Developer

HDP Developer: Java

Overview

This advanced course provides Java programmers a deep-dive into Hadoop application development. Students will learn how to design and develop efficient and effective MapReduce applications for Hadoop using the Hortonworks Data Platform, including how to implement combiners, partitioners, secondary sorts, custom input and output formats, joining large datasets, unit testing, and developing UDFs for Pig and Hive. Labs are run on a 7-node HDP 2.1 cluster running in a virtual machine that students can keep for use after the training.

Duration

4 days

Target Audience

Experienced Java software engineers who need to develop Java MapReduce applications for Hadoop.

Course Objectives

- Describe Hadoop 2 and the Hadoop Distributed File System
- Describe the YARN framework
- Develop and run a Java MapReduce application on YARN
- Use combiners and in-map aggregation
- Write a custom partitioner to avoid data skew on reducers
- Perform a secondary sort
- Recognize use cases for built-in input and output formats
- Write a custom MapReduce input and output format
- Optimize a MapReduce job
- Configure MapReduce to optimize mappers and reducers
- Develop a custom RawComparator class
- Distribute files as LocalResources
- Describe and perform join techniques in Hadoop
- Perform unit tests using the UnitMR API
- Describe the basic architecture of HBase
- Write an HBase MapReduce application
- List use cases for Pig and Hive
- Write a simple Pig script to explore and transform big data
- Write a Pig UDF (User-Defined Function) in Java
- Write a Hive UDF in Java
- Use JobControl class to create a MapReduce workflow
- Use Oozie to define and schedule workflows

Hands-On Labs

- Configuring a Hadoop Development Environment
- Putting data into HDFS using Java
- Write a distributed grep MapReduce application
- Write an inverted index MapReduce application
- Configure and use a combiner
- Writing custom combiners and partitioners
- Globally sort output using the TotalOrderPartitioner
- Writing a MapReduce job to sort data using a composite key
- Writing a custom InputFormat class
- Writing a custom OutputFormat class
- Compute a simple moving average of stock price data
- Use data compression
- Define a RawComparator
- Perform a map-side join
- Using a Bloom filter
- Unit testing a MapReduce job
- Importing data into HBase
- Writing an HBase MapReduce job
- Writing User-Defined Pig and Hive functions
- Defining an Oozie workflow

Prerequisites

Students must have experience developing Java applications and using a Java IDE. Labs are completed using the Eclipse IDE and Gradle. No prior Hadoop knowledge is required.

Format

50% Lecture/Discussion

50% Hands-on Labs

Certification

Hortonworks offers a comprehensive certification program that identifies you as an expert in Apache Hadoop. Visit *hortonworks.com/training/certification* 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.

About Hortonworks

Hortonworks develops, distributes and supports the only 100 percent open source distribution of Apache Hadoop explicitly architected, built and tested for enterprise-grade deployments.

US: 1.855.846.7866
International: +1.408.916.4121
www.hortonworks.com

5470 Great America Parkway
Santa Clara, CA 95054 USA