

SILVER OAK UNIVERSITY



College of Technology

Bachelor of Technology

Information Technology

Course Name: Big Data Analytics

Course Code: 101013436

Semester: 7th

Prerequisite:

Basics of Programming and DBMS

Objective:

1. The course aims to enable students to analyze large datasets using advanced tools, fostering the application of critical thinking to optimize processes and innovate within their fields.
2. To gain knowledge of writing mapper and reducer for a given problem and to familiarize concepts of Pig and Hive to process big data.

Teaching Scheme:

Teaching Scheme				
L	T	P	Contact Hours	Credit
3	0	2	5	4

Contents:

Unit	Topics	Teaching Hours	Weightage %
1	Introduction to Big Data: Importance of Big data, Big data Characteristics, When to consider Bigdata as a solution? Traditional Vs Big Data Business approach. Big data Use cases	5	18
2	Hadoop: HDFS file system. Design of HDFS, Java interfaces to HDFS Basics, How map reduce works? Anatomy of Map reduce, job scheduling, Shuffle and sort and Task execution Cluster setup and installation, Hadoop configuration and Hadoop in cloud.	10	25
3	No SQL Databases: Review of Traditional Databases. Need for NoSQL Databases, No SQL Case Studies, and NoSQL data Architecture patterns. Key value stores, Graph Stores, Column Family (Big Table) Stores, and Document stores. Variation of NoSQL architectural patterns, Using NoSQL to manage big data. What is a big data NOSQL solution? Choosing distribution models: Master –Slave versus peer-to-peer. Handling big data problems through SQL. Understanding the types of bigdata problems. Analyzing bigdata with shared-nothing Architecture distribution models. Four ways that NOSQL systems can handle bigdata.	10	22

4	Big Data Streaming: Introduction to Stream Concepts, Stream Data, Stream Computing, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream Real time Applications -Case Studies, Real Time Sentiment Analysis, Stock Market Predictions.	10	17
5	Pig and Hive: What is pig in hadoop, Pig Architecture and components, Comparison with Databases, pig Latin basics. Pig Latin basics. Twitter case study. Why apache hive, Installing hive. An example, hive architecture. Hive clients and services. Metastore	7	18

Course Outcomes:

Sr. No.	CO Statement	Unit
CO-1	Describe process of Big Data for large datasets.	1
CO-2	Develop application utilising the MapReduce framework and the Hadoop framework to address issues in the real world.	2
CO-3	Analyze data by applying selected technique.	3
CO-4	Illustrate the streaming and Integration strategies in big data process.	4
CO-5	Develop scripts using Pig to process large datasets and querying using hive from a data warehouse.	5

Teaching & Learning Methodology:

The various methods or tools follows by the faculties to teach the above subject are:

1. Problem Identification
2. Problem Based Learning

List of Experiments:

Total Hours: 28

Sr. No.	Practical Name
1	Prepare a case study on Big Data Use cases
2	Understanding and using basic HDFS commands
3	To understand the overall programming architecture using Map Reduce API
4	Word count application using Mapper Reducer on single node cluster
5	Analyze the weather dataset using Mapper Reducer single node cluster
6	Set up a MongoDB database and perform basic operations
7	Prepare a case study on Big Data Streaming
8	Basic Pig operations and commands
9	Daily show analysis using pig
10	Word count application using pig
11	Perform basic queries to retrieve and analyze information using hive
12	Creating the HDFS tables and loading them in Hive and learn joining of tables in Hive

Books Recommended:

1. Tom White, "Hadoop: The Definitive Guide", O'Reilly Media Inc.
2. Tanmay Deshpande, "Hadoop Real-World Solutions Cookbook", Packt Publishing.
3. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer.
4. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", Wiley.
5. Seema Acharya, Subhashini Chhellaappan, "BIG Data and Analytics", Willey.

List of Open-Source Software/learning website:

1. Cassandra: The Definitive Guide - 3rd Edition | O'Reilly | DataStax
2. Data Science at Scale | Coursera

CO-PO-PSO Matrix:

CO No.	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO-1	1											2	1	1
CO-2	2	2	3	1	3	2			2		2	2	2	1
CO-3	1	2	2	1	2				2		3	2	1	1
CO-4	1	1	2	1	2	2			1			2	1	1
CO-5		1	3		3							1	1	1