ROI TRAINING
MAXIMIZE YOUR TRAINING INVESTMENT

Spark Program

COURSE SUMMARY

Spark Program

ROITRAINING
MAXIMIZE YOUR TRAINING INVESTMENT

11-1

Course Summary

In this course, we have:

➔ Explored how HDFS and NoSQL clusters store data

➔ Compared processing distributed data with MapReduce vs. Spark

➔ Learned how to use Python to create Spark programs that use RDDs and DataFrames to parallel process data

➔ Introduced the differences between batch vs. streaming engines in Spark

➔ Transformed data into a shape necessary to perform Machine Learning algorithms on it

➔ Introduced the core Machine Learning models of Cluster, Regression, and Classification

➔ Processed natural language into a form suitable for Machine Learning and analysis

➔ Taken a survey of the Big Data ecosystem and its various common tools

ROITRAINING
MAXIMIZE YOUR TRAINING INVESTMENT

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®

Ways to Stay in Touch

ROITraining.com

Follow us!
@ROITraining

Connect with us!
/company/roi-training

Like us!
/ROITraining

Follow us!
/roitraining