CSPE3016 BIG DATA ANALYTICS (3-0-0)

Course Objectives:

- To familiarize students with big data analysis as a tool for analyzing large complex dataset.
- Provide an overview of Apache Hadoop.
- Explore tools and practices for working with big data.
- Understand the applications using Map Reduce Concepts.
- Understand how big data analytics can leverage into a key component.

Module-I: (08 HOURS)

Introduction:

Dawn of the Big Data Era, Definition and Features of Big Data, Big Data Value, The Development of Big Data, Challenges of Big Data.

Related Technologies: Cloud Computing - Cloud Computing Preliminaries, Relationship Between Cloud Computing and Big Data, IoT - IoT Preliminaries, Relationship Between IoT and Big Data, Data Center, Hadoop - Hadoop Preliminaries, Relationship between Hadoop and Big Data.

Module-II: (08 HOURS)

Big Data Analysis:

Traditional Data Analysis, Big Data Analytic Methods, Architecture for Big Data Analysis - Real-Time vs. Offline Analysis, Analysis at Different Levels, Analysis with Different Complexity, Tools for Big Data Mining and Analysis.

Module-III: (08 HOURS)

Mining Data Streams:

Introduction to Streams Concepts, Stream Data Model and Architecture - Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream –Real time Analytics Platform (RTAP) applications, Case Studies, Real Time Sentiment Analysis- Stock Market Predictions, Big Data Generation-Enterprise Data, IoT Data, Internet Data, Big Data Acquisition- Data Collection, Data Transportation, Data Pre-processing.

Module-IV: (08 HOURS)

Big Data Storage:

Storage System for Massive Data, Distributed Storage System, Storage Mechanism for Big Data - Database Technology, Design Factors, Database Programming Model

Hadoop & Map Reduce: Data Storage and Analysis, Comparison with Other Systems, A Brief History of Hadoop, Apache Hadoop and the Hadoop Ecosystem, A Weather Dataset, Analyzing the Data with Unix Tools, Analyzing the Data with Hadoop (Map and Reduce, Java MapReduce), Scaling Out, Hadoop Streaming, Hadoop Pipes.

Module-V: (8 HOURS)

Big Data Applications:

Application Evolution, Big Data Analysis Fields - Structured Data Analysis, Text Data Analysis, Web Data Analysis, Multimedia Data Analysis, Network Data Analysis, Mobile Traffic Analysis, Key Applications - Application of Big Data in Enterprises, Application of IoT Based Big Data, Application of Online Social Network-Oriented Big Data, Applications of Healthcare and Medical Big Data, Collective Intelligence, Smart Grid.

Course Outcomes:

At the end of the course, the students will be able to:

- Process data in Big Data platform and explore the big data analytics techniques for business applications.
- Analyze Map Reduce technologies in big data analytics.
- Develop Big Data solutions using Hadoop Eco System.
- Design efficient algorithms for stream data mining on big data platform.
- Access and Process Data on Distributed File System.

Text Books:

1. Jure Leskovec, Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2014.
2. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", 2013.

Reference Books:

1. Tom White, “Hadoop- The Definitive Guide”, O’reilly, 2nd Edition.
2. EMC Education Services, "Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data", Wiley publishers, 2015.