[07 BENG - 4117]

IV/IV B.Tech. DEGREE EXAMINATION

Computer Science and Engineering

First Semester

Professional Elective

BIG DATA ANALYTICS

(Effective from the admitted batch of 2020-2021)

Time : 3 hours Max. Marks : 70

Question No. 1 is compulsory.
Answer any FOUR from the remaining.
All questions carry equal marks.
Answer all parts of any question at one place.

1. Write short notes on the following:
   (a) List the components of Hadoop system.
   (b) Describe few salient differences between SQL and NoSQL.
   (c) What is the role of data node in HDFS?
   (d) Describe HiveQL data definition commands.
   (e) List the advantages of using Apache Spark over Hadoop.
   (f) What is JSON used for? Give an example.
   (g) What are the components of stream processing?

2. (a) Why is big data analytics important? What are the various types of analytics? What are the key questions to be answered by all organizations stepping into analytics?

   (b) Discuss big data in healthcare, transportation, and medicine.

3. (a) Discuss about aggregate data models in NoSQL.

   (b) Discuss the following:
   (i) Materialized views
   (ii) Schema less databases

4. (a) Explain the concept of Blocks and Federation in HDFS architecture. What are the benefits of block transfer?

   (b) Why can't we use databases with lots of disks to do large-scale batch analysis? Why is MapReduce needed? Explain.

5. (a) With neat diagram explain the relationship between Hive clients and Hive services.

   (b) Write HiveQL queries to perform the following tasks with suitable examples.
   (i) Creating a table
   (ii) Loading data into a table
   (iii) Joining tables

6. (a) Discuss about a typical data flow sequence for programming spark.

   (b) What is resilient distribution dataset in Apache Spark? Explain in detail. Make a note on why RDD is better than Map Reduce data storage?

7. (a) Discuss about grouping with expressions and grouping with maps.

   (b) Explain the execution modes of spark application.

8. (a) Discuss about sliding windows on event time.

   (b) How to count distinct elements in s Stream? Explain.

* * *