

Hall Ticket Number:

--	--	--	--	--	--	--	--	--

IV/IV B.Tech (Supplementary) DEGREE EXAMINATION

November, 2019

Eight Semester

Time: Three Hours

CSE&IT

Advanced Data Analytics

Maximum : 60 Marks

Answer Question No.1 compulsorily.

(1X12 = 12 Marks)

Answer ONE question from each unit.

(4X12=48 Marks)

1 Answer all questions

(1X12=12 Marks)

- What is Big Data?
- What are the applications of big data?
- Define namenodes and datanodes in HDFS.
- Why blocks are large in HDFS?
- What is the default scheduler in Hadoop?
- What is the role of Job tracker and Task tracker in Map Reduce?
- List the failures in classical Map Reduce.
- List the built-in counters in Map Reduce.
- Purpose of Filter function in Pig.
- List the complex data types in Hive.
- Write the scala program to find maximum temperature.
- Write any one application of Apache Swoop.

UNIT I

- Show how a client read and write data in HDFS with suitable code.
- Discuss in detail about parallel copying with distcp in HDFS.

6M

6M

(OR)

- Discuss the steps involved in designing Hadoop Distributed File System and Give the design of HDFS.

12M

UNIT II

- Analyze the data with Hadoop using Map and Reduce with an example.
- Explain in detail about MapReduce data flow with single and multiple reduce tasks.

6M

6M

(OR)

- Explain briefly about shuffle and sort in Map Reduce.

12M

UNIT III

- Explain about Pig Latin expressions and Pig Latin types.
- Explain about data processing operators in Pig.

6M

6M

(OR)

- Demonstrate briefly about Hive services.
- Write the difference between the Traditional Databases and Hive.

6M

6M

UNIT IV

- Explain the Anatomy of spark job run.
- Explain briefly about task scheduling and task execution in spark.

6M

6M

(OR)

- Describe Sqoop import command.
- How Sqoop export command works?

6M

6M

