

# Hadoop for Systems Administrators (GL660) H6C60S

<b>HPE course number</b>	H6C60S
<b>Course length</b>	3 days
<b>Delivery mode</b>	ILT, VILT
<b>View schedule, local pricing, and register</b>	<a href="#">View now</a>
<b>View related courses</b>	<a href="#">View now</a>

## 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

This course covers the essentials of deploying and managing an Apache™ Hadoop® cluster. The course is lab intensive with each participant creating their own Hadoop cluster using either the CDH (Cloudera's Distribution, including Apache Hadoop) or Hortonworks Data Platform stacks. Core Hadoop services are explored in depth with emphasis on troubleshooting and recovering from common cluster failures. The fundamentals of related services such as Ambari, Zookeeper, Pig, Hive, HBase, Sqoop, Flume, and Oozie are also covered. The course is approximately 60% lecture and 40% labs.

## Audience

Systems Administrators who will be responsible for managing and administering Hadoop clusters.

## Prerequisites

Qualified participants should be comfortable with the Linux commands and have some systems administration experience, but do not need previous Hadoop experience.

## Supported Distributions

Red Hat Enterprise Linux 7

## Detailed Course Outline

<b>Module 1: Hadoop Overview</b>	<ul style="list-style-type: none"> <li>• Data Analysis</li> <li>• Big Data</li> <li>• Origins of Hadoop</li> <li>• Hadoop Marketplace</li> <li>• Hadoop Core</li> <li>• Hadoop Ecosystem:</li> <li>• Hadoop Ecosystem (cont):</li> <li>• Hadoop Ecosystem (cont):</li> </ul>	<ul style="list-style-type: none"> <li>• Hadoop Ecosystem (cont):</li> <li>• Hadoop Ecosystem (cont):</li> <li>• Cluster Architecture</li> <li>• Hardware/Software Requirements</li> <li>• Running Commands on Multiple Systems</li> </ul> <p><b>Lab Tasks</b></p> <ul style="list-style-type: none"> <li>• Running Commands on Multiple Hosts</li> <li>• Preparing to Install Hadoop</li> </ul>
<b>Module 2: HDFS</b>	<ul style="list-style-type: none"> <li>• Design Goals</li> <li>• Design</li> <li>• Blocks</li> <li>• Block Replication</li> <li>• Namenode Daemon</li> <li>• Secondary Namenode Daemon</li> <li>• Datanode Daemon</li> <li>• Accessing HDFS</li> </ul>	<ul style="list-style-type: none"> <li>• Permissions and Users</li> <li>• Adding and Removing Datanodes</li> <li>• Balancing</li> </ul> <p><b>Lab Tasks</b></p> <ul style="list-style-type: none"> <li>• Single Node HDFS</li> <li>• Multi-node HDFS</li> <li>• Files and HDFS</li> <li>• Managing and Maintaining HDFS</li> </ul>
<b>Module 3: YARN</b>	<ul style="list-style-type: none"> <li>• YARN Design Goals</li> <li>• YARN Architecture</li> <li>• Resource Manager</li> <li>• Node Manager</li> <li>• Containers</li> </ul>	<ul style="list-style-type: none"> <li>• YARN: Other Important Features</li> <li>• Slider</li> </ul> <p><b>Lab Tasks</b></p> <ul style="list-style-type: none"> <li>• YARN</li> </ul>
<b>Module 4: MapReduce</b>	<ul style="list-style-type: none"> <li>• MapReduce</li> <li>• Terminology and Data Flow</li> </ul>	<p><b>Lab Tasks</b></p> <ul style="list-style-type: none"> <li>• Mapreduce</li> </ul>
<b>Module 5: Installing Hadoop with Ambari</b>	<p><b>Lab Tasks</b></p> <ul style="list-style-type: none"> <li>• CDH Uninstall</li> </ul>	<ul style="list-style-type: none"> <li>• Installing Hadoop with Ambari</li> <li>• Tez</li> </ul>
<b>Module 6: Data Ingestion</b>	<ul style="list-style-type: none"> <li>• Sqoop</li> <li>• Flume</li> <li>• Kafka</li> </ul>	<p><b>Lab Tasks</b></p> <ul style="list-style-type: none"> <li>• Sqoop</li> </ul>
<b>Module 7: Data Lineage and Governance</b>	<ul style="list-style-type: none"> <li>• Falcon</li> <li>• Atlas</li> </ul>	<ul style="list-style-type: none"> <li>• Oozie</li> </ul>
<b>Module 8: Data Processing Frameworks</b>	<ul style="list-style-type: none"> <li>• The Bane of MapReduce</li> <li>• Tez overview</li> <li>• Pig</li> <li>• Hive</li> <li>• Spark</li> </ul>	<ul style="list-style-type: none"> <li>• Storm</li> <li>• Solr</li> <li>• Solr (cont)</li> </ul> <p><b>Lab Tasks</b></p> <ul style="list-style-type: none"> <li>• Pig</li> </ul>
<b>Module 9: NoSQL Implementations</b>	<ul style="list-style-type: none"> <li>• HBase</li> </ul>	<ul style="list-style-type: none"> <li>• Phoenix</li> </ul>
<b>Module 10: Cluster Management</b>	<ul style="list-style-type: none"> <li>• Ambari Metrics System (AMS)</li> </ul>	<ul style="list-style-type: none"> <li>• Zookeeper</li> </ul>

Learn more at  
[hpe.com/ww/learnbigdata](http://hpe.com/ww/learnbigdata)

**Follow us:**



---

© Copyright 2017 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft is either a registered trademark or trademark of Microsoft Corporation in the United States and/or other countries. The OpenStack Word Mark is either a registered trademark/service mark or trademark/service mark of the OpenStack Foundation, in the United States and other countries and is used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation or the OpenStack community. Pivotal and Cloud Foundry are trademarks and/or registered trademarks of Pivotal Software, Inc. in the United States and/or other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions.

c04577721, November 2017, H6C60S C.02