

Analyse

- BigQuery
- Dataflow
- Dataproc

BigQuery

Prix du stockage des données :

BigQuery- STOCKAGE					
Durée du stockage	1 Go	10 Go	100 Go	1024 Go	10 To
Moins de 3 mois	0,02	0,2	2	20,48	204,8
Plus de trois mois	0,01	0,1	1	10,24	102,4

Prix des requêtes des données :

BIGTABLE - Données Requêtées					
Taille des requêtes	1 Go	10 Go	100 Go	1024 Go	10 To
Prix	0,0048828125	0,048828125	0,48828125	5	50

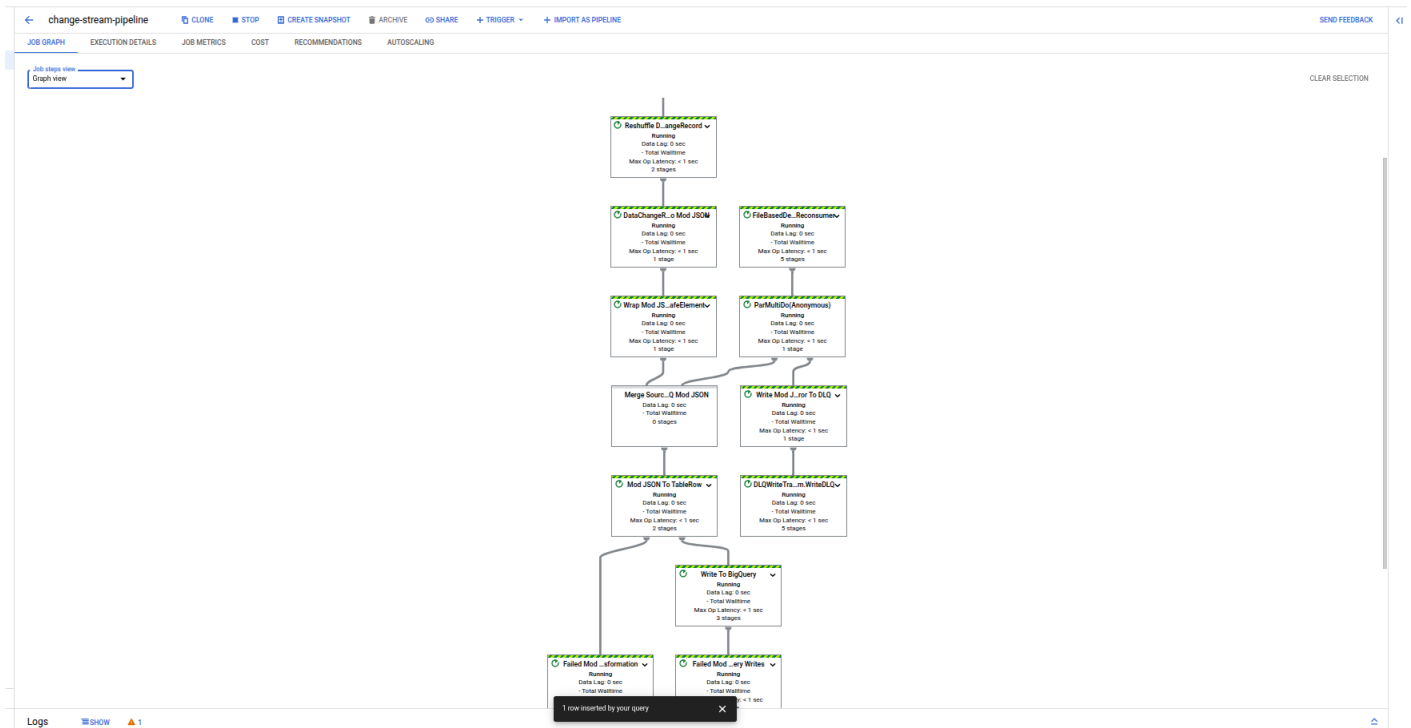
Commandes :

Lors de l'importation d'un fichier on peut spécifier `--autodetect` pour que BigQueryry comprène automatiquement le schéma.

Exemple d'insertion de donnée dans BigQuery :

```
INSERT INTO DATASET.test (colonne1, colonne2, colonne3, colonne4) VALUES (1,2,3,4)
```

Dataflow



Dataproc

Clusters

CREATE CLUSTER

REFRESH

START

STOP

DELETE

REGIONS

+ 5 RECOMMENDED ALERTS

LEARN

SHOW INFO PANEL

Filter

Search cluster by properties, press Enter

Sorry, the server was not able to fulfill your request.

☐	Name ↑	Status	Region	Zone	Total worker nodes	Flexible VMs?	Scheduled deletion	Cloud Storage staging bucket	Created	Labels
☐	example-cluster	Running	us-central1	us-central1-f	4	No	Off	dataproc-staging-us-central1-590053044234-qnywjids	Jul 22, 2025, 1:02:14 PM	goog-dataproc-...: example-cl... goog-dataproc-...: eaab6b01-8... goog-dataproc-...: cluster-ea... goog-dataproc-...: us-central1

Cluster details

SUBMIT JOB

REFRESH

START

STOP

DELETE

VIEW LOGS

Consider using Auto Zone rather than selecting a zone manually. See <https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/auto-zone>

example-cluster

Cluster UUID: eaab6b01-8e21-489e-9984-f980688d5cb9

Type: Dataproc Cluster

Status: Running

MONITORING

JOB

VM INSTANCES

CONFIGURATION

WEB INTERFACES

SAVE AS DASHBOARD

RESET ZOOM

YARN memory

YARN pending memory

YARN NodeManagers

HDFS capacity

CPU utilization

Network Bytes

Network Packets

Disk bytes

Disk operations

Request to update cluster example-cluster submitted

Cluster details

SUBMIT JOB

REFRESH

START

STOP

DELETE

VIEW LOGS

Consider using Auto Zone rather than selecting a zone manually. See <https://cloud.google.com/dataproc/docs/concepts/configuring-clusters/auto-zone>

example-cluster

Cluster UUID: eaab6b01-8e21-489e-9984-f980688d5cb9

Type: Dataproc Cluster

Status: Running

MONITORING

JOB

VM INSTANCES

CONFIGURATION

WEB INTERFACES

Filter

Filter instances

☒	Name	Role	Machine type
☒	example-cluster-m	Master	e2-standard-2
☒	example-cluster-w-0	Worker	e2-standard-2
☒	example-cluster-w-1	Worker	e2-standard-2
☒	example-cluster-w-2	Worker	e2-standard-2
☒	example-cluster-w-3	Worker	e2-standard-2

EQUIVALENT REST

Job details

CLONE

DELETE

STOP

REFRESH

Job ID

Job UUID

Type

Status

job-d965796e

ae3017b1-806a-45cc-ef5c-6b6e08a8fcc2

Dataproc Job

Succeeded

MONITORING

CONFIGURATION

The charts below represent the metrics from the cluster this job ran on, scoped to the time that this job was running. It is possible for more than one job to run on a cluster at a time, so these metrics may not reflect this job's resource usage accurately. Metrics for a job may lag behind the job run by several minutes.

SAVE AS DASHBOARD

RESET ZOOM

YARN memory

YARN pending memory

YARN NodeManagers

HDFS capacity

CPU utilization

1 hour6 hours12 hours1 day2 days4 days7 days14 days30 daysCustom 1:03 PM - 1:08 PM

150GB

100GB

10GB

2

1.5

1

30GB

20GB

60%50%40%

Output

LINE WRAP: OFF

Spark jobs take ~60 seconds to initialize resources.

DISMISS

25/07/22 11:07:27 INFO SparkEnv: Registering MapOutputTracker

25/07/22 11:07:27 INFO SparkEnv: Registering BlockManagerMaster

25/07/22 11:07:27 INFO SparkEnv: Registering BlockManagerMasterHeartbeat

25/07/22 11:07:27 INFO SparkEnv: Registering OutputCommitCoordinator

25/07/22 11:07:29 INFO MetricsConfig: Loaded properties from hadoop-metrics2.properties

25/07/22 11:07:29 INFO MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).

25/07/22 11:07:29 INFO MetricsSystemImpl: google-hadoop-file-system metrics system started

25/07/22 11:07:30 INFO DataprocSparkPlugin: Registered 188 driver metrics

25/07/22 11:07:31 INFO DefaultHadoopHwilerProxyProvider: Connecting to ResourceManager at example-cluster-m-us-central1-f.c.qwiklabs-gcp-01-b6ed1e98aac.internal./10.128.0.3:8032

25/07/22 11:07:32 INFO HSFProxy: Connecting to Application History server at example-cluster-m-us-central1-f.c.qwiklabs-gcp-01-b6ed1e98aac.internal./10.128.0.3:10200

25/07/22 11:07:34 INFO Configuration: resource-types.xml not found

25/07/22 11:07:34 INFO ResourceUtils: Unable to find 'resource-types.xml'.

25/07/22 11:07:38 INFO YarnClientImpl: Submitted application application_1753182328888_0001

25/07/22 11:07:40 INFO DefaultHadoopHwilerProxyProvider: Connecting to ResourceManager at example-cluster-m-us-central1-f.c.qwiklabs-gcp-01-b6ed1e98aac.internal./10.128.0.3:8030

25/07/22 11:07:44 INFO RequestTracker: Detected high latency for [url=https://storage.googleapis.com/storage/v1/b/dataproc-temp-us-central1-590853844234-hwervp2h/s/eaabb01-8e21-489e-9984-f980688d5cb9?spark-job-history?fields=bucket_name,timeCreated,updated,generation,metageneration,size,o

25/07/22 11:07:44 INFO GoogleCloudStorageImpl: Ignoring exception of type GoogleJsonResponseException: verified object already exists with desired state.

25/07/22 11:07:45 INFO GoogleHadoopOutputStream: hFlush(): No-op due to rate limit (RateLimiter[stableRate=0.2ops/s]): readers will "not" yet see flushed data for gs://dataproc-temp-us-central1-590853844234-hwervp2h/eaabb01-8e21-489e-9984-f980688d5cb9/spark-job-history/application_17531823288

25/07/22 11:08:02 INFO DhfsGlobalStorageStatistics: periodic connector metrics: (action_http_post_request=4, action_http_post_request_duration=374, action_http_post_request_max=140, action_http_post_request_mean=93, action_http_post_request_min=46, action_http_put_request=2, action_http_put_

CONTEXT rateLimit_period=5 MINUTES"]

25/07/22 11:08:21 INFO DhfsGlobalStorageStatistics: Detected potential high latency for operation op.hflush. latencyMs=612; previousMaxLatencyMs=443; operationCount=15; context-gs://dataproc-temp-us-central1-590853844234-hwervp2h/eaabb01-8e21-489e-9984-f980688d5cb9/spark-job-history/applica

25/07/22 11:08:21 INFO DataprocSparkPlugin: Shutting down driver plugin. metrics=[action_http_patch_request=0, files_created=1, gcs_api_server_timeout_count=0, op_get_list_status_result_size=0, op_open=0, action_http_delete_request=2, gcs_api_time=3563, gcs_backoff_count=0, gcs_api_client_un

EQUIVALENT COMMAND LINE

Request to update cluster example-cluster submitted