Hewlett Packard Enterprise

Hortonworks Data Platform Developer—Windows (EDU-PRIV-DEV-WINDOWS-200) H7G71S

HPE course number	H7G71S
Course length	4 days
Delivery mode	ILT
View schedule, local pricing, and register	**View now**
View related courses	**View now**

This 4-day hands-on training course teaches students how to develop applications and analyze Big Data stored in Apache Hadoop on Windows® using Pig and Hive. Students will learn the details of Hadoop 2.x, YARN, the Hadoop Distributed File System (HDFS), an overview of MapReduce, and a deep dive into using Pig and Hive to perform data analytics on Big Data. Other topics covered include using Sqoop to transfer data between Hadoop and Microsoft SQL Server, and connecting Microsoft Excel to Hadoop using the HiveODBC Driver. This course is 50 percent instructor-led lecture and 50 percent hands-on labs.

Audience

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

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

Course objectives

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

- Explain Hadoop and YARN
- Explain use cases for Hadoop
- Explain the various tools and frameworks in the Hadoop 2.x ecosystem
- Explain the components of the Hortonworks Data Platform on Windows
- Explain the deployment options for HDP on Windows
- Explain the architecture of the Hadoop Distributed File System (HDFS)
- Use the Hadoop client to input data into HDFS
- Use Sqoop to transfer data between Hadoop and Microsoft SQL Server
- Explain the architecture of MapReduce
- Explain the architecture of YARN
- Run a MapReduce job on YARN
- Write a Pig script to explore and transform data in HDFS
- Define advanced Pig relations

Why HPE Education Services?

- IDC MarketScape leader 4 years running for IT education and training*
- Recognized by IDC for leading with global coverage, unmatched technical expertise, and targeted education consulting services*
- Key partnerships with industry leaders OpenStack®, VMware®, Linux®, Microsoft®, ITIL, PMI, CSA, and (ISC)²
- Complete continuum of training delivery options—self-paced eLearning, custom education consulting, traditional classroom, video on-demand instruction, live virtual instructor-led with hands-on lab, dedicated onsite training
- Simplified purchase option with HPE Training Credits

*Realize Technology Value with Training, IDC Infographic 2037, Sponsored by HPE, January 2016

- Use Pig to apply structure to unstructured Big Data
- Invoke a Pig User-Defined Function
- Use Pig to organize and analyze Big Data
- Understand how Hive tables are defined and implemented
- Use the new Hive windowing functions
- Explain and use the various Hive file formats
- Create and populate a Hive table that uses the new ORC file format
- Use Hive to run SQL-like queries to perform data analysis
- Use Hive to join datasets using a variety of techniques, including Map-side joins and Sort-Merge-Bucket joins
- Write efficient Hive queries
- Create ngrams and context ngrams using Hive
- Perform data analytics like quantiles and page rank on Big Data using the DataFu Pig library
- Explain the uses and purpose of HCatalog
- Use HCatalog with Pig and Hive
- Install and configure the HiveODBC Driver for Windows
- Import data from Hadoop into Microsoft Excel
- Define a workflow using Oozie

Benefits to you

- This course will provide in depth explanation on how to develop applications and analyze Big Data stored in Apache Hadoop on Windows using Pig and Hive

Detailed course outline

Day 1	• Unit 1: Understanding Hadoop • Unit 2: The Hadoop Distributed File System (HDFS) • Unit 3: Inputting Data into HDFS • Unit 4: The MapReduce Framework Day 2
Day 2	• Unit 5: Introduction to Pig • Unit 6: Advanced Pig Programming Day 3
Day 3	• Unit 7: Hive Programming • Unit 8: Using HCatalog • Unit 9: Advanced Hive Programming Day 4
Day 4	• Unit 10: The Hive ODBC Driver • Unit 11: Hadoop 2 and YARN • Appendix A: Defining Workflow with Oozie
Hands-On Labs	Students will work through the following lab exercises using the Hortonworks Data Platform 2.1 on Windows: • Start HDP on Windows • Use HDFS commands to add/remove files and folders from HDFS • Use Sqoop to transfer data between HDFS and Microsoft SQL Server • Run a MapReduce job • Explore and transform data using Pig • Split a dataset using Pig • Join two datasets using Pig • Use Pig to transform and export a dataset for use with Hive • Use HCatLoader and HCatStorer to retrieve HCatalog schemas from within a Pig script • Understand how a Hive table is stored in HDFS • Use Hive to discover useful information in a dataset • Understand how Hive queries get executed as MapReduce jobs • Perform a join of two datasets with Hive • Use advanced Hive features like windowing, views and ORC files • Use the Hive analytics functions (rank, dense_rank, cume_dist, row_number) • Analyze and sessionize clickstream data using the Pig DataFu library • Compute quantiles of NYSE stock prices • Use Hive to compute ngrams on Avro—formatted files • Connect Microsoft Excel to Hadoop using the HiveODBC Driver • Run a YARN application • Define an Oozie workflow

Learn more at
hpe.com/ww/learnbigdata

Follow us:

c04577747, December 2016, Rev. 1