Syniti

Syniti Replicate

Setup Guide for Replicating to Cassandra
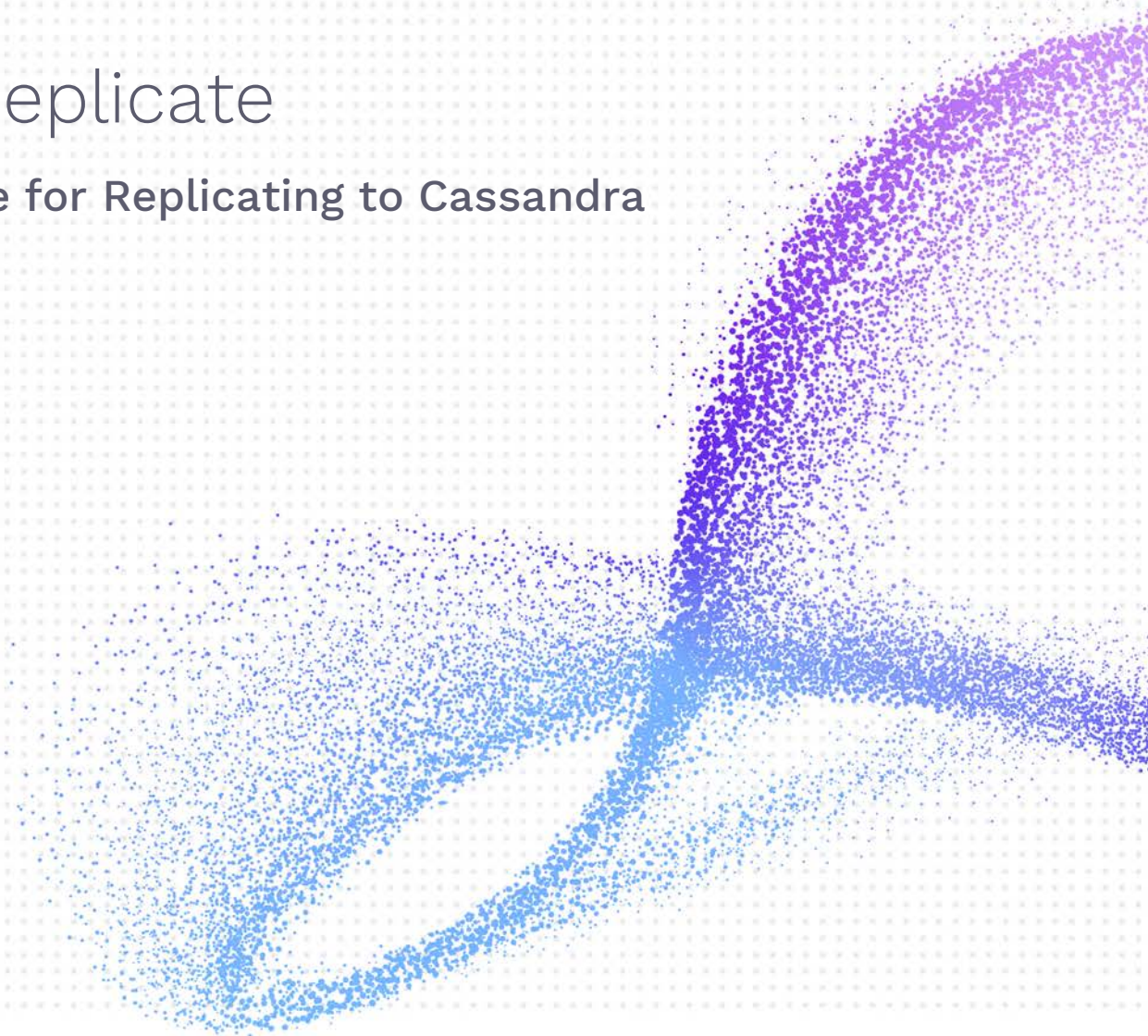
Version 10.3

Syniti

Syniti Replicate

Table of Contents

Syniti Replicate

Syniti Replicate allows you to replicate data from relational database tables to **Cassandra** in the following ways:

- **Refresh (Snapshot replication)**
  A one-time complete replication from any major relational database source to **Cassandra** as a target, 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. Source databases include Oracle, Microsoft SQL Server, IBM Db2 for i, IBM Db2 LUW, Sybase, Informix, MySQL.

  Continuous refresh
  A regularly scheduled refresh replication as described above. The schedule is defined in the replication settings.

- **One-way mirroring (Incremental Replication)**
  A continuous update of replicated tables based on changes to the source database that have been recorded in the database server log. Typically, this involves an initial refresh operation, as described above, to set up the target table. Then you can define the replication settings to check the transaction log on the source database at regular intervals. Any changes found in the log are applied to data on the **Cassandra** platform.

Syniti Replicate also allows you to use Cassandra as a **source database** for a **Refresh** replication.

For complete details on the setup process, check the *Syniti Replicate User Guide* available from the Management Center **Help** menu or the *Syniti Replicate Setup Guide*, available for download in the Help Center.

Requirements for Cassandra Targets

Connection Type

CData ADO.NET Provider for Cassandra. The driver can be downloaded from Cassandra ADO Driver: ADO Driver for Cassandra - CData Software, please use version 21.0.7850 or later.

Assembly: System.Data.CData.Cassandra (file name: System.Data.CData.Cassandra.dll)

Sample path:
C:\Program Files\CData\CData ADO.NET Provider for Cassandra 2021\lib\4.0\System.Data.CData.Cassandra.dll

Use of Primary Keys

Syniti Replicate manages INSERT, UPDATE and DELETE operations using primary keys to identify records. For Cassandra, this works as follows:

- Cassandra partition keys are used as primary keys for inserting/updating/deleting data in the Cassandra target.

- Syniti Replicate does not store clustering key values for a table in the metadata.

- If an insert is performed on a table with a primary key that already exists in a table, Cassandra does not throw a primary key violation error, rather it overwrites the existing row with the new values.

- If replicating using Refresh with a **Refresh Target Filter** property in the Replication Properties, you must use a primary key column in the WHERE clause with an "=" or "IN" operator. This is because arbitrary WHERE clauses are not supported in CQL(Cassandra Query Language).

- Cassandra does not allow updating a primary key column value via an UPDATE statement. Updates on only non-primary key columns in the source table will be replicated to the target Cassandra table via Syniti Replicate.

Syniti Replicate

Replicating Data from a Relational Database to Cassandra

- 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.
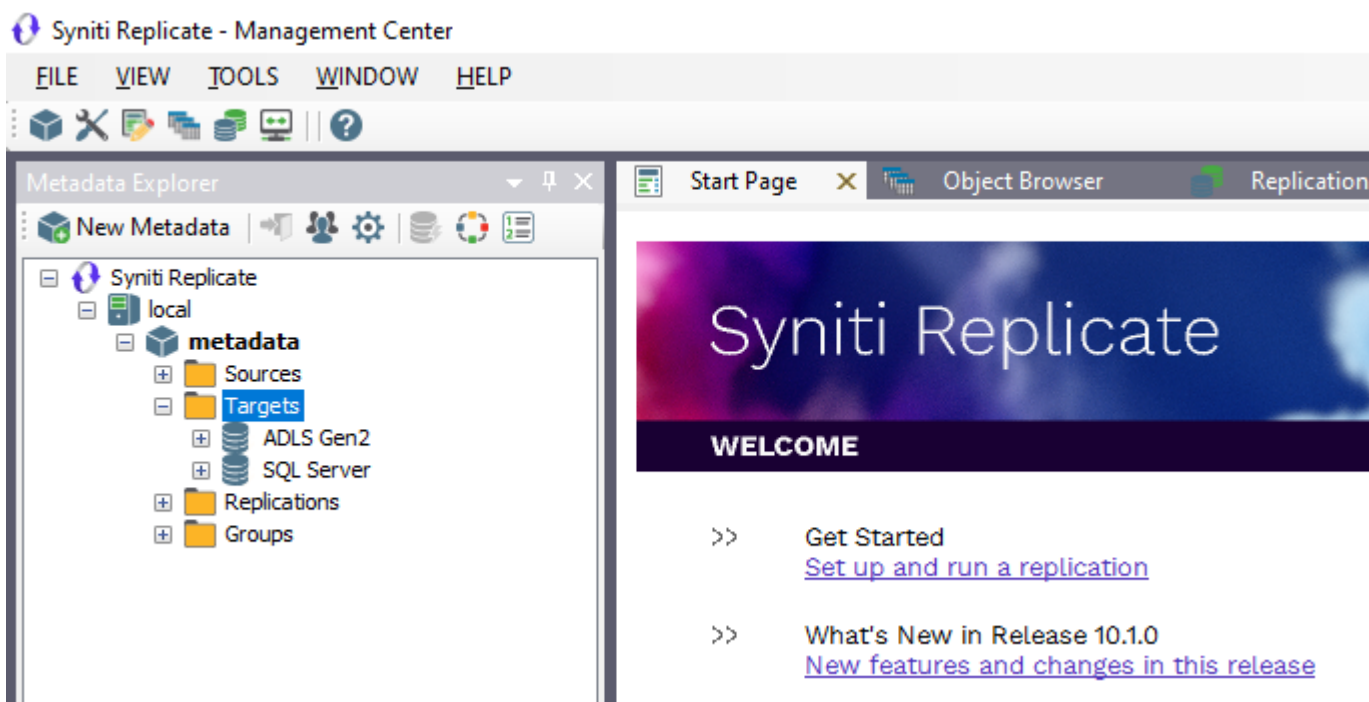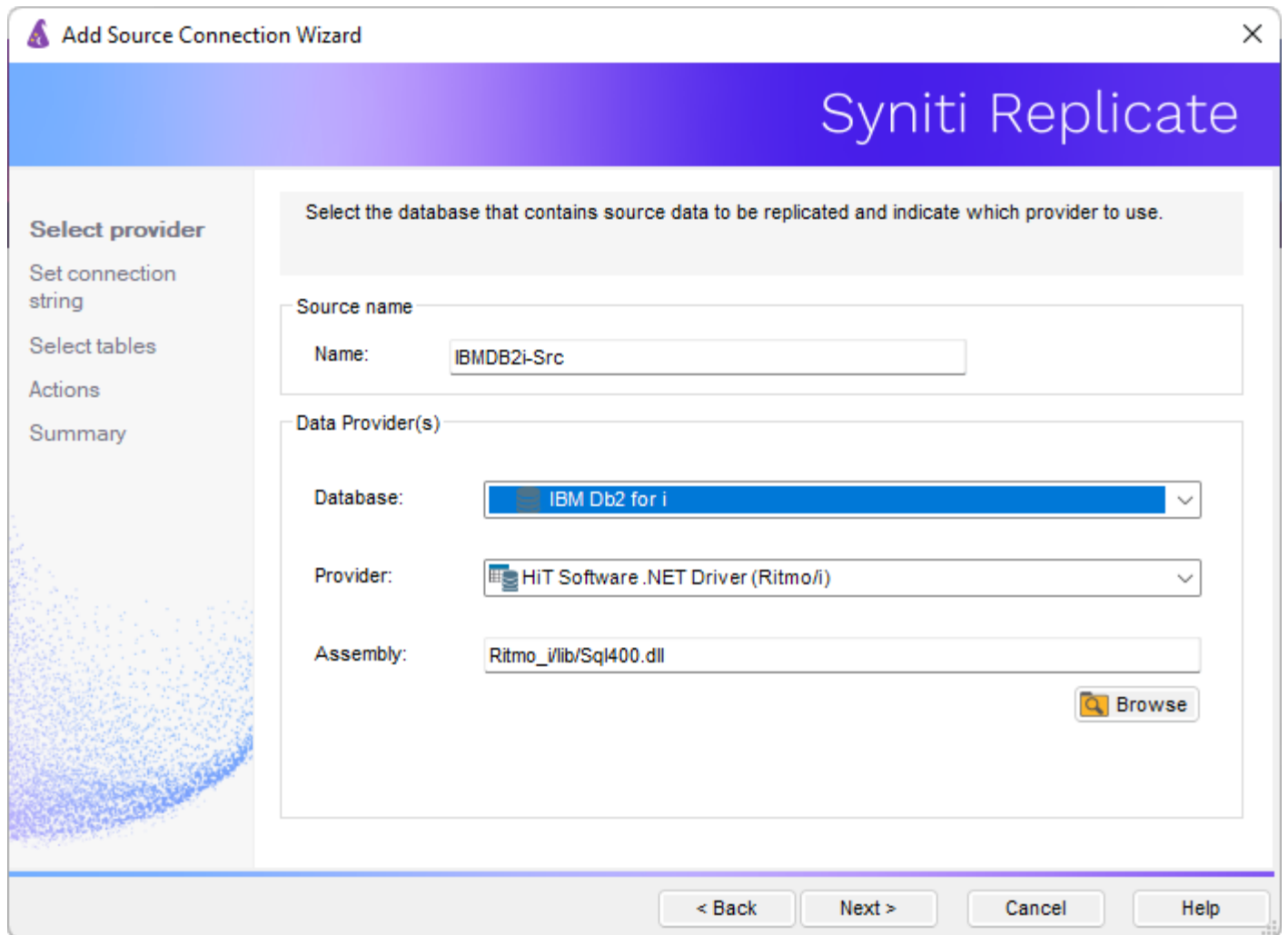

Create a Source Connection

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**.

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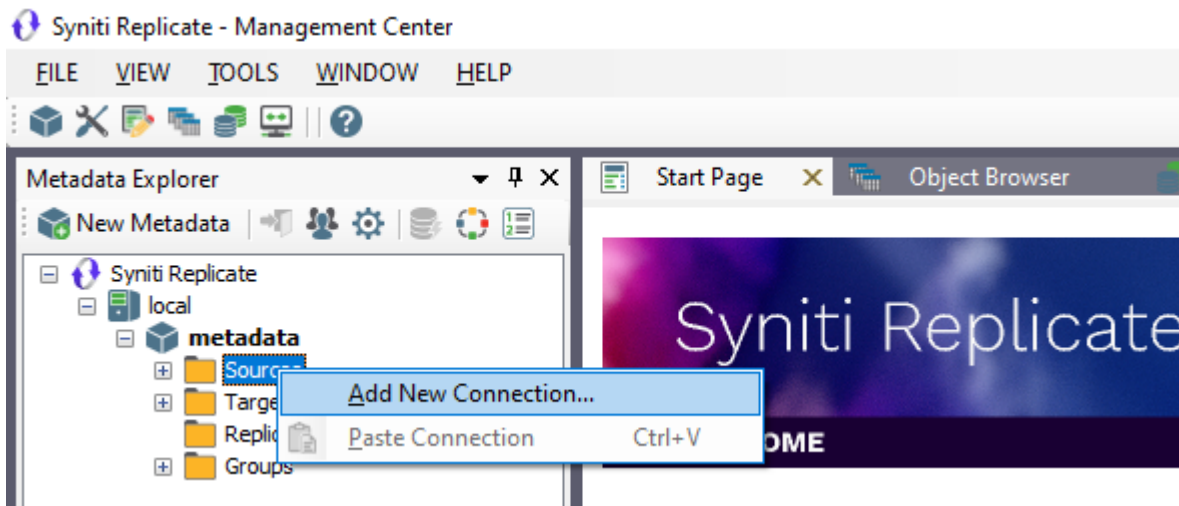
4. In the Add Source Connection Wizard, follow steps to add a connection string and test the connection to the database.
   Check the Supported Provider List in the Help Center before entering a value in the **Assembly** field.

Syniti Replicate

Add Source Connection Wizard ✕

Syniti Replicate

Select the database that contains source data to be replicated and indicate which provider to use.

- Select provider
- Set connection string
- Select tables
- Actions
- Summary

Source name

Name: IBMDB2i-Src

Data Provider(s)

Database: IBM Db2 for i

Provider: HiT Software .NET Driver (Ritmo/i)

Assembly: Ritmo_i/lib/Sql400.dll

🔍 Browse

< Back   Next >   Cancel   Help

5. In the **Select Tables** screen, choose the tables that you plan to replicate.

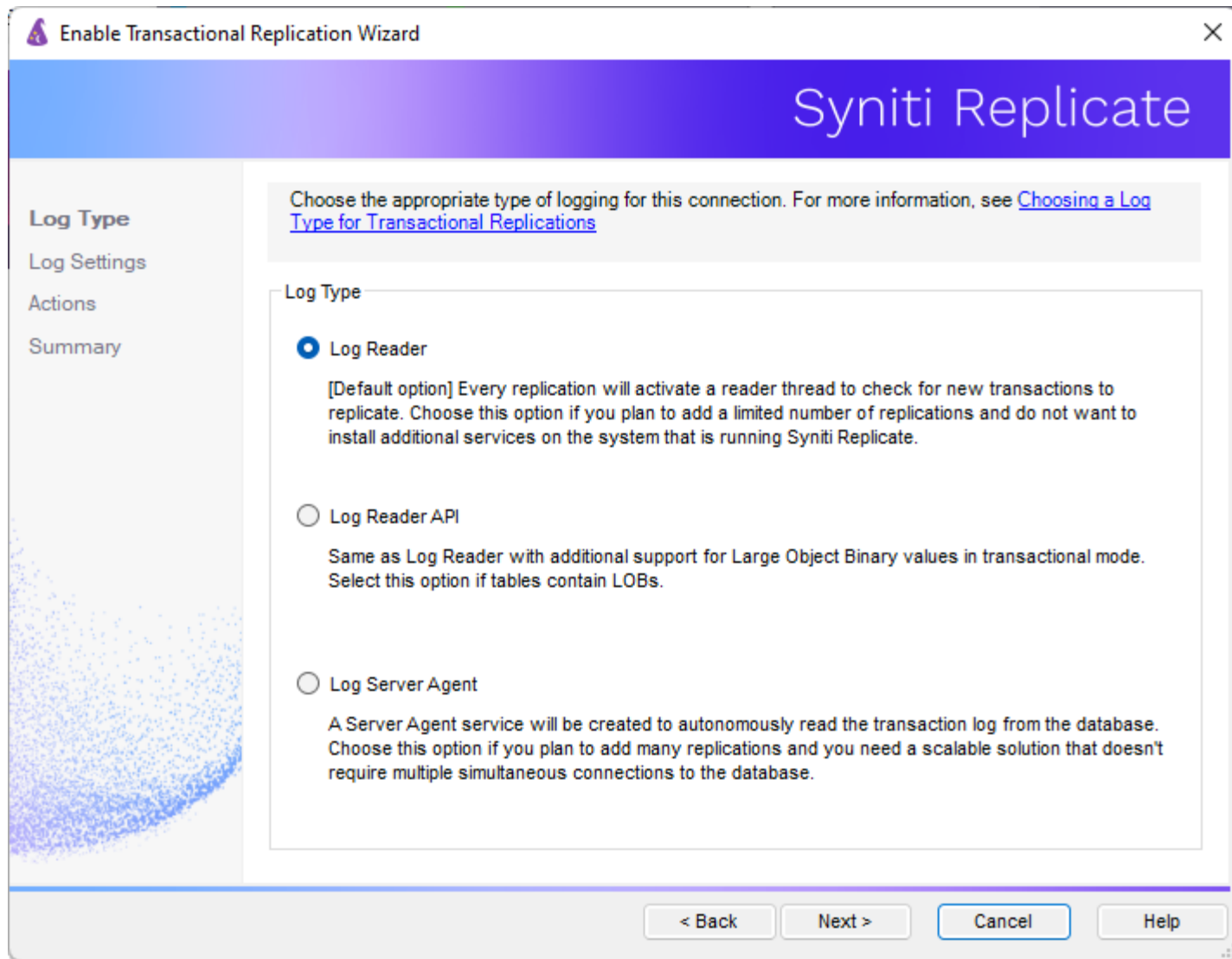6. In the **Actions** screen, check the option **Launch the Enable Transactional Setup Wizard.**

Syniti Replicate
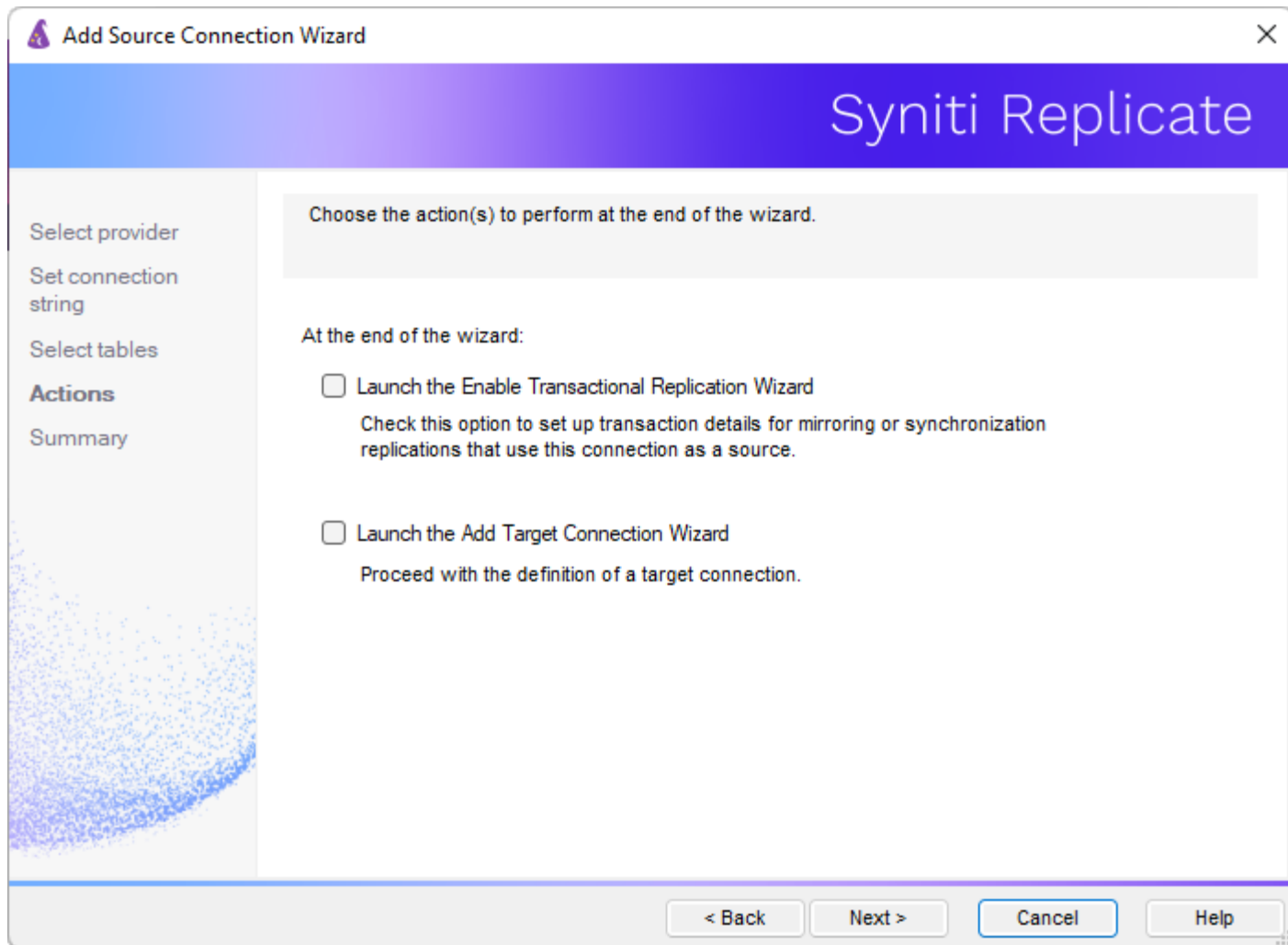
7. Complete the wizard.

Enable Transactional Replication

This section assumes you are planning to replicate data to a Cassandra 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 help for more information on the best option.
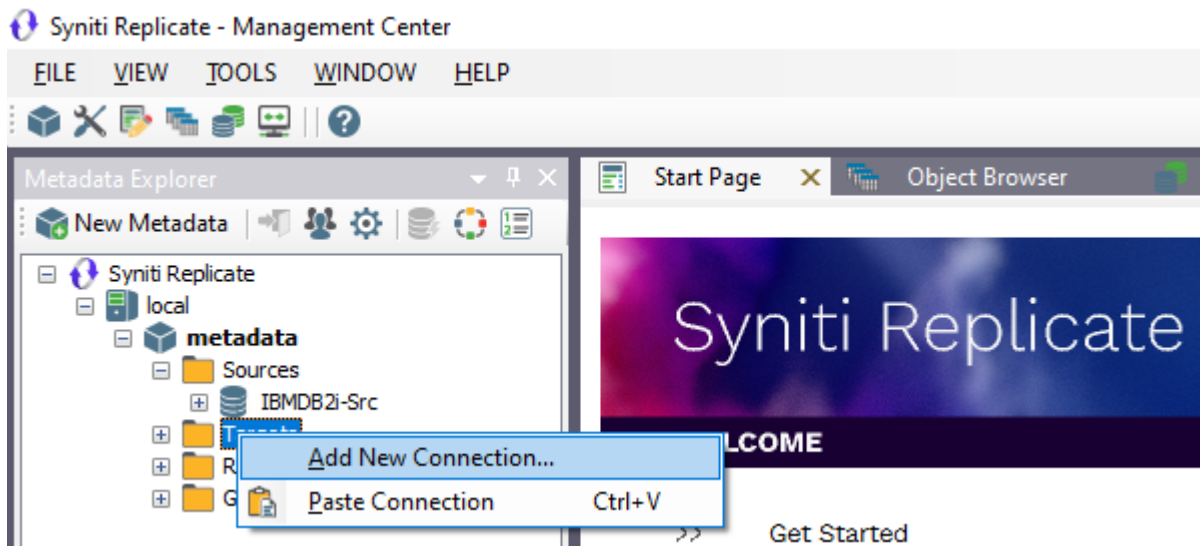
Syniti Replicate

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 Cassandra

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...**

Syniti Replicate

Syniti Replicate

1. In the Target Connection Wizard **Provider** field, select the **Cassandra** option.

2. Choose Cassandra from the drop-down list. The default provider is Cassandra CDATA ADO.NET Provider.

3. In the Assembly field, locate the file System.Data.CData.Cassandra.dll, typically in the installation folder for the CData ADO.NET Provider for Cassandra.

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

5. Add the Cassandra server name or IP Address and Port.

6. Enter a user ID and password for Cassandra.

7. If using SSL, set Use SSL to True and provide client and server certificate information.

8. Click **Next** to display the Select Tables screen.

9. Click **Next** to display the **Summary** page.

10. Click **Finish** to complete the wizard.

Syniti Replicate

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...**

NOTE: This section describes how to create a single replication. However, many situations call for setting up replications on a set of source tables. Use the **Create Multiple Replications** wizard accessible from the source connection in the Metadata Explorer for this purpose.

1. In the **Define Replication Type** screen, type a name to identify the replication.

2. Optionally provide a description of the replication.

3. In the Replication Mode area, choose **Continuous Mirroring**.

4. Click **Next** to go to the **Select Source Connection** screen.
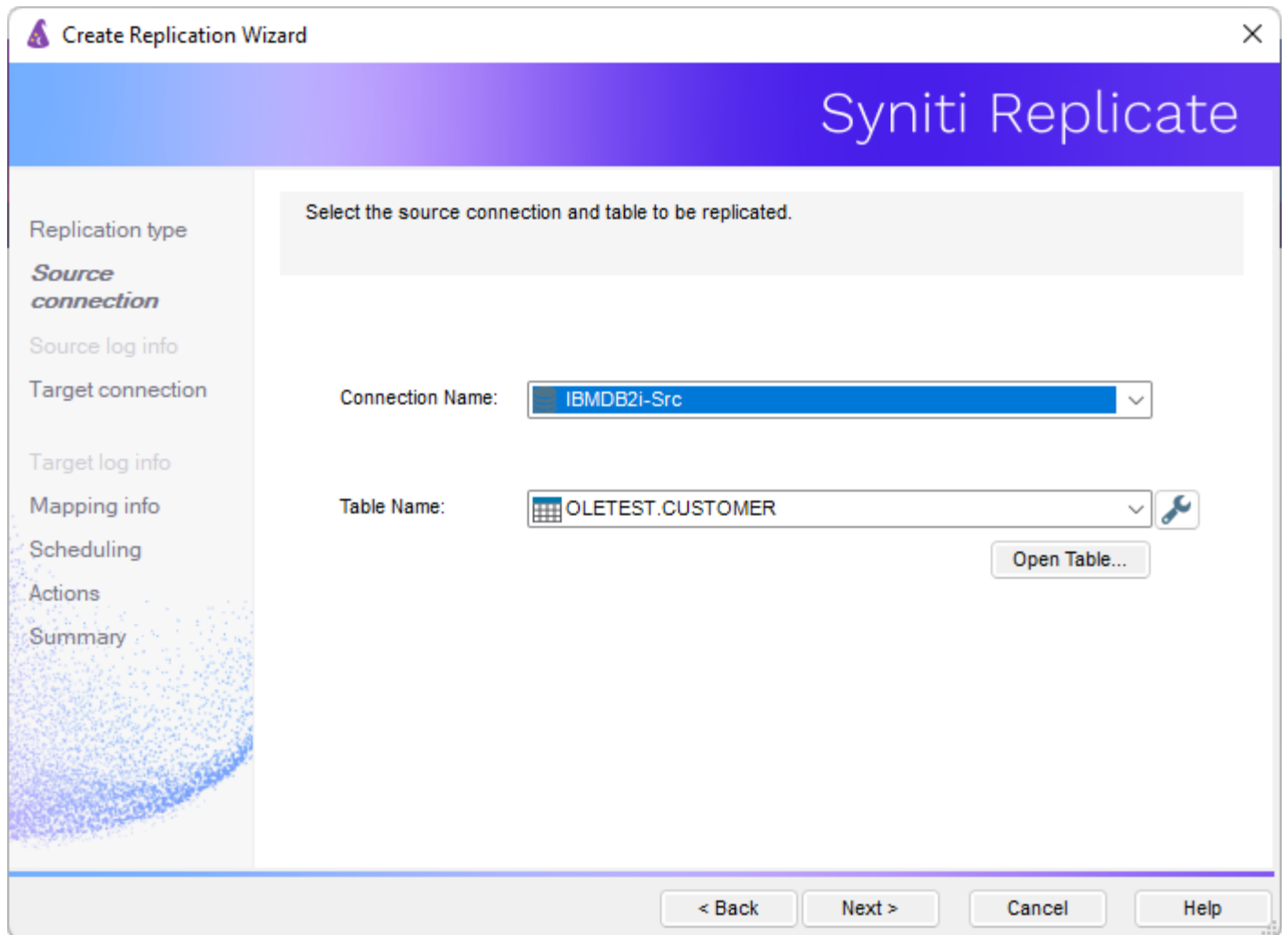
Syniti Replicate

5. Choose the source name from the drop-down list that includes all the source connections you have created in Syniti Replicate.

6. Choose the table that you want to replicate from the drop-down list.

7. If you want more information about the table before proceeding, click **Open Table...**.

8. Click **Next** to go to the **Source Log Info** screen.

Syniti Replicate

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.

9. In the Transaction ID field, click **Read** to open the Read Point dialog.

10. Choose either the current transaction or a transaction ID based on a time and date.

11. Click **OK** to add the value to the **Source Log Info** screen.

12. 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.

13. Click **Next** to go to the **Select Target Connection** screen.

14. Choose the target connection for Cassandra from the drop-down list that includes all the target connections you have created in Syniti Replicate.

15. 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.

16. Click **Next** to go to the **Set Mapping Info** screen.
    Source columns and target data with the same name are automatically mapped.

17. 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.

18. 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.

19. Go to the Mirroring Schedule tab.

20. 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.

21. Make sure the **Enable Replication** option is checked. This is required for the replication to run.

22. Click **Next** to go to the **Summary** screen.

23. Click **Finish** to complete the wizard.

Run the Replication

If you installed the Replication Agent as a service during setup, you just need to start the service using the Service Monitor program 🛢 in the Windows Notification Area.
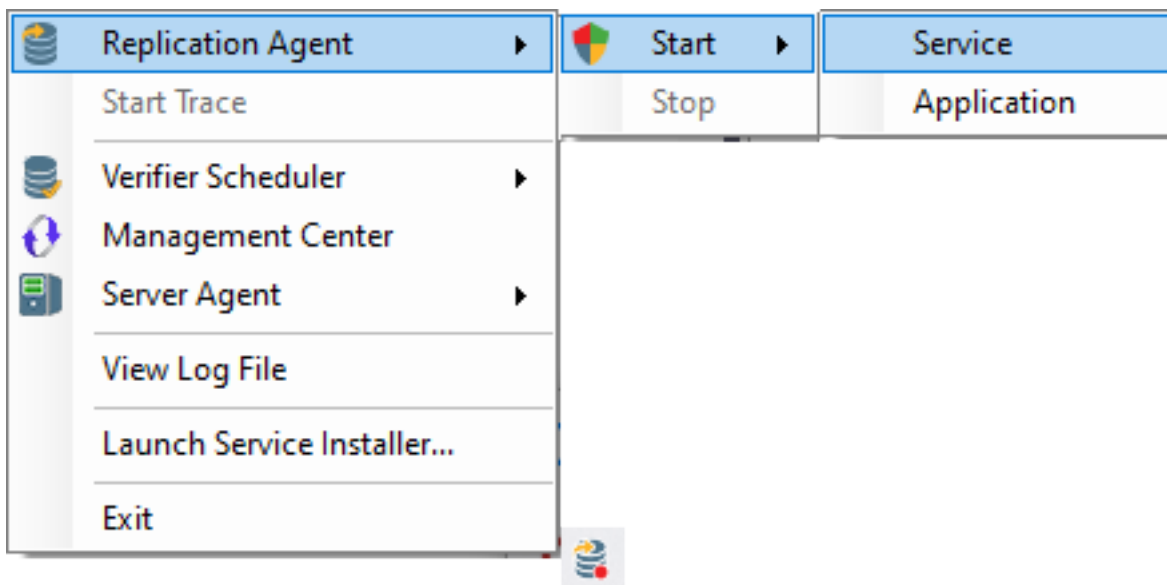
The replication that you have scheduled should start at the specified time. Use the Replication Monitor tab in the Management Center to track the progress of the replication.

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.