

# Hadoop Developer Course

100% Assurance to Get HDP Developer certification

## Course Overview

This 40 hours Instructor lead training course is aimed to teach big data application development and analysis in **Apache Hadoop** using Apache ecosystem tools like **Pig, Sqoop, Flume, Hive** and **Spark**. Students will learn the details of Hadoop, YARN, the Hadoop Distributed File System (HDFS), MapReduce, and dive deep into practical lab sessions in Pig programming, Hive programming and Spark to perform data analytics on Big Data. Data Ingestion techniques using **Sqoop** and **Flume**, and workflow definitions using **Oozie** are also covered in this course.

This training course is the best suited for certification aspirants (HDPCD and CCA developer certifications) and preparing Hadoop developers for real-world challenges.

## Prerequisites

Students should be familiar with any programming language or scripting. SQL and basic Unix knowledge is helpful but not necessary. No prior Hadoop knowledge or experience is required.

## Course Outline

<p><b>Big Data Overview</b></p> <p><b>Introduction to Apache Hadoop</b></p> <ul style="list-style-type: none"> <li>Hadoop Overview</li> <li>Hadoop ecosystem projects overview</li> </ul> <p><b>Apache Hadoop file storage</b></p> <ul style="list-style-type: none"> <li>HDFS overview</li> <li>HDFS Architecture</li> </ul> <p><b>Apache Hadoop Data Processing framework</b></p> <ul style="list-style-type: none"> <li>MapReduce Overview</li> <li>MapReduce Architecture</li> <li>YARN overview</li> <li>YARN Architecture</li> <li>Demo on MapReduce Jobs</li> </ul> <p><b>Data Ingestion</b></p> <p><u>HDFS</u></p> <ul style="list-style-type: none"> <li>Understand HDFS commands</li> <li>Move file between HDFS and Local File system</li> </ul> <p><u>Apache Sqoop</u></p> <ul style="list-style-type: none"> <li>Architecture Overview</li> <li>Sqoop programming</li> <li>Sqoop programming and free-form query</li> <li>Import &amp; export RDBMS data using sqoop</li> <li>Demonstration on Sqoop Import and export from RDBMS</li> <li>Exercises - Lab</li> </ul> <p><u>Apache Flume</u></p> <ul style="list-style-type: none"> <li>Architecture Overview</li> <li>Demonstration on HDFS commands on cluster</li> <li>Demonstration on Flume log file capture</li> <li>Exercises - Lab</li> </ul>	<p><b>Data Transformation</b></p> <p><u>Apache Pig</u></p> <ul style="list-style-type: none"> <li>Overview</li> <li>Data types in pig</li> <li>Pig modes</li> <li>Pig programming</li> <li>Pig user defined function(UDF)</li> <li>Pig TEZ MapReduce engine</li> <li>Demonstration on Pig Programming</li> <li>Exercises - Lab</li> </ul> <p><b>Data Analysis</b></p> <p><u>Apache Hive</u></p> <ul style="list-style-type: none"> <li>Hive architecture</li> <li>Data types in Hive</li> <li>Hive programming</li> <li>Partition, bucketing, Joins</li> <li>Hive User defined function (UDF)</li> <li>Demonstration on Hive</li> <li>Exercises – Lab</li> </ul> <p><u>Apache Spark</u></p> <ul style="list-style-type: none"> <li>Spark architecture overview</li> <li>Spark programming</li> <li>Demonstration on Spark programming</li> <li>Exercises - Lab</li> </ul> <p><u>Apache HCatalog overview</u></p> <ul style="list-style-type: none"> <li>Access hive tables from Pig</li> <li>Access Pig scripts from Hive query's</li> <li>Demonstration on HCatalog</li> </ul>
--	---

## Course Duration

40 hours (20 hours theory and concepts; 20 hours practical labs and demos)

This course is spread over 4 weekends (Saturday and Sunday)

## Cost

CAD 2,450 or USD 1,860 (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.

## Registration and Enquiries

For registration and any other enquiries, please contact Big Data School at [info@bigdataschool.ca](mailto:info@bigdataschool.ca) or call us on +1 647 948 9595.

Visit us on [www.Bigdataschool.ca](http://www.Bigdataschool.ca)

## About us

We are a team of independent consultants offering Big Data consulting and transformation solutions. With an average Industry experience of over 15 years and currently working on several cutting-edge technology projects on Big Data, we have extensive experience in Big data projects across several industries such as Information Technology, Telecommunications, Finance, Insurance, Healthcare, Automobile, Energy etc.

We are experienced and certified Big data corporate trainers providing Industry-standard training to several industry leaders.

## Course design philosophy

We want to bring corporate quality and industry standard training to individuals seeking a career in Big Data.

Our courses are modelled based on extensive industry experience and cater to current Industry needs to provide relevant practical experience and real-time working knowledge. Our elite courses cover core concepts in Big data as offered by corporate solution partners-Horton Works, Cloudera and Pivotal.

For those looking to certify, the course has been designed specifically to help take the certification examination with ease. Also, the courses are designed with an ideal theory to practical ratio of 50:50, ensuring learning conceptual knowledge backed by practical applicable skills suited to make you an individual contributor at work from the first day.

The courses are delivered by professional trainers who offer corporate trainings to companies, and are working as consultants and architects on big data projects.