School of Engineering and Applied Science (SEAS),

Ahmedabad University

Template for course content

Course Number: CSC521

Course Title: Big Data Analytics (BDA)

Credit Structure (Lecture - Tutorial - Credits) / (Lab Hours - Credits): 3-0-3

Category: Core

Prerequisites: Statistics, Service Development, Database Management System, Design and Analysis of Algorithms, Visualization

Instructor: Dr. Amit Ganatra

Abstract Content: Student should be able to appreciate the potential of big data analytics in the current business environment. This course will contain introduction to Big Data Analytics different tools and technologies for performing the operations on big data in efficient manner. The major part of the course will be exposure to the big data analytics technologies i.e. Map-Reduce, Hadoop ecosystem, Hive, YARN, Pig, R and data visualization.

Suggested Text book:

1. "Big Data Black book" by "dreamtech press 2015 edition".

Reference Books / Reading Material

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk deroos et al. , "Understanding Big data ", McGraw Hill, 2012.
3. Tom White, "HADOOP: The definitive Guide" , O Reilly 2012.
4. Vignesh Prajapati, "Big Data Analytics with R and Hadoop", Packet Publishing 2013.
5. Tom Plunkett, Brian Macdonald et al, "Oracle Big Data Handbook", Oracle Press, 2014.
6. http://www.bigdatauniversity.com/
7. Jy Liebowitz, "Big Data and Business analytics",CRC press, 2013.
8. Mining of Massive Datasets 3rd Edition by Jure Leskovec, Anand Rajaraman, Jeff Ullman, Cambridge University Press

Evaluation Scheme

Mid-term examination	30%
Online Course Certification, Presentations and Evaluation	10%
End-semester Theory Examination	50%
Class Room Participation and Attendance	10%

Detailed Contents:

Topic Name	Content	No of Lectures
Introduction to Data Mining	Big Data Fundamentals and Big Data Analytics. Structured Data, unstructured Data and semi Structured Data. Introduction of Big Data and Hadoop – Overview and Evolution of Big-Data Hadoop, Architecture/Framework, HDFS Architecture/Framework, Map reduce, Hadoop Environment Setup, Distributed File System(s)	6
Big Data Analytics and Big Data Analytics Techniques	Big Data and its Importance, Drivers for Big data, Optimization techniques, Dimensionality Reduction techniques, Time series Forecasting, Social Media Mining and Social Network Analysis and its Application, Big Data analysis using Hadoop, Pig, Hive, Mongodb, Spark and Mahout, Data analysis techniques like Discriminant Analysis and Cluster Analysis, Introduction to NOSQL (Neo4j) and MongoDB, Hive Architecture, HBase concepts, PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with with Zookeeper, No SQL databases: Cassandra and HBase (columnar), MongoDB and Elastic Search (document-based), Neo4j (graph based)	12
Hadoop Architecture, Hadoop Storage	HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm,	12

	Algorithms using Map Reduce, Understanding inputs and outputs of MapReduce, Map and Reduce tasks, Job, Task trackers ,Cluster Setup, SSH and Hadoop Configuration, HDFS Administering ,Monitoring and Maintenance Moving Data in and out of Hadoop, Data Serialization	
Big Data and High Dimensional Data Analysis	Introduction to Spark Framework and comparisons between Spark and Hadoop Frameworks. Apache Spark (using Scala, Java, Python), Mining streaming data, Apache Kafka, Spark MLlib, Infrastructure for Big Data, Big Data Management and Frameworks. Big Data Search., Big Data as a Service.	10
Big Data Analytics Applications/Usecases And Visualization of Big Data	Big Data Analytics in E-Governance & Society, Applications in Science, Engineering, Healthcare, Visualization, Business etc. Case Study of Existing Big Data Analytics Systems. Big Data visualization with the tools like D3, Kibana, and Grafana, Scala and Python for Data Visualization,	5