



# INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal-500043, Hyderabad

B.Tech VII SEMESTER END EXAMINATIONS (REGULAR) - DECEMBER 2023

Regulation: UG-20

**BIG DATA AND ANALYTICS**

**Time: 3 Hours**

(COMMON TO CSE | CSE(DS) | CSIT))

**Max Marks: 70**

**Answer ALL questions in Module I and II**

**Answer ONE out of two questions in Modules III, IV and V**

**All Questions Carry Equal Marks**

**All parts of the question must be answered in one place only**

## MODULE – I

1. (a) Explain in detail about nature of data and its applications. Point out the analysis tools and reporting tools used in bigdata. [BL: Understand| CO: 1|Marks: 7]
- (b) Elucidate the roles of analytical team and IT team with a detailed note on text analysis? List the out the best practices and main characteristics of bigdata. [BL: Understand| CO: 1|Marks: 7]

## MODULE – II

2. (a) Outline about Hadoop. Discuss the features of Hadoop and explain the functionalities of Hadoop cluster. [BL: Understand| CO: 2|Marks: 7]
- (b) Discuss the history of Hadoop. Explain the Hadoop ecosystem roles in big data processing. [BL: Understand| CO: 2|Marks: 7]

## MODULE – III

3. (a) Discuss about the Hadoop input and output and write a note on data integrity. [BL: Understand| CO: 3|Marks: 7]
- (b) List and explain the limitations and solutions of Hadoop for big data analytics. [BL: Understand| CO: 3|Marks: 7]
4. (a) Mention the benefits of big data processing. Discuss the features and benefits of Apache ambari. [BL: Understand| CO: 4|Marks: 7]
- (b) The default block size is 64MB and the replication factor is 3. Calculate the number of blocks allocated for a file having the size of 300MB. [BL: Apply| CO: 4|Marks: 7]

## MODULE – IV

5. (a) What is Map reduce programming model? Describe the working of Map reduce with a relevant example. [BL: Understand| CO: 5|Marks: 7]
- (b) Discuss some techniques to optimize MapReduce jobs and the points you need to consider while designing a file system in MapReduce. [BL: Understand| CO: 5|Marks: 7]

6. (a) Illustrate with diagram about how Hadoop uses HDFS staging directory as well as local directory during a job run [BL: Understand| CO: 5|Marks: 7]
- (b) Examine MapReduce job you consistently see that map tasks on your cluster are running slowly because of excessive garbage collection of JVM. How do you increase JVM heap size property to 3GB to optimize performance? [BL: Apply| CO: 5|Marks: 7]

### MODULE – V

7. (a) Differentiate Pig and Hive with various parameters. Discuss the Pig Latin relational operators. [BL: Understand| CO: 6|Marks: 7]
- (b) Illustrate different types of joins in Pig Latin with examples on different data types. [BL: Understand| CO: 6|Marks: 7]
8. (a) How will you query the data in HIVE? Explain the process of installing HIVE & features of HIVE. [BL: Understand| CO: 6|Marks: 7]
- (b) A start-up company wants to use Hive for storing its data. Discuss a shell command in Hive to list all the files in the current directory. [BL: Apply| CO: 6|Marks: 7]

– ○ ○ ○ ○ –