Big Data Analytics

Code	Credit Hours
CS-404	2-1

Course Description

Big data analytics is the process of examining data to uncover hidden patterns, unknown correlations, and other useful information that can be used to make better decisions. Big Data requires the storage, organization, and processing of data at a scale and efficiency that goes well beyond the capabilities of conventional information technologies. The course reviews the state of the art in Big Data analytics and in addition to covering the specifics of different platforms, models, and languages, students will look at real applications that perform massive data analysis and how they can be implemented on Big Data platforms.

Text Book:

1. " Mining of Massive Datasets" by Anand Rajaraman and Jeffrey Ullman. 3rd edition, 2020, Cambridge University Press. 2. Hibbeler, Russell Charles. Mechanics of materials, Latest Available Edition Pearson

2. " Data Mining Concepts and Techniques" - Jiawei han & Micheline Kamber Harcourt.

Reference Book:

1. " Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph". Elsevier by Loshin, D.2. Arthur P Boresi "Advanced Mechanics of Materials", 6th Edition, John Wiley &Sons Inc., 2003

2. Learning Spark_ Lightning-Fast Data Analytics by Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee.

Prerequisites

None

ASSESSMENT SYSTEM FOR THEORY

Quizzes	15%
Assignments	05%
Mid Terms	30%
ESE	50%

ASSESSMENT SYSTEM FOR LAB

Quizzes	10%-15%
Assignments	5% - 10%
Lab Work and Report	70-80%
Lab ESE/Viva	20-30%

Teaching Plan

Week No	Topics	Learning Outcomes
1	Introduction	Course Outline, objectives, teaching plan, assessment method, concepts review
2-8	Big Data Algorithms	Introduction to classical data mining algorithms and their draw backs in big data environments. Revising the data mining algorithms with improvements for topics such as market based analysis, nearest neighbor search, locality sensitivity hashing, bloom filters, link analysis and PageRank, recommendation systems.
9	**MID TERM IN WEEK 9**	
10-17	Big Data Technologies	This part of this course is more focused on specific big data technologies for data analytics. In this regard, learning outcomes are understanding the topics such as MapReduce computational model, Hadoop system architecture, Apache Spark in memory data analytics, machine learning libraries at scale, and text analytics.
18	**FINAL TERM IN WEEK 18**	

Practical:

Experiment No	Description
1	Weka as a tool, installation and platform workspace development
2	Understanding and application of data filters for data pre-processing
3	Rapidminer for Data Analytics.
4	Apriori and Association Mining using Python (MLXtend)
5	Locality Sensitive Hashing using Python.
6	PageRank using Python
7	Recommender System using Python

8	Project Proposal Discussions.
9	**MID TERM IN WEEK 9**
10	Hadoop installation and MapReduce
11	Big Data processing using Spark
12	Spark for ML using MLlib
13	Spark 3 operations
14	Flink/Hbase/Pig/Hive
15	**Open-Ended Lab**
16	**Open-Ended Lab**
17	**Final Lab**
18	**FINAL TERM IN WEEK 18**