www.sela.co.il   03-6176066

Microsoft Partner
Gold Learning
Gold Application Development
Gold Application Lifecycle Management
Gold Mobility
Cloud Accelerate

Big Data with Hadoop and Spark

BDHS - Version: 3

🕐 3 days Course

Description:

This course will introduce developers to the Hadoop ecosystem, focus on multiple programming models including MapReduce, Pig, Hive, and Apache Spark.

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 Java

Objectives:

Developing MapReduce programs in Java
Analyzing Data using Pig and Hive
Distributed programming with Spark
Using Machine learning algorithms

Topics:

Hadoop Introduction

- Big Data Analytics
- What is and Why Hadoop
- Comparing Hadoop with Other Technologies
- Hadoop Architecture
- Hadoop Ecosystem
- Hadoop Usage Examples

www.sela.co.il   03-6176066

Microsoft Partner
Gold Learning
Gold Application Development
Gold Application Lifecycle Management
Gold Mobility
Cloud Accelerate

Hadoop Installation

- Hadoop Operational Modes
- Hadoop Installation Options

Hadoop HDFS

- What Is and Why HDFS
- HDFS Architecture
- HDFS Features
- HDFS Commands
- HDFS Web UI
- Hue Web UI

Hadoop MapReduce

- Hadoop MapReduce Operational Architecture
- MapReduce General Concepts
- Hadoop MapReduce Programming
- Hadoop MapReduce Framework
- Hadoop MapReduce Example: Word Count
- Executing Hadoop MapReduce Application
- Input File and Input Format

Hadoop Pig

- What Is and Why Pig?
- Executing Pig
- Pig Latin Concepts
- Pig Latin Script: Operators and Functions
- Pig Latin Script: Structure
- Pig Latin Relational Operators
- Pig Examples

Hadoop Hive

- What Is and Why Hive?
- Hive Architecture
- HiveQL
- Physical Layout
- Loading Data into Hive Tables
- Partitions
- Joining
- Buckets

Spark Workshop

- Why Spark?
- Basic Concepts
- Running on Clusters
- Spark SQL
- Spark Streaming