Spark Program

INTRODUCTION

The following course materials are copyright protected materials. They may not be reproduced or distributed and may only be used by students attending the *Spark Program* course.

Welcome!

➤ ROI leads the industry in designing and delivering customized technology and management training solutions

➤ Meet your instructor
  – Name
  – Background
  – Contact info

➤ Class schedule

➤ Let's get started!

ROITRAINING
MAXIMIZE YOUR TRAINING INVESTMENT

Course Objectives

In this course, we will:

- Explore how HDFS and NoSQL clusters store data

- Compare processing distributed data with MapReduce vs. Spark

- Learn how to use Python to create Spark programs that use RDDs and DataFrames to parallel process data

- Introduce the differences between batch vs. streaming engines in Spark

- Transform data into a shape necessary to perform Machine Learning algorithms on it

- Introduce the core Machine Learning models of Cluster, Regression, and Classification

- Process natural language into a form suitable for Machine Learning and analysis

- Take a survey of the Big Data ecosystem and its various common tools

ROITRAINING
MAXIMIZE YOUR TRAINING INVESTMENT

Course Contents

Chapter 0   Introduction

Chapter 1   Introduction to HDFS & Spark

Chapter 2   DataFrames

Chapter 3   SparkSQL

Chapter 4   Spark and NoSQL

Chapter 5   Spark Streaming

Chapter 6   Cluster Analysis

Chapter 7   Classification Analysis

Chapter 8   Natural Language Processing

Chapter 9   Big Data Ecosystem

Chapter 10   Course Summary

ROI's Training Curricula

Agile Development	.NET and Visual Studio
Amazon Web Services (AWS)	Networking and IPv6
Azure	Oracle and SQL Server Databases
Big Data and Data Analytics	OpenStack and Docker
Business Analysis	Project Management
Cloud Computing and Virtualization	Python and Perl Programming
Excel and VBA	Security
Google Cloud Platform (GCP)	SharePoint
ITIL® and IT Service Management	Software Analysis and Design
Java	Software Engineering
Leadership and Management Skills	UNIX and Linux
Machine Learning and Neural Networks	Web and Mobile Apps
Microsoft Exchange	Windows and Windows Server

Please visit our website at www.ROItraining.com for a complete list of offerings

ROITRAINING
MAXIMIZE YOUR TRAINING INVESTMENT

Student Introductions

Please introduce yourself stating:

→ Name

→ Position or role

→ How many years of experience you have with Spark

→ Expectations or a question you'd like answered during this class

ROI TRAINING
MAXIMIZE YOUR TRAINING INVESTMENT