1. **Tech Mahindra** (Jan'24)
GCP Data Engineer

Round 1:
5-6 SQL Queries (Easy-Medium)
Bigquery - Optimization techniques, views

Round 2:
5-6 SQL Queries (Medium)
BQ Slots
Git commands

<Selected/>
<Offer_revoked_lt3_yoe/>

2. **Lloyds Bank** (Feb'24)
Data Engineer

Round 1:
Airflow - code
NiFi
Python - lambda exp, map,reduce,filter, classes

<Rejected/>

3. **Boston Consulting Group** (Mar'24)
Data Engineer

Round 1:
Wanted to ask Spark/Pandas - but no exp

Python - dictionary/list comprehension
- Interchange key and values in a dictionary (key becomes val, val becomes key) (Use Dict Comprehension)
- Count of substring in a string
- Characters with frequency >= 2, given a string
- Flatten a 2d list of integers and convert negative to positive (Use List Comprehension)

SQL
- Managers with more than 5 employees
- Highest salary in each department
- Top 3 unique salaries in each departments (30k, 30k, 20k, 10k, 5k -> return 30k,30k,20k,10k) (Use Dense Rank)

FastAPI - why not django

Git commands
- git rebase
- git reset --hard
- git merge

Round 1.5:
Round 1 didn't go well, so they gave an assignment
https://github.com/anandthegreat/Vehicle-Accident-Analysis/

<Rejected/>

3. **Oracle** (Apr'24)

Round 1 Oracle (Easy)

Laxmikant Chaudhari (Principal MTS, 14yoe)

Programming fundamentals:
Write a program to check if a string contains unique characters

```
def checkUniqueChar(s: str) -> bool:
    dict = {}
    for c in s:
        if c in dict:
            return False
        dict[c] = True
    return True
```

Write a program to count the frequency of characters in a string

```
def countChars(s: str) -> dict:
    for c in s:
        if c in dict:
            dict[c] = dict[c] + 1
        else:
            dict[c] = 1
    return dict
```

SQL
No of employees in each department sorted in descending order and with rank (1 for the highest no of emp)
emp -> emp_id,salary,name, dept_id
dept -> dept_id,name

```
SELECT *, DENSE_RANK() OVER(ORDER BY emp_count) AS RANK FROM(
    SELECT
    dept_id, dept_name, COUNT(*) AS emp_count
    FROM
    department d
    LEFT JOIN employee e
    ON TRIM(e.dept_id) = TRIM(d.dept_id)
    GROUP BY dept_id,dept_name
    ORDER BY 3 DESC
)
```

→ Will this work if some department has no employees?

Linux - how do you compile a code
g++ hello_world.cpp
./a.out

DSA

- linear vs non-linear data structure
- Find the middle of a linked list

Puzzle

First question: puzzle 8 stations: every station give half and take one, end with 1 (don't take 1 in the end). How many mangoes were there initially?
I told 256

Discussion about projects

- A new feature has been launched which caused test cases to fail. What would be your approach to debug it

Round 2 Oracle (Medium)

Sowmya Srinivasa (Senior MTS, 15 yoe)

Data Modelling

Design a movie ticketing system
USER

User_id, name, email, mobile_no

THEATER
Theater_id, name, screen_id, address

SCREEN
Theater_id, screen_id, movie_id, start_time, current_capacity, price

MOVIES
Movie_id, movie_name,release_date, movie_length

BOOKINGS
Booking_id, user_id, theater_id, screen_id, movie_id

PAYMENT
Payment_id, user_id, booking_id

PROMOTIONS
Theater_id, movie_id, discount

SQL

- What is Plan in SQL
- Employees not assigned any department

EMPLOYEE -> emp_id, name, department, project_id
PROJECT -> project_id, project_name, project_start_time

```
SELECT emp_id, name
FROM EMPLOYEE e
WHERE project_id IS NULL
```

DSA

Design LRU Cache

1 -> abc
2 -> xyz
3 -> pqr

Get(key) -1 O(1)
Put(key,value) O(1)
— if it is already present -> update its value and bring to front
— if not present
    — cache full: evict lru, insert key,valu
    — othrwise: insert in lru

hashmap (key, ptr::linkedlist)
Doubly Linkedlist<key,value>
    — start: most recently used
    — end: least recently used

Resume Discussion

First question: Project-level questions deep dive

- Use of APIs?
- How many tables
- Size of tables
- Challenges faced
- How much data used from the table
- Why GCP not Azure or AWS
- How do you handle errors in the transformations

Cloud functions equivalent in AWS, Azure

- How did you use Cloud functions in GCP
- I told for ticket creating in failure and trigger-based logic

What is Cold start? (I told about cloud functions cold start)
Docker

- Where have you used

- Where have you used
- Container vs VM
- Communication b/w multiple containers
- Docker Swarm

Have you implemented any data structures in your project?
Eg. stack, queue, linked list, tree etc?

Round 3 Oracle (Medium)

Johns Baby (Senior MTS, 11 yoe)

DSA

Find the first non-repeating character in a string

```
s = 'abaabcd'
def firstNonRepeatingChar(s: str):
        d = {}
        for c in s:
        if c in d:
        d[c] = d[c] + 1
        else:
        d[c] = 1

        for c in s:
        if d[c] == 1:
        return c
```

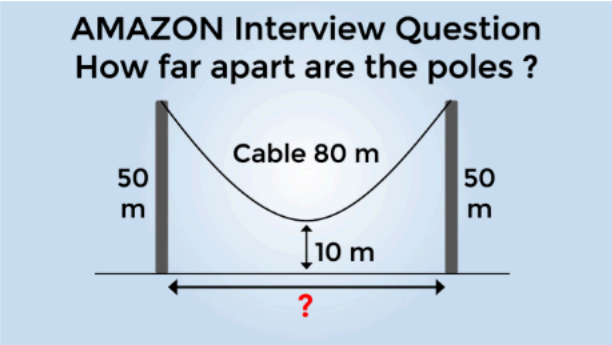
Check if the linked list has a loop (cycle)

```
  class LinkedList:
  def __init__(self,val):
  self.val = val
  self.next = None

def checkLoop(LinkedList head):
        slow = head
        fast = head

        while slow and fast.next:   #It should be fast and fast.next
        slow = slow.next
        fast = fast.next.next

        if slow == fast:
        return True

        return False
```

Puzzle

You have some oranges and you are travelling. You give away half the oranges at every station and take back 1. At the 17th station, you are left with 2 oranges. How many oranges did you have initially?
I was going in the direction ... log2(2^17) ... 2^4 for 1 additional orange

I was going in the direction -> log2(2^17) - 2^4 for 1 additional oranges
But with hints given by the interviewer, I was able to tell the correct answer = 2.
The answer is independent of the no of stations.



https://www.geeksforgeeks.org/two-poles-and-cable-puzzle/
Values given height of pole = 40, length of rope = 60, distance from bottom = 10
With a lot of hints, I was able to tell the answer 0.

Design a Quiz Game

A person can play the quiz 3 times a week.
The game runs for 4 weeks.
Per every session, the player will be asked 20 multiple choice questions with 4 options and 1 correct answer
A player will get 5 minutes per session to answer the 20 questions
Each correct is worth 1 point and the wrong answer 0 points.
Randomly choose questions from a question bank provided.
Need to print daily as well as weekly leaderboard.
The winner is the person with the highest score on the session. If scores are tied, the person who took less time to answer the quiz.

My solution:

Question_bank

Ques_id, question, options, correct_option

Persons

User_id, week_id, day, chances_left, highest_score, time_taken, session_start

Leaderboard

User_id, week_id, score, time_taken, day

Round 4 Oracle (Director Round)

Govind Lakkoju (Director, Big Data - Gateway, Scheduler, OTMM, GoldenGate Testing)

- About my role
- What is your pipeline doing that ingests data to BigQuery?
- Python/SQL rating
- How exactly Cloud functions work?
- How would you test the functionality of cloud functions?
- How would you test BigQuery?
- Frameworks that you would use for testing?
- What are you looking for?

Round 5 Oracle (Hiring Director Round)

Raghava Prasad Madhavapeddi Sai Veera (Director, Oracle)

- Postman: Testing APIs/400,500 error codes/payload format check
- PUT vs PATCH
- Faced any server side issues in APIs?
- How do you debug the BigQuery scripts?
- Exact role at Micron
- Docker commands
- B.Tech Project, CGPA etc.

Round 6 Oracle (Telephonic Senior Director Round)

Jairaj Galgali (Senior Director, Oracle)

- You have built Oracle Goldengate (data replication software) and want to test it before launching it. What tests you will write?
- Data is replicated from source to target system and data size is huge. How would you check that the data replicated is exactly same as source?
- Suppose you want to test if it will be able to handle transfer of a huge dataset, but you have limited resources in your dev environment. How will you test?
- Major challenges faced in life
- Top 3 Projects - Email writeup & discussion
- How did you verify the correctness of data after rebuilding pipelines/transformations while doing BQ Optimizations? (Ans. EXCEPT DISTINCT)

4. **CloudWerx** (Apr'24)

Round 1: Suni Mathews (Solutions Architect - AI/ML)

- Projects:
  - BQ Scheduling Framework
  - BQ Slot time
  - Dataproc cluster custom approach

- Data is getting ingested into BQ, how would you visualize it?
- Partitioning in BQ, limitations
- What are Cloud Functions? Use-cases?
- How would you show only a few columns (and secure others) of a BQ table to end users? (Using views)
- What is CDC? How does it work?
- Have you worked with relational databases?
- You are getting CSV files with changing schema, what will be your approach to store it in BQ?
- Data is coming from a source, if the records are present in the table, update it, otherwise insert it, how would you achieve this using single SQL statement?
  - Ans: MERGE statement, asked syntax and working
- CI/CD (mentioned in resume), how did you do it?
- Data is getting generated from IOT devices in real time, and being sent as a pub/sub message to GCP. How would you ingest this data to BigQuery?
  - I said using Cloud Functions/Dataflow, not sure though
- Window functions - how does it work, LEAD, LAG
- Ideal work environment?

Round 2: Jimmy Steinmetz (Team Lead/ Cloud Architect)

- Shared GCP Project link and asked to create bigquery tables from cloud storage csv files
- Queries on top of these tables
- Debugging queries
- Simple optimization techniques for given queries
- Compare current year sales with previous year sales (Self Join, LEAD/LAG window functions)
- Do you like working on multiple projects at a time or on a single project?
- My role in current company and day to day work