



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

Practical Big Data Analytics





جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

Introduction to Big Data

- *What is Big Data*
- *Big Data opportunities, Challenges*
- *Characteristics of Big Data*

Introduction to Hadoop

- *Hadoop Distributed File System*
- *Comparing Hadoop & SQL*
- *Data Locality*
- *Hadoop Architecture*
- *Map Reduce & HDFS*



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

Hadoop Distributed File System (HDFS)

- ***HDFS Design & Concepts***
- ***Blocks, Name nodes and Data nodes***
- ***HDFS High-Availability and HDFS Federation***
- ***Hadoop DFS The Command-Line Interface***
- ***Basic File System Operations***
- ***Anatomy of File Read, File Write***
- ***Metadata, FS image, Edit log, Secondary Name Node and Safe Mode***
- ***(How to add New Data Node dynamically, decommission a Data Node dynamically (Without stopping cluster***
- ***How to override default configuration at system level and Programming level***



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

Map Reduce

- ***Map Reduce Functional Programming Basics***
- ***Map and Reduce Basics***
- ***How Map Reduce Works***
- ***Anatomy of a Map Reduce Job Run***
- ***Shuffling and Sorting***
- ***Splits, Record reader, Partition, Types of partitions & Combiner***
- ***Distributed Cache and Hadoop Streaming (Python, Ruby and R)***
- ***Apache YARN***
- ***Sequential Files and Map Files***
- ***Map side Join with distributed Cache***



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

Map Reduce Programming – Java Programming

- *Hands on “Word Count” in Map Reduce in standalone and **Pseudo Distribution Mode***
- *Write some Map Reduce programs to solve some real world problems*

Apache HIVE Installation

- *Installing Hive*
- *Configuring Hive*
- *Configuring Metastore of Hive*
- *Verifying Hive Installation*
- *Database Operations*
- *Create Database Statement*
- *Drop Database Statement*
- *Table Operations*



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

- **Create Table Statement**
- **Load Data Statement**
- **Alter Table Statement**
- **Change Statement**
- **Add Columns Statement**
- **Replace Statement**
- **Drop Table Statement**

Apache Spark Basics

- **What is Apache Spark?**
- **Spark Installation**
- **Spark Configuration**
- **Spark Context**
- **Using Spark Shell**
- **Resilient Distributed Datasets (RDDs) – Features, Partitions, Tuning Parallelism**
- **Functional Programming with Spark**



جهاد دانشگاهی صنعتی شریف
مرکز آموزش های تخصصی کاربردی

Working with RDDs

- ***RDD Operations – Transformations and Actions***
- ***Types of RDDs***
- ***Key-Value Pair RDDs – Transformations and Actions***
- ***MapReduce and Pair RDD Operations***
- ***Serialization***

Spark on a cluster

- ***Overview***
- ***A Spark Standalone Cluster***
- ***The Spark Standalone Web UI***
- ***Executors & Cluster Manager***
- ***Spark on YARN Framework***