

# TRAINING OFFERING | HDP-123

# HORTONWORKS HADOOP ESSENTIALS

1 DAY | GENERAL AUDIENCE



This course provides a technical overview of Apache Hadoop. It includes high-level information about concepts, architecture, operation, and uses of the Hortonworks Data Platform (HDP) and the Hadoop ecosystem. The course provides an optional primer for those who plan to attend a hands-on, instructor-led course.

## PREREQUISITES

No previous Hadoop or programming knowledge is required. Students will need browser access to the Internet.

## TARGET AUDIENCE

Data architects, data integration architects, managers, C-level executives, decision makers, technical infrastructure team, and Hadoop administrators or developers who want to understand the fundamentals of Big Data and the Hadoop ecosystem.

## FORMAT

100% Lecture/Instructor Discussion

## AGENDA SUMMARY

**Day 1:** Hadoop Overview and Demonstrations

## **OBJECTIVES**

- Describe the case for Hadoop
- Identify the Hadoop Ecosystem and architecture
- Data Management – HDFS, YARN
- Data Access – Pig, Hive, HBase, Storm, Solr, Spark
- Data Governance & Integration – Falcon, Flume, Sqoop, Kafka, Atlas
- Security – Kerberos, Falcon, Knox
- Operations – Ambari, Zookeeper, Oozie, Cloudbreak
- Observe popular data transformation and processing engines in action: Apache Hive, Apache Pig, Apache Spark
- Detail the architecture and features of YARN
- Describe backup and recovery options
- Describe how to secure Hadoop
- Explain the fundamentals of parallel processing
- Describe data ingestion options and frameworks for batch and real-time streaming
- Detail the HDFS architecture

## **DEMONSTRATIONS**

- Operational Overview with Ambari
- Ingesting Data into HDFS
- Data Manipulation with Hive
- Risk Factor Analysis with Pig
- Risk Analysis with Spark
- Securing Hive with Ranger

Revised 10/26/2017

### **About Hortonworks**

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

### Contact

For further information visit [www.hortonworks.com](http://www.hortonworks.com) +1 408 675-0983  
+1 855 8-HORTON  
INTL: +44 (0) 20 3826 1405