POSITIVE QUADRANT
T E C H N O L O G I E S
SERVING INFORMATION WORLDWIDE

Website: www.positivequadrant.in
Email: positivequadrants@gmail.com
Contact: +91 8169150592

COMPUTER ENGINEERING

SEM VII

BIG DATA ANALYTICS

Programming & development

Course Curriculum

POSITIVE QUADRANT
TECHNOLOGIES
SERVING INFORMATION WORLDWIDE

Website: www.positivequadrant.in
Email: positivequadrants@gmail.com
Contact: +91 8169150592

BIG DATA ANALYTICS SEM VII

Module 1 : Introduction to Big Data and Hadoop

- Introduction to Big Data
  - Big Data characteristics
  - Types of Big Data
- Traditional vs. Big Data business approach
- Case Study of Big Data Solutions
- Concept of Hadoop
- Core Hadoop Components
- Hadoop Ecosystem

Module 2: Hadoop HDFS and MapReduce

- Distributed File Systems
  - Physical Organization of Compute Nodes
  - Large-Scale File-System Organization
- MapReduce
  - The Map Tasks
  - Grouping by Key
  - The Reduce Tasks
  - Combiners
  - Details of MapReduce Execution
  - Coping With Node Failures
- Algorithms Using MapReduce
  - Matrix-Vector Multiplication by MapReduce
  - Relational-Algebra Operations
  - Computing Selections by MapReduce
  - Computing Projections by MapReduce
  - Union
  - Intersection
  - Difference by MapReduce
- Hadoop Limitations

Module 3: NoSQL

POSITIVE QUADRANT
TECHNOLOGIES
SERVING INFORMATION WORLDWIDE

Website: www.positivequadrant.in
Email: positivequadrants@gmail.com
Contact: +91 8169150592

- Introduction to NoSQL
- NoSQL Business Drivers
- NoSQL Data Architecture Patterns
  - Key-value stores
  - Graph stores
  - Column family (Bigtable)stores
  - Document stores
  - Variations of NoSQL architectural patterns
  - NoSQL Case Study
- NoSQL solution for big data
- Understanding the types of big data problems
- Analyzing big data with a shared-nothing architecture
- Choosing distribution models
  - master-slave versus peer-to-peer
  - NoSQL systems to handle big data problems

Module 4: Mining Data Streams

- The Stream Data Model
  - A Data-Stream-Management System
  - Examples of Stream Sources
  - Stream Queries
  - Issues in Stream Processing
- Sampling Data techniques in a Stream
- Filtering Streams
  - Bloom Filter with Analysis
- Counting Distinct Elements in a Stream
- Count-Distinct Problem
- Flajolet-Martin Algorithm
- Combining Estimates
- Space Requirements
- Counting Ones in a Window
  - The Cost of Exact Counts
  - The Datar-Gionis-Indyk-Motwani Algorithm
  - Query Answering in theDGIM Algorithm
  - Decaying Windows

Module 5: Real-Time Big Data Models

POSITIVE QUADRANT
T E C H N O L O G I E S
SERVING INFORMATION WORLDWIDE

Website: www.positivequadrant.in
Email: positivequadrants@gmail.com
Contact: +91 8169150592

- A Model for Recommendation Systems
- Content-Based Recommendations
- Collaborative Filtering
- Case Study
    - Product Recommendation
- Social Networks as Graphs
- Clustering of Social-Network Graphs
- Direct Discovery of Communities in a social graph

Module 6: Data Analytics with R

- Exploring Basic features of R
- Exploring RGUI
- Exploring RStudio
- Handling Basic Expressions in R
- Variables in R
- Working with Vectors
- Storing and Calculating Values in R
- Creating and using Objects
- Interacting with users
- Handling data in R workspace
- Executing Scripts
- Creating Plots
- Accessing help and documentation in R
- Reading datasets and Exporting data from R
- Manipulating and Processing Data in R
- Using functions instead of script
- built-in functions in R
- Data Visualization
    - Types
    - Applications