Hortonworks University

HDP Certified Developer (HDPCD)

Certification Overview

Hortonworks has redesigned its certification program to create an industry-recognized certification where individuals prove their Hadoop knowledge by performing actual hands-on tasks on a Hortonworks Data Platform (HDP) cluster, as opposed to answering multiple-choice questions. The HDP Certified Developer (HDPCD) exam is the first of our new hands-on, performance-based exams designed for Hadoop developers working with frameworks like Pig, Hive, Sqoop and Flume.

Purpose of the Exam

The purpose of this exam is to provide organizations that use Hadoop with a means of identifying suitably qualified staff to develop Hadoop applications for storing, processing, and analyzing data stored in Hadoop using the open-source tools of the Hortonworks Data Platform (HDP), including Pig, Hive, Sqoop and Flume.

Exam Description

The exam has three main categories of tasks that involve:

- Data ingestion
- Data transformation
- Data analysis

The exam is based on the Hortonworks Data Platform 2.2 installed and managed with Ambari 1.7.0, which includes Pig 0.14.0, Hive 0.14.0, Sqoop 1.4.5, and Flume 1.5.0. Each candidate will be given access to an HDP 2.2 cluster along with a list of tasks to be performed on that cluster.

Exam Objectives

View the complete list of objectives below, which includes links to the corresponding documentation and/or other resources.

Language

The exam is delivered in English.

Take the Exam Anytime, Anywhere

The HDPCD exam is available from any computer, anywhere, at any time. All you need is a webcam and a good Internet connection.

How to Register

Candidates need to create an account at www.examslocal.com. Once you are registered and logged in, select “Schedule an Exam”, and then enter “Hortonworks” in the “Search Here” field to locate and select the HDP Certified Developer exam. The cost of the exam is $250 USD.

Duration

2 hours

Description of the Minimally Qualified Candidate

The Minimally Qualified Candidate (MQC) for this certification can develop Hadoop applications for ingesting, transforming, and analyzing data stored in Hadoop using the open-source tools of the Hortonworks Data Platform, including Pig, Hive, Sqoop and Flume.

Prerequisites

Candidates for the HPDCD exam should be able to perform each of the tasks in the list of exam objectives below. Candidates are also encouraged to attempt the practice exam. Visit www.hortonworks.com/training/class/hdp-certified-developer-hdpcd-exam/ for more details.

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.

Hortonworks

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 American Parkway
Santa Clara, CA 95054 USA

Hortonworks University

HDP Certified Developer (HDPCD) Exam Objectives

Candidates for the HPDCD exam should be able to perform each of the tasks below:

Category	Task	Resource(s)
Data Ingestion	Input a local file into HDFS using the Hadoop file system shell	http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html#put
	Make a new directory in HDFS using the Hadoop file system shell	http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html#mkdir
	Import data from a table in a relational database into HDFS	http://sqoop.apache.org/docs/1.4.5/SqoopUserGuide.html#_literal_sqoop_import_literal
	Import the results of a query from a relational database into HDFS	http://sqoop.apache.org/docs/1.4.5/SqoopUserGuide.html#_free_form_query_imports
	Import a table from a relational database into a new or existing Hive table	http://sqoop.apache.org/docs/1.4.5/SqoopUserGuide.html#_importing_data_into_hive
	Insert or update data from HDFS into a table in a relational database	http://sqoop.apache.org/docs/1.4.5/SqoopUserGuide.html#_literal_sqoop_export_literal
	Given a Flume configuration file, start a Flume agent	https://flume.apache.org/FlumeUserGuide.html#starting-an-agent
	Given a configured sink and source, configure a Flume memory channel with a specified capacity	https://flume.apache.org/FlumeUserGuide.html#memory-channel

Category	Task	Resource(s)
Data Transformation	Write and execute a Pig script	https://pig.apache.org/docs/r0.14.0/start.html#run
	Load data into a Pig relation without a schema	https://pig.apache.org/docs/r0.14.0/basic.html#load
	Load data into a Pig relation with a schema	https://pig.apache.org/docs/r0.14.0/basic.html#load
	Load data from a Hive table into a Pig relation	https://cwiki.apache.org/confluence/display/Hive/HCatalog+LoadStore
	Use Pig to transform data into a specified format	https://pig.apache.org/docs/r0.14.0/basic.html#foreach
	Transform data to match a given Hive schema	https://pig.apache.org/docs/r0.14.0/basic.html#foreach
	Group the data of one or more Pig relations	https://pig.apache.org/docs/r0.14.0/basic.html#group
	Use Pig to remove records with null values from a relation	https://pig.apache.org/docs/r0.14.0/basic.html#filter
	Store the data from a Pig relation into a folder in HDFS	https://pig.apache.org/docs/r0.14.0/basic.html#store
	Store the data from a Pig relation into a Hive table	https://cwiki.apache.org/confluence/display/Hive/HCatalog+LoadStore
	Sort the output of a Pig relation	https://pig.apache.org/docs/r0.14.0/basic.html#order-by
	Remove the duplicate tuples of a Pig relation	https://pig.apache.org/docs/r0.14.0/basic.html#distinct

Hortonworks

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 American Parkway
Santa Clara, CA 95054 USA

	Specify the number of reduce tasks for a Pig MapReduce job	https://pig.apache.org/docs/r0.14.0/perf.html#parallel
	Join two datasets using Pig	https://pig.apache.org/docs/r0.14.0/basic.html#join-inner and https://pig.apache.org/docs/r0.14.0/basic.html#join-outer
	Perform a replicated join using Pig	https://pig.apache.org/docs/r0.14.0/perf.html#replicated-joins
	Run a Pig job using Tez	https://pig.apache.org/docs/r0.14.0/perf.html#tez-mode
	Within a Pig script, register a JAR file of User Defined Functions	https://pig.apache.org/docs/r0.14.0/basic.html#register and https://pig.apache.org/docs/r0.14.0/udf.html#piggybank
	Within a Pig script, define an alias for a User Defined Function	https://pig.apache.org/docs/r0.14.0/basic.html#define-udfs
	Within a Pig script, invoke a User Defined Function	https://pig.apache.org/docs/r0.14.0/basic.html#register

Category	Task	Resource(s)
Data Analysis	Write and execute a Hive query	https://cwiki.apache.org/confluence/display/Hive/Tutorial
	Define a Hive-managed table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-Create/Drop/TruncateTable
	Define a Hive external table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-ExternalTables
	Define a partitioned Hive table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-PartitionedTables
	Define a bucketed Hive table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-BucketedSortedTables
	Define a Hive table from a select query	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-CreateTableAsSelect(CTAS)
	Define a Hive table that uses the ORCFile format	http://hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/
	Create a new ORCFile table from the data in an existing non-ORCFile Hive table	http://hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/
	Specify the storage format of a Hive table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-RowFormat,StorageFormat,andSerDe
	Specify the delimiter of a Hive table	http://hortonworks.com/hadoop-tutorial/using-hive-data-analysis/
	Load data into a Hive table from a local directory	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-

		Loadingfilesintotables
	Load data into a Hive table from an HDFS directory	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables
	Load data into a Hive table as the result of a query	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingdataintoHiveTablesfromqueries
	Load a compressed data file into a Hive table	https://cwiki.apache.org/confluence/display/Hive/CompressedStorage
	Update a row in a Hive table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Update
	Delete a row from a Hive table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Delete
	Insert a new row into a Hive table	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingvaluesintotablesfromSQL
	Join two Hive tables	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Joins
	Run a Hive query using Tez	http://hortonworks.com/hadoop-tutorial/supercharging-interactive-queries-hive-tez/
	Run a Hive query using vectorization	http://hortonworks.com/hadoop-tutorial/supercharging-interactive-queries-hive-tez/
	Output the execution plan for a Hive query	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Explain
	Use a subquery within a Hive query	https://cwiki.apache.org/confluence/display/Hive/LanguageManual+SubQueries
	Output data from a Hive query that is totally ordered across multiple reducers	https://issues.apache.org/jira/browse/HIVE-1402
	Set a Hadoop or Hive configuration property from within a Hive query	http://hortonworks.com/wp-content/uploads/downloads/2013/08/Hortonworks.CheatSheet.SQLtoHive.pdf