



Hadoop Online Training in Cloud Lab



WhatsApp: [+91-7530088009](https://wa.me/917530088009)

Call: India [+91-444-631-1234](https://tel.us/+914446311234)

USA [+1-650-265-2492](https://tel.us/+16502652492)

Email: training@sparkdatabox.com

Web: <https://sparkdatabox.com>

Spark Databox is known to be a pre-eminent platform for software certificate training and career development. Through our mission, we are a pioneer of positive change, improve productivity, increase the workforce, and creating a career opportunity for everyone. We are India's #1 software training institute. Apart from receiving excellent live training you will also receive free self-paced video courses, training materials, placement support, mock interviews and many more.

Program Key Features

- ❖ 100% Practical training
- ❖ Experienced Trainers
- ❖ 100% Placement assistance
- ❖ Small batch size
- ❖ Customized training content
- ❖ Real-time project training
- ❖ Fully equipped cloud lab
- ❖ 100% Customer support
- ❖ 100% Money back guarantee

About Course:

The **best Hadoop online course** is here. Our expert trainers at Spark Databox are here ready to guide you and launch your career. Hadoop online Training is intended to give you a certified Hadoop profession by equipping you productive hands-on training on Hadoop Ecosystem. Hadoop online training certification training is an excellent way to your Big Data drive, and you will perceive the possibility to work on multiple Hadoop projects. Hadoop Online Training will help you learn and understand Map Reduce Concepts explained by our industry experts.

This training will also help you learn on handling the Apache Hadoop ecosystems like Hive, Pig, and HBase with real-time Project training. This course gives you a clear understanding of how to develop a Hadoop application using the appropriate frameworks for the suitable Apache Project ecosystem in a real-time situation. This practice will help you to develop your own custom application. We designed this course in order to adhere to the current industry standards. We ensure that the candidates who are not familiar with enough technical knowledge can also learn & shine in this Hadoop training.

Contents

Section 1: Introduction to Big Data and Hadoop:

- Introduction to Big data
- Challenges in processing Big data
- Technologies that support Big data
- What is Hadoop?
- Why Hadoop?
- When to use Hadoop?
- Hadoop vs RDBMS
- Hadoop requirements

Section 2: HDFS - Hadoop Distributed File System:

- HDFS - Introduction
- HDFS features

Section 3: Hadoop Eco system

- Pig - Introduction
- Hive - Introduction
- HBase - Introduction
- Scoop
- Other eco systems

Section 4: Hadoop Development

- Loading Data into Hadoop
- Deleting Data from Hadoop
- Mapper Class
- Reducer Class
- Driver Class
- Basic program using MapReduce
- MapReduce internals
- MapReduce - HBase
- Hive - Introduction
- Working with Pig
- Working with Scoop
- RDBMS to Hadoop
- RDBMS to Hive
- RDBMS to Hbase
- Webserver to Hadoop
- What is Flume?
- working with Apache log viewer
- Market Basket Algorithms

Section 5: Introduction to Hive

- Introducing Hadoop Hive
- Detailed architecture of Hive
- Comparing Hive with Pig and RDBMS
- Working with Hive Query Language
- Creation of database, table, Group by and other clauses
- Various types of Hive tables, Hcatalog
- Storing the Hive Results
- Hive partitioning and Buckets
- Static partitioning
- Dynamic partitioning
- Alter Partitioned Table and MSCK Repair command (Advance)

- What is Bucketing?
- Create Bucketed Table
- Tablesampling (Advance)
- No_drop, Offline command (Advance)

Section 6: Advanced Hive and Impala

- Indexing in Hive
- The Map Side Join in Hive
- Working with complex data types
- The Hive User-defined Functions
- Introduction to Impala
- Comparing Hive with Impala
- The detailed architecture of Impala

Section 7: Working with Pig

- Apache Pig introduction and its various features
- Various data types and schema in Hive
- The available functions in Pig, Hive Bags, Tuples and Fields

Section 8: Flume, Sqoop and HBase

- Apache Sqoop introduction, overview
- Importing and exporting data
- Performance improvement with Sqoop and Sqoop limitations
- Architecture of Flume, HBase and CAP theorem
- Using Scala for writing Apache Spark applications
- Detailed study of Scala and the need for Scala
- The concept of object oriented programming and Executing the Scala code
- Programming and anonymous functions
- Bobsrockets package and comparing the mutable and immutable collections
- Scala REPL and Lazy Values
- Control Structures in Scala
- Directed Acyclic Graph (DAG)
- First Spark application using SBT/Eclipse
- Spark Web UI and Spark in Hadoop ecosystem

Section 9: Spark framework

- Detailed Apache Spark and its various features
- Comparing with Hadoop

- Various Spark components
- Combining HDFS with Spark and Scalding
- Introduction to Scala and importance of Scala and RDD

Section 10: RDD in Spark

- Understanding the Spark RDD operations and Comparison of Spark with MapReduce
- What is a Spark transformation
- Loading data in Spark
- Types of RDD operations viz. transformation and action and What is a Key/Value pair
- The detailed Spark SQL
- The significance of SQL in Spark for working with structured data processing
- Spark SQL JSON support
- Working with XML data and parquet files
- Creating Hive Context
- Writing Data Frame to Hive
- How to read a JDBC file
- Significance of a Spark Data Frame
- How to create a Data Frame
- What is schema manual inferring
- How to work with CSV files
- JDBC table reading
- Data conversion from Data Frame to JDBC
- Spark SQL user-defined functions
- Shared variable and accumulators
- How to query and transform data in Data Frames
- Data Frame execution engine

Section 11: Machine Learning Using Spark (MLlib)

- Introduction to Spark MLlib
- Understanding various algorithms
- What is Spark iterative algorithm
- Spark graph processing analysis
- Introducing Machine Learning
- K-Means clustering
- Spark variables like shared and broadcast variables and what are accumulators
- Various ML algorithms supported by MLlib
- Linear Regression and Logistic Regression
- Decision Tree and Random Forest
- K-means clustering techniques, building a Recommendation Engine

Section 12: Real-time project training

- Hadoop project environment setup
- Real-time Hadoop project
- Project demonstration
- Expert evaluation and feedback

You made it!!

Post completion of Hadoop Online Course, a proper orientation for placements is done. With this training from experienced trainers, as professionals, you will be equipped with different proficiencies. This is a chance to open up and widen your prospects.

- Spark Databox Hadoop course certification
- Interview preparation
- Mock interviews
- Resume preparation
- Knowledge sharing with industry experts
- Counseling to guide you to a right path in Hadoop development career



www.sparkdatabox.com

© 2009-2019 - Spark Databox. All rights reserved.