

# SUMMIT TRAINING OFFERING

## **HORTONWORKS DATA PLATFORM (HDP®)**

### **DEVELOPER: SPARK**

2 DAYS    
INTERMEDIATE DEVELOPER

This **2 day** course is designed for developers who need to create applications to analyze Big Data stored in Apache Hadoop using Spark. The focus will be on utilizing the Spark API from Python or Scala.

## **PREREQUISITES**

No previous Hadoop knowledge is required, though will be useful. Basic knowledge of Python or Scala is required. Previous exposure to SQL is helpful, but not required. Students will need to bring their wi-fi enabled laptop pre-loaded with Chrome or Firefox browser in order to complete hands-on labs.

## **TARGET AUDIENCE**

Developers, Architects, and Admins who would like to learn more about developing data applications in Spark, how it will affect their environment, and ways to optimize application.

## **FORMAT**

50% Lecture/Discussion

50% Demos/Hands-On Labs

## COURSE OBJECTIVES

- Describe Spark and Spark specific use cases
- Explain the differences between Spark and MapReduce
- Explore data interactively through the spark shell utility
- Explain the RDD concept
- Use the Python/Scala Spark APIs
- Create all types of RDDs: Pair, Double, and Generic
- Use RDD type-specific functions
- Explain interaction of components of a Spark Application
- Explain the creation of the DAG schedule
- Build and package Spark applications
- Use application configuration items
- Deploy applications to the cluster using YARN
- Use data caching to increase performance of applications
- Implement advanced features of spark
- Learn general application optimization guidelines/tips
- Create/transform data using dataframes
- Read, use, and save to different Hadoop file formats
- Understand the concepts of Spark Streaming
- Create a streaming application
- Use Spark MLlib to gain insights from data

### **About Hortonworks**

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

### Contact

For further information visit [www.hortonworks.com](http://www.hortonworks.com) +1 408 675-0983  
+1 855 8-HORTON  
INTL: +44 (0) 20 3826 1405

## LABS

- Create a Spark “Hello World” word count application
- Use advanced RDD programming to perform sort, join, pattern matching and regex tasks
- Explore partitioning and the Spark UI
- Increase performance using data caching
- Build/package a Spark application using Maven
- Use a broadcast variable to efficiently join a small dataset to a massive dataset
- Use an accumulator for reporting data quality issues
- Create a dataframe and perform analysis
- Load/transform/store data using Spark with Hive tables
- Create a point-in-time spark stream application
- Create a spark stream application using window functions

### **About Hortonworks**

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world’s data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

### Contact

For further information visit [www.hortonworks.com](http://www.hortonworks.com) +1 408 675-0983  
+1 855 8-HORTON  
INTL: +44 (0) 20 3826 1405