A Mini-Project in Big Data Analytics on Google Cloud in Hive

BE COMPS

TITLE:

Group Members: **Ajinkya Darshane(16102001)**

Aditya Joshi(16102017)

Dataset: *world_war_2_aircrafts*

Attributes : aircraft_type

aircraft_subtype

aircraft_name

year_in_service

country_of_origin

operators

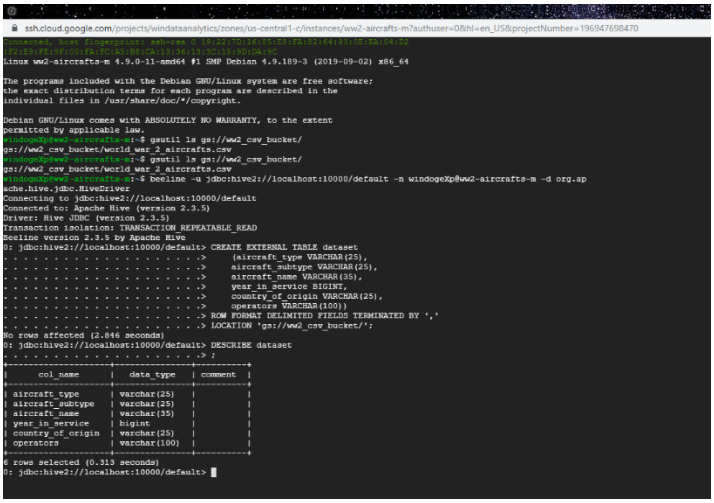
EXECUTION#

Query#1: To create and describe the table

Commands:

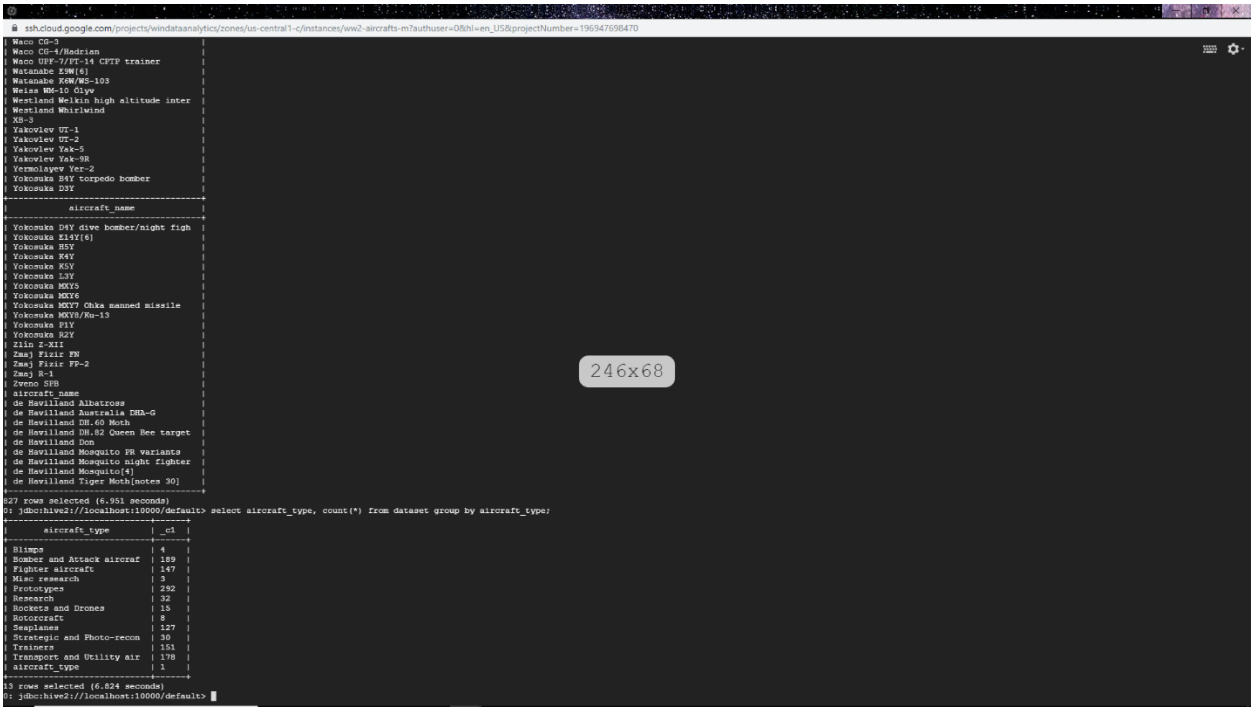
```
CREATE EXTERNAL TABLE dataset
        (aircraft_type VARCHAR(25),
        aircraft_subtype VARCHAR(25),
        aircraft_name VARCHAR(35),
        year_in_service BIGINT,
        operators VARCHAR(100))
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 'gs://ww2_csv_bucket/';

describe dataset;
```



Query#2: Display the types of aircrafts along with their count

Command: select distinct aircraft_type, count(*) from dataset group by aircraft_type;



Query#3: Display the inaugural years for each aircraft types

Command: select aircraft_type, MIN(year_in_service) as inaugural_year from dataset group by aircraft_type;

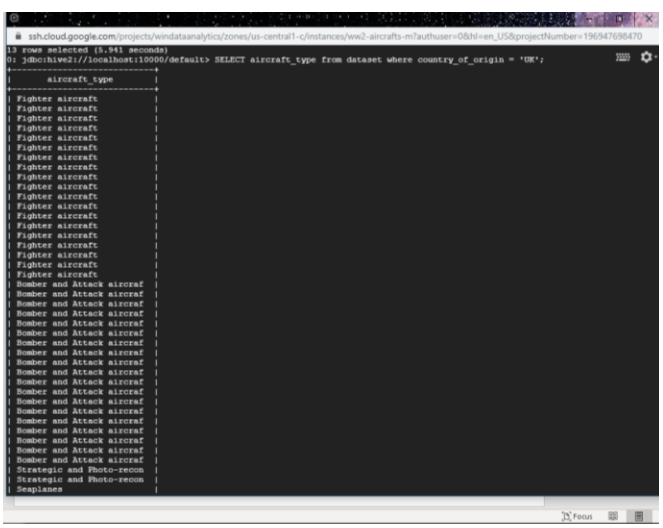
Query#4: Display the final years for each aircraft types

Command: select aircraft_type, MAX(year_in_service) as final_year from dataset group by aircraft_type;

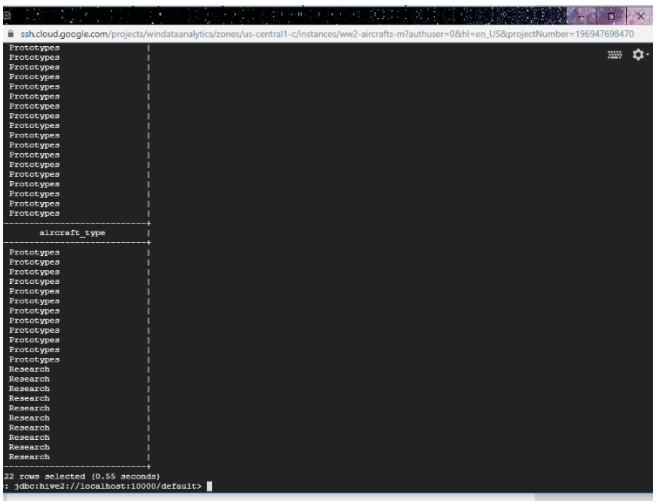
```
13 rows selected (6.216 seconds)
0: jdbc:hive2://localhost:10000/default> select aircraft_type, MAX(year_in_service) as final_year from dataset group by aircraft_
type;
+-----------------------------+-------------+
|        aircraft_type        | final_year  |
+-----------------------------+-------------+
| Blimps                      | 1944        |
| Bomber and Attack aircraf   | 1945        |
| Fighter aircraft            | 1945        |
| Misc research               | 1944        |
| Prototypes                  | 1945        |
| Research                    | 1945        |
| Rockets and Drones          | 1945        |
| Rotorcraft                  | 1945        |
| Seaplanes                   | 1945        |
| Strategic and Photo-recon   | 1945        |
| Trainers                    | 1944        |
| Transport and Utility air   | 1945        |
| aircraft_type               | NULL        |
+-----------------------------+-------------+
13 rows selected (5.941 seconds)
0: jdbc:hive2://localhost:10000/default>
```

Query#5: Select an attribute with a given condition

Command: SELECT aircraft_type from dataset where country_of_origin = 'UK';

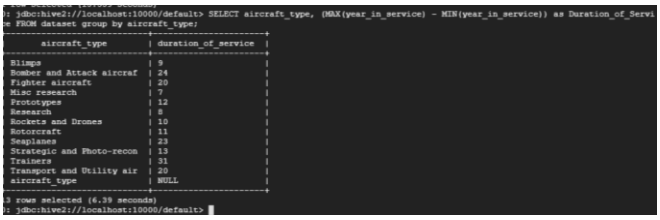


```
ssh.cloud.google.com/projects/windataanalytics/zones/us-central1-c/instances/ww2-aircrafts-m?authuser=0&hl=en_US&projectNumber=196947698470
13 rows selected (5.941 seconds)
0: jdbc:hive2://localhost:10000/default> SELECT aircraft_type from dataset where country_of_origin = 'UK';
+-----------------------------+
|        aircraft_type        |
+-----------------------------+
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Fighter aircraft            |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Bomber and Attack aircraf   |
| Strategic and Photo-recon   |
| Strategic and Photo-recon   |
| Seaplanes                   |
```



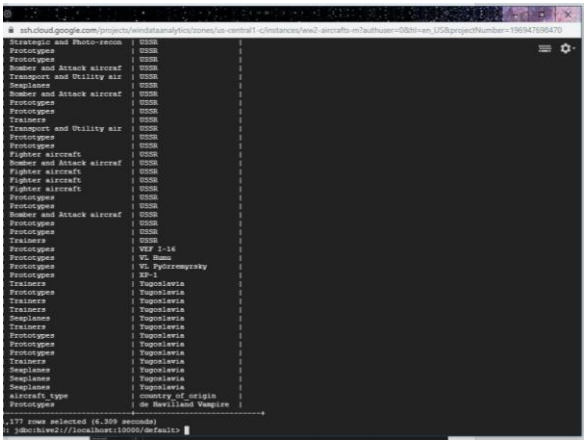
Query#6: Display the duration of service of each aircraft types

Command: SELECT aircraft_type, (MAX(year_in_service) - MIN(year_in_service)) as Duration_of_Service FROM dataset group by aircraft_type;



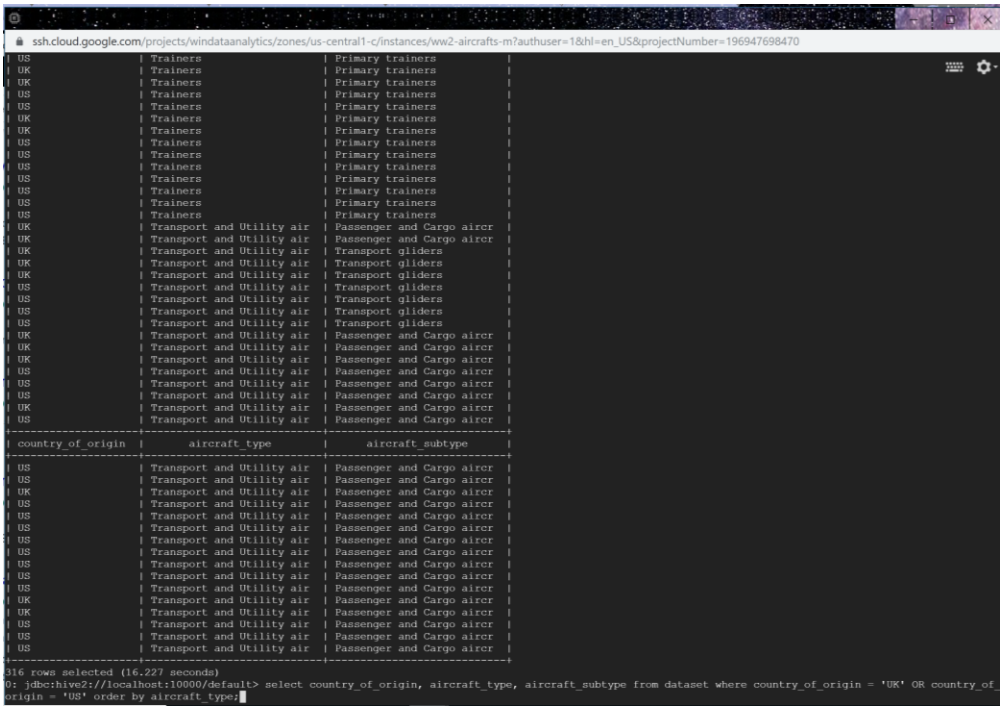
Query#6: Order aircrafts by country

Command: SELECT aircraft_type, country_of_origin FROM dataset ORDER BY country_of_origin;



Query#7: Select aircrafts common to central powers

Command: select country_of_origin, aircraft_type, aircraft_subtype from dataset where country_of_origin = 'Germany' OR country_of_origin = 'Japan' order by aircraft_type;



Query#8: Select aircrafts common to central powers

Command: select country_of_origin, aircraft_type, aircraft_subtype from dataset where country_of_origin = 'US' OR country_of_origin = 'UK' order by aircraft_type;