

UNIVERSITY OF MADRAS
M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE
SYLLABUS WITH EFFECT FROM 2023-2024

Title of the Paper	Data Science and Analytics		
Core-IX - Theory	II Year & III Semester	Credit: 4	536C3A

Objectives:

- Understand the Big Data Platform and its Use cases
- Provide an overview of Apache Hadoop
- Provide HDFS Concepts and Interfacing with HDFS
- Understand Map Reduce Jobs
- Provide hands on Hadoop Eco System

Outcomes:

1.	Identify Big Data and its Business Implications.	K1, K2
2.	List the components of Hadoop and Hadoop Eco-System	K2, K3
3.	Access and Process Data on Distributed File System	K3, K4
4.	Manage Job Execution in Hadoop Environment	K4, K5
5.	Develop Big Data Solutions using Hadoop Eco System	K6
K1-Remember;K2-Understand;K3-Apply;K4-Analyze;K5-Evaluate; K6-Create		

UNIT I:INTRODUCTION TO BIG DATA AND HADOOP

Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, AnalysingData with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big DataStrategy, Introduction to Infosphere BigInsights and Big Sheets.

UNIT II:HDFS(Hadoop Distributed File System)

The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, DataIngest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-BasedData structures

UNIT III:Map Reduce: Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Typesand Formats, Map Reduce Features.

UNIT IV: Hadoop Eco System: Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User DefinedFunctions, Data Processing operators.Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, QueryingData and User Defined Functions.

UNIT V: Hbase: HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.Big SQL: Introduction - Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering.

Text Books

1. Tom White “ Hadoop: The Definitive Guide” Third Edit on, O’reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.

UNIVERSITY OF MADRAS
M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE
SYLLABUS WITH EFFECT FROM 2023-2024

References

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
2. Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRC press (2013)
3. Anand Rajaraman and Jeffrey David Ulman, "Mining of Massive Datasets", Cambridge University Press,2012.
4. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley & sons, 2012.
6. Glen J. Myat, "Making Sense of Data", John Wiley & Sons, 2007
7. Pete Warden, "Big Data Glossary", O'Reily, 2011.
8. Michael Mineli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley Publications, 2013.
9. ArvindSathi, "BigDataAnalytics: Disruptive Technologies for Changing the Game", MC Press, 2012
10. Paul Zikopoulos ,DirkDeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan ,
11. "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGraw Hill Publications, 2012.

Related Online Contents

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	S	S	S	M	M	S
CO2	S	S	S	S	S	S	S	M	S	S
CO3	S	S	S	S	S	S	S	M	S	S
CO4	S	S	S	S	S	S	S	M	S	S
CO5	S	S	S	S	S	S	S	M	S	S

*S-Strong;M-Medium;L-Low