



Shaping Talents



Shaping Future

JOIN US

Become a Certified
Technology Expert



**BIG DATA
HADOOP DEVELOPER
TRAINING**

Big Data Hadoop Developer Training

About EnhanceLearn

EnhanceLearn offers you a complete career transition by providing training and placement programs for Students and Jobseekers looking for a Career Success. We provide best IT training and certification courses which is taken by professional certified experts. The training modules are designed as per the market requirement so that it helps student to conquer the Career Job Market and achieve their career goals with our placement assistance.



About the Big Data Hadoop Developer Training Course

Hadoop Big Data Developer Training course is a comprehensive training program designed by industry experts considering current industry job needs to assist you learn big data Hadoop and Spark modules. This is an industry recognized big data certification training course having a tremendous capability to store and process large clusters of data, it disclosed opportunities to business around the world with AI. Our expert trainers will guide you in training process and assist you in your all learning process.

Why Take Big Data Hadoop Developer Training Course?

There has been an exponential growth in the world of data since the last decade. Therefore, industries are looking for ways to handle data and get business. Here is Hadoop – jail-break – for the IT firms in order to store and retrieve large amount of data. Here are some reasons to choose Hadoop.

- A combination of online running applications on a huge-scale built of commodity hardware.
- Big companies are seeking for Hadoop professionals capable of handling data.
- Less Big Data skilled Professionals in the Market, so more job opportunities.
- Highly flexible: No pre-processing of data is required before storing the data. A large or as much amount of data you want can be stored.
- Compatible: As Hadoop is a Java-based programming, it works on all platforms.
- Streaming Job Opportunities with Big Data - Relevant jobs: After the successful completion of this course, Hadoop architect, Hadoop developer, and Data scientist are the jobs waiting for you next door.
- Great computing power: Since Hadoop has a distributed computing environment, with more computing nodes, the processing power increase.

If you are interested in joining EnhanceLearn's best Training and Placement Program team, please reach our team here:

Big Data Hadoop Developer Training

Course Content:

Module 1: Course Introduction

- Big Data introduction
- Why is it required?
- Facts and evolution
- Objectives
- Market trends
- Key features

Module 2: Introduction to Big Data

- Rise of Big Data
- Hadoop vs. traditional systems
- Hadoop master-slave introduction and architecture
- Objectives
- Types of data
- Data explosion
- Sources
- Characteristics
- Knowledge check
- Traditional IT Analytics approach
- Capabilities of Big Data technology
- Discovery and exploration of Big Data technology platform
- Handling limitations

Module 3: Hadoop Architecture

- Introduction to Hadoop
- Architecture
- History and milestones
- Objectives
- Key features
- Hadoop cluster
- Core services
- Role of Hadoop in Big Data
- Advanced Hadoop core components
- HDFS introduction
- Why HDFS
- Architecture of HDFS
- VMware player
- Real-life concept of HDFS
- Characteristics HDFS
- File system namespace
- Data block split
- Advantages of data block approach

- Replication method
- Data replication topology
- Data replication representation
- HDFS access
- Business scenario
- Error handling

Module 4: Hadoop Configuration

- Configuration
- Configuration files
- Cluster configuration
- Hadoop modes
- Terminal commands
- MapReduce in action
- Reporting
- Recovery

Module 5: Hadoop Cluster Configuration

- Overview
- Important files
- Parameters and values
- Environment setup
- “Include” and “Exclude” configuration files

Module 6: Hadoop Deployment

- Introduction
- Objectives
- Ubuntu server
 - Introduction
 - Installation
 - Business scenario
- Hadoop installation prerequisites
- Installation steps
- Hadoop multi-node installation
 - Single-node cluster
 - Multi-node cluster
- Clustering of Hadoop environment

Module 7: Introduction to MapReduce

- Introduction
- Objectives
- Why MapReduce
- Characteristics
- Analogy
- Examples
- Map execution

If you are interested in joining EnhanceLearn's best Training and Placement Program team, please reach our team here:

Big Data Hadoop Developer Training

- Map execution distributed two-node environment
- Essentials
- Jobs and associated tasks
- Business scenario
- Setup environment
- Small Data and Big Data
- Programs
- Requirements
- Steps of Hadoop MapReduce
- Responsibilities of MapReduce
- Java programming of MapReduce in Eclipse

Module 8: Deep Dive in MapReduce and Yarn

- Yarn introduction
- Objectives of Yarn
- Real-life concept of Yarn
- Application master
- Container
- Joining data-sets in MapReduce
- Infrastructure of Yarn
- Resource manager in Yarn
- Application running on Yarn
- Application start-up in Yarn
- Role of AppMaster in application start-up

Module 9: Advanced HDFS and MapReduce

- Hadoop components
- Advanced HDFS introduction
- Advanced MapReduce introduction
- Objectives
- Business scenario
- Interfaces
- Data types in Hadoop
- Input and output formats in Hadoop
- Distributed cache
- Joins in MapReduce
 - Reduce join
 - Composite join
 - Replicated join
- Errors

Module 10: Understanding the MapReduce Framework

- Overview of the framework
- Use cases of MapReduce
- Anatomy of MapReduce framework
- Mapper class
- Driver code

- Understanding partitioner and combiner

Module 11: Hadoop Administration and Maintenance

- Hardware considerations
- Potential problems and solutions
- Schedulers
- Balancers
- Directory structures and files of NameNode/DataNode
- The checkpoint procedure
- NameNode failure
- NameNode recovery
- Safe mode
- Adding and removing nodes
- Metadata and data backup

Module 12: Pig

- Introduction
 - What is Pig
 - Salient features of Pig
 - Use cases of Pig
 - Interacting with Pig
 - Real-life connect
 - Working of Pig
 - Installing Pig engine
 - Data model
 - Business scenario
 - Relations and commands
- Basic data analysis
 - Latin syntax
 - Simple data types
 - Loading data
 - Schema
 - Data filtering and sorting
 - Common functions
- Complex data processing
 - Complex data types
 - Grouping
- Multi-data-set operations
 - Combining data-sets
 - Methods used for combining
 - Set operations
 - Data-sets split
- Extended Pig
 - Processing data with Pig using other languages
 - UDFs
 - Macros and import
- Apache Pig

If you are interested in joining EnhanceLearn's best Training and Placement Program team, please reach our team here:

Big Data Hadoop Developer Training

- Pig architecture
- Pig vs. MapReduce
- Data types
- Latin relational operators
- Pig Latin join and CoGroup
- Pig Latin Group and Union
- Pig Latin file loaders and UDF

Module 13: Hive

- Introduction
 - What is Hive
 - Hive schema
 - Hive meta store
 - Data storage
 - Traditional databases
 - Use cases
 - Hive vs. Pig
- Relational data analysis
 - Databases and tables
 - Data types
 - Joining data-sets
 - Basic syntax
 - Common built-in functions
- Data management
 - Formats of Hive data
 - Loading data
 - Self-managed tables
 - Databases and tables alteration
- Optimization
 - Query performance
 - Query optimization
 - Bucketing
 - Partitioning
 - Data indexing
- Extending Hive
- User-defined functions

Module 14: HBase

- Introduction
 - Architecture
 - Objectives
 - Real-life connect with HBase
 - Characteristics
 - Components
 - HBase operations
 - Scan
 - Get
 - Delete
 - Put
 - Business scenario

- Configuration
- Fundamentals
- Installation
- HBase shell commands
- What is NOSQL
- Apache HBase
- Why HBase
 - Data model
 - Table and row
 - Cell
 - Cell versioning
 - Column qualifier
 - Column family
 - HBase master
 - HBase vs. RDBMS
 - Column families
- Performance tuning
- Java-based APIs

Module 15: NOSQL

- Introduction
- CAP theorem
- Key value stores
 - Riak
 - Memcached
 - Dynamo DB
 - Redis
- Document store
 - MongoDB
 - CouchDB
- Graph store
 - Neo4J
- Column family
 - HBase
 - Cassandra
- NOSQL vs. SQL
- NOSQL vs. RDBMS

Module 16: Flume

- Introduction
- Big Data ecosystem
- Data sources
- Core concepts
- Anatomy
- Channel selector
- Why channels
- Data ingest
- Routing and replicating
- Use cases
 - Log aggregation

If you are interested in joining EnhanceLearn's best Training and Placement Program team, please reach our team here:

Big Data Hadoop Developer Training

- Adding Flume agent
- Handling a server farm
- Data volume per agent
- Flume deployment example

Module 17: Sqoop

- Introduction
- Uses
- Benefits
- Sqoop processing
- Execution process
- Import process
- Connectors
- Sample commands
- Events
- Clients
- Agents
- Sinks
- Source

Module 18: Hue

- Introduction
- Ecosystem
- Real-world view
- Benefits
- Updating data in file browser
- User integration
- HDFS integration
- Fundamentals of Hue frontend

Module 19: Oozie

- Introduction
- Why Oozie
- Installation
- Running an example
- Workflow engine
- Workflow submission
- Workflow application
- Workflow state transitions
- Coordinator
- Bundle
- Time line of Oozie job
- Abstraction layers
- Use cases
 - Time triggers
 - Rolling window
 - Data triggers

Module 20: Zookeeper

- Introduction
- Data model
- Service
- Use cases
- Znodes
 - Types of Znodes
 - Znodes operations
 - Znodes watches
 - Reads and writes of Znodes
- Cluster management
- Leader election
- Consistency guarantees

Module 21: Impala

- Objectives
- Goals and uses
- SQL
- Architecture
- Impala state store
- Impala catalogue service
- Query execution phases

Module 22: Commercial Distribution of Hadoop

- Introduction
- Objectives of commercial distribution
- Cloudera introduction
- Downloading Cloudera
- Logging into Hue
- Cloudera manager
- Business scenario
- MapReduce data platform
- Hortonworks data platform
- Cloudera CDH
- Pivotal HD

Module 23: Ecosystem and Its Components

- Apache Hadoop ecosystem
- File system components
- Serialization components
- Data store components
- Job execution components
- Security components
- Analytics and intelligence components
- Data interactions components
- Data transfer components
- Search frameworks components
- Graph-processing framework components

If you are interested in joining EnhanceLearn's best Training and Placement Program team, please reach our team here:

Big Data Hadoop Developer Training

Module 24: Hadoop Monitoring and Troubleshooting

- Monitoring practices
- Fair scheduler
- Configuration of Fair scheduler
- Schedule of default Hadoop FIFO
- Troubleshooting and log observation
- Apache Ambari and its key features
- Hadoop security
 - Kerberos
 - Authentication mechanism
 - Configuration steps
- Data confidentiality
- Usage of trademarks

Module 25: Java Essentials for Hadoop

- Essentials of Java
- Objectives
- JVM – Java Virtual Machine
- Working of Java
- Variables in Java
- Static vs. non-static variables
- Naming conventions of variables
- Operators
 - Unary
 - Mathematical
 - Relational

- Bitwise
- Logical/conditional
- Flow control
- Statements and blocks of code
- Arrays and strings
- Classes and methods
- Access modifiers
- Java constructors
- Objectives
- Salient features
- Class objects
- Introduction to packages
- Naming conventions of packages
- Introduction to inheritance
- Types of inheritance
 - Hierarchical
 - Multilevel
- Method overriding
- Abstract classes
- Classes and exceptions in Java
- Enums of Java
- Array list
- Iterators
- Hashmaps
- Hashtable class
- Exceptions
- Error handling

If you are interested in joining EnhanceLearn's best Training and Placement Program team, please reach our team here: