

Hadoop 101

About This Course

Apache Hadoop, a free, open source, Java-based programming framework. **Why was it invented?**

- Learn about Hadoop's architecture and core components, such as MapReduce and the Hadoop Distributed File System (HDFS).
- Learn how to add and remove nodes from Hadoop clusters, how to check available disk space on each node, and how to modify configuration parameters.
- Learn about other Apache projects that are part of the Hadoop ecosystem, including Pig, Hive, HBase, ZooKeeper, Oozie, Sqoop, Flume, among others. BDU provides separate courses on these other projects, but we recommend you start here.

Course Syllabus

• Module 1 - Introduction to Hadoop

1. Understand what Hadoop is
2. Understand what Big Data is
3. Learn about other open source software related to Hadoop
4. Understand how Big Data solutions can work on the Cloud

• Module 2 - Hadoop Architecture

1. Understand the main Hadoop components
2. Learn how HDFS works
3. List data access patterns for which HDFS is designed
4. Describe how data is stored in an HDFS cluster

• Module 3 - Hadoop Administration

1. Add and remove nodes from a cluster
2. Verify the health of a cluster Start and stop a clusters components
3. Modify Hadoop configuration parameters
4. Setup a rack topology

• Module 4 - Hadoop Components

1. Describe the MapReduce philosophy
2. Explain how Pig and Hive can be used in a Hadoop environment
3. Describe how Flume and Sqoop can be used to move data into Hadoop
4. Describe how Oozie is used to schedule and control Hadoop job execution