



# MANIPAL UNIVERSITY JAIPUR

School of Computing and Information Technology

Department of Computer Science and Engineering  
Course Hand-out

Big data & Analytics | CS1701 | 4 Credits

Session: 2015-2016 | Faculty: Priyank Singh Hada

**Course Outcomes:** At the end of the course, students will be able to

[1701.1] To describe fundamentals of Data Analytics & the need of the Big Data Analytics.

[1701.2] To implement the structured lifecycle approach to data science and big data analytics projects

[1701.3] To solve the industry challenges in analyzing big data applications and create statistical models

[1701.4] To implement various theoretical models of data analytics using the R tool/R studio

[1701.5] To use analytical platforms such as Hadoop, NoSQL DBMS software for performing advanced data analytics and hence enhance the employability skills

## A. SYLLABUS

Introduction to big data: definition, need and evolution of BDA, applications of Big Data, Big Data Analytics: Analysing big data, sources of big data, characteristics of big data (4 V's), Drivers of BDA, types of data, structured vs. unstructured data, data marts, Case study based tutorial, Differences between traditional DWDM and BDA, Limitations of traditional RDBMSs to store and analyse Big Data Data science, definition and concepts, data scientists: key competencies and characteristics of data scientists, More discussions on data science: data wrangling, data munging, data jujitsu, Tutorial based on data science applications, Big Data Analytics Ecosystem, State of the Practice in Analytics: Data Analytics Lifecycle and discussions, Roles for a Successful Analytics Project; Case Study to apply the data analytics lifecycle, Analytical databases and DW appliances; Hadoop distributions – Comparing various BDA tools Analysing and Exploring the Data: Challenges when managing and analysing big data, The role of Data Virtualization in a Big Data environment; Why to visualize data. Statistics for Model Building and Evaluation: Statistics in the Analytic Lifecycle, Hypothesis Testing, Difference of means Advanced Analytics – Theory and Methods Overview: K-means clustering, Association Rules, Linear Regression, Logistic Regression, Naïve Bayesian Classifiers, Tutorial based on advanced analytics, Decision Trees, Time Series Analysis, Text Analytics; Tutorial based on analytics Big Data Platforms and Storage Options: The new multi-platform Analytical Ecosystem; Beyond the Data Warehouse - Analytical databases, Hadoop and NoSQL DBMSs

## B. TEXT BOOKS

- i. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services.
- ii. Michael Minelli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends", John Wiley, 2013

