

Course Code	Course Name	Theory	Practical	Tutorial	Theory	Practical & Oral	Tutorial	Total
ITL801	Big Data Lab	--	02	--	--	01	--	01

Course Code	Course Name	Examination Scheme							
		Theory Marks				Term Work	Oral & Practical	Oral	Total
		Internal assessment			End Sem. Exam				
		Test1	Test2	Avg. of two Tests					
ITL801	Big Data Lab	--	--	--	--	25	--	25	50

Lab Objectives: Students will try:

1. To introduce the tools required to manage and analyze big data like Hadoop, NoSql
2. To impart knowledge of Map reduce paradigm to solve complex problems Map-Reduce.
3. To introduce several new algorithms for big data mining like classification, clustering and finding frequent patterns.
4. To introduce to the students several types of big data like social media, web graphs and data streams.
5. To identify various sources of Big data
6. To enable students to have skills that will help them to solve complex real-world problems in for decision support.

Lab Outcomes: Students will be able to:

1. Demonstrate capability to use Big Data Frameworks like Hadoop
2. Program applications using tools like Hive, pig, , NO SQL and MongoDB for Big data Applications
3. Construct scalable algorithms for large Datasets using Map Reduce techniques
4. Implement algorithms for Clustering, Classifying and finding associations in Big Data
5. Design and implement algorithms to analyze Big data like streams, Web Graphs and Social Media data and construct recommendation systems.
6. Apply the knowledge of Big Data gained to fully develop a BDA applications for real life applications.

Prerequisite: Java, Python

Requirement

Hardware	Software
PC i3 or above, 8 GB RAM	Virtual Machine, Hadoop Frame work, NOSQL and MongoDB Compilers

Detailed syllabus:

Module	Detailed Content	Hours 2hrs	LO Mapping
1	Assignment on Study of Hadoop ecosystem	02	LO 1
2	Programming exercises on Hadoop Using Hive, Pig, Hbase Sqoop NOSQL, MongoDB	04	LO 2
3	Implementing simple algorithms in Map-Reduce Matrix multiplication, Aggregates, joins, sorting, searching etc.	04	LO3
4	Implementing Algorithms using MapReduce (Any 2) <ul style="list-style-type: none"> Implementing Frequent Item set Mining Implementing Clustering algorithms Implementing Classification Algorithms 	06	LO 4
5	Big Data Applications (Any 2) <ul style="list-style-type: none"> Implementing Analytics on data streams Implementing Social Network Analysis Algorithms Implementing Web Graph Algorithms Implementing recommendation Engines 	05	LO 5
6	Mini Project: One real life large data application to be implemented (Use standard Datasets available on the web) a) Twitter data analysis b) Fraud Detection c) Text Mining d) Recommendation Engines (list of datasets also given in the text book)	05	LO 5 LO 6

Text Books:

1. Radha Shankarmani, M Vijayalakshmi, "Big Data Analytics", Wiley Publications,
2. Alex Holmes "Hadoop in Practice", Manning Press, Dreamtech Press.
3. Professional NoSQL Paperback, by Shashank Tiwari, Dreamtech Press
4. MongoDB: The Definitive Guide Paperback, Kristina Chodorow (Author), Michael Dirolf, O'Reilly Publications

References:

1. Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, Bart Baesens , WILEY Big Data Series.
2. Hadoop: The Definitive Guide by Tom White, O'Reilly Publications
3. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services
4. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence by Pramod J. Sadalage, Addison Wesley

Term Work:

Term Work shall consist of at least 10 to 12 practical's based on the above list. Also Term work Journal must include at least 2 assignments.

Term Work Marks: 25 Marks (Total marks) = 15 Marks (Experiment) + 5 Marks (Assignments) + 5 Marks (Attendance)

Oral Exam: An Oral exam will be held based on the above syllabus.