BDA Lectures Taken by JMP July-Oct 2021 https://drive.google.com/drive/folders/1Q7iGqSlrnvGEloTDDrNZFjy8zHCXo5zm?usp=sharing https://classroom.google.com/c/MzUyNDg4MDQ0OTNa?cjc=pxva2mi			
Lecture#	**Date**	**Topics**	
1	05/07/2021	Syllabus and chapters overview, why big data	Sess1
		Programming paradigm - structured, oop, functional, map-reduce	
		horizontal (Scale out) vs vertical (Scale Up) scalling	
		Client Server Vs Master Slave, BDA-IoT-Cloud-ML (Sensors to cloud algos and back to Actuators)	
		Google Research paper heading and links	
		Swayam courses	
		structured, semi and unstructured	
2	09/07/2021	git source code of hadoop commons by author - Cutter	
		wordcount program; wc on linux vs wordcount as base for analytics BigData	
		Moving code to data. Locality of reference.	
		Real time traffic on roads - Big Data Analytics application	
		Crowdsouring - Big Data Analytics application; p.s. outsourcing, cloudsourcing.	
		Introduction of HDFS and MR daemons. Highlevel steps executing WordCount java program video 11mins.	
		Function as first class citizen and RTAP as latest data handling and analytics	
		Characterstics of Big Data (Many Vs)	
3	12/07/2021	Big Data definition by Gartner	
		High level aims w.r.to BigData and Analytics	
		History of Hadoop. Apache Hadoop	
		WordCount connection to WordCloud/TagCloud example	
		Introduction of Map-Reduce Framework	
		Prototype of map, reduce;	
		Assignment: Finding mutual friends between every two. Problem statement shared.	
4	16/07/2021	Data Orchestration and Big Data Architecture	
		Commerce and Big Data Analytics: Customer Churn Rate, Serial Returener, Market Basket Analysis, Promotions/Coupans	
		WordCount example explanation with combiner;	
		Oracle cloud - compute instance ssh	
5	19/07/2021	Research: Characteristics of BigData, Big Data Security and Privacy (IT Act, PDPB)	
		Introduction to problem statement: Finding Mutual Friend - MR solution discussion; Graph to Lines problem discussed.	
		Analytics and Industry Revolutions	
		Introduction to problem statement: Categorizing books example explanation	
		Anatomy of Data Science: Business Acument, Maths and Technology Expertise.	
6	23/07/2021	MR Java solution to categorizing books	
		Solution discussion to finding mutual friends MR example	
		Hadoop Eco System and Related Components	
		CAP Theorem (ACID and BASE)	
		SQL , NoSQL and NewSQL	
7	26/07/2021	NewSQL and comparisons	
		Data at rest vs data in motion, types of cloud	
		Clickstream Data Analytics as an applicatoin of BDA	
		Mapper and Reducer API - Life Cycle	
		Big Data and Data Modeling	

		Compression and Archival with Hadoop	
8	30-07-2021	Apache Maven usage with Eclipse and CLI - Hadoop Library	
		Understanding tera sort - hadoop benchmarking - Introduction	
		Importance of right question for right data in general	
		Case study - Introduction to HBase as Distributed Database and Column Oriented NoSQL	
		set up and configruation. hbase shell command line and gui to verify tables	
9	09/08/2021	Introduction to mongoDB	Sess2
		To have more sustainable solutions - "The bridge on the river Choluteca"	
		Demonstration of MongoDB daemon and connecting via mongo Shell, mongo Compass and Jupyter Notebook with imongo kernel	
		Database, collection, document mapping w RDBMS terminology	
		basic commands	
		_id to be primary key column by default with hex code or may override as your own primary key column	
		Data Model Design — MongoDB Manual (Embedded or Normalized)	
10	13-08-2021	Data Silos Vs Data Lakes	
		MongoDB architecture guide - sharding policies	
		Describe technical word resilient, workload, etc.	
		MongoDB datatypes i.e. code, regular expression, object id, etc.	
		_id internal, upsert flag, crud, pretty, sort, find, count.	
11	20-08-2021	Latency Vs Throughput	
		Schema on Write Vs Schema on Read	
		High level aims such as credict scores, bank loan defaulters, rare events	
		Approach of looking at company data repos vs looking social media, etc for feedback.	
		mongoDB MapReduce Example - orders - find total purchase amount group by cust id	
		Similar solution by mongoDB aggregation api	
		return value of reduce function - boolean. Job class with org.apache.hadoop. mapreduce package	
		return value of reduce with mongoDB	
12	23-08-2021	MongoDB crud operations	
		remove, find, pretty, upsert, multi, or, project using 0/1, pattern matching using regex, filter using find. etc	
13	27-08-2021	mongoDB adding new field, save(), relational operator, unset, null, in, regex	
		count, limit, sort, skip	
14	02-09-2021	in vs nin	
parta-from lab hours		mongo import and export	
		aggregate function	
		arrays push, pop, etc.	
partb	03-09-2021	javascript programming	
		cursors in mongodb	
		Automatic generation of unique numbers for the _id field	
		indexes and explain plan	
			Session 3
15	13-09-2021	Hive Introduction	
		Hive Architecture, Data Model - Database, table, partition and bucket	

		Pig/Hive with Apache Tez Vs Pig/Hive MapReduce	
		Managed Vs External Type of tables	
		Data Types and File Formats with Hive	
		Pig Vs Hive	
16	17-09-2021	Installation steps of hive from GettingReady Hive url	
		DDL	
		DML	
		Use of local, external, partition, clustered, skewed like keywords	
		pig philosophy, data types	
17	20-09-2021	Tumbling Windows Vs Sliding Windows - Windowing Protocol for stream data analytics	
		Example using Amazon Kinesis	
		IBM EventDetection Demo for IBM stream data analytics - Weather	
		Clustering	
		poster making - Group Activity - Movie Industry and Big Data Solutions	
18	24-09-2021	Fundamentals of Distributed Systems: architecture, inherent limitation, Distributed File System.	
		Design - Loosely coupled hardware and tightly coupled software with middleware services on top of Network OS	
		Distributed Systems Design Issues	
		MPI and RPC	
19	27-09-2021	Lamport's logical clock and background	
		Limitation of lamport's logical clock- Unable to logically connect if not causaly related	
		Distributed Shared Memory - architecture diagram with memory mapper	
		Load Sharing Vs Load Balancing	
		quorum based Maekawa's DME algo	
		Token in a DME - A data structure to hold info. Challenges involved in maintaing uniqeness of token.	
20	01-10-2021	Apriori Algo: frequent item sets, association rules, support, confidence.	
		Solution typescript view of apriori algo Oracle PL/SQL implementation	
		Diagram showing map-reduce solution to Apriori	
		Research paper brief "Apriori Versions Based on MapReduce for Mining Frequent Patterns on Big Data" - IEEE TRANSACTIONS ON CYBERNETICS, VOL. 48, NO. 10, OCTOBE	
		https://www.win.tue.nl/~mpechen/publications/pubs/Luna2018a.pdf	
		3 pillars of data science: Linear Algebra, Statistics, Optimization	
21	04-10-2021	BDA-Trends	
		Glossary part	
22	08-10-2021	high level aims' perspectives	
		visualization api by ibm watson studio	
		student submission statistics to start with	
		Google Sheet's Explore Approach - Bottom Right of Google Sheets	
		terasort partitioning solution	