



Minor in Data Science

Semester: VIII

Program: Common for All Programs (except Computer Science and Engineering (Data Science))

Course: Big Data Engineering (DJ19MN4C4)

Pre-requisite:

1. Database Management
2. SQL

Objectives: To introduce students to the fundamental concepts of Big data platforms and Hadoop Ecosystem.

Outcomes: On completion of the course, the learner will be able to:

1. Apply appropriate method to handle the building blocks of big data.
2. Apply fundamental techniques like Hadoop / Map Reduce in solving real world Big Data problems.
3. Achieve adequate perspectives of big data analytics in various applications like recommender systems, Social Networks & Streams.
4. Develop Appropriate Cloud computing solutions & deployment of service models.

Big Data Engineering (DJ19MN4C4)		
Unit	Description	Duration
1	<p>Introduction to Big Data and Hadoop: Introduction to Big Data, Characteristics of Big Data, Case study on Big data Solutions. Traditional Approach v/s Big Data Business Approach</p> <p>Big Data Hadoop - Concept of Hadoop, Core Hadoop Components; Hadoop Ecosystem. Hadoop Management: Apache Zookeeper, Oozie, HIVE, Apache Sqoop Tool for Data Exchange. Limitations of Hadoop.</p>	10
2	<p>Hadoop HDFS and Map Reduce: Big Data – Hadoop & Hadoop Ecosystem Distributed File Systems: Physical Organization of Compute Nodes, Large Scale File-System Organization. Details of Apache PIG, HBase, YARN MapReduce: Introduction to Map Reduce the Map Tasks, Grouping by Key & Value Pairs. Details of MapReduce Execution.</p>	10
3	<p>NoSQL: Introduction to NoSQL, NoSQL Properties, Business Drivers for NoSQL NoSQL Data Architecture Patterns: Key-value stores, Document Store, Graph stores, Column family (Bigtable) stores NoSQL Big Data management using MongoDB and Cassandra ACID Properties, CAP Theorem, Comparison SQL with NoSQL NoSQL for Big Data, Case Study on NoSQL</p>	08

4	Recommendation Systems and Social Networks: Recommendation System and its types: Collaborative filtering, content based. Case study on Product Recommendation Social Networks as Graphs, Clustering of Social-Network Graphs, Direct Discovery of Communities in a social graph	06
5	Data Streams Mining: The Stream Data Model: A Data Stream Management System Stream Sources examples Queries and issues in Stream Processing Real time messaging, ensuring scalability and concurrency using KAFKA.	08
6	Infrastructure Model: Virtualization, characteristics, types of Virtualization– Hosted & Bare Metal Introduction to Cloud, Cloud Architecture (NIST model), characteristics of Cloud, Cloud deployment models. Cloud Services: Iaas, Saas, Paas. Introduction to Kubernetes and Docker, Setting up applications on Kubernetes and Docker Implement Containerization using Docker.	10
Total		52

Books Recommended:

Text books:

1. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, Mining of Massive Datasets, Stanford Press, 2020.
2. Rajkumar Buyya, Christian Vecchiola, S ThamaraiSelvi, 'Mastering Cloud Computing', Tata McGraw-Hill Education, 1st Edition, 2017.
3. Alex Holmes, 'Hadoop in Practice', Dreamtech Press, 1st Edition, 2013.

Reference Books:

1. Herbert Jones, 'Data Analytics: The Ultimate Guide to Big Data Analytics for Business, Data Mining Techniques, Data Collection, and Business Intelligence Concepts', Bravex, 1st Edition, 2020.
2. Bill Franks, 'Taming The Big Data Tidal Wave: Finding Opportunities In Huge Data Streams With Advanced Analytics', Wiley India Private Limited, 1st Edition, 2012.
3. Jared Dean, 'Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners', Wiley, 1st Edition, 2014.
4. Bart Baesens, 'Analytics in a Big Data World: The Essential Guide to Data Science and its Applications', Wiley, 1st Edition, 2014.

Reference links:

1. NPTEL Course: <https://nptel.ac.in/courses/106104189>
2. NoSQL Database: <http://www.coursera.org/learn/nosql-databases#syllabus>
3. Big Data: <http://www.coursera.org/specializations/big-data#courses>

Prepared by

Checked by

Head of the Department

Principal