

Sela.

BDHS

Big Data with Hadoop and Spark

college@sela.co.il

03-6176666





Big Data with Hadoop and Spark

BDHS - Version: 3

 3 days Course

Description:

In the first training day, the course will introduce developers to the Hadoop ecosystem focusing on HDFS, Hive and HBase.

In the next two training days, the course will cover Spark including: Spark Streaming, Graphs and Machine Learning.

Intended audience:

Developers who wish to obtain the skills and knowledge to develop big data solutions with Hadoop and Spark.

Prerequisites:

Basic knowledge of programming in Python

Objectives:

- Distributed programming with Spark
- Using Machine learning algorithms
- Storing and analyzing data using HBase and Hive

Topics:

Hadoop (day 1)



- Hadoop Introduction
 - Big Data Analytics
 - What is and Why Hadoop
 - Comparing Hadoop with Other Technologies
 - Hadoop Architecture
 - Hadoop Ecosystem
 - Hadoop Usage Examples
- Hadoop HDFS
 - What Is and Why HDFS
 - HDFS Architecture
 - HDFS Features
 - HDFS Commands
 - HDFS Web UI
 - Hue Web UI
- Hadoop HBase
 - What is and Why HBase?
 - HBase Shell
 - HBase Architecture
 - HBase Data Model
 - Storing Data Into HBase table using Pig and Hive
 - Accessing data in the HBase table using Pig and Hive
 - Access HBase table via REST/Thrift
- Hadoop Hive
 - What Is and Why Hive?
 - Hive Architecture
 - HiveQL
 - Physical Layout
 - Loading Data into Hive Tables
 - Partitions
 - Joining
 - Buckets



Spark Workshop (day 2 & 3)

- Why Spark?
- Basic Concepts
- Running on Clusters
- Spark SQL
- Spark Structured Streaming
- Spark ML
- Spark Performance Topics