

# Big Data Foundation

## Module 1. Big Data Fundamentals

- 1.1 Big Data – History, Overview, and Characteristics
  - History
  - Big Data Definition
  - Big Data Benefits
  - Big Data Characteristics – Volume, Velocity & Variety
- 1.2 Big Data Technologies – Overview
- 1.3 Big Data Success Stories
- 1.4 Big Data – Privacy and Ethics
  - Privacy – Compliance
  - Privacy – Challenges
  - Privacy – Approach
  - Ethics
- 1.5 Big Data Projects
  - Who Should Be Involved?
  - What Is Involved?

## Module 2. Big Data Sources

- 2.1 Enterprise Data Sources
  - Enterprise Systems
  - Oracle
  - SAP
  - Microsoft
  - Data Warehouses
  - Unstructured Data – Introduction
  - Unstructured Data – Metadata
- 2.2 Social Media Data Sources
  - Introduction
  - Facebook – Introduction
  - Facebook – Public Feed API
  - Facebook – Keyword Insights API
  - Facebook – Graph API
  - Twitter – Introduction
  - Twitter – Streaming APIs
  - Twitter – REST APIs
  - Other Social Media
- 2.3 Public Data Sources

- Introduction
- Weather
- Economics
- Finance
- Regulatory Bodies

## Module 3. Data Mining – Concepts and Tools

### 3.1 Data Mining – Introduction

- Introduction
- Types of Data Mining – Overview
- Types of Data Mining – Classification
- Types of Data Mining – Association
- Types of Data Mining – Clustering

### 3.2 Data Mining – Tools

- Introduction
- Weka
- Modules of Weka Applications
- KNIME
- KNIME – Example
- R Language

## Module 4. Big Data Technologies – Hadoop

### 4.1 Hadoop Fundamentals

- Introduction
- Main Components of Hadoop
- Additional Components of Hadoop

### 4.2 Install and Configure

- Download
- How to Install and Configure

### 4.3 MapReduce

- Introduction
- How Does It Work?

### 4.4 Data Processing with Hadoop

- Introduction
- Twitter Sentiment Analysis – Overview
- Twitter Sentiment Analysis – Algorithm
- Network Log Analysis – Overview
- Network Log Analysis – Algorithm

## Module 5. Big Data Technologies – MongoDB

### 5.1 MongoDB Fundamentals

- Introduction

- Replication
  - Sharding
  - Sharding and Replication
  - MongoDB Ecosystem – Languages and Drivers
  - MongoDB Ecosystem – Hadoop Integration
  - MongoDB Ecosystem – Tools
- 5.2 Install and Configure
- Download
  - How to Install and Configure
- 5.3 Document Databases
- Introduction
  - Documents
  - Document Design Considerations
  - Fields
- 5.4 Data Modelling with Document Databases
- Introduction
  - Twitter Sentiment Analysis
  - Twitter Sentiment Analysis – Algorithm
  - Network Log Analysis
  - Network Log Analysis – Algorithm