Big Data Hadoop Developer Training Course

The Big Data Hadoop Architect Master's Program transforms you into a qualified Hadoop Architect. This data architect certification lets you master various aspects of Hadoop, including real-time processing using Spark and NoSQL database technology and other Big Data technologies such as Storm, Kafka and Impala. Big Data Hadoop architects are among the highest paid professionals in the IT industry.

Key Features

- 50+ In-Demand Skills & Tools
- 12+ Real-Life Projects
- Rs.19 lacs - Rs.26 lacs salaries

WHO SHOULD ATTEND

- Data Architects
- Data Scientists
- Developers
- Data Analysts
- BI Analysts
- BI Developers
- SAS Developers
- Others who analyze Big Data in Hadoop environment
- Consultants who are actively involved in a Hadoop Project
- Java software engineers who develop Java MapReduce applications for Hadoop 2.0.

Modes of Engagement

PanelcsCourses

Instructor-Led Classroom Training

4-Day Bigdata and hadoop develop rCertification exam prep classroom training workshops conducted worldwide.

Instructor-Led Live Online Training

Provided to your company's employees across global locations through Citrix GoToMeeting or Cisco WebEx.

Self-Placed E-Learning

Anywhere, anytime access to E-Learning through a Learning Management System for employees across the globe.

Enterprise Training

In-House instructor-led 4-day Bigdata & hadoop developer certification training in your office across global locations. We can also provide 2-day PMP Fundamentals training for your team to precede the PMP certification training.

Course Agenda

PanelcsCourses

1.Introduction to Big Data and Hadoop

Learning objectives:

This module will introduce you to the various concepts of big data analytics, and the seven Vs of big data—Volume, Velocity, Veracity, Variety, Value, Vision, and Visualization. Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3.

Topics:

- Understanding Big Data
- Types of Big Data
- Difference between Traditional Data and Big Data
- Introduction to Hadoop
- Distributed Data Storage In Hadoop, HDFS and Hbase
- Hadoop Data processing Analyzing Services MapReduce and spark, Hive Pig and Storm
- Data Integration Tools in Hadoop
- Resource Management and cluster management Services

Hands-on: No hands-on

2.Big Data Ecosystem

Learning Objectives:

Here you will learn the features in Hadoop 3.x and how it improves reliability and performance. Also, get introduced to MapReduce Framework and know the difference between MapReduce and YARN.

Topics:

- Need of Hadoop in Big Data
- Understanding Hadoop And Its Architecture
- The MapReduce Framework
- What is YARN?
- Understanding Big Data Components
- Monitoring, Management and Orchestration Components of Hadoop Ecosystem
- Different Distributions of Hadoop
- Installing Hadoop 3

Hands-on: Install Hadoop 3.x

3.Hadoop Cluster Configuration

Learning Objectives: Learn to install and configure a Hadoop Cluster.

Topics:

- Hortonworks sandbox installation & configuration
- Hadoop Configuration files
- Working with Hadoop services using Ambari
- Hadoop Daemons
- Browsing Hadoop UI consoles
- Basic Hadoop Shell commands
- Eclipse & winscp installation & configurations on VM

Hands-on: Install and configure eclipse on VM

4.Big Data Processing with MapReduce

Learning Objectives:

Learn about various components of the MapReduce framework, and the various patterns in the MapReduce paradigm, which can be used to design and develop MapReduce code to meet specific objectives.

Topics:

- Running a MapReduce application in MR2
- MapReduce Framework on YARN
- Fault tolerance in YARN
- Map, Reduce & Shuffle phases
- Understanding Mapper, Reducer & Driver classes
- Writing MapReduce WordCount program
- Executing & monitoring a Map Reduce job

Hands-on :Use case - Sales calculation using M/R

5.Batch Analytics with Apache Spark

Learning Objectives:

Learn about Apache Spark and how to use it for big data analytics based on a batch processing model. Get to know the origin of DataFrames and how Spark SQL provides the SQL interface on top of DataFrame.

Topics:

- SparkSQL and DataFrames
- DataFrames and the SQL API
- DataFrame schema
- Datasets and encoders
- Loading and saving data
- Aggregations
- Joins

Hands-on:

Look at various APIs to create and manipulate DataFrames and dig deeper into the sophisticated features of aggregations, including groupBy, Window, rollup, and cubes. Also look at the concept of joining datasets and the various types of joins possible such as inner, outer, cross, and so on

6.Real Time Analytics with Apache Spark

Learning Objectives:

Understand the concepts of the stream-processing system, Spark Streaming, DStreams in Apache Spark, DStreams, DAG and DStream lineages, and transformations and actions.

Topics:

- A short introduction to streaming
- Spark Streaming
- Discretized Streams
- Stateful and stateless transformations
- Checkpointing

- Operating with other streaming platforms (such as Apache Kafka)
- Structured Streaming

Hands-on: Process Twitter tweets using Spark Streaming

7.Analysis using Pig
Learning Objectives:

Learn to simplify Hadoop programming to create complex end-to-end

Enterprise Big Data solutions with Pig.

Topics:

- Background of Pig
- Pig architecture
- Pig Latin basics
- Pig execution modes
- Pig processing – loading and transforming data
- Pig built-in functions
- Filtering, grouping, sorting data
- Relational join operators
- Pig Scripting
- Pig UDF's

8.Analysis using Hive Data Warehousing Infrastructure

Learning Objectives:

Learn about the tools to enable easy data ETL, a mechanism to put structures on the data, and the capability for querying and analysis of large data sets stored in Hadoop files.

Topics:

- Background of Hive
- Hive architecture
- Hive Query Language
- Derby to MySQL database
- Managed & external tables
- Data processing – loading data into tables
- Hive Query Language
- Using Hive built-in functions
- Partitioning data using Hive
- Bucketing data
- Hive Scripting
- Using Hive UDF's

9.Working with HBase

Learning Objectives:

Look at demos on HBase Bulk Loading & HBase Filters. Also learn what Zookeeper is all about, how it helps in monitoring a cluster & why HBase uses Zookeeper.

Topics:

- HBase overview
- Data model
- HBase architecture
- HBase shell
- Zookeeper & its role in HBase environment
- HBase Shell environment
- Creating table
- Creating column families
- CLI commands – get, put, delete & scan
- Scan Filter operations

10.Importing and Exporting Data using Sqoop

Learning Objectives:

Learn how to import and export data between RDBMS and HDFS.

Topics:

- Importing data from RDBMS to HDFS
- Exporting data from HDFS to RDBMS
- Importing & exporting data between RDBMS & Hive tables

11. Oozie Workflow Management and Using Flume for Analyzing Streaming Data

Learning Objectives:

Understand how multiple Hadoop ecosystem components work together to solve Big Data problems. This module will also cover Flume demo, Apache Oozie Workflow Scheduler for Hadoop Jobs.

Topics:

- Overview of Oozie
- Oozie Workflow Architecture
- Creating workflows with Oozie
- Introduction to Flume
- Flume Architecture
- Flume Demo

12. Visualizing Big Data

Learning Objectives:

Learn to constantly make sense of data and manipulate its usage and interpretation; it is easier if we can visualize the data instead of reading it from tables, columns, or text files. We tend to understand anything graphical better than anything textual or numerical.

Topics:

- Introduction
- Tableau
- Chart types
- Data visualization tools

Hands-on: Use Data Visualization tools to create a powerful visualization of data and insights.

13. Introducing Cloud Computing

Learning Objectives:

Learn a simple way to access servers, storage, databases, and a broad set of application services over the internet.

Topics:

- Cloud computing basics
- Concepts and terminology
- Goals and benefits
- Risks and challenges
- Roles and boundaries
- Cloud characteristics
- Cloud delivery models
- Cloud deployment models

Hands-on: Implement Cloud computing and deploy models.

About PanelcsCourses

PanelcsCourses

- PanelcsCourses is a leading training provider, helping professionals across industries and sectors develop new expertise and bridge their skill gap for recognition and growth in the global corporate world. Developed with the intention of delivering high value training through innovative and practical approaches, PanelcsCourses offers a wide range of services in training, learning and development in the fields of technology and management.

- The founders of the company are zealous young entrepreneurs, who were motivated by the need to fill a niche in the IT Training industry for professionals and they are aided in their goal by industry experts who conduct the workshops; igniting minds and motivating professionals to face on-the-job challenges

- PanelcsCourses is an professional certification training provider catering its services globally across countries including USA , UK, CANADA, Australia,

  ,India, Middle East, Netherlands, Germany, France etc.

- With over 150 consultants and trainers, we have one of the largest pool of in-house experts in the industry. The training content, course material, and training methodology are developed by in-house subject matter experts and accredited by global certifying authorities to ensure the quality training experience.

Facebook Twitter LinkedIn YouTube

Instagram pinterest

PanelcsCourses

10685-B Hazelhurst Dr. #35723

Houston, TX 77043, USA

USA : **+1 (713) 287 1453**

IND : **+91 950 582 6394**

info@panelcscourses.com

corporate@panelcscourses.com