**Assignment 9: BigQuery Queries**

**Pre-requisites:**

* User must have Gmail account with GCP access.
* User must have knowledge on access BigQuery on GCP
* User must be able to access public datasets on BigQuery.

**Task 1: BigQuery Queries**

1. From the GCP console page, click on three-line (navigation menu) icon at top-left corner of the screen. Scroll to find BigQuery service, hover on it then click SQL Workspace to open BigQuery service on GCP console.

Graphical user interface, application, email, website

Description automatically generated

1. You can add data into BigQuery by clicking on “ADD DATA” button in the Explorer.
2. Now, go to the public datasets project (bigquery-public-data) and search for “ncaa\_basketball” dataset. Now, select “mbb\_pbp\_sr” table

Graphical user interface, application

Description automatically generated

1. You can also see in the above screenshot about schema previewing table contents and details.
2. Now, click on “QUERY” button and select “In New Tab” Graphical user interface, application

   Description automatically generated
3. Query Editor tab opens up with default SELECT query. You can write your own queries in the editor. After you write your query, click on run button just above the editor.

Graphical user interface, text, application, email

Description automatically generated

**Task 2: Queries:**

1. Query to fetch points scored by Wildcats team along with game time.

SELECT game\_clock, points\_scored, team\_name, event\_description FROM `bigquery-public-data.ncaa\_basketball.mbb\_pbp\_sr` WHERE season = 2014 AND home\_name = 'Wildcats' AND away\_name = 'Fighting Irish' AND points\_scored IS NOT NULL ORDER BY timestamp DESC LIMIT 10;

* 1. Game\_clock, points\_scored, team\_name, event\_description Columns are selected from the table.
  2. Data has been selected based on season, home\_name, away\_name, and points\_scored
  3. Season is 2014, home\_name is Wildcats, away\_name is Fighting Irish, Points scored must have a value.
  4. Resultant data is ordered by Timestamp column.
  5. Finally, data is limited to 10 rows.

Graphical user interface, text, application, email

Description automatically generated

1. Query to fetch scores of each time over the game clock

SELECT game\_clock, SUM( CASE WHEN team\_name = 'Wildcats' THEN points\_scored END ) OVER(ORDER BY timestamp ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS wildcats\_score, SUM( CASE WHEN team\_name = 'Fighting Irish' THEN points\_scored END ) OVER(ORDER BY timestamp ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS fighting\_irish\_score, team\_name, event\_description FROM `bigquery-public-data.ncaa\_basketball.mbb\_pbp\_sr` WHERE season = 2014 AND home\_name = 'Wildcats' AND away\_name = 'Fighting Irish' AND points\_scored IS NOT NULL ORDER BY timestamp DESC LIMIT 10;

* 1. columns are game\_clock, wildcats\_score, fighting\_irish\_score, team\_name, event\_description.
  2. we are calculating individual team points by using SUM function and OVER clause which groups rows in the table.
  3. Here, we are using UNBOUNDED PRECEDING to the CURRENT ROW to group the rows from start of the game to the current row.
  4. Resultant is ordered by Timestamp
  5. Results are limited to 10 rows for this activity