

Course Code	Course Title	L	T	P	C
OLMCA606	Big Data Analytics	3	1	0	4
Pre-requisite	NIL	Syllabus version			
		1.0			
Course Objectives:					
1. To understand the challenges confronting the enterprises embracing big data					
2. To use open-source framework like Hadoop and analyze using MapReduce					
3. To introduce big data store technologies to store different formats of data					
Course Outcomes:					
1. Identify the types of data and analysis methods					
2. Use Hadoop for large data storage					
3. Adapt efficient MR algorithms for analysis					
4. Create and manipulate data in NoSQL Db					
5. Operate on unstructured data and streaming data					
Module:1 Big Data Cycle					
Evolution of Big data – types and sources of data-characteristics- analytics cycle-roles in analytic projects- Big data challenges and applications in Industries-Different types of dataset and analytics methods					
Module:2 Big Data Technologies-Hadoop					
Features, advantages, History, HDFS-file read/write, HDFS commands, Hadoop ecosystem					
Module:3 MapReduce Programming					
Stages of MR-reducer, combiner and partitioner, classic framework and YARN, MapReduce Daemons, Example, MapReduce based analysis algorithms-Matrix multiplication, SQL-join query					
Module:4 Big Data Technologies-NoSQL					
NoSQL-Types, advantages, SQL Versus NoSQL, HBase and Hive, Hive architecture, file formats, HQL-DDL, DML, HBase architecture, HBase and RDBMS, HBase commands-DDL, DML					
Module:5 Analysing Unstructured Data					
Structure of Mongo DB, replication, sharding, JSON files, import/Export, CRUD operations- create/drop DB, find documents, aggregate functions, MR functions, indexes					
Module:6 Analysing Streaming Data					
Features of Cassandra-replication factor, streaming data and indexes, CQL-key space, datatypes, collection, import /export csv file, CRUD operations-create or alter table/ key space, read or modify or delete collection and other types					
Module:7 Big Data and Machine Learning					
Machine Learning- algorithms, difference between ML and DA, algorithms using MR in clustering and association rule mining, Big Data trends- edge computing, chatbots, visualization					
		Total Lecture hours:		120 hours	
Text Book(s)					
1.	Seema Acharya, Subhashini Chellappan " Big Data and Analytics", 2019,2 nd edition, Wiley India Pvt Ltd.				

Reference Books			
1.	DT Editorial Services, "Big Data (covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization) Black Book" Dreamtech Press, 2017.		
2.	Tom White, Hadoop, the Definitive guide, O'Reilly Media, 2015.		
3.	Bart Baesens, "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", Wiley Publishers, 2015.		
Mode of Evaluation: Quiz / Assignment / Mid-Term, FAT			
Recommended by Board of Studies		01-11-2023	
Approved by Academic Council		No. 72	Date: 13-12-2023