TASK - 3

OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

# Project Description:

This project talks about Operational Analytics which I understand to be the analysis of a company's entire activities. This helps the business identify the areas where it needs to make enhancements. In such an analysis we closely collaborate with different teams involved like the ops team, the support staff, the sales team, etc. and assist them in drawing conclusions from the data they gather. Improved workflows, better communication and better automation are the results. As we are training to be a Data Analyst, one must be able to comprehend or help other teams understand questions, therefore analyzing metric spikes is also a crucial component of operational analytics. This type of study is often used to estimate the general rise or fall of a company's fortune because it is also one of the essential components of a business.

# Approach:

I would like to go with the Divide and Conquer Approach i.e. group all the similar problems and solve them together sequentially one after the other.

There happen to be two case studies:

Case study 1 is on dataset called as “job\_data” where different types of areas are tested

Case study 2 is on three tables: users, events and email\_events. All the problems are based on these three datasets.

# Tech Stack:

Language: Structured Query Language [S.Q.L.]

Software used: MySQL

Version: 8.0.30

Original Author: MySQL AB

Developer: ORACLE Corporation

Latest Stable Release: 6th July 2022

Purpose: To run and execute SQL Commands.

# Task/Solution:

## CASE STUDY 1:

1. **Number of jobs reviewed:** Amount of jobs reviewed over time.  
   **Your task:** Calculate the number of jobs reviewed per hour per day for November 2020?

Sol: The task can be solved using the following query:

select count(\*), DATE\_PART('day',ds) as day, DATE\_PART('hour',ds) as hour

from job\_data

Group by hour,day;

1. **Throughput:** It is the no. of events happening per second.  
   Let’s say the above metric is called throughput.

Sol: Calculate 7 day rolling average of throughput?

select avg(event) as avg\_of\_events over()

from job\_data

Where ds>now()-interval ‘7 day’;

For throughput, do you prefer daily metric or 7-day rolling and why?

Number of events happening per second can hugely be affected by the day of the week. Moreover on a longer run, lots of data needs to be processed everyday if we take the daily metric as the option and it also takes away a lot of time. However when we choose a 7-day metric we can assess the interactions weekly. Hence choosing a 7 day rolling metric is more suggestable.

1. **Percentage share of each language:** Share of each language for different contents.  
   Calculate the percentage share of each language in the last 30 days?

Sol: The percentage share of each language in the last 30 days is:

select (count(language)/(select count(\*) from job\_data))\*100

From job\_data

Where ds>now()-interval ‘30 day’

Group by language;

1. **Duplicate rows:** Rows that have the same value present in them.  
   Let’s say you see some duplicate rows in the data. How will you display duplicates from the table?

Sol: The duplicate data in the table can be represented using a rank function which will give the same rank to duplicate data and using the having clause we can simply return those records which are duplicate.

Select \*, rank() over() as rank, count(rank) as rank\_count

From job\_data

Group by rank

Having rank\_count>1;

## CASE STUDY 2:

1. **User Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service.  
   Calculate the weekly user engagement?

Sol: The user engagement for the week can be given by

select user\_id, location, event\_name

from events

Where (event\_type=’engagement’) and (occured\_at between dateadd(d,-7,convert(nvarchar(10),getdate(),101)) and convert(nvarchar(10),getdate(),101))

Order by user\_id;

1. **User Growth:** Amount of users growing over time for a product.  
   Calculate the user growth for the product?

Sol: (Did not understand the question exactly)

Select \* over() from users;

1. **Weekly Retention:** Users getting retained weekly after signing-up for a product.  
   **Your task:** Calculate the weekly retention of users-sign up cohort?

Sol: The weekly retention of users-sign up cohort: i presume user\_signup here means complete\_signup in table 2

select events.user\_id, occured\_at, state, event\_name

from users, events

where event\_name=’complete\_signup’ and occured\_at>now()-interval ‘7 day’;

1. **Weekly Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.  
   Calculate the weekly engagement per device?

Sol**:** The weekly engagement per device can be calculated using:

select \*

from events

where event\_type=’engagement’ and occured\_at>now()-interval ‘7 day’

group by device;

1. **Email Engagement:** Users engaging with the email service.  
   **Your task:** Calculate the email engagement metrics?

Select \* from email\_events

Where occured\_at>now()-interval ‘7day’;

# Insights:

Insights refers to accurate understanding of something. These points helps in an insightful understanding of the problems:

* All the problems refer to real-life situations which any data analyst would face while dealing with data. The attributes may differ but the application or approach will not change.
* There are 2 case studies
* Case Study 1 is data collected for a week and called job\_data over different features.
* Case Study 2 has 3 tables: users, events, email\_events.
* The tasks can be solved using concepts of window function (over()), date and time functions, also our previous DDL and DML commands.

# Result:

To recapitulate, the results are elaborately discussed above, moreover this project/task helped me in better understanding of SQL commands and working with databases. It also enhanced my Critical Thinking and Problem-Solving skills. (I could not solve all the questions by using joins. However I managed to draw conclusions using other concepts which are hopefully right).

Thank You.