



Shri Shamrao Patil(Yadraykar) Educational & CharitableTrust's
Sharad Institute of Technology College of Engineering
(An Autonomous Institute)

NBA Accredited, Accredited by NAAC 'A' Grade, ISO 9001:2015 Certified

Introduction to Big Data

OEXXX	OEC	Introduction to Big Data	3-0-0	3Credits
-------	-----	--------------------------	-------	----------

Teaching Scheme	Examination Scheme
Lecture:3 hrs/week	CA I: 10 Marks CA II:10 Marks Mid Semester Exam:30 Marks End Semester Exam: 50 Marks

Pre-Requisites: Nil

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

CO1	Outline big data with challenges and nature of data.
CO2	Explain data streaming and sampling with its process.
CO3	Demonstrate different case studies in big data.
CO4	Summarize framework Hadoop to store and process big data.
CO5	Illustrate the applications of big data using MapReduce.
CO6	Illustrate the applications of big data using Pig and Hive.



f. Jadhav
Head of the Department
Computer Science & Engineering
SHARAD INSTITUTE OF TECHNOLOGY
COLLEGE OF ENGINEERING
Yadray (Ichalkaranji) Dist. Kolhapur.



Shri Shamrao Patil(Yadravkar) Educational & CharitableTrust's


Sharad Institute of Technology College of Engineering
(An Autonomous Institute)

NBA Accredited, Accredited by NAAC 'A' Grade, ISO 9001:2015 Certified

Course Contents:

Unit1: Introduction to Big Data Platform, Challenges of Conventional Systems, Intelligent data analysis, Nature of Data.	[6]
Unit2: Analytic Processes and Tools - Analysis vs. Reporting.Introduction To Streams Concepts, Stream Data Model and Architecture, Stream Computing, Sampling Data in a Stream, Filtering Streams, Counting Distinct, Elements in a Stream, Estimating Moments.	[6]
Unit3: Counting Oneness in a Window, Decaying Window Real time Analytics Platform (RTAP) Applications - Case Studies -Real Time Sentiment Analysis- Stock Market Predictions.	[6]
Unit4: Introduction of Hadoop- the Hadoop Distributed File System, Components of Hadoop, Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS Java interfaces to HDFS Basics- Developing a Map Reduce Application.	[6]
Unit5: How Map Reduce Works-Anatomy of a Map Reduce Job run-Failures-Job Scheduling-Shuffle and Sort – Task execution - Map Reduce Types and Formats- Map Reduce FeaturesHadoop environment.	[6]
Unit6: Applications on Big Data Using Pig and Hive – Data processing operators in Pig – Hive services – HiveQL – Querying Data in Hive - fundamentals of HBase and ZooKeeper	[6]
Reference/Textbooks:- 1) Frank J Ohlhorst, “Big Data Analytics: Turning Big Data into Big Money”, Wiley and SAS Business Series, 2013. 2) Colleen Mccue, “Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis”, Elsevier, Second Edition, 2015. 3) Michael Berthold, David J. Hand, “Intelligent Data Analysis”, Springer, Secondv Edition, 2007.	




Head of the Department
Computer Science & Engineering
SHARAD INSTITUTE OF TECHNOLOGY
COLLEGE OF ENGINEERING
Yadrav (Ichalkaranji) Dist. Kolhapur.