CAP456:INTRODUCTION TO BIG DATA

L:3 T:0 P:0 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: define the need of Big Data Analytics in real world

CO2 :: understand Big Data Concepts and its relevance in the present scenario

CO3 :: use Big Data analytics in an integrated manner to improve analysis skills

CO4 :: analyze the Hadoop-based Big Data framework to effectively store and analyze Big Data and produce analytics

Unit I

The Fundamentals of Big Data : understanding big data concepts and terminology, datasets data analysis, data analytics, descriptive analytic, diagnostic analytics, predictive analytics, prescriptive analytics, business intelligence (BI), key performance indicators (KPI), big data characteristics volume, velocity ,variety veracity value, different types of data :structured data ,unstructured data ,semi-structured data, metadata case study, identifying data characteristics volume velocity variety veracity

Unit II

Business Motivations and Drivers for Big Data Adoption : marketplace, dynamics business architecture, business process, management information and communications technology, data analytics and data science, digitization, affordable technology and commodity hardware, social media hyper-connected communities and devices, internet of everything (IoE) case study example

Unit III

Big Data Adoption Considerations : organization prerequisites, data procurement, privacy, security, provenance limited realtime support, distinct performance challenges, distinct governance requirements, distinct methodology, clouds, big data analytics lifecycle business case evaluation

Unit IV

Big Data Storage Concepts : clusters file systems and distributed file systems nosql sharding replication master-slave peer-to-peer sharding and replication combining sharding and master-slave replication, clusters file systems and distributed file system, nosql sharding, replication, master-slave, peer-to-peer sharding and replication, combining sharding and master-slave replication

Unit V

Introduction to Hadoop : Hadoop and its Ecosystem, Hadoop Distributed File System, Map Reduce Framework and programming Model, Hadoop Yarn, HDFS Design Features, HDFS Components, HDFS User commands, Introduction to Hadoop Tools, APACHE Pig, Sqoop, Flume, Oozie, HBase

Unit VI

Big Data Analysis Techniques : Quantitative Analysis, Qualitative analysis, Data Mining, Statistical Analysis, Machine Learning, Semantic Analysis, Visual Analysis

Text Books:

1. BIG DATA SIMPLIFIED by SOURABH MUKHERJEE SAYAN GOSWAMI AMIT KUMAR DAS, PEARSON

References:

1. BIG DATA, BLACK BOOK by DT EDITORIAL SERVICES, DREAMTECH PRESS