GITAM (Deemed to be University)
[20SDS803]
M.Sc. Degree Examination

Data Science
III SEMESTER

BIG DATA ANALYTICS
(for the admitted batch 2021–22)

Time: 3 Hours   **Max.Marks: 60**

Instructions: Answer all units choosing one question from each unit.
All parts of the unit 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)   List out important features of Hadoop.
   b)   List some methods for the verification of credit cards.
   c)   What is metadata? What information does it provide?
   d)   Discuss the term streaming information access.
   e)   Write short notes on visualization layer.
   f)   Explain the use of Job History Service.
   g)   Define Bucketing in hive.
   h)   What is the purpose of bag and tuple functions supported by pig?
   i)   Define resilient distributed datasets.
   j)   Explain the use of .leftOUterJoin(…) transformation method.

Section-B

Answer the following: 5×8=40)

UNIT-I

2.   Briefly discuss the evolution of Big Data.

OR

3.   List some common types of financial frauds prevalent in the current business scenario.

UNIT-II

4. Explain HBase Architecture with a neat diagram also explain all the component of HBase.

OR

5. Explain the phases of MapReduce Model with a neat diagram.

UNIT-III

6. Explain the working of resource manager and applications manager.

OR

7. Discuss the key building blocks of the Hadoop platform management layer.

UNIT-IV

8. Demonstrate the following Data Retrieval Queries in HIVE. Select, Where, Group by, Having, Limit.

OR

9. Explain the operations supported by Pig along with syntax.

UNIT-V

10. Explain all the actions available on RDDs.

OR

11. Discuss about creation of data frames.

[32/III S/222]