# CAP456:INTRODUCTION TO BIG DATA

**Course Outcomes:** Through this course students should be able to

CO1 :: understand big data concepts

CO2 :: define need of big data analytics in real world

CO3 :: develop interest in the area of Hadoop Cluster Mechanism

CO4 :: apply the big data learning in research

**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 system, nosql sharding, replication, master-slave, peer-to-peer sharding and replication, combining sharding and master-slave replication

**Unit V**

**Big Data Processing Concepts** : parallel data processing, hadoop processing, distributed data processing, workloads processing in batch mode, batch processing with map reduce, map and reduce tasks, map combine partition shuffle and sort reduce, a simple map reduce example

**Unit VI**

**Big Data Planning Considerations** : business case evaluation, data identification, data acquisition and filtering, data extraction, data validation and cleansing, data aggregation, data analysis, data visualization, utilization of analysis results case study example

**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