Hortonworks University | Hortonworks CERTIFIED DEVELOPER

Data Sheet

Developing Apache Hadoop Applications in Java

This 4 days hands-on training course takes a deep-dive into developing Java MapReduce applications for Big Data deployed on the Hadoop Distributed File System (HDFS). Students who attend this course will learn how to harness the power of Apache Hadoop™ and MapReduce to manipulate, analyze and perform computations on their Big Data.

Prerequisites

This course assumes students have experience developing Java applications and using a Java IDE. Labs are completed using the Eclipse IDE and Maven.

Target Audience:

Experienced Java developers responsible for developing MapReduce applications and performing analysis of Big Data stored on Apache Hadoop.

Course Objectives

At the completion of the course, student will be enabled to perform the following:

- Write a Java MapReduce application using Eclipse and Maven
- Develop a Combiner to perform map aggregation
- Customize input and output formats of a MapReduce job
- Compute mathematical computations on your Big Data files
- Use best practices to optimize MapReduce jobs
- Create JUnit tests for a MapReduce job
- Discover trends in your Big Data
- Define an Oozie workflow
- Access Apache HBase™ data from a Java MapReduce job
- Write custom, user-defined functions for Apache Pig™ and Apache Hive™

Lab Content

Students will work through the following exercises using Eclipse, Maven and the Hortonworks Data Platform:

- Configuring a Hadoop™ Development Environment
- Word Count
- Distributed Grep
- Inverted Index
- Using a Combiner
- Computing an Average
- Writing a Custom Partitioner
- Using a TotalOrderPartitioner
- Custom Sorting
- Writing a Custom InputFormat
- Customizing Output
- Simple Moving Average
- Using Data Compression
- Defining a RawComparator
- A Map-Side Join
- Using a Bloom Filter
- Unit Testing
- Defining an Oozie Workflow
- Term Frequency–Inverse Document Frequency (TF-IDF)
- Accessing HBase from Java MapReduce
- Writing a User-Defined Pig Function
- Writing a User-Defined Hive Function

Agenda

Day 1

- Understanding Hadoop and MapReduce
- Writing MapReduce Applications
- Map Aggregation

Day 2

- Partitioning and Sorting
- Input and Output Formats

Day 3

- Optimizing MapReduce Jobs
- Advanced MapReduce Features
- Unit Testing
- Finding Trends

Day 4

- Controlling Workflow with Oozie
- Apache HBase Development
- Integrating MapReduce with Apache Pig and Apache Hive

Hortonworks CERTIFIED DEVELOPER

Hortonworks certification identifies you as an expert in the Apache Hadoop ecosystem. Hortonworks offers a comprehensive certification program for students that attend a Hortonworks public or private on-site training course. Please visit hortonworks.com/training for more information.

Hortonworks University

Hortonworks University is your expert source for Apache Hadoop training and certification. Public and private on-site courses are available for developers, administrators, data analysts and other IT professionals involved in implementing big data solutions. Classes combine presentation material with industry-leading hands-on labs that fully prepare students for real-world Hadoop scenarios.

Hortonworks

About Hortonworks

Hortonworks is a leading commercial vendor of Apache Hadoop, the preeminent open source platform for storing, managing and analyzing big data. Our distribution, Hortonworks Data Platform powered by Apache Hadoop, provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy big data solutions. Hortonworks is the trusted source for information on Hadoop, and together with the Apache community, Hortonworks is making Hadoop more robust and easier to install, manage and use. Hortonworks provides unmatched technical support, training and certification programs for enterprises, systems integrators and technology vendors.

US: 1.855.846.7866
International: 1.408.916.4121
www.hortonworks.com

3460 West Bayshore Road
Palo Alto, CA 94303 USA