it the next time Data Integration processes a Lookup transformation configured to use the cache.

You can specify the file name prefix to use with persistent lookup cache files in the **Cache File Name Prefix** field.

If the lookup table changes occasionally, you can enable the **Re-cache from Lookup Source** property to rebuild the lookup cache.

### Dynamic Lookup Cache

Use a dynamic lookup cache to keep the lookup cache synchronized with the target. By default, the dynamic lookup cache is disabled and represents static cache.

If the cache is static, the data in the lookup cache does not change as the mapping task runs.

If the task uses the cache multiple times, the task uses the same data. If the cache is dynamic, the task updates the cache based on the actions in the task, so if the task uses the lookup multiple times, downstream transformations can use the updated data.

### Pre-build Lookup Cache

Use a pre-build lookup cache to build the lookup cache before the Lookup transformation receives data. Multiple lookup cache files can be built at the same time to improve performance.

For information about lookup caching, see *Transformations* in the Data Integration documentation.

## Process SQL queries and stored procedures using an SQL transformation

You can configure an SQL transformation to process SQL queries and stored procedures midstream in a Microsoft Azure Synapse SQL mapping. When you add an SQL transformation, you define the database connection and the type of SQL that the transformation processes.

The SQL transformation can process the following types of SQL statements:

### Stored procedure

You can configure an SQL transformation to call a stored procedure in Microsoft Azure Synapse SQL. The stored procedure must exist in the Microsoft Azure Synapse SQL database before you create the SQL transformation. When the SQL transformation processes a stored procedure, it passes input parameters to the stored procedure. The stored procedure returns the values to the output fields of the transformation.

### SQL Query

You can configure an SQL transformation to process an entered query that you define in the SQL editor. The SQL transformation processes the query and returns the rows. The SQL transformation also returns any errors that occur from the underlying database or if there is an error in the user syntax.

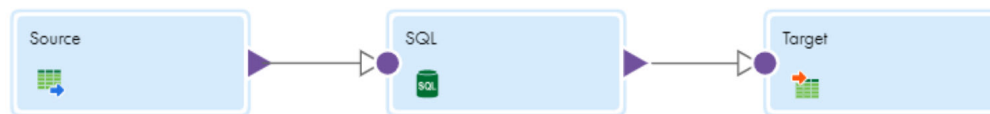
**Note:** Saved query type is not applicable.

For more information about SQL queries and stored procedures, see *Transformations* in the Data Integration documentation.

## Configuring an SQL transformation

Your mapping includes user IDs in the data flow. You want to include user names in addition to user IDs. You have a stored procedure that matches user IDs with user names in the database. You add an SQL transformation to your mapping, select the stored procedure, and map the `userId` incoming field with the `userId` input field in the stored procedure. Add an SQL transformation in a Microsoft Azure Synapse SQL mapping.

You check the Output Fields tab for the SQL transformation to confirm that it includes the `username` field. When you run the mapping, the `username` value is returned with the user ID.



This example lists the tasks required to configure an SQL transformation that calls a stored procedure in Microsoft Azure Synapse SQL:

1. Enter a name and description for the SQL transformation.
2. In the **Incoming Fields** tab, define field rules that determine the data to include in the transformation.
3. In the Properties panel of the SQL transformation, click the **SQL** tab.
4. In the **SQL** tab, perform the following tasks:
  - a. Select the connection to the database. You can select the connection or use a parameter.
  - b. Set the SQL type to **Stored Procedure**.
  - c. Click **Select** to select the stored procedure from the database, or enter the exact name of the stored procedure to call.  
The stored procedure name is case-sensitive.  
**Note:** If you add a new stored procedure to the database while you have the mapping open, the new stored procedure does not appear in the list of available stored procedures. To refresh the list, close and reopen the mapping.
5. In the Field Mapping tab, specify how to map incoming fields to the input fields of the selected stored procedure.
6. Define advanced properties for the transformation according to your requirement.

To configure an SQL transformation using the SQL type as SQL entered query, see *Transformations* in the Data Integration help.

## Rules and guidelines for SQL transformations

Consider the following rules and guidelines when you configure an SQL transformation in mappings:

- The RETURN\_VALUE parameter is not applicable for Microsoft Azure Synapse SQL Connector.
- A stored procedure returns output fields that contain only integer values. For example, fields of int, bigint, tinyint, smallint, and bit data type are returned as output fields.
- You cannot have input fields of datetimeoffset, binary, or varbinary data type in a stored procedure.
- When you configure a mapping to call a stored procedure in Microsoft Azure Synapse SQL, do not configure an override for the schema. The Secure Agent fails to fetch the metadata from the stored procedure when you specify an override to the schema name.
- You cannot read from a stored procedure whose name contains Unicode characters. The mapping fails with an SQL compilation error.
- Do not specify more than one SQL query in an SQL transformation. A mapping with multiple SQL queries fails.
- You cannot call a stored procedure from an entered query.
- If you want to parameterize the connection in an SQL transformation, perform the following steps:
  1. Select a valid non-parameterized connection for the SQL transformation.
  2. Configure the stored procedure or SQL query.
  3. Change the connection to a parameterized connection.
- When you use a parameterized connection for a SQL query, you cannot add new output fields.
- When the connection and the SQL query point to different schemas, the agent considers the schema specified in the SQL query.

# Dynamic schema handling

You can choose how Data Integration handles changes that you make to the data object schemas. To refresh the schema every time the mapping task runs, you can enable dynamic schema handling in the task.

A schema change includes one or more of the following changes to the data object:

- Fields added, deleted, or renamed.
- Fields updated for data type, precision, or scale.

Configure schema change handling on the **Schedule** page when you configure the task.

The following table describes the schema change handling options:

Option	Description
Asynchronous	Default. Data Integration refreshes the schema when you edit the mapping or mapping task, and when Informatica Intelligent Cloud Services is upgraded.
Dynamic	Data Integration refreshes the schema every time the task runs. You can choose from the following options to refresh the schema: <ul style="list-style-type: none"><li>- <b>Alter and apply changes.</b> Data Integration applies the following changes from the source schema to the target schema:<ul style="list-style-type: none"><li>- New fields: Alters the target schema and adds the new fields from the source.</li><li>- Renamed fields: Adds renamed fields as new columns in the target.</li><li>- Data type and precision updates. Applies these changes to the target.</li><li>- Deleted fields: Ignores deleted fields.</li></ul></li><li>- <b>Don't apply DDL changes.</b> Data Integration does not apply the schema changes to the target.</li><li>- <b>Drop current and recreate.</b> Drops the existing target table and then recreates the target table at runtime using all the incoming metadata fields from the source.</li></ul>

For more information, see the "Schema change handling" topic in *Tasks* in the Data Integration help.

## Rules and guidelines for dynamic schema handling

Consider the following rules and guidelines when you enable dynamic schema change handling:

- When you enable dynamic schema change handling, you cannot make changes to the columns of Uniqueidentifier data type.
- When you read data from or write data to a Microsoft Azure Synapse SQL and override the schema name or table name, do not use the **Alter and apply changes** or **Drop current and recreate** option.
- When you alter the datetime data type to another, Data Integration always maps the altered field to the `datetime` field in the Microsoft Azure Synapse SQL target.  
For example, if you alter the source `datetime` field to `datetime2`, Data Integration maps the `datetime2` field to `datetime` in the Microsoft Azure Synapse SQL target.
- When you read data from Microsoft Azure Synapse SQL, any table or column constraints in the source table are not retained in the target.

## IDENTITY column

You can use the IDENTITY column in mappings and mapping tasks.

Consider the following instructions when you use the IDENTITY column:

- You can use the IDENTITY column only for insert operation.
- Ensure that the metadata of the target object including the IDENTITY column matches the metadata fetched in an override.
- You cannot enable dynamic schema change handling for the IDENTITY column.
- You cannot define an IDENTITY column when you create a new target at runtime .
- The IDENTITY column is not highlighted in the imported objects.

## Parameterization

You can parameterize the connection, objects, and the advanced runtime properties in mappings.

Consider the following rules and guidelines when you use parameterization:

### General guidelines

- When you create a mapping with a parameterized target that you want to create at runtime, set the target field mapping to automatic.
- You cannot parameterize the field mapping.

### Mappings

When you use input parameters, specify the parameter name in the following format:

- When you use input parameters, specify the parameter name in the following format:
  - Format in a mapping task: \$name\$
  - Format in a parameter file: \$name
- When you use in-out parameters, specify the parameter name in the following format in a mapping task or a parameter file: \$\$name.

## Rules and guidelines for mappings and mapping tasks

Consider the following rules and guidelines for mappings and mapping tasks:

### General rules and guidelines

- When you use Microsoft Azure Data Lake Storage Gen2 or Microsoft Azure Blob Storage to stage files in Microsoft Azure Synapse SQL mapping, consider the following guidelines:
  - You can specify the Microsoft Azure Data Lake Storage Gen2 file system name and Microsoft Azure Blob Storage container name at the connection level for all mappings that use the Microsoft Azure Synapse SQL connection. You can override file system name and container name specified in the connection from the advanced source or target properties.
  - You must specify the file system name and container name either in the connection or in the advanced source or target properties. If you specify at both the places, the agent considers the values specified in the advanced properties.

- When you connect to a Microsoft Azure Synapse SQL case-sensitive database and if a mapping uses multiple sources or targets, ensure that the sources or targets do not have the same name.
- The session load summary in the session log is not captured for each commit interval.
- When you use an ODBC connection to connect to Microsoft Azure Synapse SQL, do not use the MERGE statement in the pre-SQL or post-SQL query. Else, the mapping task fails.
- When the source fields are empty, the Secure Agent writes the empty values as NULL in the target.
- When you use the \$\$\$SESSSTARTTIME variable in an SQL query, the variable returns the session start time as a string value.

Use the following syntax to convert the string values to datetime:

```
SELECT CAST(CONVERT(VARCHAR,SUBSTRING('$$$SESSSTARTTIME',0,20),113) as datetime)
```

Use the following syntax to convert the string values to date:

```
SELECT CAST(CONVERT(VARCHAR,SUBSTRING('$$$SESSSTARTTIME',0,20),113) as date)
```

- When you use a custom query, SQL override, or update override for Uniqueidentifier fields, you must convert the Uniqueidentifier data type to string data type. Else, the mapping fails.

For example, when you use `Select * from uid.NEWID_TEST2;`, the mapping fails.

Use the following SQL query:

```
Select CONVERT(CHAR(36),ID) AS ID, TESTCOLUMN from uid.NEWID_TEST2;
```

#### Source transformation

- You cannot preview data for binary or varbinary columns for Microsoft Azure Synapse SQL objects in a mapping.
- When read a JSON file from a Microsoft Azure Data Lake Storage Gen2 source and write to a Microsoft Azure Synapse SQL target created at runtime, ensure that the source data does not have null values.
- When you read data from multiple Microsoft Azure Synapse SQL sources, ensure that the table names for the source object and the related object are not the same, regardless of the name case.
- When you read external tables, ensure that two or more tables are not in the same location in Microsoft Azure Blob Storage or Microsoft Azure Data Lake Storage Gen2.
- When you read data from a source that contains varbinary columns, the Secure Agent reads the empty values from varbinary columns as NULL.
- Custom query as source
  - If the query starts with the WITH clause, ensure that there is no semicolon (;) before WITH.  
For example, `;With TEMP_OBJ AS (Select 'abc' AS FIELD1,123 AS Field2) Select Field1,Field2 from Temp_OBJ` is not a valid query.
  - Do not use a DECLARE statement in a source custom query.
  - The parameters in the query must have a default value that can be overridden at runtime.
  - When you use a custom query to read from multiple tables that have similar column names, you must use aliases for such column names while joining the tables.
  - When you read data from Microsoft Azure Synapse SQL, you must configure the following property in the JVM options of the Secure Agent:  
`-DAzureSynapseDisableStagingForSort=true`
  - When you write data to Microsoft Azure Synapse SQL, you cannot use the ORDER BY clause.

- If you use a custom query in a Lookup transformation, you must perform the following tasks:
  - Configure the `-DAzureSynapseDisableStagingForSort=true` property in the JVM options of the Secure Agent.
  - Enable the **Sorted Input** option in the advanced lookup properties.
  - The query must first contain the ORDER BY clause for the condition fields are included in the lookup condition in the ascending order and then the other fields in any desired order.
  - When you configure the Lookup transformation to return multiple rows, the **Return All Rows** option does not return the values in the expected order.

#### Target transformation

- Behavior of target operations

Operation property	Treat Source Rows As property	Behavior
Insert, delete, or update	None or data driven	Value of the Operation property is given precedence.
Upsert	None or data driven	Update operation is performed. In this case, you must set both the Operation property and the Treat Source Rows As property to Upsert to perform an upsert operation.
Insert, delete, update, or upsert	Insert, delete, update, or upsert	Value of Treat Source Rows As property is given precedence.

- When you use the copy command, adhere to the following rules:
  - When you stage files in Microsoft Azure Data Lake Gen2, you can use the copy command only with Service Principal Authentication.
  - You cannot completely parameterize a multi-line copy command using a parameter file.
- When you create a Microsoft Azure Synapse SQL target at run time, ensure that the field names in the source do not contain special characters.
- When you create a new target, the fields of the Uniqueidentifier data type in the source are converted to Nvarchar data type in the target.  
To retain the Uniqueidentifier data type, edit the metadata for the target fields and change the native type of the fields to Uniqueidentifier.
- When you write data to a Microsoft Azure Synapse SQL target and if the source table has columns of the Datetime or Datetime2 data type, the seconds value is rounded to the nearest value.
- When you write data to Microsoft Azure Synapse SQL and override the default update SQL statement, ensure that the parameter query does not exceed 600 characters.
- When you upsert or update data to a Microsoft Azure Synapse SQL target and multiple rows in the source table contain the same value for the column on which the primary key is defined, the first row with the value is correctly updated but the subsequent rows with the same value are incorrectly updated in the target.
- When you use an ODBC connection to connect to Microsoft Azure Synapse SQL and the target table has an IDENTITY column, you cannot update data in the target.

#### Lookup transformation

- When you use an uncached lookup, you cannot look up data with unicode characters.

- Guidelines for dynamic lookup cache
  - The Lookup transformation must be a connected transformation.
  - You can specify only an equal operator in a lookup condition.
  - You cannot use dynamic lookup cache in pushdown optimization.



## CHAPTER 4

# Pushdown optimization

You can use pushdown optimization to push the transformation logic to Microsoft Azure Synapse SQL.

Use pushdown optimization to improve task performance by using the database resources. When you run a task configured for pushdown optimization, the task converts the transformation logic to a PolyBase query or SQL query. The task sends the query to the database, and the database executes the query.

Configure pushdown optimization for a mapping in the tasks properties. Full pushdown optimization is enabled by default in mapping tasks.

## Pushdown optimization types

When you apply pushdown optimization, the task pushes transformation logic to the source or target database based on the optimization type you specify in the task properties.

You can configure the following types of pushdown optimization in a mapping:

### **None**

The task does not push down the transformation logic to the Microsoft Azure Synapse SQL database.

### **Full**

The task pushes as much of the transformation logic as possible to process in the target database.

### **Source**

The task pushes down as much as the transformation logic as possible to process in the source database. When you select source pushdown optimization, the task pushes the transformation logic for all the supported transformations downstream in the mapping.

## Data Integration behavior with source and full pushdown optimization

When you select full or source pushdown optimization for a mapping that reads from or writes to Microsoft Azure Synapse SQL, Data Integration analyzes the mapping from the source to the target or until it reaches a downstream transformation to determine whether to push down only partial or the entire mapping logic for processing to the database.

If all the transformations in the mapping are supported for pushdown optimization, the task pushes down the entire mapping logic to the database.

When a transformation is not supported in the mapping, the task partially pushes down the mapping logic to the point where the transformation is supported for pushdown optimization. Data Integration generates and

executes a SELECT statement for the transformation logic that needs to be pushed down. Then, it reads the results of this SQL query and processes the remaining transformations in the mapping.

## Pushdown optimization scenarios

You can configure pushdown optimization using the Microsoft Azure Synapse SQL Connector in mappings.

You can use configure pushdown optimization for the following scenarios:

Source and target endpoints	Supported pushdown scenarios in mappings	Pushdown optimization type
Microsoft Azure Synapse SQL source Microsoft Azure Synapse SQL target	Reads from and writes to Microsoft Azure Synapse SQL using the Microsoft Azure Synapse SQL connection.	Source, Full
Microsoft Azure Data Lake Storage Gen2 source Microsoft Azure Synapse SQL target	Reads from Microsoft Azure Data Lake Storage Gen2 using a Microsoft Azure Data Lake Storage Gen2 connection and writes to Microsoft Azure Synapse SQL using a Microsoft Azure Synapse SQL connection.	Full

**Note:** You can also configure pushdown for a mapping that uses a Microsoft Azure Synapse SQL ODBC connection to read from and write to Microsoft Azure Synapse SQL. Informatica recommends that you use the Microsoft Azure Synapse SQL connection in mappings to configure pushdown optimization. If you cannot push specific transformation logic using the Microsoft Azure Synapse SQL connection, you can explore configuring pushdown optimization using the Microsoft Azure Synapse SQL ODBC connection.

## Pushdown optimization using a Microsoft Azure Synapse SQL connection

You can configure pushdown optimization for a mapping that contains a Microsoft Azure Synapse SQL connection. Pushdown optimization enhances the mapping performance.

Use the **Pushdown Optimization** advanced session property to configure pushdown optimization in a mapping task.

When you run a task that reads from Microsoft Azure Data Lake Storage Gen2 and writes to Microsoft Azure Synapse SQL with pushdown optimization, the task converts the transformation logic to a PolyBase query.

When you can configure pushdown optimization in a mapping to read from and write to Microsoft Azure Synapse SQL using a Microsoft Azure Synapse SQL connection, you can read data from tables, external tables, custom queries, and views.

## Previewing pushdown optimization

Before you can run a mapping task configured for pushdown optimization, you can preview if pushdown optimization is possible when you create the mapping. You can preview pushdown optimization from the **Pushdown Optimization** panel in the Mapping Designer.

After you select the required pushdown optimization options and run the preview, Data Integration creates and runs a temporary pushdown preview mapping task. When the job completes, Data Integration displays the SQL queries to be executed and any warnings in the **Pushdown Optimization** panel. The warning messages help you understand which transformations in the configured mapping are not applicable for pushdown optimization. If pushdown optimization fails, Data Integration lists any queries generated up to the point of failure. You can edit the mapping and fix the required transformations before you run the mapping for pushdown optimization.

You can also view the temporary job created under **My Jobs** and download the session log to view the queries generated.

For more information about how to preview pushdown optimization, see the topic "Pushdown optimization preview" in *Mappings* in the Data Integration help.

## Configuring optimization for a Microsoft Azure Synapse SQL mapping task

Perform the following steps to configure pushdown optimization for a Microsoft Azure Synapse SQL mapping task:

1. Create a mapping based on your requirement:
  - To read from Microsoft Azure Data Lake Storage Gen2 and write to Microsoft Azure Synapse SQL, use the Microsoft Azure Data Lake Storage Gen2 connection in the Source transformation and Microsoft Azure Synapse SQL connection in the Target transformation.
  - To read from or write to Microsoft Azure Synapse SQL, use the Microsoft Azure Synapse SQL connection in the Source and Target transformations.
2. Create a mapping task.
  - a. Select the configured mapping.
  - b. In the **Pushdown Optimization** section on the Schedule tab, set the pushdown optimization value to **Full** or **To Source**.

Full pushdown optimization is enabled by default in mapping tasks.

You can use Source pushdown optimization only when you read from a Microsoft Azure Synapse SQL source.
  - c. To cancel the task when pushdown optimization does not work, select **If pushdown mode is not possible, cancel the task**. If you do not select this option, the task runs without pushdown optimization.

**Note:** The Optimization context type option is not applicable for a Microsoft Azure Synapse SQL mapping task.
  - d. Save the task and click **Finish**.

When you run the mapping task, the transformation logic is pushed to the Microsoft Azure Synapse SQL. To verify that the pushdown optimization has taken place, you can check the session log for the job. You can monitor the jobs that you initiated on the **My Jobs** page. You choose to stop or clean stop a job. When you use clean stop, Data Integration terminates all the issued statements and processes spawned by the job.

## Supported functions

You can push functions in a transformation to Microsoft Azure Synapse SQL using pushdown optimization.

When you use pushdown optimization, the Secure Agent converts the expression, filter, or aggregator in the transformation by determining equivalent functions in the database.

The following table summarizes the availability of pushdown functions when you configure pushdown optimization for a mapping that contains a Microsoft Azure Synapse SQL connection.

Columns marked with an X indicate that the function can be pushed to Microsoft Azure Synapse SQL. Columns marked with a dash (-) symbol indicate that the function cannot be pushed to the database.

Function	Pushdown	Function	Pushdown	Function	Pushdown
ABORT()	-	INSTR()	X	REPLACECHR()	X
ABS()	X	IS_DATE()	X	REPLACESTR()	X
ADD_TO_DATE()	X	IS_NUMBER()	X	REVERSE()	-
AES_DECRYPT()	-	IS_SPACES()	X	ROUND(DATE)	-
AES_ENCRYPT()	-	ISNULL()	X	ROUND(NUMBER)	X
ASCII()	X	LAST()	X	RPAD()	X
AVG()	X	LAST_DAY()	X	RTRIM()	X
CEIL()	X	LEAST()	-	SESSSTARTTIME	X
CHOOSE()	-	LENGTH()	X	SET_DATE_PART()	X
CHR()	X	LN()	X	SHA-256 ()	X
CHRCODE()	-	LOG()	X	SIGN()	X
COMPRESS()	-	LOOKUP	-	SIN()	X
CONCAT()	X	LOWER()	X	SINH()	X
COS()	X	LPAD()	X	SOUNDEX()	X
COSH()	-	LTRIM()	X	SQRT()	X
COUNT()	X	MAKE_DATE_TIME()	X	STDDEV()	X
CRC32()	-	MAX()	X	SUBSTR()	X
CUME()	-	MD5()	-	SUM()	X
DATE_COMPARE()	-	MEDIAN()	-	SYSDATE()	X
DATE_DIFF()	X	METAPHONE()	-	SYSTIMESTAMP()	X
DECODE()	X	MIN()	X	TAN()	X
DECODE_BASE64()	-	MOD()	X	TANH()	X

Function	Pushdown	Function	Pushdown	Function	Pushdown
DECOMPRESS()	-	MOVINGAVG()	-	TO_BIGINT	X
ENCODE_BASE64()	-	MOVINGSUM()	-	TO_CHAR(DATE)	X
EXP()	X	NPER()	-	TO_CHAR(NUMBER)	X
FIRST()	X	PERCENTILE()	-	TO_DATE()	X
FLOOR()	X	PMT()	-	TO_DECIMAL()	X
FV()	-	POWER()	X	TO_FLOAT()	X
GET_DATE_PART()	X	PV()	-	TO_INTEGER()	X
GREATEST()	-	RAND()	-	TRUNC(DATE)	X
IIF()	X	RATE()	-	TRUNC(NUMBER)	X
IN()	X	REG_EXTRACT()	-	UPPER()	X
INDEXOF()	-	REG_MATCH()	-	VARIANCE()	X
INITCAP()	-	REG_REPLACE	-	-	-

The following table describes the syntax for the pushdown functions:

Function	Syntax	Description
ADD_TO_DATE	ADD_TO_DATE (date, format, amount)	<b>Supported formats:</b> <ul style="list-style-type: none"> <li>- yyyy</li> <li>- mm</li> <li>- dd</li> <li>- hh</li> <li>- mi</li> <li>- ss</li> <li>- ms</li> <li>- us</li> <li>- ns</li> </ul>
DATE_DIFF	DATE_DIFF (date1, date2, format)	<b>Supported formats:</b> <ul style="list-style-type: none"> <li>- yyyy</li> <li>- mm</li> <li>- dd</li> <li>- hh</li> <li>- mi</li> <li>- ss</li> <li>- ms</li> <li>- us</li> <li>- ns</li> </ul>
FIRST	FIRST (value)	

Function	Syntax	Description
GET_DATE_PART	GET_DATE_PART (date, format)	Supported formats: <ul style="list-style-type: none"> <li>- YYYY</li> <li>- mm</li> <li>- dd</li> <li>- hh</li> <li>- mi</li> <li>- ss</li> <li>- ms</li> <li>- us</li> <li>- ns</li> </ul>
IIF	IIF (condition, value1 [,value2])	
IN	<ul style="list-style-type: none"> <li>- <b>Aggregator and Expression transformation</b>  IIF(IN(search_value, value1, [value2, ..., valueN,]), 'TRUE', 'FALSE')</li> <li>- <b>Filter transformation</b>  IN(search_value, value1, [value2, ..., valueN,])</li> </ul>	
INSTR	INSTR (string, %search_value%)	The search value must be a regular expression enclosed in %.
IS_DATE	<ul style="list-style-type: none"> <li>- <b>Aggregator and Expression transformation</b> <ul style="list-style-type: none"> <li>- IS_DATE(column name)</li> <li>- IIF(IS_DATE(colname), value1 [,value2])</li> <li>- IIF(IS_DATE(colname)=0 or 1, value1 [,value2])</li> </ul> </li> <li>- <b>Filter and Router transformation</b> <ul style="list-style-type: none"> <li>- IS_DATE(column name)=0 or 1</li> <li>- IIF ( IS_DATE(colname), 0 or 1, 0 or 1)</li> <li>- IIF(IS_DATE(column name)=0 or 1, 0 or 1, 0 or 1)</li> </ul> </li> </ul>	
ISNULL	<ul style="list-style-type: none"> <li>- <b>Aggregator and Expression transformation</b> <ul style="list-style-type: none"> <li>- IIF(ISNULL(colname), value1 [,value2])</li> <li>- IIF(ISNULL(colname)=0 or 1, value1 [,value2] )</li> </ul> </li> <li>- <b>Filter and Router transformation</b> <ul style="list-style-type: none"> <li>- ISNULL(column_name)</li> <li>- IIF(ISNULL(column name), 0 or 1, 0 or 1)</li> </ul> </li> </ul>	
IS_NUMBER	<ul style="list-style-type: none"> <li>- <b>Aggregator and Expression transformation</b> <ul style="list-style-type: none"> <li>- IS_NUMBER(column name)</li> <li>- IIF(IS_NUMBER(colname), value1 [,value2])</li> <li>- IIF(IS_NUMBER(colname)=0 or 1, value1 [,value2])</li> </ul> </li> <li>- <b>Filter and Router transformation</b> <ul style="list-style-type: none"> <li>- IS_NUMBER(column name)=0 or 1</li> <li>- IIF(IS_NUMBER(column name), 0 or 1, 0 or 1)</li> <li>- IIF(IS_NUMBER(column name)=0 or 1, 0 or 1, 0 or 1)</li> </ul> </li> </ul>	

Function	Syntax	Description
IS_SPACES	<ul style="list-style-type: none"> <li>- <b>Aggregator and Expression transformation</b> <ul style="list-style-type: none"> <li>- IS_SPACES(column name)</li> <li>- IIF(IS_SPACES(colname), value1 [,value2])</li> <li>- IIF(IS_SPACES(colname)=0 or 1, value1 [,value2])</li> </ul> </li> <li>- <b>Filter and Router transformation</b> <ul style="list-style-type: none"> <li>- IS_SPACES(colname) = 0 or 1</li> <li>- IIF (IS_SPACES(colname), 0 or 1, 0 or 1)</li> <li>- IIF(IS_SPACES(column name)=0 or 1, 0 or 1, 0 or 1)</li> </ul> </li> </ul>	
LAST	LAST (value)	
LAST_DAY	LAST_DAY (date)	
LTRIM and RTRIM	LTRIM (string) RTRIM (string)	You can pass only a single argument in the LTRIM or RTRIM function.
TANH	Use the EXP() function in the following formula to calculate TANH: $\text{TANH}(x) = (\text{EXP}(2 * x) - 1) / (\text{EXP}(2 * x) + 1)$	Maximum value of x: 354 Minimum value of x: -354 If the value of x is greater than 354 or less than -354, use the following expression: <pre> CASE WHEN X &gt; 354 THEN 1 WHEN X &lt; -354 THEN -1 ELSE TANH(X) END </pre>
TO_BIGINT	TO_BIGINT (numeric value) TO_BIGINT truncates the decimal portion. To avoid truncation, use the ROUND function. TO_BIGINT (ROUND (numeric value [,0]))	You can pass only a single argument in the function.
TO_CHAR (Date)	TO_CHAR (date [,format]) Specify a supported format. The format defines the format of the return value, not the format for the values in the date argument.	Supported formats: <ul style="list-style-type: none"> <li>- dd/mm/yyyy</li> <li>- dd.mm.yyyy</li> <li>- dd-mm-yyyy</li> <li>- dd mon yyyy</li> <li>- mm/dd/yyyy</li> <li>- hh:mi:ss</li> <li>- yyyy.mm.dd</li> <li>- yyyy/mm/dd</li> <li>- yyyymmdd</li> <li>- yyyy-mm-dd hh:mi:ss</li> </ul>

Function	Syntax	Description
TO_DATE	TO_DATE (string [,format]) The format must match the parts of the string argument.	Supported formats: <ul style="list-style-type: none"> <li>- dd/mm/yyyy</li> <li>- dd.mm.yyyy</li> <li>- dd-mm-yyyy</li> <li>- dd mon yyyy</li> <li>- mm/dd/yyyy</li> <li>- hh:mi:ss</li> <li>- yyyy.mm.dd</li> <li>- yyyy/mm/dd</li> <li>- yyyymmdd</li> <li>- yyyy-mm-dd hh:mi:ss</li> </ul>
SUBSTR	SUBSTR (string, start [,length ])	The start argument must be a positive number. The length must be an integer greater than 0.

## Supported operators

When you use pushdown optimization, the Secure Agent converts the expression in the transformation by determining equivalent operators in the database. If there is no equivalent operator, the Secure Agent processes the transformation logic.

The following table summarizes the availability of pushdown operators when you configure pushdown optimization for a mapping that contains a Microsoft Azure Synapse SQL connection.

Columns marked with an X indicate that the operator can be pushed to Microsoft Azure Synapse SQL.

Operator	Pushdown
+ - * /	X
%	X
	-
= > < >= <= <>	X
!=	X
^=	-
NOT AND OR	X

## Supported transformations

The following table lists the transformation logic that is supported by pushdown optimization.



Columns marked with an X indicate that the transformation can be pushed to Microsoft Azure Synapse SQL. Columns marked with a dash (-) symbol indicate that the transformation cannot be pushed to the database.

Transformation	Pushdown
Aggregator	X
Expression	X
Filter	X
Joiner	X
Lookup	X
Normalizer	X
Rank	X
Router	X
Sequence Generator	-
Sorter	-
SQL	X
Union	X

## Expression transformation

You can configure full pushdown optimization to push an Expression transformation to process in Microsoft Azure Synapse SQL.

You can add an Expression transformation to each of the sources in the mapping, followed by a join downstream in the mapping. Additionally, you can add multiple Expression transformations that branch out from a transformation and then branch in into a transformation downstream in the mapping.

When you configure an Expression transformation, consider the following rules to include variables in the expression:

- You cannot use variables where you are using the value assigned while processing a previous row for calculations in the current row. If you do, the mapping runs without pushdown optimization.
- The variables can be nested, but you cannot refer to a variable before it is defined in the expression. If the variables are not defined in that order, the mapping runs without pushdown optimization.

For example,

```
var: AGEPLUS2 = AGEPLUS1 + 1
var: AGEPLUS1 = AGE + 1
out: NEXTAGE = AGEPLUS2 + 1
```

Here, AGE + 1 is defined later. AGEPLUS2 in the first variable refers to AGEPLUS1 and remains unresolved. To resolve this, specify the variables in the following order:

```
var: AGEPLUS1 = AGE + 1
var: AGEPLUS2 = AGEPLUS1 + 1
out: NEXTAGE = AGEPLUS2 + 1
```

- The variables cannot have an expression that is cyclic or refers to itself:

For example,

```
var: AGEPLUS1 = AGEPLUS2 + 1
var: AGEPLUS2 = AGEPLUS1 + 1
out: NEXTAGE= AGEPLUS2
```

Here, AGEPLUS1 refers to AGEPLUS2 and remains unresolved.

## Lookup transformation

You can configure full pushdown optimization to push a Lookup transformation to process in Snowflake. You can push both a connected and an unconnected lookup.

Consider the following rules when you configure lookups:

- When you look up data and the lookup condition finds multiple matches, the lookup returns all rows. In a mapping with Microsoft Azure Synapse SQL as target, set the **Multiple Matches** option for the lookup object to **Return all rows**. If you enabled **Multiple Matches** to any option other than **Return all rows**, the agent ignores it.
- When you lookup data with null values in a Microsoft Azure Synapse SQL source and write to the target, there is a difference in the number of records in the target for a mapping that runs with and without pushdown optimization.
- When you read from and write to Microsoft Azure Synapse SQL, you cannot use filters in a Lookup transformation.

## SQL transformation

You can configure an SQL transformation to process SQL queries in a Microsoft Azure Synapse SQL mapping enable for pushdown optimization.

### Use functions to run queries

You can include functions in an entered query in an SQL transformation and run queries with the Microsoft Synapse SQL target endpoint. You must use only the SELECT clause SQL statement to push a function. Specify the column name in the select query or function.

Do not use FROM or WHERE clause in the SQL statement. For example, do not push functions using statements like "SELECT \* FROM TABLE".

You can use the following functions in an entered query:

- CURRENT\_USER
- HAS\_DBACCESS
- RAND
- SESSION\_USER
- SIGN
- SUSER\_NAME
- SUSER\_SNAME
- SYSTEM\_USER
- USER
- USER\_NAME

For more information about the supported functions, see the Microsoft Azure Synapse SQL documentation.

### User defined functions

You can configure a custom query in an SQL transformation to read from SQL user-defined functions (UDF) in Microsoft Azure Synapse SQL. You cannot read UDFs that have newline characters in the UDF name.

## Rules and guidelines

When you configure an SQL transformation, consider the following rules and guidelines:

- You cannot use an unconnected SQL transformation.
- You can use queries that are static. Dynamic queries are not applicable.
- You cannot use functions that return multiple rows.
- You cannot use the **Track the number of rows affected by this query** option for the output fields.
- If the mapping fails, the SQL error is not logged.
- You cannot use the **Max output row count** and **Continue on SQL Error within row** options in the advanced properties.

## Supported features for Microsoft Azure Synapse SQL mappings

When you configure pushdown optimization, the mappings support the following Microsoft Azure Synapse SQL properties in the Source, Lookup, and Target transformations:

If you configure properties that are not supported, the mapping runs with pushdown optimization but the properties are ignored. If the pushdown query generation fails, the mapping runs without pushdown optimization.

### Source

When you configure pushdown optimization, the mappings support the following properties for a Microsoft Azure Synapse SQL source:

- Source type - Single object, multiple objects, query, and parameter.
- Parameter - Allow source object to be overridden at run time
- Query options - Filter. Supports both simple and advanced filter conditions.
- Schema Name Override
- Table Name Override
- Pre-SQL
- Post-SQL
- SQL Override
- On Pre-Post SQL Error
- Edit Metadata

### Target

When you configure pushdown optimization, the mappings support the following properties for a Microsoft Azure Synapse SQL target:

- Target type - Single object and parameter.
- Target object - Existing target, Create new at runtime.
- Operation - Insert, Update, Upsert, Delete, Data driven
- Azure Government endpoints
- Schema Name Override
- Table Name Override
- Truncate Table
- Pre-SQL

- Post-SQL
- On Pre-Post SQL Error
- Treat Source Rows As
- Reject Threshold. Applicable only when you read from a Microsoft Azure Data Lake Storage Gen2 source.
- Edit metadata when you create target at the runtime
- Update override

### Lookup

When you configure pushdown optimization, the mappings support the following properties for Microsoft Azure Synapse SQL connected and unconnected lookup:

- Source type
  - Single Object - table, views, external table
  - Custom Query
  - Parameter
- Multiple matches
  - Connected lookup - Return all rows
  - Unconnected lookup - Report error
- Parameter - Allow source object to be overridden at run time
- Schema Name Override
- Table Name Override
- Pre-SQL
- Post-SQL
- SQL Override
- On Pre-Post SQL Error
- Edit Metadata
- Dynamic Lookup Cache for connected lookup
- Output Old Value On Update for connected lookup
- Insert Else Update for connected lookup
- Pre-build Lookup Cache - Auto, Value

**Note:** When a mapping runs without pushdown, the Azure Blob Container Name or ADLS FileSystem Name are required fields. When you run a mapping with Source pushdown optimization, Azure Blob Container Name is a required field.

## Supported features for Microsoft Azure Data Lake Storage Gen2 source

When you configure pushdown optimization, the mappings support the following properties for a Microsoft Azure Data Lake Storage Gen2 source:

### Source

- Azure Government endpoints

- Source Type
  - Single object
  - Parameter
- Format Type
  - Delimited
  - ORC
  - Parquet
- Formatting options
  - Delimiter
  - Qualifier
  - First Data Row
  - Import from schema file
- Filesystem Name Override
- Directory Override
- File Name Override
- Compression Format
- Edit metadata

**Note:** You can also read data from a Microsoft Azure Data Lake Storage Gen2 case-sensitive database.

#### Lookup

When you configure pushdown optimization, the mappings support the following properties for Microsoft Azure Data Lake Storage Gen2 connected and unconnected lookup:

- Azure Government endpoints
- Source type
  - Single Object
  - Parameter
- Multiple matches
  - Connected lookup - Return all rows
  - Unconnected lookup - Report error
- Format Type
  - Delimited
  - ORC
  - Parquet
- Formatting options
  - Delimiter
  - Qualifier
  - First Data Row
  - Import from schema file
- Filesystem Name Override
- Directory Override

- File Name Override
- Compression Format
- Edit metadata

## Configuring a custom query for the Microsoft Azure Synapse SQL source object

You can push down a custom query to Microsoft Azure Synapse SQL.

Before you run a task that contains a custom query as the source object, you can set the **Create Temporary View** property in the mapping task properties.

Perform the following task to set the property:

1. In the mapping task, navigate to the **Pushdown Optimization** section on the **Schedule** tab.
2. Set the **Create Temporary View** option based on your requirement.
  - When the **Create Temporary View** option is disabled, the task incorporates the custom query as a sub-query in the pushdown query.
  - When the **Create Temporary View** option is enabled, the task creates a temporary view based on the provided custom query and then generates the pushdown query with the temporary view.
3. Click **Finish**.

## Rules and guidelines for pushdown optimization

Certain rules and guidelines apply for pushdown optimization to a Microsoft Azure Synapse SQL database.

### General rules and guidelines

Consider the following rules and guidelines when you enable a mapping for pushdown optimization:

- When you use Rank transformation, the data is not stored as per the Rankindex order in the database.
- When you read varchar or nvarchar fields and use the Rank transformation, the target data is sorted as per collation of the database.
- When you use Rank transformation, the session sort order that you specify in the advanced session property is not considered in a Microsoft Azure Synapse SQL mapping.
- When you select multiple objects as the source type, you cannot read more than two objects.
- When you use a custom query as the source type, the first three letters of the table name alias is always `INF`.
- To write date/time fields to a Microsoft Azure Synapse SQL target created at the runtime, edit the metadata for the target fields and change the native data type of all the date/time fields to `datetime2`. Else, the mapping fails.
- When you use a custom query as the source type, ensure that the table name or column name that you specify in the query does not contain unicode characters. Else, the mapping fails.
- When you use a Filter transformation, ensure that the filter value does not contain special characters. Else, the mapping does not write any rows to the target.
- You can use the update and delete operations only for an existing target.
- When you enable the **Create New at Runtime** option in a Microsoft Azure Synapse SQL target to write data from file such as Delimited, ORC, or Parquet, delete the **FileName** field.

- The reject threshold for a Microsoft Azure Synapse SQL target is considered for the whole data set, not for a batch of data.
- Do not use the Aggregator transformation with update operation.
- Do not configure a nested aggregate function in an Aggregator transformation.
- Do not configure parallel transformations in a mapping.
- Ensure that the Expression transformation does not contain a variable field.
- Use the ISNULL function within the IIF clause in Aggregator and Expression transformations.
- Use the IN function in a Filter transformation or within the IIF clause in Aggregator and Expression transformations.
- When you use STDDEV, VARIANCE, and TRUNC functions in an Aggregator transformation, data corruption is possible if the table has a defined decimal column with precision more than 28, but the table contains data with precision less than 28.
- When you use the `lastruntime` variable in an Expression transformation in full pushdown optimization, you must use the `yyyy-mm-dd hh:mi:ss` date format.  
For example, `to_char($lastruntime, 'yyyy-mm-dd hh:mi:ss')`
- You cannot use the `ISNULL(column name1)<>ISNULL(column name2)` syntax.  
Use the following syntax for Filter and Router transformations:  

```
(ISNULL(column name1) AND ISNULL(column name2)) OR ( NOT ISNULL(column name1) AND NOT ISNULL(column name2))
```

  
Use the following syntax for Expression and Aggregator transformations:  

```
IIF((NOT ISNULL(column name1) AND ISNULL(column name 2)) OR (ISNULL(column name1) AND NOT ISNULL(column name 2)), 1, 0)
```
- When you use a Filter or Router transformation and specify a string value in the filter condition, the mapping appends the character N before the condition.  
For example,  

```
Select*from [Azure].[employee] where (10125 <= "INF1"."EMPLOYEE_ID_NEW") AND ("INF1"."JOB_ID_NEW" = N'FI_ACCOUNT');
```
- When you pass columns with Null values in a Normalizer transformation, Null values are not written to the target.

#### Mapping with Microsoft Azure Synapse SQL source and target

Use the following rules and guidelines when you configure pushdown optimization in a mapping that reads from and writes to Microsoft Azure Synapse SQL:

- The source and target table must be in the same database.
- When you configure filters in a Source transformation, consider the following guidelines:
  - If a mapping contains a Filter transformation and also a filter in the Source transformation, the mapping consolidates the filter conditions from both these transformations to filter the records. However, it is recommended that you use only one of these filters at a time in a mapping.
  - The column names in the filter condition must not contain unicode characters.
  - You cannot use system variables in filters.
  - You cannot use filters when the source object type is query.
- Ensure that each key column for the update operation is mapped, else the query generation fails and the mapping runs without pushdown optimization.

- When you use a custom query as the source type, ensure that there is no semicolon at the beginning of or within the query.
- When you use a custom query as the source type, ensure that the table name or column name that you specify in the query does not contain special characters.
- You cannot completely parameterize a multi-line custom query using a parameter file.
- If the custom query starts with the WITH clause, you must enable the **Create Temporary View** option to run the mapping with pushdown optimization.
- You cannot use the ORDER BY clause in a source custom query unless you also specify a TOP, OFFSET, or FOR XML clause in the query.
- You cannot use the OPTION clause in a custom query or SQL override.
- When you run a mapping to write data to new target created at runtime and you stop the job using clean stop, the target table created is not dropped.
- Source pushdown optimization
  - To run a mapping with pushdown optimization, the mapping must have at least one transformation between the source and target.
  - When you select source pushdown optimization, the task pushes the transformation logic for all the supported transformations downstream in the mapping.
  - When a function within a transformation is not supported, the query generation fails and the mapping runs without pushdown optimization.
  - You can connect the Source transformation only to a single downstream transformation.

#### **Mapping with Microsoft Azure Data Lake Storage Gen2 source and write to a Microsoft Azure Synapse SQL target**

Use the following rules and guidelines when you configure pushdown optimization in a mapping that reads from a Microsoft Azure Data Lake Storage Gen2 source and write to a Microsoft Azure Synapse SQL target:

- The Microsoft Azure Data Lake Storage Gen2 account and Microsoft Azure Synapse SQL account must be associated with the same Azure Active Directory tenant.
- The Azure Government endpoints for the source and target must be associated with the same Azure Active Directory tenant.
- You cannot configure a private endpoint or a virtual network to connect to Microsoft Azure Data Lake Storage Gen2. You can use managed identity authentication instead to securely connect to Microsoft Azure Data Lake Storage Gen2.
- You cannot read objects stored in subdirectories in Microsoft Azure Data Lake Storage Gen2.
- Ensure that the Microsoft Azure Data Lake Storage Gen2 source object is not parameterized.
- When you read data from Microsoft Azure Data Lake Storage Gen2, do not use escape characters. Else, it might lead to incorrect results.
- When you push down a mapping that reads from a Microsoft Azure Data Lake Storage Gen2 source and writes to a Microsoft Azure Synapse SQL target, the mapping task fails to generate the push down query if the task has all the following configurations:
  - The Microsoft Azure Synapse target connection is parameterized.
  - The target operation and Treat Source Rows As property in the mapping is set to Insert.
  - The Treat Source Rows As property is changed to Update in the mapping task.



## CHAPTER 5

# Data type reference

Data Integration uses the following data types in Microsoft Azure Synapse SQL mappings and mapping tasks:

- Microsoft Azure Synapse SQL native data types appear in the source and target transformations when you choose to edit metadata for the fields.
- Transformation data types. Set of data types that appear in the transformations. These are internal data types based on ANSI SQL-92 generic data types, which the Secure Agent uses to move data across platforms. They appear in all transformations in a mapping.

When the Secure Agent reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When the Secure Agent writes to a target, it converts the transformation data types to the comparable native data types.

## Microsoft Azure Synapse SQL and transformation data types

The following table compares the Microsoft Azure Synapse SQL native data type to the transformation data type:

Microsoft Azure Synapse SQL Data Type	Transformation Data Type	Range and description
bigint	bigint	-9,223,372,036,854,770,000 to 9,223,372,036,854,770,000
binary	binary	1 to 8000 bytes
bit	integer	0,1,NULL
char	string	1 to 8000 characters
date	date/time	January 1, 1753 00:00:00 to 12/31/9999
datetime	date/time	January 1, 1753 00:00:00 to December 31, 9999 23:59:59.997
datetime2	date/time	1 to 8000 characters

Microsoft Azure Synapse SQL Data Type	Transformation Data Type	Range and description
datetimeoffset	string	
decimal	decimal	Precision 28, scale 0 to 5 Data corruption is possible when the table has a defined decimal column with precision more than 28, but the table contains data less than 28.
float	decimal	Precision 7, scale 0 to 7
int	integer	-2,147,483,648 to 2,147,483,647
money	decimal	-922,337,203,685,477.0000 to 922,337,203,685,477.0000
nchar	string	1 to 8000 characters
nvarchar	string	1 to 4000 characters
real	double	Precision 7, scale 0 to 7
smalldatetime	date/time	1/1/1900 0:00 to 6/6/2079 23:59
smallint	integer	-32,768 to 32,767
smallmoney	decimal	-214,748.3648 to 214,748.3647
time	string	00:00:00.0000000 to 23:59:59.9999999
tinyint	integer	0 to 255
Uniqueidentifier	string	0 to 36 characters
varbinary	binary	1 to 8000 bytes
varchar	string	1 to 8000 characters

### Rules and guidelines for data types

- The `datetimeoffset` data type is supported only for passthrough mappings that read data from and write data to Microsoft Azure Synapse SQL objects.
- When you read or write `varchar` or `nvarchar` data types, do not use `\n` or `\r` character in the string data.
- When you create a new target, the fields of `Uniqueidentifier` data type in the source are converted to `Nvarchar` data type in the target.  
To retain the `Uniqueidentifier` data type, edit the metadata for the target fields and change the native type of the fields to `Uniqueidentifier`.

## CHAPTER 6

# Troubleshooting

Use the following section to troubleshoot errors in Microsoft Azure Synapse SQL Connector.

## Troubleshooting a Microsoft Azure Synapse SQL connection

### Unable to connect to Microsoft Azure Blob Storage or Microsoft Azure Synapse SQL VNet endpoints

When the Secure Agent machine is outside the virtual network (VNet) and you connect to Microsoft Azure Blob Storage or Microsoft Azure Synapse SQL VNet endpoints, the **Test Connection** fails.

You must configure the Secure Agent IP in both the **Virtual Network** section and **Firewall** section in your account and test the connection again.

## Troubleshooting a mapping or mapping task

### When the staging property is enabled, the mapping fails even though the rows are successfully written to the target.

When you enable the staging property in a mapping that reads from Microsoft Azure Synapse SQL, the mapping might fail with one of the following errors even though the rows are successfully written to the target:

- [FATAL] Error : HdfsBridge::recordReaderFillBuffer - Unexpected error encountered filling record reader buffer: HadoopSQLException: Error converting data type VARCHAR to DATETIME.
- [FATAL] Error : HdfsBridge::recordReaderFillBuffer - Unexpected error encountered filling record reader buffer: HadoopExecutionException: Too long string in column [-1]: Actual len = [1015]. MaxLEN=[1000]

To resolve this issue, increase the reject threshold in the advanced target properties and run the mapping again.

### [FATAL] Exception: com.microsoft.sqlserver.jdbc.SQLServerException: 100001;Failed to generate query plan.

When you use an ORDER BY clause in a custom query or SQL override, the mapping fails with the following error:

```
[FATAL] Exception: com.microsoft.sqlserver.jdbc.SQLServerException: 100001;Failed to generate query plan.
```

To use the ORDER BY clause in a query, you must configure the following property in the JVM options of the Secure Agent:

```
-DAzureSynapseDisableStagingForSort=true
```

If you use custom query in a Lookup transformation, you must perform the following additional tasks:

- Enable the **Sorted Input** option in the advanced lookup properties.
- The query must first contain the ORDER BY clause for the condition fields are included in the lookup condition in the ascending order and then the other fields in any desired order.
- When you configure the Lookup transformation to return multiple rows, the **Return All Rows** option does not return the values in the expected order.

#### **Mapping failed with a Java heap space error**

When you run a mapping that reads a large volume of data from or writes a large volume of data to Microsoft Azure Synapse SQL, the mapping fails with a Java heap space error.

Set the JVM options for type Tomcat JRE to increase the -Xms and -Xmx values in the system configuration details of the Secure Agent. You must then restart the Secure Agent.

#### **[ERROR] com.informatica.cci.cloud.client.impl.CCIClientExceptionImpl: Invalid expression string for filter condition.**

If you configure an uncached Lookup transformation to look up data from a Microsoft Azure Synapse SQL source and the source contains a column of DateTime or Timestamp data type, the mapping fails for certain values with the following error:

```
[ERROR] com.informatica.cci.cloud.client.impl.CCIClientExceptionImpl: Invalid expression  
string for filter condition.
```

# INDEX

## A

administration [8](#)  
Azure active directory authentication [9](#)  
Azure Blob Container Name [18](#), [24](#)

## C

cache  
    enable lookup cache [33](#)  
CDC source  
    Microsoft Azure Synapse SQL mapping task [28](#)  
Cloud Application Integration community  
    URL [5](#)  
Cloud Developer community  
    URL [5](#)  
connections  
    Microsoft Azure Synapse SQL [13](#)

## D

Data Integration community  
    URL [5](#)  
data types [57](#)

## E

error file directory [24](#)

## F

field delimiter [18](#), [24](#)  
flat file  
    staging data [11](#)  
forward rejected rows [24](#)

## I

IDENTITY column [36](#)  
Informatica Global Customer Support  
    contact information [6](#)  
Informatica Intelligent Cloud Services  
    web site [5](#)

## L

lookup caches  
    dynamic [33](#)  
Lookup transformation  
    lookup caching [33](#)

## M

maintenance outages [6](#)  
mappings  
    overview [18](#)  
    source properties [18](#)  
    target properties [24](#)  
Microsoft Azure Synapse SQL  
    connection properties [13](#)  
    introduction [7](#)

## N

native data type [57](#)  
number of concurrent sessions to Blob Storage [18](#), [24](#)

## P

properties  
    in mappings [18](#), [24](#)  
Pushdown optimization  
    preview [43](#)  
Pushdown optimization preview [43](#)

## R

runtime environments [8](#)

## S

Secure Agents [8](#)  
staging data  
    flat file [11](#)  
status  
    Informatica Intelligent Cloud Services [6](#)  
success file directory [24](#)  
system status [6](#)

## T

tracing level [18](#)  
transformation data type [57](#)  
trust site  
    description [6](#)

## U

upgrade notifications [6](#)

W

web site [5](#)