Data Sheet

Developing Solutions Using Apache Hadoop

This four-day course provides Java programmers the necessary training for creating enterprise solutions using Apache Hadoop. It consists of an effective mix of interactive lecture and extensive hand-on lab exercises.

After successfully completing this course each student will receive one free voucher for the Hadoop Certified Developer exam.

Course Objectives

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

- Write a MapReduce program using Hadoop API
- Utilize HDFS for effective loading and processing of data with CLI and API.
- Understand best practices for building, debugging and optimizing Hadoop solutions.
- Use Pig, Hive, HBase and HCatalog effectively

Extensive Hands-On Lab Experience

Students will work through the following lab exercises using the Hortonworks Data Platform:

- Running a Hadoop Solution
- MapReduce Programming
- HDFS
- MapReduce in Operation
- MapReduce with Combiner
- MapReduce with Partitioner
- MapReduce with a Secondary Sort and a Custom Comparator
- MapReduce with Distributed Cache
- MapReduce with Data Handling
- MapReduce with Streaming

- Pig Into
- Pig Data Operations
- Pig ETL Features
- Pig Clustering Solution
- Hive Introduction
- Hive Features
- Combined Hive and Pig Solution
- HBase
- HCatalog
- MapReduce, Pig, Hive and HCatalog in a Combined Solutions

Course Outline

Day 1
- Overview
- MapReduce Code
- HDFS
- MapReduce - JobTracker, TaskTracker and Running Jobs

Day 2
- MapReduce Combiner
- MapReduce Partitioner
- MapReduce Distributed Cache
- MapReduce Streaming
- MapReduce Data Handling

Day 3
- Pig Into
- Pig Data Model
- Pig Scripting Language
- Hive - Part 1

Day 4
- Hive - Part 2
- HCatalog
- HBase
- Enterprise Integration
- Future of Hadoop

Target Audience

Java programmers who want to better understand how to create Apache Hadoop solutions.

Additional Information

- For availability for individual seats in our open enrollment classes please visit us at www.hortonworks.com/training
- Customized, onsite training is also available to be hosted at your offices.
- For more information around pricing and private training opportunities contact us at hwuniversity@hortonworks.com

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 accelerating the adoption of Apache Hadoop by bridging the technology and knowledge gaps that exist today. We are driving much of the design and development of both the current and future generations of Apache Hadoop and leveraging our development expertise to provide unmatched expert technical support, training and certification programs for enterprises, systems integrators and ISVs.

US: 1.855.846.7866
International: 1.408.916.4121
www.hortonworks.com

455 W. Maude Avenue, Suite 200
Sunnyvale, CA 94085 USA