

HDP Developer Training: Pig & Hive

OVERVIEW

This course is designed for developers who need to create applications to analyze Big Data stored in Apache Hadoop using Pig and Hive. Topics include:

- > Hadoop, YARN, HDFS, MapReduce
- > Data ingestion
- > Workflow definition and using Pig
- > Using Hive to perform data analytics on Big Data

METHODOLOGY

Duration

4-days

Format

- > 50% Lecture/Discussion
- > 50% Hands-on Labs

Target Audience

Software developers who need to understand and develop applications for Hadoop

COURSE OBJECTIVES

- | | |
|---|---|
| > Describe Hadoop, YARN and use cases for Hadoop | > Create and populate a Hive table that uses ORC file formats |
| > Describe Hadoop ecosystem tools and frameworks | > Use Hive to run SQL-like queries to perform data analysis |
| > Describe the HDFS architecture | > Use Hive to join datasets using a variety of techniques, including Map-side joins and Sort-Merge-Bucket joins |
| > Use the Hadoop client to input data into HDFS | > Write efficient Hive queries |
| > Transfer data between Hadoop and a relational database | > Create ngrams and context ngrams using Hive |
| > Explain YARN and MapReduce architectures | > Perform data analytics like quantiles and page rank on Big |
| > Run a MapReduce job on YARN | > Data using the DataFu Pig library |
| > Use Pig to explore and transform data in HDFS | > Explain the uses and purpose of HCatalog |
| > Use Hive to explore and understand how Hive tables are defined, implemented and used to analyze data sets | > Use HCatalog with Pig and Hive |
| > Use the new Hive windowing functions | > Define a workflow using Oozie |
| > Explain and use the various Hive file formats | > Schedule a recurring workflow using the Oozie Coordinator |

PRE-REQUISITES

- > Attendees should have a basic understanding of big data and Hadoop principles
- > Students should be familiar with programming principles and have experience in software development
- > Experience with at least one ETL tool is a plus
- > Programming or scripting language is a plus
- > Familiarity with a Unix Shell (i.e Bash) is a plus
- > SQL and relational DB knowledge is a plus

LIMITATIONS

- > Classroom size is limited to 10 student so that personalized instruction can be provided to maximize learning

HANDS-ON LABS

- | | |
|---|--|
| > Use HDFS commands to add/remove files and folders | > Describe how Hive queries get executed as MapReduce jobs |
| > Use Sqoop to transfer data between HDFS and a RDBMS | > Perform a join of two datasets with Hive |
| > Run MapReduce and YARN application jobs | > Use advanced Hive features: windowing, views, ORC files |
| > Explore and transform data using Pig | > Use Hive analytics functions |
| > Split and join a dataset using Pig | > Write a custom reducer in Python |
| > Use Pig to transform and export a dataset for use with Hive | > Analyze and sessionize clickstream data |
| > Use HCatLoader and HCatStorer | > Compute quantiles of NYSE stock prices |
| > Use Hive to discover useful information in a dataset | > Use Hive to compute ngrams on Avro-formatted files |
| | > Define an Oozie workflow |

DELIVERABLES

- > Students will be provided reference material
 - o Hortonworks will ship printed copies of a student guide and a lab manual
 - o You will also receive electronic versions of the same student guide and a lab manual beforehand
 - o A VM will be made available for the course and is available for later use
- > Students that complete the training will be awarded a Certificate of Competition by Hortonworks

RELATED SERVICES

- > OnX HDP Administration Training
- > OnX HDP Data Science Training
- > OnX Hadoop Appliance
- > OnX Hadoop Quick Start
- > OnX Hadoop Development
- > OnX Business Analytics and Data Science
- > OnX Hadoop Administration

ACCEPTANCE

Service is deemed accepted upon notification to customer from OnX of completion of service scope activities listed above.

TERMS AND CONDITIONS

Unless otherwise agreed to in writing by the parties in a document effective as of the date hereof, these products and services are offered subject to the applicable terms, conditions and agreements and the OnX Standard Terms and Conditions of Sale (U.S) available at: www.onx.com/TermsAndConditionsOfSale/ . Prices may not include taxes, or insurance charges, if any.