

H-4/22/22

Roll No.

IV Semester Examination, 2022

M.Sc.

INFORMATION TECHNOLOGY

Paper III

(Big Data Analytics)

Time : 3 Hours]

[Max. Marks : 100

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/Multiple Choice Questions)

Choose correct answer :

- 1.** Identify the term used to define the multi-dimensional model of the data warehouse :
(a) Table (b) Data cube
(c) Tree (d) Data structure
- 2.** What are the main components of Big Data ?
(a) Map Reduce (b) HDFS
(c) YARN (d) All of these

P.T.O.

[2]

- 3.** The type of Big data are :
(a) Structured data
(b) Semi-structured data
(c) Unstructured data
(d) All of the above
- 4.** What are the fourv's of big data ?
(a) Volume (b) Velocity
(c) Variety (d) All of these
- 5.** Hadoop is open sources :
(a) Always true
(b) True only for apache hadoop
(c) True only for apache and cloudera hadoop
(d) Alaways false
- 6.** Identify among the options below which is general purpose computing model and runtime system for distributed data analysics.
(a) HDFS (b) Mapreduce
(c) Oozie (d) All of these
- 7.** Which of the following is not a valid NOSQL database ?
(a) Cassandra (b) Scylla
(c) Hadoop/Hbase (d) PostgreSQL

H-4/22/22

[3]

8. Which of the following format is supported by mongoDB ?

- (a) XML (b) SQL
(c) BSON (d) All of these

9. You can run pig in batch mode using :

- (a) Pig shell command
(b) Pig scripts
(c) Pig options
(d) All of the above

10. The results of a hive query can be stored :

- (a) Local file (b) HDFS file
(c) both (a) and (b) (d) cannot be stored

SECTION B

6×5=30

(Short Answer Type Questions)

Note : Attempt *one* question from each unit.

Unit-I

1. Describe KDD process.

Or

Explain data warehouse architecture.

Unit-II

2. Write the difference between bigdata & small data.

H-4/22/22

P.T.O.

[4]

Or

Explain classification of data.

Unit-III

3. Define Hadoop. Write the features of Hadoop.

Or

Describe HDFS data storage ?

Unit-IV

4. Explain datatypes of mongoDB.

Or

Discuss SQL Vs NOSQL ?

Unit-V

5. Write the application of Apache Pig.

Or

What is Hive ? Write the characteristics and limitation of Hive.

SECTION C

12×5=60

(Long Answer Type Questions)

Note : Attempt *one* question from each unit.

Unit-I

1. Explain the following :

- (a) OLAP operations.
(b) Architecture of data mining system.

H-4/22/22

[5]

Or

What is data cube ? Explain multidimensional data model with example.

Unit-II

2. (a) Write the challenges of conventional system.
(b) Why we need Big data ? Explain.

Or

- (a) What is the importance of Big data ? Write Big data characteristics.
(b) Write short notes on the following :
(i) Big Data technologies
(ii) Big Data future scope

Unit-III

3. What is Hadoop Ecosystem ? Explain various HDFS commands with examples.

Or

Explain the following :

- (i) Map Reduce Frame work
(ii) Hadoop physical organization

Unit-IV

4. Explain the following :
(i) Types of NOSQL database.
(ii) NOSQL data architecture pattern.

[6]

Or

Explain mongoDB query language and database command with example.

Unit-V

5. Write short notes on the following :

- (i) Pig Latin Data model
(ii) Hive Data type and fill formats

Or

- (a) Explain Hive integration and workflow steps.
(b) Explain Hive Built-in functions and HiveQL

★ ★ ★ ★ ★ c ★ ★ ★ ★ ★