

Deloitte

A photograph of a modern office building with a large, white, three-dimensional 'Deloitte' logo mounted on its roof. The building features large glass windows and a light-colored facade. A crane is visible behind the logo, suggesting it was recently installed. The sky is clear and blue. The image is framed by a blue border on the left and a yellow border on the right, with a green and white striped pattern at the bottom.

Tell me about yourself

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

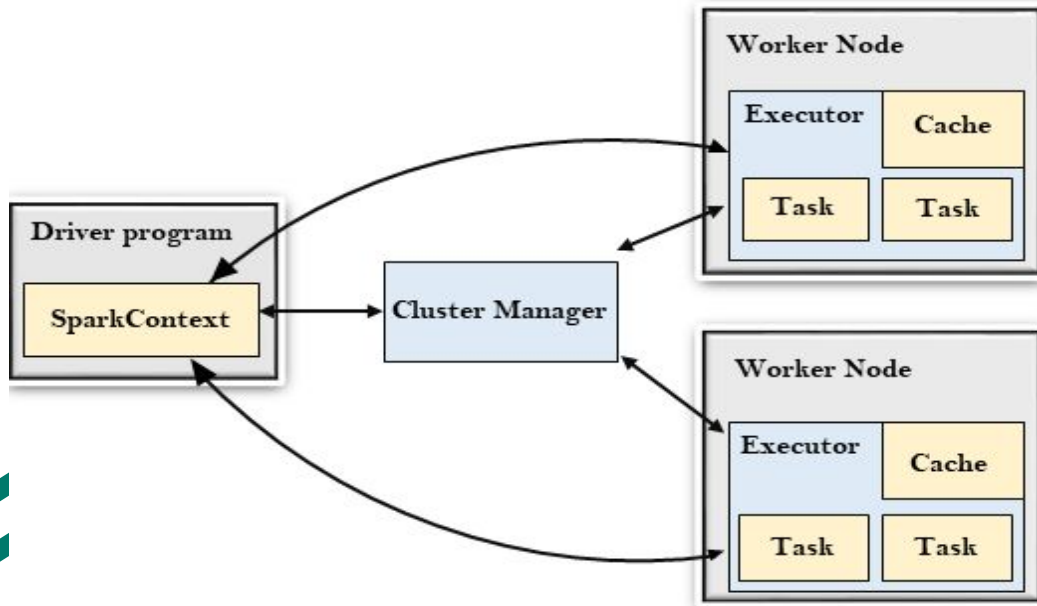
1. What all services i have used in GCP?
2. How you extracted the data from On-Premise DW and loaded to GCP?
3. Which Object Storage you used, and how it was helpful?
4. Did i used any Relational Database on GCP ?

GCP Related Questions

1. What is cloud SQL?
2. What is DataFlow?
3. What is DataProc?
4. Are you aware of DataPrep?
5. Have you ever used Pub/Sub?

Spark Related Questions

Are you aware of Spark Architecture



SQL Questions

Difference between RANK and DENSE_RANK

RANK

Rank	Student	Score
1	David	75
2	Alice	85
2	Charlie	85
4	Bob	90
5	Emily	95

DENSE_RANK

Dense Rank	Student	Score
1	David	75
2	Alice	85
2	Charlie	85
3	Bob	90
4	Emily	95

SQL Questions

What is SCD and how it is used?

SCD stands for "Slowly Changing Dimension." It's a concept used in data warehousing to manage changes to dimensional data over time. Dimensional data typically represents descriptive attributes of business entities, such as customers, products, or locations.

1. TYPE 1 SCD (Overwrite)
2. TYPE 2 SCD (Add New Row)
3. TYPE 3 SCD (Add New Column)

Spark and Related Questions

1. What is Lazy Execution in Spark?
2. What is Google Transfer Appliance?
3. What is Shuffling in Spark?
4. Are you aware of Docker or Kubernetes?

SQL Question

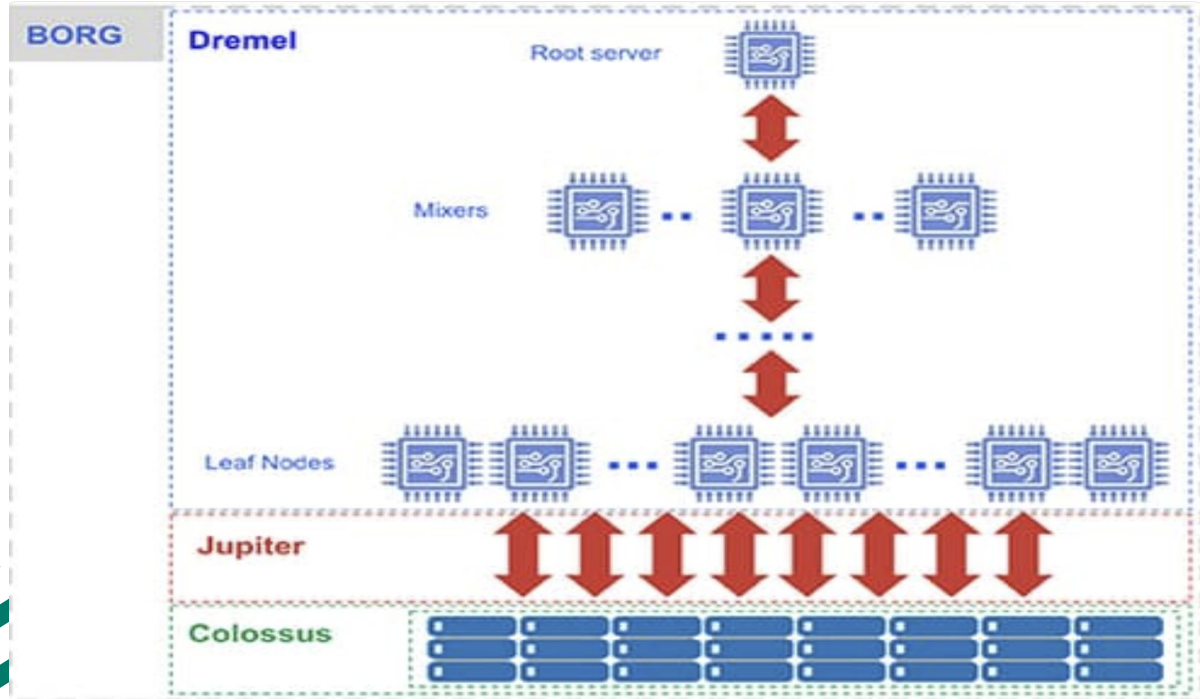
Query to write the second highest salary

sql

 Copy code

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

BigQuery Architecture



Challenges in the Project

1. Asked about Tell some challenges in the project?
2. Why the project went for 3 years?
3. Why We didn't used any data transfer service?
4. Have you worked on Spark Job configurations?