

## HDP Certified Developer (HDPD) Exam

### 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 (HDPD) 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.

### 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.

### Language

The exam is delivered in English

### 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



**System  
Admin**

**Data  
Analyst**

**Developer**

## HDP Certified Developer (HDP-CD) Exam Objectives

Candidates for the HDP-CD exam should be able to perform each of the tasks below:

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

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



### 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 America Parkway  
Santa Clara, CA 95054 USA

	HDFS	<a href="#">store</a>
	Store the data from a Pig relation into a Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/HCatalog+LoadStore">https://cwiki.apache.org/confluence/display/Hive/HCatalog+LoadStore</a>
	Sort the output of a Pig relation	<a href="https://pig.apache.org/docs/r0.14.0/basic.html#order-by">https://pig.apache.org/docs/r0.14.0/basic.html#order-by</a>
	Remove the duplicate tuples of a Pig relation	<a href="https://pig.apache.org/docs/r0.14.0/basic.html#distinct">https://pig.apache.org/docs/r0.14.0/basic.html#distinct</a>
	Specify the number of reduce tasks for a Pig MapReduce job	<a href="https://pig.apache.org/docs/r0.14.0/perf.html#parallel">https://pig.apache.org/docs/r0.14.0/perf.html#parallel</a>
	Join two datasets using Pig	<a href="https://pig.apache.org/docs/r0.14.0/basic.html#join-inner">https://pig.apache.org/docs/r0.14.0/basic.html#join-inner</a> and <a href="https://pig.apache.org/docs/r0.14.0/basic.html#join-outer">https://pig.apache.org/docs/r0.14.0/basic.html#join-outer</a>
	Perform a replicated join using Pig	<a href="https://pig.apache.org/docs/r0.14.0/perf.html#replicated-joins">https://pig.apache.org/docs/r0.14.0/perf.html#replicated-joins</a>
	Run a Pig job using Tez	<a href="https://pig.apache.org/docs/r0.14.0/perf.html#tez-mode">https://pig.apache.org/docs/r0.14.0/perf.html#tez-mode</a>
	Within a Pig script, register a JAR file of User Defined Functions	<a href="https://pig.apache.org/docs/r0.14.0/basic.html#register">https://pig.apache.org/docs/r0.14.0/basic.html#register</a> and <a href="https://pig.apache.org/docs/r0.14.0/udf.html#piggybank">https://pig.apache.org/docs/r0.14.0/udf.html#piggybank</a>
	Within a Pig script, define an alias for a User Defined Function	<a href="https://pig.apache.org/docs/r0.14.0/basic.html#define-udfs">https://pig.apache.org/docs/r0.14.0/basic.html#define-udfs</a>
	Within a Pig script, invoke a User Defined Function	<a href="https://pig.apache.org/docs/r0.14.0/basic.html#register">https://pig.apache.org/docs/r0.14.0/basic.html#register</a>

Category	Task	Resource(s)
<b>Data Analysis</b>	Write and execute a Hive query	<a href="https://cwiki.apache.org/confluence/display/Hive/Tutorial">https://cwiki.apache.org/confluence/display/Hive/Tutorial</a>
	Define a Hive-managed table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-Create/Drop/TruncateTable">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-Create/Drop/TruncateTable</a>
	Define a Hive external table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-ExternalTables">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-ExternalTables</a>
	Define a partitioned Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-PartitionedTables">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-PartitionedTables</a>
	Define a bucketed Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-BucketedSortedTables">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-BucketedSortedTables</a>
	Define a Hive table from a select query	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-</a>

**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 America Parkway  
Santa Clara, CA 95054 USA



**System Admin    Data Analyst    Developer**

		<a href="#">CreateTableAsSelect(CTAS)</a>
	Define a Hive table that uses the ORCFile format	<a href="http://hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/">http://hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/</a>
	Create a new ORCFile table from the data in an existing non-ORCFile Hive table	<a href="http://hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/">http://hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/</a>
	Specify the storage format of a Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-RowFormat,StorageFormat,andSerDe">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-RowFormat,StorageFormat,andSerDe</a>
	Specify the delimiter of a Hive table	<a href="http://hortonworks.com/hadoop-tutorial/using-hive-data-analysis/">http://hortonworks.com/hadoop-tutorial/using-hive-data-analysis/</a>
	Load data into a Hive table from a local directory	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables</a>
	Load data into a Hive table from an HDFS directory	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables</a>
	Load data into a Hive table as the result of a query	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingdataintoHiveTablesfromqueries">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingdataintoHiveTablesfromqueries</a>
	Load a compressed data file into a Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/CompressedStorage">https://cwiki.apache.org/confluence/display/Hive/CompressedStorage</a>
	Update a row in a Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Update">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Update</a>
	Delete a row from a Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Delete">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Delete</a>
	Insert a new row into a Hive table	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingvaluesintotablesfromSQL">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingvaluesintotablesfromSQL</a>
	Join two Hive tables	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Joins">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Joins</a>
	Run a Hive query using Tez	<a href="http://hortonworks.com/hadoop-tutorial/supercharging-interactive-queries-hive-tez/">http://hortonworks.com/hadoop-tutorial/supercharging-interactive-queries-hive-tez/</a>
	Run a Hive query using vectorization	<a href="http://hortonworks.com/hadoop-tutorial/supercharging-interactive-queries-hive-tez/">http://hortonworks.com/hadoop-tutorial/supercharging-interactive-queries-hive-tez/</a>
	Output the execution plan for a Hive query	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Explain">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Explain</a>
	Use a subquery within a Hive query	<a href="https://cwiki.apache.org/confluence/display/Hive/LanguageManual+SubQueries">https://cwiki.apache.org/confluence/display/Hive/LanguageManual+SubQueries</a>
	Output data from a Hive query that is totally ordered across multiple reducers	<a href="https://issues.apache.org/jira/browse/HIVE-1402">https://issues.apache.org/jira/browse/HIVE-1402</a>



**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



	Set a Hadoop or Hive configuration property from within a Hive query	<a href="http://hortonworks.com/wp-content/uploads/downloads/2013/08/Hortonworks.CheatSheet.SQLtoHive.pdf">http://hortonworks.com/wp-content/uploads/downloads/2013/08/Hortonworks.CheatSheet.SQLtoHive.pdf</a>
--	--	---



**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