Shri Vile Parle Kelavani Mandal's

Dwarkadas J. Sanghvi College of Engineering

(Autonomous College Affiliated to the University of Mumbai)

Scheme and detailed Syllabus (DJS22)

of

Minors Degree Program

in

Data Science

(Semester VIII)

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Scheme of Semester VIII (DJS22) MINORS for Department of Computer Science and Engineering (Data Science)

Sr. No.	Course Code	Course	Teaching Scheme				Semester End Examination (SEE) - A						Continuous Assessment (CA) - B					Aggregate (A+B)	Credits Earned
			Theory (Hrs)	Practical (Hrs)	Tutorial (Hrs)	Credits	Duration (Hrs)	Theory	Oral	Pract	Oral & Pract	SEE Total (A)	Term Test 1 (TT1)	Term Test 2 (TT2)	Term Test Total (TT1 + TT2)	Term Work	CA Total (B)		
1	DJS22MN4C4	Big Data Engineering	4	-	-	4	2	65	-	-	-	**65**	20	15	35	-	**35**	100	**4**
		Total	**4**	**-**	**-**	**4**	**2**	**65**	**-**	**-**	**-**	**65**	**20**	**15**	**35**	**-**	**35**	**100**	**4**

Minor in Big Data Engineering **Semester: VIII**

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

Course: Big Data Engineering (DJS22MN4C4)

Pre-requisite: Database Systems, 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. Analyze 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 (DJS22MN4C4)

Unit	Description	Duration
1	**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 **Big Data Hadoop** - Concept of Hadoop, Core Hadoop Components. Hadoop Management: Apache Zookeeper, Oozie, HIVE, Apache Sqoop Tool for Data Exchange. Limitations of Hadoop.	**10**
2	**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.	**10**
3	**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	**10**
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 Virtualizations– 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.	**12**
	Total	**56**

Books Recommended:

Text books:

1. Alex Holmes ―Hadoop in Practice‖, Manning Press, Dreamtech Press.
2. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, Mining of Massive Datasets, 3rd Edition — Cambridge University Press, 2023.
3. RajkumarBuyya, Christian Vecchiola, S ThamaraiSelvi, "Mastering Cloud Computing", Tata 2nd McGraw-Hill Education, 2024

Reference Books:

1. Bill Franks, ―Taming the Big Data Tidal Wave: Finding Opportunities In Huge Data Streams With Advanced Analytics‖, Wiley India Private Limited.
2. Joe Reis, Matt Housley, Fundamentals of Data Engineering: Plan and Build Robust Data Systems O'Reilly, 2022
3. Bart Baesens --Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, , WILEY Big Data Series.

Web links:

1. https://nptel.ac.in/courses/106104189
2. http://www.coursera.org/learn/nosql-databases#syllabus
3. http://www.coursera.org/specializations/big-data#courses

Prepared by Checked by Head of the Department Principal