Source - Web Server

Frequency - Real time / Batch

Target - HDFS

We will be considering here Batch to execute the project.

Dataset has been loaded to mypath. You can load to your own path and give in location while creating table.

================= CREATE EXTERNAL TABLE QueryLogs

CREATE EXTERNAL TABLE QueryLogs

(

querytime timestamp,

query varchar(255)

)

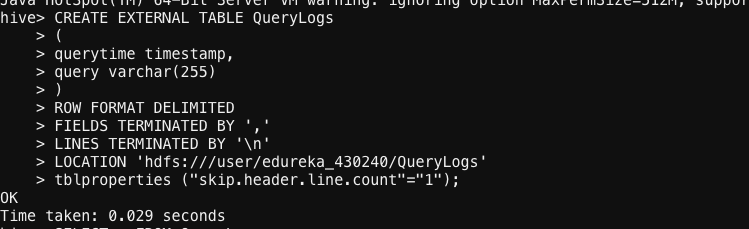
ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

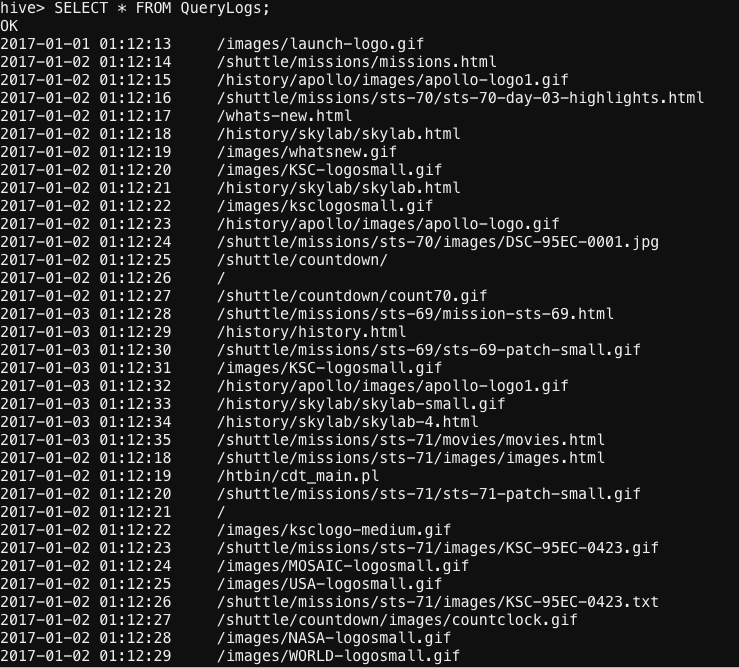
LINES TERMINATED BY '\n'

LOCATION 'hdfs:///user/edureka\_430240/QueryLogs'

tblproperties ("skip.header.line.count"="1");



=====================CHECKING RESULTS



====================== Query To Get Results

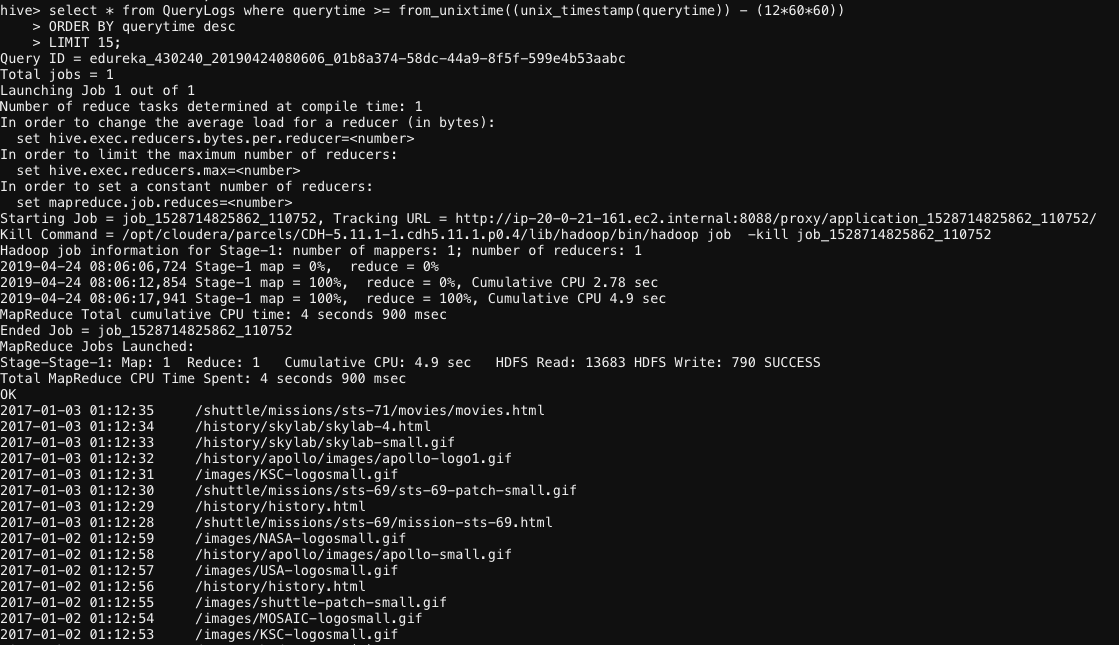
SELECT \* FROM QueryLogs

WHERE

querytime >= from\_unixtime((unix\_timestamp(querytime)) - (12\*60\*60))

ORDER BY querytime desc

LIMIT 15;



======================= ALternative Query to Get Results

SELECT \*

FROM (

SELECT query, RANK() OVER (ORDER BY querytime) query\_time\_rank

FROM query\_logs

WHERE querytime BETWEEN (SELECT min(querytime) FROM query\_logs) AND

(SELECT FROM\_UNIXTIME(UNIX\_TIMESTAMP(min(querytime)) +

(12\*60\*60)) FROM query\_logs)

) a

WHERE query\_time\_rank <= 15;