Hall Ticket No | | | | | | | | | | | Question Paper Code: AITB14

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal-500043, Hyderabad

B.Tech VII SEMESTER END EXAMINATIONS (REGULAR/SUPPLEMENTARY) - DECEMBER 2022

Regulation: R18

BIG DATA ANALYTICS

Time: 3 Hours **(INFORMATION TECHNOLOGY)** **Max Marks: 70**

Answer FIVE Questions choosing ONE question from each module
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

MODULE – I

1. (a) What is big data? List the characteristics of big data. Describe the terminologies used in big data. [BL: Understand| CO: 1|Marks: 7]

   (b) List the various steps involved in big data analytics. Identify the difference between traditional data and big data. [BL: Understand| CO: 1|Marks: 7]

2. (a) Identify the different applications of big data. Explain how big data is useful in health care. [BL: Understand| CO: 1|Marks: 7]

   (b) If an organization retains large data which is not evaluated or analyzed, it will not benefit the organization. Identify how big data analytics would benefit the organization in this scenario. [BL: Understand| CO: 1|Marks: 7]

MODULE – II

3. (a) Illustrate Hadoop core components with a neat diagram. Discuss in brief about the basic building blocks in Hadoop. [BL: Understand| CO: 2|Marks: 7]

   (b) Explain the relationship between the Hadoop and big data. How big data processing differs from distributed processing. [BL: Apply| CO: 2|Marks: 7]

4. (a) Classify the various kinds of NoSQL data stores. When should we use a NoSQL database instead of a SQL database (relational database)? [BL: Understand| CO: 2|Marks: 7]

   (b) Describe the four major categories of NoSQL databases with relevant products. Differentiate RDMS and NOSQL. [BL: Apply| CO: 2|Marks: 7]

MODULE – III

5. (a) Demonstrate command line interface using HDFS files and give a brief note on Hadoop-specific file system types and HDFS commands. [BL: Understand| CO: 4|Marks: 7]

   (b) Mention the different HDFS file operations. Explain in detail about anatomy of File Read and write in HDFS [BL: Apply| CO: 4|Marks: 7]

6. (a) Enlist the limitations of Hadoop 1.0. Explain Hadoop 2: HDFS and Hadoop 2: YARN? [BL: Understand| CO: 4|Marks: 7]

(b) Write about data node. How many instances of data node run on a Hadoop cluster?
[BL: Apply| CO: 4|Marks: 7]

MODULE – IV

7. (a) Interpret the various methods in MapReduce job class. Discuss in brief the role of mapper and reducer in Hadoop Map reduce. [BL: Understand| CO: 5|Marks: 7]

   (b) In Map-Reduce implementation of word count, write types of keys and values mapper, reducer, combiner(if any) and which of the types need to writable. [BL: Apply| CO: 5|Marks: 7]

8. (a) Demonstrate map reduce workflow with the help of an example. Write the differences between Mapreduce combiner and reducer. [BL: Understand| CO: 5|Marks: 7]

   (b) Develop the map reduce application to count the frequency of each word in an input text.
   [BL: Apply| CO: 5|Marks: 7]

MODULE – V

9. (a) Demonstrate Hive meta store. Which classes are used by the Hive to read and write HDFS files? Explain in detail. [BL: Understand| CO: 6|Marks: 7]

   (b) Consider the student data file (st.txt), data in the following format Name, District, age, gender
   i) Write a PIG script to display names of all female students
   ii) Write a PIG script to find the number of students from Prakasham district
   iii) Write a PIG script to display district wise count of all male students.
   [BL: Apply| CO: 6|Marks: 7]

10. (a) How the PIG programs can be packaged and explain the modes of running a PIG script with a neat sketch. [BL: Understand| CO: 6|Marks: 7]

    (b) Write the various Hive built-in functions with syntax. Illustrate the architecture of HIVE with the help of a neat diagram. [BL: Understand| CO: 6|Marks: 7]