Course Code	CSE510A
Course Name	Big Data Analytics
Credits	2
Course Offered to	UG/PG
Course Description	Distributed processing frameworks have emerged as a feasible and cost effective way of analyzing the increasing volume of data. This course provides a solid understanding of two of the most popular of distributed processing frameworks - Hadoop and a more recent incarnation called Spark. Examples and hands-on exercises will prepare those taking this course to be able to apply these frameworks in practice.

Pre-requisites

Pre-requisite (Mandatory)	Pre-requisite (Desirable)	Pre-requisite(other)
DBMS	Datamining, Machine Learning	Working knowledge of Programming in C,C++ or Java and experience with Unix

*Please insert more rows if required

Post Conditions*(For suggestions on verbs please refer the second sheet)

CO1	CO2	CO3	CO4
Ability to solve big data problems using MAP-reduce Framework.	Ability to solve big data analytics problems using Spark Framework.	Ability to analyze performance of a problem implemented in map-reduce framework.	

Weekly Lecture Plan

Week Number	Lecture Topic	COs Met	Assignment/Labs/Tutorial
1	Introduction to Big Data Problems	CO1, CO2	Installation of Hadoop
2,3	Introduction to Hadoop, HDFS	CO2	
4,5,6,7	Map Reduce Design Patterns	CO1	A lab assignment on Map-Reduce
8,9,10	Introduction to Spark and its APIs	CO3	A lab assignment on Spark
11,12,13	Discussion of Research Papers published in KDD, VLDB that use Map-Reduce and Spark	CO1, CO2, CO3	A Project on Spark/MapReduce

*Please insert more rows if required

Assessment Plan

Type of Evaluation	% Contribution in Grade
End-sem	30%
Laboratory and Project	70%

*Please insert more row for other type of Evaluation

Resource Material

Type	Title
Textbook	Massive Data Analytics by Rajaraman and Ullman
Textbook	Big Data: Principles and best practices of scalable real time data systems by Nathan Marz