Syniti

Data Replication

Setup Guide for Replicating to Hadoop

Version 9.8.2



Table of Contents

Setup Guide for Replicating to Hadoop	1
Introduction	1
Basic Configuration Steps	2
Connection Type	2
Setup Summary	2
Download and Install Syniti Data Replication	3
Steps for Replicating Tables	3
1. Set Up a Source Connection to a Relational Database	3
2. Enable Transactional Replication	3
3. Set up a Target Connection to Hadoop HDFS ϵ	3
4. Add Table Information to the Target Connection)
5. Define Replications)
Start Replications	ł
Stop Replications14	ł

Introduction

Syniti Data Replication 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



Specialized Libraries

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 DR transfers data as CSV files then uses the Syniti DR File Broker service to communicate with HDFS.

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

Copyright© 2022 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 DR 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 DR. 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 DR	 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 DR Setup	 In the Syniti DR Management Center: In the Metadata Explorer, create a source connection to your RDBMS. Create a target connection using the provider "Hadoop HDFS".

(See <u>Steps for</u> <u>Replicating Tables</u> below)	3. Create a replication.
Start Replicating	In the Syniti DR Service Monitor:
(See <u>Start Replications</u>)	1. Start the Syniti Replication Agent.

Download and Install Syniti Data Replication

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.

- 1. Syniti Knowledge Base
- 2. Enter a generic support ticket.

Steps for Replicating Tables

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

Copyright© 2022 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.

		Syniti Data Replication
Select provider	Select the dat	abase that contains source data to be replicated and indicate which provider to use.
Set connection string Select tables	Source name	IBMDR21 Sre
Actions Summary	Data Provider(s)
	Database:	BM Db2 for i
	Provider:	Fight HiT Software .NET Driver (Ritmofi)
	Assembly:	Ritmo_Vilb/Sql400.dli
Statistics and the second		

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 Connectic	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</back>	p

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 DR help for more information on the best option.

💰 Enable Transactional	Replication Wizard	×
	Syniti Data Replication	
Log Type Log Settings	Choose the appropriate type of logging for this connection. For more information, see <u>Choosing a Log</u> <u>Type for Transactional Replications</u>	
Actions Summary	Log rype I 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.	
	 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. 	
Same	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.

O-lastid			
Select provider	Select the dat	abase target where to replicate data and indicate which provider to use.	
Set connection string	Target name		
	Name:	HDFSTgt	
Select tables	Data Provider(\$)	
Summary	Database:	S MySQL	l
	Provider:	Oracle PostgreSQL SAP Sybase ASE SolidDB Tibero TmaxSoft	
Contractor of		Analytics Or Data Warehouse Actian Vector Amazon Redshift Google BigQuery IBM PureData for Analytics (Netezza)	
		Microsoft Azure Data Warehouse SAP HANA SAP Sybase IQ SAP Sybase IQ	۶lp
		Teradata Vertica	
		Big Data	
		DocumentDB	
		Hadoop HDFS Microsoft Azure Data Lake	
		Microsoft Azure Data Lake Gen2	
		MongoDB Atlas	

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

Add larger connection	wizard			
	Sy	niti Data R	eplicat	ion
elect provider	Specify the connection parameters for	the target connection.		
et connection tring	Connection properties			
	✓ Required			^
	Output Folder			
	Hostname			
elect tables	Username			
tions	Password			
	Private Key File Path			
mmary	Path To Binary			
	Target Directory			
	Working Directory			
	Output Folder Archive			
	Use Column Names	True		~
	Output Folder Path for the schema and output files.			
State Balanta			👱 Edit 📑	Test

Output Folder	An existing folder on the system that is running Syniti Data Replication for files associated with replications to Hadoop.
Hostname	The 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 DR.
Output Folder Archive	Optional. Provides a local copy of data replicated to Hadoop. Data is not managed by Syniti DR, so the files must be managed manually and could grow quickly.

- 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.
- 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.



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.

		Syniti Dat	ta Replicati	on
Replication type Source connection	Mirroring mode will defin server. Reolication Name	ne a one-way transactional replication	, from the source server to the ta	rget
Source log info Target connection	Replication Name Description:	leave blank to use source table not source table not	names>	
Target log info Mapping info	Use Group:		✓ <u>C</u> reat	te
Scheduling Actions Summary	Replication Type C Refresh Continuous Mirro Synchronization	ring		

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

		Syniti	Data R	eplicat	ion
Replication type	Select the source conn	ection and table to be replica	ted.		
connection					
larget connection	Connection Name:	SQLServer			\checkmark
Mapping info					
Scheduling	Table Name:	DEMODB.dbo.PRODUC	TS		~ 5
lotions Summary				<u>O</u> pen Table.	
and the second second					

- 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....

 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.

Replication type Source connection Source log info Target connection Target log info Publication: Distributor: DBMotoPUB_DEMODB Transaction ID: 000000850000018b0008 Transaction Timestamp: 9/17/2020 6:20:39 PM Read Interval (sec): 60			Syniti D	ata Rep	olication
Target connection Distributor: DBRS_distribution Read TD Target log info Publication: DBMotoPUB_DEMODB Mapping info Transaction ID: 000000850000018b0008 Scheduling Transaction Timestamp: 9/17/2020 6:20:39 PM Actions Read Interval (sec): 60	Replication type Source connection Source log info	Click Next to use the curr TID to set the transaction	ent transaction read point from ID from which to replicate.	the SQL Server serve	r. To override, click Read
	Target log info Mapping info Scheduling Actions Summary	Distributor: Publication: Transaction ID: Transaction Timestamp: Read Interval (sec):	DBRS_distribution DBMotoPUB_DEMODB 000000850000018b0008 9/17/2020 6:20:39 PM 60]	Read TID

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

💰 Create Replication Wiza	rd					_
		Sy	/niti D	ata Re	eplicat	ion
Replication type Source connection	Select the target conne	ction and table	to be replicated.			
Source log info Target connection	Connection Name:	🛢 Hadoop-1	ſgt			~
Farget log info Mapping info	Table Name:	l			Open Table	~ 6 50
Scheduling						
Summary						
C. C						
			< 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 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.

		Sy	/niti E	Data I	Replica	atio	n
Replication type	The mapping has I target fields. Drag	been automatically ge source fields to targ	enerated using the let fields or use t	ne default map; the toolbar to e	bing rule to match so dit mappings.	urce and	1
source connection		- PH -					
Source log info	ISOL Server1 DEMOD	B dbo PRO		1	Hadoop-Totl PRODU	CTS	
Target connection	Field name	Ordinal A			Field name	Ordinal	Т. /
-		01				01	de
	PRODUCTNA	02				02	st
	SUPPLIERID	03			SUPPLIERID	03	de
Mapping Into	CATEGORYID	04		•	CATEGORYID	04	de
Scheduling	QUANTITYPE	05		•	QUANTITYPE	05	st
Actions		06 -		•		06	de
Summary	<	>			<		>
	Field name	Target Table	Ordinal	Туре	Size		5
	PRODUCTID	[Hadoop-Tgt] PR	1	decimal	-1		F
	PRODUCTNA	[Hadoop-Tgt] PR	2	string	-1		F
San Start	SUPPLIERID	[Hadoop-Tgt] PR	3	decimal	-1		5
							-

16. Click Next to go to the Scheduling screen.

	Syniti Data Replication
Replication type	Set scheduling information for the replication.
ource log info	Enable Replication Execute Initial Refresh Start Time: 9/18/2020 3:10:07 PM
arget log info Mapping info Scheduling Iuctions Summany	Refresh Schedule Mirroring Schedule Verifier Schedule Run One Time Only Run Recurrently:
C. C	Add Remove Edit

- 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 🔤 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.