Getting Started with Qlik Replicate on Amazon EC2
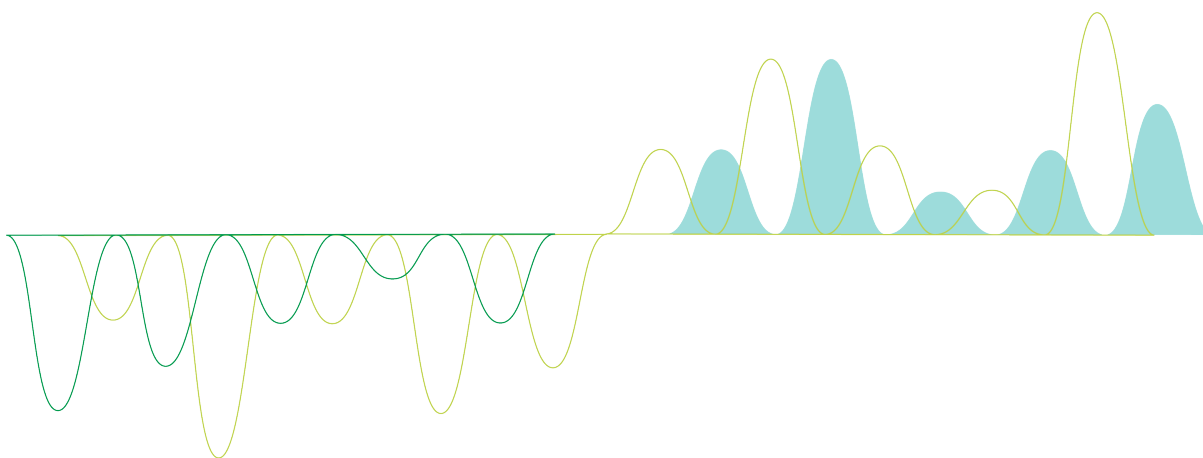
Qlik Replicate™

April 2020 (Version 6.6)

Last Updated: October 13, 2021

LEAD WITH DATA   Qlik Q

Contents

1   Overview

Qlik Replicate on Amazon EC2 facilitates data replication across a wide range of heterogeneous sources and targets (listed below). In this tutorial, you will learn how to set up a basic replication task that replicates data from an on-premises Oracle database to an Amazon Redshift database in the cloud.

For more detailed information about topics covered in this tutorial or if you simply wish to learn more about Qlik Replicate's capabilities, click the **Help** link in the top tight of the Qlik Replicate console.

1.1   Supported Source Databases

Qlik Replicate on Amazon EC2 can replicate data from the following source databases:

- Oracle
- Microsoft SQL Server
- MySQL (including Amazon Aurora MySQL, Maria DB, and Percona)
- PostgreSQL
- IBM DB2 for LUW
- Amazon RDS for MySQL
- Amazon RDS for Oracle
- Amazon RDS for SQL Server

1.2   Supported Target Databases

Qlik Replicate on Amazon EC2 can replicate data to the following target databases:

- Oracle (including Amazon RDS for Oracle)
- Microsoft SQL Server (including Amazon RDS for SQL Server)
- MySQL (including Amazon RDS for MySQL, Amazon RDS for MariaDB, and Amazon Aurora MySQL)
- PostgreSQL (including Amazon Aurora PostgreSQL and Amazon RDS for PostgreSQL)
- Teradata
- Amazon Redshift
- Hadoop
- Snowflake on Azure
- Snowflake on AWS
- Amazon RDS for MySQL
- Amazon RDS for Oracle
- Amazon RDS for SQL Server

1.3   Preinstalled Third-Party Clients

The following third-party clients are preinstalled on the Qlik Replicate Amazon EC2 instance:

- Oracle
- Microsoft SQL Server
- Amazon Redshift
- IBM DB2 for LUW
- Snowflake
- Hadoop

As these clients are already installed on the Qlik Replicate Amazon EC2 Instance, you can replicate both to and from the corresponding database servers without needing to install any client-side software. All other supported sources and targets require you to install the appropriate client-side software before replication can be performed.

2   What You Need

Before starting this tutorial, make sure that the following prerequisites have been met:

- *Oracle Source Prerequisites (page 6)*
- *Amazon Redshift Target Prerequisites (page 6)*

2.1   Oracle Source Prerequisites

Before you can use Oracle as a source endpoint in this tutorial, make sure the following prerequisites have been met:

- Oracle 10.x, 11.x, 12.1, 12.2.0.1, or 18.3 installed in your network and accessible from the Qlik Replicate machine
- Port 1521 open for outbound communication on the Qlik Replicate Amazon EC2 instance.
- `system/<password>` for an admin user

2.2   Amazon Redshift Target Prerequisites

Before you can use Amazon Redshift as a target endpoint in this tutorial, make sure the following prerequisites have been met:

- **Amazon Redshift Cluster:** If you haven't already done so, set up an Amazon Redshift cluster and make sure that following information about your Amazon Redshift Cluster is readily available:
    - Amazon Redshift Cluster Name
    - Amazon Redshift Cluster Port
    - Amazon Redshift User Name and Password
    - Amazon Redshift Database Name

- **Port:** Make sure that port 5439 (the Amazon Redshift Cluster port) is open for outbound connections from your Qlik Replicate Amazon EC2 instance.
- **Amazon S3 Bucket Details:** During the first stage of the replication process, Qlik Replicate moves the data files created by the source database into an Amazon S3 bucket. The files are then loaded into the proper tables in the Amazon Redshift data warehouse (using the "copy" command). Therefore, in addition to setting up an Amazon Redshift Cluster, you also need to set up and Amazon S3 Bucket.
  Make sure that the following information about your Amazon S3 Bucket is readily available:
    - Amazon S3 Bucket Name
    - Amazon S3 Bucket Region
    - Amazon S3 Access Key and Secret Key

- **Amazon S3 Bucket Permissions:** The Qlik Replicate user must have the following permission on the Amazon S3 staging bucket:
    - On the Amazon S3 bucket itself: 'List Bucket Objects'
    - On the Amazon S3 bucket folder: 'Read, Write, Delete of Objects'

The following is a sample policy with these permissions:

```
{

"Statement": [

{

"Effect": "Allow",

"Action": [

"s3:ListAllMyBuckets"

],

"Resource": "arn:aws:s3:::*"

},

{

"Effect": "Allow",

"Action": [

"s3:ListBucket",

"s3:GetBucketLocation"

],

"Resource": "arn:aws:s3:::demo-virginia"

},

{

"Effect": "Allow",

"Action": [

"s3:PutObject",

"s3:GetObject",

"s3:DeleteObject"

],

"Resource": "arn:aws:s3:::demo-virginia/*:

}

]

}
```

3   Set up a Replication Task

After fulfilling the prerequisites, the next stage is to set up your replication task.

This consists of the following steps:

- *Step 1: Open the Qlik Replicate Console (page 9)*
- *Step 2: Add a Replication Task to the Qlik Replicate Console (page 10)*
- *Step 3: Define your Source and Target Endpoint Connections (page 12)*
- *Step 4: Add the Source and Target Endpoint Connections to the Task (page 14)*
- *Step 5: Select Source Tables for Replication (page 16)*

3.1   Step 1: Open the Qlik Replicate Console

From the Windows Start menu, select **All Programs > Qlik Replicate > Qlik Replicate Console**.

You can access Qlik Replicate from any computer in your network.

To access the Console from a remote computer, type the following address in the address bar of your Web browser:

```
https://<computer name>/attunityreplicate
```

where `<computer name>` is the name or IP address of the computer (including the Windows domain name) on which the Qlik Replicate for Amazon EC2 instance is installed.

3.2   Step 2: Add a Replication Task to the Qlik Replicate Console

After opening the console, the next step is to add a replication task.

Note that in actual production environments, some customers prefer to [set up their endpoint connections](#) *before* adding tasks. The order that you choose will depend on your specific replication needs. For instance, setting up the endpoint connections first may be preferable when setting up several tasks that use different endpoints.

To add a replication task:

1. Make sure that you are in **TASKS** view. The word "TASKS" should be displayed in the upper left corner of the Qlik Replicate Console.
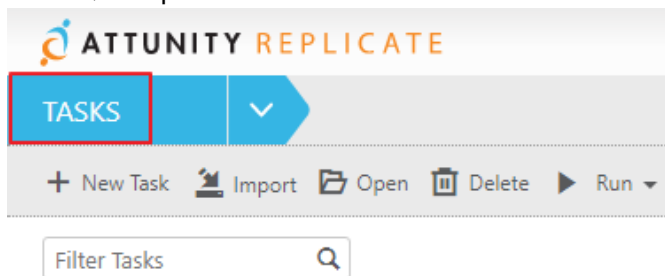
2. Click **New Task** to open the **New Task** dialog box.
3. In the **Name** field, enter `oracletoredshift` and then click **OK**.
4. Make sure that the **Apply Changes** and **Store Changes** replication options are disabled. When these options are enabled, Replicate captures changes to the source tables in real time and applies them to the target tables (**Apply Changes**), and/or stores them in special tables on the target (**Store Changes**). As the **Apply Changes** and **Store Changes** options requires additional source database preparation/permissions, for the sake of simplicity, the scope of this tutorial will be limited to setting up a **Full Load** only task. For information on setting up a task with Apply Changes, please refer to the *Qlik Replicate Setup and User Guide*.
5. Click **OK** to add the task.

The task is displayed in a new tab.
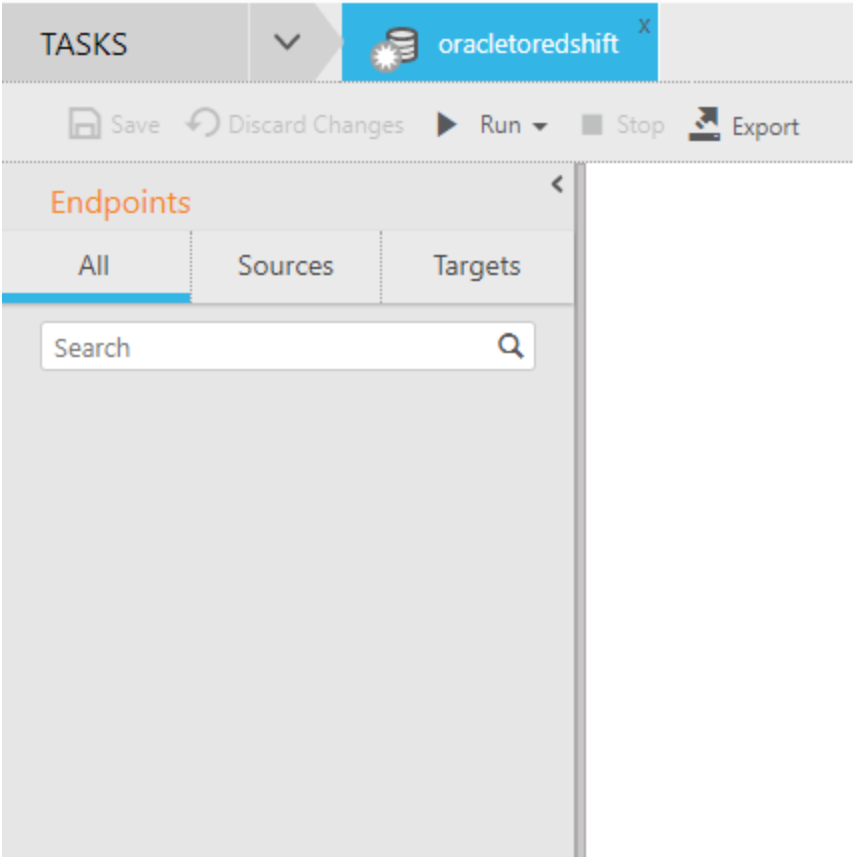
3.3   Step 3: Define your Source and Target Endpoint Connections

After adding a new task, the next step is to set up your source and target endpoint connections. An endpoint connection is basically a set of connection parameters that enables Replicate to extract data from the source database and replicate it to the target database. Connection parameters specify information such as the database server name, the database name, the user name, and so on.

In this tutorial, we will be using Oracle as our source and Amazon Redshift as our target.

Add an Oracle Source Endpoint Connection

This section guides you through process of defining your Oracle endpoint connection.

Before beginning this section, make sure that Oracle prerequisites have been met.

To add an Oracle source endpoint connection:

1. In **TASKS** view, click the **Manage Endpoint Connections** toolbar button.
   The **Manage Endpoint Connections** dialog box opens.
2. Click the **New Endpoint Connection** toolbar button.
3. Provide the following information:
   - **Name**: Enter any name that will help you identify the endpoint.
   - **Description**: Optionally, enter a description.
   - **Role**: Select **Source**.
   - **Type**: Select **Oracle**.
   - **Connection string**: Enter the connection string to your Oracle database, in any Oracle format. For example, to connect to an Oracle database on a computer called **tomato** with the default Oracle port and service name, enter the following:
     ```
     tomato:1521/orcl
     ```
   - **User Name**: Enter the user name for your Oracle database.
   - **Password**: Enter the password for your Oracle database.
4. Click **Test Connection** to verify that the information you entered is correct.
5. Click **Save** to add the endpoint.

Add an Amazon Redshift Target Endpoint Connection

This section guides you through process of defining your Amazon Redshift endpoint connection.

Before beginning this section, make sure that Amazon Redshift Target Prerequisites have been met.

To add an Amazon Redshift target endpoint connection:

1. In **TASKS** view, click the **Manage Endpoint Connections** toolbar button.
   The **Manage Endpoint Connections** dialog box opens.
2. Click the **New Endpoint Connection** toolbar button.
3. Provide the following general information:
   - **Name**: Enter any name that will help you identify the endpoint.
   - **Description**: Optionally, enter a description.
   - **Role**: Select **Target**.
   - **Type:** Select **Amazon Redshift**.

4. Provide the following **Amazon Redshift target** information:
   - **Redshift cluster:** Your Amazon Redshift cluster name
   - **User name:** Your Amazon Redshift user name
   - **Password:** Your Amazon Redshift password
   - **Database name**: The target Amazon Redshift database name
5. Provide the following **Amazon S3 staging** information. You may need to click the **Amazon S3 staging** header to see the information:
   - **Bucket name:** The name of the Amazon S3 bucket where you are copying files to.
   - **Bucket region:** The Amazon S3 region where the S3 buckets and folders you are using are hosted.
   - **Access key:** The access key information for your Amazon S3 bucket
   - **Secret key:** The secret key information for your Amazon S3 bucket
   - **Folder:** The S3 folder to which the files will be copied
6. Click **Test Connection** to verify that the information you entered is correct.
   If the connection is successful a message in green is displayed. If the connection fails, an error message is displayed at the bottom of the dialog box.
   To view the log entry if the connection fails, click **View Log**. The server log is displayed with the information for the connection failure. Note that this button is not available unless the test connection fails.
7. Click **Save** to add the endpoint.

3.4   Step 4: Add the Source and Target Endpoint Connections to the Task

After defining your source and target endpoint connections, the next step is to add them to the replication task.

The **Endpoints** pane is located in the left of the console. Replicate supports several endpoint types (database, cloud, Hadoop, data warehouse, file, NoSQL) with each endpoint type having its own icon. As the source endpoint type is a database and the target endpoint type is cloud-based, the following icons will be displayed:
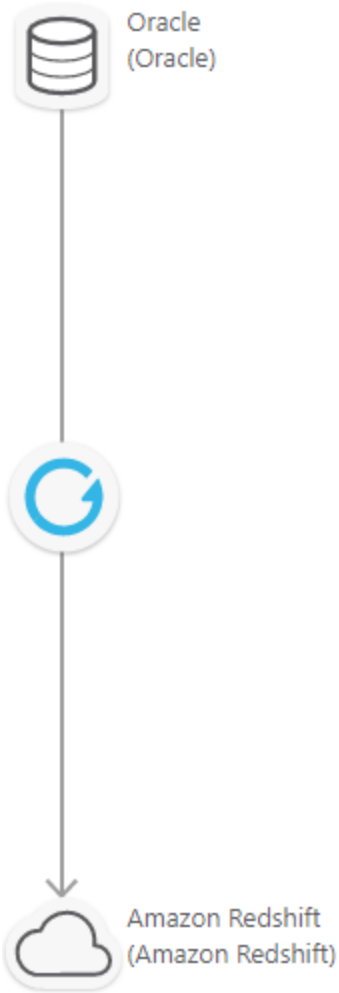
The source endpoint.

The target endpoint.

To add the source and target endpoints to the task

1. In the **Endpoints** pane, select the **All** tab.
2. Add the Oracle source endpoint to the task using any of the following methods:
   - Drag the endpoint to the **Drop source endpoint here** area in the endpoints diagram to the right of the **Endpoints** pane.
     -OR-
   - Hover your mouse cursor over the endpoint name and click the arrow that appears to its right.
3. Add the Amazon Redshift target endpoint to the task using any of the following methods:
   - Drag the endpoint to the **Drop target endpoint here** area in the endpoints diagram to the right of the **Endpoints** pane.
     -OR-
   - Hover your mouse cursor over the endpoint name and click the arrow that appears to its right.

The endpoint connections diagram should now appear similar to this:

Oracle
(Oracle)

Amazon Redshift
(Amazon Redshift)

3.5 Step 5: Select Source Tables for Replication

After adding the source and target endpoints, you now need to select the Oracle source tables that you want to replicate to Amazon Redshift.

To add Oracle tables to the replication task:

1. In the right pane of the Qlik Replicate Console, click the **Table Selection** button.
   The **Select Tables** dialog box opens.
2. In the **Select Tables** dialog box, do the following:
    1. From the **Schema** list, select the schema containing your source tables, and then click **Search**.
    2. From the **Table List**, either:
       Hold down the [Ctrl] button and select the tables you want to replicate. Then, click the [>] button to move the tables to the **Selected Tables** list.
       -OR-
       Click the [> >] button to move all the tables to the **Selected Tables** list.
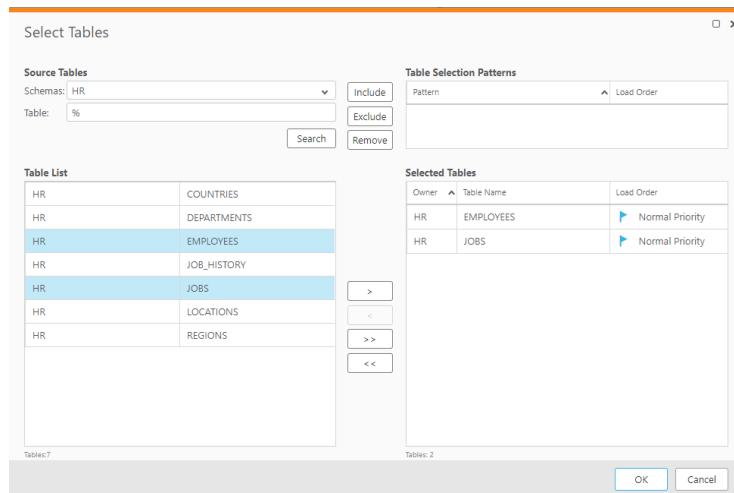
    3. Click **OK**.
3. On the main toolbar, click **Save**. The task is now ready to be run.

4   Run and Monitor the Replication Task

In this part of the tutorial, you'll learn how to run a replication task and monitor its progress.

To run and monitor the replication task

1. Click the **Run** button in the main toolbar.
    The **Starting task** dialog is displayed and the console automatically switches to Monitor view:
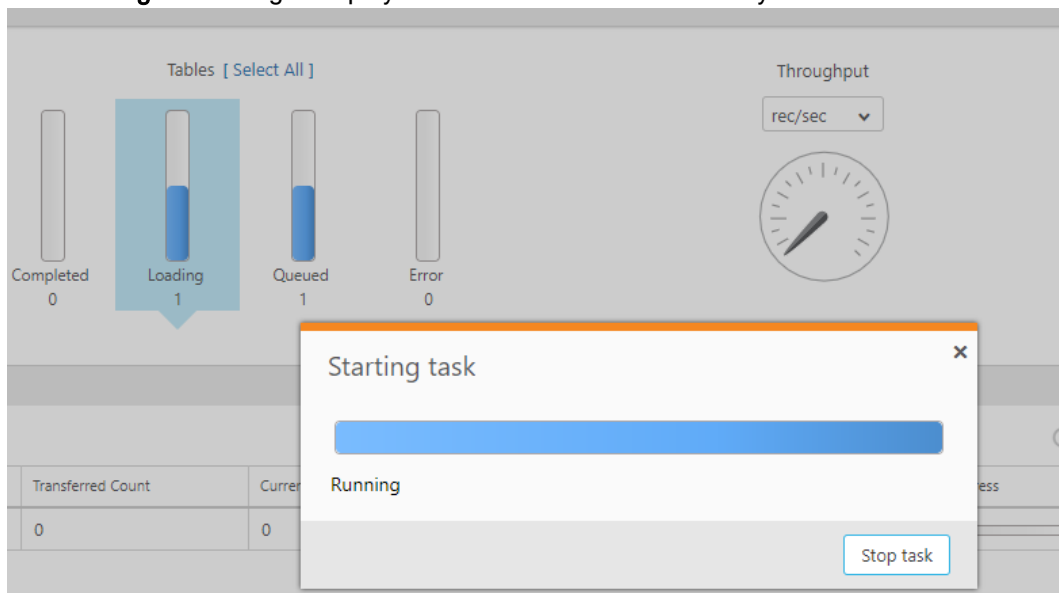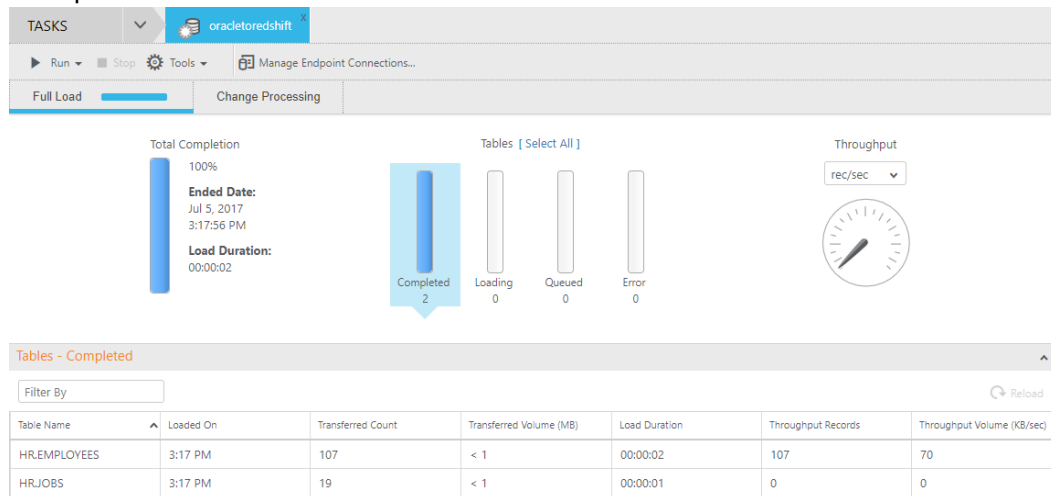

    Monitor view contains the following tabs:

    - **Full Load** tab: Uses status bars to provide a visual indication of the loading status during the full load operation.

    - **Change Processing** tab: Shows any changes that occur after the Full Load completes. Since this a Full Load only task, this tab is not relevant.

2. Click the **Completed** status bar. Replicate displays a table below the graphs with statistics related to the replicated tables.

Table Name	Loaded On	Transferred Count	Transferred Volume (MB)	Load Duration	Throughput Records	Throughput Volume (KB/sec)
HR.EMPLOYEES	3:17 PM	107	< 1	00:00:02	107	70
HR.JOBS	3:17 PM	19	< 1	00:00:01	0	0

3. Click the **Total Completion** bar to see a summary of the task.

5   What Next?

Now that you've successfully completed the tutorial, you'll probably want to waste no time applying what you've learned to your own environment.

For detailed setup and usage instructions, simply click the **Help** link in the top tight of the Qlik Replicate console.