



AVD GROUP
EDUCATION · EXCELLENCE · EMPLOYMENT

SCENARIO BASED QUESTIONS

1. Suppose we have a HDFS utility where you have to work with java. HDFS utility help us to create folder, delete folder . can you work on that?
2. Let's say you have table, daily partitioned with date. And in that table we have five years worth of history. so we will have close to 1800 partitions. if you want to perform bucketing on this , how will you do that.
3. Let's say you have 50 products that bank offer and it is 5 years worth of data and if your partitioning on daily basis, whatever bucketing you do will have 18x partitions. In this case aren't we adding the burden to name node by creating separate partitions for each bucket?
4. if your using vectorization, what is the target file format?
5. Project explanation
6. What is the data pipeline? how do you build it?
7. How do you decide the number of buckets for the given data. and if it is continuously increasing then how will you decide?
8. Let's say your getting 500 gb of data. And you tuned according to that. Now after sometime, may be a week after or a month you got data which is 1000gb . Now earlier you used few parameters (executors,drivers,cores etc...), will that work now.

how will you configure your cores, executors and memory.
9. If any job, if it fails, then what will be your approach.
10. Have you faced any errors in hive? And how you overcome with that. What kind of issues have you faced.

11. Tell the basic transformations and actions you worked on.
12. Hive table creation syntax, and how to load the data into it
13. What types of joins have you used in your project
14. How do you do select statement on partition table
15. If you do "select * from table " on partition table, will you be getting the benefit of partitions.
16. Select * from table --> which will run map or reducer? what is a key and value for this here.
17. Select * from table where id=1 ---> which will run here? map or reducer?
18. If you don't give any split or mappers in Sqoop what will happen?
19. Can we do partition using employee id.
20. How to configured any reducers in Sqoop?