

SAP Analytics on BigQuery with Qlik Replicate

How to extract SAP data into BigQuery

December 9, 2020

Table of contents

Abstract	5
Introduction	5
Architecture	6
Before you begin	7
Set up Google Cloud with a project and permissions	7
Set up SAP by installing Qlik Transports	9
Set up Qlik with client libraries, drivers, and SDK	9
Replicate data	10
Create your target endpoint	11
Create your database source endpoint	13
Create your SAP application source endpoint	14
Create your task	15
Work within quotas	19
View your replicated data	22

Abstract

- Performing analytics on SAP data directly in the transactional datastore is painful, and might impact SAP itself.
- BigQuery is part of an ecosystem of capabilities that enables Business Intelligence and Machine Learning.
- Qlik Replicate not only copies selected data from SAP into BigQuery, but keeps BigQuery synchronized as new transactions occur.

Introduction

You've realized that seeking business insights or performing machine learning on SAP enterprise resource planning (ERP) central component (ECC) data in-situ conflicts with the use of SAP as part of your core business, and requires SAP-proprietary tooling.

You're already using various services from Google Cloud, and want to leverage the data management capabilities of BigQuery for Business Intelligence analytics on your SAP data.

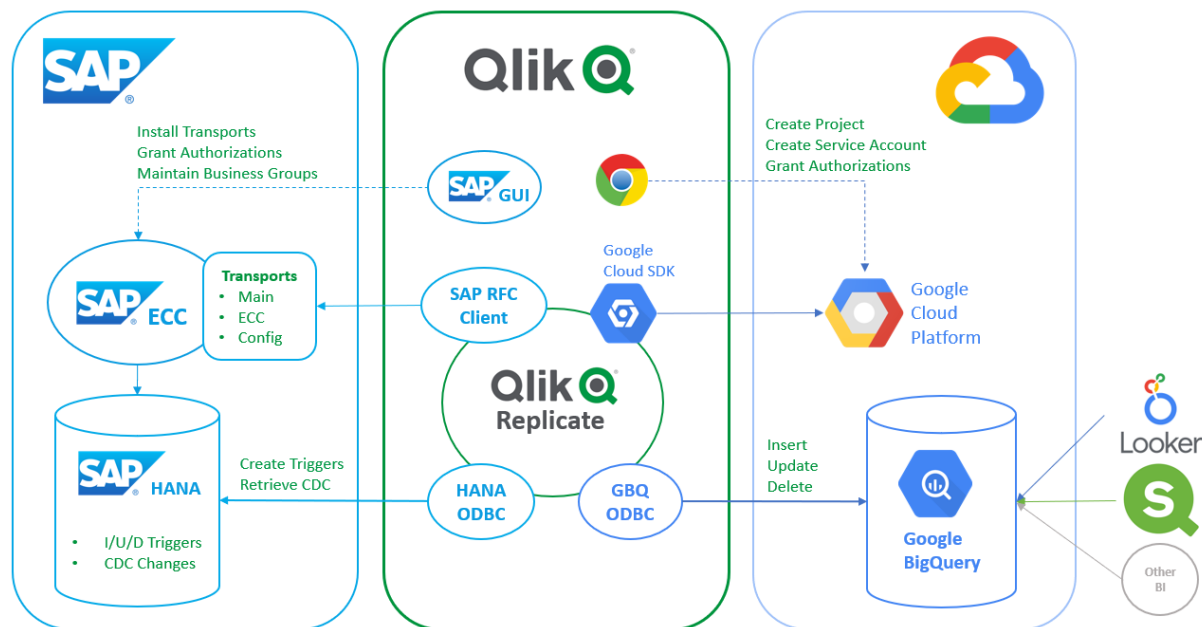
Qlik Replicate supports several change data capture mechanisms and many source and target systems. It is easy to deploy and use, and is far more powerful than its simple user interface indicates.

This whitepaper describes how to copy data from SAP—where it might sit in SAP HANA, Oracle[®], SQL Server, or IBM DB2—into BigQuery, and then keep BigQuery synchronized with SAP by copying transactions as they occur. The example in the paper assumes you use SAP HANA and have Qlik Replicate deployed on Windows.

This whitepaper describes the replication of data from SAP ERP/ECC 6.0 into BigQuery using Qlik Replicate v6.6. The paper is current as of May 2020. All SAP Modules are supported except for HR, including CRM, SRM, GTS, and MDG SAP Applications. This whitepaper does not apply to additional SAP products and services such as SAP Business ONE or SAP SuccessFactors.

Architecture

As shown in the following architecture diagram, Qlik Replicate connects to the database underlying SAP and creates triggers, which are used to capture data manipulation language (DML) actions (such as insert, update, and delete) to a log table as those transactions occur.



The initial **Full Load** task from SAP into BigQuery copies the selected tables. Subsequent **Apply Changes** tasks replicate DML transactions into BigQuery as they occur.

Qlik Replicate mostly copies data directly from the storage tier, but some SAP objects, such as cluster tables, are meaningful only at the application tier. Qlik Replicate retrieves these tables through a remote function call (RFC) connection to SAP. SAP maintains a mapping table that groups SAP tables into meaningful business groups (/nzs4sap), which Qlik Replicate queries through RFC. Qlik Replicate refers to this table when configuring a new data replication task.

Before you begin

This section describes the software and settings you need to perform the example in this whitepaper, and notes alternatives you might need if you are using an underlying database other than SAP HANA.

Set up Google Cloud with a project and permissions

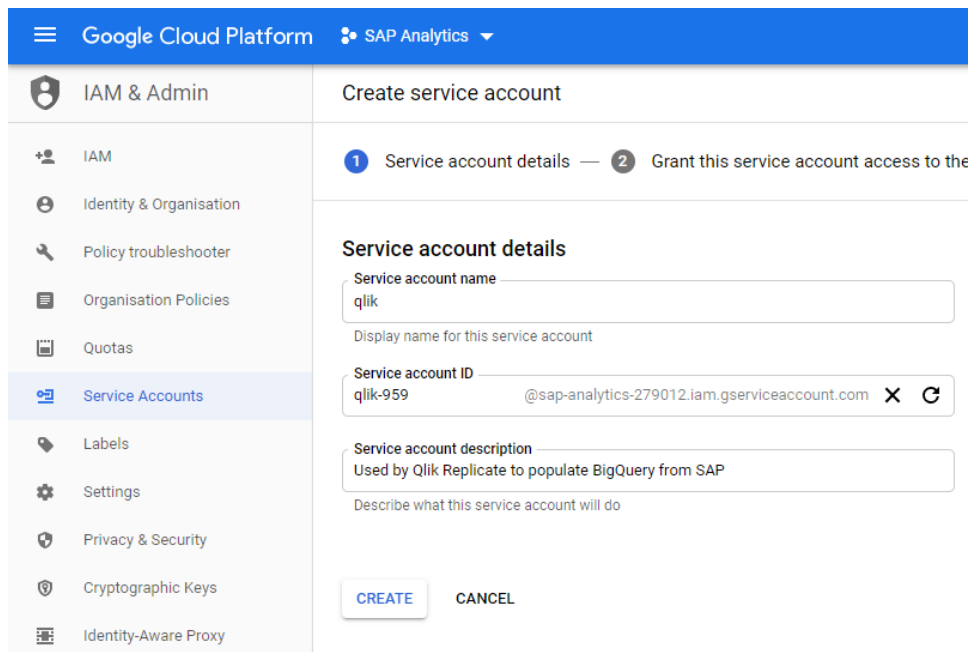
You need a Google Cloud project to be the target of your Qlik Replicate tasks, and a service account that you can grant the following roles:

- Project > Owner or Editor
- BigQuery > BigQuery Data Owner

You need a Service Account Key from the Google Cloud Console that Qlik Replicate can use for authentication into BigQuery.

For more information, see [Using BigQuery as a Target](#) in the Qlik Replicate documentation.

1. Open the [Google Cloud Console](#).
2. Create a project called **SAP Analytics**.
3. Go to **IAM & Admin**, click **Service Accounts**, and then create a service account named **qlik**.



Google Cloud Platform SAP Analytics

IAM & Admin Create service account

1 Service account details — 2 Grant this service account access to the

Service account details

Service account name
qlik

Display name for this service account

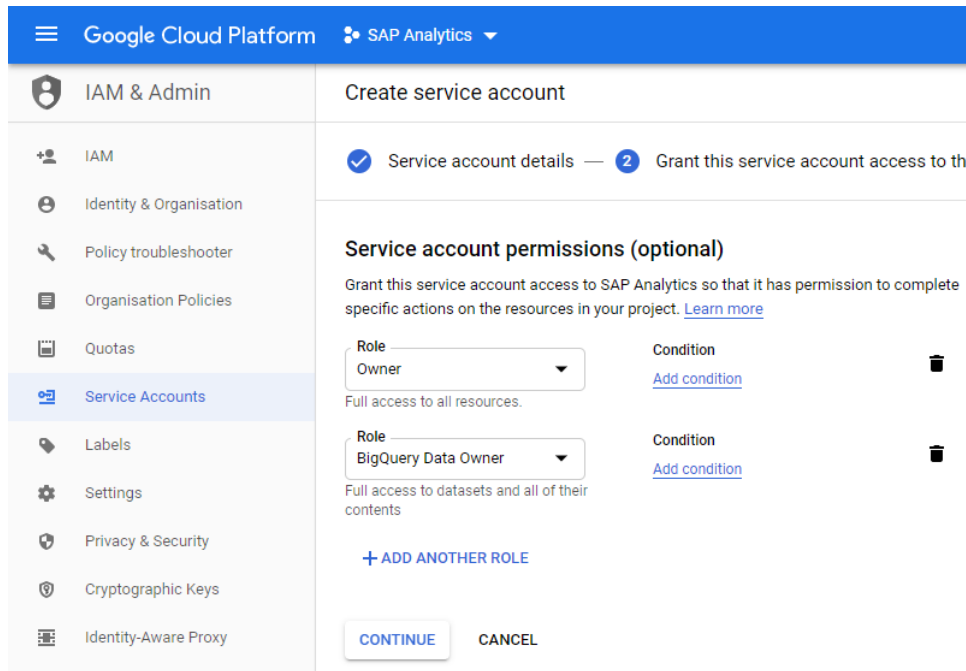
Service account ID
qlik-959 @sap-analytics-279012.iam.gserviceaccount.com

Service account description
Used by Qlik Replicate to populate BigQuery from SAP

Describe what this service account will do

CREATE CANCEL

- Grant the **qlik** service account the Project > Owner and BiqQuery > BigQuery Data Owner roles.



The screenshot shows the Google Cloud Platform IAM & Admin console. The main heading is 'Create service account'. Below this, there are two steps: '1. Service account details' (checked) and '2. Grant this service account access to th...'. Under the heading 'Service account permissions (optional)', there is a note: 'Grant this service account access to SAP Analytics so that it has permission to complete specific actions on the resources in your project. [Learn more](#)'. There are two role assignments listed:

Role	Condition	Action
Owner	Add condition	🗑️
BigQuery Data Owner	Add condition	🗑️

Below the roles, there is a '+ ADD ANOTHER ROLE' button and 'CONTINUE' and 'CANCEL' buttons at the bottom.

- Create a security key in JSON format, and save it for use later, when you connect from Qlik Replicate to BigQuery.

Set up SAP by installing Qlik Transports

Qlik Replicate supports connectivity primarily to SAP ERP/ECC 6.0 + all EhP levels. All modules are supported except HR, including support for CRM, SRM, GTS, and MDG SAP Applications.

Connectivity from Qlik Replicate to SAP primarily accesses the underlying database directly, but also utilizes RFC connections to Qlik Transports. You must install the Qlik Transports into SAP. These consist of the following:

- `InstallCode.zip`, the main transport file
- `InstallCodeECC.zip` (ECC-specific, where applicable)
- `InstallConfig.zip`, the initial configuration supplied as a transport

You need to enable additional authorizations for the Qlik Replicate software. Qlik Replicate for SAP delivers its own authorization object, ZR4SAP, and you must ensure that a dialog user and a communication user exist in SAP. The dialog user is required to access the Qlik Replicate for SAP Client GUI in SAP, and the communication user is required to support the RFC calls from

Qlik Replicate to the SAP system. For more information, see [Install the Qlik Replicate for SAP client on the SAP machine in the Qlik Replicate](#).

You'll also need the connectivity details for the database underlying your SAP instance, such as what database it is (HANA, Oracle[®], SQL Server, or IBM DB2), your credentials for accessing it, and verification that it is configured to support change data capture (CDC). In some cases, CDC is performed using database triggers, which must be installed into the database for the tables to be replicated. If Qlik Replicate has sufficient rights, it creates those triggers in the database itself for the tables to be replicated.

This example assumes you're using SAP HANA.

Set up Qlik with client libraries, drivers, and SDK

This whitepaper assumes you have installed a licensed copy of Qlik Replicate on a Windows host, ideally near your source data to avoid firewall considerations and to minimize latency.

- To connect from Qlik Replicate to SAP, you must extract the **SAP Netweaver RFC Client** libraries (SDK v7.20 or v7.50) from the NWRFC_XXX.SAR file and copy them from `nwrfc_sdk/lib` into the Qlik Replicate bin directory.
- To connect from Qlik Replicate to BigQuery, you must download and install the **Simba ODBC** driver v2.2.3.1000 and the **Google Cloud SDK** 253.0.0 onto the Qlik Replicate machine.
- You must rename certain irregularly named SAP schema/column objects in the Qlik Replicate task before you write them to BigQuery.

The different SAP source endpoints in Qlik Replicate require an already-configured source endpoint that connects directly to the HANA, Oracle, SQL Server, or IBM DB2 database underlying SAP. Each of these endpoints has its own requirements, and might require you to enable CDC or implement it manually. At a minimum, you must have the corresponding database driver installed on the Qlik Replicate machine. Certain rights are required in the database.

Required drivers for each database:

- HANA – SAP HANA ODBC (x64) Driver 2.x HDBODBC
- Oracle – Oracle Instant Client (x64) v11.2.0.3.0 (or greater). [Download and install](#).
- SQLServer – Microsoft ODBC Driver 17.4.2 (or greater). [Download and install](#).
- IBM DB2 – Your requirements vary depending on what platform DB2 is hosted on, but at a minimum you require the IBM Data Server Client 11.1.

For information about the connectivity requirements of your SAP database, see the [Qlik Replicate](#) documentation.

Replicate your data

The Qlik Replicate fundamental objects are *source* and *target endpoints*, and *tasks*.

Tasks include a one-time **Full Load** task, which copies data from source to target and then completes, and the **Apply Changes** task, which continues to run, looking for transactions in the selected source tables to propagate into the target.

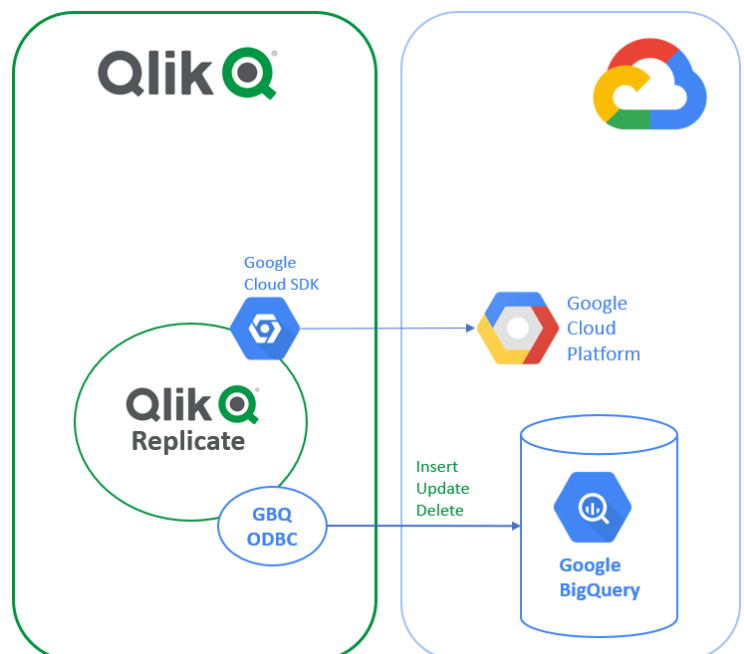
The SAP source in Qlik Replicate contains *business groups*, which are collections of SAP objects that are used within the SAP modules. These groups simplify how you select the data you need to copy into BigQuery to satisfy your analytic objectives. The list of table objects in each Business Group is maintained in SAP and accessed using the /nzs4sap transaction.

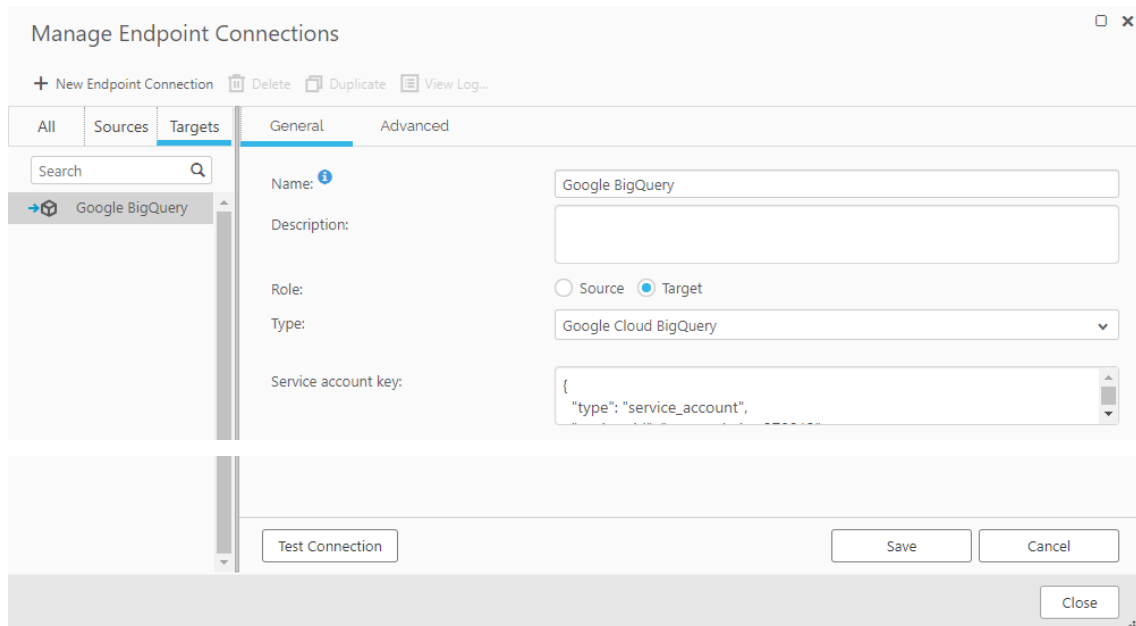
First, you create your source and target endpoints, and then you select tables from the source to be replicated into the target in a new task.

Start by browsing to Qlik Replicate (<https://yourServer.example.com/attunityreplicate>) and authenticating.

Create your target endpoint

1. In Qlik Replicate, go to **Manage EndPoint Connections** and click **New EndPoint Connection**.
2. Name the new endpoint **Google BigQuery** or something meaningful to you. Ensure that its role is **target**, and that the type is **Google Cloud BigQuery**.
3. Paste the contents of the security key that you created earlier into the **Service Account Key** field.

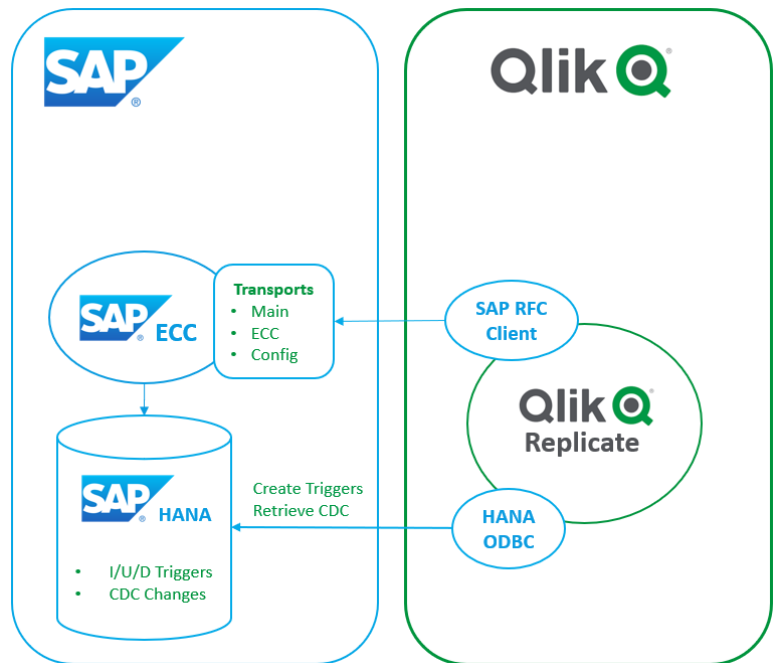




4. Click **Test Connection**, resolve any errors, and then save your new target endpoint.

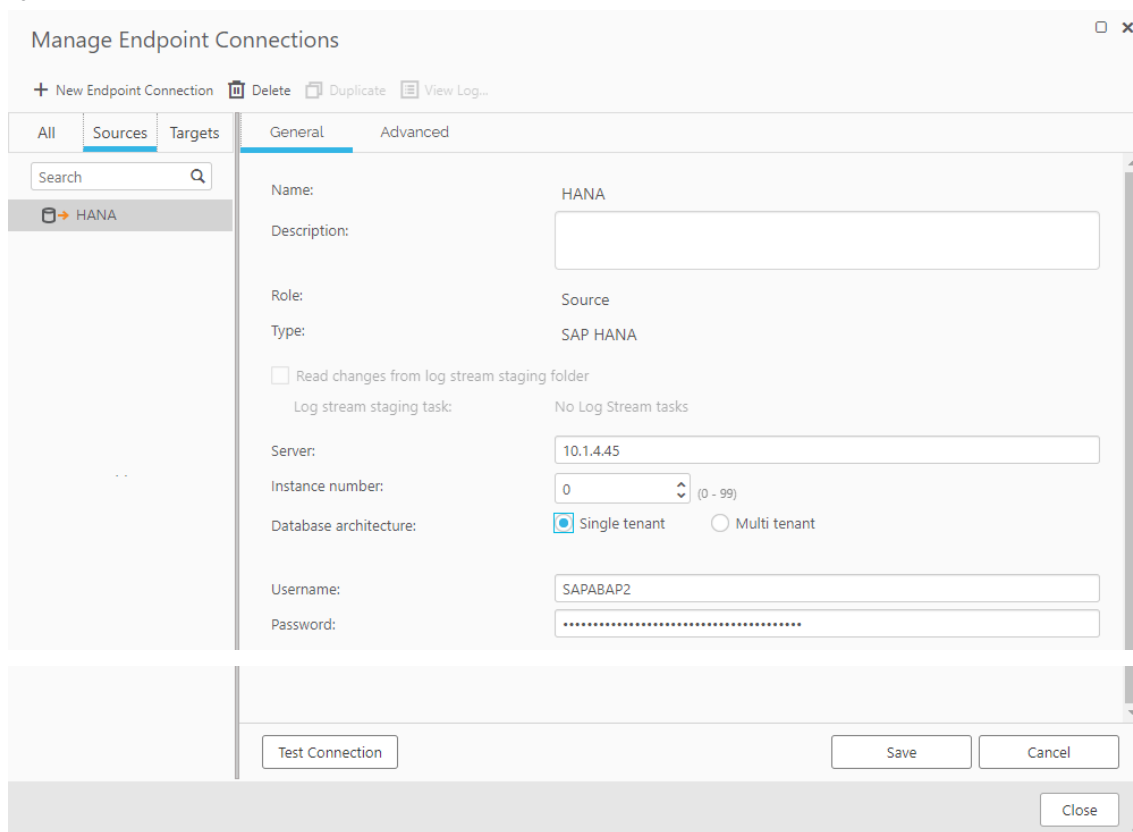
Next, you create a source endpoint for your SAP HANA database. Then you create one of two possible SAP application endpoints to leverage the database source endpoint:

- SAP Application, requiring:
 - Server name
 - Instance number (e.g. 00)
 - Client (e.g. 800)
 - Username & password
- SAP Application (DB), requiring:
 - Client (e.g. 800)
 - SAP Schema



Create your database source endpoint

1. Go to **Manage EndPoint Connections**, and click **New Endpoint Connection**.
2. Name the new endpoint **HANA** or something meaningful to you. Ensure that its role is **source**, and that the type is **SAP HANA**.
3. Specify the **server** where HANA is hosted and the **Instance number**.
4. Type the credentials used to access HANA.



The screenshot shows the 'Manage Endpoint Connections' dialog box with the 'Sources' tab selected. The configuration for a new endpoint named 'HANA' is as follows:

- Name:** HANA
- Description:** (empty text box)
- Role:** Source
- Type:** SAP HANA
- Read changes from log stream staging folder
- Log stream staging task:** No Log Stream tasks
- Server:** 10.1.4.45
- Instance number:** 0 (range 0 - 99)
- Database architecture:** Single tenant Multi tenant
- Username:** SAPABAP2
- Password:** (masked with dots)

Buttons at the bottom include 'Test Connection', 'Save', 'Cancel', and 'Close'.

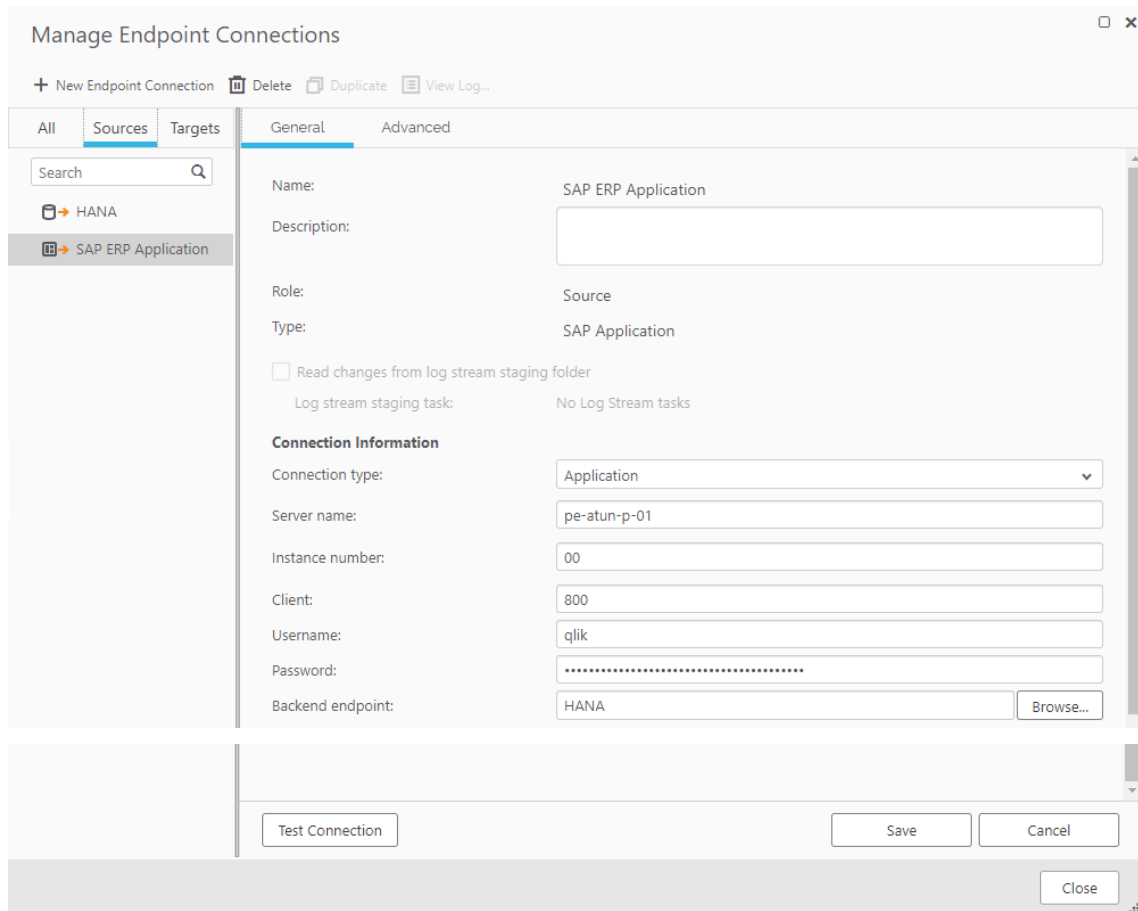
5. Click **Test Connection**, resolve any errors, and then save your new source endpoint.

Now that you've created the database source endpoint, you must create the SAP Application source endpoint to leverage it.

Create your SAP application source endpoint

1. Click **New Endpoint Connection**.

2. Name the new endpoint **SAP ERP Application** or something meaningful to you. Ensure that its role is **source**, and that the type is **SAP ERP Application**.
3. Provide the following information to enable the RFC connection from Qlik Replicate to SAP:
 - Connection type
 - Server name where SAP can be accessed
 - Instance number
 - Client
4. Record credentials within SAP that have been granted the required authorizations.
5. For **Backend endpoint**, select the **HANA** source endpoint that you created earlier.



Manage Endpoint Connections

+ New Endpoint Connection Delete Duplicate View Log...

All Sources Targets General Advanced

Search

HANA

SAP ERP Application

Name: SAP ERP Application

Description:

Role: Source

Type: SAP Application

Read changes from log stream staging folder

Log stream staging task: No Log Stream tasks

Connection Information

Connection type: Application

Server name: pe-atun-p-01

Instance number: 00

Client: 800

Username: qlik

Password:

Backend endpoint: HANA Browse...

Test Connection Save Cancel

Close

6. Click **Test Connection**, resolve any errors, and then save your new source endpoint.

Create your task

1. Click **New Task**. Give it a descriptive name, such as **SAP to BigQuery**.
2. For **Replication Profile**, select **Unidirectional**. (This means the replication goes from one source to one target.)
3. Select at least one value under **Task Options**.
 - Select **Full Load** to start the initial replication of existing data from SAP to BigQuery.
 - Select **Apply Changes** to turn on CDC and replicate new transactions as they occur.
4. Click **OK**.
5. Drag the **SAP ERP Application** source endpoint to the top bubble, and drag the **BigQuery** target endpoint to the bottom bubble.

New Task
✕

Name i

Description


Replication Profile
Select the profile that best describes your replication use case

Unidirectional


Bidirectional i

Log Stream i


Task Options
Select the replication options for this task



Full Load



Apply Changes



Store Changes

Perform a full load of target tables, then keep them up-to-date by applying changes.



6. Click **Task Settings**.

This example provides a default target table schema that contains the tables in BigQuery.

7. Enter **SAP_SD** in the target metadata.

8. Click **Table Selection**. For **Business groups**, select **A – QLIK SALES & DISTRIBUTION**, and then click **Search**.

The *business group* maintained in SAP is a collection of logical and physical tables within SAP. A list of SAP Table objects appears for **Results**.

Select Tables

Source Tables ⓘ

Business groups: A - QLIK SALES AND DISTRIBUTION ▼

Table: %

Search

Results

A - QLIK SALES AND DISTRIB...	DD04T
A - QLIK SALES AND DISTRIB...	ISEG
A - QLIK SALES AND DISTRIB...	KNA1
A - QLIK SALES AND DISTRIB...	KNB1
A - QLIK SALES AND DISTRIB...	KNVV
A - QLIK SALES AND DISTRIB...	LIKP
A - QLIK SALES AND DISTRIB...	LIPS
A - QLIK SALES AND DISTRIB...	MAKT
A - QLIK SALES AND DISTRIB...	MARA
A - QLIK SALES AND DISTRIB...	RBKP
A - QLIK SALES AND DISTRIB...	RSEG

Tables:35

9. Select the tables you want to replicate into BigQuery, and click the > button to move them to the **Selected** window on the right side of the screen. Click **OK**.

Next, you create a global transformation. *Global* means “across all tables within the scope of your Replicate Task.” The global transformation renames the schema to something meaningful. (Don’t use spaces in the global transformation name.) without spaces in its name.

10. Click **Global Transformations**, then click **New Global Transformation** and use the following settings:
 - **Name:** SAP_SD
 - **Table or Schema:** Choose **Rename schema**, and then click **Next**.
 - **Schema name is like %:** Click **Next** to skip past this option.
 - **Schema name:** Rename the schema to SAP_SD.
11. Click **Finish**.

Some SAP column names contain a forward-slash (/), which is not a permitted character in BigQuery or many other databases. Replace forward-slashes with underscores (_) using an expression in another Global Transformation.

New Transformation Rule

Which Global Transformation?
Select Global Transformation type

Name:

Table or Schema

Column

Rename column

Add column

Drop column

Convert data type

12. Click **New Global Transformation** and use the following settings:

- **Name:** Remove slashes
- **Expand Column:** Select **Rename column**, and click **Next**.
- **% of Column name is like %:** Replace with %/% (percent-slash-percent), and click **Next**.
- **How to Transform:** Choose **Rename column to expression**, and enter the following:

```
replace( $AR_M_SOURCE_COLUMN_NAME, ' / ', '_ ' )
```

- Click **Next**, then **Finish**, then **OK**.

Global Transformations Rules □ ×

+ New Global Transformation... 🗑 Delete... 🔗 Open...

Name	Description
SAP_SD	Rename schema %.% with column % to SAP_SD
Remove Slashes	Rename column %.% with column %/% to replace(\$AR_M_SOURCE_COLU...

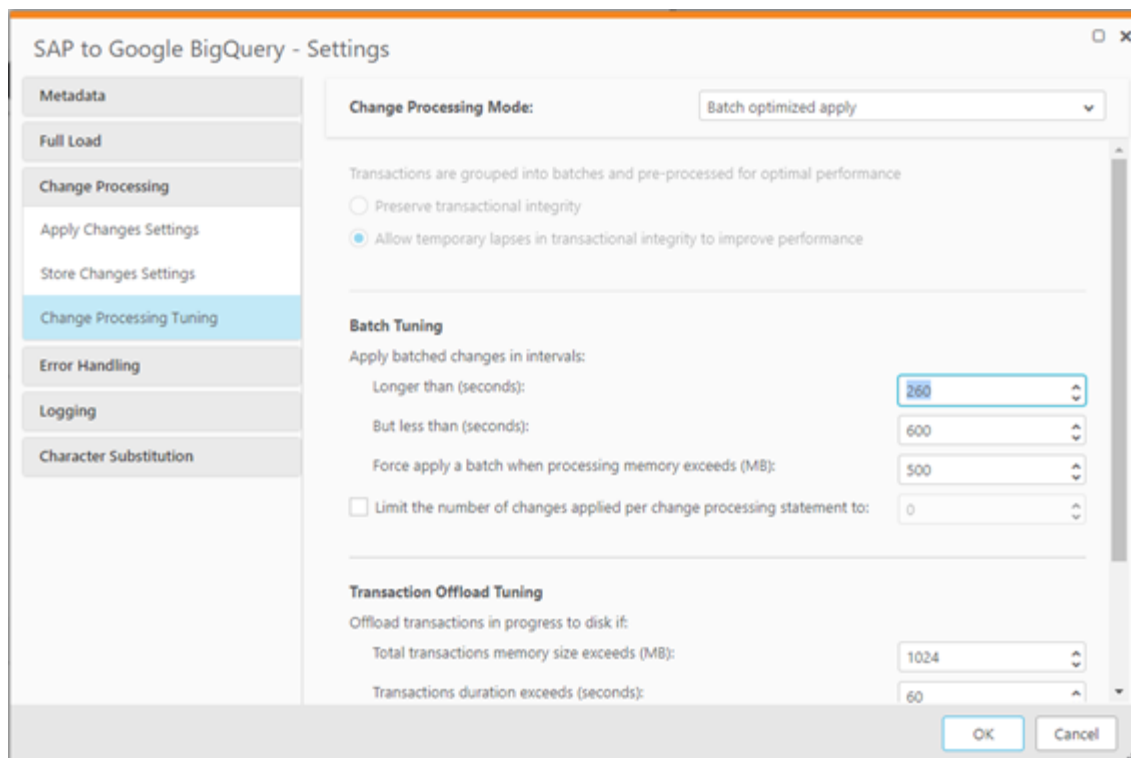
Work within quotas

BigQuery enforces quotas on the maximum number of operations that can be performed per day. These quotas are as follows:

- A maximum of 1000 load operations per table, per day
- A maximum of 1000 DML operations per table, per day

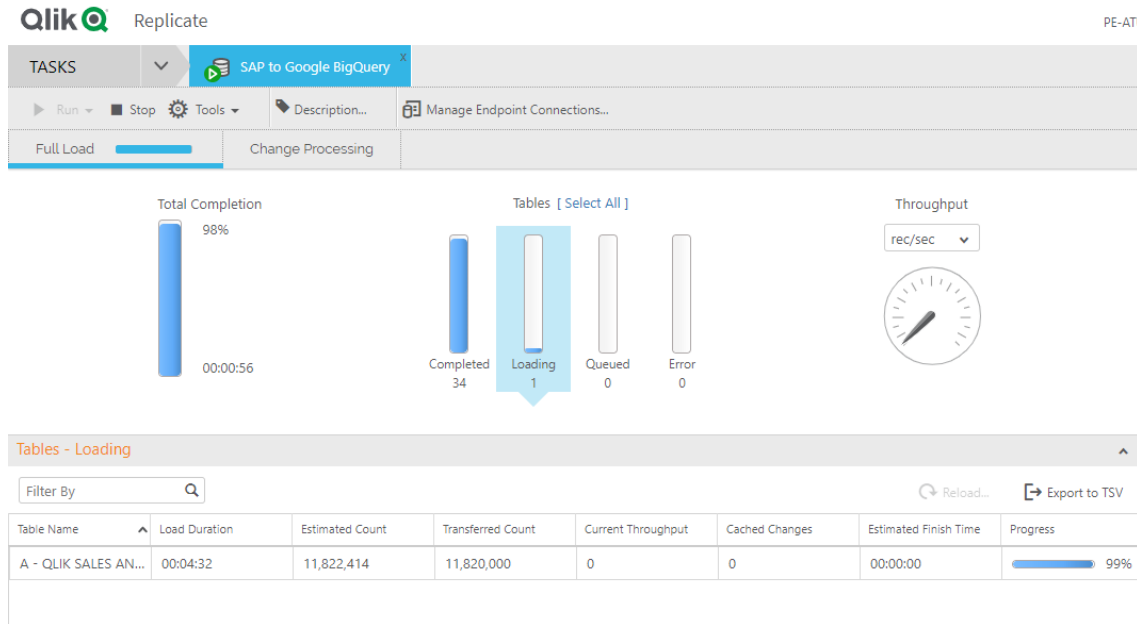
To work around these limitations, edit the following settings:

1. In the **Advanced Tab** of the target endpoint, increase the **Max file size (MB)** to a value greater than 1000.
2. Open **Task Settings** and go to **Change Processing Tuning**. Do one of the following:
 - To capture all DML operations, set **Apply batched changes in intervals > Longer than (seconds)** to **260** seconds.
 - To capture INSERT operations only, set **Apply batched changes in intervals > Longer than (seconds)** to **85** seconds.



The task is ready to execute. The initial **Full Load** will have some impact on the database, so wait to click **Run** until there'll be minimum impact on SAP users. Subsequent CDC transactions are replicated with barely any impact on SAP.

When the **Full Load** task is complete, you can see and interact with that data in BigQuery:



The screenshot shows the Qlik Replicate interface for a task named 'SAP to Google BigQuery'. The task is currently in the 'Full Load' phase, which is indicated by a blue progress bar at 98% completion. The interface includes a toolbar with 'Run', 'Stop', and 'Tools' buttons, along with 'Description...' and 'Manage Endpoint Connections...' options. Below the toolbar, there are three main sections: 'Total Completion' showing 98% progress and a timer at 00:00:56; 'Tables [Select All]' showing a bar chart with 34 Completed, 1 Loading, 0 Queued, and 0 Error tables; and 'Throughput' showing a gauge and a dropdown menu set to 'rec/sec'. Below these sections is a table titled 'Tables - Loading' with columns for Table Name, Load Duration, Estimated Count, Transferred Count, Current Throughput, Cached Changes, Estimated Finish Time, and Progress. The table contains one entry: 'A - QLIK SALES AN...' with a load duration of 00:04:32, an estimated count of 11,822,414, a transferred count of 11,820,000, and a progress bar at 99%.

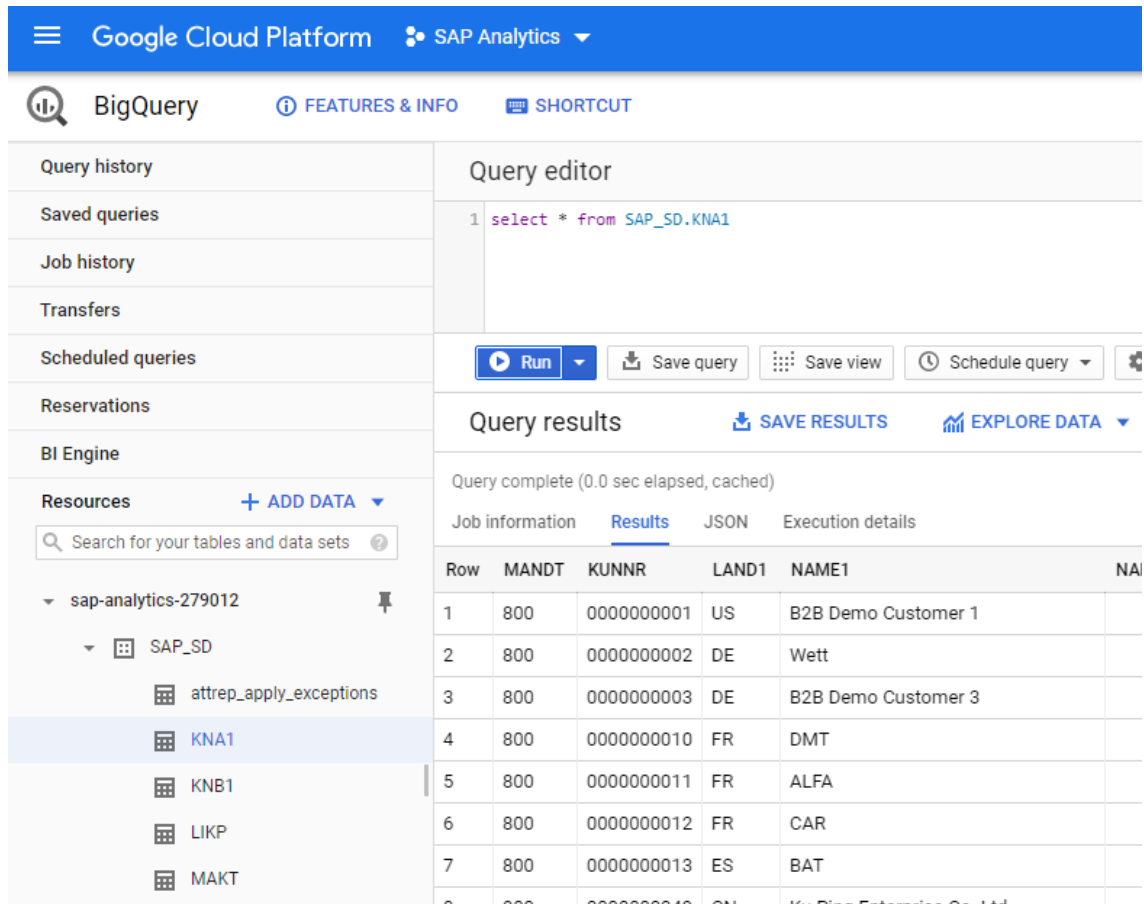
Table Name	Load Duration	Estimated Count	Transferred Count	Current Throughput	Cached Changes	Estimated Finish Time	Progress
A - QLIK SALES AN...	00:04:32	11,822,414	11,820,000	0	0	00:00:00	99%

If the task continues to run, you can monitor the **Change Processing** task to observe the replication of new transactions that affect the selected tables and are propagated from SAP into BigQuery.

View your replicated data

Navigate to the BigQuery console to see the data that has been replicated. Repeated queries will reveal transactional data that is replicated as it happens!

The following example shows that the SAP_SD table **KNA1** is created and populated with data.



The screenshot shows the Google Cloud Platform BigQuery console interface. The top navigation bar includes 'Google Cloud Platform' and 'SAP Analytics'. The main area is divided into a left sidebar and a main content area. The sidebar contains navigation options like 'Query history', 'Saved queries', 'Job history', 'Transfers', 'Scheduled queries', 'Reservations', 'BI Engine', and 'Resources'. The 'Resources' section is expanded to show a tree view of data sets under 'sap-analytics-279012', with 'SAP_SD' selected and 'KNA1' highlighted. The main content area shows the 'Query editor' with the SQL query `select * from SAP_SD.KNA1`. Below the editor are buttons for 'Run', 'Save query', 'Save view', and 'Schedule query'. The 'Query results' section shows the query is complete and displays a table of results.

Row	MANDT	KUNNR	LAND1	NAME1	NAI
1	800	0000000001	US	B2B Demo Customer 1	
2	800	0000000002	DE	Wett	
3	800	0000000003	DE	B2B Demo Customer 3	
4	800	0000000010	FR	DMT	
5	800	0000000011	FR	ALFA	
6	800	0000000012	FR	CAR	
7	800	0000000013	ES	BAT	
8	800	0000000014	CN	Ku Ping Enterprise Co., Ltd	



About Qlik

Qlik's vision is a data-literate world, one where everyone can use data to improve decision-making and solve their most challenging problems. Only Qlik offers end-to-end, real-time data integration and analytics solutions that help organizations access and transform all their data into value. Qlik helps companies lead with data to see more deeply into customer behavior, reinvent business processes, discover new revenue streams, and balance risk and reward. Qlik does business in more than 100 countries and serves over 50,000 customers around the world.

qlik.com

© 2020 QlikTech International AB. All rights reserved. All company and/or product names may be trade names, trademarks and/or registered trademarks of the respective owners with which they are associated.

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), and code samples are licensed under the [Apache 2.0 License](https://www.apache.org/licenses/LICENSE-2.0/). For details, see our [Site Policies](#). Oracle and Java are registered trademarks of Oracle and/or its affiliates.