1.Describe any five characteristics of Big Data.

2. Explain Hadoop Ecosystem

3. What is a data stream?

4. How Big Data and Hadoop are linked?

5. What is MapReduce ? Explain How Map and Reduce Work? What is Shuffling in MapReduce?

6. Explain Word-Count mapreduce implementation.

7. How to count distinct elements in a stream? Explain Flajolet-Martin Algorithm.

8. What is Real Time Analytics? Discuss their technologies in detail

9. Explain Hadoop. List the core components of Hadoop.

10. What are Combiners? When should one use combiner in MapReduce job?

11.Explain Recommendation system and its use.

12.Explain HDFS and its basic commands

13. how to install hadoop sigle node/multi node cluster on ubuntu.

14. What is the difference between Hadoop and Cloudera?

15. What is hortonworks sandbox?

16.How to activate all services on cloudera?

17. What are the relational algebra operations in MapReduce? explain all

18. Explain Matrix Multiplication using Mapreduce.

19 . Difference between NoSql,RDBMS

20.Explain CURD operation in MongoDB?

21. Which are the different languages supported by MongoDB?

22. Is MongoDB better than other SQL databases? If yes then how?

23. Explain how to count ones in a window using DGIM algorithm

24.What are the filters in big data? Explain all.

25. Define the difference between Hive and HBase?

26.What kind of applications is supported by Apache Hive?

27. Where does the data of a Hive table gets stored? Why Hive does not store metadata information in HDFS?

28. Explain the uses of Map Reduce in Pig.

NOTE :

- **Prepare : Exp 1, Exp 3 and Exp 8 (check notes and Videos)**
- **Oral question based on Mini project topics**