Syniti

Data Replication

Google BigQuery Setup Guide

Version 9.7.0.22

Software Release Date: 4/15/2020

Syniti | Formerly HiT SOFTWARE®

https://support.hitsw.com

Syniti Data Replication

Document History

Version	Author	Date	Reviewer	Date	Comments
1	JHLorenzin	4/14/20	VFarruggio	2/26/20	

Syniti Data Replication

Table of Contents

Syniti Data Replication

Introduction

This guide describes how to set up your Syniti Data Replication environment to replicate data to Google™ BigQuery™ cloud data warehouse. When replicating **from a relational database to BigQuery**, Syniti DR allows you to replicate data in two different ways, or modes:

- **Refresh**
  A one-time complete replication from source table to BigQuery, 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 BigQuery 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.

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, MySQL, Microsoft SQL Server, IBM Db2 for i, IBM Db2 LUW. For complete details on available source databases and the setup process, check the *Syniti Data Replication User Guide* available from the Management Center **Help** menu or the *Syniti Data Replication Setup Guide*, available for download in the Help Center.

Source tables loaded into a BigQuery connection are associated with an existing BigQuery dataset to indicate the location where the data will be loaded. The table definition for the BigQuery target is a virtualization of the data loaded to the target and does not correspond to existing physical objects. The tables displayed in Syniti DR show columns that match the columns from the source table (to facilitate the definition of mappings, expressions, scripting etc.), but the actual target table on BigQuery will have a slightly different schema, including some metadata tables (Transaction ID, Transaction Timestamp, User Id, Operation Type) and two subsets of record images (Before and After) to show the complete image of the record.

Data is loaded to BigQuery in **append** mode. This means that UPDATEs and DELETEs as well as INSERTs will actually insert new records. Each record is flagged with the proper operation type and filled with the proper record image values.

Basic Configuration Steps

Use Syniti DR Management Center to:

- Create source connections to RDBMS tables
- Create BigQuery targets
- Map RDBMS sources to BigQuery targets
- Enable replication

Subsequent data management on the BigQuery side depends upon your application needs.

Requirements for BigQuery Targets

Google client library to access the BigQuery API

1. Download the Google.Cloud.BigQuery.V2 (version 1.4.0) nuget package from:
   https://www.nuget.org/packages/Google.Cloud.BigQuery.V2/.

2. Copy the assemblies to the Syniti Data Replication Plugins/BigQuery folder. See below for the list of files.

BigQueryPlugin.dll
BigQueryPlugin.pdb
Google.Api.Gax.dll
Google.Api.Gax.Rest.dll
Google.Apis.Auth.dll
Google.Apis.Auth.PlatformServices.dll
Google.Apis.Bigquery.v2.dll
Google.Apis.Core.dll
Google.Apis.dll
Google.Apis.PlatformServices.dll
Google.Cloud.BigQuery.V2.dll
Newtonsoft.Json.dll
System.Interactive.Async.dll

AVRO™ Data Serialization

Optional – only needed if using AVRO serialization.

3. Download the Confluent.Apache.Avro (version 1.7.7.7) from:
   https://www.nuget.org/packages/Confluent.Apache.Avro/

1. Copy the assemblies to the Syniti Data Replication Plugins/AvroSerializer folder. See below for the list of files.

AvroSerializer.dll
Confluent.Apache.Avro.dll
Confluent.SchemaRegistry.dll
Confluent.SchemaRegistry.Serdes.dll

Replicating Data from a Relational Database to BigQuery

- Make sure you have database connections via a .NET data provider to the supported databases for source database tables.

- Start Syniti DR Management Center.
  Syniti DR provides a default database (Microsoft SQL Server CE) for your metadata, all the information that Syniti DR needs to store about your replication setup.

Syniti Data Replication
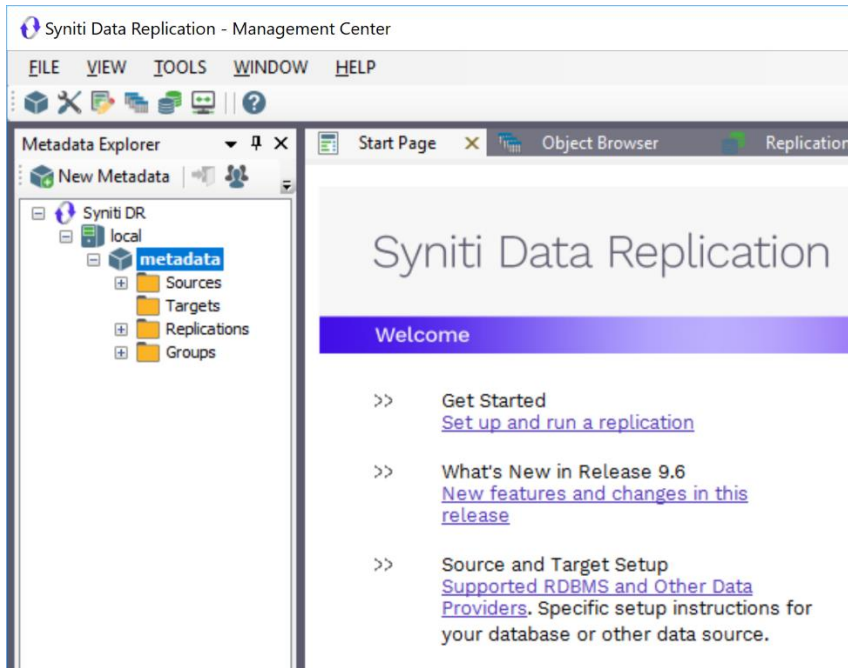
Create a Source Connection

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**.
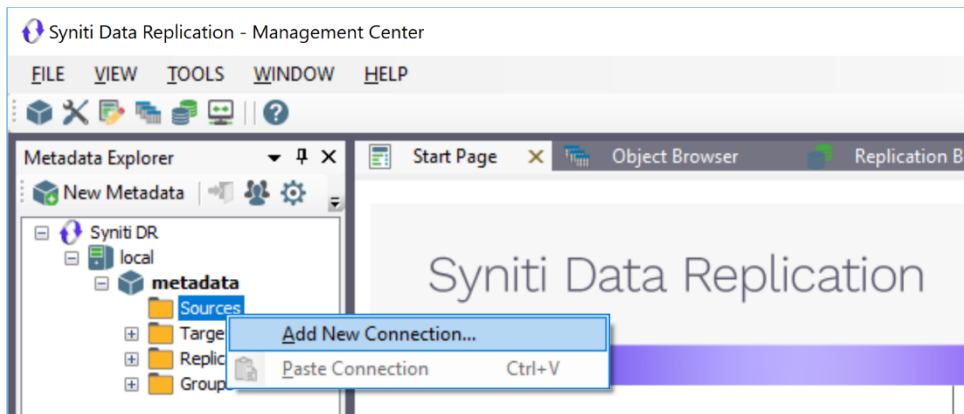

4. In the Add Source Connection Wizard, follow steps to add a connection string and test the connection to the database.
Check the Supported Provider List in the Help Center before entering a value in the **Assembly** field.

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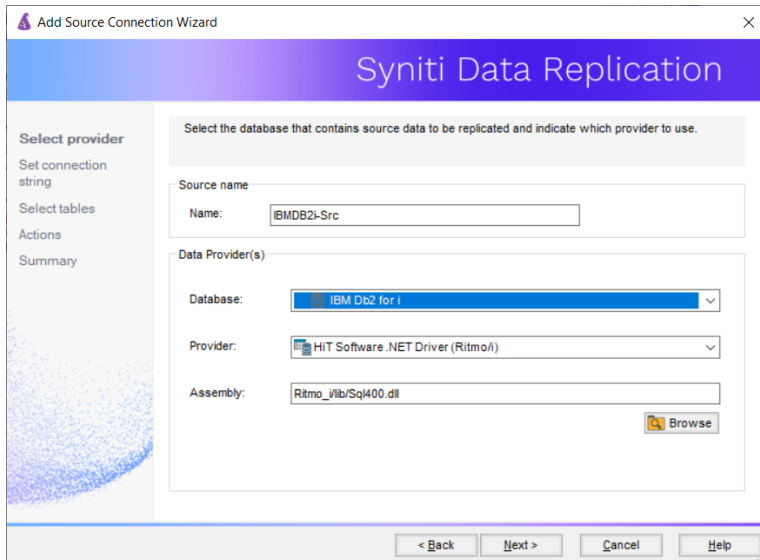
5. In the **Select Tables** screen, choose the tables that you plan to replicate.

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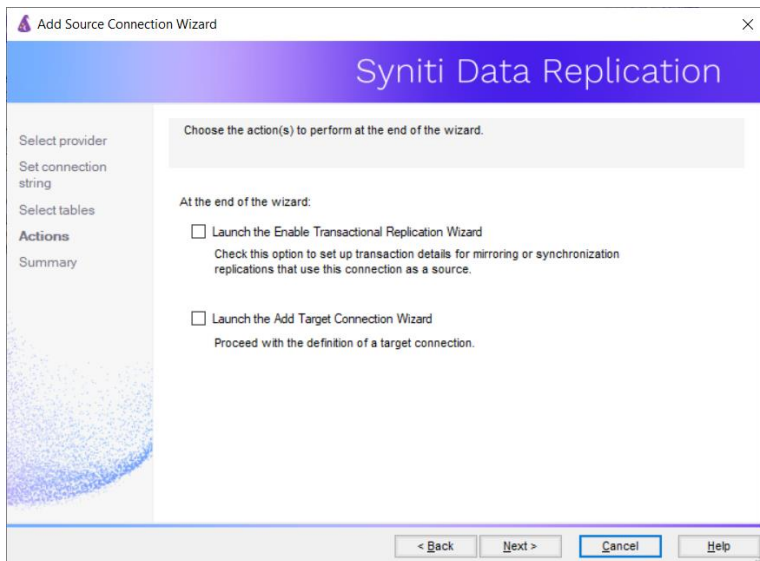

7. Complete the wizard.

Enable Transactional Replication

This section assumes you are planning to replicate data to a BigQuery 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**:

Syniti Data Replication

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.
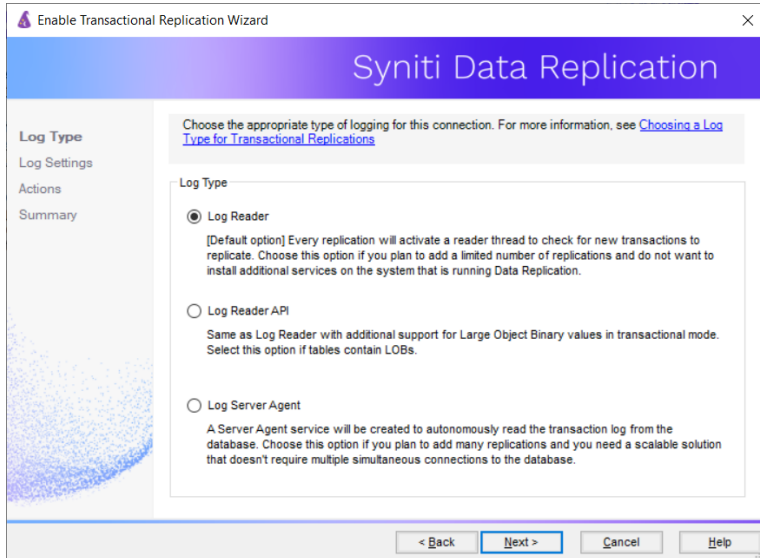
2. Click **Next** to enter the log settings. The fields and appropriate values depend on the database and log type. You can obtain a setup guide for a specific relational database by making a request in the Help Center.

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 BigQuery

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...**
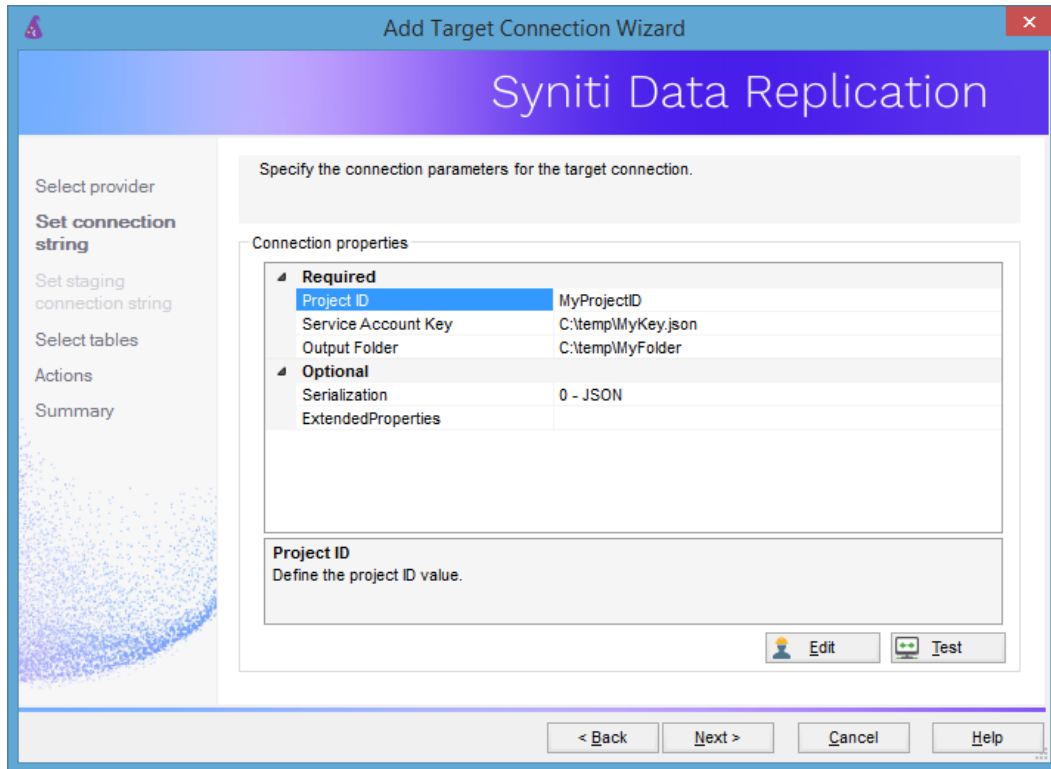
Syniti Data Replication
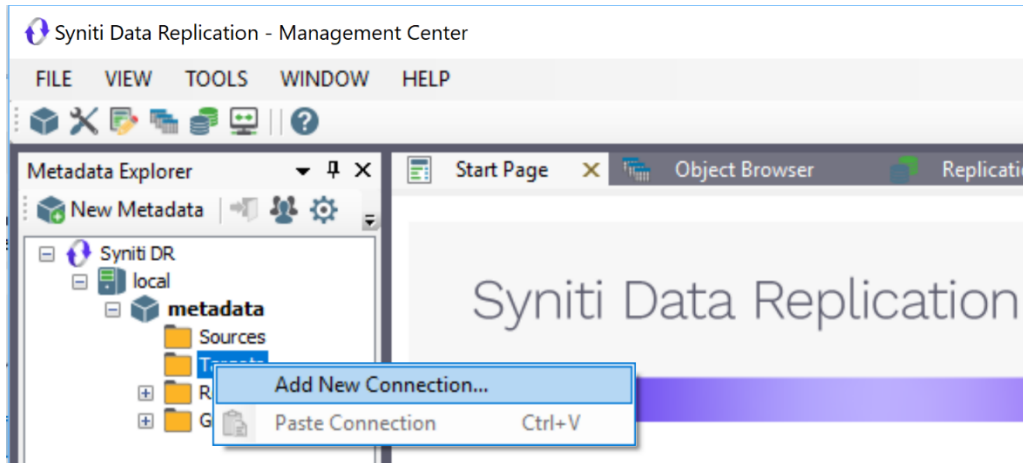
1. In the Target Connection Wizard **Provider** field, select the **Google BigQuery** option.
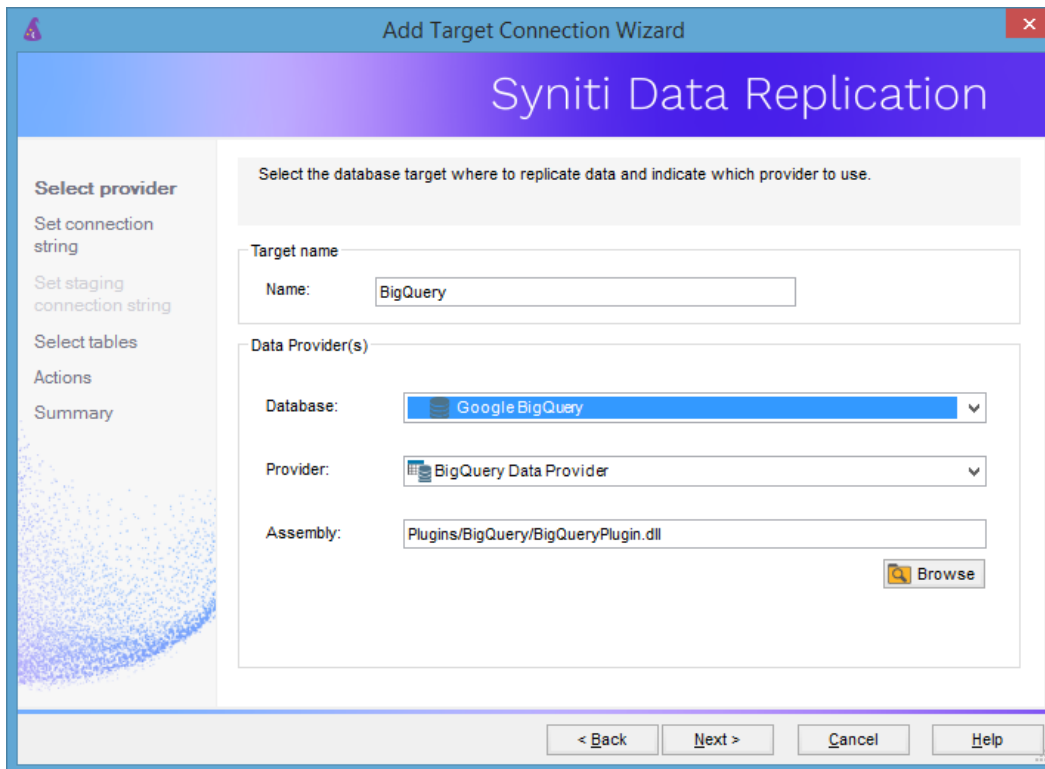
2. Click **Next** to display the **Set connection string** screen.

Syniti Data Replication

3. Set values for the following fields:

Project ID	The ID of the BigQuery project under which you want to upload data.
Service Account Key	Path to the local service account key file.
Output folder	Local folder where to create the temporary files before loading them to BigQuery.
Serialization	The type of serialization used to send data to BigQuery. Options available are: • CSV (Comma Separated Values) • JSON • AVRO (Typically recommended for performance and to maintain and support compatibility with different versions of schemas.)

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.

Syniti Data Replication

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 DR 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...**

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**.

Syniti Data Replication
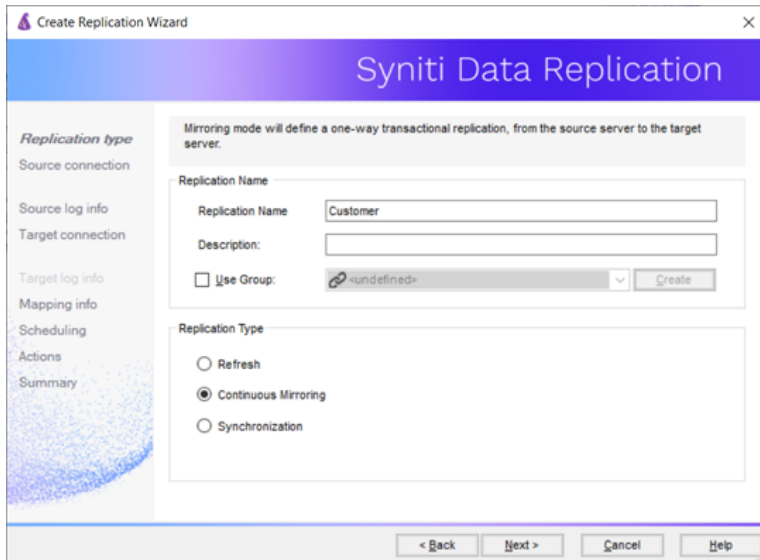
4. Click **Next** to go to the **Select Source Connection** screen.
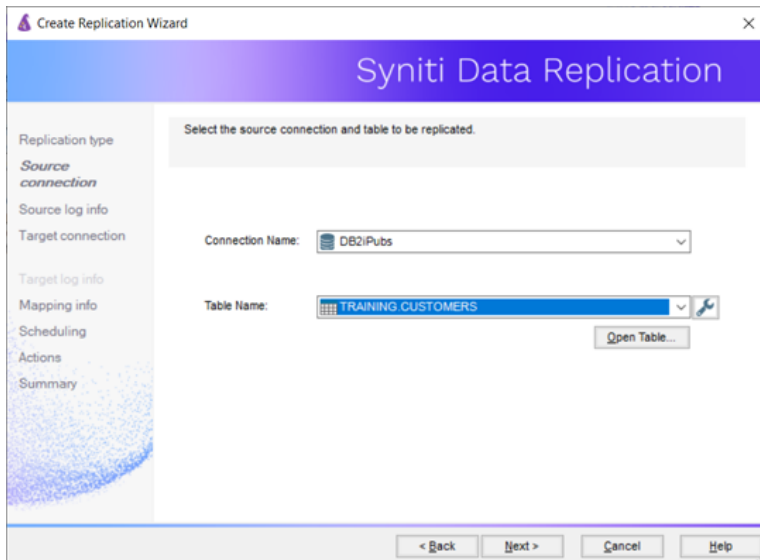
5. Choose the source name from the drop-down list that includes all the source connections you have created in Syniti DR.

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 Data Replication

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 DR 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 BigQuery from the drop-down list that includes all the target connections you have created in Syniti DR.

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.

Syniti Data Replication

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.
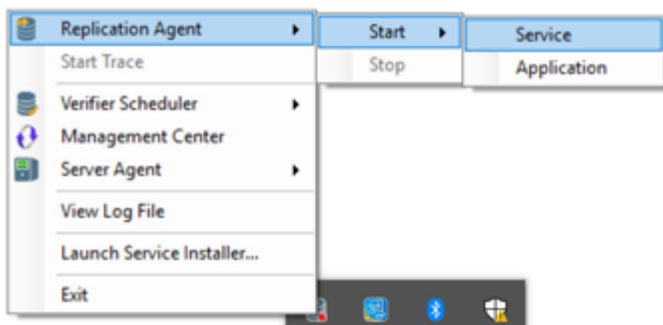
The replication that you have scheduled should start at the specified time. Use the Replication Monitor tab in the Management Center to track the progress of the replication.

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

- From the Windows Desktop **Start** menu, choose **Syniti Data Replication**, then **Service Installer**.

- Manage the service from Service Monitor program (located in the Syniti DR 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.

JSON/AVRO File Structure

If the target data has been serialized using JSON or AVRO, the files are named with either a.ref or a .mir extension. For JSON format, output is a text file; for AVRO, the file is in binary format. Both follow the structure shown below.

Column	Value
1	Transaction ID
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)
AFTER	
6 …	All columns from the database table with the after image of the record
BEFORE	
6	All columns from the database table with the before image of the record

Flat File Structure

Data files replicated to BigQuery using mirroring with CSV serialization 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

Syniti Data Replication

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"
```

Last Updated on 4/14/20