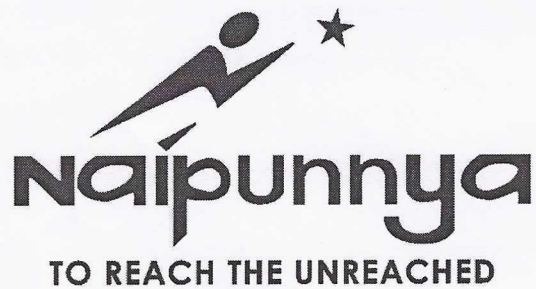


Naipunnnya School of Management Cherthala



Department of Computer Science and Applications

ADD-CSA-05: Certificate in Big Data


Head of the Department
Dept of Computer Science and Applications
NAIPUNNYA SCHOOL OF MANAGEMENT
CHERTHALA




FR. BAIJU GEORGE
PRINCIPAL
NAIPUNNYA SCHOOL OF MANAGEMENT

ADD-CSA-05: Certificate in Big Data (30 Hours)

Aim

- The goal of this course is to develop and apply innovative, state-of-the-art practices and technologies and provide sustainable solutions to the Computer Science and Engineering Problems.

Course Objective

- To understand the need of Big Data, challenges and different analytical architectures
- Installation and understanding of Hadoop Architecture and its ecosystems
- Processing of Big Data with Advanced architectures like Spark.
- Describe graphs and streaming data in Spark

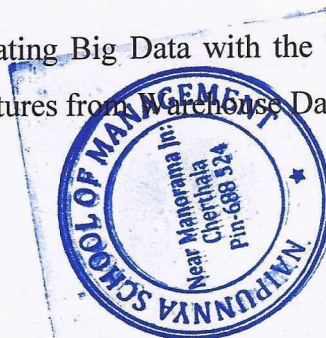
Syllabus

Module 1: Introduction to Big Data – Data, Big Data – Examples, Types – Structured, Unstructured, semi structured. Characteristics of Big Data – Volume, Variety, Velocity, Variability and Value. Advantages of Big Data processing, Benefits of Big Data Processing. Real Time Application of Big Data, Distributed Computing – basics, Cloud Computing – basics.

Module 2: Hadoop – Introduction, Hadoop Ecosystem & Components, Hadoop MapReduce, Hadoop, Distributed File System (HDFS), Hadoop Architecture – NameNode, DataNode, MasterNode, SlaveNode. Features of Hadoop – suitable for Big data Analysis, Scalability, Fault Tolerance, Network Topology in Hadoop.

Module 3: HDFS – Introduction, HDFS Architecture, Read Operation in HDFS, Write Operation in HDFS. MapReduce in Hadoop – Phases of MapReduce – Map Phase, Reduce Phase, working of MapReduce – Input splits – mapping, shuffling, reducing – MapReduce Architecture – Work Organization using MapReduce.

Module 4: Big Data Warehouse – Integrating Big Data with the Traditional Warehouse – Optimizing, Differentiating Big Data Structures from Warehouse Data. Data Analytics.




Course Outcome

- Discuss the challenges and their solutions in Big Data
- Understand and work on Hadoop Framework and eco systems.
- Explain and Analyse the Big Data using Map-reduce programming in Both Hadoop and Spark framework.
- Demonstrate spark programming with different programming languages.
- Demonstrate the graph algorithms and live streaming data in Spark

References

1. Mike Frampton, "Mastering Apache Spark", Packt Publishing, 2015.
2. Tom White, "Hadoop: The Definitive Guide", O'Reilly, 4th Edition, 2015.
3. Nick Pentreath, "Machine Learning with Spark", Packt Publishing, 2015.
4. Mohammed Guller, "Big Data Analytics with Spark", Apress, 2015
5. Donald Miner, Adam Shook, "Map Reduce Design Pattern", O'Reilly, 2012


Head of the Department
Dept of Computer Science and Applications
NAIPUNNYA SCHOOL OF MANAGEMENT
CHERTHALA


Principal
Naipunnya School of Management
Cherthala

