MANIPAL INSTITUTE OF TECHNOLOGY

Manipal University, Manipal
Karnataka -576104

COURSE PLAN

Department : **CSE**
Course Name & code : **CSE 449 Big Data Analytics**
Semester & branch : **7th CSE**
Name of the faculty : **Dr. Dinesh Acharya U, Mr. Roshan David**
No of contact hours/week : **4**

ASSESSMENT PLAN:

1. **In Semester Assessments - 50 %**
   - Written tests : 40%
   - Surprise quizzes : 10%

2. **End Semester Examination - 50 %**
   - Written examination of 3 hours duration (Max. Marks: 50)

Portions for Assignment	
Assignment no.	**Topics**
1	**L1-L13**
2	**L14-L25**
3	**L26-L40**
Portions for Sessional Test	
Test no.	**Topics**
1	**L1-L17**
2	**L18-L35**

Course Plan

L. No.	Topics
L0	Introduction to the Course
L1	Types of digital data, Introduction to Big Data
L2	Introduction to Big Data Analytics
L3	Core Hadoop components
L4	Hadoop Ecosystem
L5	Introduction to NoSQL
L6	Types and Advantages of NoSQL , Comparison of SQL, NoSQL and NewSQL
L7	MongoDB: Features
L8	MongoDB: Data Types
L9	MongoDB: Query Language
L 10	Cassandra: Features
L11	Cassandra: Data types
L12	Cassandra: Query Language
L 13	MapReduce with NOSQL as a data source
L 14	Hive: Basics
L 15	Hive: Architecture
L16	Hive: Data Types
L 17	Hive: File Formats
L18	Hive: Query Language
L19	Hive: User Defined Function
L 20	Pig: Introduction
L 21	Pig: Pig Latin
L 22	Pig: Data Types
L23	Pig: Data Processing operators
L 24	Word count example using Pig
L25	Data stream processing with Spark
L 26	Manipulating and Processing Data in R

L27	Programming in R
L 28	Graphical Analysis in R
L29	Statistical Analysis in R
L 30	RHadoop Architecture
L31	Understanding the Data Analytics Project Life Cycle
L32	Data Analytics Case Studies: Exploring Web paes categorization
L33	Data Analytics Case Studies: Stock market Analysis
L 34	Data Analytics Case Studies: Predicting the sale price
L 35	Machine Learning with R and Hadoop – Linear Regression
L 36	Machine Learning with R and Hadoop – Logistic Regression
L37	Machine Learning with R and Hadoop – Recommendation Algorithms
L38	Applying MapReduce patterns to Big Data: Map-side Joins
L39	Applying MapReduce patterns to Big Data: Reduce-side Joins
L40	Applying MapReduce patterns to Big Data: Sorting
L41	Applying MapReduce patterns to Big Data: Sampling
L42	YARN Overview
L43	YARN and MapReduce
L44	Understanding I/O in MapReduce
L45	Processing common serialization formats
L46	Big data serialization formats
L47	Organizing data in HDFS
L48	Optimizing data in HDFS

References:

1. Seema Acharya, Subhashini Chellappan, "Big Data and Analytics", Wiley India Pvt. Ltd., 2015

2. Alex Holmes, "Hadoop in Practice", 2nd Edition, Manning Publications, 2015

3. Vignesh Prajapati, "Big Data Analytics with R and Hadoop", 1st Edition, Packt Publishing, 2013

4. Tom White, "Hadoop: The definitive guide", 4th Edition, O'reilly, Yahoo

Press, 2015
5. Pramod J Sadalage , Martin Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence", 1st Edition, Addison-Wesley , 2012

Submitted by: Dr. Dinesh Acharya, Mr. Roshan David

(Signature of the faculty)

Date: 31-07-2015

Approved by:

(Signature of HOD)

Date:

****