# Company: Credit Suisse

- General Questions
    1. Introduce yourself.
- Machine learning
    1. Explain your project.
    2. Name of ml model you used and explain that model and why you this model?
- Big Data
    1. Name the big data technologies you know.
    2. Explain hadoop architecture? and explain the map reduce flow?
    3. Explain HDFS. Is hadoop and hdfs is different?
    4. Explain spark and what is spark?
    5. Why we industry used spark nowdays, why not hadoop?
    6. What is rdd?
    7. What is transformations? explain different types of transformations?
    8. What is lazy evaluation?
    9. What is action in rdd?
    10. What is difference between dataframe and rdd?
    11. Why we used dataframe why not rdd as it is a wrapper on rdd?
    12. Explain spark architecture?
    13. What is driver? and work of driver?
    14. Which type of spark cluster do you used during cdac?
- Java
    1. Explain OOP in java with the help of examples?(tip: keep ready minimum 2 real life example of oop).
    2. How to implement encapsulation?
    3. What is difference between interface and abstract class?
    4. Explain functional interface and its types?
    5. What is abstract method?
    6. What are collections are available in java collection framework?
    7. What is lambda function in java?

8. How data type of input and output variable is decided in lambda function in java?
9. What is exception? How to handle the exception?
10. What are different types of exception?
11. What are the different types of access specifier?
12. Can we declare a private constructor? Is it accessible outside the class?
13. Write a function to find out the power of given number in java or python?

- Database
  1. What is normalization in sql?
  2. Explain different type of normalization forms?
  3. Difference between primary key and unique key?
  4. What are the different types of constraint?
  5. Write a query to find 3rd highest salary?
  6. Write a query to find out the names of emp who are not a manager?

---

## Company: Nitor Infotech

- General
  1. Introduce about yourself.
  2. What was your project?
- SQL
  1. What is index,views,joins,subquery?
  2. How can you optimized the query?
  3. Select * from emp where ename="King"? <-- order of execution in query
  4. How will you print the values from table B which are not matching to table A?
  5. What is difference in output of left and inner join?
- Python:
  1. Input: [None, 1, 2, 3, None, 4, 5, None, None]. Write code to convert the list a as except first none, other none must be replace with previous num.
     Output: [None, 1, 2, 3, 3, 4, 5, 5, 5]
  2. list = "aaabjguyvjkbcskbks". Print all the letters with there count

---

# Company: Dataeaze Infotech

- General
    1. Introduce about yourself.
    2. What was your project?
- Database
    1. Consider table structure and write given queries.
        - Table 1 : orders (order_id, product_id, customer_id, order_timestamp, order_amount)
        - Table 2 : product_master(product_id, product_name, product_category)
        - Query to get total count of unique customers on 5th Jan 2021
        - Query to get product category wise total count of orders and total transaction amount
        - Query to get top 5 products by total transaction amount
        - Find out those products which have entry in master but do not have any order at all
    2. What is window function in sql?
    3. How will be row_number in results when use partition in window?
- Python
    1. Find the 5th highest num from the list without inbulit func.
    2. Reverse the string
    3. Differnce in numpy array and list? which one is fast and why?
    4. If we install pandas as pip install pandas, can we use numpy array?
    5. Full form of CSV?
- Cloud and linux:
    1. How to create EC2 instance?
    2. How to connect to server?
    3. What is S3?
    4. Command to check background process?
    5. Command to find the file from directory?
- Bigdata:
    1. Why bigdata came into picture?
    2. What is namenode and datanode?
    3. What is default replication?

4. What is hive?
5. What is OLAP and OLTP?
6. What is Managed and External table? Which one is fast?
7. What is partitioning and bucketing?
8. What is Spark?
9. Differnce in spark and hadoop?
10. What is Spark session and Spark context?
11. What is Kafka? Applications?
12. What is HBase? Main components?

---

## Company: Bajaj Finserv

1. Self-Introduction.
2. Project in details. (It seems that the interviewer was not interested in the project, as it was on deep learning.)
3. What is python?
4. Which packages you used in your project? (As Flask mentioned in JD I focused on flask Python web framework)
5. OOPs concept. (Inheritance is-a relationship and has-a relationship with example. (Give example like Dog is-a animal and Dog has-a tail)
6. Major pillar of oops.
7. Difference between Java and python language.
8. Which language you like, why? (Post is for Python developer and Data engineer Hence answer is Python)
9. If you got chance to work on java project then can you able to do that? (Yes, I have done the modular course of core java from SunBeam and also, I learned java in my DBDA course. And as I told you I am quick learner, If I found that I am lagging in java then I will brush-up my java skills.)
10. What is Exception handling. (Explain try, except, finally and else block)
11. List and tuple difference. (Along with mutable and immutable also told memory difference in tuple and list. List consume more memory and mutable whereas tuple consume less memory and immutable)
12. List and array different in python.
13. Have you worked on python? (You can tell him about hackathon activity and PG-DBDA project)
14. Why you used Jupiter notebook in your project.
15. What is RDBMS?
16. What is SQL.

17. Views also asked to write down syntaxes (Simple View, Complex view and Materialized views)
18. What is Trigger.
19. Explain Stored procedure. (Very important questions).
20. Difference between stored procedure and function in SQL.
21. Indexing in python. Why indexing is used and how will you create index on column.
22. Clustering index and not clustering index. (Unfortunately, I didn't answer such easy questions)
23. Join in SQL (inner join, left outer join, right outer join and full outer join. Explained by vein diagram.)
24. And they had given two table asked to find left- and right-hand joins output row number.
25. Public, protected and private in python.
26. And you done big data then what is big data.
27. What is Hadoop.

- Also asked few HR type questions during tech interview:
    1. So, as you mentioned you worked as branch team leader so what was your work in Previous company.
    2. You did electrical engineering, then job in finance and now switching in IT sector. Why? Didn't you get any opportunity in your electrical field?

---

## Company: Mixed

- Bajaj Finserv, Amdocs, Bank of America, Neiron India, Relience Jio, Societe Generale.

- Python

    1. Difference between list and tupple.
    2. Difference between list and array.
    3. Can we implement interface in python like java does?
    4. Fibonacci Series upto n number.
    5. Sum of first n prime numbers.
    6. Whether given no is Armstrong or not
    7. Find no of occurrences of specific letter in a string.
    8. How will you find perticular letter in string.

- RDBMS

1. Difference between mysql and nosql.
2. How to insert records in nosql collections.
3. Types of joins in mysql and explain with example.
4. Explain ACID properties in mysql.
5. What are TCL in mysql.
6. Difference between inner join and outer join.
7. Difference between Union and inner join.
8. Explain Cartesian product with real life example.
9. Exaplin different types of index in mysql.
10. Find employees from table whose salary is greater than manager salary.
11. Find duplicate records from table.
12. Find employee from table with particular name.
13. How to select first or last 3 letters of name.
14. Demonstration of inner join.
15. Find employees from table who work as assistant manager and having salary greater than 50000.
16. Change gender of employees from M to F or F to M whose salary is gt than 50000.
17. Two separate table were given 1. Employees details 2. Date of joining. Change job role to manager from assistant manager whose date of joing is in 2020.
18. In How many ways we can find 2nd highest salary from table. Explain each one.

- Big Data Technologies

    1. Difference between Hadoop and Spark.
    2. Explain partitioning and bucketing with example.
    3. Explain Use case of repartitioning and coalesce in spark.
    4. Write spark code for word count program.
    5. How do you perform join on two dataframes in spark.
    6. Explain Architecture of spark.
    7. What do you mean by parallel processing?
    8. How to copy data from one dir to another on hdfs.

- Linux and cloud computing

1. How to rename a file.
2. Advantages of linux.
3. What do you mean by open source?

- Java

1. Explain oops concepts with example.
2. Explain polymorphism with java code.
3. What do you mean by run time polymorphism.
4. Why we implement interface in java.
5. Advantages of java over other languages.

- Project

1. Explain your project in details.
2. Explain which technologies have you used in your project.
3. How you deployed your model.
4. Why you have used aws and not azure.
5. Why you have used only this technologies and not other.
6. Your model if in offline learning right now. How will you convert it into real time data.
7. What are future scope of your project that you not implemented now.
8. Which cleansing process you have followed in your project.
9. Any database used in your project.

- HR Round

1. Tell me something about your self in brief.
2. Why do you want to come in IT.
3. Why you left your old branch?
4. What is your dream company.
5. What do you know about our company.
6. Which are the projects you have completed till right now.
7. Tell me something about your previous job.

8. What are your hobbies.
9. Are you ready to relocate.
10. Do you have any questions.

---

## Company: Mixed

- Strong sql concepts
- Questions related to duplicate records -- Find duplicates, Delete duplicates
- Questions related to Data warehousing.

---

## Company: Ecelrx

- JD: Web scraping

1. What are joins and types of joins write its output for 2 given tables ?
2. what are acid properties?
3. nth highest salary?
4. what is unique key?

---

## Company: Vanderlande

- Difference between list and tupple ?
- Difference between OLAP and OLTP?
- As you mentioned in your resume about data collection so tell me how you collect data?
- Which one will you prefer if you have access to dbms and webpage to collect data ?(same amount of data on both database and web page )
- Emp table (id,first name,last,name sal) has primary key on which column you will put index?
    - In emp table find/count records having same first name ?
- Do you know Redshift?

---

## Company: Stack nexus

- print diamond pattern of stars.

---

## Company: exucia

- Count no of duplicate records
- Display only duplicate records exclude orignals

---

## Company: C-DAC

- Data types of python
- What is F1 score?
- What do you think is most important part of big data?

---

## Company: SAS

- Introduction
- OOP
    1. What are pointers?
    2. What is overloading and overriding?
    3. What is abstraction?
    4. What are sorting algorithms?
    5. What are the limitations of inheritance?
- Python
    1. What are lists and tuples?
- DBMS
    1. What is the difference between primary key and composite key?
    2. What are different joins?

    3. How to get 2nd Highest salary from a salary column?
    4. What is the importance of indexing?
- Cloud
    1. What were the services you learn from cloud?
    2. what is SAAS?
- Data Visualization
    1. What is row level security?
    2. What DV techniques you have used?
- Machine Learning
    1. What is supervised and unsupervised learning?
    2. What is Decision tree ?
- UNIX:
    1. How to execute python code using UNIX?
- Project (ML Project):
    1. What was your role?
    2. As a developer which things you considered while doing project?
    3. What algorithms you have used in project?
    4. What OOP concepts were implemented in project?
    5. Which problems you faced during implementing your project?
- BE Project:
    1. What was your role?
    2. What algorithms you have used?
- Scenario based:
    1. As a developer if you are implementing Whatsapp, which things you'll consider?
    2. If you get a hands-on project, what will be your main concern as a developer?

---

## Company: Unknown

- Python
    1. What is list and tuples?Explain difference between them.

2. What is urllib?
3. What is request?
4. What do you prefer to use when you have to perform insert or delete operations on sequence of elements?(List, Tuple or set)
5. What is dictionary?
6. How do you declare dictionary?
7. How do you declare set?
8. What is iterator?
9. What is lambda function?
10. What is difference between a normal function and a lambda function?
11. If you have sequence of 1 million elements and you want to insert an element at its first position what will you prefer?(list, tuple or set)
12. What is caching?
13. What is JWT?
14. What is dataframe?
15. Can you traverse through dataframe?
16. Can you add a new column to dataframe?
17. How will you create a new column in dataframe that have addition of two others columns in the dataframe? Suppose you have first_name and last_name as two different columns and create a new column that contains full_name?
18. Can you convert dataframe to csv and xlsx file?
19. How can convert dataframe to csv and xlsx file? (functions and packages used).
20. What are different functions for performing EDA?
21. What is the function to get the dimensions of dataframe?
22. What is pandas?
23. What is numpy?
24. Name the data visualization packages in Python.
25. What is difference between set and dictionary?
26. What is webscraping?
27. Name packages used for web scraping.
28. Which methods are provided by BeautifulSoup for scraping the data?
29. Can we scrape data in BeautifulSoup using xpath?
30. Can we scrape data by using CSS?

31. Is scraping legal?

- SQL
    1. What are joins? Explain them in detail.
    2. What are different functions provided by SQL?(Aggregate and Scalar functions)
    3. What are stored procedures?
    4. What is view?
    5. Difference between group by, having and order by.
    6. What is SQL injection?
    7. Write a SQL query to print emp_name whose first name starts with 'A'.
    8. Write a SQL query to display second highest record in the table.
    9. What is cursor?

---

## T systems

1. Write a spark code to load csv and project 2 columns of table?
2. write a spark code to fetch employee name with its manager name? Hint: using spark sql/dataframes.
3. Explain your project?
4. Have you done any project in spark?
5. Write a Spark code to join 2 table and fetch results?
    - Join -- shuffle join?
    - Join -- broadcast join?

```
emp.join([emp.deptno == dept.deptno], broadcast(dept), "inner")
```

6. What is Checkpoint in Spark?
7. Differnce between managed and external table?
8. Explain Architecture of Spark ?
9. How a DAG is processed in spark?
10. write() is transformation or action? spark.write.....save();

11. In Hive, if in a table there are 4 columns and you are load data having only 3 columns in it, so what will be the output?
12. What is Bucketing ?
13. What is ETL and ELT ?
14. Explain Mapreduce in Yarn ?
15. what is Lazy evaluation in RDD?
16. During transformation, How previous RDD is stored?

---

## Demandhelm

1. What is Spark?
2. Explain Spark Architecture. What is Spark Master, Worker and Driver?
3. Hadoop Computing Engine vs Spark.
4. Spark vs Pandas. Why are people switching from Pandas to Spark?
5. Explain Spark RDDs and DAG in detail. Why is RDD fault tolerant? How exactly DAG works?
6. Python vs Java, Which language do you prefer?
7. What do you know about Data engineerng?

---

## Motifworks

1. Have you Done Data Modeling for Data warehouses?
2. What is Star & SnowFlake Schema? Explain both in detail. Which one is more denormalized and why?
3. What is Structure behind RDBMS and Data warehouse? Row based storage or column based storage?
4. What is Facts and Dimensions?
5. Can we have Primary and Foreign Key relation in Star Schema?
6. Can a single primary key from Fact table be present in more than 2 Dimension table?
7. What is airflow?
8. What are operators and connectors?
9. Can operators be scheduled parallely?
10. Can a single operator be used for multiple customers?

---

# Pattern Technologies

1. what is airflow? How to use airflow? Can you write a code for it?
2. Explain airflow architecture.
3. What are operators in airflow? types of operators.
4. Schedule task1, task2, task3, task4 such that task2 executes after task1. Task3, task4 executes after task 2.
    - What if a task in a dag fails?
    - Can you make sure that the successor task should be evaluated even if the predecessor fails?
5. Differnce between Hadoop 1.x and 2.x.
6. What is YARN?
7. Explain spark rdd and dataframes. Can you write a code in any of the two and which one will you choose and why?
8. What are facts and dimensions. Give real life examples.
9. What is slowly changing dimensions (SCD) and explain it.
10. what is amazon s3?
11. what is route 53 (Project related)?
12. which AMI have you used in project in ec2?
13. which instance type you have used for project?
14. what is a serverless functionality?