

# Spark Developer Course:

## Course Overview:

This is an 8 hours instructor lead crash 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 a foundational basics and whole stack overview

## 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.

## Prerequisites

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

## Course Agenda

- Big Data overview
- Big Data Use case
- Hadoop Overview
- HDFS Overview
- HDFS commands
- Yarn Architecture/ Overview
- Core Spark
- SPARK Overview
- Spark RDD
  - The purpose & function of RDD
  - Spark programming basics
  - Spark transformation
  - Spark actions
  - Multiple RDD's
  - Pair RDD
- Spark Streaming
  - Describe Spark Streaming
  - Create and view basic data streams
  - Perform basic transformations on streaming data
  - Utilize window transformations on streaming data

- Spark SQL
  - Spark SQL components
  - DataFrames & tables
  - Creating Data frames
  - Manipulating Dataframes
- Spark Job monitoring
  - Spark Job structure
  - Spark Application UI
  - Spark performance Tuning
  - Broadcast Variables
  - Joining strategies
- Spark Application submission
  - YARN client mode
  - Yarn cluster mode
  - Spark configuration
- Spark Machine learning Overview

#### Course Duration

8 hours (complete Hadoop coverage and demos)

#### Cost

CAD 450 or USD 300 (includes taxes)

The costs include the training, material and cloud based lab costs.

Note: Please enquire with us for ongoing promotions, group and early bird prices.