

BIG DATA HADOOP ARCHITECT



simpli|learn

Table of Contents



About
the Course

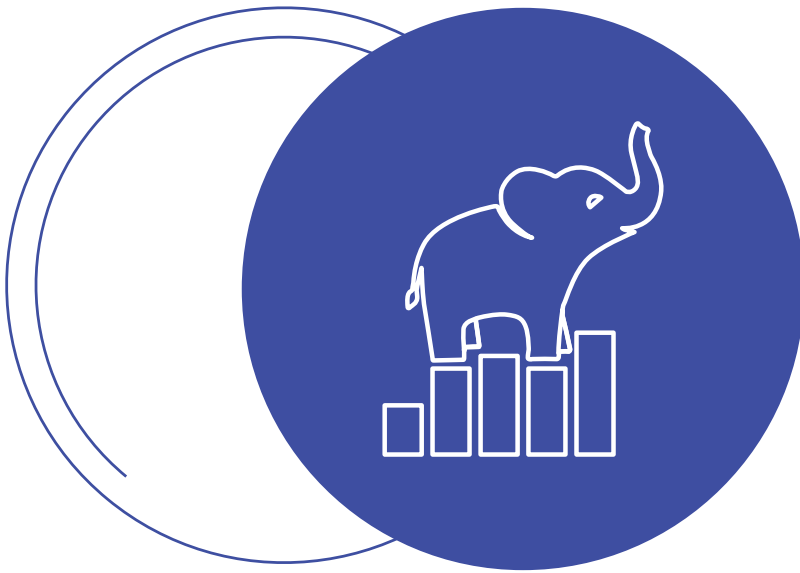


Key Features of Big
Data Hadoop Architect
Masters Program



Learning Path

- 06 • **Step 1** : Begin your big data journey with Apache Hadoop
- 08 • **Step 2** : Harness the power of real time processing with Apache Spark
- 10 • **Step 3** : Store and query big data with NoSQL Databases
- 12 • **Step 4** : Big Data and Hadoop Administrator
- 14 • **Step 5** : Apache Storm
- 16 • **Step 6** : Apache Kafka
- 17 • **Step 7** : Apache Cassandra
- 18 • **Electives**










About the Course

The Big Data Hadoop Architect is the perfect training program for an early entrant to the Big Data world. With a number of required skills required to be a big data specialist and a steep learning curve, this program ensures you get hands on training on the most in-demand big data technologies. The learning is complemented with projects on a cloud based environment, Cloumlabs, for real world experience. The program covers hadoop, spark, NoSQL databases and other hadoop ecosystem components and makes sure you are ready for your next Big Data assignment.

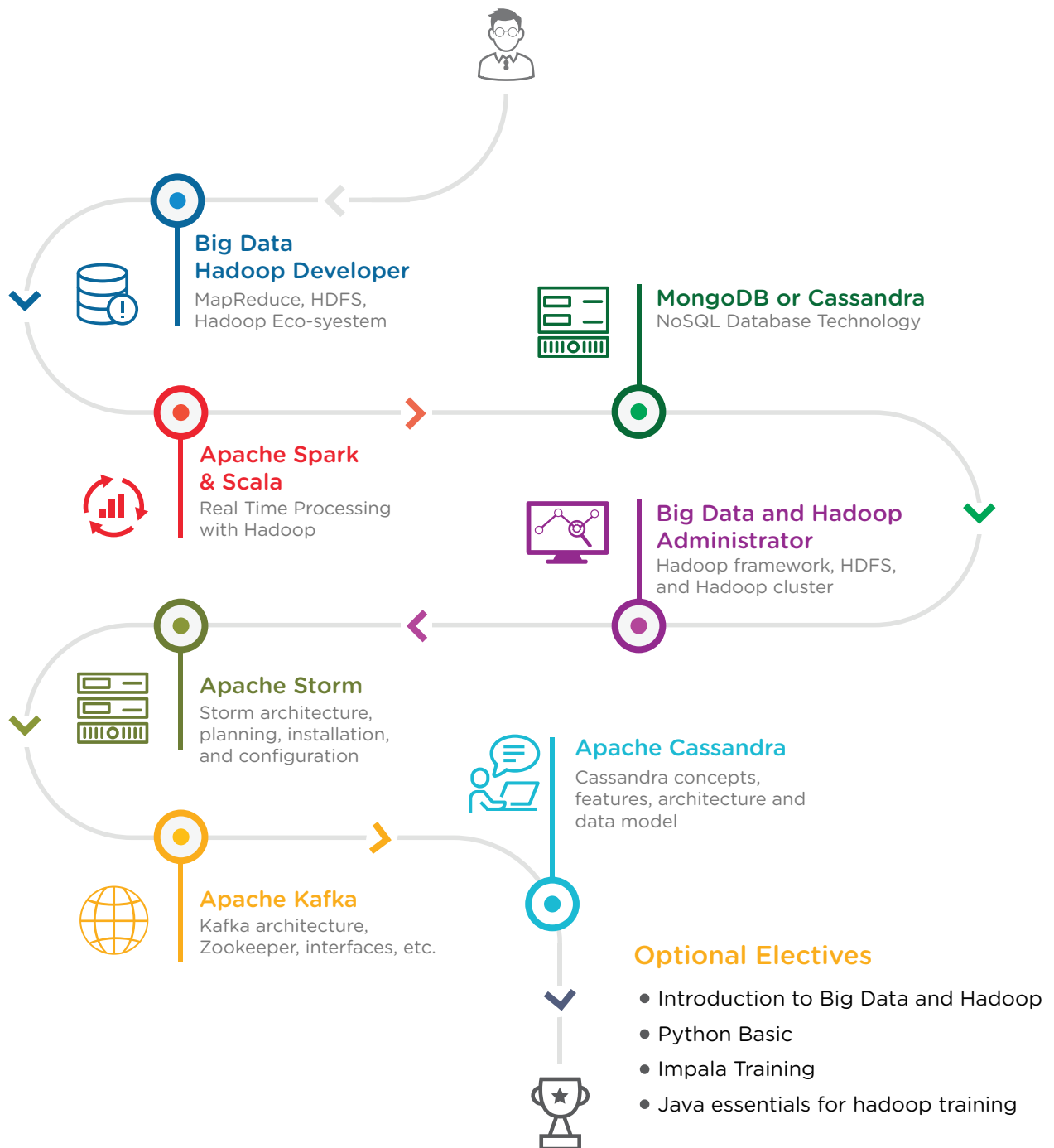
The Big Data Hadoop Architect is the perfect program for an early entrant to upskill and success in the Big Data world.



Key Features

-  Industry Recommended Learning path
-  Access to 100+ hours of Instructor Led training
-  Over 150 hours of high quality E-learning
-  Hands on project execution on CloudLabs
-  Prepares you for the most in demand Cloudera CCA175 Certification
-  Industry recognized Simplilearn Masters Certificate on completion
-  10+ industry projects

Learning Path

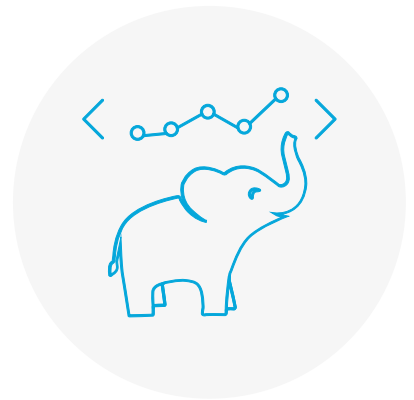


STEP 1 2 3 4 5 6 7

Begin your big data journey

Apache Hadoop

This is a Cloudera aligned deep dive into Hadoop and all its ecosystem components including MapReduce, HDFS, Yarn, HBase, Impala, Sqoop and Flume. It also provides an introduction to Apache Spark which is a next step after Hadoop. After completing this program not only will you be ready to enter the Big Data domain but will also be able to clear the in demand Cloudera CCA175 certification.

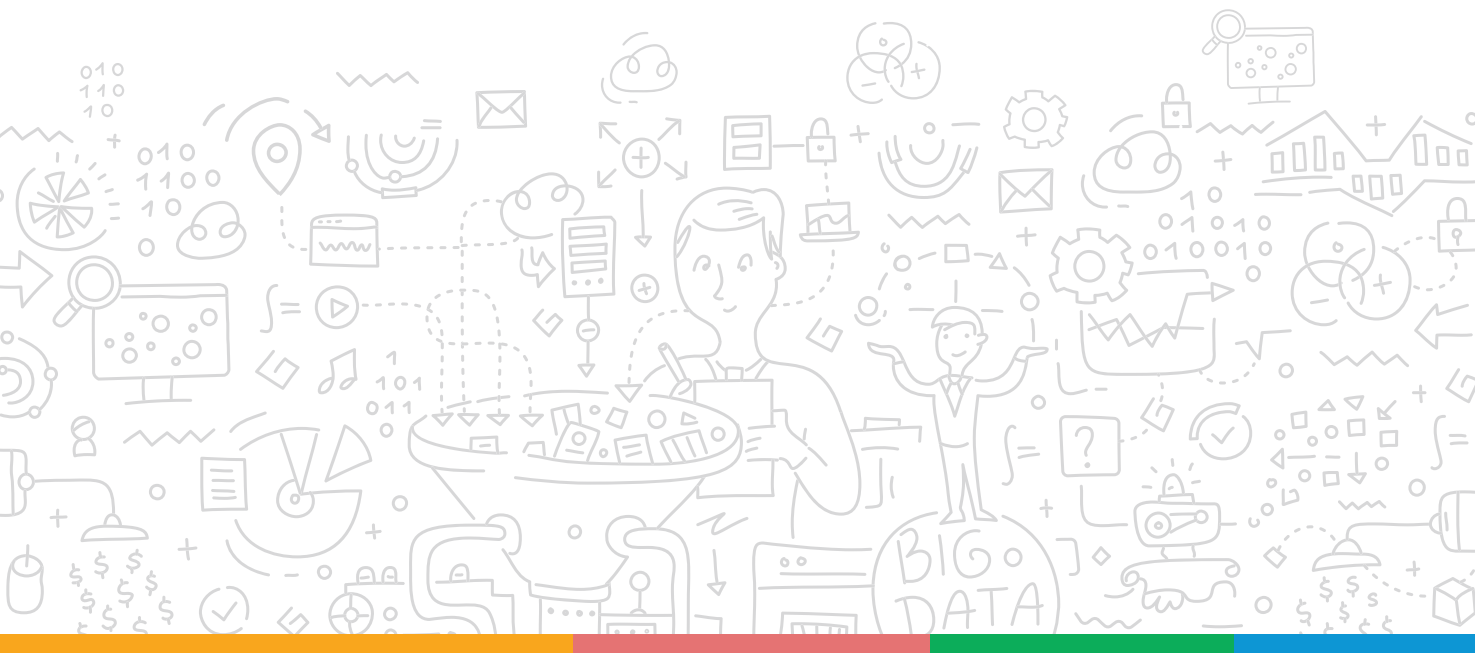


Key Learning Objectives

- ✓ "Master the concepts of the Hadoop framework and its deployment in a cluster environment
- ✓ Understand how the Hadoop ecosystem fits in with the data processing lifecycle
- ✓ Learn to write complex MapReduce programs
- ✓ Describe how to ingest data using Sqoop and Flume
- ✓ Get introduced to Apache Spark and its components
- ✓ List the best practices for data storage
- ✓ Explain how to model structured data as tables with Impala and Hive"

Course Curriculum

- ✓ "Introduction to Bigdata and Hadoop Ecosystem"
- ✓ HDFS and Hadoop Architecture
- ✓ MapReduce and Sqoop
- ✓ Basics of Impala and Hive
- ✓ Working with Impala and Hive
- ✓ Type of Data Formats
- ✓ Advance HIVE concept and Data File Partitioning
- ✓ Apache Flume and HBase
- ✓ Apache Pig
- ✓ Basics of Apache Spark, RDDs in Spark and Applications



STEP 1 2 3 4 5 6 7

Harness the power of real time processing

Apache Spark

The Apache Spark and Scala training is a deep dive into Spark which is a very fast in-memory big data processing engine. Spark is quickly becoming the big-data technology of choice after Hadoop due to its real time applications and streaming ability. This course ensures you are not restricted to Hadoop and gain experience on the most in demand big data technology.



Key Learning Objectives

- ✓ "Get clear understanding of the limitations of MapReduce and role of Spark in overcoming these limitations
- ✓ Understand fundamentals of Scala Programming Language and it's features
- ✓ Explain & master the process of installing Spark as a standalone cluster
- ✓ Expertise in using RDD for creating applications in Spark
- ✓ Mastering SQL queries using SparkSQL
- ✓ Gain thorough understanding of Spark Streaming features
- ✓ Master & describe the features of Spark ML Programming and GraphX Programming"

Course Curriculum

- ✓ "Introduction to Spark
- ✓ Introduction to Programming in Scala
- ✓ Using RDD for Creating Applications in Spark
- ✓ Running SQL queries Using SparkSQL
- ✓ Spark Streaming
- ✓ Spark ML Programming
- ✓ Spark GraphX Programming"

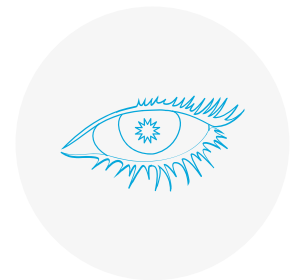
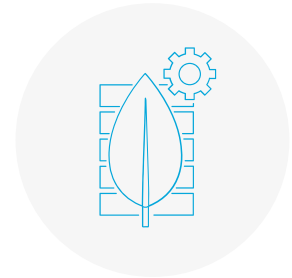


STEP 1 2 3 4 5 6 7

Store and query big data

NoSQL Databases

The third and final step to a big data architect is knowledge of a NoSQL database. These databases work with dynamic schema which increases flexibility and scalability. They have been developed keeping big data uses in mind and are good for handling large amounts of non-uniform data. MongoDB is the most used NoSQL Database and another popular option is Cassandra. Hence this is the final skill required for a big data architect



MongoDB - Key Learning Objectives

- ✓ Develop an expertise in writing Java and Node JS applications using MongoDB
- ✓ Master the skills of Replication and Sharding of data in MongoDB to optimize read / write performance
- ✓ Perform installation, configuration and maintenance of MongoDB environment
- ✓ Get hands-on experience in creating and managing different types of indexes in MongoDB for query execution
- ✓ Develop skillsets in processing huge amounts of data using MongoDB tools
- ✓ Proficiently store unstructured data in MongoDB
- ✓ Gain proficiency in MongoDB configuration, backup methods as well as monitoring and operational strategies
- ✓ Acquire in-depth understanding of managing DB Notes, replica set & Master-Slave concepts

Apache Cassandra - Key Learning Objectives

- ✓ Describe the need for big data and NoSQL
- ✓ Explain the fundamental concepts of Cassandra
- ✓ Describe the architecture of Cassandra
- ✓ Demonstrate data model creation in Cassandra
- ✓ Use Cassandra database interfaces
- ✓ Demonstrate Cassandra database configuration"

Course Curriculum

MongoDB Developer and Administrator

- ✓ NoSQL Database Introduction
- ✓ MongoDB - A Database for the Modern Web
- ✓ CRUD Operations in MongoDB
- ✓ Indexing and Aggregation
- ✓ Replication and Sharding
- ✓ Developing Java and Node JS Application with MongoDB
- ✓ Administration of MongoDB Cluster Operations

Apache Cassandra

- ✓ Overview Big Data and NoSQL Databaases
- ✓ Introduction to Cassandra
- ✓ Cassandra Architecture
- ✓ Cassandra Installation and Configuration
- ✓ Cassandra Data Model
- ✓ Cassandra Interfaces
- ✓ Cassandra Advanced Architecture
- ✓ Apache Ecosystem around Cassandra"

STEP

1

2

3

4

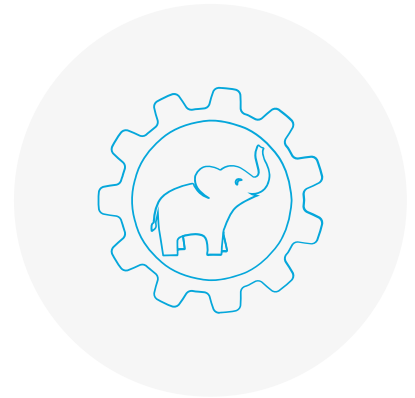
5

6

7

Big Data and Hadoop Administrator

This Big Data and Hadoop Administrator training course will furnish you with the aptitudes and methodologies necessary to excel in the Big Data Analytics industry. With this Hadoop Admin training, you'll learn to work with the adaptable, versatile frameworks based on the Apache Hadoop ecosystem, including Hadoop installation and configuration; cluster management with Sqoop, Flume, Pig, Hive and Impala and Cloudera; and Big Data implementations that have exceptional security, speed and scale.



Key Learning Objectives

- ✓ Understand the fundamentals and characteristics of Big Data and various scalability options available to help organizations manage Big Data
- ✓ Master the concepts of the Hadoop framework, including architecture, the Hadoop distributed file system and deployment of Hadoop clusters using core or vendor specific distributions
- ✓ Use Cloudera manager for setup, deployment, maintenance and monitoring of Hadoop clusters
- ✓ Understand Hadoop Administration activities and computational frameworks for processing Big Data
- ✓ Work with Hadoop clients, nodes for clients and web interfaces like HUE to work with Hadoop Cluster

Key Learning Objectives

- ✓ Use cluster planning and tools for data ingestion into Hadoop clusters, and cluster monitoring activities
- ✓ Utilize Hadoop components within Hadoop ecosystem like Hive, HBase, Spark and Kafka
- ✓ Understand security implementation to secure data and clusters.

Course Curriculum

- ✓ Course Introduction
- ✓ Big Data and Hadoop - Introduction
- ✓ HDFS Hadoop Distributed File System
- ✓ Hadoop Cluster Setup and Working
- ✓ Hadoop Configurations and Daemon Logs
- ✓ Hadoop Cluster Maintenance and Administration
- ✓ Hadoop Computational Frameworks
- ✓ Scheduling: Managing Resources
- ✓ Hadoop Cluster Planning
- ✓ Hadoop Clients and Hue Interface
- ✓ Data Ingestion in Hadoop Cluster
- ✓ Hadoop Ecosystem ComponentsServices
- ✓ Hadoop Security
- ✓ Hadoop Cluster Monitoring

STEP

1

2

3

4

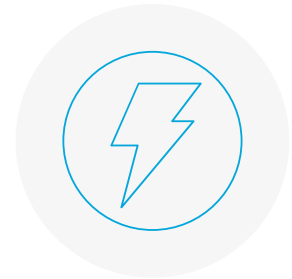
5

6

7

Apache Storm

Apache Storm Certification Training from Simplilearn enables you to can handle stream processing with big data technology of Apache Storm. This is one of the most popular stream processing engine beside Spark Streaming.



Key Learning Objectives

- ✓ Get clear understanding of the limitations of MapReduce and role of Spark in overcoming these limitations
- ✓ Understand fundamentals of Scala Programming Language and it's features
- ✓ Explain & master the process of installing Spark as a standalone cluster
- ✓ Expertise in using RDD for creating applications in Spark
- ✓ Mastering SQL queries using SparkSQL
- ✓ Gain thorough understanding of Spark Streaming features
- ✓ Master & describe the features of Spark ML Programming and GraphX Programming

Course Curriculum

- ✓ Introduction to Spark
- ✓ Introduction to Programming in Scala
- ✓ Using RDD for Creating Applications in Spark
- ✓ Running SQL Queries Using Spark SQL
- ✓ Spark Streaming
- ✓ Spark Structured Streaming
- ✓ Spark ML Programming
- ✓ Graph Processing using GraphX and GraphFrames

STEP

1

2

3

4

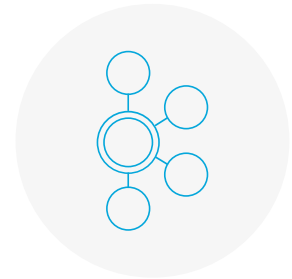
5

6

7

Apache Kafka

This course introduces you to Apache Kafka which is an important high-performance real-time messaging system that can process millions of messages per second. It provides a distributed and partitioned messaging system and is highly fault tolerant.



Key Learning Objectives

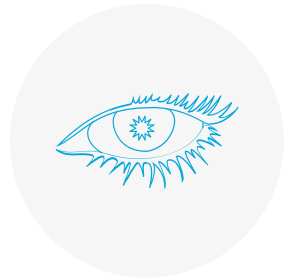
- ✓ Describe the importance of big data
- ✓ Describe the fundamental concepts of Kafka
- ✓ Describe the architecture of Kafka
- ✓ Explain how to install and configure Kafka
- ✓ Explain how to use Kafka for real-time messaging

Course Curriculum

- ✓ Course introduction
- ✓ Big Data Overview
- ✓ Introduction to Zookeeper
- ✓ Introduction to Kafka
- ✓ Installation and Configuration
- ✓ Kafka Interfaces

Apache Cassandra

This Apache Cassandra Certification Training will develop your expertise in working with the high-volume Cassandra database management system as part of the Big Data Hadoop framework. With this Cassandra training, you will learn Cassandra concepts, features, architecture and data model, and how to install, configure and monitor open-source databases. The Casandra course is ideal for software developers and analytics professionals who wish to further their careers in the Big Data field.



Key Learning Objectives

- ✓ Describe the need for Big Data and NoSQL
- ✓ Explain the fundamental concepts of Cassandra and its architecture
- ✓ Describe the architecture of Cassandra
- ✓ Demonstrate data model creation in Cassandra
- ✓ Use Cassandra database interfaces
- ✓ Demonstrate Cassandra database configuration

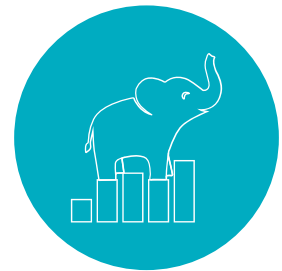
Course Curriculum

- ✓ Course Overview
- ✓ Introduction to big data and No-SQL Databases
- ✓ Introduction to Cassandra
- ✓ Architecture of Cassandra
- ✓ Installation and Configuration of Cassandra
- ✓ Cassandra Data Model
- ✓ Cassandra Interfaces
- ✓ Advanced Architecture and Cluster Management
- ✓ Hadoop Ecosystem around Cassandra

Other Electives:

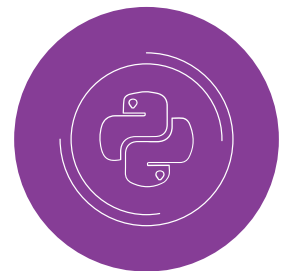
Introduction to Big Data and Hadoop

This is an ideal course for individuals who want to understand the basic concepts of Big Data and Hadoop. On completing this course, learners will be able to interpret what goes behind the processing of huge volumes of data as the industry switches over from excel-based analytics to real-time analytics.



Python Basics

This course is ideal for you to understand the basics of Python Programming Language.



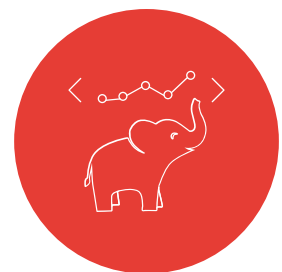
Impala Training

The course focuses on the basics of Impala. It further provides an overview of the superior performance of Impala, against other popular SQL-on-Hadoop systems. On completing this course, you will be able to interpret the role of Impala in the Big Data Ecosystem.



Java Essentials for Hadoop Training

This course will help you comprehend the Java concepts that is required for our Big Data Hadoop & Spark Developer course.



Advisory Board Members:



Named by Analytica as one of the 3 most influential people in Big Data, Ronald is an author for a number of leading Big Data & Data Science websites, including Dataflog, Data Science Central, and The Guardian. He is also a renowned speaker at industry events.

Ronald Van Loon

[Big Data Expert, Director Adversitement](#)



Sina has over 10 years of experience in Big Data. He's worked as a Platinum Level Big Data trainer and as a Big Data Architect at Bell Labs. He is passionate about building a Big Data education ecosystem, and has written a number of magazine articles and journal publications about the field.

Sina Jamshidi

[Big Data Lead at Bell Labs](#)





simplilearn

USA

Simplilearn Americas, Inc.

201 Spear Street, Suite 1100, San Francisco, CA 94105

United States

Phone No: +1-844-532-7688

INDIA

Simplilearn Solutions Pvt Ltd.

53/1 C, Manoj Arcade, 24th Main, Harlunkunte

2nd Sector, HSR Layout

Bangalore - 560102

Call us at: 1800-102-9602

www.simplilearn.com