G-4/427/21 Roll No. ....

M.Sc. IV Semester Examination, 2021
INFORMATION TECHNOLOGY
Paper III

(Big Data Analytics)

Time : 3 Hours]   [Max. Marks : 100
[Min. Marks : 40

Note : *All questions are compulsory. Question Paper comprises of 3 sections. Section A is objective type/multiple choice questions with no internal choice. Section B is short answer type with internal choice. Section C is long answer type with internal choice.*

SECTION A   1×10=10
(Objective Type Questions)

Choose the correct answer :

1. The analysis performed to uncover the interesting statistical correlation between associated-attributes value pairs are known as the ....

   (a) Mining of cluster

   (b) Mining of correlation

   (c) Mining of association

   (d) All of the above.

2. The issues like efficiency, scalability of data mining algorithms comes under ....

   (a) Mining methodology and User interaction

   (b) Diverse data type issues

   (c) Performance issues

   (d) All of the above.

3. The size of data in Big Data is of .... bytes.

   (a) Giga   (b) Mega

   (c) Peta   (d) Terra

4. Big Data Analysis performs the following except :

   (a) Spreads data   (b) Analyzes data

   (c) Collects data   (d) None of these.

5. .... is general purpose computing model and runtime system for distributed data analytics.

   (a) Drill   (b) Mapreduce

   (c) Oozie   (d) None of these.

6. HDFS supports the .... command to fetch Delegation taken and store in a file on the local system.

   (a) fetdt   (b) rec

   (c) fsk   (d) fetchdt.

7. Which is the most popular NoSQL database for scalable big data store with Hadoop ?

(a) MongoDB   (b) Cassandra

(c) Hbase   (d) None of these.

8. MongoDB has been adopted as .... software by a number of major websites and services.

(a) backend   (b) frontend

(c) proprietary   (d) All of the above.

9. You can run Pig in batch mode using ....

(a) Pig option   (b) Pig shell command

(c) Pig script   (d) All of the above.

10. Which of the following operator executes a shell command from the Hive shell ?

(a) ^   (b) +

(c) | (d) !

SECTION B   6×5=30

(Short Answer Type Questions)

Note : *Attempt one question from each unit.*

Unit-I

1. Discuss Data Mining Functionality.

Or

Write the applications of Data warehouses.

Unit-II

2. Why we need Big Data ?

Or

Write note on big data challenges.

Unit-III

3. Explain any three component of HDFS.

Or

Explain any three HDFS commands with example.

Unit-IV

4. Explain with example any three operations with MongoDB.

Or

Explain the use of NoSQL database system.

Unit-V

5. Explain the benefits of Apache HIVE in Hadoop.

Or

What is the significance of Apache pig in Hadoop context ?

SECTION C   12×5=60

(Long Answer Type Questions)

Note : *Attempt one question from each unit.*

Unit-I

1. Explain the schemas of multidimensional data model with suitable example.

Or

Define Data Mining. Describe the architecture of Data Mining system.

Unit-II

2. What do you mean by Big Data ? Explain with example. Describe the classification of data with example.

Or

Differentiate big data and small data with example. Discuss the big data characteristics.

Unit-III

3. What is Hadoop YARN ? With block diagram discuss the various framework that run under YARN.

Or

Define Apache Hadoop. Discuss the Hadoop Ecosystem and Hadoop Ecosystem tools.

Unit-IV

4. Explain NoSQL data architecture and NoSQL to manage big data.

Or

What is MongoDB ? Write the features of MongoDB and explain data types in MongoDB with examples.

Unit-V

5. Describe Hive datatypes and Hive built-in function with suitable examples.

Or

Explain Pig Latin scripts and commands.

✧ ✧ ✧ ✧ ✧ c ✧ ✧ ✧ ✧ ✧