Roll No. .....

E–1065

M. Sc. (IT) (Fourth Semester) (Main/ATKT)

EXAMINATION, May-June, 2021

(New Course)

BIG DATA ANALYTICS

(403)

Time : Three Hours] [*Maximum Marks : 100*

[*Minimum Pass Marks : 40*

Note : Attempt all Sections as directed.

Section—A 1 each

(Objective/Multiple Choice Questions)

Note : Attempt all questions.

Choose the correct answer :

1. The ability to economically store and manage petabytes of data online is part of :

   (a) Theoretical Science

   (b) Data Science

   (c) Computational Science

   (d) Empirical Science

2. Which process includes data cleaning, data integration, data selection, transformation, data mining, pattern evaluation ?

   (a) Web Mining

   (b) Data Mining

   (c) Data Warehouse

   (d) Knowledge Discovery Databases

3. Which is not part of Post-Processing ?

   (a) Pattern selection

   (b) Pattern interpretation

   (c) Pattern discovery

   (d) Pattern evaluation

4. Multiple fact tables share dimension tables, viewed as a collection of stars, is called :

   (a) Fact Constellation

   (b) Snowflake schema

   (c) Star schema

   (d) None of the above

5. "Data extracted from IoT devices provides a mapping of device inter-connectivity", is part of which application ?

   (a) Crime Prediction and Prevention

   (b) Big Data Applications in IoT

   (c) Big Data Applications in Healthcare

   (d) Big Data Applications in Manufacturing

6. Which is not a type of analytics ?
   (a) Prescriptive Analytics
   (b) Predictive Analytics
   (c) Descriptive Analytics
   (d) Processing Analytics

7. Big data analytic technologies are necessary to :
   (a) Formulate eye-catching charts and graphs
   (b) Extract valuable insights from the data
   (c) Integrate data from internal and external sources
   (d) None of the above

8. Which is called to Information that is yet to be transformed into digital format ?
   (a) Dark data
   (b) Big data
   (c) Web Data
   (d) None of the above

9. All of the following accurately describe Hadoop, except :
   (a) Java-based
   (b) Real-time
   (c) Distributed computing approach
   (d) Open source

10. Which was inspired by Google MapReduce and Google File System papers ?
   (a) Hadoop
   (b) Cruise Control
   (c) YARN
   (d) MapReduce

11. Which is the architectural center of Hadoop that allows multiple data processing engines ?

    (a) Chuckwa

    (b) Hive

    (c) Incubator

    (d) YARN

12. Which Tracker is allocates work to the tracker nearest to the data with an available slot ?

    (a) Task

    (b) Process

    (c) Job

    (d) None of the above

13. Which is a not characteristic of MongoDB ?

    (a) Easy scalability

    (b) High memory usage

    (c) High performance

    (d) High availability

14. Which is used for storing documents and making remote procedure calls in MongoDB ?

    (a) BSON

(b) JSON

(c) Native

(d) Divers

15. In this process distributes data across multiple physical partitions is called ..... .

(a) Shards

(b) GridFS

(c) Servers

(d) Replication

16. Which is not wide-column stores database ?

(a) Big Table

(b) Accumulo

(c) Redis

(d) Hbase

17. The runtime engine is a compiler that produces sequences of which programs ?

(a) User Defined Functions

(b) MapReduce

(c) Pig Latin

(d) Pig Engine

18. Which is data size of Hive ?

(a) Petabytes

(b) Gigabytes

(c) Megabytes

(d) Kilobytes

19. Which of the following is a relational operator in Pig ?
    (a) DESCRIBE
    (b) DISTINCT
    (c) DUMP
    (d) All of the above

20. Which is not part of Hive Data Model ?
    (a) Table
    (b) Bag
    (c) Partition
    (d) Buckets

Section—B 2 each

(Very Short Answer Type Questions)

Note : Attempt all questions. Answer using 2-3 sentences.

1. What is Data Warehouse ?
2. What is attributes ?
3. What is Web Data ?
4. What are Big data type ?
5. What is Task Tracker ?
6. What is Hadoop ?
7. What is NoSQL ?
8. What are data type of MongoDB ?
9. What is HiveQL ?
10. What is Apache Pig ?

Section—C 3 each

(Short Answer Type Questions)

Note : Attempt all questions. Answer precisely using < 75 words.

1. Write the Data Mining Functionality.

2. What is OLAP operation ?
3. What are the challenges of Conventional System ?
4. What is difference between big data and small data ?
5. What are the features of Hadoop ?
6. What is YARN ?
7. What are the advantages of NoSQL ?
8. What are difference between RDBMS and MongoDB ?
9. What are the applications of Apache Pig ?
10. What are Hive Data Model ?

Section—D 6 each

(Long Answer Type Questions)

Note : Attempt all questions. Answer precisely using 150 words.

1. What do you mean by Data Mining ? Explain the Architecture of Data Mining System.

Or

Explain the following :

(i) Data Warehouse Applications

(ii) Data Cube

2. What do you mean by Big Data ? Write the characteristics of Big Data.

Or

Explain the following :

(i) Classification of Data

(ii) Big Data Technologies

3. What do you understand by Hadoop Distributed File System ? Write any *five* HDFS commands.

Or

Explain the following :

(i) Features of Hadoop

(ii) MapReduce Framework

4. What do you mean by MongoDB ? Write any *five* MongoDB Database Commands.

Or

Explain the following :

(i) Features of MongoDB

(ii) Difference between SQL and NoSQL

5. What is big data programming languages ? Explain the architecture of Apache Pig.

Or

Explain the following :

(i) Hive Data Type

(ii) Pig Latin Data Model