

Introduction

2CS702 Big Data Analytics



Jai Prakash Verma

NF4-2

Jaiprakash.verma@nirmauni.ac.in

9427621081

Course Outcomes

After successful completion of this course, student will be able to

1. outline the significance and challenges of big data
2. model big data using different tools and frameworks
3. apply big data techniques for useful business analytic applications
4. design algorithms for mining the data from large volumes

Syllabus

Unit I

Introduction to Big Data: Evolution of Big Data, Types of Digital Data, Classification of Digital Data, Structured Data, Semi-Structured Data, Unstructured Data, Definition of Big Data, Challenges of Conventional Systems, Big data platforms and data storage

Unit II

Big Data Analytics: Importance of Big data analytics, Classification of Analytics, Top Challenges Facing Big Data, Technologies to meet the Challenges Posed by Big Data, Terminologies Used in Big Data Environment

Unit III

Hadoop: Introducing Hadoop, comparisons of RDBMS and Hadoop, Distributed Computing Challenges, Hadoop Overview, Business Value of Hadoop, Hadoop Distributed File System, Processing Data with Hadoop, working with Map Reduce, Hadoop YARN, Hadoop in the Cloud, Applications on Big Hadoop Ecosystem, Fundamentals of Pig, Hive, HBase and ZooKeeper, Basic concepts of Apache Spark

Unit IV

The Big data technology landscape: CAP Theorem - BASE Concept, NoSQL, Types of No SQL databases, Introduction to MongoDB, Data Types in MongoDB, CRUD, Apache Cassandra, Features of Cassandra, CRUD

Unit V

Big data analytics Algorithm: Applying Linear Regression, Clustering, Association rule mining, Decision tree on Big Data.

Self-study: Frameworks: Applications on Big Data Using Pig and Hive

References:

1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
2. **Tom White, Hadoop: The Definitive Guide, Third Edition, O'reilly Media**
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Publishing
4. **Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press**
5. Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley & sons
6. **Glenn J. Myatt, Making Sense of Data, John Wiley & Sons**
7. **Pete Warden, Big Data Glossary, O'Reilly**
8. Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, Second Edition, Elsevier
9. Da Ruan, Guoqing Chen, Etienne E.Kerre, GeertWets, Intelligent Data Mining, Springer
10. Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, Harness the Power of Big Data The IBM Big Data Platform, Tata McGraw Hill Publications
11. Michael Minelli, Michele Chambers, Ambiga Dhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley Publications
12. Zikopoulos, Paul, Chris Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Tata McGraw Hill Publications
13. **Seema Acharya and Subhashini C, Big Data and Analytics, Wiley India**

Lab session guidelines

InClassQuestion#1

- What is the need to learn this subject?

Big Data popular case study

1)

Big Data Case Study – Walmart



Big Data popular case study

2)

Big Data Case Study – Uber



Big Data popular case study

3)

Big Data Case Study – Netflix



Big Data popular case study

4)

. Big Data Case Study – eBay



Big Data popular case study

5)

Big Data Case Study – Procter & Gamble



InClassQuestion#2

- How can we apply Big data Analytics in your research?