

Round 1:

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
3. Explain your project ?
4. Have you done any project in spark ? ans: I said movie recommendation system that was told by sir in class. so, explained it.
5. Why you decided to switch to IT Domain ?

Round 2:

1. Write a Spark code to join 2 table and fetch results ?
2. Write a SQL Query to find avg of salary department wise with rank given to them ?
3. What is Fact and Dimension ?
4. What is Checkpoint in Spark ?
5. Difference between managed and external table ?
6. Explain Arch of Spark ?
7. How a dag is processed?
8. Write() is transformation or action ?
9. In hive if in a table there are 4 columns and you are load data having only 3 columns so what will be the output ?
10. What is Bucketing ?
11. What are the services in GCP ?
12. Explain Compute services in GCP?
13. Write a python code reverse a list without using list slicing and reverse() ?
14. What is ETL and ELT ?
15. What is primary key and foreign key in SQL ?
16. What is Hlog in Hive ?
17. Have you used any AWS services ?
18. Explain Mapreduce in Yarn ?
19. In which Domain you are interested ?
20. Why you have used EC2 in your project?