

Big Data Analytics & Hadoop Workshop

Introduction

The workshop is for professionals desiring to an understanding of Big Data Analytics domain.

The program takes an individual through all aspects of this emerging technology and will benefit participants from all streams wanting to get a career.

Duration: 40 hours

No	Title	Topics Covered	Duration
Day 1 – Introduction to Big Data and Hadoop			
01	Introduction to Big Data	<ul style="list-style-type: none"> • What is Big Data? • Data Evolution Through Past 4 Decades • Nature of Big Data – Volume, Variety & Velocity • Factors Driving Use of Big Data • Challenges faced by Conventional Data Analytics system • Need for Big Data Technologies • Overview of Big Data Technologies – Hadoop 	1.5 Hours
02	Applications of Big Data	<ul style="list-style-type: none"> • Real life business problems, which can't be answered without data • Case Study – Increasing Customer Life-time Value in Web Retail Business • Case Study – Increasing Machine utilization and asset life in a Manufacturing Organization • Case Study – Real-time Credit Card Fraud Detection 	1.5 Hours
03	Hadoop Architecture	<ul style="list-style-type: none"> • Principles of Distributed Processing • Overview of Hadoop Architecture • Hadoop Core Components – HDFS & Map/Reduce 	1.5 Hours
04	HDFS	<ul style="list-style-type: none"> • HDFS Overview • Distributed File Storage Concepts • Importance of Block Size • Data Fault-tolerance using Replication • Understanding Name node Metadata • Cluster Fault-tolerance using Secondary Name node 	2.5 Hours
Day 2– Hadoop Setup and Big Data Processing with Map / Reduce			
05	Setting Up a Hadoop Cluster	<ul style="list-style-type: none"> • Downloading, installing & Configuring Hadoop • Setting up a Pseudo-Distributed Cluster • Setting Up a Fully-distributed Cluster • Working with Hadoop – DFS shell commands • Hadoop User Interface • Understanding Hadoop Log Files • Changing HDFS block sizes and Replication Factors • Adding and Removing Nodes from Hadoop Cluster 	2 Hours
06	Introduction to Map/Reduce	<ul style="list-style-type: none"> • What is Map/Reduce • Map/Reduce examples from real world • Basic principles of Map/Reduce • Understanding various Map/Reduce Phases • Running our first Map/Reduce program using Wordcount example 	1 Hours
07	Working with Map/Reduce	<ul style="list-style-type: none"> • Map/Reduce Framework • Roles of Developer and Framework 	1.5 Hours

No	Title	Topics Covered	Duration
		<ul style="list-style-type: none"> • Hadoop Data Types – Writable & Comparable Interfaces • Hadoop Input & Output Formats • Writing a Map Class • Writing a Reduce Class • Creating a Job Configuration Object • Building a Jar and running a Map/Reduce program 	
08	Map/Reduce in Detail	<ul style="list-style-type: none"> • How Map/Reduce works • Roles of Developer and Framework • Hadoop Data Types – Writable & Comparable Interfaces • Hadoop Input & Output Formats • Writing a Map Class • Writing a Reduce Class • Creating a Job Configuration Object using inner classes • Building a Jar and running a Map/Reduce program 	2.5 Hours
Day 3 – Big Data Processing Simplified with PIG & HIVE			
09	Introduction to Pig	<ul style="list-style-type: none"> • What is Pig? Why Pig • Pig Architecture & Modes of Operations • Pig Latin Script • Pig Data Types • Pig Operators • Word Count Example & Other Hands-on Exercises • User Defined Functions and Data Types • PIG Data Joins 	3 Hours
10	Introduction to Hive	<ul style="list-style-type: none"> • Hive Introduction • Differences between Hive and traditional RDBMS • Hive Architecture – components of Hive • Hive Data Model – database, partitions, tables, fields, indexes • Hive Schema and Meta store • Input / Output Formats – SerDe's • HQL - Hive Query Language • HQL – Hands On Exercises • Project - Building a Tweeter Sentiment Analysis Application 	4 Hours
Day 4 – Real-Time Analytics with Hadoop			
11	Introduction to No-SQL Databases	<ul style="list-style-type: none"> • Why HBase? • HBase Architecture • Role of Zookeeper • Overview of Cassandra • Cassandra Exercise 	3 Hours
12	Data Import. Export with Flume and Sqoop	<ul style="list-style-type: none"> • Bringing unstructured data into Hadoop using Flume • Flume Exercise • Bringing Relational Data from SQL into Hadoop using Sqoop • Sqoop Exercises 	2 Hours
13	Real-Time Analytics with Spark	<ul style="list-style-type: none"> • Why Spark • Spark Architecture • Spark Application Development Environment 	2 Hours