



Data Sheet

## Developing Solutions for Apache™ Hadoop® on Windows

Students will learn to develop applications and analyze big data stored in Apache Hadoop running on Microsoft Windows. Students will learn the details of the Hadoop Distributed File System (HDFS™) architecture and MapReduce framework, as well as learn how to develop applications on Hadoop® using tools like C#, Pig™, Hive™, HCatalog, Sqoop, Oozie and Microsoft Excel.

### Duration:

4 days

### Prerequisites

Students should have programming experience, preferably with Visual Studio and SQL, as well as familiarity with the Windows Server operating system. No prior Hadoop knowledge is required.

### Target Audience:

.NET Developers and Data Analysts responsible for developing applications and performing analysis on big data using the Hortonworks Data Platform for Windows.

### Course Objectives

At the completion of the course students will be able to:

- Explain the various tools and frameworks in the Hadoop ecosystem
- Recognize use cases for HDP for Windows and Big Data
- Explain the architecture of the Hadoop Distributed File System (HDFS™)
- Transfer data between HDFS and Microsoft SQL Server using Sqoop
- Explain the architecture of MapReduce
- Run a MapReduce job on Hadoop
- Use Hadoop streaming
- Use the Microsoft .NET API for Hadoop to write a C# MapReduce job
- Recognize use cases for Pig
- Write a Pig script to explore and transform data in HDFS
- Define advanced Pig relations
- Use Pig to apply structure to unstructured Big Data
- Join large datasets using Pig
- Invoke a Pig User-Defined Function
- Write a Hive query using Hive QL
- Understand how Hive tables are defined and implemented
- Use Hive to run SQL-like queries to perform data analysis
- Explain the uses and purpose of HCatalog
- Use an HCatalog schema within a Pig script
- Explain the purpose of the Hive ODBC driver
- Connect Microsoft Excel to HDFS using Hive ODBC
- Import Hive query results into Excel
- Explain the usages of Oozie
- Write and execute an Oozie workflow

## Agenda

### Day 1

- Understanding Big Data and Hadoop
- The Hadoop Distributed File System (HDFS)
- Inputting Data into HDFS

### Day 2

- The MapReduce Framework
- Developing MapReduce Applications in .NET
- Introduction to Pig

### Day 3

- Advanced Pig Programming
- Hive Programming

### Day 4

- Using HCatalog
- The Hive ODBC Driver
- Defining Workflow with Oozie

## Lab Content

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

- Access HDFS using the HDFS commands
- Import SQL Server data into HDFS using Sqoop
- Export HDFS data from HDFS into SQL Server using Sqoop
- Run a MapReduce Job
- Monitor a MapReduce Job

- Develop a .NET MapReduce application in C#
- Explore data using Pig
- Split and join datasets using Pig
- Transform unstructured for use with Hive
- Analyze Big Data with Hive

- Understanding MapReduce with Hive
- Joining datasets with Hive
- Use HCatalog with Pig
- Use Hive ODBC with Microsoft Excel
- Define an Oozie Workflow



**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](http://hortonworks.com/training) for more information.



**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](http://www.hortonworks.com)**

3460 West Bayshore Road  
Palo Alto, CA 94303 USA