Subject Code	Subject Name	Credits
26CS607	DATA SCIENCE AND BIG DATA ANALYTICS	4

Course Objectives

1. To understand the basics of Data Science
2. To learn about big data
3. To understand and learn the Hadoop framework

Learning Outcomes

After the completion of the course, the graduate will be able to

1. To learn the advanced concepts of Big Data Technology
2. To learn data visualization techniques and uses.

Unit 1 - Data Science in the big world (12 Hrs.)

Data Science in the big world – Benefits and uses of data science and big data – Facets of data – The data science process – The big data ecosystem and data science – Data Science Process

Unit 2 - Problem Solving Methods (12 Hrs.)

Big Data Overview-State of the practice in Analytics-Key Roles for New Data Big Data Ecosystem-Data Analytics lifecycle overview-Discovery-Data Preparation-Model Planning Model Building-Case study-GINA

Unit 3 - Operating System for Big Data-Basic Concepts (12 Hrs.)

An Operating System for Big Data-Basic Concepts-Hadoop Architecture-Working with Distributed File System and Distributed Computation-Hadoop Streaming-A framework for Map reduce

Unit 4 - Data visualization (12 Hrs.)

Data visualization-Why use Data Visualization-Programming packages-Understand the Data-Data Dictionary-Types of Data-Preparing your Data-Explore the data visually-Design Standards-Data Integrity

Unit 5 - Big Data Analytics Implementation (12 Hrs.)

Big Data Analytics Implementation-Data Acquisition-Big Data Collection Systems-Big Data Storage-Real time Analysis – Clustering-K means-Classification and Regression

References:

1. Arshdeep Bahga,"Big Data Science & Analytics A Hands On Approach, Published by Arshdeep Bahga & Vijay Madisetti
2. David Dietrich, Barry Heller, and Beibei Yang,"Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data," Published by John Wiley & Sons, In
3. Benjamin Bengfort,"Data Analytics with Hadoop,Published by O'Reilly Media, Inc.,
4. Kristen Sosulski," Data Visualization Made Simple, Insights into Becoming Visual Taylor & Francis
5. Cielen, Davy, Arno DB Meysman, and Mohamed Ali. Introducing data science: big data, machine learning, and more, using Python tools. Manning Publications Co.,, 2016.