PART-B

4 x 15 = 60M

UNIT-I

2. a. Briefly discuss about characteristics of Big Data. **7M**
   b. Differentiate between Data vs. Information vs. Big Data. **8M**

(or)

3. a. What is Hadoop? Categorize various tools of Hadoop framework. **8M**
   b. Explain Hadoop ecosystem with examples. **7M**

UNIT-II

4. a. Write short notes on name node and data node. **7M**
   b. Discuss how to read data from a Hadoop URL. **8M**

(or)

5. a. Discuss in detail about basic file system operations in HDFS. **7M**
   b. Define HDFS? Explain in brief about the basic building blocks of Hadoop? **8M**

UNIT-III

6. a. Discuss in brief about Mapper and Reducer in MapReduce. **8M**
   b. Explain Pig's built-in types in detail. **7M**

(or)

7. a. List and explain MapReduce input and output formats. **7M**
   b. Differentiate between local and distributed modes in Pig scripts. **8M**

UNIT-IV

8. a. Write Hive commands to create a Student table with fields: roll number, name and address. Also, insert two rows into that table. **8M**
   b. Differentiate HiveQL with traditional SQL. **7M**

(or)

9. a. Elaborate the procedure to create and manage tables in Hive. **7M**
   b. Discuss various data types supported by HiveQL with an example. **8M**

* * *

VR17 Reg. No: | | | | | | | | | | |

VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

III/IV B.Tech. DEGREE EXAMINATION, JUNE, 2022

Sixth Semester

INFORMATION TECHNOLOGY

17IT4604A BIG DATA

Time: 3 hours ***Max. Marks: 70***

Part-A is compulsory

Answer One Question from each Unit of Part - B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. List applications of Big Data.
   b. Name the latest versions of Hadoop releases.
   c. Define Big Data.
   d. What is the role of job tracker in HDFS?
   e. How to copy a file from the HDFS to local file system?
   f. Which is the default input formats defined in MapReduce?
   g. What is the key-value pair in Hadoop MapReduce?
   h. List different complex data types in Pig.
   i. What is a metastore in Hive?
   j. List any two Pig commands.

VR17 Reg. No:

VELAGAPUDI RAMAKRISHNA
SIDDHARTHA ENGINEERING COLLEGE
(AUTONOMOUS)
III/IV B.Tech. DEGREE EXAMINATION, AUGUST, 2021
Sixth Semester
INFORMATION TECHNOLOGY
17IT4604A BIG DATA

Time: 3 hours ***Max. Marks: 70***
Part-A is compulsory
Answer One Question from each Unit of Part - B
Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. Define big data.
   b. List out different types of data in big data.
   c. What is FUSE?
   d. HDFS has the concept of a block, but it is a much larger unit 64MB by default? Justify your answer.
   e. List any two major concepts of Hadoop.
   f. Identify two operations of MapReduce.
   g. What are the two properties that can be set in HDFS configuration through Command-Line interface?
   h. What are big data sources?
   i. What is HIVE?
   j. Explain execution modes in Pig.

PART-B

4 x 15 = 60M

UNIT-I

2. a. Discuss about the three characteristics defined in big data in detail. **8M**
   b. Contrast between traditional fraud detection patterns and modern-day fraud detection with a neat sketch in fraud detection patterns for big data deployment. **7M**

(or)

3. a. Discuss the history of Hadoop and its ecosystem. List out Hadoop latest release. **8M**
   b. Explain the solutions taken to overcome the problem for data storage and analysis in Hadoop. **7M**

UNIT-II

4. a. What is HDFS? List the two types of nodes operating in HDFS cluster. **8M**
   b. Discuss the term commodity hardware in design of HDFS. **7M**

(or)

5. a. Discuss different interfaces in HDFS in detail. **7M**
   b. Explain reading data from a Hadoop URL and reading data using the file system API in Java interface. **8M**

UNIT-III

6. a. Explain about Hadoop streaming and pipes. **7M**
   b. Explain MapReduce dataflow with single, multiple and no reduce tasks with a neat sketch. **8M**

(or)

7. a. Demonstrate Pig Latin structure and schema with an example. **7M**
   b. List out the functions of Pig and give description of each function in a neat table. **8M**

UNIT-IV

8. a. List out different types of Joins in Hive and explain with an example. **7M**
   b. What is Hive shell? Explain the steps to configure Hive and setting its engine and logging into Hive. **8M**

(or)

9. a. Compare Hive with traditional databases. **8M**
   b. List out different types of Hive services. **7M**

* * *

VR17 Reg. No: | | | | | | | | | | |

VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

III/IV B.Tech. DEGREE EXAMINATION, OCTOBER, 2020

Sixth Semester

INFORMATION TECHNOLOGY

17IT4604A BIG DATA

Time: 3hours ***Max. Marks: 70***

Part-A is compulsory

Answer One Question from each Unit of Part-B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. What are the characteristics of big data?
   b. Name any 2 Hadoop projects.
   c. Which java abstract class represents a file system in Hadoop?
   d. Define FUSE.
   e. Write the function of Map phase and Reduce phase.
   f. What are the two execution modes of Pig?
   g. Mention the different services of Hive.
   h. What are the two formats that govern table storage in Hive?
   i. Define Hadoop.
   j. List the two types of nodes that control the job execution process.

PART-B

4 x 15 = 60M

UNIT-I

2. a. Explain about social media pattern of big data. **7M**
   b. How can you store and analyze big data? **8M**

(or)

3. a. Write about brief history of Hadoop. **7M**
   b. Briefly explain about Hadoop ecosystem. **8M**

UNIT-II

4. a. Discuss about different Hadoop file systems. **7M**
   b. Explain about the following HDFS concepts **8M**
      i) Blocks
      ii) Name node and data node

(or)

5. a. Explain about the coherency model. **7M**
   b. Discuss about the sequence of events when a client writing data to HDFS. **8M**

UNIT–III

6. a. Write a unix shell program for finding the maximum recorded temperature. **7M**

   b. Discuss about hadoop pipes. **8M**

(or)

7. a. Explain about different types of functions used in Pig. **7M**
   b. Compare the following **8M**
      i) PigLatin and SQL
      ii) Pig and RDBMS

UNIT–IV

8. a. Write the procedure for installing HIVE. **7M**
   b. Explain how HIVE is able to create and store the tables? **8M**

(or)

9. a. Explain about HIVE services. **7M**
   b. Discuss about HIVE architecture. **8M**

* * *

VR17 Reg. No: | | | | | | | | | | |

VELAGAPUDI RAMAKRISHNA
SIDDHARTHA ENGINEERING COLLEGE
(AUTONOMOUS)
III/IV B.Tech. DEGREE EXAMINATION, MARCH, 2021
Sixth Semester
INFORMATION TECHNOLOGY
17IT4604A BIG DATA
(CBCS)

Time: 3 hours ***Max. Marks: 70***

Part-A is compulsory

Answer One Question from each Unit of Part - B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. Define HDFS.
   b. List out any two advantages of Hadoop.
   c. What is NFS?
   d. What is the default block size in HDFS?
   e. Explain Megastore in Hive.
   f. Define MapReduce. Why MapReduce matters?
   g. What is the main work of shuffle and sort phase in MapReduce?
   h. What is Pig Latin?
   i. What is Hive Shell?
   j. Explain Coherency Model.

PART-B

4 x 15 = 60M

UNIT-I

2. a. Explain about Risk Modeling and Management in Big Data usage pattern. **8M**
   b. What is the importance of Big Data solutions and when to consider it? **7M**

(or)

3. a. Explain different processing patterns that work with Hadoop. **7M**
   b. Compare RDBMS, Grid Computing and Volunteer Computing with Hadoop. **8M**

UNIT-II

4. a. 'Design of HDFS runs on clusters of commodity hardware', Justify this statement in detail. **8M**
   b. What is a block in HDFS and why a block in HDFS is so large? **7M**

(or)

5. a. Demonstrate with a neat diagram the Anatomy of a file write. **8M**
   b. List out different File systems in hadoop and give its description in a neat table. **7M**

UNIT-III

6. a. Explain analyzing of data with Unix Tools and analyzing of data with Hadoop with a neat example. **8M**
   b. Explain the working of MapReduce with a neat Logical data flow example. **7M**

(or)

7. a. Discuss the execution modes of Pig and explain the three ways of executing Pig programs. **7M**
   b. Explain the two Data Processing operators in PIG. **8M**

UNIT-IV

8. a. Demonstrate Hive architecture, with a neat sketch. **8M**
   b. Compare and contrast between SQL and HiveQL in a neat table. **7M**

(or)

9. a. List out the different tables in Hive and explain. **7M**
   b. Explain sorting and aggregating and MapReduce Scripts in querying data in Hive to retrieve data with an example. **8M**

* * *

VR17 Reg. No:

V.H. Siddhartha Engineering College 520 007

VEL AGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

III/IV B.Tech. DEGREE EXAMINATION, FEBRUARY, 2022

Sixth Semester

INFORMATION TECHNOLOGY

17IT4604A BIG DATA

(CBCS)

Time: 3 hours *Max. Marks: 70*

Part-A is compulsory

Answer One Question from each Unit of Part - B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. List applications of Big Data analytics.
   b. Name the latest versions of Hadoop Releases.
   c. Define Big Data.
   d. What is the role of Job tracker in HDFS?
   e. How to copy a file from the HDFS to local file system?
   f. Which is the default Input Formats defined in MapReduce?
   g. What is the key-value pair in Hadoop MapReduce?
   h. List different complex data types in pig.
   i. What is a metastore in HIVE?
   j. List any two PIG commands.

PART-B

4 x 15 = 60M

UNIT-I

2. a. Define Big Data and explain its characteristics briefly. **7M**
   b. Differentiate between Data vs. Information vs. Big Data. **8M**

(or)

3. a. What is Hadoop? Categorize various tools of Hadoop framework. **8M**
   b. Explain Hadoop Ecosystem with examples. **7M**

UNIT-II

4. a. Write short notes on Name Node and Data Node. **7M**
   b. Explain the term commodity hardware in design of HDFS. **8M**

(or)

5. a. Discuss in detail about basic file system operations in HDFS. **7M**
   b. Define HDFS? Explain in brief about the basic building blocks of Hadoop? **8M**

UNIT-III

6. a. Discuss in brief about Mapper code and Reducer code. **8M**
   b. Explain Pig's built-in types in detail. **7M**

(or)

7. a. List and Explain MapReduce input and output formats. **7M**
   b. Differentiate Local and Distributed mode in pig scripts. **8M**

UNIT-IV

8. a. Write the Hive commands to create a sample table of a student with roll number, name and address and insert two rows into that table. **8M**
   b. Differentiate HiveQL with traditional SQL. **7M**

(or)

9. a. Model the procedure for create and manage the data bases in Hive. **7M**
   b. Discuss various data types supported by HiveSQL with an example. **8M**

* * *

VR14 Reg. No:

VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

III/IV B.Tech. DEGREE EXAMINATION, MARCH/APRIL, 2019

Sixth Semester

INFORMATION TECHNOLOGY

14IT3602 BIG DATA

Time: 3hours *Max. Marks: 70*

Part-A is compulsory

Answer One Question from each Unit of Part-B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. List the examples of big data.
   b. Outline the considerations where big data technologies used.
   c. What is HDFS?
   d. What are namenodes and datanodes in HDFS?
   e. What is a coherency model?
   f. What is the command for determining the user identity that hadoop uses for permissions in HDFS?
   g. What is a mapper?
   h. Give the use of shuffle and sort.
   i. What is hive shell?
   j. What is a metastore?

PART-B

4 x 15 = 60M

UNIT-I

2. a. Discuss why is big data important? **8M**
   b. Explain the characteristics of big data in detail. **7M**

(or)

3. a. Discuss hadoop ecosystem. **8M**
   b. Explain the context of big data in risk modelling and management. **7M**

UNIT-II

4. a. Discuss querying the filesystem. **7M**
   b. How do we read data using the filesystem API? **8M**

(or)

5. Demonstrate with a neat diagram the anatomy of a file write. **15M**

UNIT-III

6. a. Explain hadoop's configuration API. **8M**
   b. Write short notes on GenericOptionsParser and ToolRunner. **7M**

(or)

7. Discuss with an application running a job in a local job runner. **15M**

UNIT-IV

8. a. Compare hive with traditional databases and comment. **8M**
   b. Discuss the data types supported by hive. **7M**

(or)

9. a. Discuss about configuring hive. **8M**
   b. Explain about hiveQL. **7M**

* * *

PART-A

10 x 1 = 10M

1. a. What is Big Data?
   b. Discuss about data in warehouse and data in Hadoop?
   c. What are the advantages of Hadoop?
   d. List the components of a map reduce application that we can develop.
   e. List the writable wrapper classes for java primitive.
   f. List the main features of MapReduce.
   g. What do you mean by a block in file system and specify its size?
   h. Explain Metastore in Hive.
   i. What is Spark?
   j. List Hive services.

PART-B

4 x 15 = 60M

UNIT-I

2. a. List and discuss the four elements of Big Data. 8M
   b. Discuss brief history of Hadoop. 7M

(or)

3. a. List the important features of Hadoop. 8M
   b. Discuss how MapReduce is different from RDBMS? 7M

UNIT-II

4. a. Explain the Hadoop distributed file system architecture with a neat sketch. 8M
   b. Describe the areas where HDFS is not suitable for data handling? 7M

(or)

5. a. Describe any two file system interfaces that are used with HDFS. 8M
   b. Explain about coherency model. 7M

UNIT-III

6. a. Discuss about running a job in a Local Job Runner. 8M

   b. Distinguish between the old and new versions of Hadoop API for MapReduce frame work. 7M

(or)

7. a. Explain about configuring the development environment. 7M
   b. Explain working of following phases of MapReduce with one common example 8M
      i) Map Phase
      ii) Combiner Phase
      iii) Shuffle and Sort Phase
      iv) Reducer Phase

UNIT-IV

8. a. Explain about the various data types supported by HIVEQL with an example. 8M
   b. How to create a table by using HIVEQL? 7M

(or)

9. a. Explain about running and configuring Hive. 8M
   b. What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example. 7M

* * *

10 x 1 = 10M

1. a. List the characteristics of big data.
   b. What is social media analytics?
   c. What is Hadoop?
   d. What is FUSE?
   e. What is a block in HDFS?
   f. What is the classical tool for processing line oriented data?
   g. What does Hadoop streaming use?
   h. Give the use of reduce in mapreduce.
   i. What is Hive?
   j. Why do we need HiveQL?

PART-B

4 x 15 = 60M

UNIT-I

2. a. Discuss on data in the warehouse and data in Hadoop. **8M**
   b. Explain the different processing patterns that work with Hadoop. **7M**

(or)

3. a. Discuss how Hadoop is different from other systems? **8M**
   b. Discuss about characteristics of Big Data. **7M**

UNIT-II

4. a. Illustrate in detail Hadoop filesystem. **7M**
   b. With a neat diagram demonstrate accessing HDFS over HTTP directly and via a bank of HDFS proxies. **8M**

(or)

5. Demonstrate with a neat diagram the anatomy of a file read. **15M**

UNIT-III

6. a. Give an example of determining maximum temperature and explain how mapreduce works? **8M**
   b. Discuss about writing a Unit Test. **7M**

(or)

7. a. Explain about running a Job in a Local Job Runner. **8M**
   b. Explain about Hadoop Streaming and Pipes. **7M**

UNIT-IV

8. a. Explain about HiveQL. **8M**
   b. Discuss about Hive Services. **7M**

(or)

9. a. Explain Tables and Quering Data in Hive. **9M**
   b. Discuss about Configuring Hive. **6M**

* * *