Google Cloud

BigQuery Remote Functions

Course reading

This is a brief document that describes how you can integrate Cloud Run functions with BigQuery using BigQuery Remote Functions.

What are BigQuery Remote Functions

BigQuery is a fully managed, serverless enterprise data warehouse that helps you manage and analyze your data with features like machine learning, geospatial analysis, and business intelligence.

BigQuery's architecture lets you use SQL queries to query your data. It’s highly scalable and distributed, and lets you query terabytes of data in seconds and petabytes of data in minutes.

A BigQuery remote function lets you incorporate Google Standard SQL functionality with software outside of BigQuery by providing a direct integration with Cloud Run functions.

With BigQuery remote functions, you deploy your functions in Cloud Run functions, and then invoke them from Google Standard SQL queries.

To use BigQuery Remote Functions, create and deploy an HTTP function in Cloud Run functions.

In BigQuery create a connection of type CLOUD_RESOURCE, then create a remote function and use it in a query.

Configuring BigQuery Remote Functions

To use BigQuery Remote Functions, you first create and configure a BigQuery connection with these steps:

1. Enable the BigQuery Connection API.
2. Ensure that you have the necessary IAM role permissions. (for example, roles/bigquery.admin)
3. Create a connection of type CLOUD_RESOURCE using the Google Cloud console, bq CLI, or the connection API.
4. Grant the service account that is associated with the connection, the Cloud Functions Invoker role on your 1st gen function, or the Cloud Run Invoker role on your 2nd gen function.

Here is a sample bq command that creates a connection:

```
$ bq mk --connection --display_name='friendly name' \
--connection_type=CLOUD_RESOURCE \
--project_id=my_project_id --location=US my-connection
```

The next step is to create a remote function in BigQuery:

1. Ensure that you have the required role permissions on the BigQuery dataset where you create the remote function, and on the BigQuery connection which is used. (for example, `roles/bigquery.admin`)
2. Create a remote function in BigQuery with the CREATE FUNCTION statement and specify the BigQuery connection name and function URL endpoint.

Here is a sample BigQuery command to create a remote function:

```
CREATE FUNCTION my_project_id.my_dataset.function_name(x INT64, y
INT64) RETURNS INT64 \
REMOTE WITH CONNECTION ‘my_project_id.us.my-connection’ \
OPTIONS (endpoint =
'https://us-east1-my_gcf_project.cloudfunctions.net/function_name
')
```

Invoking Cloud Run Functions from BigQuery

To invoke your function from BigQuery:

- Ensure that you have the required role permissions on the dataset. (for example, `roles/bigquery.dataViewer`), and on the connection (`roles/bigquery.connectionUser`) that is used by the remote function.
- Use the BigQuery remote function in a query providing any required function arguments.

The sample function shown adds the second argument to the value of the first argument and returns the result for each row.

```
SELECT val, my_project_id.my_dataset.function_name(val, 2) FROM
UNNEST([NULL,2,3,5,8]) AS val;

Output:
Val | f0_
---------
Null| 2
2   | 4
3   | 5
5   | 7
8   | 10
```

Related documentation

For more information, view the documentation at the links provided:

Managing BigQuery connections

BigQuery Remote Functions