



Introduction

My Full name is Anurag Srivastava, i have been working as Data Engineer from past 3 years, i started my career in 2021, where i got mapped to a project of Cloud Data Warehouse that was basically a data migration project, where i got opportunities to work in 2 different teams. Both the team i worked was core Data Team.

Then explain about the project description in brief, quantify the data example worked on loading of 975 PB of Data.

Roles and Responsibilities

As mentioned i worked in 2 Teams -:

DDL Team-: where we were responsible for conversion of the objects like table,views, procedures, macros, triggers to GCP compatible and also support the cross functional teams if there were any issues with the testing and ETL jobs, or object definitions.

DMU Team-: Here we were responsible to load the history and incremental data, there were tables divided as Full load($\leq 100\text{GB}$) and Full Large Load($> 100\text{GB}$), and also help the teams for incremental load of data in batch. Also helped cross functional teams for data validation and data analysis.

Follow-up Questions on Projects

1. What was the Tech stacks you were using in the project?
2. How you kept the data after extracting?
3. What was the intermediate layer before loading to target?
4. How fast was the load and volume of data?

SQL Questions

1. What is Indexing in SQL, how it helps, is it logical or physical?
2. What is Nested field in BigQuery?
3. What is Repeated Fields in BigQuery?
4. What optimisation Techniques should be used in writing an complex select queries?
5. What is an Materialised View in SQL?

SQL Questions

Tell the difference between them?


1. `select count(*) from emp;`
2. `select count(1) from emp;`
3. `select count(distinct 2) from emp;`
4. `select 1 from emp;`
5. `select 1;`

emp	ename	salary
1	Akash	50000
2	Neha	60000
3	Rahul	55000
4	Priya	58000
5	Aarav	62000
2	Neha	60000
3	Rahul	55000

Python Questions

Write a Python Code to do fibonacci series by adding previous 3 numbers

python

 Copy code

```
def fibonacci(n):  
    # Initialize the first three numbers of the series  
    fib_series = [0, 1, 1]  
  
    # Generate the Fibonacci series up to n  
    for i in range(3, n):  
        next_number = fib_series[-1] + fib_series[-2] + fib_series[-3]  
        fib_series.append(next_number)  
  
    return fib_series[:n]  
  
# Test the function with n = 10  
n = 10  
print(f"Fibonacci series up to {n}:", fibonacci(n))
```

SQL Question

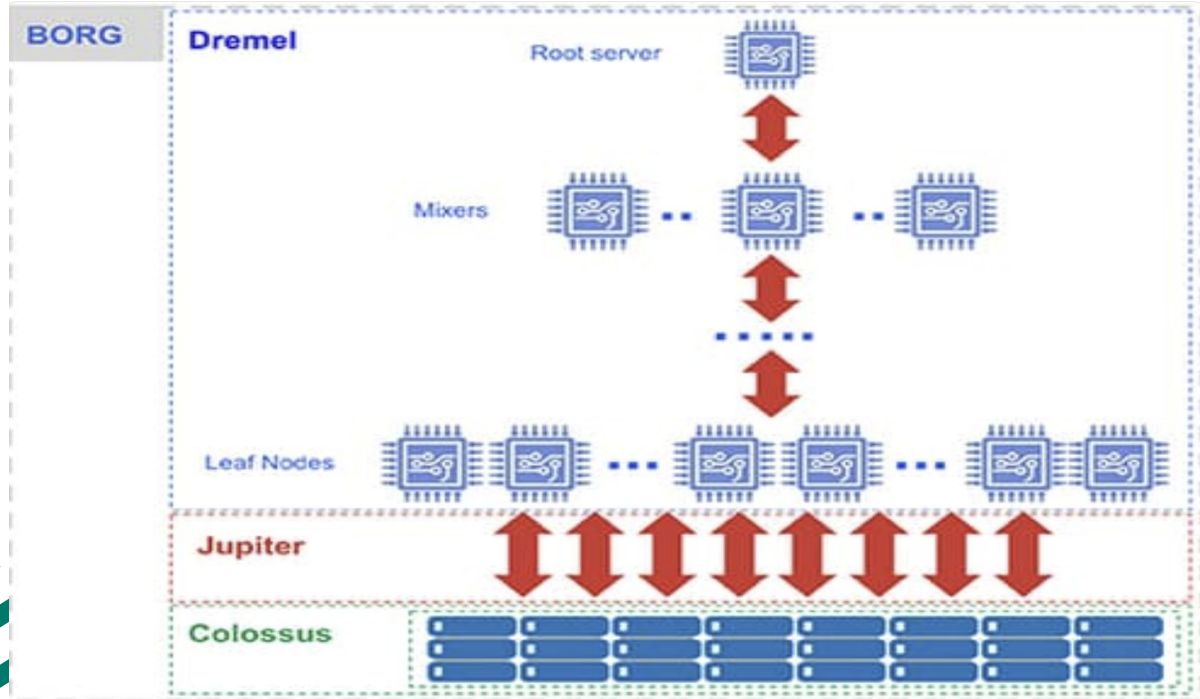
Query to write the second highest salary

sql

 Copy code

```
SELECT DISTINCT salary
FROM employees
ORDER BY salary DESC
LIMIT 1 OFFSET 1;
```


BigQuery Architecture



Questions on Migration

1. Why Migration Happened?
2. Disadvantages of using Teradata over GCP?
3. Why We didn't used any Data Transfer Tools?

Pyspark Question

1. What is Pyspark? Why it is used?
2. How to do Data Processing using GCP?
3. How do we read csv in spark?
4. If we don't give `inferSchema = True` what is the case?