

[Nov-22]

**GITAM (Deemed to be University)**  
**[20SST-843]**  
**M.Sc. Degree Examination**

**Statistics**  
**III Semester**

**BIG DATA ANALYTICS**

(Effective from the admitted batch 2021–22)

**Time: 3 Hours**

**Max.Marks: 60**

---

**Instructions:** All parts of the module must be answered in one place only.  
Figures in the right hand margin indicate marks allotted.

---

**Section-A**

1. Answer all the questions: (10×2=20)
- a) How Big Data is used in retail industry?
  - b) List five vs of Big data.
  - c) Distinguish between Master and Slave nodes.
  - d) Signify the importance of resource manager.
  - e) Distinguish between map() and reduce().
  - f) What is Ingestion in Big Data processing?
  - g) Interpret Application Manager.
  - h) What is container?
  - i) What are the Hive Datatypes?
  - j) What is Join?

**Section-B**

Answer the following: (5×8=40)

**UNIT-I**

2. Discuss various types of data supported by Big Data with suitable example for each type.

**OR**

3. Interpret the use of Big Data in preventing fraudulent activities.

## **UNIT-II**

4. Signify the role of Cloud Services play in the Big Data.

**OR**

5. List the important features of Hadoop.

## **UNIT-III**

6. Explain the role of Partitioner and Combiner in MapReduce Application.

**OR**

7. Explain Security Layer and Physical infrastructure layer of Big Data Architecture.

## **UNIT-IV**

8. Explain Log Management in YARN.

**OR**

9. Illustrate working of YARN with neat illustration.

## **UNIT-V**

10. Elaborate Evaluation, Math and String functions in Hive.

**OR**

11. What do you mean by statistical significance.

[IIS/222]