BTEC-909A-18	Credits	L	T	P	Int	Ext
Introduction to Big Data	3	3	0	0	40	60

Course Outcomes

At the end of this course students will demonstrate the ability to:

1. Understand the Evolution and basics of Big Data.
2. Understand the Architecture of Hadoop with its file system and its Programming.
3. Explain the Advanced analytical theory and methods.
4. Describe the challenges in handling streaming data from the real world.

Unit 1 - Evolution & Introduction of Big data: Best Practices for Big data Analytics, Big data characteristics, Validating – The Promotion of the Value of Big Data, Big Data Use Cases, Characteristics of Big Data Applications, Perception and Quantification of Value, Understanding Big Data Storage,

Unit 2 - A General Overview of High Performance Architecture: HDFS, Map Reduce and YARN – Map Reduce Programming Model. Big Data Overview Analysis of data at Rest- Hadoop analytics: Limitations of existing distributing systems, Hadoop Approach, Hadoop Architecture, Distributed file system: HDFS and GPFS, Internals of Hadoop MR engine, Hadoop cluster components, Hadoop Ecosystem, Evaluation criteria for distributed Map Reduce runtimes, Enterprise-grade Hadoop Deployment, Hadoop Implementation

Unit 3 - Advanced Analytical Theory and Methods: Overview of Clustering – K-means, Use Cases, Overview of the Method, Determining the Number of Clusters, Clustering, Classification, Segmentation, Linear regression, ML Search: Indexing and Indexing Techniques, Create inverted index using JAQL, Data Explorer Bundling Hadoop job: Application, Diagnostics, Reasons to Choose and Cautions, Classification: Decision Trees, Overview of a Decision Tree, The General Algorithm – Decision Tree Algorithms, Evaluating a Decision Tree

Unit 4 - Real time analytics: Introduction to streams computing, Challenges/limitations of conventional Systems, Solving a real time analytics problem using conventional system, Challenges to be solved - scalability, thread pooling, etc., Understanding the challenges in handling streaming data from the real world and how to address those using stream computing, Benefits of stream computing in Big Data world, Realtime Analytics Platform (RTAP), Real Time Sentiment Analysis.

Recommended Books:

1. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, by Chris Eaton, Paul Zikopoulos, Wiley Publication 2015.
2. Big Data Analytics: Turning Big Data into Big Money By Frank J. Ohlhorst, McGraw Hill 2012.
3. Ethics of Big Data: Balancing Risk and Innovation By Kord Davis, 2011.