Avinashilingam Institute for Home Science and Higher Education for Women

(Deemed to be University) Coimbatore-641043

Master's Degree Examination –November 2018

III – semester

Class : II PG **Time: 3 Hours**

Major: Information Technology **Max. marks: 60**

17MITC15 Big Data Analytics

Part A **10 x 1/2=5**

Choose the correct answer

1. The primary benefit behind a big data platform is to reduce_____.
   a) Storage b) Complexity c) Services d) All

2. The _____ different features of big data analytics.
   a) Open source b) Data recovery c) Scalability d)All

3. The _____ site is creates and have real big data in the world.
   a) Facebook b)Twitter c) Google d) YouTube

4. _____ is a framework for performing remote procedure calls and data serialization.
   a) Drill b) BigTop c)Avro d) Chukwa

5. _____ is the most popular high-level Java API in Hadoop Ecosystem
   a) Scalding b) HCatalog c) Cascalog d) Cascading

6. What license is Hadoop distributed under?
   a) ApacheLicense2.0 b) Mozilla Public License
   c) Shareware d) Commercial

7. _____job are optimized for scalability but not latency.
   a) Map reduce b) Drill c) Oozie d) Hive

8. _____ is a unit of data that flows through a flume agent.
   a) Record b) Row c) Log d) Event

9. The Hadoop Distributed File System is specially designed to be highly_____
   a) Fault-tolerant b) Server rack c) Master/slave d) Hadoop cluster

10. In _____mode HiveServer2 only accepts valid Thrift calls.
    a)Remote b)HTTP c)Embedded d) Interactive

Part B 5 x 4= 20

Answer ALL questions

Each answer should not exceed 200 words or one page

11.a) Define Big data? Write about Big Data Platform (Or)

11 b) Discriminate on Analysis VS Reporting.

12.a) Short note on Estimating Moments in Stream computing (Or)

12.b) Explain the concept of Real time Sentiment Analysis.

13.a) Describe the History of Hadoop. (Or)

13.b) Explain the concept of Task Execution in Hive.

14.a) How to do cluster Setup and Installation ? (Or)

14.b) Illustrate the Hadoop benchmarks.

15.a) Explain the concept of Hive services, Hive QL, Querying data. (Or)

15.b) Discuss about ZooKeeper.

Part C 5 x 7 = 35

Answer ALL questions

Each answer should not exceed 600 words or three pages

16.a) Describe the concept Intelligent data analysis and Nature of data. (Or)

16.b) Explain about Prediction Error with example.

17.a) State the Stream data model and Architecture with diagram. (Or)

17.b) Describe about the Real Time Analytics platform and recitate.

18.a) Summarized Analyzing the data with Hadoop, Scaling Out and Streaming. (Or)

18.b) Discuss about Map Reduce Types and Formats.

19.a) Explain Hadoop Configuration and security in Hadoop. (Or)

19.b) Illustrate HDFS and Monitoring, Maintenance.

20.a) List out Application on Big data using Pig and Hive (Or)

20.b) Explain about Visualization techniques in Big Data