BigQuery Health Check Metadata Download Guide

Product Version	4.8.2
Document Type	Health Check Preparation Guide
Authors	BigQuery Data source Team
Reviewer	Red Team & Architects
Approver	CTO
Total Pages	6
Document Status	Draft

Table Of Contents

1.1	**Objectives**	**2**
1.2	**Architecture**	**2**
1.3	**Prerequisite**	**2**
1.4	**Download BigQuery Metadata for health check.**	**3**
1.5	**Output Files**	**5**

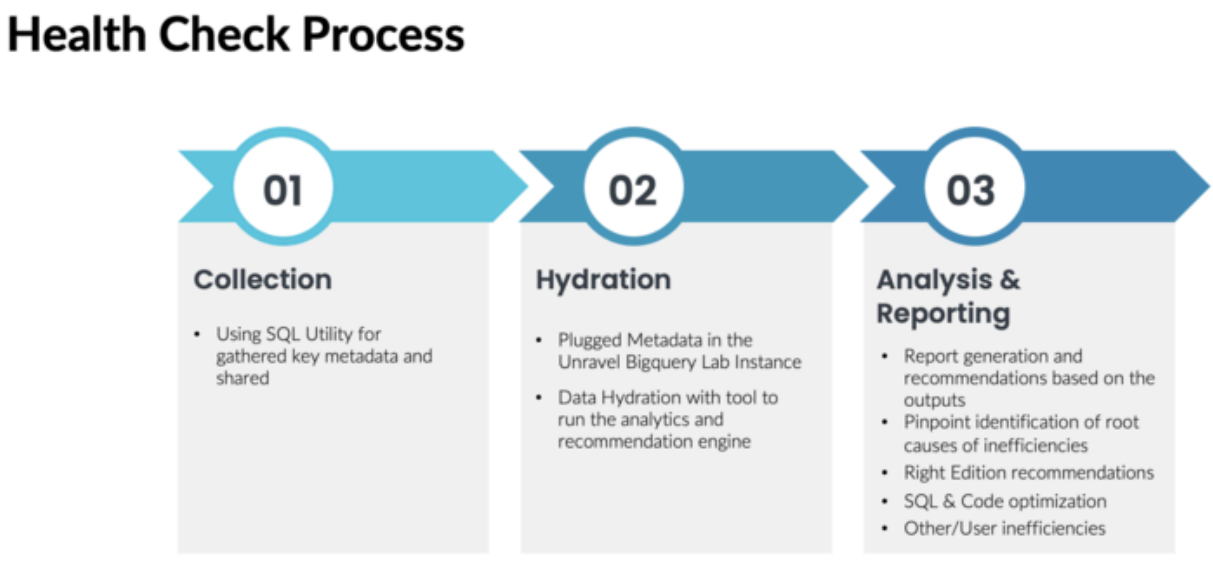
Document Version Record

Date	Version #	Author	Remarks / Reason
05-Jan-24	1.0	Dev Team	New Document

1.1 Objectives

Health check download for bigquery unravel product.

1.2 Architecture



1.3 Prerequisite

A. **BigQuery Permissions/roles required to download metadata :**
   a. Permissions required to download the JOBS, JOBS_TIMELINE metadata :
      i. bigquery.jobs.create
      ii. bigquery.jobs.listAll
      iii. storage.objects.create
      iv. storage.objects.list
   b. Permissions required to download Billing Data :
      i. bigquery.jobs.create
      ii. bigquery.tables.getData
      iii. storage.objects.create
      iv. storage.objects.list

1.4 Download BigQuery Metadata for health check.

Do the following to download BigQuery metadata :

1. Clone this repo: https://github.com/unraveldata-org/BigQuery-data-loader.git
2. Follow below Steps to create a Service Account key that has above roles which will authenticate to download the JOBS, JOBS_TIMELINE and Billing metadata :
   a. cd terraform/healthcheck
   b. cp input.tfvars.example input.tfvars
   c. open input.tfvars
   d. Under monitoring_project_ids add the project ids for which you want to download JOBS and JOBS_TIMELINE Data (perform health check)
   e. Under billing_project_ids add the project in which you have billing data exported
   f. Under svc_account_project_id add the project on which you want the service account to be created to execute the download script
   g. Now, open local.tf
   h. Under monitoring_project_role_permission add the below permissions, comma separated (ignore if the same permissions are already present) :
      i. "bigquery.jobs.create"
      ii. "bigquery.jobs.listAll"
      iii. "storage.objects.create"
      iv. "storage.objects.list"
   i. Under billing_project_role_permission add the below permissions, comma separated (ignore if the same permissions are already present) :

i. "bigquery.jobs.create"
ii. "bigquery.tables.getData"
iii. "storage.objects.create"
iv. "storage.objects.list"

j. Before using this project, you need to authenticate with Google Cloud using gcloud. Follow the instructions provided at https://cloud.google.com/sdk/docs/install-sdk for a one-time configuration. You can find the installation instruction based on the Machine Arch and OS installed in the above link.
k. To authenticate gcloud, execute the following commands:
    i. gcloud init
    ii. gcloud auth application-default login
l. Download and install terraform (Documentation)
m. Now, to run terraform execute the following :
    i. terraform init
    ii. terraform plan --var-file=input.tfvars
    iii. terraform apply --var-file=input.tfvars
n. Now, there must be a credential file created inside terraform/healthcheck/keys folder, you will need this credential file while providing configuration for the download script

3. cd Data Downloader Script
4. cp download_config.yaml.example download_config.yaml
5. Open download_config.yaml file and edit it with your configuration :

a. gcs_bucket_path: "add the gcs bucket path where you want to download the csv files" (To know how to create a gcs bucket check the Steps for creating a gcs bucket)
b. locations: "add comma separated regions where you execute your queries on the projects"
c. projects: "replace this with comma separated project_ids for which you want to download JOBS data / perform health check"
d. billing_project: "replace this with the project_id where billing data is exported"
e. billing_dataset: "replace this with the dataset_id where billing data is exported"
f. billing_table: "replace this with the table_name where billing data is exported"
g. credential: "replace this with path of the credential file created in previous step through terraform"

Note : Please use a double backslash when defining credential path in download_config.yaml file for Windows Machine

Example :

Windows :

credential: "C:\\Users\\user1\\path\to\\authentication\\key\\project-1.json"

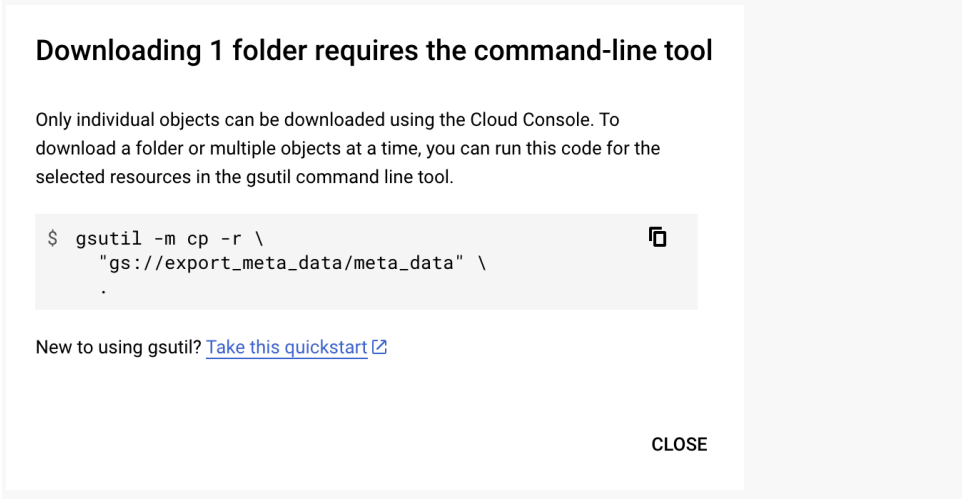
Mac/Linux :

credential: "/path/to/authentication/key/project-1.json"

6. Install required python packages using : pip3 install -r requirements.txt
7. Run download script with below command :

```
python3 /path/to/download_metadata.py --config_file
/path/to/download_config.yaml
```

8. CSV files containing the JOBS, JOBS_TIMELINE and BILLING metadata will be downloaded to the gcs bucket defined in download_config.yaml file
9. Now, download these files to your local by using google-sdk by following given steps :
    a. Go to the gcs bucket where metadata is downloaded
    b. Tick the folder where all the downloaded csv files are present
    c. Click on download button, you will get a dialog box like this :

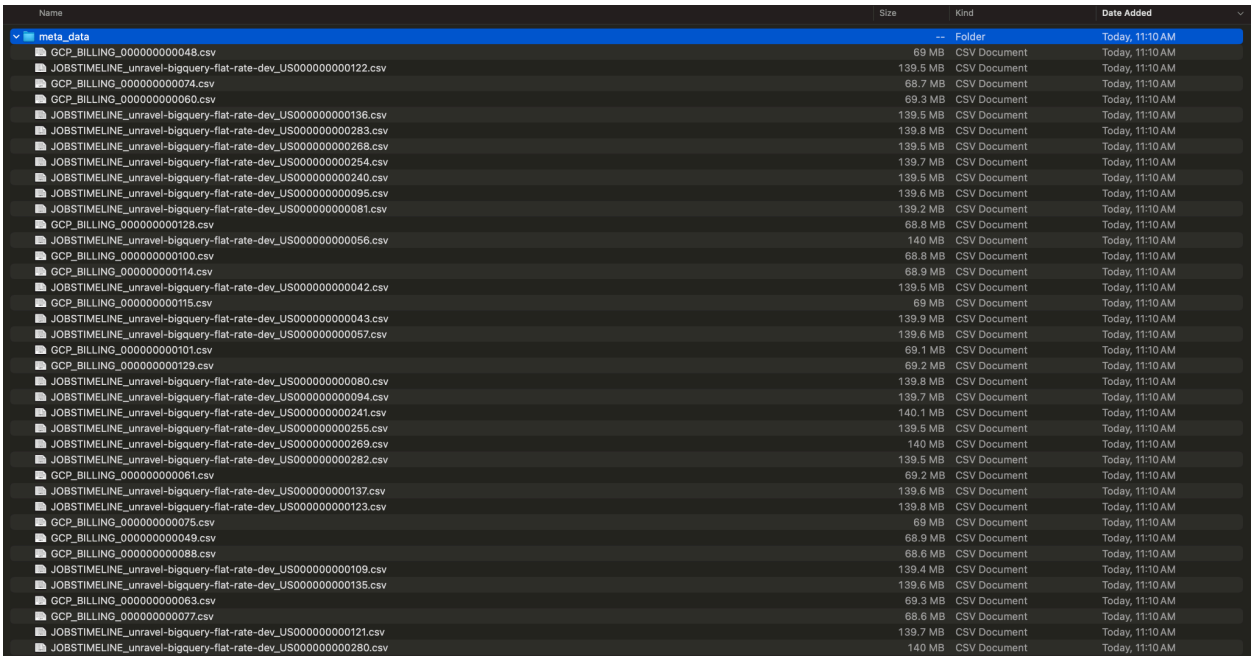


    d. copy the command and run it on gcloud cli installed previously, you will get all the files downloaded to the current directory on your local
10. To destroy the roles and key created by terraform, run :
    a. cd ../terraform/healthcheck/
    b. terraform destroy --var-file=input.tfvars
11. Share these files through Google Drive to the the Unravel POC

Steps for creating a gcs bucket :

1. Go to the Cloud Storage Buckets page: https://console.cloud.google.com/storage/browser in the Google Cloud console.
2. Click + Create.
3. On the Create a bucket page, enter the following information:
    a. Name your bucket: Enter a unique name for your bucket. The name must start with a lowercase letter or number, and it can contain up to 63 characters. It can also contain dashes and periods.
    b. Choose where to store your data: Select a location for your bucket. You can choose a location in the same region as your project, or you can choose a different region.
    c. Choose a storage class for your data: Select a storage class for your bucket. The storage class determines how long your data is kept and how much it costs to store.
4. Click Create.
5. Create new folder in bucket

1.5 Output Files (Sample)



Name	Size	Kind	Date Added
meta_data	--	Folder	Today, 11:10 AM
GCP_BILLING_000000000048.csv	69 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000122.csv	139.5 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000074.csv	68.7 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000060.csv	69.3 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000136.csv	139.5 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000283.csv	139.8 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000268.csv	139.5 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000254.csv	139.7 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000240.csv	139.5 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000095.csv	139.6 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000081.csv	139.2 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000128.csv	68.8 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000056.csv	140 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000100.csv	68.8 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000114.csv	68.9 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000042.csv	139.5 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000115.csv	69 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000043.csv	139.9 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000057.csv	139.6 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000101.csv	69.1 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000129.csv	69.2 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000080.csv	139.8 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000094.csv	139.7 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000241.csv	140.1 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000255.csv	139.5 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000269.csv	140 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000282.csv	139.5 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000061.csv	69.2 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000137.csv	139.6 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000123.csv	139.8 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000075.csv	69 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000049.csv	68.9 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000088.csv	68.6 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000109.csv	139.4 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000135.csv	139.6 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000063.csv	69.3 MB	CSV Document	Today, 11:10 AM
GCP_BILLING_000000000077.csv	68.6 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000121.csv	139.7 MB	CSV Document	Today, 11:10 AM
JOBSTIMELINE_unravel-bigquery-flat-rate-dev_US000000000280.csv	140 MB	CSV Document	Today, 11:10 AM

1.6 Video(s) :

- https://drive.google.com/file/d/1U3nhRFyKnfjrhUacoV1V-nHD12hb19CQ/view?usp=sharing