How we will remove duplicates in Hive?

What is data structure?

What is your role in your project?

Can we do incremental job in Sqoop?

Can we do First bucketing then partitioning in Hive?

Will you do full loading or incremental loading in Sqoop?

Explain workflow of your project?

How much are you getting, size of cluster, configuration of data node?

Write the syntax of incremental loading?

can we change the password in sqoop command ?

Explain spark architecture?

Which type of tables are you creating in Hive?

Which type of data format you have used in your project?

What is partition and type of partition?

What is bucketing?

What is RDD?

What is transformation and actions?

What is dynamic partition?

What is persistence cache in spark?

Which syntax will you use to get the update data in which added in sqoop?

How many columns do you have in your project?

Which ecosystem you are using in your project?

What is static partitioning and dynamic partitioning?

Difference between order by and sort by?

What are the optimization technique you used in your project?

Which partition are you using in your project?

How you start your work everyday in your company?

If your record is in your HDFS,I want to delete or execute some ,tell me syntax.

If data is in web server how will you be transfer to HDFS?

From where you are importing and where you are storing?

What is the need of exporting?

In HDFS in which format data will stored?

Tell me about the two types of tables in Hive?

How to copy data from one cluster to another cluster?

What are the technical challenges you faced in your project?

What type of join you have used in your project?

What is the diff between partitioning and bucketing and why this is used ?

How to see size of data in HDFS?

Have you deal with XML data?

What is Oozie?

Why spark and why not map reduce?

Diff between RDD and Dataset?

Find sum and average marks of 3 students in spark?

Why RDD is lazily evaluated?

Which Hive version you are using?

How to reduce the number of mapper and reducer in hive?

Diff between external and managed table?

How many records will be loaded if daily you are receiving 50 GB data?

If we delete the source path of external table, then still will it be there in HDFS ?

Explain the components of Kafka and why kafka is using?

What is offset in Kafka?

In kafka the flow of data will be synchronous or asynchronous?

How to get one row in Hbase, and why Hbase and how data is stored in Hbase?

What is work of a broker?

Which word is used fro bucketing?

What are the spark API's?

Create 2 table A and B, place the data in B from table A which is not there in B?

What least comprehension technique in python?

What are the data types in python?

What are the basic diff between python and scala?

I have 100 tables ,i don't want 10 tables write the sqoop command for this?

What is Hive analytics?

WAP in python to find fibonisys series?

WAP in python to find prime number?

Are you doing import or export or both in your project?

What is configuration file in Hive?

In which format data will be stored in bucketing?

Developer don't have the authority to delete one table then what the exact need of e external table?

Diff between Hadoop 1.x version and Hadoop 2.x version

How to load the reference table in Hive?

Tell me some command in HDFS?

I want to see all tables with all columns in sqoop?

Tell me some real time example of partitioning and bucketing?

How to delete one particular row in Hive?

How to add a column in Hive?

In which format data will be stored in hive and tell me path to see the content?

How to change block size of HDFS?

Why we need partitioning?

Explain architecture of HBase?

In sqoop how to reduce the number of mapper and reducer in Sqoop?

While executing sqoop only mapper file will be create or reducer also?

Can we change the syntax of table if already one table is created in Hive?

Diff between copy,copy from local and move from local and tell me the diff and e uuse of each one ?

Can we create primary key in Hive?