Nirma University

Institute of Technology

Semester End Examination (IR), February - 2022

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 Consistency and Partition tolerance (CP) or Availability and Partition tolerance (AP) is possible in NoSQL databases. Why Consistency and Availability (CA) is not possible in NoSQL?

2 Define "big data". Share your understanding of big data for the application in education sector. Also explain how big data analytics help higher authorities in education sector to take important decisions.

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

OR

3 "Variety of data generates big data". Explain this statement by taking suitable example.

4 Clarify the need of combiner function and partitioner in map reduce programming. Take the case of word count problem and discuss in detail.

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

6 Discuss the detail scenario of writing data to HDFS with diagram.

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

1 If you are given the data of year and its temperature. How would you specify the partitioner function in order to achieve the maximum global temperature of the year? Discuss the importance of partitioner function.

OR

1 If you are given the data of year and its temperature. How would you specify the combiner function in order to achieve the maximum global temperature of the year? Discuss the importance of combiner function.

2 Write down **one line** importance of following.

a) Data Queue

b) Resource Manager in YARN

c) FSImage

d) hdfs-site.xml

OR

2 Explain CAP theorem in detail. Compare NOSQL database with RDBMS.

3 Compare and contrast horizontal scaling and vertical scaling.

OR

3 Explain the concept of network topology in HDFS.

4 Describe the term "fault tolerant"? Which platforms should we choose for fault tolerant applications? Why?

Q:3 Consider the movie dataset in mongoDB. Each collection contains movie title, writer, year, actors, franchise, ratings, and comments. Insert appropriate documents into database. Write down the query for following scenario. Assume necessary data for following query. **[16]**

1) Get all movies released before the year 2000 or after 2010.
2) Add an actor named "Samuel L. Jackson" to the movie "Pulp Fiction".
3) find all movies that have a synopsis that contains the word "Bilbo" and not the word "Gandalf"
4) delete the movie "Avatar"
5) find all movies that have a synopsis that contains the word "Bilbo"
6) get all documents where actors include "Brad Pitt"
7) find all movies that have a synopsis that contains the word "dwarves" or "hobbit"
8) Sort all the documents according to movie name.