

D-1142
M. Sc. (I.T.) (Fourth Semester) New
EXAMINATION, JUNE, 2020

Big Data Analytics
M. Sc. (I.T.) (402)

Time: Three Hours]

[Maximum Marks: 100
[Minimum Pass Marks:40

Note: Attempt all sections as directed.

SECTION – A
(Objective Type Questions,)

Note: Attempt all questions.

each 1 mark.

Q1 . Which part is not in Post processing?

- | | |
|---------------------------|----------------------|
| A. Pattern evaluation | B. Pattern selection |
| C. Pattern interpretation | D. Pattern discovery |

Q2. Discovering interesting patterns and knowledge from massive amount of data is called:

- | | |
|----------------------------------|------------------|
| A. Data warehouse | B. Data Mining |
| B. Knowledge Discovery Databases | D. Data Analysis |

Q3. Structured and systematic analysis at each step before proceeding to the next is called

- | | |
|----------------|--------------|
| A. Spiral | B. Waterfall |
| C. Interactive | D. Prototype |

Q4. Data in a data warehouse

- | | |
|---------------------------------------|---|
| A. In a flat file format | B. Must be in normalized form to at least 3NF |
| C. Can be normalized but often is not | D. Must be in normalized form to at least 2NF |

Q5. Which is collective term it refers to any type of data you might pull from internet whether to for research purposes:

- | | |
|----------------|------------------|
| A. Mining Data | B. Ware Data |
| C. Web Data | D. Data Analysis |

Q6. What are the main components of Big Data?

- | | |
|--------------|-----------------|
| A. MapReduce | B. HDFS |
| C. YARN | D. All of above |

Q7. Big data analytics 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 these |

Q8. The big data analysis happens not in the data center or cloud, but close to the IoT devices and sensors.

- | | |
|---------------------|----------------------|
| A. Edge Computing | B. Quantum Computing |
| C. Mobile Computing | D. Data Computing |

Q9. All of the following accurately describe Hadoop, EXCEPT

- | | |
|-----------------------------------|----------------|
| A. Real-time | B. Java-based |
| C. Distributed computing approach | D. Open source |

Q10. Which of the following platforms does Hadoop run on?

- | | |
|---------------|-------------------|
| A. Unix-like | B. Cross-platform |
| C. Bare metal | D. Debian |

Q11. Which is the architectural center of Hadoop that allows multiple data processing engines.

- | | |
|--------------|------------|
| A. Incubator | B. YARN |
| C. Hive | D. Chuckwa |

- Q12. The output of the _____ is not sorted in the MapReduce framework for Hadoop.
- A. Mapper
B. Scalding
C. Cascader
D. None of the mentioned
- Q13. In which of the following Data life cycle is manage in MongoDB Schema Design?
- A. Insert. Update and Delete
B. Grant, permission and View
C. Read and write ratio, Query Type
D. Application Side
- Q14. which is used for storing documents and making remote procedure calls in MongoDB.
- A. BSON
B. Divers
C. JSON
D. Native
- Q15. Which document is stores in Object data type?
- A. Database documents
B. Embedded documents
C. Web documents
D. Query documents
- Q16. Which is not session commands in MongoDB?
- A. commitTransaction
B. killCursors
C. endSession
D. killSessions
- Q17. Which is not part of Pig Latin Data Model?
- A. Relationship
B. Bag
C. Tuple
D. Field
- Q18. Which is data size of Hive ?
- A. Gigabytes
B. Megabytes
C. kilobytes
D. Petabytes
- Q19. Which of the following is relational operators in Pig .
- A. DESCRIBE
B. DISTINCT
C. DUMP
D. All of the above
- Q20. Which is not Apache Pig Execution Mechanism?
- A Interactive Mode
B. Embedded Mode
C. Collection Mode
D. Batch Mode

SECTION – B

(Very Short Answer Type Questions)

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

each 2 marks

- Q.1 What is attribute types?
- Q.2 What is OLAP Operations?
- Q.3 What is Quantum Computing?
- Q.4 What is Big Data?
- Q.5 What is Hadoop Distributed File System?
- Q.6 What is Hadoop ecosystem ?
- Q.7 What is NoSQL?
- Q.8 How many types of data types in MangoDB?
- Q.9 What is Local Mode in Apache Pig?
- Q.10 What is HiveQL?

SECTION – C

(Short Answer Type Questions)

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

each 3 marks.

- Q.1 What do you mean by classification of data mining function?
- Q.2. What do you mean by Data Mining Primitives?
- Q.3. What is importance of big data?
- Q.4. What is difference between big data and small data?
- Q.5. What is the difference between JobTracker and TaskTracker?
- Q.6. What is MapReduce framework?
- Q.7. What do you mean by Key-Value Store Database?
- Q.8 What is difference between SQL and NoSQL?
- Q. 9 Write the application of Apache Pig?
- Q.10. Write the Hive Datatype.

SECTION – D

(Long Answer Type Questions)

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

each 6 marks.

Q.1. What do you understand by Data Warehouse? Explain the Architecture of Data Warehouse.

OR

What do you mean by KDD process? Explain the which type of data mined.

Q.2. Why data analytics is important in Big data? Explain the different types of Analytics.

OR

What do you understand data? Write the classification of Data.

Q.3. What do you understand by Open Source Big Data Tools? Write name any five Big Data tools and Explain the one tool Hadoop 2.0.

OR

Explain the following: -

- (i) HDFS data Storage
- (ii) MapReduce Execution

Q.4. What is the structure of MongoDB? Explain the feature of MongoDB.

OR

Explain the following: -

- (ii) NoSQL Datastore
- (ii) NoSQL Data Architecture pattern

Q.5. What is big data programming language? Explain the Pig data Model with each component.

OR

Explain the following: -

- (i) Hive Architecture
- (ii) Hive Data model
