SARDAR VALLABHBHAI PATEL INSTITUTE OF TECHNOLOGY, VASAD.

Information Technology Department

Subject Name: **BIG DATA ANALYTICS (2171607)**

1 What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram.

2 What is Big Data? Explain characteristics of Big Data.

3 What are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities.(Discuss one application)

4 Draw HDFS Architecture. Explain any two commands of HDFS from following commands with syntax and al least one example of each.
(i) copyFromLocal
(ii) setrep
(iii) checksum

5 Write Map Reduce code for counting occurrences of specific words in the input text file(s). Also write the commands to compile and run the code.

6 Explain working of following phases of Map Reduce with one common example.
(i) Map Phase
(ii) Combiner Phase
(iii) Shuffle and Sort Phase
(iv) Reducer Phase

7 Explain Job Scheduling in Map Reduce. How it is done in case of
(i) The Fair Scheduler
(ii) The Capacity Scheduler

8 What is Big data? Discuss it in terms of four dimensions, volume, velocity, variety and veracity.

9 Define HDFS. Discuss the HDFS Architecture and HDFS Commands in brief.

10 Discuss big data in healthcare, transportation and medicine.

11 Discuss Hadoop YARN in detail with failures in classic Map-reduce

12 Explain Map-reduce framework in detail. Draw the architectural diagram for Physical Organization of Compute Nodes.

13 What is big data analytics? Explain four 'V's of Big data. Briefly discuss applications of big data.

14 What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram.

15 What is Hadoop Ecosystem? Discuss various components of Hadoop Ecosystem.

16 What is data serialization? With proper examples discuss and differentiate structured, unstructured and semi-structured data. Make a note on how type of data affects data serialization.

17 With suitable block diagram explain architecture of HDFS. Discuss role of Data node and Name node in HDFS. Give commands with appropriate arguments to perform data transfer between local file system and HDFS.