**Comparing results of Partitioning**

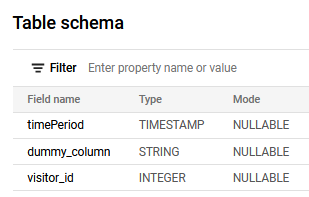
1. **Uploading CSV file and applying Partitioning**
2. T[able partitioned by Day on a timestamp field](https://stackoverflow.com/questions/62388473/query-bigquery-table-partitioned-by-day-on-a-timestamp-field)

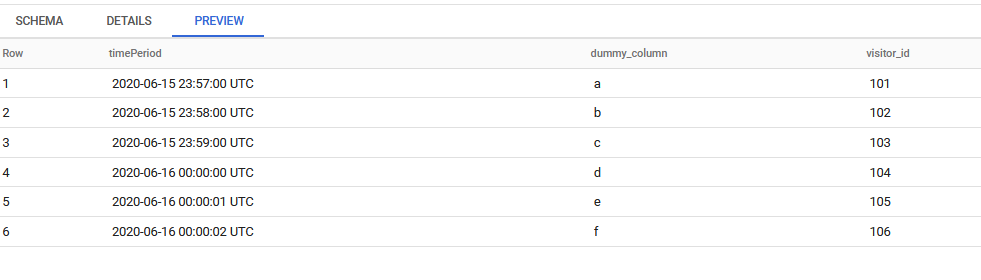
|  |  |
| --- | --- |
| Table type | partitioned |
| Partitioned by | Day |
| Partitioned on field | Time |

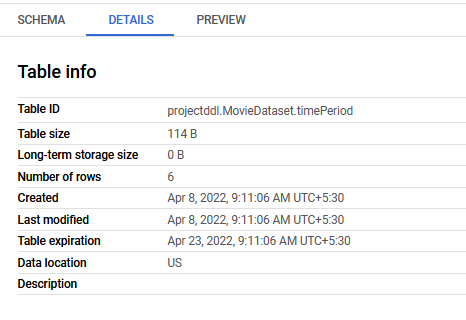
**Condition**: column(timestamp data type) …partition option….Partitioning type(day)

Table is partitioned on \_time but it is partitioned by Day

Use the CSV file( timePeriod.csv)







Otherwise we can use the below syntax to create a table exactly like timePeriod.CSV file

Query statement for partition by ***day***

create table dataset.table\_name(timePeriod timestamp, dummy\_column string, visitor\_id integer) partition by date(\_time)

as select timestamp '2020-06-15 23:57:00 UTC' as \_time, "a" as dummy\_column union all

select timestamp '2020-06-15 23:58:00 UTC' as \_time, "b" as dummy\_column union all

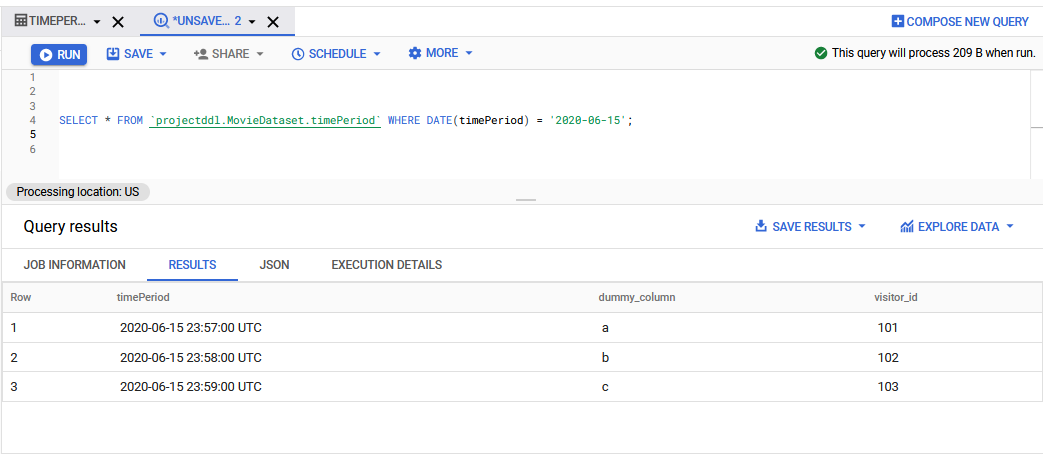
select timestamp '2020-06-15 23:59:00 UTC' as \_time, "c" as dummy\_column union all

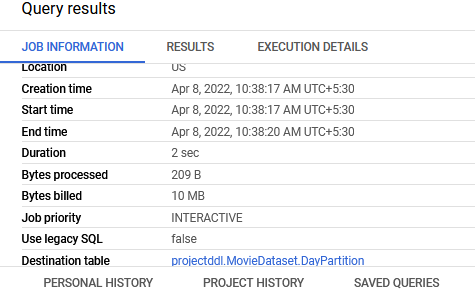
select timestamp '2020-06-16 00:00:00 UTC' as \_time, "d" as dummy\_column union all

select timestamp '2020-06-16 00:00:01 UTC' as \_time, "e" as dummy\_column union all

select timestamp '2020-06-16 00:00:02 UTC' as \_time, "f" as dummy\_column

SELECT \* FROM `projectddl.MovieDataset.timePeriod` WHERE DATE(timePeriod) = '2020-06-15';





The query takes time duration of 2 sec and bytes processed 209B

{CREATE TABLE | CREATE TABLE IF NOT EXISTS | CREATE OR REPLACE TABLE}

[[project\_name.]dataset\_name.]table\_name

[(

column\_name column\_schema[, ...]

)]

[PARTITION BY partition\_expression]

[CLUSTER BY clustering\_column\_list]

[OPTIONS(table\_option\_list)]

[AS query\_statement]

CREATE TABLE IF NOT EXISTS

  `projectddl.MovieDataset.DayPartition`

PARTITION BY

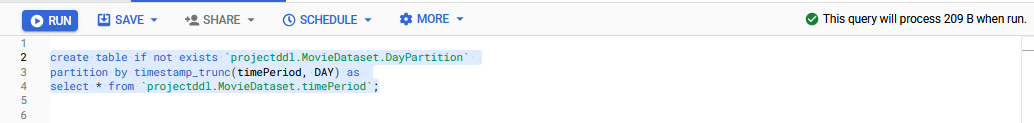
  TIMESTAMP\_TRUNC(timePeriod, DAY) AS

SELECT

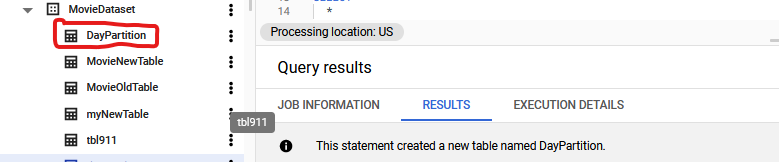
  \*

FROM

  `projectddl.MovieDataset.timePeriod`;



**Result**

****

Use *partition table* to retrieve information from timePeriod column.

SELECT

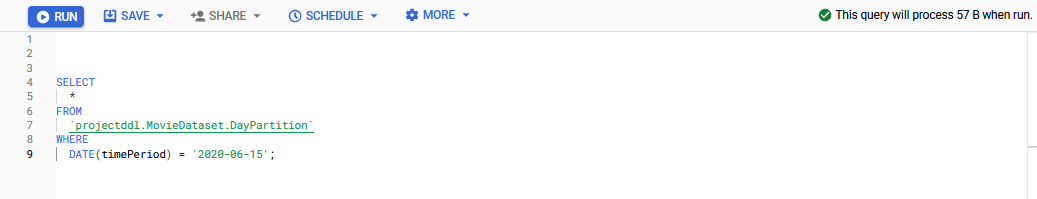
  \*

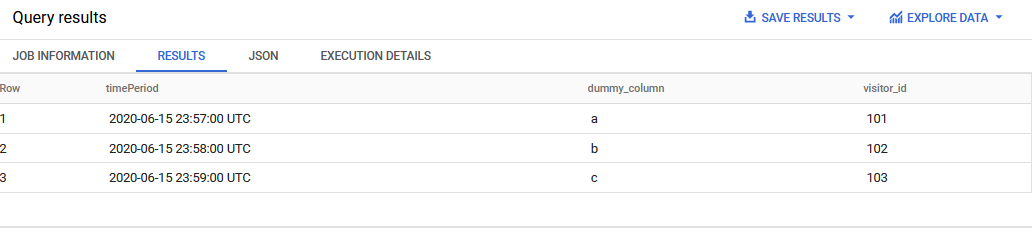
FROM

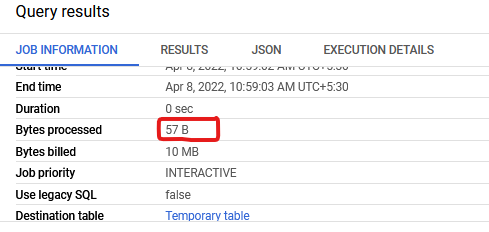
  `projectddl.MovieDataset.DayPartition`

WHERE

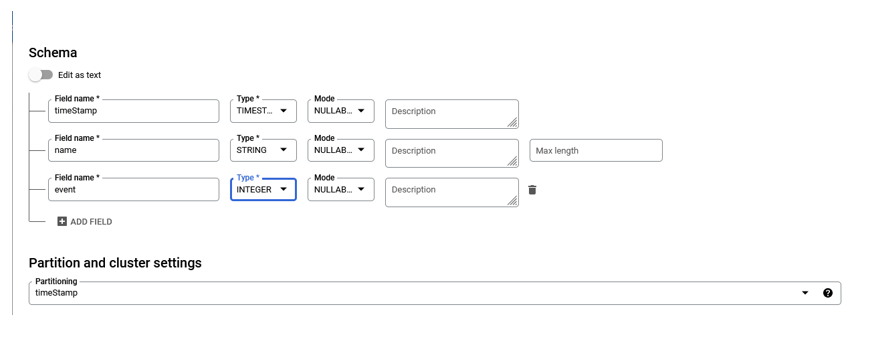
  DATE(timePeriod) = '2020-06-15';







1. **Creating Empty table for Partitioning**
2. **Write the query statement and apply partitioning**

****



Insert values into **EMPTY TABLE**

INSERT INTO

  `projectddl.MovieDataset.DayPartition\_EmptyTable`

VALUES

  ('2020-06-15 23:58:00 UTC','a',101),

  ('2020-06-15 23:58:00 UTC','b',102),

  ('2020-06-15 23:59:00 UTC','c',103),

  ('2020-06-16 00:00:00 UTC','d',104),

  ('2020-06-16 00:00:01 UTC','e',105),

  ('2020-06-16 00:00:02 UTC','f',106),

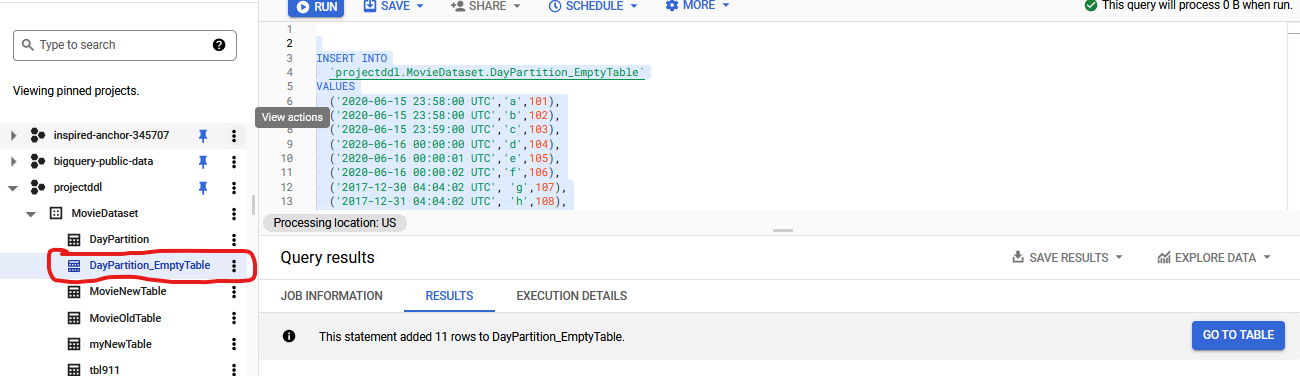
  ('2017-12-30 04:04:02 UTC', 'g',107),

  ('2017-12-31 04:04:02 UTC', 'h',108),

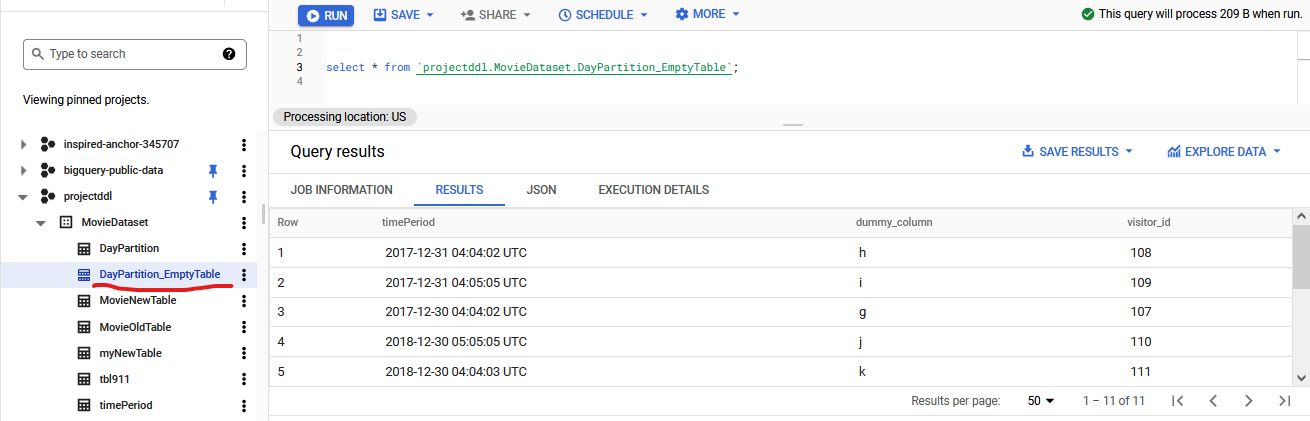
  ('2017-12-31 04:05:05 UTC', 'i',109),

  ('2018-12-30 05:05:05 UTC', 'j',110),

  ('2018-12-30 04:04:03 UTC', 'k',111);



select \* from `projectddl.MovieDataset.DayPartition\_EmptyTable`;



Use query statement to select result from *DayPartition\_EmptyTable* on ‘**Day**’ basis.

SELECT

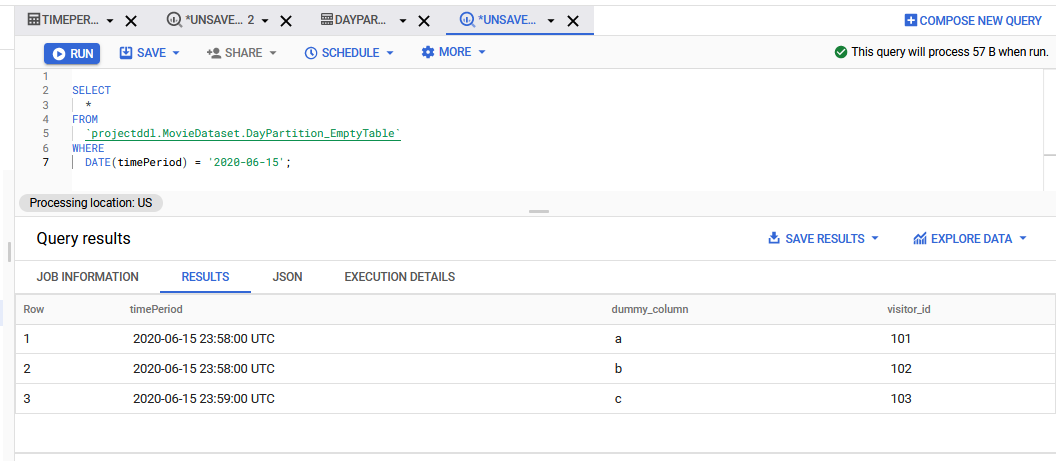
  \*

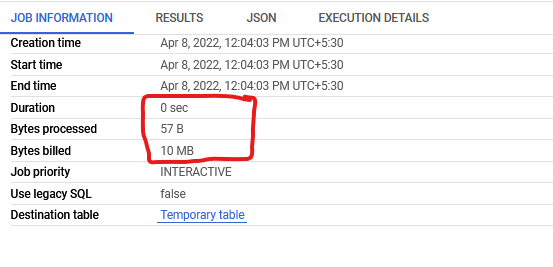
FROM

  `projectddl.MovieDataset.DayPartition\_EmptyTable`

WHERE

  DATE(timePeriod) = '2020-06-15';





1. **Without using any query statement( as we have already chosen the partitioning options(***hour,daya,month,year***) and apply partitioning.**

SELECT

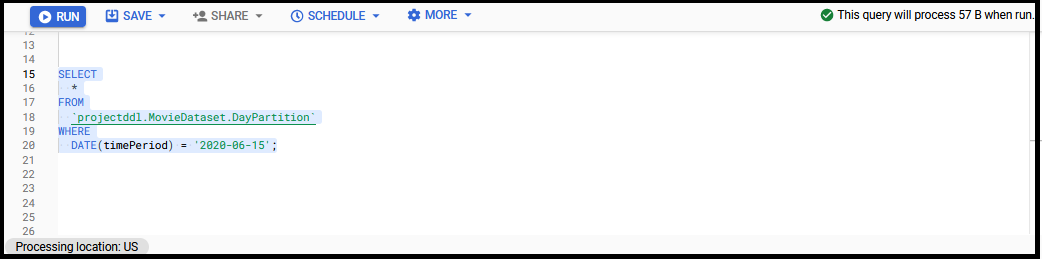
  \*

FROM

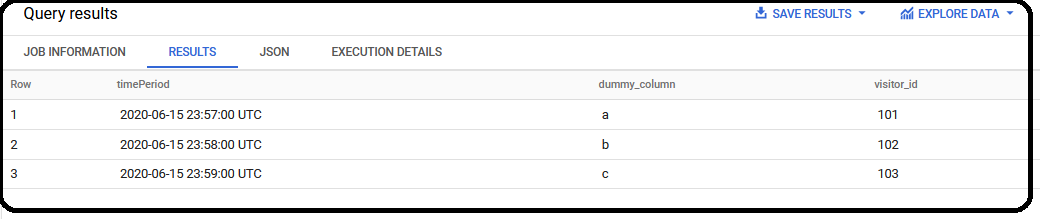
  `projectddl.MovieDataset.DayPartition`

WHERE

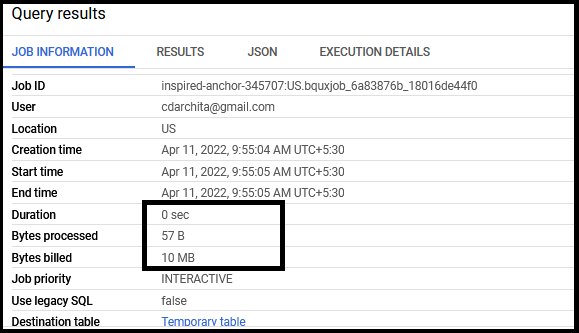
  DATE(timePeriod) = '2020-06-15';

****

**RESULT**



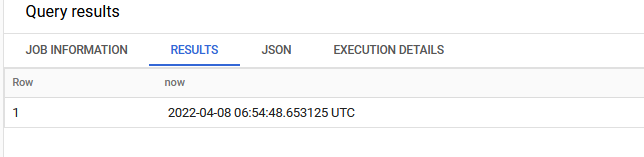
**Details about the time and bytes taken by the query**

****

|  |  |  |
| --- | --- | --- |
| Time taken by BigQuery to process | Uploading an External table | Creating an Empty table(columns with data  type) |
|  | Without partition(2 sec, 209B) | 0 sec , Bytes processed:57B, bytes billed:10MB |
|  | With partition(0 sec , 57B) |  |
|  |  |  |

1. T[able partitioned by Hour on a timestamp field](https://stackoverflow.com/questions/62388473/query-bigquery-table-partitioned-by-day-on-a-timestamp-field)

SELECT CURRENT\_TIMESTAMP() as now;



**TIMESTAMP\_SUB**

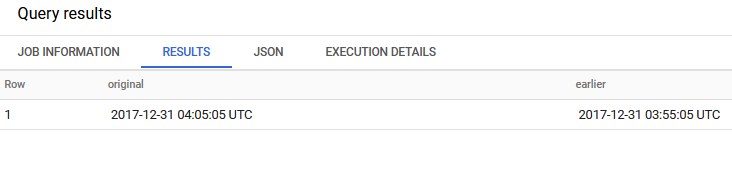
TIMESTAMP\_SUB supports the following values for date\_part:

* MICROSECOND
* MILLISECOND
* SECOND
* MINUTE
* HOUR. Equivalent to 60 MINUTEs.
* DAY. Equivalent to 24 HOURs.

SELECT

  TIMESTAMP("2017-12-31 04:05:05 UTC") AS original,

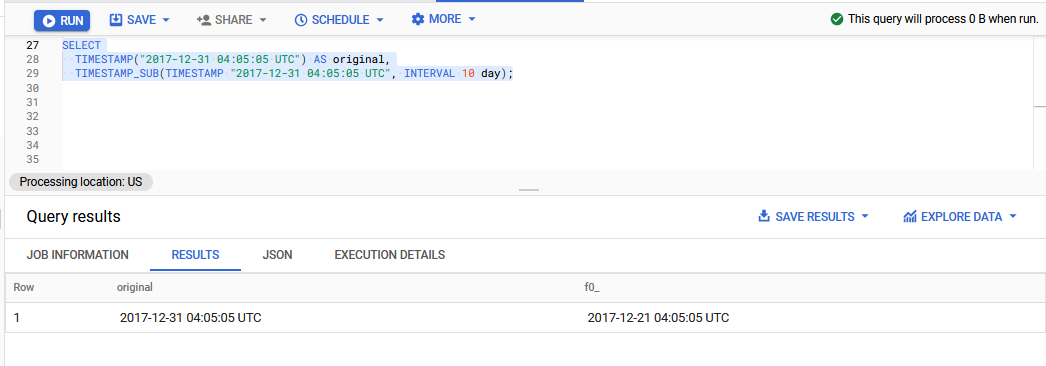
  TIMESTAMP\_SUB (TIMESTAMP "2017-12-31 04:05:05 UTC", INTERVAL 10 MINUTE) AS earlier;



SELECT

  TIMESTAMP("2017-12-31 04:05:05 UTC") AS original,

  TIMESTAMP\_SUB(TIMESTAMP "2017-12-31 04:05:05 UTC", INTERVAL 10 day);



CREATE TABLE IF NOT EXISTS

  `projectddl.MovieDataset.HourPartition2`

PARTITION BY

  TIMESTAMP\_TRUNC(timePeriod, HOUR) AS

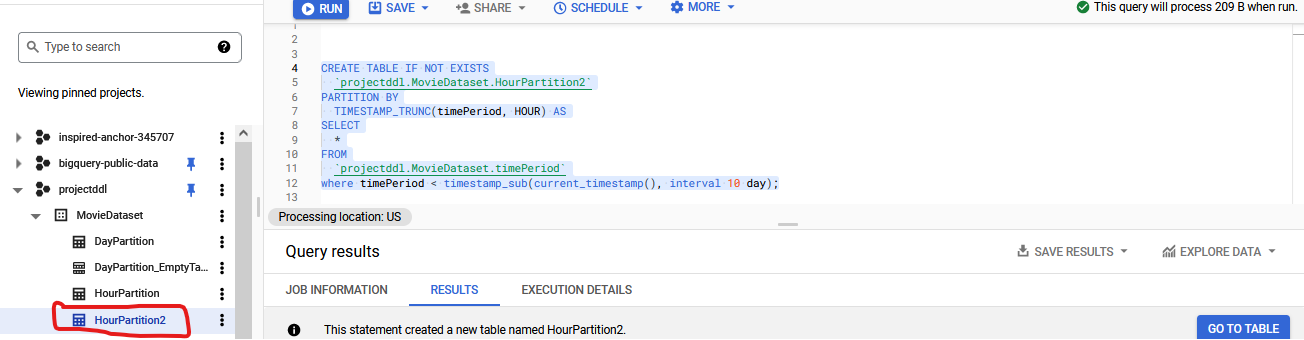
SELECT

  \*

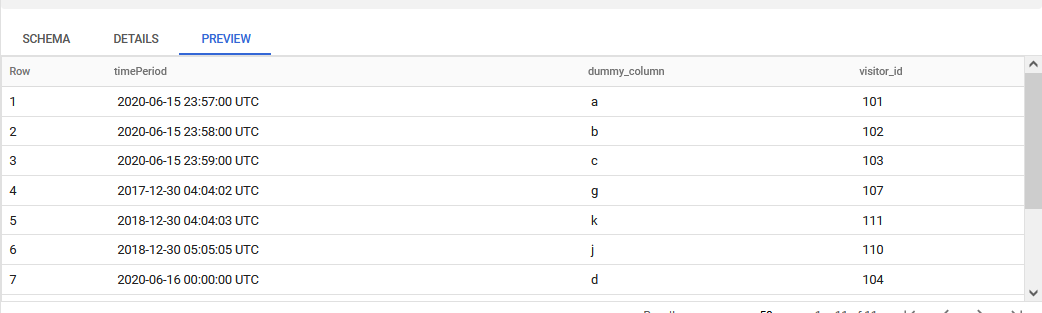
FROM

  `projectddl.MovieDataset.timePeriod`

where timePeriod < timestamp\_sub(current\_timestamp(), interval 10 day);



**Result**

****

**Searching for a range of Hours**

SELECT

  \*

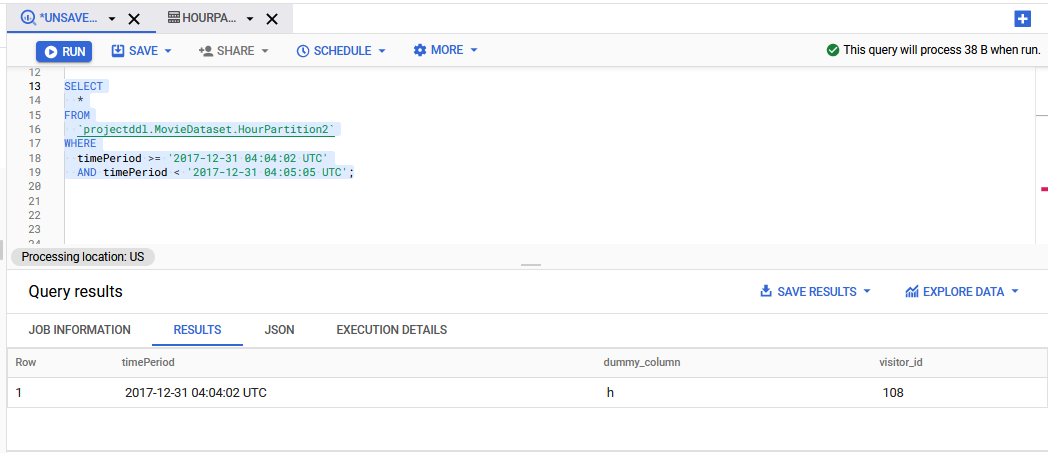
FROM

  `projectddl.MovieDataset.HourPartition2`

WHERE

  timePeriod >= '2017-12-31 04:04:02 UTC'

  AND timePeriod < '2017-12-31 04:05:05 UTC';

****

**Original table values**

('2020-06-15 23:58:00 UTC','a',101),

('2020-06-15 23:58:00 UTC','b',102),

('2020-06-15 23:59:00 UTC','c',103),

('2020-06-16 00:00:00 UTC','d',104),

('2020-06-16 00:00:01 UTC','e',105),

('2020-06-16 00:00:02 UTC','f',106),

('2017-12-30 04:04:02 UTC', 'g',107),

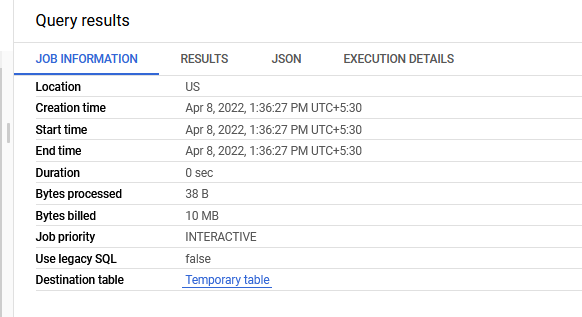
('2017-12-31 04:04:02 UTC', 'h',108),

('2017-12-31 04:05:05 UTC', 'i',109),

('2018-12-30 05:05:05 UTC', 'j',110),

('2018-12-30 04:04:03 UTC', 'k',111);

Table DETAILS



We can check the result for Hour based bigquery partitioning also by using same method (create an Empty table and choose the option by **hour**).

**END**