

\*\*\* Reading remote log from gs://us-east1-spark-airflow-comp-9c560139-bucket/logs/dag\_id=Trade\_Quote\_Analysis/run\_id=scheduled\_\_2022-11-05T00:00:00+00:00/task\_id=ingest\_trade\_data/attempt=1.log.

[2022-11-06, 03:55:36 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 03:55:36 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 03:55:36 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 03:55:36 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 03:55:36 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 03:55:36 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocCreateBatchOperator): ingest\_trade\_data> on 2022-11-05 00:00:00+00:00

[2022-11-06, 03:55:36 UTC] {standard\_task\_runner.py:52} INFO - Started process 2179 to run task

[2022-11-06, 03:55:36 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'ingest\_trade\_data', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '50', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmp5yd0s0z8', '--error-file', '/tmp/tmpj885cchz']

[2022-11-06, 03:55:36 UTC] {standard\_task\_runner.py:80} INFO - Job 50: Subtask ingest\_trade\_data

[2022-11-06, 03:55:37 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 03:55:37 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=ingest\_trade\_data

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

[2022-11-06, 03:55:37 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

[2022-11-06, 03:55:37 UTC] {dataproc.py:2022} INFO - Creating batch

[2022-11-06, 03:58:48 UTC] {taskinstance.py:1911} ERROR - Task failed with exception

Traceback (most recent call last):

File "/opt/python3.8/lib/python3.8/site-packages/airflow/providers/google/cloud/hooks/dataproc.py", line 246, in wait\_for\_operation

return operation.result(timeout=timeout)

File "/opt/python3.8/lib/python3.8/site-packages/google/api\_core/future/polling.py", line 137, in result

raise self.\_exception

google.api\_core.exceptions.Aborted: 409 Google Cloud Dataproc Agent reports job failure. If logs are available, they can be found at:

<https://console.cloud.google.com/dataproc/batches/us-east1/ingest-batch-create?project=my-project-367217>

gcloud dataproc batches wait 'ingest-batch-create' --region 'us-east1' --project my-project-367217'

<https://console.cloud.google.com/storage/browser/dataproc-staging-us-east1-665710256196-gtkvwse4/google-cloud-dataproc-metainfo/77dd9478-2848-4980-a371-622d8ec9982d/jobs/srvls-batch-e2f71042-8ef3-42bc-a60e-dad5fc7f905f/>

gs://dataproc-staging-us-east1-665710256196-gtkvwse4/google-cloud-dataproc-metainfo/77dd9478-2848-4980-a371-622d8ec9982d/jobs/srvls-batch-e2f71042-8ef3-42bc-a60e-dad5fc7f905f/driveroutput

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/opt/python3.8/lib/python3.8/site-packages/airflow/providers/google/cloud/operators/dataproc.py", line 2038, in execute

result = hook.wait\_for\_operation(timeout=self.timeout, operation=self.operation)

File "/opt/python3.8/lib/python3.8/site-packages/airflow/providers/google/cloud/hooks/dataproc.py", line 249, in wait\_for\_operation

raise AirflowException(error)

airflow.exceptions.AirflowException: 409 Google Cloud Dataproc Agent reports job failure. If logs are available, they can be found at:

[https://console.cloud.google.com/dataproc/batches/us-east1/ingest-batch-create?project= my-project-367217](https://console.cloud.google.com/dataproc/batches/us-east1/ingest-batch-create?project=splendid-planet-367217)

gcloud dataproc batches wait 'ingest-batch-create' --region 'us-east1' --project my-project-367217

<https://console.cloud.google.com/storage/browser/dataproc-staging-us-east1-665710256196-gtkvwse4/google-cloud-dataproc-metainfo/77dd9478-2848-4980-a371-622d8ec9982d/jobs/srvls-batch-e2f71042-8ef3-42bc-a60e-dad5fc7f905f/>

gs://dataproc-staging-us-east1-665710256196-gtkvwse4/google-cloud-dataproc-metainfo/77dd9478-2848-4980-a371-622d8ec9982d/jobs/srvls-batch-e2f71042-8ef3-42bc-a60e-dad5fc7f905f/driveroutput

[2022-11-06, 03:58:48 UTC] {taskinstance.py:1416} INFO - Marking task as UP\_FOR\_RETRY. dag\_id=Trade\_Quote\_Analysis, task\_id=ingest\_trade\_data, execution\_date=20221105T000000, start\_date=20221106T035536, end\_date=20221106T035848

[2022-11-06, 03:58:48 UTC] {standard\_task\_runner.py:92} ERROR - Failed to execute job 50 for task ingest\_trade\_data (409 Google Cloud Dataproc Agent reports job failure. If logs are available, they can be found at:

[https://console.cloud.google.com/dataproc/batches/us-east1/ingest-batch-create?project= my-project-367217](https://console.cloud.google.com/dataproc/batches/us-east1/ingest-batch-create?project=splendid-planet-367217)

gcloud dataproc batches wait 'ingest-batch-create' --region 'us-east1' --project my-project-367217

<https://console.cloud.google.com/storage/browser/dataproc-staging-us-east1-665710256196-gtkvwse4/google-cloud-dataproc-metainfo/77dd9478-2848-4980-a371-622d8ec9982d/jobs/srvls-batch-e2f71042-8ef3-42bc-a60e-dad5fc7f905f/>

gs://dataproc-staging-us-east1-665710256196-gtkvwse4/google-cloud-dataproc-metainfo/77dd9478-2848-4980-a371-622d8ec9982d/jobs/srvls-batch-e2f71042-8ef3-42bc-a60e-dad5fc7f905f/driveroutput; 2179)

[2022-11-06, 03:58:48 UTC] {local\_task\_job.py:156} INFO - Task exited with return code 1

[2022-11-06, 04:01:00 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:01:00 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:01:00 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:01:00 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 04:01:00 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:01:00 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocCreateBatchOperator): ingest\_trade\_data> on 2022-11-05 00:00:00+00:00

[2022-11-06, 04:01:00 UTC] {standard\_task\_runner.py:52} INFO - Started process 2306 to run task

[2022-11-06, 04:01:00 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'ingest\_trade\_data', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '52', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmpr\_8954tc', '--error-file', '/tmp/tmpwp7zsb5t']

[2022-11-06, 04:01:00 UTC] {standard\_task\_runner.py:80} INFO - Job 52: Subtask ingest\_trade\_data

[2022-11-06, 04:01:01 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 04:01:01 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=ingest\_trade\_data

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

[2022-11-06, 04:01:01 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

[2022-11-06, 04:01:01 UTC] {dataproc.py:2022} INFO - Creating batch

[2022-11-06, 04:01:02 UTC] {dataproc.py:2041} INFO - Batch with given id already exists

[2022-11-06, 04:01:02 UTC] {taskinstance.py:1416} INFO - Marking task as SUCCESS. dag\_id=Trade\_Quote\_Analysis, task\_id=ingest\_trade\_data, execution\_date=20221105T000000, start\_date=20221106T040100, end\_date=20221106T040102

[2022-11-06, 04:01:02 UTC] {local\_task\_job.py:156} INFO - Task exited with return code 0

[2022-11-06, 04:02:09 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:02:09 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:02:09 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:02:09 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 04:02:09 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:02:09 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocCreateBatchOperator): ingest\_trade\_data> on 2022-11-05 00:00:00+00:00

[2022-11-06, 04:02:09 UTC] {standard\_task\_runner.py:52} INFO - Started process 2366 to run task

[2022-11-06, 04:02:09 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'ingest\_trade\_data', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '55', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmp7x4hx83\_', '--error-file', '/tmp/tmp4ym7\_9vp']

[2022-11-06, 04:02:09 UTC] {standard\_task\_runner.py:80} INFO - Job 55: Subtask ingest\_trade\_data

[2022-11-06, 04:02:10 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 04:02:10 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=ingest\_trade\_data

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

[2022-11-06, 04:02:10 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

[2022-11-06, 04:02:10 UTC] {dataproc.py:2022} INFO - Creating batch

[2022-11-06, 04:02:11 UTC] {dataproc.py:2041} INFO - Batch with given id already exists

[2022-11-06, 04:02:11 UTC] {taskinstance.py:1416} INFO - Marking task as SUCCESS. dag\_id=Trade\_Quote\_Analysis, task\_id=ingest\_trade\_data, execution\_date=20221105T000000, start\_date=20221106T040209, end\_date=20221106T040211

[2022-11-06, 04:02:11 UTC] {local\_task\_job.py:156} INFO - Task exited with return code 0

[2022-11-06, 04:06:27 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:06:27 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:06:27 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:06:27 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 04:06:27 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:06:27 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocCreateBatchOperator): ingest\_trade\_data> on 2022-11-05 00:00:00+00:00

[2022-11-06, 04:06:27 UTC] {standard\_task\_runner.py:52} INFO - Started process 2473 to run task

[2022-11-06, 04:06:27 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'ingest\_trade\_data', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '57', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmp0x6hozoe', '--error-file', '/tmp/tmpt8l0v2mu']

[2022-11-06, 04:06:27 UTC] {standard\_task\_runner.py:80} INFO - Job 57: Subtask ingest\_trade\_data

[2022-11-06, 04:06:28 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.ingest\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 04:06:29 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=ingest\_trade\_data

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

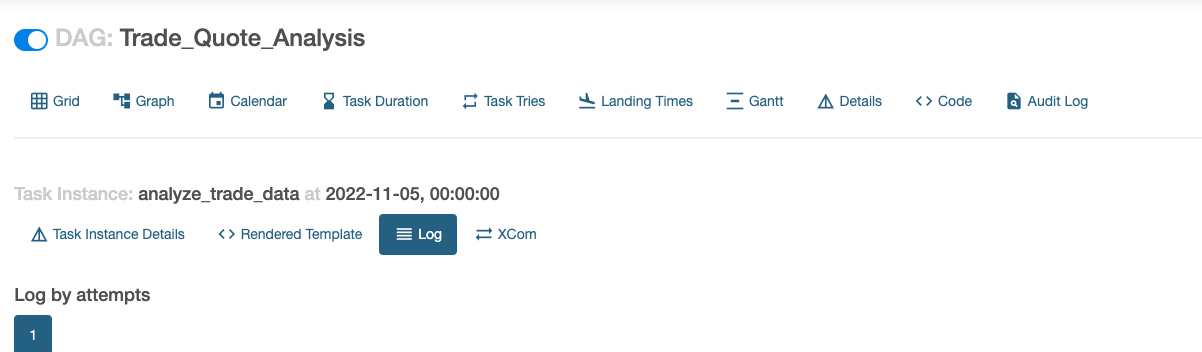
[2022-11-06, 04:06:29 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

[2022-11-06, 04:06:29 UTC] {dataproc.py:2022} INFO - Creating batch

[2022-11-06, 04:09:29 UTC] {dataproc.py:2039} INFO - Batch ingest-batch-create created

[2022-11-06, 04:09:30 UTC] {taskinstance.py:1416} INFO - Marking task as SUCCESS. dag\_id=Trade\_Quote\_Analysis, task\_id=ingest\_trade\_data, execution\_date=20221105T000000, start\_date=20221106T040627, end\_date=20221106T040930

[2022-11-06, 04:09:30 UTC] {local\_task\_job.py:156} INFO - Task exited with return code 0



\*\*\* Reading remote log from gs://us-east1-spark-airflow-comp-9c560139-bucket/logs/dag\_id=Trade\_Quote\_Analysis/run\_id=scheduled\_\_2022-11-05T00:00:00+00:00/task\_id=analyze\_trade\_data/attempt=1.log.

[2022-11-06, 04:01:05 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.analyze\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:01:06 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.analyze\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:01:06 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:01:06 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 04:01:06 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:01:06 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocCreateBatchOperator): analyze\_trade\_data> on 2022-11-05 00:00:00+00:00

[2022-11-06, 04:01:06 UTC] {standard\_task\_runner.py:52} INFO - Started process 2313 to run task

[2022-11-06, 04:01:06 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'analyze\_trade\_data', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '53', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmpvb06du22', '--error-file', '/tmp/tmpe63o4qyl']

[2022-11-06, 04:01:06 UTC] {standard\_task\_runner.py:80} INFO - Job 53: Subtask analyze\_trade\_data

[2022-11-06, 04:01:06 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.analyze\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 04:01:07 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=analyze\_trade\_data

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

[2022-11-06, 04:01:07 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

[2022-11-06, 04:01:07 UTC] {dataproc.py:2022} INFO - Creating batch

[2022-11-06, 04:01:37 UTC] {local\_task\_job.py:220} WARNING - State of this instance has been externally set to None. Terminating instance.

[2022-11-06, 04:01:37 UTC] {process\_utils.py:125} INFO - Sending Signals.SIGTERM to group 2313. PIDs of all processes in the group: [2313]

[2022-11-06, 04:01:37 UTC] {process\_utils.py:80} INFO - Sending the signal Signals.SIGTERM to group 2313

[2022-11-06, 04:01:37 UTC] {taskinstance.py:1563} ERROR - Received SIGTERM. Terminating subprocesses.

[2022-11-06, 04:03:43 UTC] {taskinstance.py:1911} ERROR - Task failed with exception

Traceback (most recent call last):

File "/opt/python3.8/lib/python3.8/site-packages/airflow/providers/google/cloud/hooks/dataproc.py", line 246, in wait\_for\_operation

return operation.result(timeout=timeout)

File "/opt/python3.8/lib/python3.8/site-packages/google/api\_core/future/polling.py", line 132, in result

self.\_blocking\_poll(timeout=timeout, \*\*kwargs)

File "/opt/python3.8/lib/python3.8/site-packages/google/api\_core/future/polling.py", line 110, in \_blocking\_poll

retry\_(self.\_done\_or\_raise)(\*\*kwargs)

File "/opt/python3.8/lib/python3.8/site-packages/google/api\_core/retry.py", line 283, in retry\_wrapped\_func

return retry\_target(

File "/opt/python3.8/lib/python3.8/site-packages/google/api\_core/retry.py", line 218, in retry\_target

time.sleep(sleep)

File "/opt/python3.8/lib/python3.8/site-packages/airflow/models/taskinstance.py", line 1565, in signal\_handler

raise AirflowException("Task received SIGTERM signal")

airflow.exceptions.AirflowException: Task received SIGTERM signal

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/opt/python3.8/lib/python3.8/site-packages/airflow/providers/google/cloud/operators/dataproc.py", line 2038, in execute

result = hook.wait\_for\_operation(timeout=self.timeout, operation=self.operation)

File "/opt/python3.8/lib/python3.8/site-packages/airflow/providers/google/cloud/hooks/dataproc.py", line 249, in wait\_for\_operation

raise AirflowException(error)

airflow.exceptions.AirflowException: None

[2022-11-06, 04:03:43 UTC] {taskinstance.py:1416} INFO - Marking task as UP\_FOR\_RETRY. dag\_id=Trade\_Quote\_Analysis, task\_id=analyze\_trade\_data, execution\_date=20221105T000000, start\_date=, end\_date=20221106T040343

[2022-11-06, 04:03:43 UTC] {standard\_task\_runner.py:92} ERROR - Failed to execute job 53 for task analyze\_trade\_data ((psycopg2.errors.UniqueViolation) duplicate key value violates unique constraint "dag\_run\_dag\_id\_run\_id\_key"

DETAIL: Key (dag\_id, run\_id)=(Trade\_Quote\_Analysis, scheduled\_\_2022-11-05T00:00:00+00:00) already exists.

[SQL: INSERT INTO dag\_run (id, dag\_id, queued\_at, execution\_date, start\_date, end\_date, state, run\_id, creating\_job\_id, external\_trigger, run\_type, conf, data\_interval\_start, data\_interval\_end, last\_scheduling\_decision, dag\_hash, log\_template\_id) VALUES (%(id)s, %(dag\_id)s, %(queued\_at)s, %(execution\_date)s, %(start\_date)s, %(end\_date)s, %(state)s, %(run\_id)s, %(creating\_job\_id)s, %(external\_trigger)s, %(run\_type)s, %(conf)s, %(data\_interval\_start)s, %(data\_interval\_end)s, %(last\_scheduling\_decision)s, %(dag\_hash)s, %(log\_template\_id)s)]

[parameters: {'id': 31, 'dag\_id': 'Trade\_Quote\_Analysis', 'queued\_at': datetime.datetime(2022, 11, 6, 4, 0, 58, 24416, tzinfo=Timezone('UTC')), 'execution\_date': datetime.datetime(2022, 11, 5, 0, 0, tzinfo=Timezone('UTC')), 'start\_date': datetime.datetime(2022, 11, 6, 4, 0, 58, 69679, tzinfo=Timezone('UTC')), 'end\_date': None, 'state': 'running', 'run\_id': 'scheduled\_\_2022-11-05T00:00:00+00:00', 'creating\_job\_id': 48, 'external\_trigger': False, 'run\_type': 'scheduled', 'conf': <psycopg2.extensions.Binary object at 0x7f85aa34d150>, 'data\_interval\_start': datetime.datetime(2022, 11, 5, 0, 0, tzinfo=Timezone('UTC')), 'data\_interval\_end': datetime.datetime(2022, 11, 6, 0, 0, tzinfo=Timezone('UTC')), 'last\_scheduling\_decision': datetime.datetime(2022, 11, 6, 4, 1, 3, 514930, tzinfo=Timezone('UTC')), 'dag\_hash': '825759430eed0eba5fbd23bf78b588af', 'log\_template\_id': 2}]

(Background on this error at: <https://sqlalche.me/e/14/gkpj);> 2313)

[2022-11-06, 04:03:43 UTC] {process\_utils.py:75} INFO - Process psutil.Process(pid=2313, status='terminated', exitcode=1, started='04:01:05') (2313) terminated with exit code 1

[2022-11-06, 04:09:33 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.analyze\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:09:33 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.analyze\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:09:33 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:09:33 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 04:09:33 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:09:33 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocCreateBatchOperator): analyze\_trade\_data> on 2022-11-05 00:00:00+00:00

[2022-11-06, 04:09:33 UTC] {standard\_task\_runner.py:52} INFO - Started process 2551 to run task

[2022-11-06, 04:09:33 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'analyze\_trade\_data', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '59', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmpdy\_313iq', '--error-file', '/tmp/tmpx7wbt5ab']

[2022-11-06, 04:09:33 UTC] {standard\_task\_runner.py:80} INFO - Job 59: Subtask analyze\_trade\_data

[2022-11-06, 04:09:33 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.analyze\_trade\_data scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 04:09:34 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=analyze\_trade\_data

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

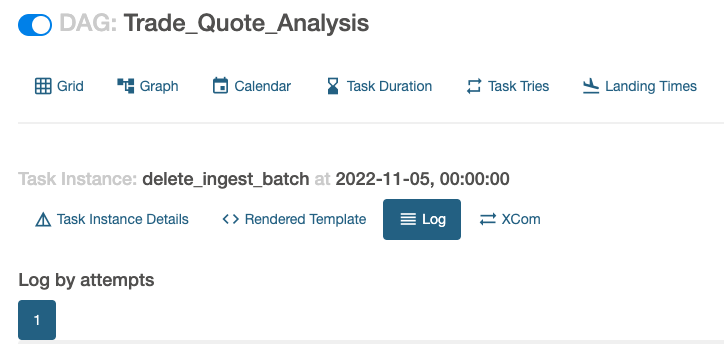
[2022-11-06, 04:09:34 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

[2022-11-06, 04:09:34 UTC] {dataproc.py:2022} INFO - Creating batch

[2022-11-06, 04:14:02 UTC] {dataproc.py:2039} INFO - Batch process-batch-create created

[2022-11-06, 04:14:02 UTC] {taskinstance.py:1416} INFO - Marking task as SUCCESS. dag\_id=Trade\_Quote\_Analysis, task\_id=analyze\_trade\_data, execution\_date=20221105T000000, start\_date=20221106T040933, end\_date=20221106T041402

[2022-11-06, 04:14:03 UTC] {local\_task\_job.py:156} INFO - Task exited with return code 0



\*\*\* Reading remote log from gs://us-east1-spark-airflow-comp-9c560139-bucket/logs/dag\_id=Trade\_Quote\_Analysis/run\_id=scheduled\_\_2022-11-05T00:00:00+00:00/task\_id=delete\_ingest\_batch/attempt=1.log.

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.delete\_ingest\_batch scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.delete\_ingest\_batch scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocDeleteBatchOperator): delete\_ingest\_batch> on 2022-11-05 00:00:00+00:00

[2022-11-06, 04:14:08 UTC] {standard\_task\_runner.py:52} INFO - Started process 2670 to run task

[2022-11-06, 04:14:08 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'delete\_ingest\_batch', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '61', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmps7fn7q6v', '--error-file', '/tmp/tmp81jnj592']

[2022-11-06, 04:14:08 UTC] {standard\_task\_runner.py:80} INFO - Job 61: Subtask delete\_ingest\_batch

[2022-11-06, 04:14:09 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.delete\_ingest\_batch scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 04:14:10 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=delete\_ingest\_batch

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

[2022-11-06, 04:14:10 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

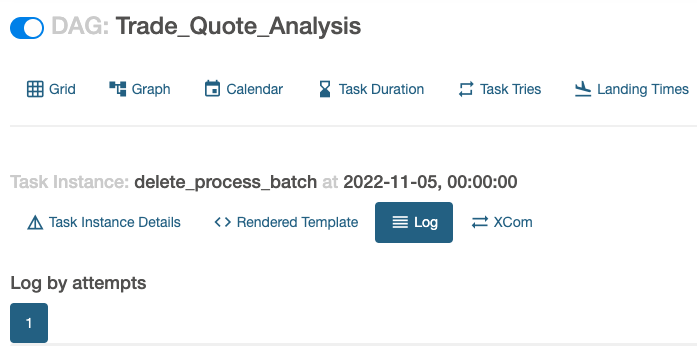
[2022-11-06, 04:14:10 UTC] {dataproc.py:2113} INFO - Deleting batch: ingest-batch-create

[2022-11-06, 04:14:10 UTC] {dataproc.py:2122} INFO - Batch deleted.

[2022-11-06, 04:14:10 UTC] {taskinstance.py:1416} INFO - Marking task as SUCCESS. dag\_id=Trade\_Quote\_Analysis, task\_id=delete\_ingest\_batch, execution\_date=20221105T000000, start\_date=20221106T041408, end\_date=20221106T041410

[2022-11-06, 04:14:10 UTC] {local\_task\_job.py:156} INFO - Task exited with return code 0

[2022-11-06, 04:14:10 UTC] {local\_task\_job.py:273} INFO - 0 downstream tasks scheduled from follow-on schedule check



\*\*\* Reading remote log from gs://us-east1-spark-airflow-comp-9c560139-bucket/logs/dag\_id=Trade\_Quote\_Analysis/run\_id=scheduled\_\_2022-11-05T00:00:00+00:00/task\_id=delete\_process\_batch/attempt=1.log.

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.delete\_process\_batch scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1180} INFO - Dependencies all met for <TaskInstance: Trade\_Quote\_Analysis.delete\_process\_batch scheduled\_\_2022-11-05T00:00:00+00:00 [queued]>

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1377} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1378} INFO - Starting attempt 1 of 3

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1379} INFO -

--------------------------------------------------------------------------------

[2022-11-06, 04:14:08 UTC] {taskinstance.py:1398} INFO - Executing <Task(DataprocDeleteBatchOperator): delete\_process\_batch> on 2022-11-05 00:00:00+00:00

[2022-11-06, 04:14:08 UTC] {standard\_task\_runner.py:52} INFO - Started process 2669 to run task

[2022-11-06, 04:14:08 UTC] {standard\_task\_runner.py:79} INFO - Running: ['airflow', 'tasks', 'run', 'Trade\_Quote\_Analysis', 'delete\_process\_batch', 'scheduled\_\_2022-11-05T00:00:00+00:00', '--job-id', '60', '--raw', '--subdir', 'DAGS\_FOLDER/trade\_analysis\_pipeline.py', '--cfg-path', '/tmp/tmpkw17wvlz', '--error-file', '/tmp/tmpv8z6nhmn']

[2022-11-06, 04:14:08 UTC] {standard\_task\_runner.py:80} INFO - Job 60: Subtask delete\_process\_batch

[2022-11-06, 04:14:09 UTC] {task\_command.py:375} INFO - Running <TaskInstance: Trade\_Quote\_Analysis.delete\_process\_batch scheduled\_\_2022-11-05T00:00:00+00:00 [running]> on host airflow-worker-trp9b

[2022-11-06, 04:14:09 UTC] {taskinstance.py:1591} INFO - Exporting the following env vars:

AIRFLOW\_CTX\_DAG\_OWNER=airflow

AIRFLOW\_CTX\_DAG\_ID=Trade\_Quote\_Analysis

AIRFLOW\_CTX\_TASK\_ID=delete\_process\_batch

AIRFLOW\_CTX\_EXECUTION\_DATE=2022-11-05T00:00:00+00:00

AIRFLOW\_CTX\_TRY\_NUMBER=1

AIRFLOW\_CTX\_DAG\_RUN\_ID=scheduled\_\_2022-11-05T00:00:00+00:00

[2022-11-06, 04:14:09 UTC] {base.py:68} INFO - Using connection ID 'google\_cloud\_default' for task execution.

[2022-11-06, 04:14:09 UTC] {dataproc.py:2113} INFO - Deleting batch: process-batch-create

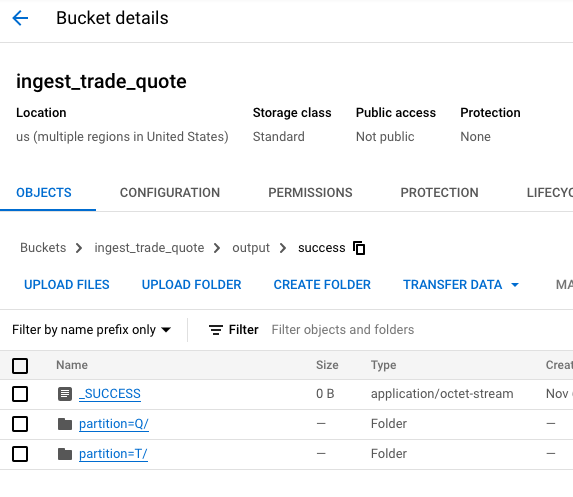
[2022-11-06, 04:14:10 UTC] {dataproc.py:2122} INFO - Batch deleted.

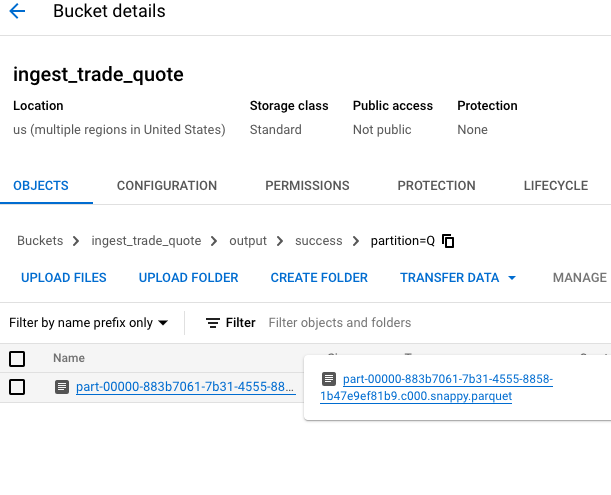
[2022-11-06, 04:14:10 UTC] {taskinstance.py:1416} INFO - Marking task as SUCCESS. dag\_id=Trade\_Quote\_Analysis, task\_id=delete\_process\_batch, execution\_date=20221105T000000, start\_date=20221106T041408, end\_date=20221106T041410

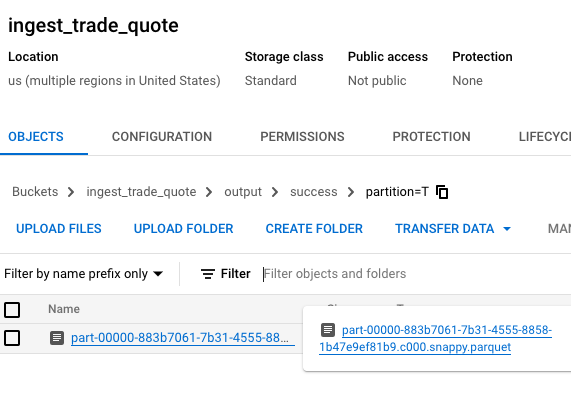
[2022-11-06, 04:14:10 UTC] {local\_task\_job.py:156} INFO - Task exited with return code 0

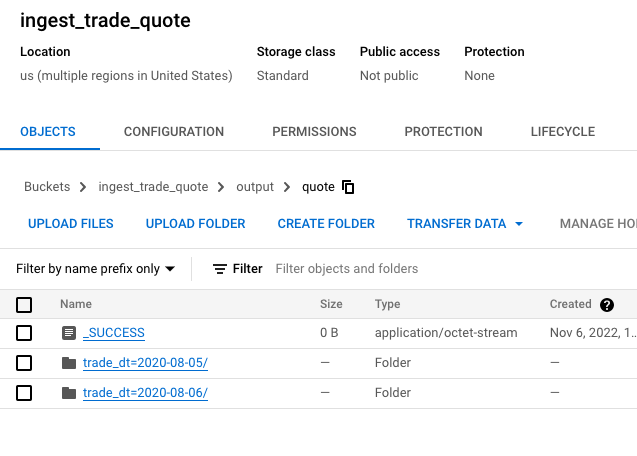
[2022-11-06, 04:14:10 UTC] {local\_task\_job.py:273} INFO - 0 downstream tasks scheduled from follow-on schedule check

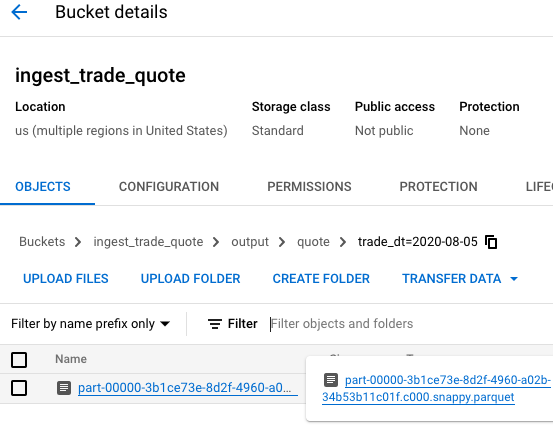
Data written by Spark in Cloud:

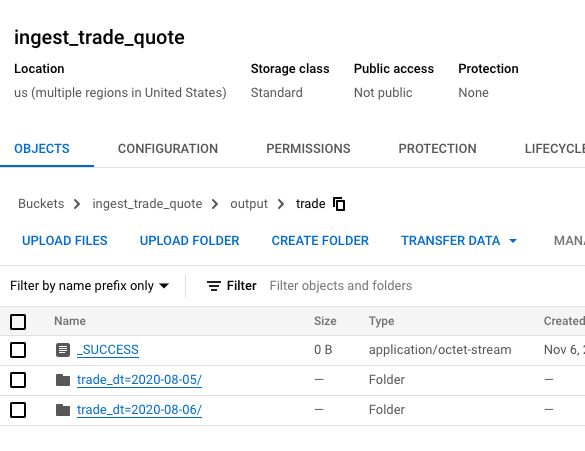


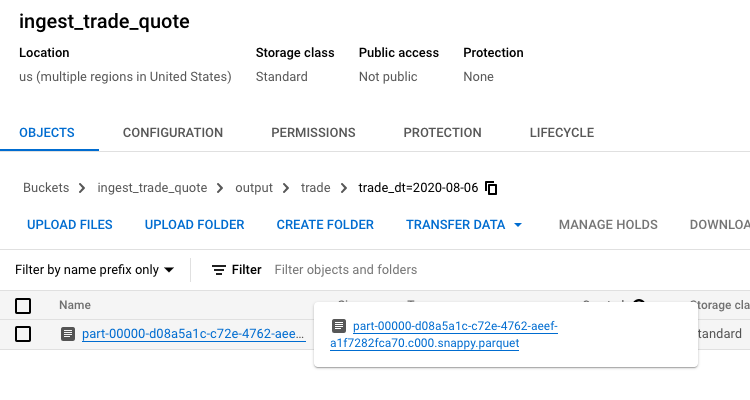


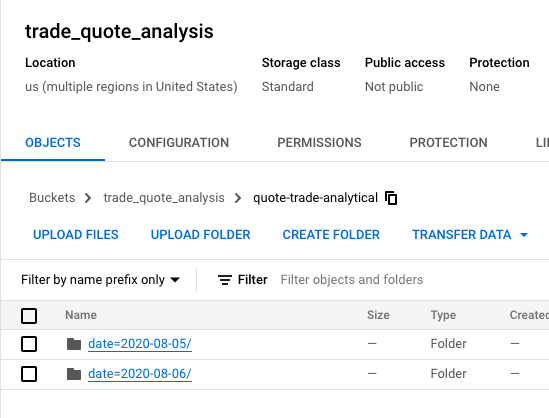


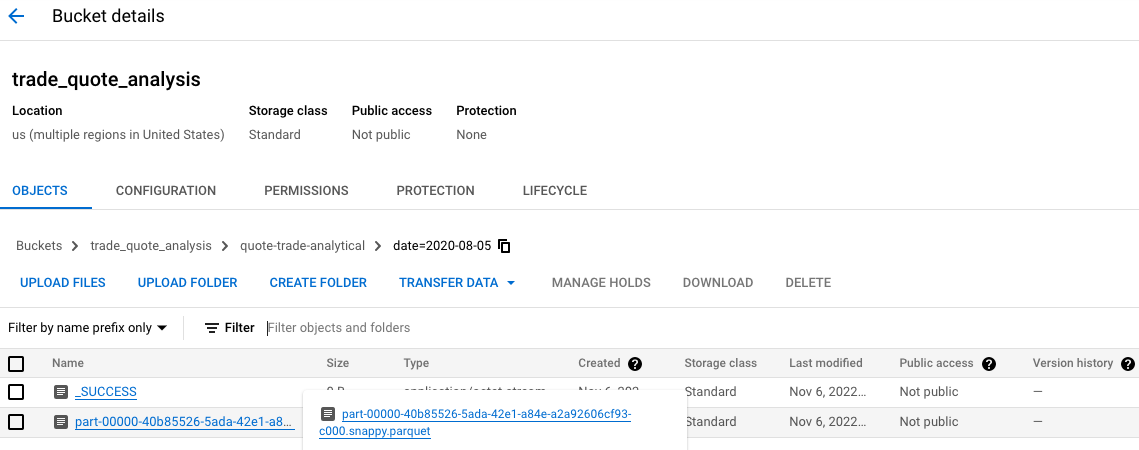




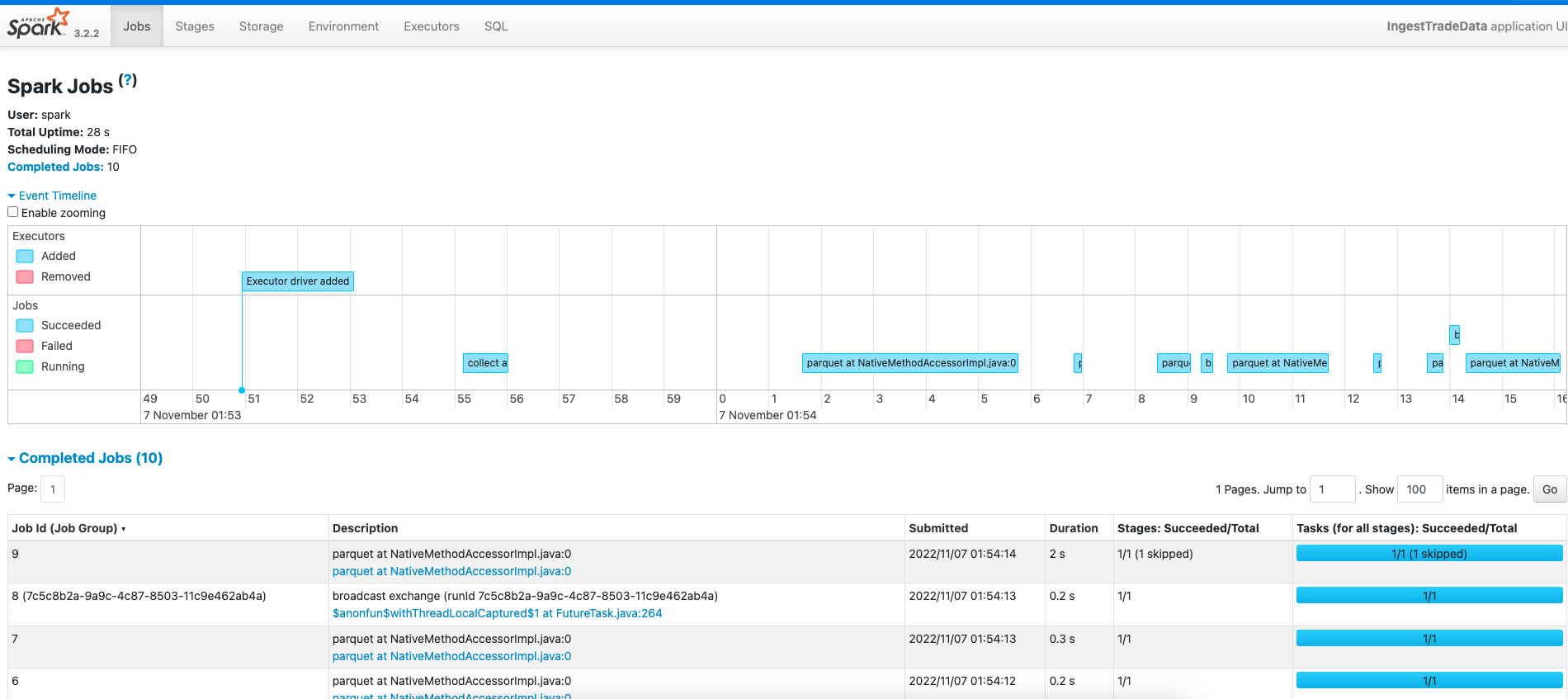


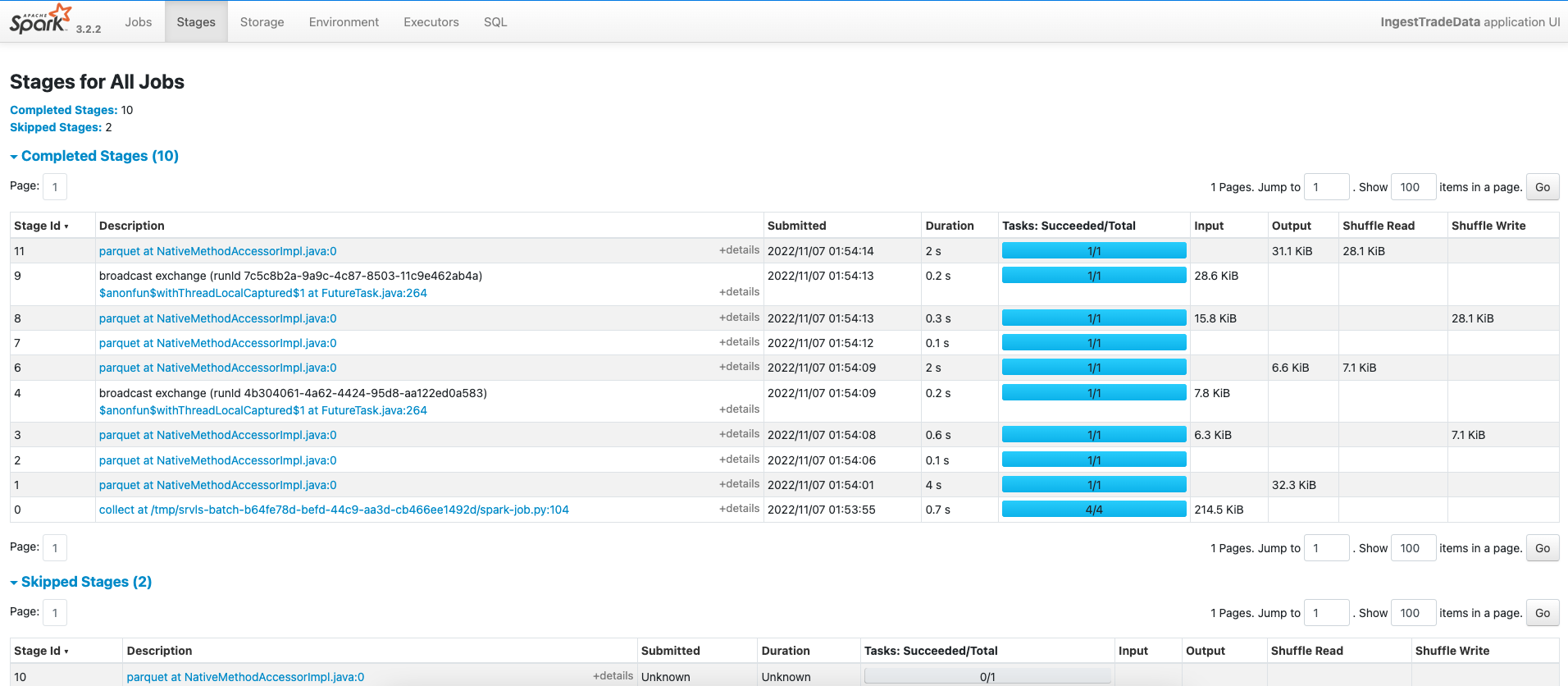


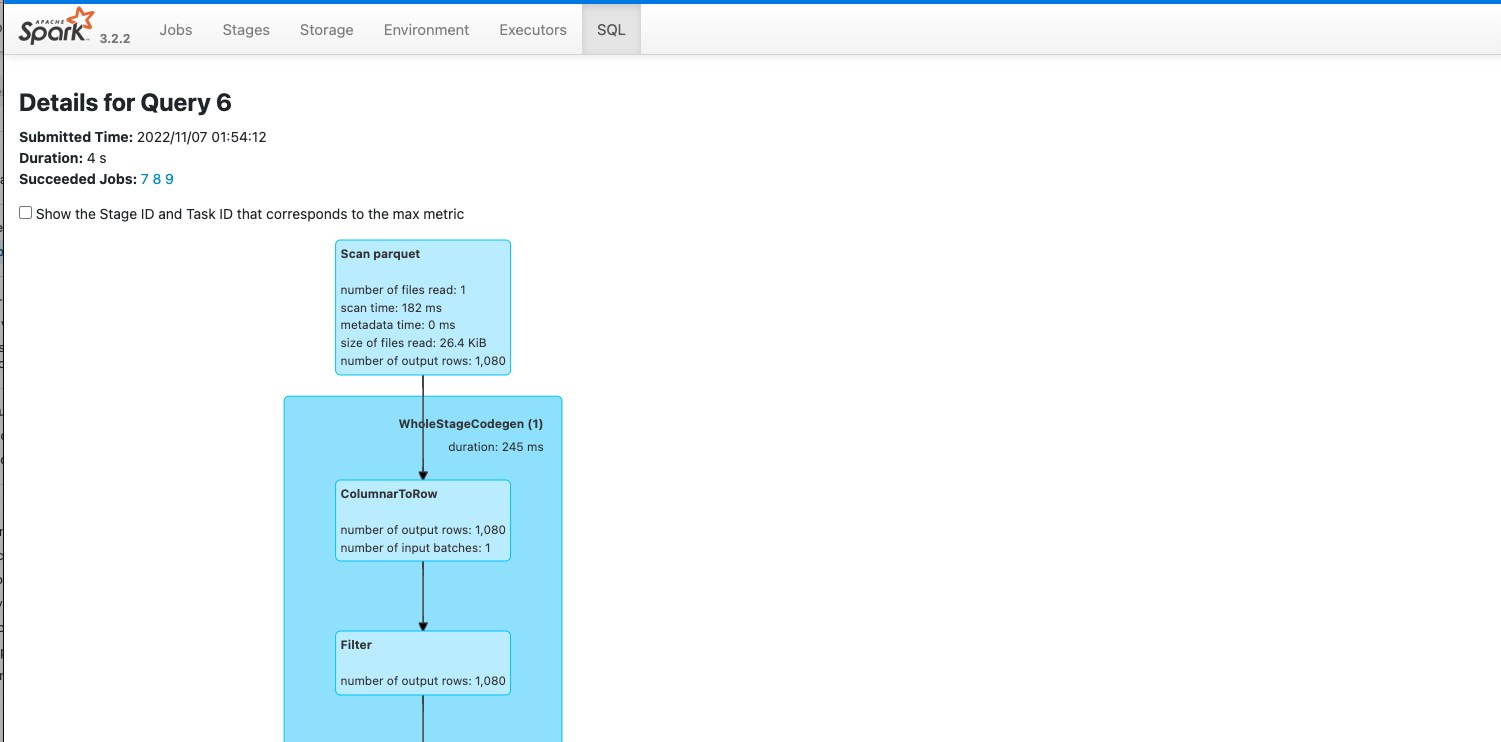


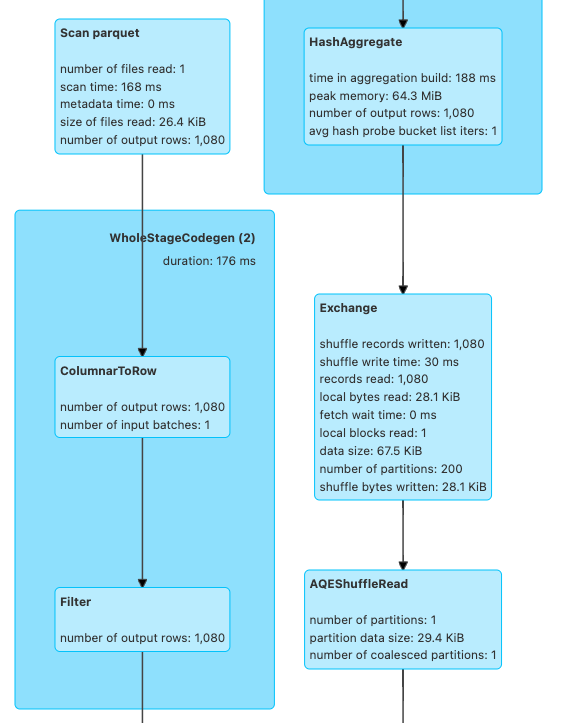


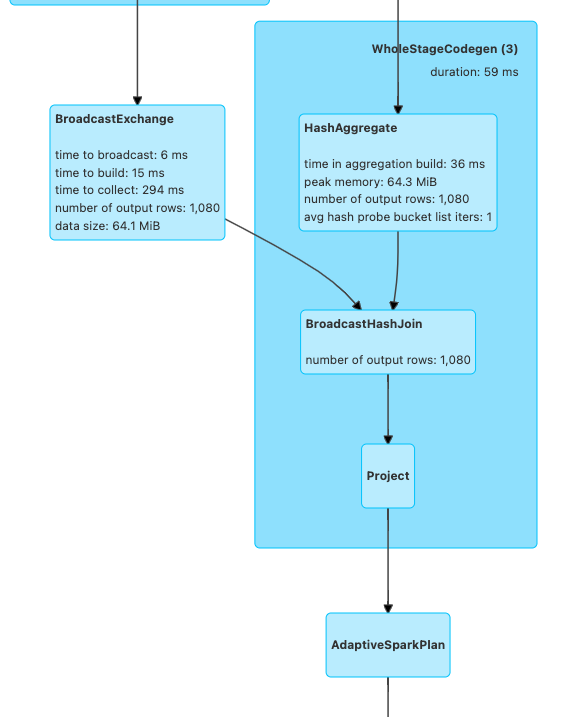
Sample Spark History Logs:

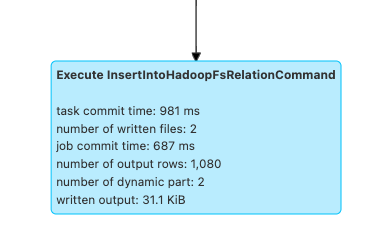












== Physical Plan ==

Execute InsertIntoHadoopFsRelationCommand (26)

+- AdaptiveSparkPlan (25)

+- == Final Plan ==

\* Project (15)

+- \* BroadcastHashJoin Inner BuildLeft (14)

:- BroadcastQueryStage (5)

: +- BroadcastExchange (4)

: +- \* Filter (3)

: +- \* ColumnarToRow (2)

: +- Scan parquet (1)

+- \* HashAggregate (13)

+- AQEShuffleRead (12)

+- ShuffleQueryStage (11)

+- Exchange (10)

+- \* HashAggregate (9)

+- \* Filter (8)

+- \* ColumnarToRow (7)

+- Scan parquet (6)

+- == Initial Plan ==

Project (24)

+- BroadcastHashJoin Inner BuildRight (23)

:- Filter (17)

: +- Scan parquet (16)

+- BroadcastExchange (22)

+- HashAggregate (21)

+- Exchange (20)

+- HashAggregate (19)

+- Filter (18)

+- Scan parquet (6)

(1) Scan parquet

Output [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127]

Batched: true

Location: InMemoryFileIndex [gs://ingest\_trade\_quote/output/success/partition=Q]

PushedFilters: [IsNotNull(trade\_dt), IsNotNull(symbol), IsNotNull(exchange), IsNotNull(event\_tm), IsNotNull(event\_seq\_nb)]

ReadSchema: struct<trade\_dt:date,rec\_type:string,symbol:string,exchange:string,event\_tm:timestamp,event\_seq\_nb:int,arrival\_tm:timestamp,trade\_pr:decimal(10,2),bid\_pr:decimal(10,2),bid\_size:int,ask\_pr:decimal(10,0),ask\_size:int>

(2) ColumnarToRow [codegen id : 2]

Input [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127]

(3) Filter [codegen id : 2]

Input [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127]

Condition : ((((isnotnull(trade\_dt#116) AND isnotnull(symbol#118)) AND isnotnull(exchange#119)) AND isnotnull(event\_tm#120)) AND isnotnull(event\_seq\_nb#121))

(4) BroadcastExchange

Input [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127]

Arguments: HashedRelationBroadcastMode(List(input[0, date, false], input[2, string, false], input[3, string, false], input[4, timestamp, false], input[5, int, false]),false), [plan\_id=271]

(5) BroadcastQueryStage

Output [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127]

Arguments: 1

(6) Scan parquet

Output [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Batched: true

Location: InMemoryFileIndex [gs://ingest\_trade\_quote/output/success/partition=Q]

PushedFilters: [IsNotNull(trade\_dt), IsNotNull(symbol), IsNotNull(exchange), IsNotNull(event\_tm), IsNotNull(event\_seq\_nb)]

ReadSchema: struct<trade\_dt:date,symbol:string,exchange:string,event\_tm:timestamp,event\_seq\_nb:int>

(7) ColumnarToRow [codegen id : 1]

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

(8) Filter [codegen id : 1]

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Condition : ((((isnotnull(trade\_dt#172) AND isnotnull(symbol#174)) AND isnotnull(exchange#175)) AND isnotnull(event\_tm#176)) AND isnotnull(event\_seq\_nb#177))

(9) HashAggregate [codegen id : 1]

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Keys [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Functions: []

Aggregate Attributes: []

Results [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

(10) Exchange

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Arguments: hashpartitioning(trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177, 200), ENSURE\_REQUIREMENTS, [plan\_id=228]

(11) ShuffleQueryStage

Output [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Arguments: 0

(12) AQEShuffleRead

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Arguments: coalesced

(13) HashAggregate

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Keys [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Functions: []

Aggregate Attributes: []

Results [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

(14) BroadcastHashJoin [codegen id : 3]

Left keys [5]: [trade\_dt#116, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121]

Right keys [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Join condition: None

(15) Project [codegen id : 3]

Output [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127, trade\_pr#123]

Input [17]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127, trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

(16) Scan parquet

Output [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127]

Batched: true

Location: InMemoryFileIndex [gs://ingest\_trade\_quote/output/success/partition=Q]

PushedFilters: [IsNotNull(trade\_dt), IsNotNull(symbol), IsNotNull(exchange), IsNotNull(event\_tm), IsNotNull(event\_seq\_nb)]

ReadSchema: struct<trade\_dt:date,rec\_type:string,symbol:string,exchange:string,event\_tm:timestamp,event\_seq\_nb:int,arrival\_tm:timestamp,trade\_pr:decimal(10,2),bid\_pr:decimal(10,2),bid\_size:int,ask\_pr:decimal(10,0),ask\_size:int>

(17) Filter

Input [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127]

Condition : ((((isnotnull(trade\_dt#116) AND isnotnull(symbol#118)) AND isnotnull(exchange#119)) AND isnotnull(event\_tm#120)) AND isnotnull(event\_seq\_nb#121))

(18) Filter

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Condition : ((((isnotnull(trade\_dt#172) AND isnotnull(symbol#174)) AND isnotnull(exchange#175)) AND isnotnull(event\_tm#176)) AND isnotnull(event\_seq\_nb#177))

(19) HashAggregate

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Keys [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Functions: []

Aggregate Attributes: []

Results [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

(20) Exchange

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Arguments: hashpartitioning(trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177, 200), ENSURE\_REQUIREMENTS, [plan\_id=201]

(21) HashAggregate

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Keys [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Functions: []

Aggregate Attributes: []

Results [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

(22) BroadcastExchange

Input [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Arguments: HashedRelationBroadcastMode(List(input[0, date, true], input[1, string, true], input[2, string, true], input[3, timestamp, true], input[4, int, true]),false), [plan\_id=204]

(23) BroadcastHashJoin

Left keys [5]: [trade\_dt#116, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121]

Right keys [5]: [trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

Join condition: None

(24) Project

Output [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127, trade\_pr#123]

Input [17]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, trade\_pr#123, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127, trade\_dt#172, symbol#174, exchange#175, event\_tm#176, event\_seq\_nb#177]

(25) AdaptiveSparkPlan

Output [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127, trade\_pr#123]

Arguments: isFinalPlan=true

(26) Execute InsertIntoHadoopFsRelationCommand

Input [12]: [trade\_dt#116, rec\_type#117, symbol#118, exchange#119, event\_tm#120, event\_seq\_nb#121, arrival\_tm#122, bid\_pr#124, bid\_size#125, ask\_pr#126, ask\_size#127, trade\_pr#123]

Arguments: gs://ingest\_trade\_quote/output/quote, false, [trade\_dt#116], Parquet, [\_\_partition\_columns=["trade\_dt"], path=gs://ingest\_trade\_quote/output/quote], Overwrite, [trade\_dt, rec\_type, symbol, exchange, event\_tm, event\_seq\_nb, arrival\_tm, bid\_pr, bid\_size, ask\_pr, ask\_size, trade\_pr]