

21CSE221T - Big Data Tools and Techniques

Course Assessment Plan

Offered to M.Tech Integrated Data Science - IV Semester

Hour #	Unit	Topic	CO	BL	Instructional Method	Assessment Method
1	1	Overview of Big Data Analytics Introduction to data analytics and big data Big data mining	1	1	Classroom lecture	Cycle Test 1 Assignment 1 - HDFS Commands
2		Technical elements of the Big Data platform Analytics Toolkit, Components of the analytics toolkit Distributed and Parallel Computing for Big Data	1	1		
3		Cloud computing and Big Data In-Memory Computing Technology for Big Data	1	2		
4		Fundamentals of Hadoop	1	2		
5		The core modules of Hadoop	2	2	Lecture, Demo, Practice, Assignment	
6						
7						
8						
		Introduction to Hadoop MapReduce	2	2		
		Introduction to Hadoop YARN	2	2		
9		2	HDFS Hadoop filesystems	2	2	
10	MapReduce Analyzing data with Unix tools and Hadoop Scaling Out		3	2	Lecture, Demo, Practice, Assignment	Cycle Test 2 Assignment 2 - ● Map Reduce Programs
11						
12	2	Data Flow, Combiner Functions	3	1	Lecture	

		Hadoop Streaming				<ul style="list-style-type: none">• Pig Commands• Hive Queries• Zookeeper Cluster Commands	
13		Java Interface to Hadoop	3	2	Lecture, Demo, Practice, Assignment		
14		YARN Job Scheduling Hadoop I/O	3	2			
15		Data Integrity Compression Serialization, File based Data Structures	3	1			Lecture
16		Developing a MapReduce Application	3	2			Lecture, Demo, Practice, Assignment
17							
18							
19		Setting up a Hadoop Cluster Cluster specification and setup	3	1	Lecture		
20		Hadoop configuration YARN configuration	3	1	Lecture		
21		Introduction to Pig Installing and running pig	4	2			
22		Basics of Pig Latin Example Programs					
23		Introduction to Hive Installing and running Hive	4	2			
24		Introduction to HiveQL Create-Drop-Alter-order by-Group by-Joins					
25							
26		Introduction to Zookeeper Installing and running Zookeeper Creating different types of Znodes	4	2	Lecture, Demo, Practice, Assignment		
27	3	Flume Architecture Introduction to Sqoop	4	2	Lecture, Demo		
28		Introducing Oozie Limitations of Hadoop and overcoming the Limitations	5	2	Lecture, Demo	Cycle Test 3	
29		Apache Spark Core components and architecture of Spark	5	2	Lecture, Demo, Practice, Assignment	Assignment 3 - <ul style="list-style-type: none">• Spark Commands• Flink Streaming Analysis• MongoDB	
30							
31							
32	4	Introduction to Apache Flink	5	2			

33		Installing Flink Batch analytics using Flink				Queries • Visualization in Python and Tableau
34						
35		Big Data Mining with NoSQL Why NoSQL?				
36		NoSQL databases Introduction to MongoDB	5	2		
37		Basic queries in MongoDB	5	2		
38		Introduction to Cassandra	5	2		
39	5	Visualizing Big Data Using Python and R for visualization	6	2	Lecture, Demo, Practice, Assignment	
40		Big Data Visualization Tools				
41		Data Visualization with Tableau		2		
42		Case Studies: Hadoop-Case Studies: Spark				
43		Case Studies: NoSQL				
44		Enterprise Data Science Overview Data Science Solutions in the enterprise Enterprise data science		1	Lecture	
45		Machine Learning and AI Enterprise Infrastructure Solutions		1	Lecture	
		5	6	1	Lecture	

Course Outcomes (CO):	At the end of this course, learners will be able to:
CO-1:	Use the various tools and techniques in big data analytics
CO-2:	Apply Hadoop and related technologies to big data analytics
CO-3:	Apply MapReduce, HDFS and YARN develop big data applications
CO-4:	Develop applications using Pig, Hive and Sqoop
CO-5:	Apply Apache Spark and Flink to applications and understand the importance of NoSQL databases
CO-6:	Understand the applications of Enterprise Data Science and data visualization tools

Syllabus:

Unit1: Overview of Big Data Analytics-Introduction to data analytics and big data-Big data mining-Technical elements of the Big Data platform, Analytics Toolkit, Components of the analytics toolkit -Distributed and Parallel Computing for Big Data-Cloud computing and Big Data-In-Memory Computing Technology for Big Data-Fundamentals of Hadoop-Hadoop Ecosystem-The core modules of Hadoop-Introduction to Hadoop MapReduce-Introduction to Hadoop YARN.. 9hrs
Unit 2: MapReduce-Analyzing data with Unix tools and Hadoop-Scaling Out – Data Flow, CombinerFunctions-Hadoop Streaming-HDFS-Hadoop filesystems-Java Interface to Hadoop-YARN-Job Scheduling-Hadoop I/O-Data Integrity-Compression-Serialization-File based Data Structures-Developing a MapReduce Application. 9hrs
Unit 3: Setting up a Hadoop Cluster-Cluster specification and setup-Hadoop configuration-YARN configuration-Introduction to Pig-Installing and running pig-Basics Pig Latin -Example Programs-Introduction to Hive-Installing and running Hive-Introduction to HiveQL-Create-Drop-Alter-order

by-Group by-Joins-Introduction to Zookeeper-Installing and running Zookeeper-Creating different types of Znodes-Flume Architecture-Introduction to Sqoop. 9hrs
Unit 4:Introducing Oozie-Apache Spark-Limitations of Hadoop and overcoming the Limitations-Core components and architecture of Spark-Introduction to Apache Flink-Installing Flink-Batch analytics using Flink-Big Data Mining with NoSQL-Why NoSQL?-NoSQL databases-Introduction to MongoDB,-Basi queries in MongoDB-Introduction to Cassandra. 9hrs
Unit 5: Enterprise Data Science Overview-Data Science Solutions in the enterprise-Enterprise data science – Machine Learning and AI-Enterprise Infrastructure Solutions-Visualizing Big Data-Using Python and R for visualization-Big Data Visualization Tools-Data Visualization with Tableau-Case Studies: Hadoop-Case Studies: Spark-Case Studies: NoSQL. 9hrs

Course Assessment Table:

	Bloom's Level of Thinking	CLA – 1 (50%)	CLA – 2 (10%)	Final exam (40% Weightage)
		Theory	Theory	Theory
Level 1	Remember	50%	50%	50%
Level 2	Understand	50%	50%	50%
Level 3	Apply	-	-	-
Level 4	Analyze	-	-	-
Level 5	Evaluate	-	-	-
Level 6	Create	-	-	-
	Total	100 %	100 %	100 %