ITCS 6100

Project Group 1

Rachel William, Gnanasri Pusunuru, Lakshmi Srujitha Gali, Sai Radhika Koneru, Colin McNeil

Deliverable 2

SUMMARY

We were able to successfully complete the following steps in our project:

* Launch and configure an AWS Cloud9 integrated development environment (IDE) instance.
* Run SQL queries against a single file by using the Amazon S3 Select feature of Amazon Simple Storage Service (Amazon S3).
* Transform CSV-formatted data files to the Apache Parquet format and upload them to Amazon S3.
* Create an AWS Glue crawler to infer the structure of the data.
* Use Amazon Athena to query the data.
* Create an Athena view.
* Use Amazon QuickSight to visualize the data.

CODE

Task 3:

SELECT year, fishing\_entity AS Country, CAST(CAST(SUM(landed\_value) AS DOUBLE) AS DECIMAL(38,2)) AS ValuePacificWCSeasCatch

FROM <FMI\_1>

WHERE area\_name LIKE '%Pacific%' and fishing\_entity='Fiji' AND year > <FMI\_2>

GROUP BY year, fishing\_entity

ORDER By year

To find value in US dollars of all fish caught by the country Fiji from the Pacific, Western Central high seas area since 2001 we changed FMI\_1 to the data-source bucket and changed FMI\_2 to the year 2000.

SELECT year, fishing\_entity AS Country, CAST(CAST(SUM(landed\_value) AS DOUBLE) AS DECIMAL(38,2)) AS ValueAllHighSeasCatch

FROM fishdb.data\_source\_xxxxx

WHERE area\_name LIKE '%Pacific%' and fishing\_entity='Fiji' AND year > 2000

GROUP BY year, fishing\_entity

ORDER By year

Task 4:

# Change the names of the 'fish\_name' and 'country' columns to match the column names where this data appears in the other data files already in your data-source bucket

This code merges the data in the fish\_name column with the data in the common\_name column from the HighSeas dataset. It also merged the data in the country column with the data in the fishing\_entity column.

df.rename(columns = {fish\_name": "common\_name", "country": "fishing\_entity"}, inplace = True)

Task 5:

Using this query we were able to find the 6 countries or fishing entities that caught the highest mackerel catch each year.

SELECT year, Country, MAX(TotalWeight) AS Weight

FROM fishdb.mackerelscatch

GROUP BY year, Country

ORDER BY year, Weight DESC;

We modified the MackerelsCatch view we created earlier to adjust it to years 2013-2018 and only include the 6 fishing entities that had the highest mackerel catch.

CREATE OR REPLACE VIEW MackerelsCatch AS

SELECT year, area\_name AS WhereCaught, fishing\_entity as Country, SUM(tonnes) AS TotalWeight

FROM fishdb.data\_source\_09876

WHERE common\_name LIKE '%Mackerels%' AND year >= 2013 AND year <=2018

AND fishing\_entity IN ('USA', 'Fiji', 'Indonesia','Viet Nam','Japan','Taiwan')

GROUP BY year, area\_name, fishing\_entity, tonnes

ORDER BY tonnes DESC

The First Data chart in QuickSight changed to display the countries and years we specified.

