Big Data Analytics Lab

PC 752 CSM

Instruction :	**2 per week**
Duration of SEE :	**2 hours**
CIE :	**25 marks**
SEE :	**50 marks**

Credits :1

Course Objectives:

Students will try:

1. To provide the knowledge to setup a Hadoop Cluster
2. To impart knowledge to develop programs using MapReduce Technique
3. To learn file handling in HDFS
4. To introduce Pig, PigLatin and HiveQL to process big data
5. To learn machine learning operations using Mahout Hadoop
6. To introduce NoSQL databases

Course Outcomes:

Student will able to:

1. Understand Hadoop working environment
2. Work with big data applications in multi node clusters
3. Write scripts using Pig to solve real world problems
4. Write queries using Hive to analyse the datasets
5. Apply big data and echo system techniques for real world

List of Experiments to be performed

1. Understanding and using basic HDFS commands
2. Word count application using Mapper Reducer on single node cluster
3. Working with files in Hadoop file system: Reading, Writing and Copying
4. Writing User Defined Functions/Eval functions for filtering unwanted data in Pig
5. Retrieving user login credentials from /etc/passwd using Pig Latin
6. Working with HiveQL.
7. Writing User Defined Functions in Hive

Suggested reading:

1. Tom White, "Hadoop: The Definitive Guide", 4th Edition, O'Reilly Media Inc, April 2015.
2. Alan Gates, "Programming Pig", O'Reilly Media Inc, 2011.