

# HADOOP DEVELOPER FOUNDATIONS

Course Code: 100682

Learn about the Hadoop ecosystem and how to process large data streams.

Apache Hadoop is a framework for processing Big Data, and Spark is a new in-memory processing engine. This course will introduce you to the Hadoop ecosystem and Spark.

This course explores processing large data streams in the Hadoop ecosystem. Working in a hands-on learning environment, you'll learn techniques and tools for ingesting, transforming, and exporting data to and from the Hadoop ecosystem for processing. You'll also process data using Map/Reduce and other critical tools, including Hive and Pig. Towards the end of the course, we'll review other useful tools such as Oozie and discuss security in the ecosystem.

## What You'll Learn

Join an engaging hands-on learning environment, where you'll explore:

- Introduction to Hadoop
- HDFS
- YARN
- Data Ingestion
- HBase
- Oozie
- Working with Hive
- Hive advanced
- Hive in Cloudera/Hortonworks Distribution (or tools of choice)
- Working with Spark
- Spark Basics
- Spark Shell
- RDDs
- Spark Dataframes and Datasets
- Spark SQL
- Spark API programming
- Spark and Hadoop
- Machine Learning (ML/MLlib)
- GraphX

- Spark Streaming

This course has a 50% hands-on labs to 50% lecture ratio with engaging instruction, demos, group discussions, labs, and project work.

## Who Needs to Attend

Experienced Developers and Architects seeking to be proficient in Hadoop, Hive, and Spark within an enterprise data environment.

## Prerequisites

Before attending this course, you should be:

- Familiar with a programming language
- Comfortable in Linux environment (be able to navigate Linux command line, edit files using vi or nano)



# HADOOP DEVELOPER FOUNDATIONS

Course Code: 100682

PRIVATE GROUP TRAINING

4 Day

Visit us at [www.globalknowledge.com](http://www.globalknowledge.com) or call us at 1-866-716-6688.

Date created: 8/30/2025 4:12:14 PM

Copyright © 2025 Global Knowledge Training LLC. All Rights Reserved.