Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 10511

M.E./M.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2019.

Second Semester

Biometrics and Cyber Security

CP 5293 — BIG DATA ANALYTICS

(Common to M.E. Computer Science and Engineering/M.E. Computer Science and Engineering (With Specialization in Networks)/M.E. Mobile and Pervasive Computing/M.E. Software Engineering/M.Tech. Information Technology)

(Regulation 2017)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is V^3?
2. List the reasons why data has to be cleansed before analyzing.
3. How are big data and Hadoop related to each other?
4. What is fsck?
5. How is Hash Table Collisions is Avoided?
6. Name any two statistical methods that are useful for data scientist.
7. What is Discrete and Continuous data in Data mining world?
8. List the issues in stream processing.
9. Identify ways in which pig programs can be executed.
10. What is Data purging?

PART B — (5 × 13 = 65 marks)

11. (a) Discuss about the available tools for databases, MapReduce, storage, servers and processing used in Big-Data scenarios with their uses. (13)

Or

(b) (i) Discuss any three applications of big data. (6)

(ii) How is big data analysis helpful in increasing business revenue? (7)

12. (a) (i) Describe the components of HDFS. (6)

(ii) What are the different configuration files in Hadoop? (7)

Or

(b) (i) Explain the process that overwrites the replication factors in HDFS. (6)

(ii) What are the configuration parameters in a "MapReduce" program? (7)

13. (a) (i) Distinguish Data Mining and Data Analysis. (6)

(ii) Discuss how the association rules are classified based on various criteria. (7)

Or

(b) (i) Find out the advantages of using R in data analysis. (6)

(ii) Identify the properties that clustering algorithms should satisfy. Discuss them in brief. (7)

14. (a) Discuss about stock market predictions. (13)

Or

(b) (i) List the rules to be followed when representing a stream by buckets. (5)

(ii) Discuss about the moments of streams, and how they are estimated? (8)

15. (a) (i) Discuss about the different tombstone markers used for deletion purposes in HBase. (7)

(ii) List Pros and Cons of Hive. (6)

Or

(b) Write short notes on Hive components with a neat diagram. (13)

PART C — (1 × 15 = 15 marks)

16. (a) (i) What are the three types of User Defined Functions in Hive? Differentiate them. (7)

(ii) What is clustering? Discuss about the requirements and types of data in cluster analysis. (8)

Or

(b) Write short notes on real time sentiment analysis. (15)