Syniti

Syniti Replicate

Setup Guide for Replicating to Hadoop

Version 10.2



Table of Contents

Setup Guide for Replicating to Hadoop1	1
Introduction1	1
Basic Configuration Steps	2
Connection Type2	2
Setup Summary2	2
Steps for Replicating Tables	3
1. Set Up a Source Connection to a Relational Database3	3
2. Enable Transactional Replication6	6
3. Set up a Target Connection to Hadoop HDFS7	7
4. Add Table Information to the Target Connection11	11
5. Define Replications	2
Start Replications	8
Stop Replications	9

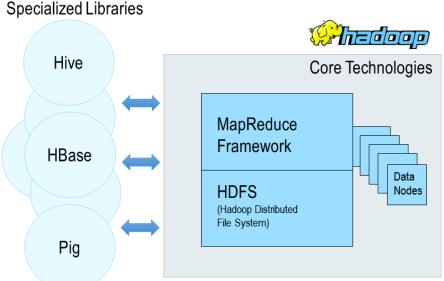
Introduction

Syniti Replicate allows you to replicate data from relational database tables to the Hadoop Distributive File System (HDFS) using:

- **Refresh**, or **Snapshot**, replication: a one-time complete replication from any major relational database source to HDFS as a target, according to replication settings and scripts.
- **Mirroring**, or **Change Data Capture**, replication: a continuous read of changes to the source database that have been recorded in the database server log. Any changes found in the log are applied to HDFS as a target, according to replication settings and scripts.

The approach to replicating data provides a flexible, configurable solution which takes advantage of stable core Hadoop technology (HDFS) to:

- By-pass evolving specialized libraries
- Work on all distributions of Hadoop
- Allow Hadoop to manage data distribution

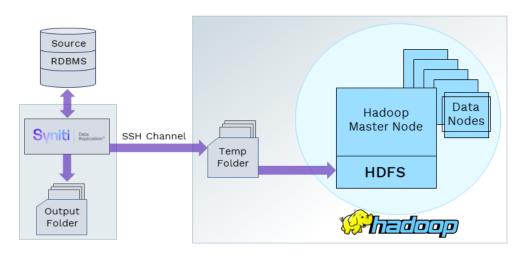


You can control the timing of the replication, identify the columns to be replicated and add scripts to transform data during replication. Source databases include Oracle, Microsoft SQL Server, IBM DB2 for i, IBM DB2 LUW, Sybase, Informix, MySQL. When HDFS is used as a target, Syniti Replicate transfers data as CSV files then uses the

Syniti Replicate File Broker service to communicate with HDFS.

Syniti Replicate replication to Hadoop is supported in Hadoop Version 2.6.0 or above.

Copyright© 2023 by BackOffice Associates, LLC d/b/a Syniti and/or affiliates. All Rights Reserved. This document contains confidential and proprietary information and reproduction is prohibited unless authorized by Syniti. Names appearing within the product manuals may be trademarks of their respective owners.



Basic Configuration Steps

Use Syniti Replicate Management Center to:

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

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

Connection Type

Syniti proprietary connector: Hadoop Data Provider. There is no need to install any additional connection software.

Setup Summary

This section provides a summary of all the steps required for setting up and using Syniti Replicate. Use the link for each step for more information.

Install. NET Provider for Source Database	Help Center Database Access Provider List (.NET and ODBC)				
Download and Install Syniti Replicate	 The <u>Knowledge Platform Product Suites article</u> acts as a hub to point to various resources. To download and/or register Syniti Replicate, log in to the support site, then click the relevant link in the Replicate section of the article. <u>Syniti Knowledge Base</u> <u>Enter a generic support ticket</u> 				
Syniti Replicate Setup	In the Syniti Replicate Management Center:				
	1. In the Metadata Explorer, create a source connection to your RDBMS.				

(See <u>Steps for</u> <u>Replicating Tables</u> below)	 Create a target connection using the provider "Hadoop HDFS". Create a replication.
Start Replicating	In the Syniti Replicate Service Monitor:
(See <u>Start Replications</u>)	1. Start the Syniti Replication Agent.

Steps for Replicating Tables

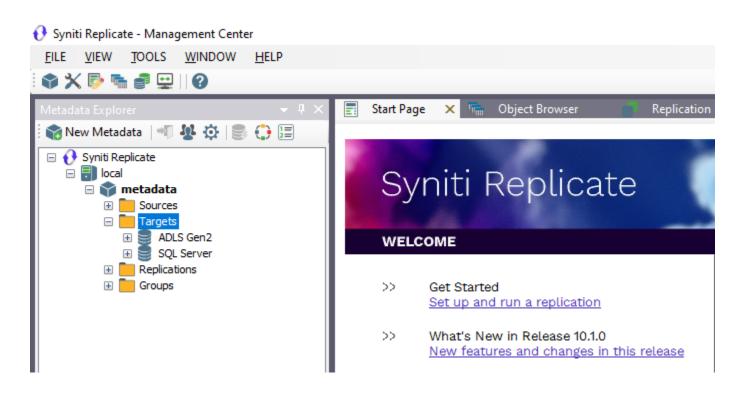
Syniti Replicate support for replicating relational data to Hadoop HDFS allows you to set up replications using either **Refresh** or **Mirroring** modes.

The steps below explain how to replicate data from a relational database to the Hadoop HDFS environment. Check <u>the Help Center</u> for the latest list of supported databases.

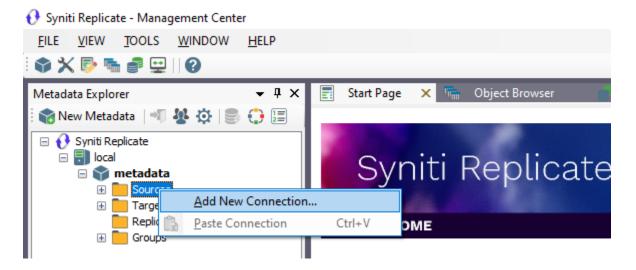
1. Set Up a Source Connection to a Relational Database

- 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 your .NET Provider.
 - From the provider, 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 Replicate. Contact the support team via the <u>Help Center</u> for specific requirements for your database.
- 2. 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.



- 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 Supported Provider List in the <u>Help Center</u> before entering a value in the **Assembly** field.

💰 Add Source Connectio	on Wizard	×
		Syniti Replicate
Select provider	Select the datat	ase that contains source data to be replicated and indicate which provider to use.
string	Source name	
Select tables	Name:	IBMDB2i-Src
Actions Summary	Data Provider(s)	
	Database:	IBM Db2 for i
	Provider:	HiT Software .NET Driver (Ritmo/i)
	Assembly:	Ritmo_i/lib/Sql400.dll
Constantine and		
		< Back Next > Cancel Help

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

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

💧 Add Source Connect	tion Wizard X
	Syniti Replicate
Select provider Set connection	Choose the action(s) to perform at the end of the wizard.
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 Help

10. Complete the wizard.

2. Enable Transactional Replication

This section assumes you are planning to replicate data to an HDFS 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.

💰 Enable Transactional I	Replication Wizard X	
	Syniti Replicate	
Log Type Log Settings	Choose the appropriate type of logging for this connection. For more information, see <u>Choosing a Log</u> Type for Transactional Replications	
Actions	Log Type	
Actions 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 Syniti Replicate. 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. Log Server Agent 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. 	
	< Back Next > Cancel Help	

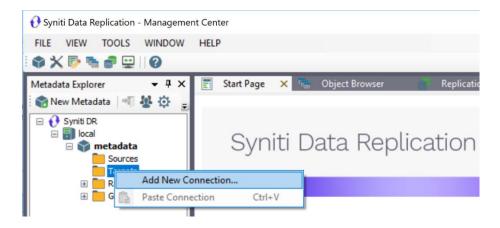
- 2. Click **Next** to enter the log settings. The fields and appropriate values depend on the database and log type. You can obtain a setup guide for a specific relational database by making a request in the <u>Help Center</u>.
- 3. Click **Next** to verify your settings against the source connection 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.

3. Set up a Target Connection to Hadoop HDFS

1. Select the Targets node.

2. From the right mouse button menu, choose Add New Connection.



3. In the Add Target Connection Wizard **Provider** field, select the **Hadoop HDFS** option.

💰 Add Target Connection	Wizard	×
		Syniti Replicate
Select provider	Select the databa	ase target where to replicate data and indicate which provider to use.
Set connection string	Target name	
Set staging connection string	Name:	Hadoop
Select tables	Data Provider(s)	
Actions	Database:	
Summary	Database.	Hadoop HDFS V
	Provider:	Hadoop Data Provider 🗸 🗸
	Assembly:	
		C Browse
S. S. Salahara		
		< Back Next > Cancel Help

4. In the **Set Connection String** page, set properties as described in the table below.

💰 Add Target Connecti	on Wizard X
	Syniti Replicate
Select provider	Specify the connection parameters for the target connection.
string	Connection properties
Set staging connection string Select tables Actions Summary	Required Output Folder Hostname Username Password Private Key File Path Path To Binary Target Directory Working Directory Output Folder Archive Use Column Names
and the second	Output Folder Path for the schema and output files. <u> <u> Edit</u> <u> <u> Edit</u> <u> <u> </u> </u></u></u>
	< <u>B</u> ack <u>N</u> ext > <u>C</u> ancel <u>H</u> elp

Output Folder	An existing folder on the system that is running Syniti Replicate for files associated with replications to Hadoop.				
Hostname	server name for the system running Hadoop.				
Username	The user name for the Hadoop instance.				
Password_KeyFile	Either a password or more typically a key file (.ppk extension)				
Path to Binary	Pathname to the Hadoop binary, ending in hadoop e.g. /home/hadoop-2.7.7/bin/hadoop				

	Use the command 'which hadoop' from the SSH command line to find the location of the executable file.				
Target Directory	HDFS directory where files will be uploaded.				
Working Directory	Temporary server directory where files will be stored before moving them to HDFS. Files are managed by Syniti Replicate.				
Output Folder Archive	Optional. Provides a local copy of data replicated to Hadoop. Data is not managed by Syniti Replicate, so the files must be managed manually and could grow quickly.				
Add Transactional Info	Set to Yes if performing mirroring replications				

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

4. 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.

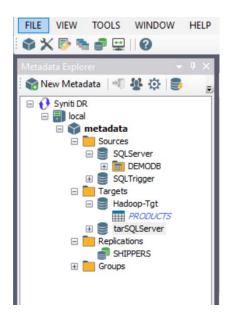
Although you are not actually creating tables, you can use this wizard to create a representation of the file data.

- 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.

- 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 target connection in the Metadata Explorer.



5. 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 Refresh or Continuous Mirroring.

💰 Create Replication Wizard X				
		Syniti Replicate		
Replication type Source connection	server. Replication Name	e a one-way transactional replication, from the source server to the target		
Source log info Target connection Target log info	Replication Name Description: Use Group:	CUSTOMER		
Mapping info Scheduling Actions Summary	Replication Type Refresh Continuous Mirror	ng		
Contraction of the second	Synchronization			
		< Back Next > Cancel Help		

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

Copyright© 2023 by BackOffice Associates, LLC d/b/a Syniti and/or affiliates. All Rights Reserved. This document contains confidential and proprietary information and reproduction is prohibited unless authorized by Syniti. Names appearing within the product manuals may be trademarks of their respective owners.

💰 Create Replication Wiz	zard				×
			Syniti	Replica	ate
Replication type	Select the source conne	ection and table to be replicated.			
<i>connection</i> Source log info					
Target connection	Connection Name:	IBMDB2i-Src		~	
Target log info Mapping info	Table Name:	OLETEST.CUSTOMER		~	a contraction of the second se
Scheduling				Open Table	_
Summary					
Starting and a second					
		< 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 Replicate.
- 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 (if using Continuous Mirroring). Complete the fields in this screen only if you are setting up a mirroring replication. The fields displayed

depend on the source database log type.

💰 Create Replication Wiz	ard			×
			Syniti Re	plicate
Replication type Source connection <i>Source log info</i> Target connection		ent transaction read point from ID from which to replicate.	n the IBM i/AS400 server. To ov	verride, click Read
Target connection Target log info Mapping info Scheduling Actions Summary	Journal: Receiver: Transaction ID: Transaction Timestamp: Read Interval (sec):	OLETEST.QSQJRN OLETEST.QSQJRN3424 5550 6/28/2022 9:42:41 PM 60		Read TID
		< Back	Next > Canc	el Help

12. Click Next to go to the Select Target Connection screen.

Copyright© 2023 by BackOffice Associates, LLC d/b/a Syniti and/or affiliates. All Rights Reserved. This document contains confidential and proprietary information and reproduction is prohibited unless authorized by Syniti. Names appearing within the product manuals may be trademarks of their respective owners.

🗴 Create Replication Wiz	zard				×
			Synit	i Repli	cate
Replication type Source connection	Select the target conner	ction and table to be replicated.	i -		
Source log info <i>Target</i> <i>connection</i> Target log info	Connection Name:	Hadoop-Tgt		~	
Mapping info Scheduling Actions Summary	Table Name:			∽ Open Table	F
		< Back	Next >	Cancel	Help

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

💧 Create Replication Wiz	zard	×
		Syniti Replicate
Replication type Source connection Source log info	The mapping has been automatically generated using target fields. Drag source fields to target fields or use	
Target connection	Field name Ordinal	Field name Ordinal Ty
Target log info <i>Mapping info</i> Scheduling Actions Summary	CUSTOMERID 01 TERRITORYID 02 ACCOUNTNU 03 CUSTOMERT 04 ROWGUID 05 MODIFIEDDATE 06	CUSTOMERID 01 ini TERRITORYID 02 ini ACCOUNTNU 03 va CUSTOMERT 04 nc ROWGUID 05 bi MODIFIEDDATE 06 da
Constant of the second	Field name Target Table Ordinal CUSTOMERID [SQLServer-Tgt] 1 TERRITORYID [SQLServer-Tgt] 2 ACCOUNTNU [SQLServer-Tgt] 3	TypeSize§int4(int41varchar104
	< Back	Next > Cancel Help

16. Click Next to go to the Scheduling screen.

💰 Create Replication Wiz	ard X
	Syniti Replicate
Replication type Source connection	Set scheduling information for the replication.
Source log info Target connection	 ✓ Enable Replication ✓ Execute Initial Refresh Start Time: 6/29/2022 ∨ 2:10:45 PM ÷
Target log info Mapping info <i>Scheduling</i> Actions Summary	Refresh Schedule Mirroring Schedule Verifier Schedule Image: Run One Time Only Run Recurrently: Image: Run Recurrently:
States and	Add Remove Edit Sack Next > Cancel Help

- 17. Make sure the **Enable Replication** option is checked. This is required for the replication to run.
- 18. Set a start time for the replication. The **Start Time** field indicates the time at which the Data Replicator will begin considering the replication for execution.
- 19. Check the option to **Execute Initial Refresh**. A full replication will be performed from the source table to the data file.
- 20. Click **Next** to go to the **Summary** screen.
- 21. Click Finish to complete the wizard.

Start Replications

If you installed the Replication Agent as a service during setup, you just need to start the service using the

ServiceMonitor 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 set up the Replication Agent as a service:

- From the ServiceMonitor program ⁵⁵ in the Windows Notification Area, choose Launch Service Installer.
- Manage the service from the Service Monitor program (located in the Windows Notification Area 🛸).
- 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.

Stop Replications

Stop the Replication Agent from the Service Monitor in the Windows Notification Area.