

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;
```

```

$ cd /root/.aws
$ aws s3 cp s3://aws-logs-913611831361-us-east-1/s3/aws-logs-agent-installers/ /usr/share/doc/ --recursive
Linux aws2-aircrafts-m 4.9.0-11-amd64 #1 SMP Debian 4.9.189-3 (2019-09-02) x86_64

The programs included with the Debian GNU/Linux system are free software;
the distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.

root@aws2-aircrafts-m:~# gsu2 ls gsu1/wm2_csw_bucket/
gsu1/wm2_csw_bucket/world_war_2-aircrafts.csv
root@aws2-aircrafts-m:~# gsu2 ls gsu1/wm2_csw_bucket/
gsu1/wm2_csw_bucket/world_war_2-aircrafts.csv
root@aws2-aircrafts-m:~# ssh root@aws2-aircrafts-m -i /root/.ssh/jdbc-hive2 //localhost:10000/default -n winlogexpgwm2-aircrafts-m -d org.apache.hive.jdbc.HiveDriver
Connecting to jdbc:hive2://localhost:10000/default
Connected to Apache Hive (version 2.3.5)
Transaction isolation: TRANSACTION_READ_COMMITTED
Debian version 2.3.5 by Apache Hive
root@aws2-aircrafts-m:~# jdbc:hive2://localhost:10000/default; CREATE EXTERNAL TABLE dataset
(
  aircraft_type VARCHAR(25),
  aircraft_subtype VARCHAR(25),
  aircraft_name VARCHAR(35),
  year_in_service HIGINT,
  country_of_origin VARCHAR(25),
  operators VARCHAR(100)
)
ROW FORMATTED DELIMITED FIELDS TERMINATED BY ','
LOCATION 'gsu1/wm2_csw_bucket';
root@aws2-aircrafts-m:~#
root@aws2-aircrafts-m:~# jdbc:hive2://localhost:10000/default; DESCRIBE dataset
+-----+
| col_name | data_type | comment |
+-----+
| aircraft_type | varchar(25) |
| aircraft_subtype | varchar(25) |
| aircraft_name | varchar(35) |
| year_in_service | bigint |
| country_of_origin | varchar(25) |
| operators | varchar(100) |
+-----+
6 rows selected (0.113 seconds)
root@aws2-aircrafts-m:~# jdbc:hive2://localhost:10000/default;

```

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

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

```
1 htdcloud@google.com: projects/wendataanalytics/tables/central -> (dataset/central) aircraft: m/author=3001-en_us&projectId=196347036470
| Waco CG-3
| Waco CG-4/Hedrian
| Waco OVF-7/PT-14 CPTF trainer
| Watanabe E2W(1)
| Watanabe K6W/WS-103
| Weiss WB-10 Clwy
| Westland Welkin high altitude inter
| Westland Whirlwind
| XP-3
| Yakovlev UT-1
| Yakovlev UT-2
| Yakovlev Yak-5
| Yakovlev Yak-9
| Yakovlev Yak-18
| Yeremeyev Ye-2
| Yokosuka M4Y torpedo bomber
| Yokosuka SH
|
+-----+
| aircraft_name
+-----+
| Yokosuka M4Y dive bomber/night figh
| Yokosuka Z44(e)
| Yokosuka SHV
| Yokosuka K4Y
| Yokosuka K4Y
| Yokosuka L3Y
| Yokosuka M10
| Yokosuka M10
| Yokosuka M10V Gha named missile
| Yokosuka M10V/No-13
| Yokosuka F1V
| Yokosuka K2Y
| Zilin Z-111
| Zmaj Fizir F3
| Zmaj Fizir F3-2
| Zmaj K-1
| Zveno SF8
|
+-----+
| aircraft_name
+-----+
| de Havilland Albatross
| de Havilland Australia OMA-G
| de Havilland DH.60 Moth
| de Havilland DH.82 Queen Bee target
| de Havilland Don
| de Havilland Mosquito PR variants
| de Havilland Mosquito night fighter
| de Havilland Mosquito(4)
| de Havilland Tiger Moth[notes 30]
|
+-----+
627 rows selected (6.951 seconds)
01 jdbc:hive2://localhost:10000/default> select aircraft_type, count(*) from dataset group by aircraft_type;
+-----+
| aircraft_type | _c1
+-----+
| Blimps | 4 |
| Bomber and Attack aircraft | 149 |
| Fighter aircraft | 147 |
| Misc research | 2 |
| Prototypes | 282 |
| Research | 32 |
| Rockets and Drones | 15 |
| Rotorcraft | 4 |
| Seaplanes | 127 |
| Strategic and Photo-recon | 10 |
| Trainers | 151 |
| Transport and Utility air | 178 |
| aircraft_type |
+-----+
13 rows selected (6.824 seconds)
01 jdbc:hive2://localhost:10000/default>
```

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;

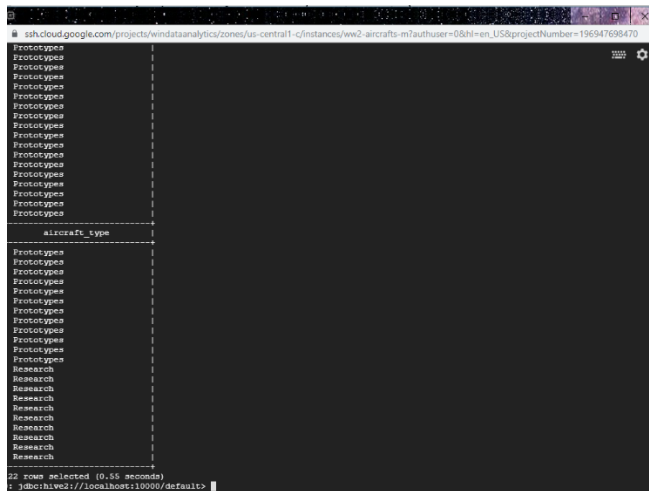
```
01 jdbc:hive2://localhost:10000/default> select aircraft_type, MIN(year_in_service) as inaugural_year from dataset group by aircraft_type;
+-----+
| aircraft_type | inaugural_year
+-----+
| Blimps | 1935 |
| Bomber and Attack aircraft | 1921 |
| Fighter aircraft | 1925 |
| Misc research | 1937 |
| Prototypes | 1933 |
| Research | 1937 |
| Rockets and Drones | 1935 |
| Rotorcraft | 1934 |
| Seaplanes | 1922 |
| Strategic and Photo-recon | 1952 |
| Trainers | 1913 |
| Transport and Utility air | 1925 |
| aircraft_type | NULL |
+-----+
13 rows selected (6.216 seconds)
```

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 aircraft | 1945       |
| Fighter aircraft | 1945       |
| Misc research  | 1944       |
| Prototypes    | 1945       |
| Research       | 1945       |
| Rockets and Drones | 1945       |
| Rotocraft     | 1945       |
| Seaplanes     | 1945       |
| Strategic and Photo-recon | 1945       |
| Trainers      | 1944       |
| Transport and Utility air | 1945       |
| aircraft_type | NULL       |
+-----+-----+
13 rows selected (5.541 seconds)
0: jdbc:hive2://localhost:10000/default:
```

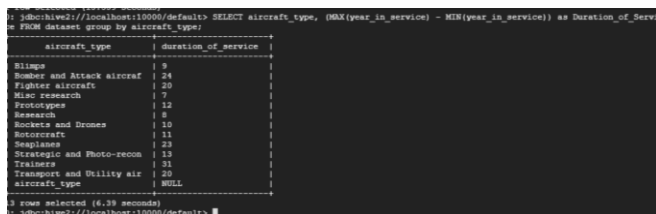
Command: SELECT aircraft_type from dataset where country of origin = 'UK';

[illegible]



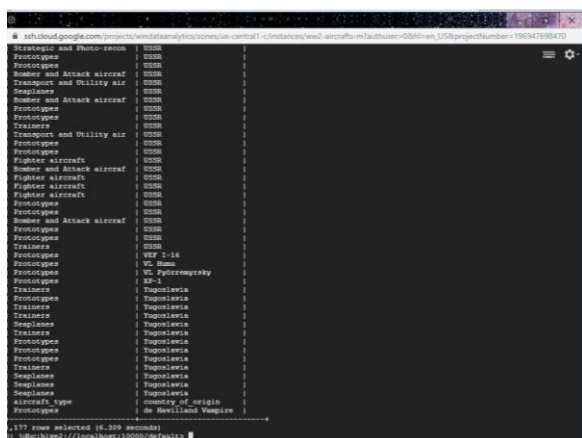
Query#6: Display the duration of service of each aircraft types

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;



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;

```
ssh.cloud.google.com/projects/windataanalytics/zones/us-central-1-c/instances/ww2-aircraft-m?authuser=1&hl=en_US&projectNumber=196947698470
```

US	Trainers	Primary trainers	
UK	Trainers	Primary trainers	
UK	Trainers	Primary trainers	
US	Trainers	Primary trainers	
US	Trainers	Primary trainers	
UK	Trainers	Primary trainers	
US	Trainers	Primary trainers	
US	Trainers	Primary trainers	
US	Trainers	Primary trainers	
US	Trainers	Primary trainers	
US	Trainers	Primary trainers	
US	Trainers	Primary trainers	
UK	Transport and Utility air	Passenger and Cargo aircraft	
UK	Transport and Utility air	Passenger and Cargo aircraft	
UK	Transport and Utility air	Transport gliders	
UK	Transport and Utility air	Transport gliders	
US	Transport and Utility air	Transport gliders	
US	Transport and Utility air	Transport gliders	
US	Transport and Utility air	Transport gliders	
US	Transport and Utility air	Passenger and Cargo aircraft	
UK	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
UK	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	

country_of_origin	aircraft_type	aircraft_subtype	

US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
UK	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
UK	Transport and Utility air	Passenger and Cargo aircraft	
UK	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	
US	Transport and Utility air	Passenger and Cargo aircraft	

116 rows selected (16.227 seconds)

```
% !sh -c "psql -h localhost -U root -d windataanalytics -e 'select country_of_origin, aircraft_type, aircraft_subtype from dataset where country_of_origin = 'UK' OR country_of_origin = 'US'; 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;

country_of_origin	aircraft_type	aircraft_subtype
Germany	Trainers	Advanced trainers
Japan	Trainers	Advanced trainers
Japan	Trainers	Advanced trainers
Japan	Trainers	Advanced trainers
Japan	Trainers	Advanced trainers
Germany	Trainers	Glider trainers
Germany	Trainers	Glider trainers
Germany	Trainers	Primary trainers
Germany	Trainers	Primary trainers
Germany	Trainers	Primary trainers
Germany	Trainers	Primary trainers
Japan	Trainers	Primary trainers
Japan	Trainers	Primary trainers
Germany	Trainers	Advanced trainers
Germany	Trainers	Advanced trainers
Japan	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
country_of_origin	aircraft_type	aircraft_subtype
Germany	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Germany	Transport and Utility air	Transport gliders
Germany	Transport and Utility air	Transport gliders
Japan	Transport and Utility air	Transport gliders
Germany	Transport and Utility air	Transport gliders
Japan	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr
Japan	Transport and Utility air	Passenger and Cargo aircr

22 rows selected (6.488 seconds)

j@debian2:~/localhost:10000/default> select country_of_origin, aircraft_type, aircraft_subtype from dataset where country_of_origin = 'Germany' OR countr
of_origin = 'Japan' order by aircraft_type;