BigQuery에서 Google Analytics 데이터 쿼리하기

SQL 을 배워보겠습니다!

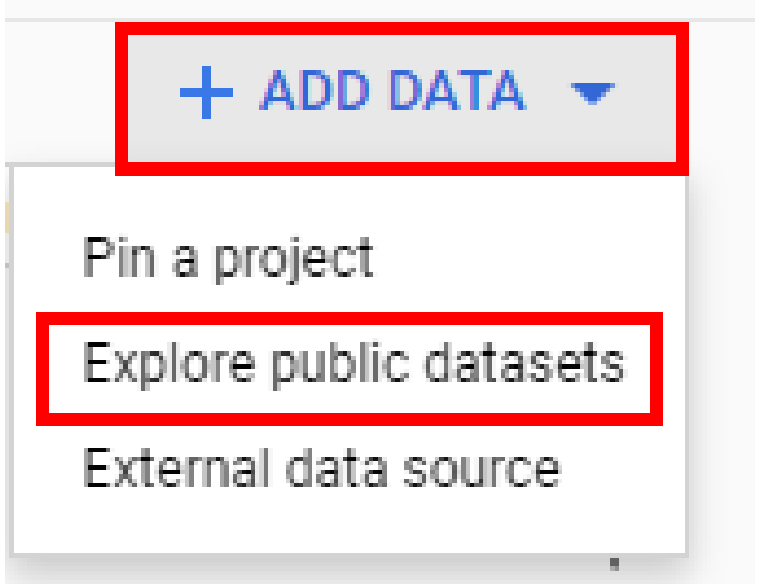
- 여러 종류의 DB 마다 sql 문법이 조금씩 다르지만 다 배울 필요 없음. (영어로 비유하자면~~~)

- DBA가 될 게 아니니 실용적으로 어떻게 사용하는지에 초점을 맞춰서 실제 업무에서 쓰는 문법들 위주로 배워봅시다.

Mission :

Google Analytics 퍼블릭 데이터 추가하기,
GA 테이블 특징 : 파티션(partition) 이해하기

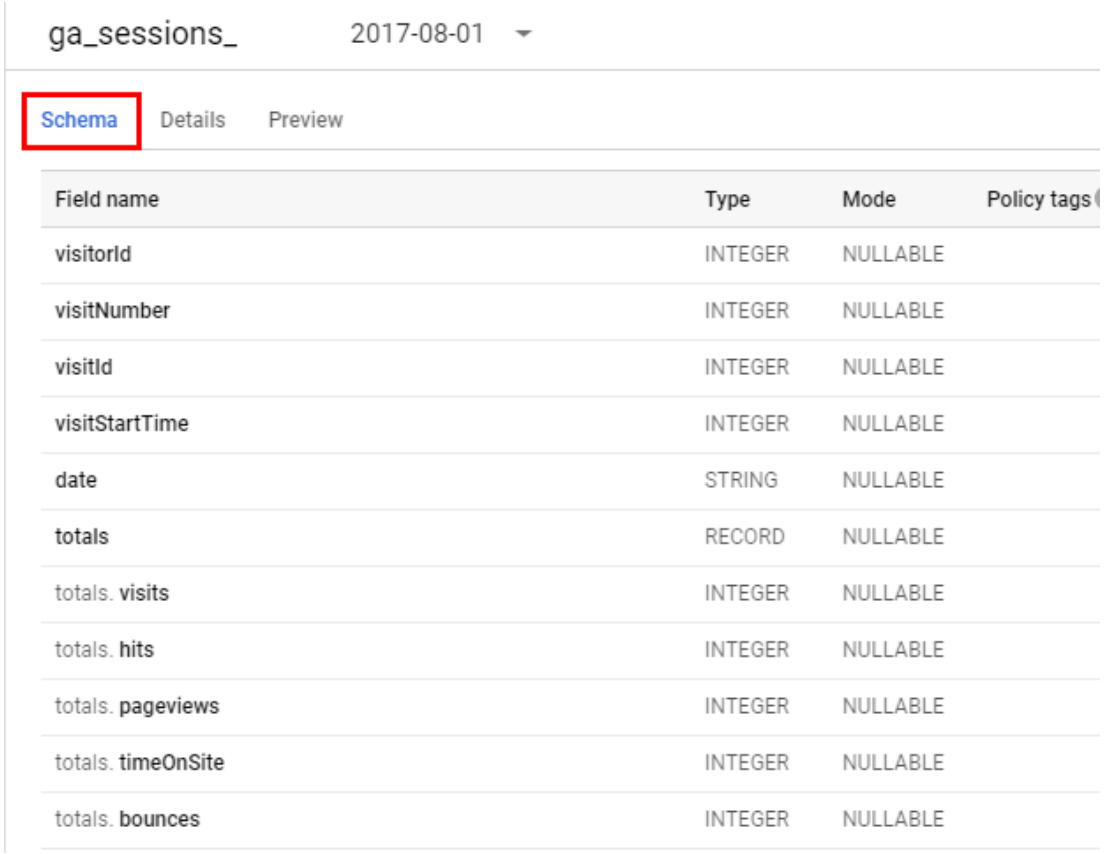
- Hint :



Mission :

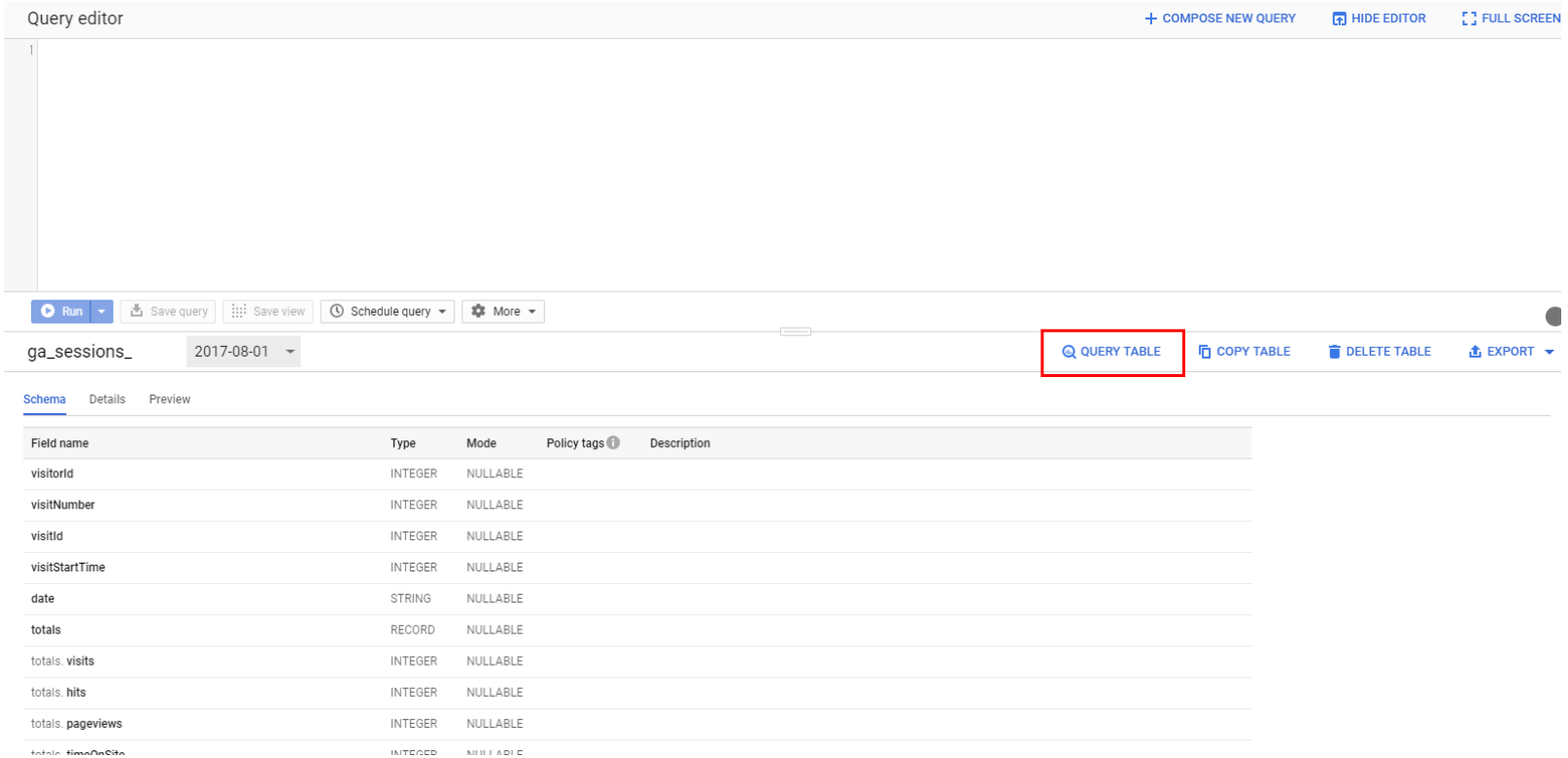
테이블 스키마, 상세정보, 데이터 조회하기

- Hint :



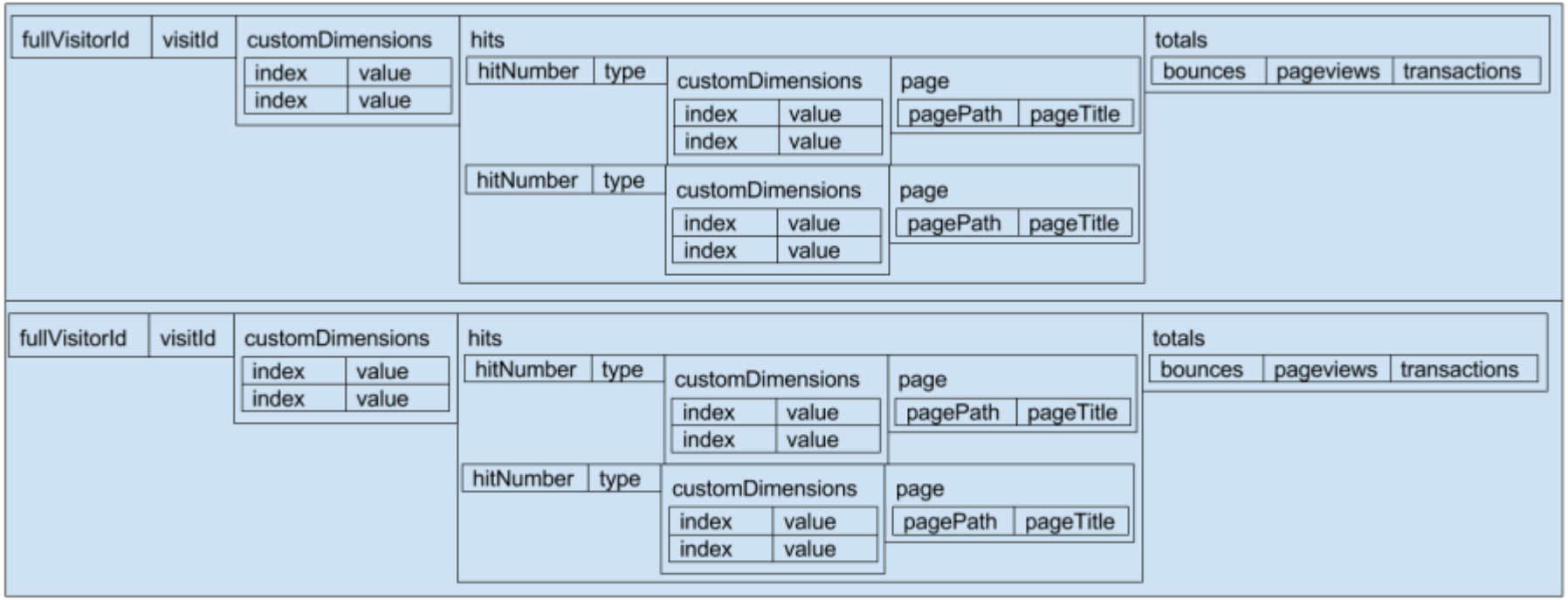
Mission :
테이블 쉽게 쿼리하기, `와 '의 차이, LIMIT 이해하기

- Hint :



Mission :

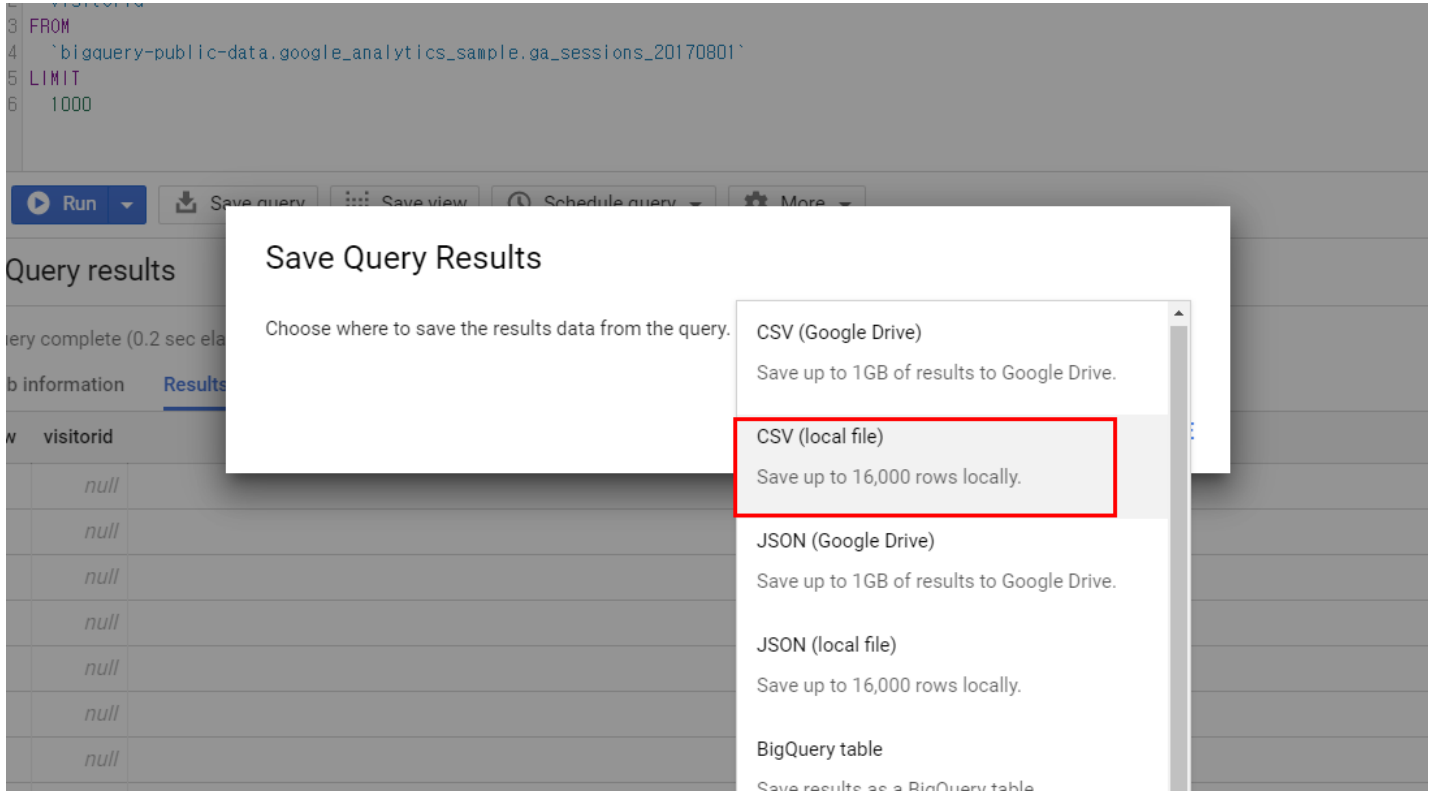
- *(와일드카드)로 조회하기
- 쿼리 크기 비교하기
- 빅쿼리 데이터 구조 이해하기



Mission :

- Format 기능
- device.browser 쿼리 후 시각화하기
- 쿼리 결과 csv로 내려받기

- Hint :



Mission :

- 다운받은 csv 파일 빅쿼리에 다시 적재(load)하기

- Hint : 1. 데이터셋 생성 2. create table

Mission :

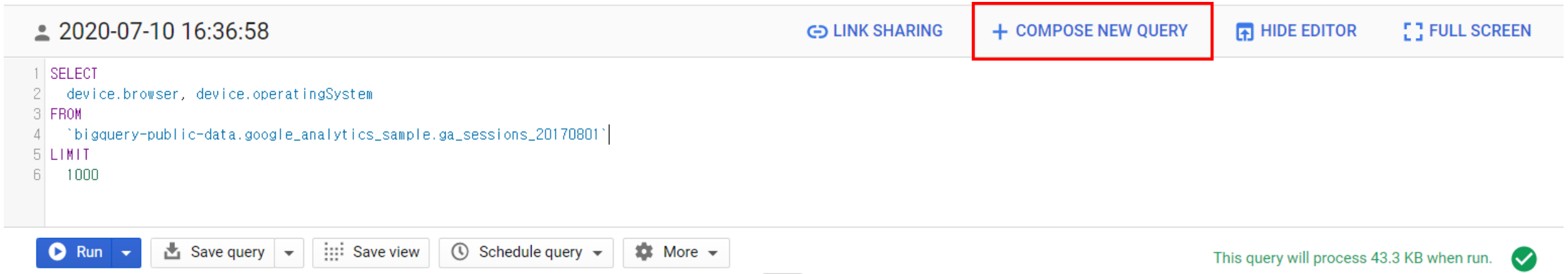
- 쿼리 세이브하기
- 과거 쿼리 내역 조회하기
- Job 이랑 비교하기

- Hint : 화면 왼쪽 메뉴 중 하나!

Mission :

- 새로운 쿼리 작성 편하게 하는 법
- 테이블 삭제하기
- 데이터 삭제 하기

- Hint :

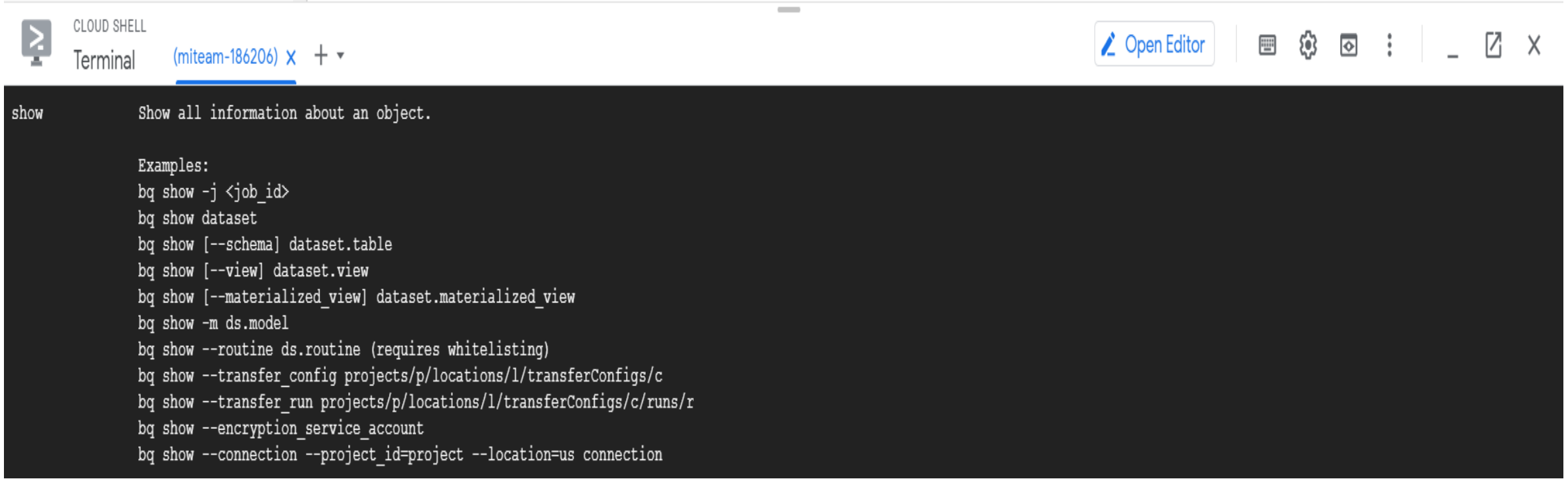


Congratulations! (1 단계 미션 클리어 : 경험치 XP 100 상승 up!!!)

다음 Stage 로 고고씽~~~!!!!

Bonus :
- 가끔 카페에서 멋있게 보이고 싶을 때 멋있게 쿼리하는 법

- Hint :



```
bq query --nouse_legacy_sql 'select * from bigquery-public-data.samples.shakespeare'
```

Query syntax (SQL 2011 + extensions)

BUILT-IN FUNCTIONS: SUM, IF, COUNT

<PROJECT>.<DATASET>.<TABLE>

CLAUSE, BOOLEAN OPERATIONS

GROUP BY

```
SELECT     SQL-LIKE SYNTAX
   airline,
   SUM(IF(arrival_delay > 0, 1, 0)) AS num_delayed,
   COUNT(arrival_delay) AS total_flights

FROM
  `bigquery-samples.airline_ontime_data.flights`

WHERE
  arrival_airport='OKC'
  AND departure_airport='DFW'

GROUP BY
  airline
```

Query Results

Row	airline	num_delayed	total_flights
1	AA	10312	23060
2	OO	198	552
3	EV	756	1912
4	MQ	3884	7903

Google Cloud

https://cloud.google.com/bigquery/docs/reference/standard-sql/

number of user interactions on 10th September 2013 before a purchase

```
SELECT one.hits.item.productSku AS ProductSku, ( sum_of_hit_number / total_hits ) AS
avg_hit_number
FROM (
 SELECT hits.item.productSku, SUM(hits.hitNumber) AS sum_of_hit_number
 FROM [GoogleStore.ga_sessions_20130728]
 WHERE hits.item.productSku IS NOT NULL
  AND totals.transactions>=1
 GROUP BY hits.item.productSku  ) AS one
JOIN (
 SELECT hits.item.productSku, COUNT( fullVisitorId ) AS total_hits
 FROM [GoogleStore.ga_sessions_20130728]
 WHERE hits.item.productSku IS NOT NULL
  AND totals.transactions>=1
 GROUP BY hits.item.productSku ) AS two
ON one.hits.item.productSku = two.hits.item.productSku;
```

select 실습

1. 퍼블릭 데이터 세트에서 samples 의 Shakespeare 테이블을 선택
2. Query table 선택
3. * 로 모든 컬럼 검색
4. order by word_count desc 해보기

Mission :

- 테이블 만들기

create table miteam-186206.test1234.jin (ID int64, LastName string, FirstName string, address string)

Mission :

- UI에서 테이블 만들기

Mission :

- 데이터 insert 하기

insert into miteam-186206.test1234.jin (id, lastname, firstname, address) values (1, "Lee", "Jin", "테헤란로")

Mission :

- 데이터 10개 정도를 넣어보자!

Mission :

- 주석 (comment) 를 사용해보자

Mission :

- 데이터 select 하기

```
SELECT *  FROM `miteam-186206.test1234.jin`
```

Mission :

- 특정 컬럼 select 하기

```
SELECT address  FROM `miteam-186206.test1234.jin`
```

Mission :

- 특정 조건의 데이터 select 하기

```
SELECT *  FROM `miteam-186206.test1234.jin`
where address =‘테헤란로’
```

Mission :

- 특정 데이터 delete 하기

delete from 테이블명 where id = 1

Mission :

- 테이블 drop 하기

drop table `miteam-186206.test1234.jin`

Congratulations! (2 단계 미션 클리어 : 경험치 XP 1000 상승 up!!!)

다음 Stage 로 고고씽~~~!!!!

중 간 고 사

어이쿠 당했구나!

직접 본인의 데이터베이스를 만들어보자!

- 데이터베이스를 만들고 강사에게 자랑하기! (제한시간 20분)
- 테이블 스키마를 그럴싸하게! 데이터도 그럴싸하게!
- 다하신 분은 복습, 휴식, 질문, DB타임(?) 중에 ㄱㄱ

(심화) _table_suffix 를 사용해 봅시다.

쿼리 편집기

```
1 SELECT fullvisitorid
2 FROM `bigquery-public-data.google_analytics_sample.ga_sessions_*`
3 where _table_suffix between "20170701" and "20170801"
```

(개심화) 빅쿼리에서 가장 어려운 개념 (Array, Struct) 이해하기

- Record, repeated
- totals
- totals.*
- hits
- hits.hitNumber
- unnest

기 말 고 사

아니. 이 양반이?!

오늘 배운 것 앞으로 업무에 잘 활용하기

졸업 축하합니다. A+ 드립니다!!!!