SPRING END SEMESTER EXAMINATION-2018

6th Semester B.Tech

DATA ANALYTICS

IT-3002

[For 2016(L.E.), 2015 & Previous Admitted Batches]

Time: 3 Hours | Full Marks: 60

Answer any SIX questions including question No.1 which is compulsory.
The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. Answer the following questions. [2 × 10]

   (a) How do you describe the Big Data?

   (b) Is data visualization always needed in a data analytics system? Justify using suitable example.

   (c) Volume or Variety, which is challenging dimension in Big Data Analytics?

   (d) What are the modern day challenges of information visualization?

   (e) Choose a social networking site of your choice, and justify the need of Big Data analytics?

   (f) Illustrate, how triples can be used to optimally store and count pairs in a frequent itemset mining?

   (g) How to count distinct elements in a stream data?

   (h) How NoSQL is different from conventional database approach for data analysis?

   (i) What is data stream management system? Explain with Block diagram.

(j) 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

2. (a) What are the V's of Big Data? Also list various real-world big data and specify the V's of each? [4]

(b) What are the benefits of Big Data Analytics? How Big Data Analytics can be useful in the development of smart cities in India. [4]

3. (a) Explain different data model used in Hive. [4]

(b) Discuss the different distance measures, that can be used to measure the similarity or dissimilarity between data point in data analytics of big data set. [4]

4. (a) List the various challenges in stream computing? Explain any one algorithm to count the number of distinct elements in a data stream? [4]

(b) Explain the Big Data Stack and objective of each layer? Also justify the role of analytics engine in big data stack. [4]

5. (a) List the various sampling techniques for big data? Describe any two techniques with the help of real-world examples? [4]

(b) Given a 1 dimension dataset {1,5, 8, 10, 2}. Use Agglomerative clustering algorithm with Euclidean distance to establish hierarchical grouping relationship. Draw the dendrogram. [4]

6. (a) Explain four HDFS commands with its syntax & example. [4]

(b) Show Map reduce implementation for the following scenario, [4]

(i). Multiplication of two matrices with one Map reduce step.

(ii).Natural join of two relations with example.

7. (a) What is difference between Data Profiling and Data Mining? Explain by considering any 'Societal' use case. [4]

(b) List down the key responsibilities of a data analyst? What are the requirements needed to become a data analytics? [4]

8. Write short notes on the following (Any four) [4 × 2]

(a) How Data Analytics over Citizen dataset (e.g. AADHAAR) could be useful?

(b) What is the motivation to count triangles in social graph?

(c) What are the basic differences between relational database and HDFS?

(d) What is a 'Community' in 'Social networking' and how these are created?

(e) What is Zookeeper? List the benefits of it?

*****