LNMIIT, Jaipur Department of Computer Science & Engineering	LNMIIT The LNM Institute of Information Technology

Programme: B. Tech. CSE/CCE	**Course Title:** Introduction to Big Data			**Course Code:**
Type of Course: SLI Program Elective	**Prerequisites:** Admissibility to SLI			**Total Contact Hours:** 2(L)+2(Lab) Equivalent
Year/Semester: 4 /8	**Lecture** **2**	**Tutorial Hrs/Week:** **0**	**Average Practical Hrs/Week: 2**	**Credits:** **2-0-2-3**

Course Context and Overview:
Extreme data volume, velocity, and variety challenge conventional data-processing platforms and practices. Big data discipline trades some advantages of the established approaches to surmount the limitations of conventional storage infrastructures, data structures, databases and algorithms.
The course provides an understanding of the needs, purposes, and characteristics of the Big Data domain. The students will gain an understanding of the platforms for executing big data applications, algorithms, and analytical libraries.
Hadoop and Spark frameworks will guide the students in learning about the execution platforms that grow linearly with the problem size. The students will also learn how these systems stay resilient and tolerant against failures. The programming language Scala will be introduced as it provides the base for building Apache Spark Analytical libraries. The libraries contain algorithms and techniques for solving big data problems.
On successful completion, the students will be ready to continue learning big data tools, algorithms, and libraries for handling Streaming data, NoSQL and SQL databases, Machine Learning, Frequent Pattern Growth Algorithms, and Graph-based Analytics.
This subject is a hands-on self-study elective course requiring the students to demonstrate independent learning, regular on-computer exercises and program implementations.

Prerequisites Courses:
Operating systems, Programming, Introduction to Data Sciences, Design and Analysis of Algorithms.

Course outcomes (COs):

On completion of this course, the students will have the ability to:		**Bloom taxonomy Level**
CO-1	Explain the purpose, concepts, and characteristics of big data applications	2
CO-2	Implement big data applications appropriate to the maturity level of an undergraduate student.	3
CO-3	Explain multi-computer clusters available for big-data needs using open-source software (for example, Hadoop and Apache Spark).	2
CO-4	Understand solutions using Analytics libraries available through Apache Spark	2

Week (Date of intro session)	**Contents Topics**	**Lectures + Labs (Equivalent) Lectures + Labs**	
Preliminary/Introduction			
1 13 Jan 2024	Meaning and implications of "big" in big data. Three Vs: Volume, Velocity, Variety. Other properties of big data.	**2**	**0**
2 20 Jan 2024	Multi-computer processing. Java RMI.	**2**	**4**
3 27 Jan 2024	Examples of big data applications. Prepare and submit a big data proposal.	**2**	**2**
Hadoop Infrastructure			
4, 5 3 & 10 Feb 2024	Hadoop framework – HDFS, MapReduce paradigm, Combiner	**5**	**0**
6 17 Feb 2024	Single-node Hadoop setup	**1**	**6**
7 24 Feb 2024	Running word count problem on Hadoop setup. Demonstration.	**2**	**2**
Midterm Examination			
8 16 Mar 2024	Fault-tolerance in Hadoop. Pseudo-distributed setup and word count problem		
9 23 Mar 2024	Frequent Item Set problem; approaches to run the algorithms as a big data exercise	2	**2**
10 30 Mar 2024	Overview of YARN	**2**	0
Resilience Distributed Datasets and SPARK			
11 6 April 2024	Introduction to Scala. Install Scala and practice some Scala code	**2**	**2**
12 13 April 2024	Spark basis, Spark execution model, Install Spark and run examples	**2**	**3**
13 20 Apr 2024	RDD, RDD Frames, RDD Sets	**2**	**0**
14 27 Apr 2024	Wrap-up report: Prepare the final report describing the lessons learned.		**5**

Textbook References (IEEE format) :

Text Book:

1. [BD] Rathinaraja Jeyaraj, Ganeshkumar Pugalendhi, Anand Paul, Big Data with Hadoop Map Reduce: A Classroom Approach, Apple Academic Press,2020.
2. [DL] Doug Lea, Concurrent Programming in Java: Design Principles and Patterns, 2nd Edn, The Java Series, Addison-Wesley, Boston, 2000.
3. [HADOOP] http://hadoop.apache.org/
4. [SPK] Bill Chambers and Matei Zaharia, SPARK: The Definitive Guide, O'Reilly Media, Inc, 2017.
5. [SCALA] https://www.scala-lang.org/
6. [SPARK] https://spark.apache.org/
7. [TW] Tom White, Hadoop: The Definitive Guide, 4th Edition, O'Reilley, 2015.

Evaluation Method:

Evaluation Method	
Item	**Weightage (%)**
Quiz (2)	12
Progress reports and Assigned Essays (3)	8+7+8=23
Midterm	25
Final Examination	40

CO and PO Correlation Matrix

For CSE Students

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	3	2	3	1	1		3	3	3	3	3	3	3
CO2	3	3	3	2	3	1	1		3	3	3	3	3	3	3
CO3	3	3	3	2	3	1	1		3	3	3	3	3	3	3
CO4	2	2	2	1	3	1	1		2	2	2	3	2	2	2

For CCE Students

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	3	2	3	1	1		3	3	3	3	1	2	3
CO2	3	3	3	2	3	1	1		3	3	3	3	1	2	3
CO3	3	3	3	2	3	1	1		3	3	3	3	2	2	3
CO4	2	2	2	1	3	1	1		2	2	2	3	2	2	2

Last Updated On: 20 December 2023

Updated By:

Approved By: