Big Data Analytics With Hadoop

Course Summary

Description

Apache Hadoop is a popular framework for processing Big Data. Hadoop provides rich and deep analytics capability, and it is making in-roads in to traditional BI analytics world. This course will introduce an analyst to the core components of Hadoop eco system and its analytics.

Objective

At the end of this course, students will

- Understand Hadoop ecosystem
- Data storage using HDFS
- Data warehousing and querying using Hive

Topics

- Hadoop overview
- Hive

Audience

This course is designed for Business Analysts and Developers.

Prerequisite

- Programming background with databases / SQL
- Basic knowledge of Linux

Duration

Two Days
Big Data Analytics With Hadoop

Course Outline

I. Section 1: Hadoop eco system
   A. Hadoop overview
      1. distributions
      2. high level architecture
      3. hardware / software
      - Labs: first look at Hadoop
   B. HDFS Overview
      1. concepts (horizontal scaling, replication, data locality)
      2. architecture (Namenode, Data node)
      3. Demo: Interacting with HDFS
   C. YARN Overview
      1. YARN operating system
      2. Demo: Running applications on YARN program

II. Section 2: Hive
   A. hive concepts & architecture
   B. SQL support in Hive
   C. Data warehousing in Hive
   D. data types
   E. table creation and queries
   F. partitions
   G. joins
   H. modern data formats
   I. text analytics
   J. Hive performance
      - Labs (multiple)