Hall Ticket Number:

IV/IV B.Tech (Supplementary) DEGREE EXAMINATION

November, 2019	**CSE&IT**
Eight Semester	**Advanced Data Analytics**
Time: Three Hours	**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)
- a) What is Big Data?
- b) What are the applications of big data?
- c) Define namenodes and datanodes in HDFS.
- d) Why blocks are large in HDFS?
- e) What is the default scheduler in Hadoop?
- f) What is the role of Job tracker and Task tracker in Map Reduce?
- g) List the failures in classical Map Reduce.
- h) List the built-in counters in Map Reduce.
- i) Purpose of Filter function in Pig.
- j) List the complex data types in Hive.
- k) Write the scala program to find maximum temperature.
- l) Write any one application of Apache Swoop.

UNIT I

2   a) Show how a client read and write data in HDFS with suitable code. 6M
    b) Discuss in detail about parallel copying with distcp in HDFS. 6M

(OR)

3   Discuss the steps involved in designing Hadoop Distributed File System and Give the design of HDFS. 12M

UNIT II

4   a) Analyze the data with Hadoop using Map and Reduce with an example. 6M
    b) Explain in detail about MapReduce data flow with single and multiple reduce tasks. 6M

(OR)

5   a) Explain briefly about shuffle and sort in Map Reduce. 12M

UNIT III

6   a) Explain about Pig Latin expressions and Pig Latin types. 6M
    b) Explain about data processing operators in Pig. 6M

(OR)

7   a) Demonstrate briefly about Hive services. 6M
    b) Write the difference between the Traditional Databases and Hive. 6M

UNIT IV

8   a) Explain the Anatomy of spark job run. 6M
    b) Explain briefly about task scheduling and task execution in spark. 6M

(OR)

9   a) Describe Sqoop import command. 6M
    b) How Sqoop export command works? 6M