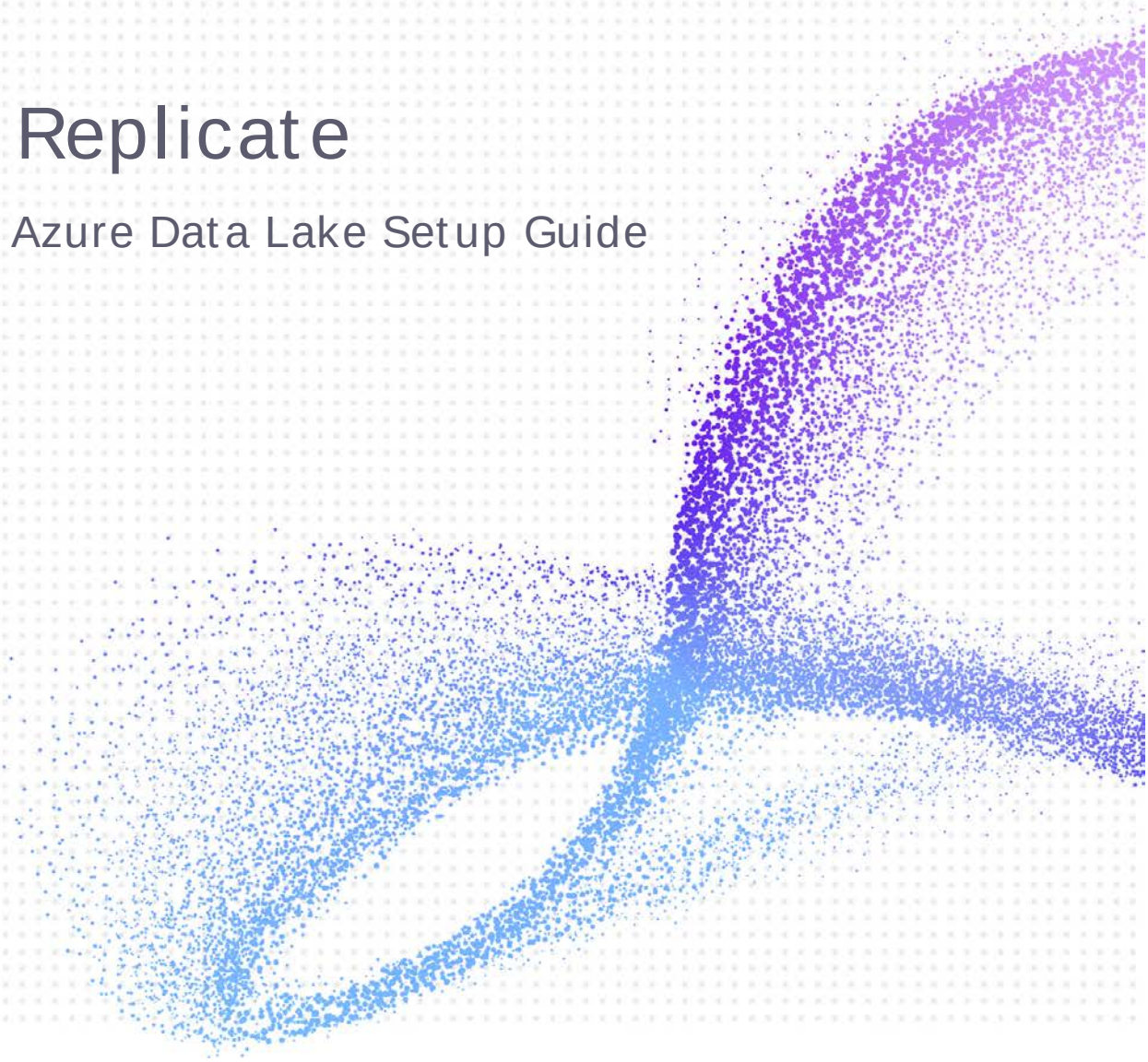
Syniti

Syniti Replicate

Microsoft Azure Data Lake Setup Guide

Version 10.2



Syniti

Syniti Replicate

Table of Contents

Syniti Replicate

This guide describes how to set up your Syniti Replicate environment to replicate data to a Microsoft Azure Data Lake Storage (ADLS) Gen 1 or Gen 2 target. The replicated data is stored in flat file format in ADLS. When replicating **from a relational database to ADLS**, Syniti Replicate allows you to replicate data in two different ways, or modes:

- **Refresh**
  A one-time complete replication from source table to ADLS, according to replication settings and scripts. You can control the timing of the replication, identify the columns to be replicated and add scripts to transform data during replication.
- **One-way mirroring**
  A continuous update of a replicated table to ADLS based on changes to the source database that have been recorded in the database server log. You can define the replication settings to check the transaction log on the source database at regular intervals. Any changes found in the log would be applied to the target.

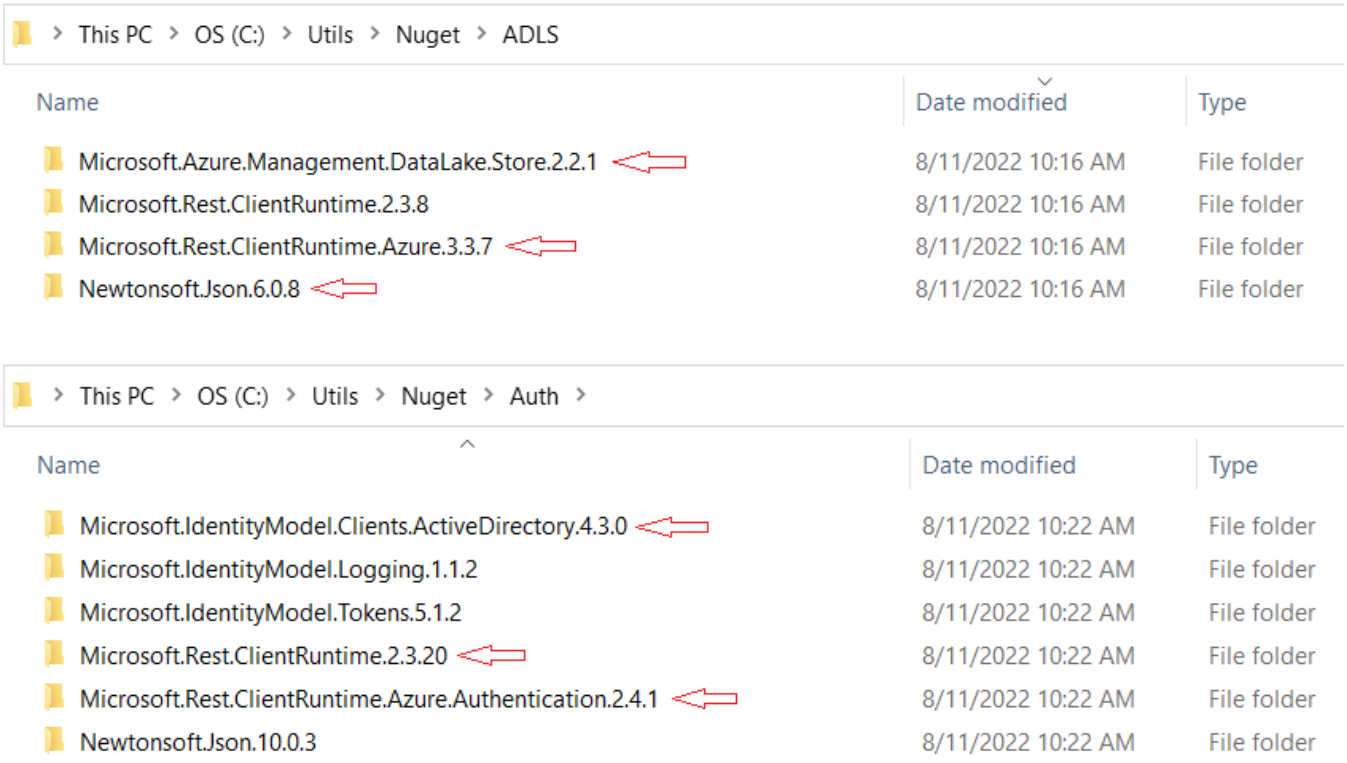
Requirements for ADLS Gen 1 Targets

1. Download the nuget packages:
   - Microsoft.Azure.Management.DataLake.Store ver 2.2.1

     https://www.nuget.org/packages/Microsoft.Azure.Management.DataLake.Store
   - Microsoft.Rest.ClientRuntime.Azure.Authentication ver 2.4.1

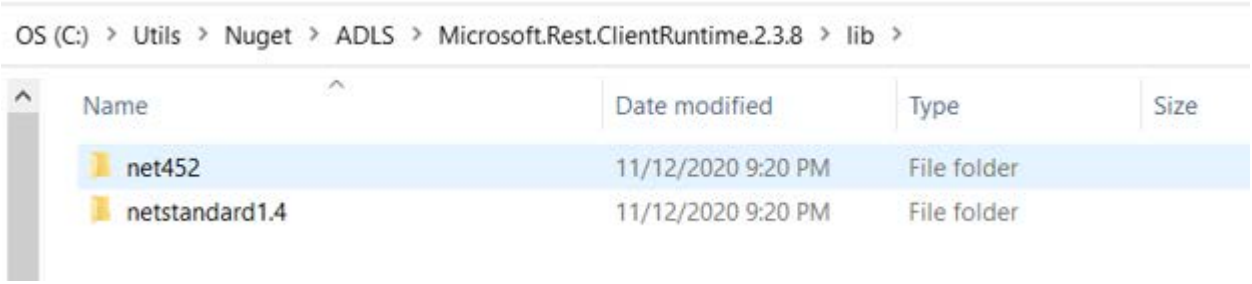
     https://www.nuget.org/packages/Microsoft.Rest.ClientRuntime.Azure.Authentication/
   - Microsoft.Rest.ClientRuntime (ver. 2.3.24)
   - Microsoft.Rest.ClientRuntime.Azure (ver. 3.3.19)
   - Newtonsoft.Json (ver. 13.0.1)

2. Download the nuget.exe command line tool from www.nuget.org/downloads.
3. Run the nuget.exe command line tool for each of the above packages. For example:
   ```
   nuget install Microsoft.Azure.Management.DataLake.Store -Version 2.2.1 -Framework net480 -OutputDirectory C:\Utils\Nuget\ADLS
   ```
   ```
   nuget install Microsoft.Rest.Clientruntime.Azure.Authentication -Version 2.4.1 -OutputDirectory C:\Utils\Nuget\Auth
   ```
4. This command downloads all assemblies into the specified folder. However, there are separate folders for each of the assemblies:



5. Open each folder highlighted in above screenshots and extract the version of the .dll that is:
   - Valid for .NET framework 4.8 or earlier (**netxx**), if available OR
   - Use the latest .NET Standard framework (**netstandardxx**) version

   For instance:



6. Copy the following assemblies to the DataLake plugin path in the Syniti Replicate folder Plugins/DataLake/
   - Microsoft.Azure.Management.DataLake.Store.dll
   - Microsoft.Azure.Management.DataLake.Store.xml
   - Microsoft.IdentityModel.Clients.ActiveDirectory.dll

- Microsoft.IdentityModel.Clients.ActiveDirectory.xml
- Microsoft.IdentityModel.Logging.dll
- Microsoft.IdentityModel.Tokens.dll
- Microsoft.Rest.ClientRuntime.Azure.Authentication.dll
- Microsoft.Rest.ClientRuntime.Azure.Authentication.xml
- Microsoft.Rest.ClientRuntime.Azure.dll
- Microsoft.Rest.ClientRuntime.Azure.xml
- Microsoft.Rest.ClientRuntime.dll
- Microsoft.Rest.ClientRuntime.xml
- Newtonsoft.Json.dll

Requirements for ADLS Gen 2 Targets

1. Download the nuget package:

   https://www.nuget.org/packages/Microsoft.IdentityModel.Clients.ActiveDirectory/5.2.8

2. Download the nuget.exe command line tool from www.nuget.org/downloads.

3. Run the nuget.exe command line tool for the above package. For example:

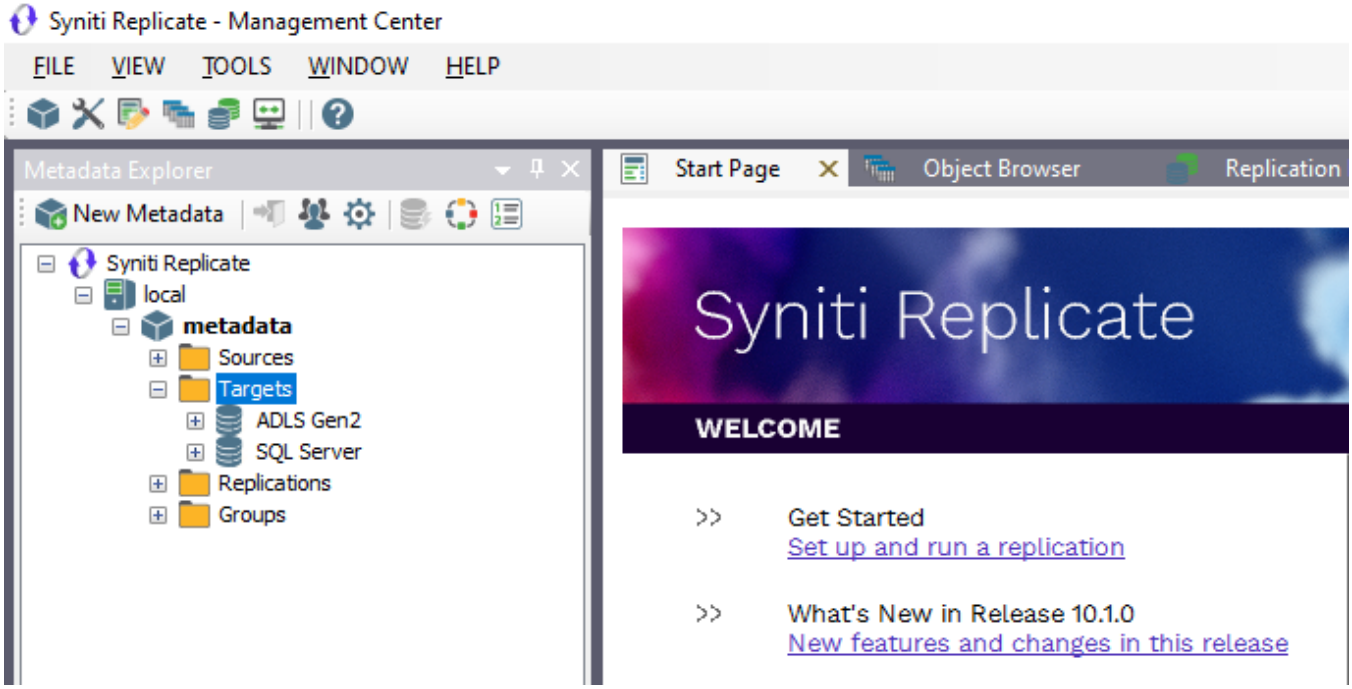
   ```
   nuget install Microsoft.IdentityModel.Clients.ActiveDirectory -Version 5.2.8 -Framework net480 -OutputDirectory C:\Utils\Nuget\ADLS
   ```

4. Copy the following assemblies to the DataLakeGen2 plugin path in the Syniti Replicate folder Plugins/DataLakeGen2/

   Microsoft.IdentityModel.Clients.ActiveDirectory.dll

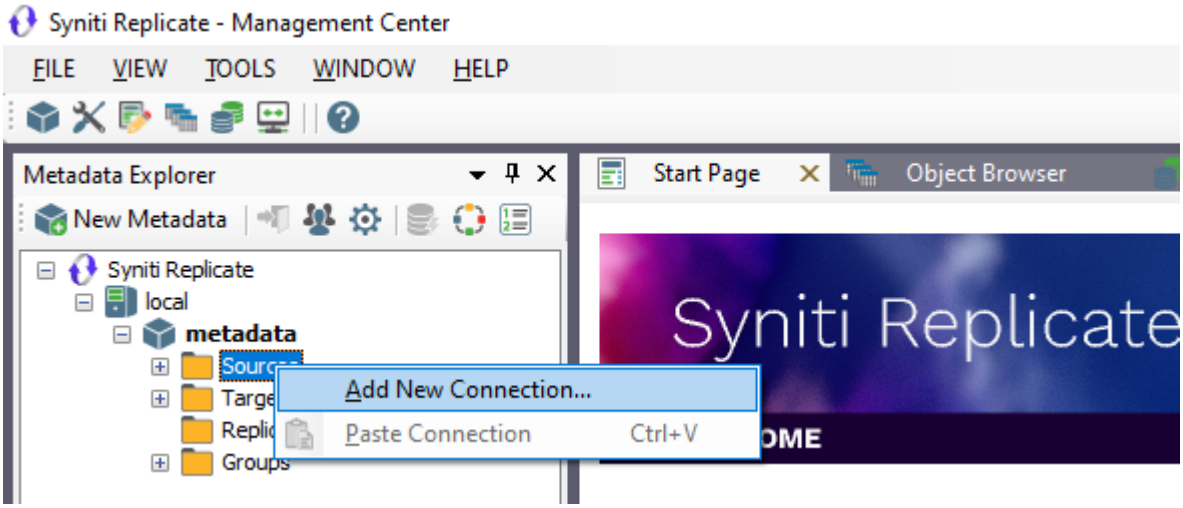
Replicating Data from a Relational Database to ADLS

- Make sure you have database connections via a .NET data provider to the supported databases for source database tables.
- Start Syniti Replicate Management Center.
  Syniti Replicate provides a default database (Microsoft SQL Server CE) for your metadata, all the information that Syniti Replicate needs to store about your replication setup.

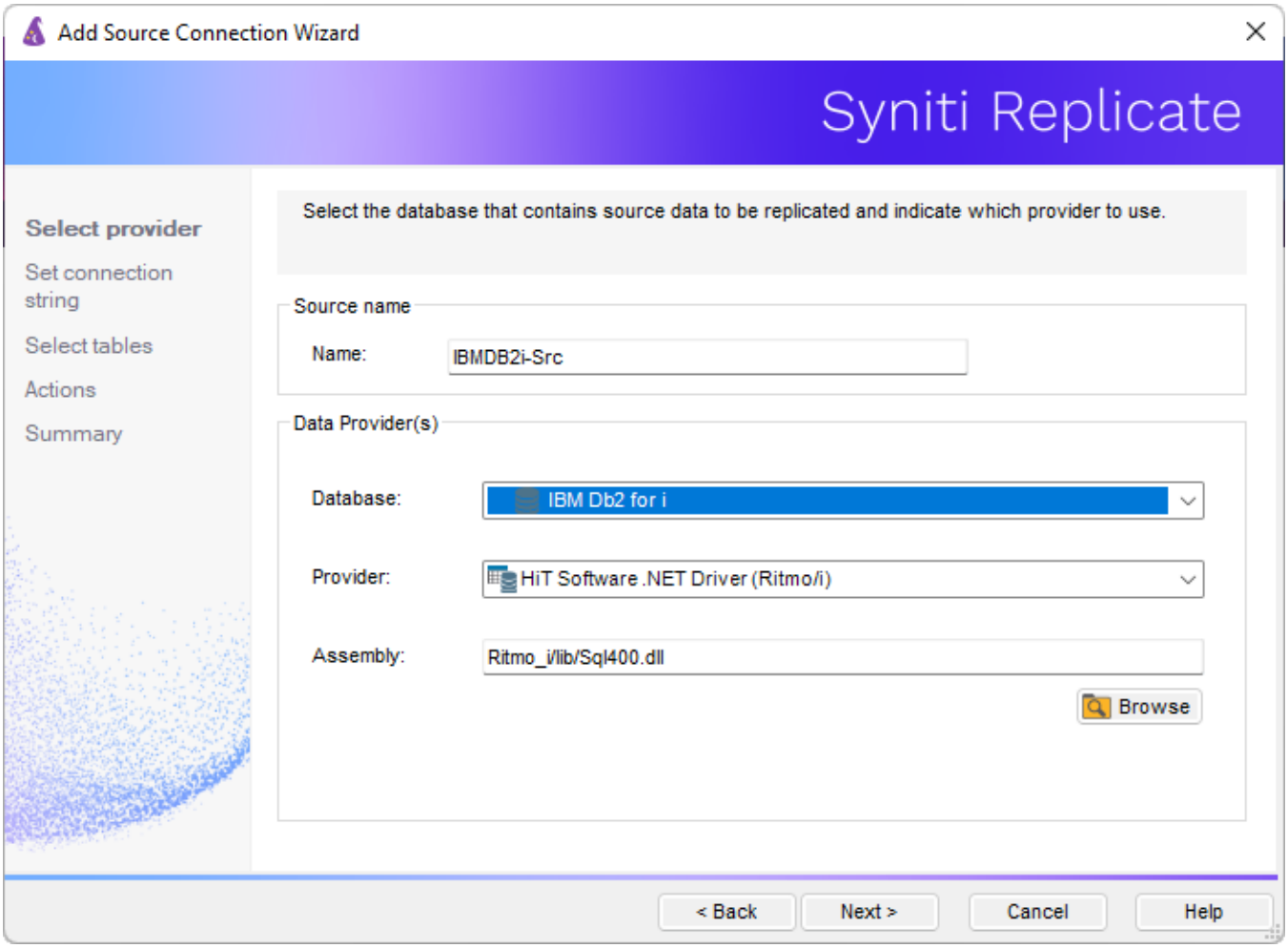


Create a Source Connection

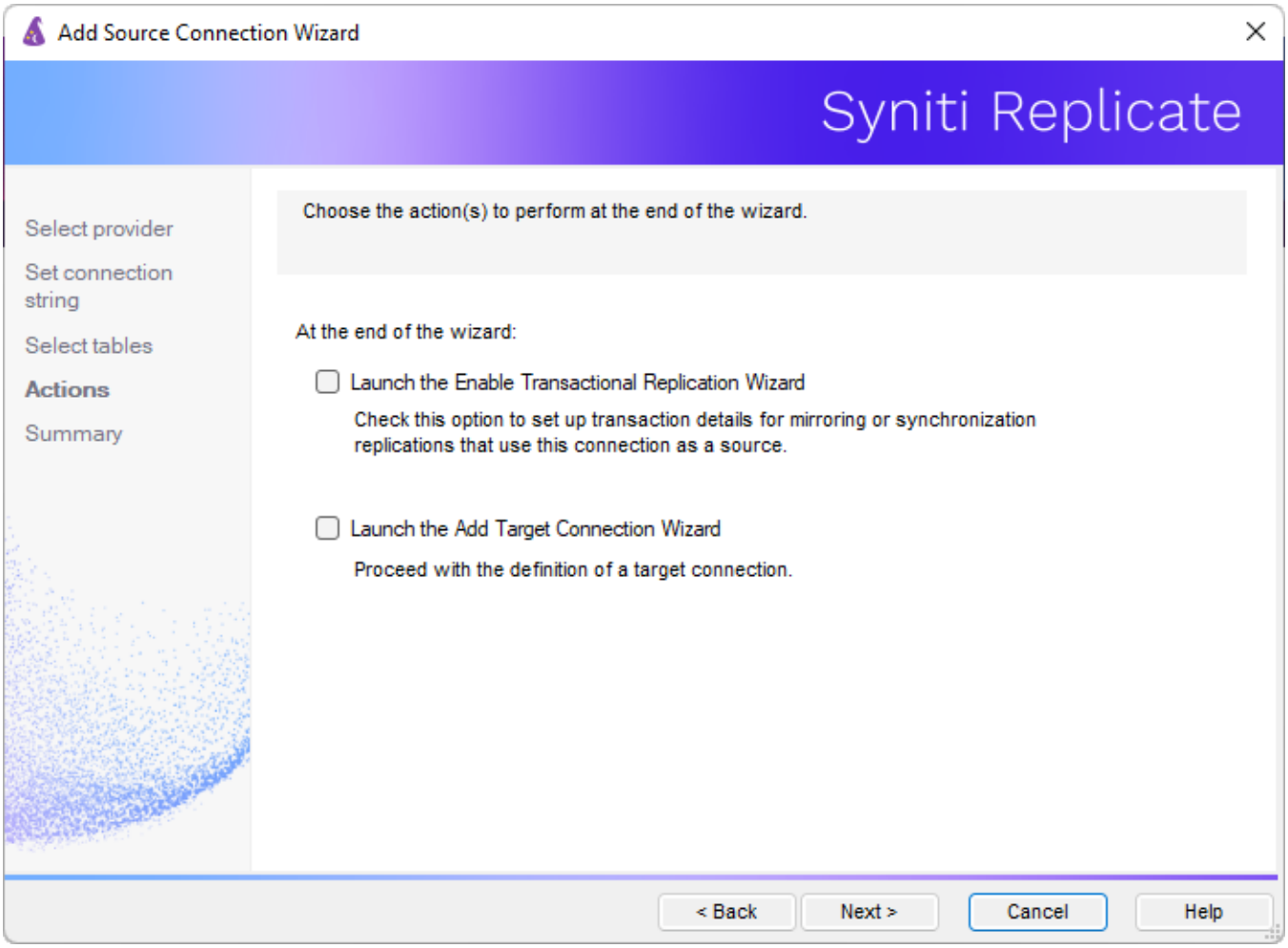
1. In the Metadata Explorer, expand the metadata node to view the **Sources** and **Targets** nodes.
2. Select the **Sources** node.
3. From the right mouse button menu, choose **Add New Connection**.



4. In the Add Source Connection Wizard, follow steps to add a connection string and test the connection to the database.
   Check the Supported Provider List in the Help Center before entering a value in the **Assembly** field.



5. In the **Select Tables** screen, choose the tables that you plan to replicate.
6. In the **Actions** screen, check the option **Launch the Enable Transactional Setup Wizard**.



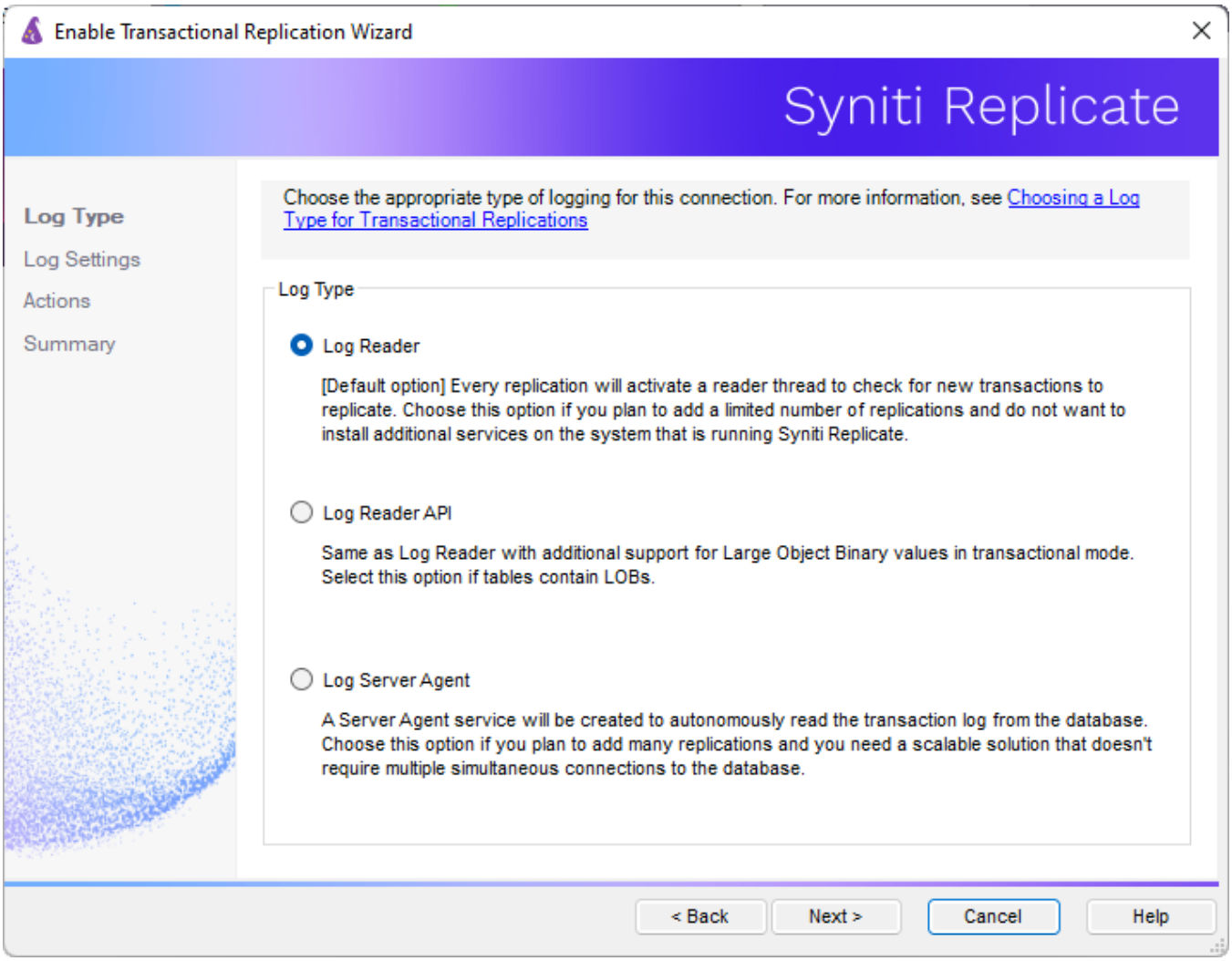
7. Complete the wizard.

Enable Transactional Replication

This section assumes you are planning to replicate data to an ADLS flat file target using mirroring (rather than Refresh) from a relational database. For Refresh replications, you do not need to complete the Enable Transactional Replication wizard. If you have checked the Source Connection wizard option to launch the Enable Transactional Replication wizard, follow the steps below. To open the wizard from the Management Center, choose the connection in the Metadata Explorer, then right-click to choose Transactional Setup > Enable...

In the **Enable Transactional Replication wizard**:

1. Select the type of transactional replication to use. The options depend on the source database and can include log reader, log server agent, triggers, plus log reader API (for IBM Db2 for i only). Check the Syniti Replicate help for more information on the best option.

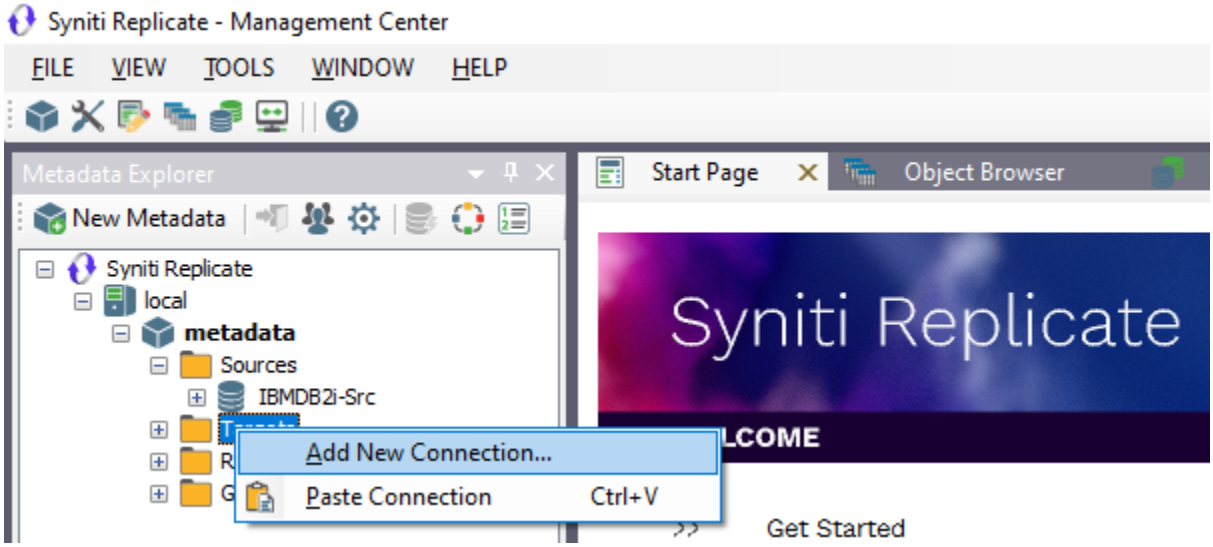


2. Click **Next** to enter the log settings. The fields and appropriate values depend on the database and log type. You can obtain a setup guide for a specific relational database by making a request in the Help Center.
3. Click **Next** to verify your settings against the source connection to the database. If any information is missing, you will not be able to proceed.
4. In the Actions screen, check the option to launch the **Add Target Connection wizard**.
5. Click **Next** to review your changes.
6. Click **Finish** to complete the wizard.

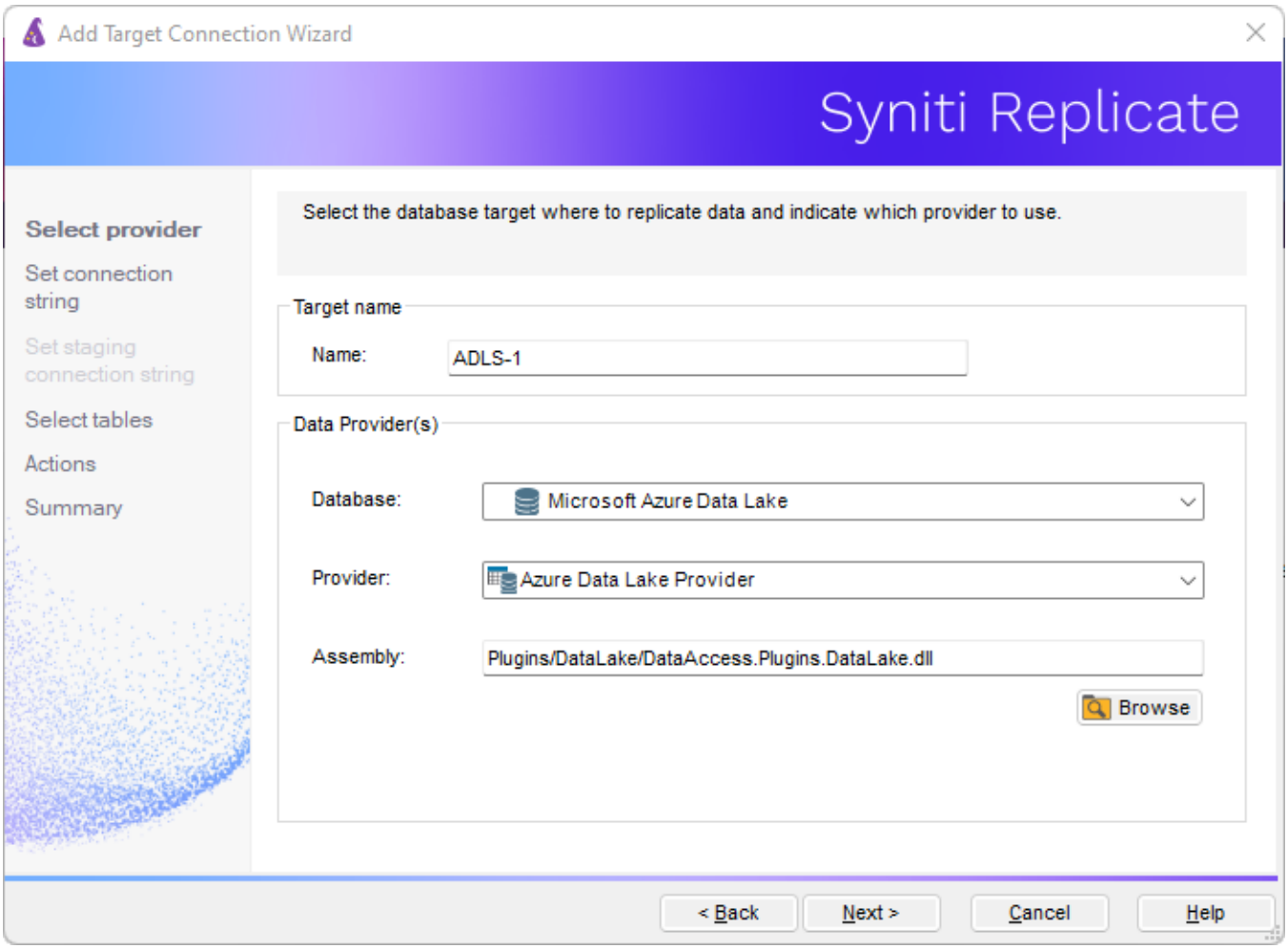
The source connection is now set up for transactional replications.

Set Up Target Connection to ADLS

This section assumes you have checked the **Enable Transactional Replication wizard** option to launch the Add Target Connection wizard. If not, to open the wizard from the Management Center, choose **Targets** in the Metadata Explorer, then right-click to choose **Add New Connection...**



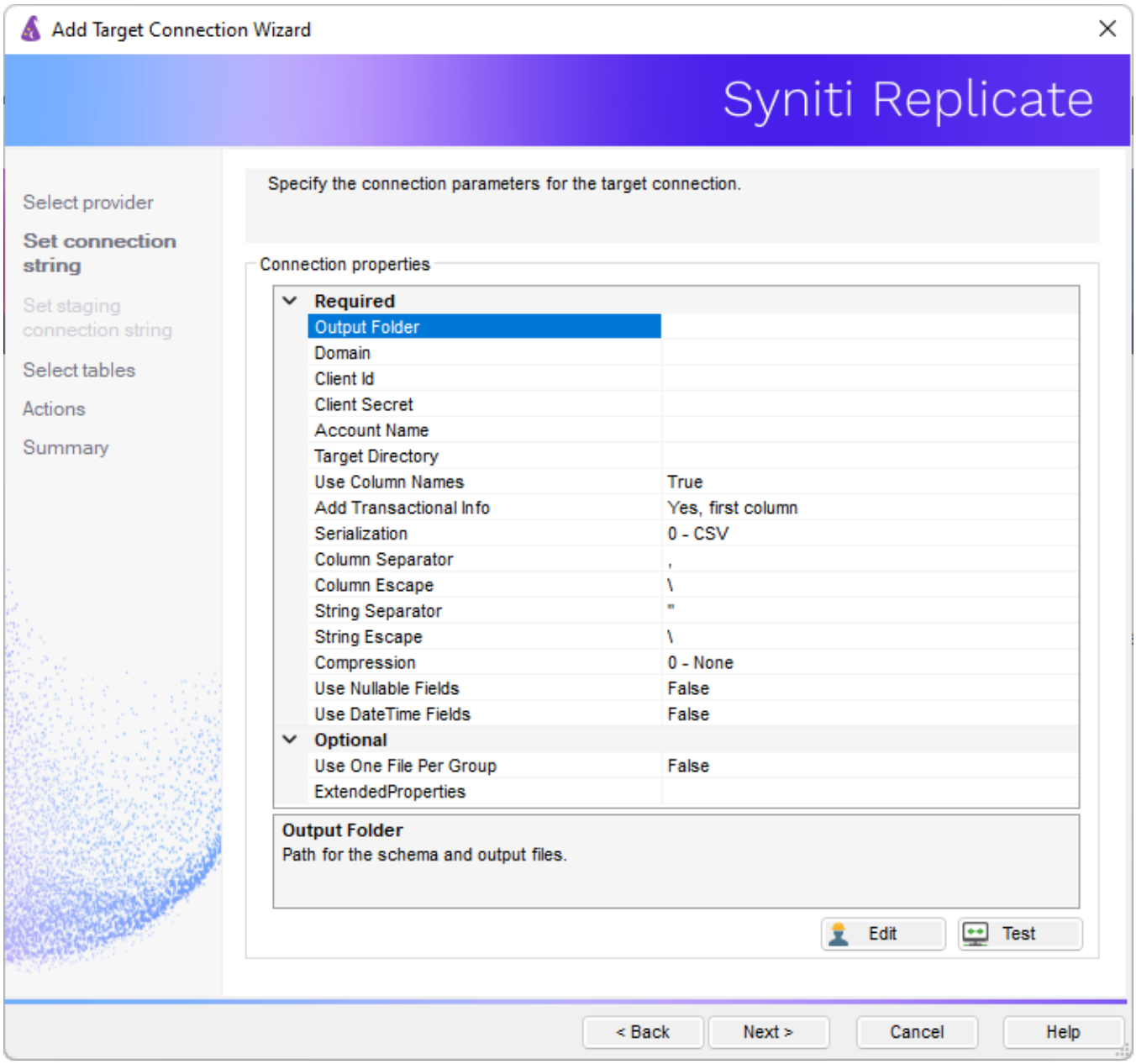
1. In the Target Connection Wizard **Provider** field, select the **Microsoft Azure Data Lake** option for Gen 1 connections or the **Microsoft Azure Data Lake Gen2** option for Gen 2 connections.



2. Click **Next** to display the **Set connection string** screen.

3. **For Gen 1 connections**, set values for the following fields:

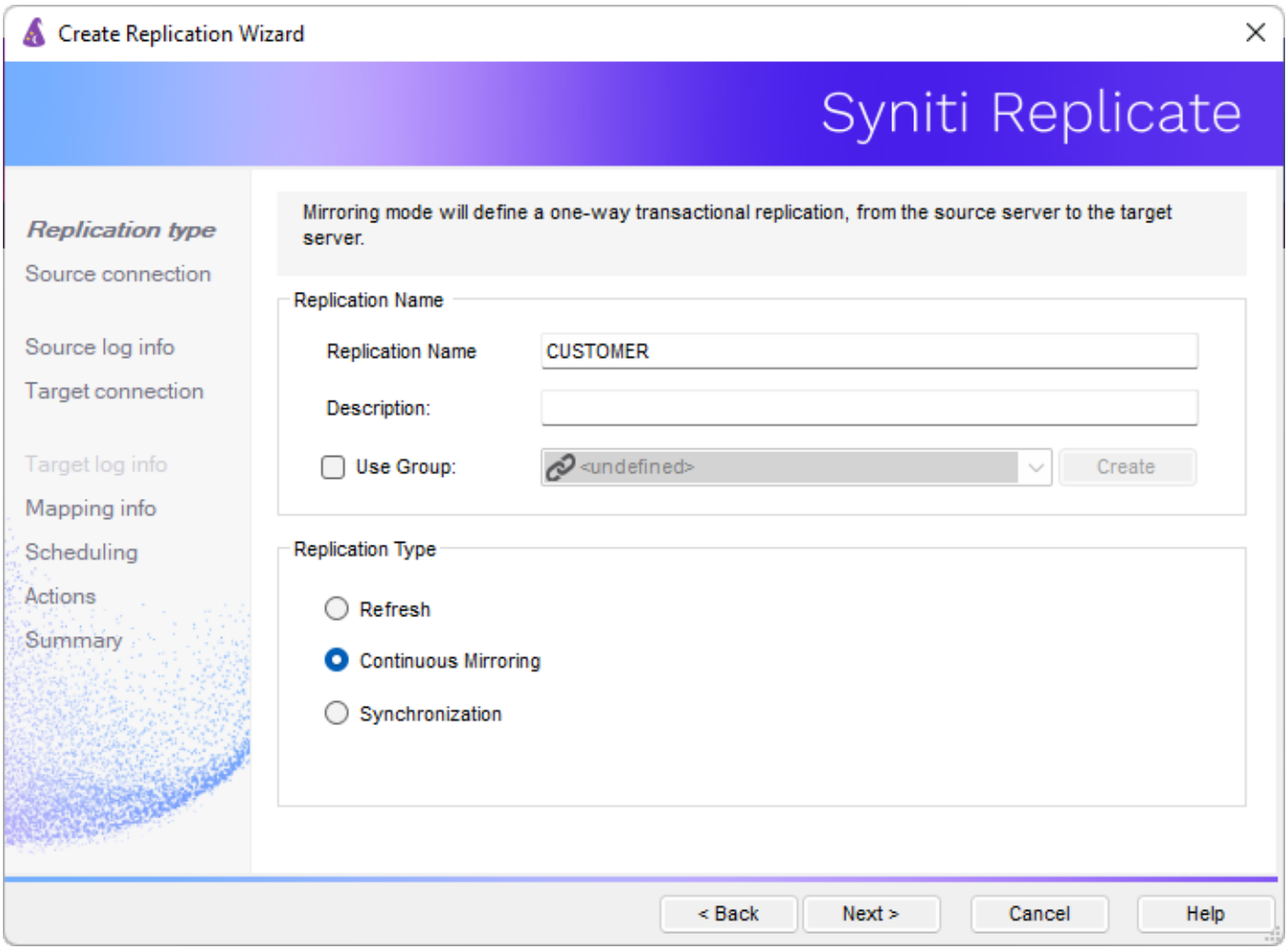
Output Folder	The location of the schema for config files for the ADLS objects
Domain	The name of the domain associated with the Azure Active Directory App used to upload data to Azure Data Lake
Client Id	Client ID in GUID format that identifies the client application
Client Secret	Client secret token used from the client application
Account Name	The name of the Data Lake Storage Gen1 account used to load data
Target Directory	Target location where to load data in Data Lake



Or, for Gen2 connections, set the following fields:

Output Folder	The location of the schema for config files for the ADLS objects
Client Id	Client ID in GUID format that identifies the client application. (Application added in **App registrations** in the Active Directory)

Tenant Id	Tenant(Directory) Id in guid format representing the Active Directory Tenant.
Client Secret	Client secret token used from the client application. It can be generated by clicking on "**Certificates & Secrets**" for the App added in **App registrations**.
Storage Account Name	The name of the Data Lake Storage Gen2 account used to load data.
Target Filesystem Name	Target FileSystem/Container name to use in loading data in ADLS.



4. Click **Next** to display the Select Tables screen.
   At this point, there is no text output structure available to display. You can add the information after completing the Target Connection wizard.
5. Click **Next** to display the **Summary** page.
6. Click **Finish** to complete the wizard.

Add Table Information to the Target Connection

The target connection is displayed in the Metadata Explorer, but you still need to add the file representation for source table data so that when you create replications below, you can specify a source table and target "file."

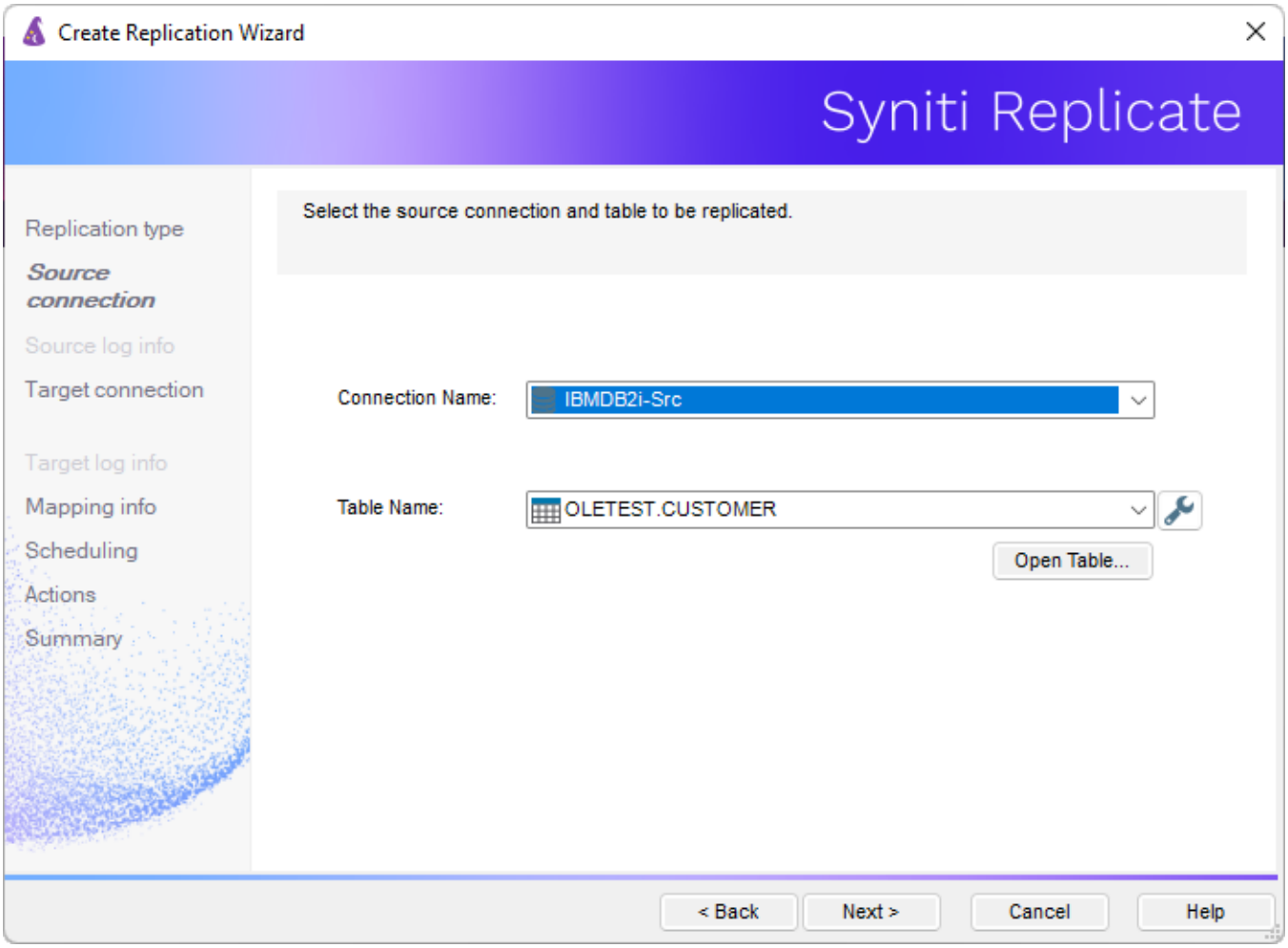
1. In the Metadata Explorer, expand the source connection you created above.
2. Select and drag a source table to the target file connection.
3. The Create Target Table wizard is displayed.
   Although you are not actually creating tables, you can use this wizard to create a representation of the file data.
4. In the **Source Connection** screen, you should see the source table you selected above.
5. Click **Next** to display the **Target Connection** screen.
6. Verify that the screen displays the correct target connection name, and table name.
7. Click **Next** to review the table structure.
   At this point, you can modify data types, null values and so on, if you want to modify the data eventually sent to a file.
8. Click **Next** to display the **SQL Script** screen.
   The contents of this screen are inactive because there is no editable SQL script to create a table. Instead, Syniti Replicate outputs the table information to a file.
9. Click **Next** to display the **Summary** screen.
10. Click **Finish** to create the table representation in the Metadata Explorer.
11. Click **Yes** to add the table name to the target connection entry in the Metadata Explorer
12. Repeat steps 2 through 11 for each source table that you want to replicate to a file.

At the end of this process, you should have a list of table representations under the Files target connection in the Metadata Explorer.

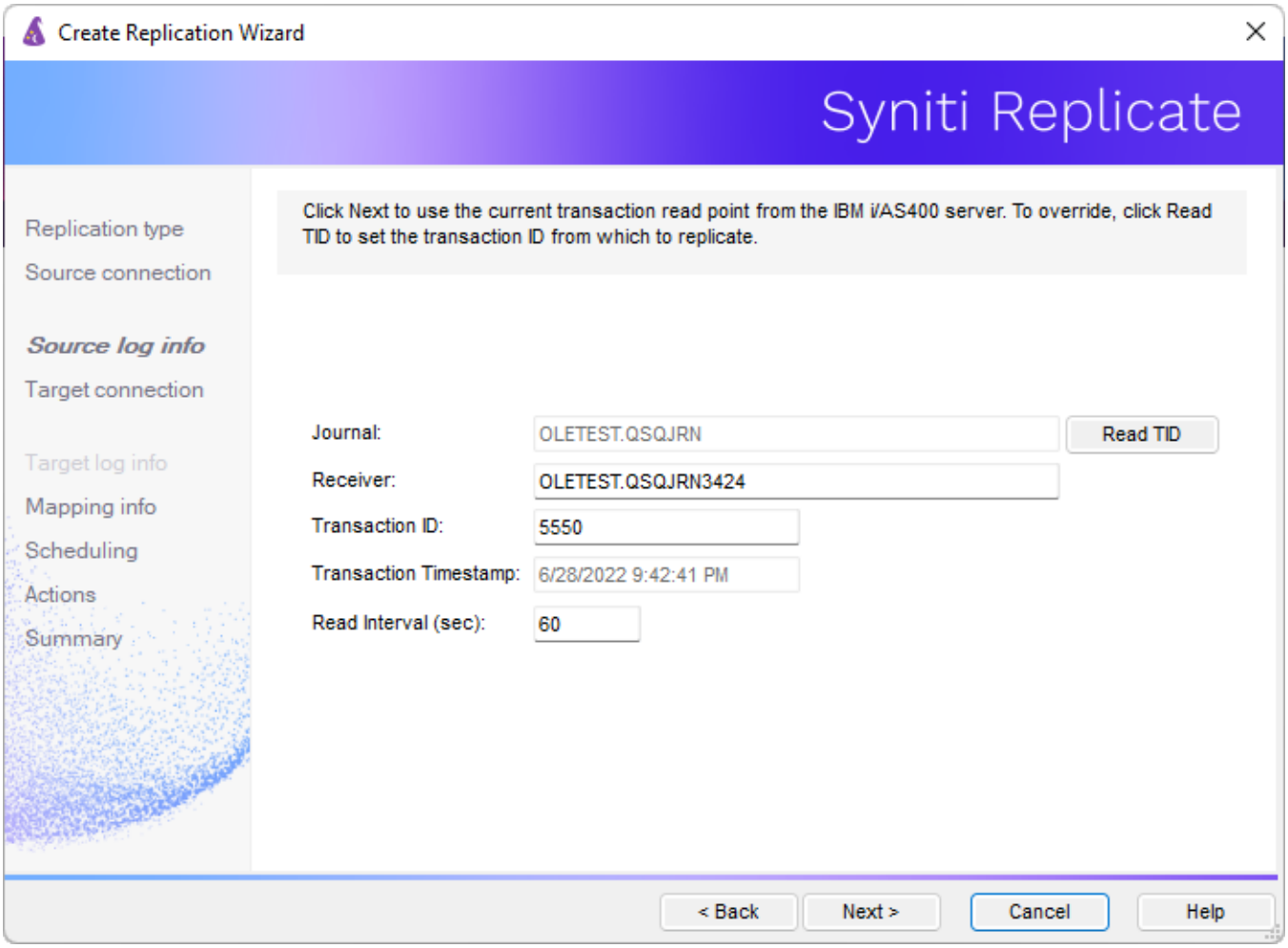
Define the Replication

This section assumes you have checked the Create Replication wizard option to launch the Create Replication wizard. If not, to open the wizard from the Management Center, choose the table you want to replicate in the Metadata Explorer, then right-click to choose **Replication > Create New Replication...**

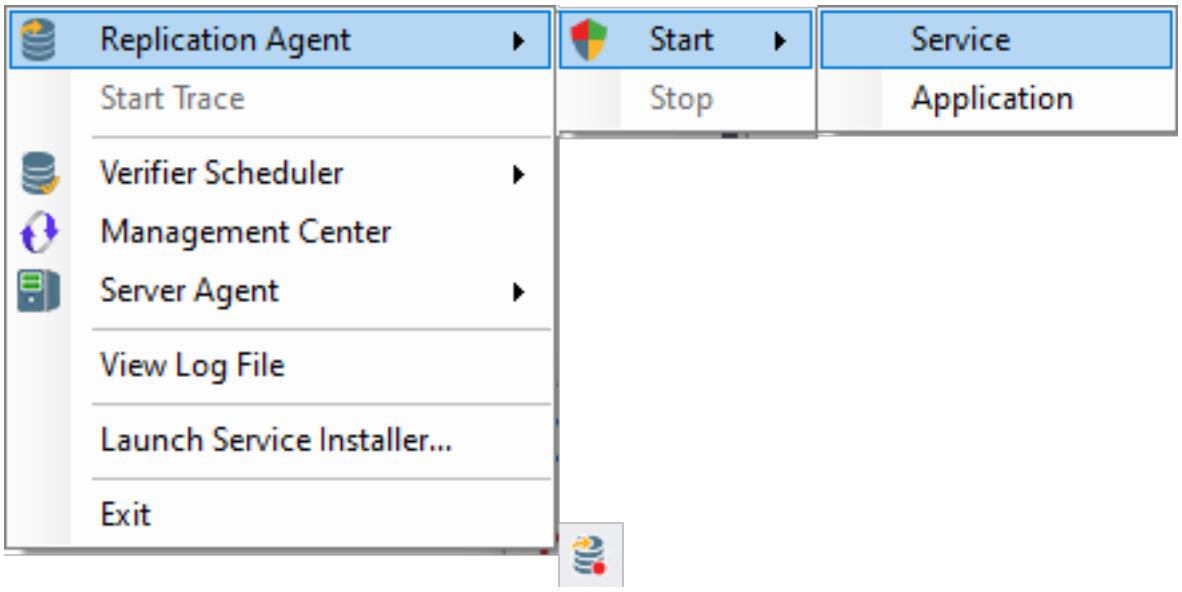
13. In the **Define Replication Type** screen, type a name to identify the replication.
14. Optionally provide a description of the replication.
15. In the Replication Mode area, choose **Continuous Mirroring**.



16. Click **Next** to go to the **Select Source Connection** screen.



17. Choose the source name from the drop-down list that includes all the source connections you have created in Syniti Replicate.
18. Choose the table that you want to replicate from the drop-down list.
19. If you want more information about the table before proceeding, click **Open Table...**.
20. Click **Next** to go to the **Source Log Info** screen.



The first two fields on this screen depend on the source database you are using. In this explanation, we assume you are using IBM Db2 for i (iSeries or AS/400). Check the help for the Replication Wizard for details on the other databases.

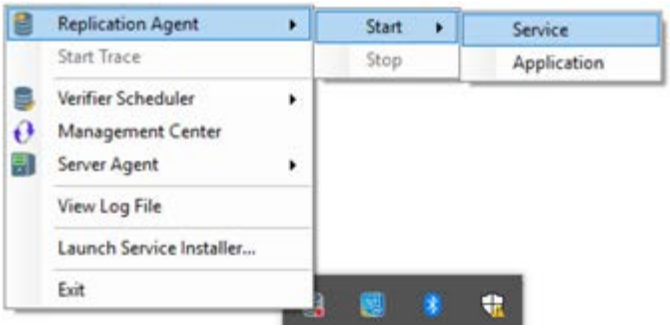
The Journal field is automatically filled in by retrieving the information from Db2 for i. The Receiver field will be automatically filled in after setting the Transaction ID. You do not need to modify these fields. However, if the library that you have specified as a source is not journaled, you will need to ask your system administrator to journal the library.

21. In the Transaction ID field, click **Read** to open the Read Point dialog.
22. Choose either the current transaction or a transaction ID based on a time and date.
23. Click **OK** to add the value to the **Source Log Info** screen.
24. Set the value of the **Read Interval** field to the frequency with which you want Syniti Replicate to check the transaction log for new events to mirror.
25. Click **Next** to go to the **Select Target Connection** screen.
26. Choose the target connection for ADLS from the drop-down list that includes all the target connections you have created in Syniti Replicate.
27. Choose the data set you want to replicate from the drop-down list.
    If the drop-down list is empty, exit the wizard and add or create a target data set.
28. Click **Next** to go to the **Set Mapping Info** screen.
    Source columns and target data with the same name are automatically mapped.
29. Set a start time for the replication. The **Start Time** field indicates the time at which the Replication Agent will begin considering the replication for execution.

30. Check the option to **Execute Initial Refresh**.
    A full replication will be performed from the source table to the data file, prior to starting the mirroring process where only incremental changes will be replicated.
31. Go to the Mirroring Schedule tab.
32. Select how you want to run the replication:
    - Run Continuously: the transaction log will be checked for changes to the table using the frequency that you specified on the Set Log Info tab. Any changes will be replicated to the target table.
    - Schedule Interruptions: the replication process will run as above, except for interruptions specified when you click the Schedule button in the Scheduler dialog.
33. Make sure the **Enable Replication** option is checked. This is required for the replication to run.
34. Click **Next** to go to the **Summary** screen.
35. Click **Finish** to complete the wizard.

Run the Replication

If you installed The Replication Agent as a service during Syniti Replicate setup, you just need to start the service using the Service Monitor program in the Windows Notification Area.



The replication that you have scheduled should start at the specified time. Use the Replication Monitor tab in the Management Center to track the progress of the replication. As an ADLS target consists of files, it is not possible to open the content of a target table to verify if the replication has been successful. To do that, you need to connect to the Azure portal or download the Azure Storage Explorer from here: https://azure.microsoft.com/it-it/features/storage-explorer/.

If you would like to install the Replication Agent as a service:

- From the Windows Desktop **Start** menu, choose **Syniti Replicate**, then **Service Installer**.
- Manage the service from Service Monitor program (located in the Syniti Replicate install folder or on the Windows **Start** > **Programs** > **Startup** menu).
- Use the Replication Monitor tab in the Management Center to track the progress of the replication.

To run the Replication Agent interactively:

- In the Windows Notification Area, select the Service Monitor icon.

- From the right mouse button menu, choose **Replication Agent**, then **Start** then **Application**. The replication that you have scheduled should start at the specified time.
- Use the Replication Monitor tab in the Management Center to track the progress of the replication.

File Structure

Data files replicated to ADLS using mirroring have the following structure. Files have a .mir extension when used for mirroring and a .ref extension when used for snapshot, or refresh replications.

Column	Value
1	Sequence number
2	Timestamp of transaction
3	User, if provided by the database
4	Table name
5	Operation code – one of I (Insert), U (Update) or D (Delete)
6	Previous primary key value for the database column. This is used for update operations to identify the correct record
7	Next primary key value. This is used for update operations to identify the correct record in cases where the primary key is changed during the update operation
8	Database column value

Example File

Example filename: `corp_20150115T161123155064_1.mir`

```
10,2015-01-22 15:32:39.000000,,"glb","I",,2,"test insert"
11,2015-01-22 15:32:40.000000,," glb ","D",10,,
12,2015-01-22 15:32:42.000000,," glb ","U",104,104,"my updated"
13,2015-01-22 15:32:44.000000,," glb ","U",1,12,"sd"
```