

COURSE CODE	COURSE TITLE	L	T	P	C
1152IT111	BIG DATA ANALYTICS	3	0	0	3

Course Category:

~~Foundation (0) / Program Core (1) / Program Elective (2) / Allied Elective (3) / University Elective (4) / Value Education Elective (5) / Independent Learning (6) / Industry Higher Learning Institute Interaction (7)~~

a. Preamble :

This course covers foundational techniques and tools required for data science and big data analytics. The course focuses on concepts, principles, and techniques applicable to any technology environment and industry and establishes a baseline that can be enhanced by further formal training and additional real-world experience.

b. Prerequisite Courses:

Sl. No	Course Code	Course Name
1		Operating System
2		Python
3		C# and .NET
4		Java Programming

c. Related Courses:

Sl. No	Course Code	Course Name
1		Linux
2		Distributed computing

d. Course Educational Objectives :

Learners are exposed to

- To explore the fundamental concepts of big data analytics.
- To learn to analyze the big data using intelligent techniques.
- To understand the various search methods and visualization techniques.
- To learn to use various techniques for mining data stream.
 - To understand the applications using Map

Reduce Concepts.

e. Course Outcomes :

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	Knowledge Level (Based on revised Bloom's Taxonomy)
CO1	Work with big data platform	K2
CO2	Analyze the big data analytic techniques for useful business applications.	K2, S3

CO3	Design efficient algorithms for mining the data from large volumes.	K2, S3
CO4	Analyze the HADOOP and Map Reduce technologies associated with big data analytics	K3, S3
CO5	Explore on Big Data applications Using Pig and Hive	K3, S3

f. Correlation of COs with POs :

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	M											
CO2	M											
CO3	M	M								M		
CO4	M	M								M		
CO5	M	H								M		

H- High; M-Medium; L-

Low

g. Course Content:

Theory

UNIT I INTRODUCTION TO BIG DATA 8 Introduction to BigData Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions - Re-Sampling - Statistical Inference - Prediction Error.

UNIT II MINING DATA STREAMS

9

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 – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications - Case Studies - Real Time Sentiment Analysis, Stock

Market Predictions.

UNIT III HADOOP

10

History of Hadoop- The Hadoop Distributed File System – Components of HadoopAnalyzing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFSBasics- Developing a Map Reduce Application-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 Features

UNIT IV HADOOP ENVIRONMENT

9

Setting up a Hadoop Cluster - Cluster specification - Cluster Setup and Installation – Hadoop Configuration-Security in Hadoop - Administering Hadoop – HDFS - MonitoringMaintenance-

Hadoop benchmarks- Hadoop in the cloud

UNIT V FRAMEWORKS

9

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 - IBM InfoSphere BigInsights and Streams. Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications

Total: 45 Hours

h. Learning

Resources i.Text

Books:

1. Big Data: A Revolution That Will Transform How We Live, Work, and Think by Viktor Mayer-Schoenberger & Kenneth Cukier
2. MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems