



## HDP Developer: Windows

### Overview

This course is designed for developers who create applications and analyze Big Data in Apache Hadoop on Windows using Pig and Hive. Topics include: Hadoop, YARN, the Hadoop Distributed File System (HDFS), MapReduce, Sqoop and the HiveODBC Driver.

### Duration

4 days

### Target Audience

Software developers who need to understand and develop applications for Hadoop 2.x on Windows.

### Course Objectives

- Describe Hadoop and Hadoop and YARN
- Describe the Hadoop ecosystem
- List Components & deployment options for HDP on Windows
- Describe the HDFS architecture
- Use the Hadoop client to input data into HDFS
- Transfer data between Hadoop and Microsoft SQL Server
- Describe the MapReduce and YARN architecture
- Run a MapReduce job on YARN
- Write a Pig script
- Define advanced Pig relations
- Use Pig to apply structure to unstructured Big Data
- Invoke a Pig User-Defined Function
- Use Pig to organize and analyze Big Data
- Describe how Hive tables are defined and implemented
- Use Hive windowing functions
- Define and use Hive file formats
- Create Hive tables that use the ORC file format
- Use Hive to run SQL-like queries to perform data analysis
- Use Hive to join datasets
- Create ngrams and context ngrams using Hive
- Perform data analytics
- Use HCatalog with Pig and Hive
- Install and configure HiveODBC Driver for Windows
- Import data from Hadoop into Microsoft Excel
- Define a workflow using Oozie

### Hands-On Labs

- Start HDP on Windows
- Add/remove files and folders from HDFS
- Transfer data between HDFS and Microsoft SQL Server
- Run a MapReduce job
- Using Pig to analyze data
- Retrieve HCatalog schemas from within a Pig script
- Using Hive tables and queries
- Advanced Hive features like windowing, views and ORC files
- Hive analytics functions using the Pig DataFu library
- Compute quantiles
- Use Hive to compute ngrams on Avro-formatted files
- Connect Microsoft Excel to Hadoop with HiveODBC Driver
- Run a YARN application
- Define an Oozie workflow

### Prerequisites

Students should be familiar with programming principles and have experience in software development. SQL knowledge and familiarity with Microsoft Windows is also helpful. 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](http://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](http://www.hortonworks.com)

5470 Great America Parkway  
Santa Clara, CA 95054 USA



Hortonworks<sup>®</sup>  
**UNIVERSITY**



Developer

---



**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)

5470 Great America Parkway  
Santa Clara, CA 95054 USA