



https://support.hitsw.com

Document History

Version	Author	Date	Reviewer	Date	Comments
1	JHLorenzin	9/20/21	AFerrari	9/20/21	

Table of Contents

Connection Type	3
Replicating Data from a Relational Database to Parquet	3
1. Set Up a Source Connection to a Relational Database	3
Enable Transactional Replication	6
2. Set up a Target Connection for Parquet	6
3. Add Table Information to the Target Connection	9
4. Define Replications	9
5. Run the Replication1	2
Parquet File Structure	3

This guide describes how to set up your Syniti Data Replication environment to replicate data to Apache Parquet files.

When replicating **from a relational database to a Parquet file**, Syniti DR allows you to replicate data in two different ways, or modes:

- **Refresh** A one-time complete replication from source table to a Parquet file, 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 a Parquet file 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 file.

Connection Type

Parquet.NET provider 3.9.0

- 1. Go to https://www.nuget.org/packages and search for "Parquet.NET".
- 2. Download version 3.9.0. The easiest way to do it is by clicking the Downloads menu on top of the page and get the nuget.exe command line tool.
- 3. From a DOS prompt with admin permissions, execute the following command.
 nuget install Parquet.Net -Version 3.9.0 -Framework net480 -OutputDirectory
 "C:\Temp\Parquet 3.9.0"

The command will download the entire provider into the specified output directory.

- 4. Open this directory to view the folders where the assemblies are located:
- 5. Collect the following assemblies selecting them from the shown path:

Parquet\IronSnappy.1.3.0\lib\netstandard2.0.IronSnappy.dll

Parquet\Parquet.Net.3.9.0\lib\netstandard2.0.Parquet.dll

Parquet\System.Buffers.4.4.0\lib\netstandard2.0.**System.Buffers.dll**

Parquet\System.Memory.4.5.3\lib\netstandard2.0.System.Memory.dll

Parquet\System.Numerics.Vectors.4.4.0\lib\netstandard2.0.System.Numerics.Vectors.dll

Parquet\System.Runtime.CompilerServices.Unsafe.4.5.2\lib\netcoreapp2.0.**System.Runtime.CompilerServices.**

6. Copy these files into the Syniti Data Replication install folder Plugins/ParquetSerializer folder.

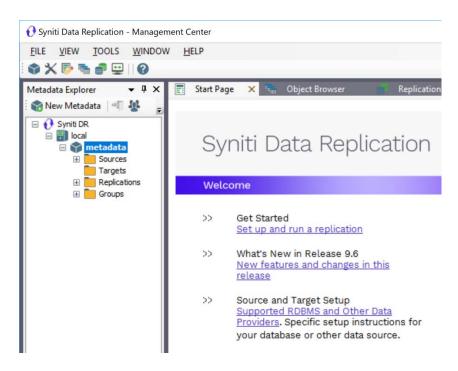
Replicating Data from a Relational Database to Parquet

1. Set Up a Source Connection to a Relational Database

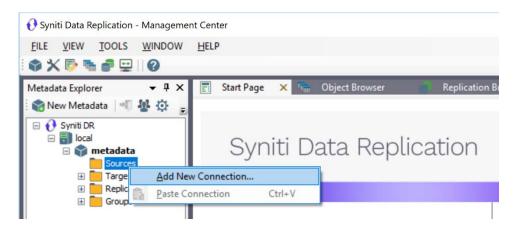
- 1. Make sure you have database connections via a .NET data provider to your source database. For each database you are planning to use in your replication project:
 - Install and configure one of the supported data access products. See <u>the list of supported .NET data</u> <u>providers in the knowledge base</u> for a current list of supported providers.
 - From the data access product, test the connection to the database.
 - Create a connection string for the data access product/database you are using. Check the documentation for the data access product for information on how to do this.

- Check that the user ID you are planning to use has sufficient permissions to complete all operations in Syniti DR. Contact the technical support team via the <u>Help Center</u> for specific requirements for your database.
- 2. 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.



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



6. In the Source Connection Wizard, follow steps to add a connection string and test the connection to the database.

Check the knowledge base article on data providers before entering a value in the Assembly field.

		Syniti Data Replication	
elect provider	Select the dat	abase that contains source data to be replicated and indicate which provider to use.	
et connection ring elect tables	Source name Name:	IBMDB2i-Src	
ctions ummary	Data Provider(3)	
	Database:	IBM Db2 for i	
	Provider:	HIT Software .NET Driver (Ritmo/i)	
	Assembly:	Ritmo_Vilb/Sql400.dll	
and the state			

7. In the Set Connection String screen, fill out the following fields:

Connection Properties

Edit at least the **Required** connection properties by clicking in the property value field and typing a new value. **Note**: If using an Oracle 11 client, in the Data Source field, be sure to add the **Oracle Service Name** after the IP address.

- 8. Choose the tables that you plan to replicate.
- 9. In the Actions screen, check the option Launch the Enable Transactional Setup Wizard.

💰 Add Source Connectio	n Wizard	×
	Syniti Data Replication	
Select provider	Choose the action(s) to perform at the end of the wizard.	
Set connection string Select tables Actions Summary	At the end of the wizard: Launch the Enable Transactional Replication Wizard Check this option to set up transaction details for mirroring or synchronization replications that use this connection as a source.	
	Launch the Add Target Connection Wizard Proceed with the definition of a target connection.	
	< Back Next > Cancel Hel)

10. Complete the wizard.

Enable Transactional Replication

This section assumes you are planning to replicate data to Parquet 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.

💰 Enable Transactional	Replication Wizard	×
	Syniti Data Replication	
Log Type	Choose the appropriate type of logging for this connection. For more information, see <u>Choosing a Log</u> Type for Transactional Replications	
Actions	Log Туре	
Summary	Log Reader	
	[Default option] Every replication will activate a reader thread to check for new transactions to replicate. Choose this option if you plan to add a limited number of replications and do not want to install additional services on the system that is running Data Replication.	
	O Log Reader API	
	Same as Log Reader with additional support for Large Object Binary values in transactional mode. Select this option if tables contain LOBs.	
No. 1 All	O Log Server Agent	
Constant of	A Server Agent service will be created to autonomously read the transaction log from the database. Choose this option if you plan to add many replications and you need a scalable solution that doesn't require multiple simultaneous connections to the database.	
	< <u>Back</u> <u>N</u> ext > <u>C</u> ancel <u>H</u> el	•

- 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 <u>Help Center</u>.
- 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.

2. Set up a Target Connection for Parquet

- 1. Select the **Targets** node.
 - 2. From the right mouse button menu, choose **Add New Connection**.

😯 Syniti Data Replication - Manageme	ent Center
FILE VIEW TOOLS WINDOW	HELP
🕸 🗙 🖻 🐂 💣 🖳 1 🛛	
Metadata Explorer 🔹 🖣 🗙	📑 Start Page 🗙 🐃 Object Browser 👘 Replication
🛙 😭 New Metadata 🛛 📲 🍄 🍦	
Syniti DR Synit DR Syniti DR Syniti DR Syniti DR Syniti DR Syniti DR	Syniti Data Replication
Sources	onnection

3. In the Target Connection Wizard **Provider** field, select the **Files - Parquet** option.

		Syniti Data Replication	
Select provider	Select the da	abase target where to replicate data and indicate which provider to use.	
Set connection string	Target name		
	Name:	ParquetConn	
Select tables	Data Provider(s)	
lctions			
Summary	Database:	Files - Parquet V	
	Provider:	Parquet Data Provider 🗸	
	Assembly:		
		Q Browse	
and the second	i i		
and the state			
STATE Conserved			

4. In the **Set Connection String** page, set a directory path for the output from replications between the source relational database and the Parquet file target.

Add Target Connect			
	Sy	niti Data Replication	
elect provider	Specify the connection parameters for	the target connection.	
et connection			
tring	Connection properties		
	✓ Required		
	Output Folder		
elect tables	Add Transactional Info	Yes, at the beginning	
iect tables	✓ Optional		
tions	Compression	0 - None	
immary	Use Nullable fields	0 - False	
internary	Use DateTime fields	0 - False	
	Use One File Per Group	False	_
	ExtendedProperties		
	Optional		
Sector State			_
		👱 Edit 😐 Test	
Charlen			
		< Back Next > Cancel H	elp

5. Complete the connection properties using the information below.

Output Folder	The schema name and location to the Parquet files. Set an Output Folder available to the system where Syniti DR is running.
Add transactional info	Make sure that the Add Transactional Info field is set to Yes, at the beginning.
Compression	The compression method desired, Currently supported 0 - None, 1 – Gzip, 2 – Snappy
	• None for no compression. This is the fastest way to write files, however they may end up slightly larger.
	 Snappy is the default level and is a perfect balance between compression and speed.
	• Gzip, using gzip compression, is the slowest, but should produce the best results if maximum compression is your top priority.
Use Nullable Fields	Set to 1 - True to allow nullable fields in Parquet files.
Use DateTime Fields	Set to 1 – True to allow datetime fields in Parquet files. If 0 - False is selected, Strings are used instead of datetime values.
Use One File Per Group	This property is not implemented for V. 9.8.1

6. Click **Next** to display the **Select tables** page.

At this point, there is no output structure available to display. You can add the information after completing the Target Connection wizard.

- 7. Click Next to display the Summary page.
- 8. Click **Finish** to complete the wizard.

3. Add Table Information to the Target Connection

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

- I. 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.
- 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.
- 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 target connection in the Metadata Explorer.

4. Define Replications

- 1. Expand the Metadata Explorer tree to display the table that contains the data you want to replicate.
 - 2. Select the table.
 - 3. From the right mouse button menu, choose Replication then Create New Replicaton....
 - 4. In the **Define Replication Type** screen, type a name to identify the replication.
 - 5. Optionally provide a description of the replication.
 - 6. In the **Replication Mode** area, choose **Mirroring** to capture changed data only.

		Syniti D	ata Re	eplicat	ion
Replication type Source connection	Mirroring mode will defi server. Replication Name	ne a one-way transactional repl	cation, from the sou	irce server to the t	arget
Source log info	Replication Name	Customer			
Target connection	Description:				
	Use Group:	🔗 sundefined>		√ <u>C</u> re	ate
Mapping info					
Scheduling	Replication Type				
Actions	O Refresh				
Summary	Continuous Mirro	rina			
-	O Synchronization				
		< Back	Next >	Cancel	Help

7. Click Next to go to the Select Source Connection screen.

Create Replication W	izard				
		Syniti I	Data Re	eplicat	tion
Replication type	Select the source conne	ection and table to be replicate	ŀd.		
Source connection					
Source log info					
Target connection	Connection Name:	DB2iPubs			~
Mapping info	Table Name:	TRAINING CUSTOME	RS		~ &
Scheduling				Open Table	
Actions					
Summary					
and a state of the					
State and					
		< Back	Next >	Cancel	Help

- 8. Choose the source connection name from the drop-down list that includes all the source connections you have created in Syniti DR.
- 9. Choose the table that you want to replicate from the drop-down list.
- 10. If you want more information about the table before proceeding, click **Open Table...**.
- 11. Click Next to go to the Source Log Info screen.

		Syniti L	Data Repl	ication
Replication type Source connection <i>Source log info</i> Target connection		ent transaction read point fror ID from which to replicate.	m the IBM VAS400 server. To	override, click Read
Target log info	Journal: Receiver:	TRAINING.QSQJRN		Bead TID
Mapping info Scheduling Actions Summary	Receiver. Transaction ID: Transaction Timestamp: Read Interval (sec):	TRAINING QSQJRN0001 23 7/15/2019 8:08:20 PM 60		

The first two fields on this screen depend on the source database you are using. In this explanation, we assume you are using Microsoft SQL Server. Check the help for the Replication Wizard for details on the other databases.

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

🔟 Transaction Read Point	×	
Current Sequence	<u>о</u> к	
○ Date (dd/mm/yyyy) and Time	<u>C</u> ancel	
Date 8/30/2012 V Time 12:00:00 AM		
<u>R</u> efresh Transactional Replication Objects		
Keep Current Sequence		

- 13. Choose either the current transaction or a transaction ID based on a time and date.
- 14. Click **OK** to add the value to the **Source Log Info** screen.
- 15. 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.
- 16. Click Next to go to the Select Target Connection screen.
- 17. Choose the target connection for text output from the drop-down list that includes all the target connections you have created in Syniti DR.
- 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.
- Click Next to go to the Set Mapping Info screen.
 Source columns and target data with the same name are automatically mapped.
- 20.Click Next to go to the Scheduling screen.

- 21. Make sure the **Enable Replication** option is checked. This is required for the replication to run.
- 22.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.

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.

23.Go to the Mirroring Schedule tab.

24. 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.

25. Click Next to go to the Summary screen.

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

5. 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 sin the Windows Notification Area.

8	Replication Agent	•	Start	•	Service
	Start Trace		Stop		Application
8	Verifier Scheduler	•		-	
9	Management Center				
1	Server Agent	•			
	View Log File				
	Launch Service Installer				
	Exit				and the second se

The replication that you have scheduled should start at the specified time. Use the <u>Replication Monitor</u> 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 <u>Replication Monitor</u> 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.

Parquet File Structure

The file format generated by the process will have a .parquet extension. In addition, files used as a replication target have the following structure. The file has a .mir.parquet extension when used for mirroring and a .ref.parquet 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), UB and UA (Update before and after image) or D (Delete)
6	Record image

Last Updated on 11/30/21