sublime 으로 하자나 정리가 안되니 ggwp monday night gcp study 퇴근후 공부 ㅎㅎ

anaconda python distribution doesn't work with gcp sdk so have to install separate python distribution…. Took over to fix this…

**• 2021-05-03: section2 loading data**

-- set path so that cmd can pick up schema definition in json file from local

cd C:/Users/Kim Jungyoon/Documents/2.study/gcp-de/

-- ls buckets (<https://cloud.google.com/storage/docs/gsutil/commands/ls>)

gsutil ls gs://test-etl

-- table created through web UI but this is the command?

Load gs://test-etl/2019-04-27.json to gcp-etl-prac:data\_test.flights\_temp

## 윈도우 쉘이라고 \ 안먹히는거 실화 ㅋㅋㅋㅋㅋㅋㅋ

## bq mk -t --schema schema.json data\_test.flight\_delays\_json\_non\_partitioned

-- create empty schema from a json file

bq mk -t --schema schema.json ^

data\_test.flight\_delays\_json\_non\_partitioned ## &&

-- load data into table

bq load --source\_format=NEWLINE\_DELIMITED\_JSON ^

data\_test.flight\_delays\_json\_non\_partitioned ^

gs://test-etl/2019-04-27.json

-- otherwise auto detect/define schema from the input

bq load --source\_format=NEWLINE\_DELIMITED\_JSON --autodetect ^

data\_analysis.flight\_delays\_json\_non\_partitioned ^

gs://test-etl/2019-04-27.json

required vs nullable on the json schema

csv file basically same

Partitioning data

<https://cloud.google.com/bigquery/docs/partitioned-tables?_ga=2.203658896.-1095196622.1620094460>)

gcp billing based on data scan so partition table e.g. based on column (date-column) or partition based on ingestion time, not part of the table but will add “createddate” like column

-- cmd for creating table with partition on date and loading data

bq mk -t --schema schema.json --time\_partitioning\_field flight\_date data\_test.flight\_delays\_partitioned

bq load --source\_format=NEWLINE\_DELIMITED\_JSON data\_test.flight\_delays\_partitioned gs://test-etl/2019-04-27.json

bq load --source\_format=NEWLINE\_DELIMITED\_JSON data\_test.flight\_delays\_partitioned ^

gs://test-etl/2019-04-28.json

++ can also query with bq? Doesn’t work? Because word “” format… lol

bq query --use\_legacy\_sql=false "SELECT \* FROM `data\_test.flight\_delays\_partitioned` WHERE flight\_date = "2019-04-27" LIMIT 10"

I guess this can use programmatically to load to table, query report and so on ㅇㅅㅇ DE stuff hehe

**2021-05-04**