2. Tell me about your current project

3. How do u load data into HDFS

4. Which hive table is used internal or external? and why?

5. which Hadoop distribution used?

6. Which version?

7. What is cluster size?

8. where is your  cluster located ?

9. How do u access your  cluster? (from namenode or from client)

10. can we import to hive from sqoop? how?

11. what is input split? How to configure it?

12. How to read file formats(other than textfile) in pig?

13. What is Partitioning in hive?

-----------------------------------------------------------------

14. Generics in java, Collections

15. Difference between java serialization and Hadoop serialization?

16. What is Immutable Classes ?

. What is Dependency Injection

2. How do you inject the values from one object to another.

3. What are Runnable Interface and Callable Interface

4.There are 10 billion records. And the new data adds up every day. How do you develop a system to store, read,

update the data.

5. Give an use case for a Tree collection framework.

6. Can we operate HBase on any other file system other than HDFS.

7. In Hive, while partitioning, how the data is stored. In the same folder or separate folders.

8. What is the use of Combiner. Difference between combiner and reducer. In what cases we don't use Combiner.

9. Difference between Join and Inner join.

10. Can we combine two relational datasets without using Joins

11. Have you used UDFs in PIG Latin and how did you use them.

12. A Hash code method always returns 1. What will happen in that case.

13. Have you worked on Springs.