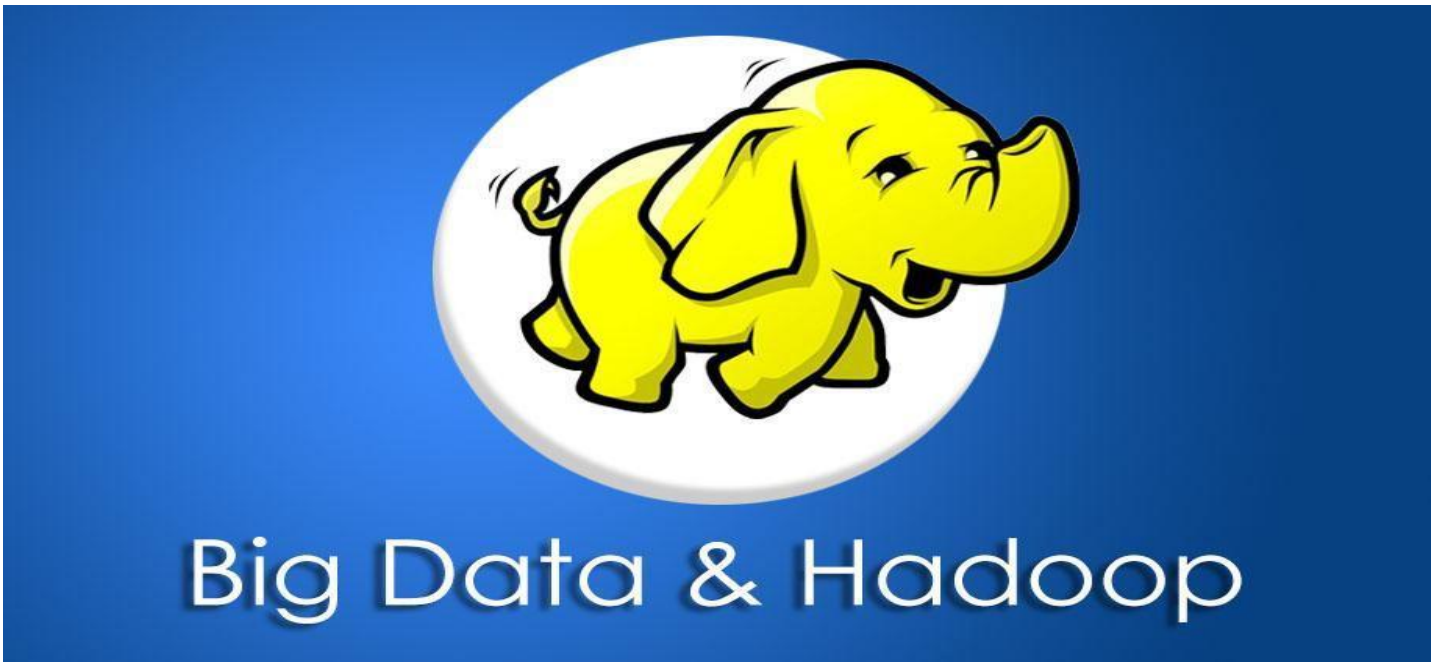
LEARN, THINK, INNOVATE



Overview

Let eMexo Technologies **Best Big Data Hadoop Training in Electronic City Bangalore** take you from the fundamentals of Hadoop to Advance Hadoop and make you an expert in developing real-time Hadoop applications. Here are the major topics we cover under this Hadoop course Syllabus, **Apache Hadoop, MapReduce Framework, Apache Hive, Apache Pig, Apache HBase, Apache Sqoop, Apache Flume, Apache HUE, Apache Zookeeper, and Administration concepts**. Each topic will be covered in a practical way with examples.

All the topics will be covered with Practical and hands-on training. Our trainers have industry experience with live project experience in cutting-edge technologies which they teach. We hire only the Best Hadoop industry specialists as trainers for our **Big Data Hadoop Training Course in Electronic City Bangalore**.

If you are looking for **Big Data Hadoop Certification Training in Electronic City Bangalore**, eMexo Technologies is the **Best Big Data Hadoop Training Institute in Electronic City Bangalore**. Come over to our training institute for a free demo class. Let our trainer give you a demo on Hadoop and only then you take the decision to enroll in the training program.

Training Features

Real-life Case Studies

Do a real-life case study to understand the usage in real-world scenarios.

Assignments

Each class will be followed by a practical assignment switch that can be completed before the next class.

Preparation for interview

Our trainers are professionals working in multinational corporations. They are experts in their field and know exactly what the interviewer will look for in the candidate. Experienced trainers not only share interview questions but also conduct mock interviews to help prepare for the actual interview.

Key Features

eMexo Technologies offers the **Best Big Data Hadoop Training Course in Electronic City Bangalore** with the TOP industry expert trainers.

Here are the key features.

★ Free Demo Class Available

★ Practical Approach

★ Expert & Certified Trainers

★ 100% Job Oriented Training

★ Real World use cases and Scenarios

★ Completed 500+ Batches

★ Certification Guidance

Unit 1: Apache Hadoop

➢ Introduction to Big Data & Hadoop Fundamentals
➢ Dimensions of Big Data
➢ Type of Data generation
➢ Apache ecosystem & its projects
➢ Hadoop distributors
➢ HDFS core concepts
➢ Modes of Hadoop Employment
➢ HDFS Flow architecture
➢ HDFS MrV1 vs. MrV2 architecture
➢ Types of Data compression techniques
➢ Rack topology
➢ HDFS utility commands
➢ Min h/w requirements for cluster & property files changes

Unit 2: MapReduce Framework

Goal: In this module, you will understand the Hadoop MapReduce framework and the working of MapReduce on data stored in HDFS. You will understand concepts like Input Splits in MapReduce, Combiner & Partitioner, and Demos on MapReduce using different data sets.

Objectives – Upon completing this module, you should be able to understand MapReduce involves processing jobs using the batch processing technique.

➢ MapReduce can be done using Java programming.
➢ Hadoop provides with Hadoop-examples jar file which is normally used by administrators and programmers to perform testing of the MapReduce applications.
➢ MapReduce contains steps like splitting, mapping, combining, reducing, and output.

Topics:

➢ Introduction to MapReduce
➢ MapReduce Design flow
➢ MapReduce Program (Job) execution
➢ Types of Input Formats & Output Formats
➢ MapReduce Datatypes
➢ Performance tuning of MapReduce jobs
➢ Counters techniques
➢ Hands-on

Unit 3: Apache Hive

Goal: This module will help you in understanding Hive concepts, Hive Data types, Loading and Querying Data in Hive, running Hive scripts, and Hive UDF.

Objectives – Upon completing this module, you should be able to understand Hive is a system for managing and querying unstructured data into a structured format.

- ➢ The various components of Hive architecture are meta store, driver, execution engine, and so on.
- ➢ Metastore is a component that stores the system catalog and metadata about tables, columns, partitions, and so on.
- ➢ Hive installation starts with locating the latest version of the tar file and downloading it to the Ubuntu system using the wget command.
- ➢ While programming in Hive, use the show tables command to display the total number of tables.

Topics:

- ➢ Introduction to Hive & features
- ➢ Hive architecture flow
- ➢ Types of hive tables flow
- ➢ DML/DDL commands explanation
- ➢ Partitioning logic
- ➢ Bucketing logic
- ➢ Hive script execution in shell & HUE
- ➢ Hands-on

Unit 4: Apache Pig

Goal: In this module, you will learn Pig, the types of use cases we can use Pig, tight coupling between Pig and MapReduce, Pig Latin scripting, PIG running modes, PIG UDF, Pig Streaming, and Testing PIG Scripts. Demo on healthcare dataset.

Objectives – Upon completing this module, you should be able to understand Pig is a high-level data flow scripting language and has two major components: Runtime engine and Pig Latin language.

- ➢ Pig runs in two execution modes: Local mode and MapReduce mode. Pig script can be written in two modes: Interactive mode and Batch mode.
- ➢ The pig engine can be installed by downloading the mirror web link from the website: pig.apache.org.

Topics:

- ➢ Introduction to Pig Concepts
- ➢ Pig modes of execution/storage concepts
- ➢ Pig program logics explanation
- ➢ Pig basic commands
- ➢ Pig script execution in shell/HUE

- Hands-on

Unit 5: Apache HBase

Goal: This module will cover Advanced HBase concepts. We will see demos on Bulk Loading, Filters. You will also learn what Zookeeper is all about, how it helps in monitoring a cluster, and why HBase uses Zookeeper.

Objectives – Upon completing this module, you should be able to understand HBaseha's two types of Nodes—Master and RegionServer. Only one Master node runs at a time. But there can be multiple RegionServersat a time.

- The data model of Hbasecomprises tables that are sorted by rows. The column families should be defined at the time of table creation.
- There are eight steps that should be followed for the installation of HBase.
- Some of the commands related to HBaseshell are create, drop, list, count, get, and scan.

Topics:

- Introduction to Hbase concepts
- Introduction to NoSQL/CAP theorem concepts
- Hbase design/architecture flow
- Hbase table commands
- Hive + Hbase integration module/jars deployment
- Hbase execution in shell/HUE
- Hands-on

Unit 6: Apache Sqoop

Goal: Sqoop is an Apache Hadoop Eco-system project whose responsibility is to import or export operations across relational databases. Some reasons to use Sqoop are as follows:

- SQL servers are deployed worldwide
- Nightly processing is done on SQL servers
- Allows to move a certain part of data from traditional SQL DB to Hadoop
- Transferring data using the script is inefficient and time-consuming
- To handle large data through the Ecosystem
- To bring processed data from Hadoop to the applications

Objectives – Upon completing this module, you should be able to understand Sqoop is a tool designed to transfer data between Hadoop and RDBs including MySQL, MS SQL, Postgre SQL, MongoDB, etc.

- Sqoop allows the import of data from an RDB, such as SQL, MySQL, or Oracle into HDFS

Topics:

- Introduction to Sqoop concepts
- Sqoop internal design/architecture
- Sqoop Import statements concepts
- Sqoop Export Statements concepts
- Quest Data connectors flow
- Incremental updating concepts

- ➢ Creating a database in MySQL for importing to HDFS
- ➢ Sqoop commands execution in shell/HUE
- ➢ Hands-on

Unit 7: Apache Flume

Goal: Apache Flume is a distributed data collection service that gets the flow of data from their source and aggregates them to where they need to be processed.

Objectives – Upon completing this module, you should be able to understand Apache Flume is a distributed data collection service that gets the flow of data from its source and aggregates the data to sink.

- ➢ Flume provides a reliable and scalable agent mode to ingest data into HDFS

Topics:

- ➢ Introduction to Flume & features
- ➢ Flume topology & core concepts
- ➢ Property file parameters logic
- ➢ Hands-on

Unit 8: Apache HUE

Goal: Hue is a web front end offered by the ClouderaVM to Apache Hadoop.

Objectives – Upon completing this module, you should be able to understand how to use hue for Hive, Pig, and Oozie.

Topics:

- ➢ Introduction to Hue Design
- ➢ Hue architecture flow/UI interface
- ➢ Hands-on

Unit 9: Apache Zookeeper

Goal: Following are the goals of ZooKeeper:

- ➢ Serialization ensures avoidance of delay in reading or writing operations.
- ➢ Reliability persists when an update is applied by a user in the cluster.
- ➢ Atomicity does not allow partial results. Any user update can either succeed or fail.
- ➢ A simple Application Programming Interface or API provides an interface for development and implementation.

Objectives – Upon completing this module, you should be able to understand ZooKeeper provides a simple and high-performance kernel for building more complex clients.

- ➢ ZooKeeper has three basic entities—Leader, Follower, and Observer.
- ➢ The watch is used to get the notification of all followers and observers to the leaders.

Topics:

- ➢ Introduction to zookeeper concepts
- ➢ Zookeeper principles & usage in the Hadoop Framework
- ➢ Basics of Zookeeper
- ➢ Hands-on

Unit 10: Administration concepts

Goal: Explain different configurations of the Hadoop cluster

- ➢ Identify different parameters for performance monitoring and performance tuning
- ➢ Explain the configuration of security parameters in Hadoop.

Objectives – Upon completing this module, you should be able to understand Hadoop can be optimized based on the infrastructure and available resources.

- ➢ Hadoop is an open-source application and the support provided for complicated optimization is less.
- ➢ Optimization is performed through XML files.
- ➢ Logs are the best medium through which an administrator can understand a problem and troubleshoot it accordingly.
- ➢ Hadoop relies on the Kerberos-based security mechanism.

Topics:

- ➢ Principles of Hadoop Administration & its Importance
- ➢ Hadoop admin commands explanation
- ➢ Balancer concepts
- ➢ Rolling upgrade mechanism explanation
- ➢ Hands-on

FAQs

1. How is the training organized? How much percentage is theoretical and how much is practical hands-on?

We at eMexo believe nothing beats hands-on practice when it comes to learning a concept. Our teaching methodology is 100% practical and hands-on oriented. You learn a concept, you practice it then and there with the trainers. We also give you assignments for each topic which you can practice at home and any doubts regarding the topic can be cleared with the trainer the next day.

2. What is the course duration? How and when do you plan to complete the course?

We generally cover our courses in 60 hours, however, we know that we can't put a hard- stop to learning with a number. Our trainer will make sure that you have learned everything that is part of the curriculum. This could mean 48 hours or 60 hours, it doesn't matter.

3. What is the material provided in the training?

We have industry standard course material which is used by our trainers to train you. At the end of the training, apart from the notes which you have taken during the course, we will also provide you with the training material which was used. This training material includes the training content, interview questions, etc.

4. Do you help in preparing for the interview?

Our trainers are working professionals who work in MNCs. They are the experts in their domain and they know exactly what an interviewer looks into a candidate. Our expert trainers apart from sharing the interview questions will also conduct mock interviews to help you prepare for the real interview.

5. Who are your trainers?

Our trainers are industry experts who work in their respective technologies day in and day out. They work in MNCs and are technology experts within their organizations.

6. **What is the total batch size per course?**

We maintain a strict batch size of a maximum of 5 students. We also provide exclusive one-to-one training as well. Talk to our training partner to get more details.

7. Do you provide certification for the course?

Yes, at the end of the training, we provide a certification of completion.

8. Will I be joining a new batch or being merged with another batch?

You will be added to a new batch.

9. Is fast-track training available?

Yes, we also provide fast-track training for those who want to complete the course faster. The curriculum and the total hours required to complete the course will remain the same. However, the trainer will be spending more hours with you to complete the course.

10. Do you assist in job placement?

Our trainers are expert professionals in their organizations and they often act as the interviewer to hire new candidates. Our trainers will help you prepare your resume with industry standards. After all, they know exactly what to look for in a resume.

11. Timings for training - Regular training/weekend training?

We provide both regular and weekend training. Talk to our training partner to learn more about the timings.

12. Will you be working on a live project during training?

Yes, apart from doing the hands-on practice our trainer will also be taking a real-world project and working with you on the implementation.

13. What happens if I miss a class?

If you miss a class the content of that class will be taught to you again. With us, you might miss a class but not the content.

14. Can I attend a demo before the actual class?

Yes, absolutely! Talk to our training counselor on the phone at **+91-9513216462** or email us at **info@emexotechnologies.com** to arrange a free demo. You can also fill in the contact us form below and we will call you to discuss your training requirements.