






BDA Lectures Taken by RS-July2022-Oct2023		
Sr#	Date	Topics
 1	04/07	Introduction to big data and big data Analytics
		Data Analytics v/s Data Analysis, Example of Big data in different domains
		Types of digital data, sources of data and V's of bigdata
 2	07/07/2022	Why Big Data (Topic no 2.7), Types of analytics
		Data Warehouse, When to use big data Analytics and when to not use it
		Intro to NoSQL, types and Vendors, Intro to OLAP,OLTP
 3	12/07/2022	Master Slave Architecture
		Introduction to Map Reduce ,WordCount Example
		Categorizing books into "Large Books" and "Small Books" by page numbers
 4	18/07/2022	MapReduce Phases: Mapper(Record Reader, Map, Combiner, Partitioner)
		Reducer (Shuffle, sort, Reduce)
		Input format, Output Format, Input splits, Byte offset
		SQL vs NoSQL, Hadoop vs SQL, SQL vs MapReduce
 5	19/07/2022	Hadoop version 1 vs Hadoop version 2 LEFT
		YARN components, Application submission in YARN
		Hadoop Ecosystem and its components
		NewSQL and, SQL vs NoSQL vs NewSQL
Sessional 2		
 6	23/08/2022	Sessional 1 paper discussion
		Introduction to mongoDB, RDBMS and MongoDB terminology mapping
		Replication, Sharding and in-place updation in mongoDB
		Different datatypes available in MongoDB & Unique key generation
		Basic commands (insert, save)
 7	29/08/2022	MongoDb basic CRUD operation (update, delete, remove, find, pretty)
		Array in mongoDb, upsert functionality
		Map reduce mongoDB example
Session3		
8	13/09/2022	Working with MongoDB arrays, sort, limit, skip and count methods












		Cursors, embedded documents example
		Denormalized document (using reference) example
9	20/09/2022	Data preprocessing steps(Data Cleaning , Data Integration
		Data reduction, Data transformation), Complex types of data
		Types of clustering methods and clustering requirements
10	03/10/2022	CLIQUE-Grid based clustering Algorithm
		Decision Tree Induction Algorithm
		Dark Data,Social Impacts of Data Mining
11	04/10/2022	Attribute Oriented Induction
		Time series Analysis,Data summarization techniques, Sequence Pattern Mining GSP (Generalized Sequential Pattern Mining) SPADE (Sequential Pattern Discovery using Equivalence Class)
		Issues with Data Mining

BDA Lectures Taken by RJ-July2022-Oct2023				
Sr#	Date	Topics		
1	06/07/22	Introduction to big data and BDA Data Analytics v/s Data Analysis, Example of BDA in different domains Types of digital data, sources of data and characteristics of bigdata overview of syllabus		
2	11/7/22	Hadoop introduction, features, versions, advantages, overview of Hadoop ecosystem, hadoop high level architecture, HDFS, NameNode, DataNode, anatomy of File read & write. Example to understand the need of Hadoop		
3	25/7/22	CAP theorem, Difference between ACID and BASE theorem MapReduce Job Execution Flow, Failed task vs Killed task Key value pairing, Input files, Types of Input format, InputSplits, Block Vs Inputsplit, Job vs Task		
4	27/07/2022	human generated data vs machine generated data, importance of right question. Introduction to Apriori algorithm, Mapreduce : Apriori, Apriori example		
Sessional 2				
5	08-08-2022	Introduction to Cassandra, Comparison of NoSQL Databases, Data Center Features of Cassandra, Replication Factor, Peer-to-peer Network, Gossip protocol Cassandra's consistency level, Anti Entropy, Read repair, CQL data types		
6	22-08-2022	Hinted Handoffs, Writes in Cassandra, Cassandra Read Path, Partitioner, Partition key, Tokens, Virtual node, Replication strategies, Keyspaces, Brief steps for Installation of Cassandra, Commands to create, drop, describe, use Keyspace Commands to create, drop, describe Table	Left Cassandra read paths	
7	30-08-2022	Demonstration of commands like help, Create, drop, describe, use on Keyspace and table, Primary key, Table sorting based on one column and Two columns, Imbalanced Partition and how to handle it, Read and Write Consistency level and setting them.		

		Insert and select operations, Timestamps, Update Operation, TTL, CRUD.		
	Session3			
8	12-09-2022	collections in Cassandra (set, list, map), Alter commands, Counter		
		Import from CSV, STDIN		
		Export to CSV, STDOUT, Querying system tables,		
9	19-09-2022	Introduction of Mahout, Machine Learning and its applications ,		
		installation of Mahout, Maven repository and dependency, clustering, classification,		
		Recommendation: Engine, Architecture, and building recommender with example(lab 9)		
10	26-09-2022	Hive : Introduction (case study), Features, Data units, Architecture		
		Data types, collections, RCFile, HQL Commands :Database & table creation,		
		Querying table, Views, Aggregation		
11	27-09-2022	Partition (static, dynamic), Bucketing, Sub Query, Joins(Inner , full, Left, Right)		
		Serde, User-defined Function, Hive vs Pig		

BDA Lectures Taken by JMP-July2022-Oct2023-MMH				
Sr#	Date	Topics		
1	06-07-2022	Programming paradigm - structured, oop, functional, map-reduce		
		horizontal (Scale out) vs vertical (Scale Up) scaling		
		Google Research paper heading and links		
		Moving code to data. Locality of reference.		
2	07-07-2022	Data importance - quotes		
		Swayam/NPTEL Courses by IITs; Codebasics by Dhaval Patel		
		git source code of hadoop commons by author - Dough Cutting		
		hadoop framework setup pre-requisite - JDK		
3	13-07-2022	pwd; whoami; hostname; cat /etc/hosts; echo \$PATH; cat ~/.bashrc; echo \$HADOOP_HOME; tree /opt/hadoop; start_dfs.sh; jps;		
		Distributed file system, block, replication factor.		
		R script, install packages and demo word/tagcloud of scrapped data		
		Hadoop start-stop-monitor		
4	14-07-2022	similarity/difference between master-worker and client-server architecture		
		Psuedo Distributed mode hadoop setup		
		Distributed file system interaction (very much like use of google drive) help		
		hadoop dfs -ls /		
5	20-07-2022	hadoop dfs -mkdir -p /user/hadoop		
		hadoop dfs -put /home/hadoop/file.txt /user/hadoop		
		hadoop dfs -cat /user/hadoop/file.txt		
		gui based file utility and dr.who		
6	21-07-2021	uri with hdfs://, file:///		
		absolute and relative ways of accessing files		
		Highlevel steps of running map-reduce Word Count java program		
		demonstration, hadoop classpath		
7	27-07-2022	start-yarn.sh 8088 and start historyserver 19888		
		rpc remote procedure call demo		
		commerce and big data analytics		
		customer churn, serial returner, market basket analysis		
7	27-07-2022	Knowing your data; Why?		

		central tendency: mean, mode, median, range, midrange, iqr, quartiles		
		Skewness (positive/negative)		
		histogram and scatter plot		
		five-number summary using five-number summary api and summary api		
		$y = mx + c$ example for data prediction		
8	28-07-2022	boxplot draw		
		positively and negatively co-related data		
		unrelated data		
		types of data sets		
		types of attributes		
Sessional 2				
9	10-08-2022	HBase case study		
		Columnar database purpose, column family example		
		problem statement: mapping of every student's old number and new roll number provided student id with each.		
10	17-08-2022	HBase setup steps and related concepts		
		quorum, zookeeper, regionserver(worker)		
		column family example income (salary+incentives(bonus+stocks))		
		HBase API and HBase Shell		
11	18-08-2022	HBase shell command example		
		apache maven as a build tool for Test Driven Development		
		Introduction to terasort competition and benchmarking		
12	24-08-2022	internal sorting vs external sorting vs terasort		
		demonstration of TDD and maven using eclipse and command line		
		exercise: Bringing together student's properties using join and map-reduce		
13	01-09-2022	MongoDB data modeling (denormalized/embedded and normalized)		
		MongoDB aggregate api		
		sort		
Session 3				

	14	10-09-2022	Introduction to Apache Spark		
			Architecture of Spark		
	15	14-09-2022	Spark streaming, Windowing protocols		
			Spark SQL		
	16	22-09 Thur	Jaspersoft Report		
			drilldown, pagination, sorting, downloads, etc.		
	17	27-09 Tues	Spark installation and pyspark/spark-shell with scala demo		
			spark deployment options, rdd and parallelize		
	18	29-09 Thur	spark - ways to create RDD (parallelize, file read, from existing RDD and api		
			cricket match simple analytics using spark		
			transformation, action, lazy evaluation, repartitioning, coalesce.		
	19	10-04 Tue	Spark architecture (driver, application, context, resource manager, executor)		
			partition, job, stage, task w.r.to DAG		
			Purpose of DAG in lazy evaluation		
			wordcount example using spark		
			Narrow and wide transformations		
	20	6/10/2022 Thur	Wordcount example using spark demonstration with HDFS data storage		