Nirma University

Institute of Technology

Semester End Examination (IR), December - 2021

B. Tech. in Computer Science and Engineering, Semester-VII

2CS702 Big Data Analytics

Roll / Exam No. [ ]

Supervisor's Initial With Date [ ]

Time: 2 Hours

Total Marks: 50

Q:1 Answer the following questions: (6 X 3) **[18]**

1 Discuss the detail scenario of reading data from Hadoop Distribute File System with diagram.

2 Define "Big Data". Share your understanding of big data for the application in parliament (government sector). Also explain how big data analytics help higher authorities in parliament to take important decisions.

3 Why do you need scaling of data? Discuss any two platforms available for vertical scaling of big data.

OR

3 Explain the need for Apache Cassandra. Also describe its features by taking suitable application.

4 Velocity of data generates big data. Explain this statement by taking suitable example.

5 Explain the need for Apache Spark. Also describe its features by taking suitable application.

6 Describe the role of record reader in map reduce program execution. Also explain the types of record reader with example.

Q:2 Answer the following questions: (4 X 4) **[16]**

1 If you are given the data of age of the person. **(Any one)**

1) Write down the reducer logic/pseudo code for generating even numbers and odd numbers in a separate output files.
2) Write down the combiner logic/pseudocode in order to reduce the burden of reducer for the above case.
3) Write down the partitioner logic/pseudocode in order to reduce the burden of reducer for the above case.

2 Write down one line importance of following.

a) Application Master in YARN
b) yarn-site.xml
c) Edit Log
d) hdfs-site.xml

3 Describe the nature of applications for which Apache Hive is most suitable. Also define the functionality of apache hive.

4 Explain the role of number of iterations in choice of platforms. Which platforms are most suitable for iterative nature of program?

[P.T.O.]

OR

4 Write down the role and responsibilities of sorting and shuffling phase in map reduce program execution.

Q:3 Consider the employee dataset in mongodb. Each collection contains employee id, name, department, joining date, pay scale, promotions and description. Insert appropriate documents into database. Write down the query for following scenario. Assume necessary data for following query. **[16]**

1) Get all employee data who joined before the year 2000 or after 2010.
2) Add an employee named "Sunil" to the employee "Anil ".
3) find all employees that have a description that contains the word "Good" and not the word "bad"
4) delete the employee record of "Ramesh"
5) find all employee details who get at least three(3) promotions till date.
6) get all documents where department include "Production & Control"
7) find all employee details whose pay scale greater than 40k and joining in last three years.
8) Sort all the documents according to employee joining date.