VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Big Data Analytics ITA6008

Prof. Ramesh Ragala

July 8, 2018

Course Objective

1. Understand the fundamentals of various big data analysis techniques.
2. Analyze the big data analytic techniques for useful business applications.
3. Perform map-reduce analytics using Hadoop and related tools.

Expected Outcomes

On Completion of the course, the students will be able to

1. Analyze Big data, create statistical models, identify insights that can lead to actionable results.
2. Performs map-reduce analytics using Hadoop.
3. Implement software tools such as R and Hadoop for big data analytics.

- Big Data Overview
- Characteristics of Big Data
- State of practice in analytics
- Role of Data Scientists
- Examples of Big Data Analytics
- Data Analytics Lifecycle

- Hadoop Components
- Design Principle of Hadoop
- Analyzing Big data with Hadoop
- Design of HDFS
- Developing a Map reduce Application

- Distributed File System(DFS)
- Map Reduce
- Algorithms using Map Reduce
- Communication cost Model
- Graph Model for Map Reduce Problem

- Setting up a Hadoop Cluster
- Hadoop Configuration
- Security in Hadoop
- Administering Hadoop
- Hadoop Benchmarks
  item Hadoop in the cloud

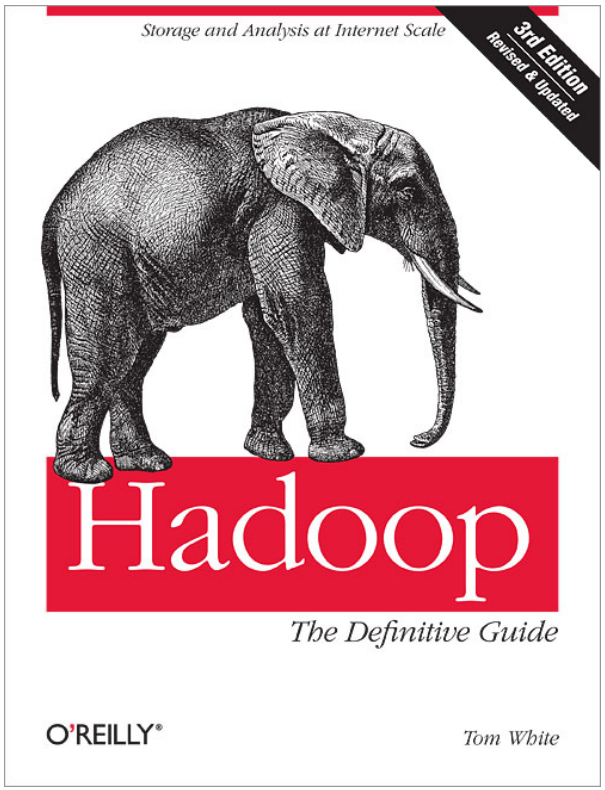
- Introduction to R-Attributes
- R Graphical user interfaces
- Data import and export
- attribute and Data Types
- Descriptive Statistics
- Exploratory Data Analysis

- Hypothesis Testing
- Difference of Means
- Wilcoxon Rank-Sum Test
- Type I and Type II errors
- power and sample size
- ANOVA

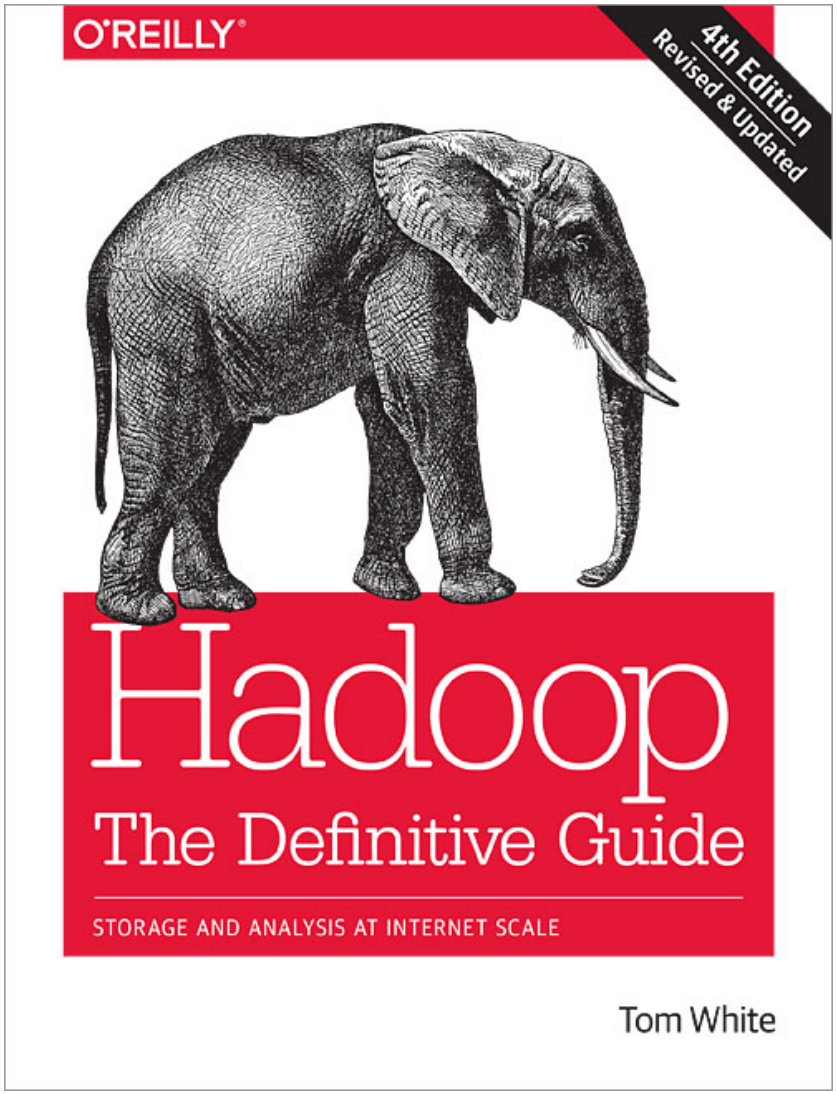
- Analytics for unstructured data
- The Hadoop ecosystem
  - pig
  - Hive
  - Hbase
  - Mahout
  - NoSQL

- **Guest Lecture from Industry experts**

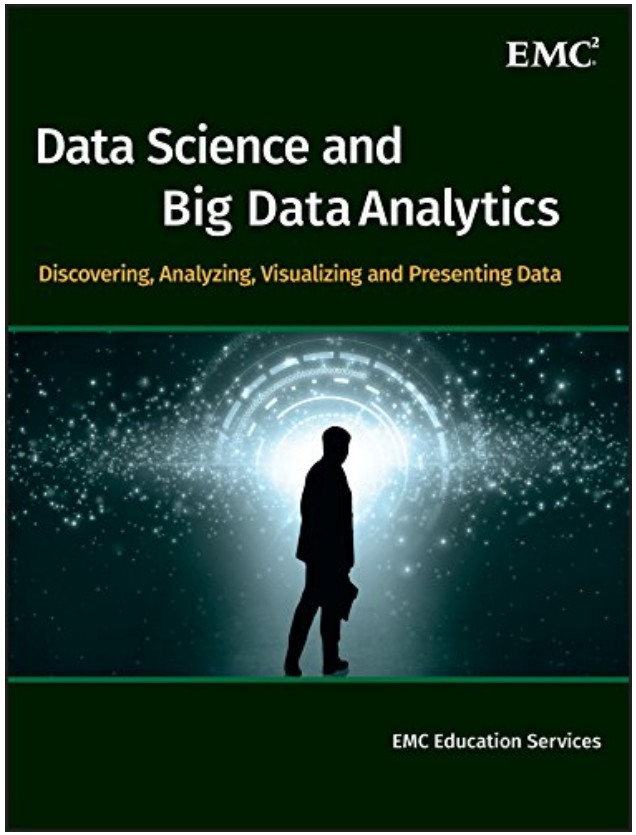
- **Hadoop: The Definitive Guide, 3rd Edition** by Tom White



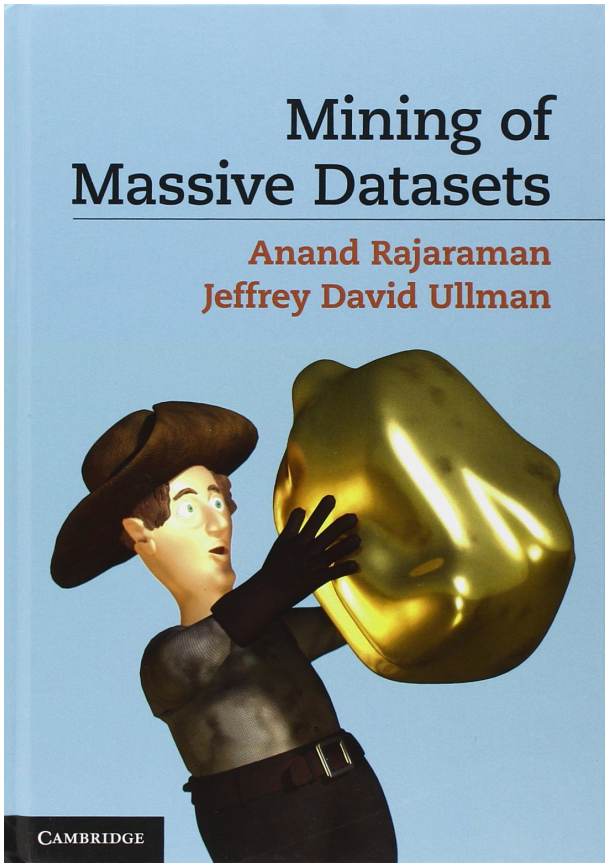
- **Hadoop: The Definitive Guide, 4^{th} Edition** by Tom White



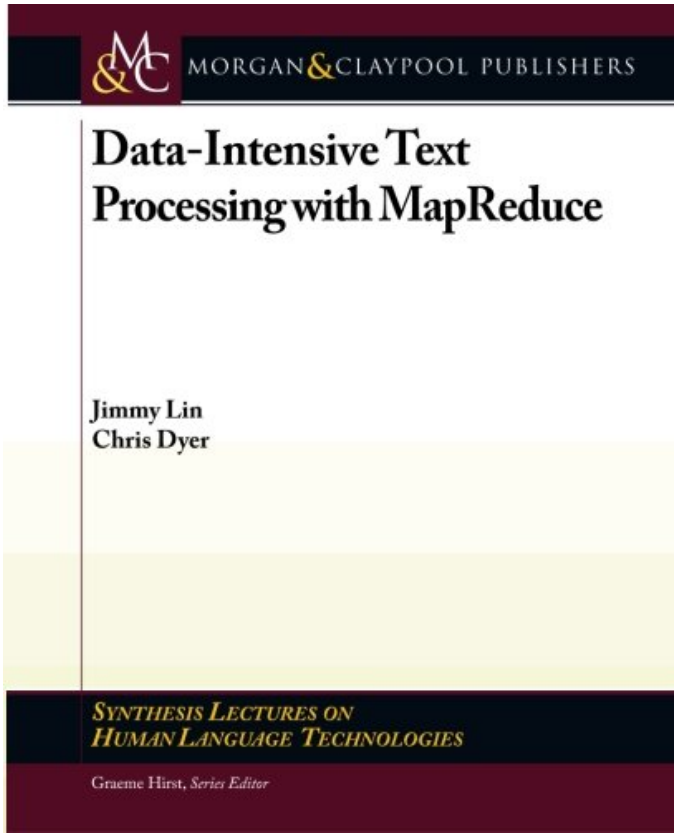
- **Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data 2015** by by EMC Education Services



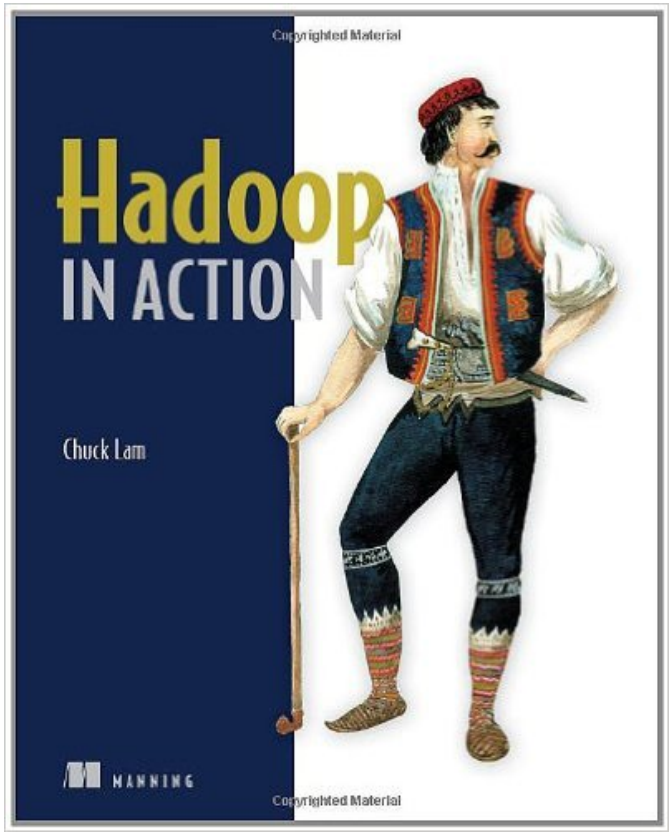
- **Mining of Massive Datasets** by Anand Rajaraman and Jeffrey David Ullman



- **Data-Intensive Text Processing with MapReduce** by Jimmy Lin, Chris Dyer and Graeme Hirst



- **Hadoop in Action** by Chuck Lam



Communication

- **email ID:** ramesh.ragala@vit.ac.in
- **Mobile No:** 9087277270
- **Room No:AB1-604, Cabin No: 8**