1. what is map and reduce.
2. **Tell about your experience and job profile? your role in the project. where you have worked and what you have worked.**
3. where you have used spark streaming in the project?
4. what is the hadoop ecosystem you have in your project?
5. what is the size of your cluster/node size.
6. have you created any udf's in hive. how can you create them.(what class it extends and what it implements )
7. what are the problems you faced in scheduling the oozie jobs?

**Progapes:**

1. what are the parameters of map functions?
2. how to set no. of reducers?
3. how mapper can find the data locality?
4. What is Speculative execution, Does the slower older mapper will be stoped or run continuesly if the new mapper is re-launched by the application master?
5. how do file write happens on hdfs(explain procedure)?
6. what is hive?
7. what is default metastore database?(derby)
8. can we change the default database in hive. if so how can we change it what are the configuration changes need to make.
9. what is size of your data you get daily?
10. what are different types of tables are there what are their difference?
11. how many types you can create RDD ? what are they?
12. how do you define namenode high availability?
13. what type of file formats you get in your project?
14. what methodology you used in your project?
15. **what is your daily activity in your project?**
16. sparksql
17. how can you recognize failure in your spark job?
18. where does the mapper's output is stored?
19. what is casewhen, casewhen-else in sql?

**Polaris:**

1. how you rate your java language?
2. collection frame work
3. hashmap and hashtable difference?
4. object,static, difference between exception and error?
5. spark dframes object. how you used
6. what is the mapreduce framework?explain in terms of mapper reducer.
7. what are different oozie schedulers.
8. what is the yarn scheduler. what is capacity scheduler?
9. how do you optimize hive.
10. what is external table in hive.
11. can you control the no of mappers ,reducers.how?
12. if I have two reducers what will be my output?
13. can we create two spark context objects in a jvm?
14. what is row key.
15. what is the difference between column family and column?
16. what is hcatalog?
17. what is the difference between broker and topic.
18. can we run kafka server without zookeeper.
19. what is qarum(zookepeer).
20. what is persist() in scala.
21. which comes first transformation or action?
22. what is the difference between transformation and action?
23. can you perform aggregation on hbase?
24. what are the noSql databases you know?
25. network optimization in spark can be achieved.
26. what is broadcast variable?
27. how to you start your spark scala(initial step)
28. difference between map and flatmap in scala.
29. what is lamda in scala.
30. what is the difference between scan and get in hbase?

**HCL**

1. What is lazy evaluation?
2. what is an rdd?
3. hive v/s rdbms, hive v/s pig
4. how do you rate yourself in all the technologies?
5. what is distributed cache?
6. diff mapreduce v/s spark,
7. does your data store in database or hdfs?
8. how you perform incremental dataload in sqoop
9. reduce is transformation or action?
10. can we compress data while importing?

**SYNECRON**

1. diff if-else in scala and java, traits in scala and interface in java, mapreduce and spark, mapreduce and rdbms, hive and hbase, val and var in scala.
2. what are the classes in scala?
3. different modes of hbase?
4. what is hbase?
5. what is the initial phase of reducer?
6. what are the libraries in spark and what do you like to work with.
7. what are the configuration setting you do in mapreduce program?

**WIPRO**

1. Tell about your latest project?
2. where did you use spark in your project?
3. how did you ingest data into spark from hive?
4. how can you tune the performance in spark?

**SpringerNature**

1. How do you rate in Scala And Java ?
2. How do design this:- we have an elevator and 6 buttons along with that to go to the particular floor, so how do you class level and higher level design for this?
3. what are case classes?
4. what are object?
5. What is apply and unapply method?
6. If I have traits a,b,c having the same function name , will it give me the problem?
7. how does the case class work means if we talk in higher level.
8. how to get data to the Hdfs from any where?
9. where do we you object, case classes, traits?
10. are working in functional scala or only scala.?

**Cisco Client Round**

1. Challenges you find in your carrier?
2. describe your project?
3. we are having different-2 loan detail from deifferent-2 sources so,

how you get the information?

1. How you get data from cibil?
2. how spark is different than Hive?
3. what is partitioning in hive?
4. how hive use partition to fetch the result?
5. can we use multiple partitions in hive?
6. what type of model that propensity is?
7. what fields did you use to process propensity model?

INCEDO Technologies:-06/07/17

1. Case class and simple class
2. performance tuning in spark
3. difference b/w Traits and Interface
4. how to add column to DataFrame
5. how to alter column in Dataframe
6. Memory optimization.

**Infosys(02/11/2017):-**

1. About you and your technologies?

2. About your project?

3. What your cluster configurations?

4. Explain the flow of your project in detail.?

5. How do you connect kafka to API(Job Attempt API).

6. How much data is there?

7. What is the difference between input splits and Block in HDFS?

8. How to did you maintain partitions?

9. What happen if static modifier removed from main method?