UE18CS322: Big Data (4:0:0:0:4)

# of Hours: 56

Class #	Chapter Title/Reference Literature	Topics to be Covered	% of Portions Covered	
			Reference Chapter	Cumulative
1.	**Unit: I** **Introduction** **T1**	Big Data definition, Challenges and opportunities with Big Data	21.4%	**21.4%**
2.		Data intensive scientific discovery and the role of Big Data, History		
3.		Map Reduce – Storage (HDFS)		
4.		Map Reduce – Computation model, Map Reduce architecture,		
5.		Demo class: Map-Reduce – Hands on programming		
6.		Case Study: Google. YARN introduction.		
7.	**Unit: II** **Big Data Infrastructures for Compute/Storage** T2	Overview of Hadoop Ecosystem	21.4%	**42.8%**
8.		Introduction to sample Big Data Algorithms – matrix multiplication.		
9.		Introduction to sample Big Data Algorithms - Pagerank computations		
10.		Relational operators on Map-reduce,		
11.		HIVE with hands on		
12. .		case study: Other storage - Hbase/Cassandra		
	Unit : III **In Memory Computation** T3	Issues with Hadoop, Spark and Scala	21.4%	**64.2%**
13.		PySpark programming model		
14.		Transformations and Actions, Spark SQL		
15.		Spark architecture – RDD, DataFrames, Wide and Narrow dependencies,		
16.		Complexity of Big Data algorithms – Communication Cost complexity model.		
17.		Spark HandsOn		
18.	**Unit : IV** **Streaming analysis** **T1,T2**	Streaming analytics use cases, Streaming Spark,	17.8%	**82.1%**
19.		Kafka – use cases, architecture		
20.		Streaming Algorithms - Sampling, set membership		
21.		Kafka with HandsOn		
22.		Streaming Algorithms - Bloom Filters, Counting Counting unique elements – Flajolet Martin Algorithm.		
23.	**Unit : V** **Advanced Analytics on Big Data**	Clustering Algorithms - kmeans and collaborative filtering	17.8%	**100%**
24.		Scaling Neural Networks for Big Data, case study MLLib.		
25.		Project Work		
26.		Project Work		
27.		Project Evaluations		
(Note: Each class is of 2 Hour duration.)				

Literature

Book Type	Code	Title & Author	Publication Information		
			Edition	Publisher	Year
Text Book	T1	Big Data Analytics, Rajkamal, Preeti Saxena,	1st	McGraw Hill Education	2019
	T2	Big Data Simplified, Sourabh Mukherjee, Amit Kumar Das, Sayan Goswami	1st	Pearson	2019
Reference Book/Papers	R1	Mining of Massive Datasets, Anand Rajaraman, Jure Leskovec, Jeffrey D. Ullman	2nd	Cambridge University Press	2014
	R2	Big Data Analytics Beyond Hadoop: Real-Time Applications with Storm, Spark, and More Hadoop Alternatives, Vijay Srinivasa Agneeswaran	1st	Pearson	2014
	R3	Hadoop: The Definitive Guide, Tom White	4th	O'Reilly	2009