TechyEdz Solutions

Training | Consulting | Developement | Outsourcing

Big Data Engineer Masters Program

9032803832

9032803832

contact@techyedz.com

www.techyedz.com

Big Data Engineer Masters Program

Course Overview:

Big Data Masters Program to professionals who seek to dependant on their knowledge in the field of Big Data. It makes you proficient in tools and systems used by Big Data experts. It includes training on Hadoop and Spark, Java Essentials, and SQL. The program is customized based on current industry standards that comprise of major sub-modules as a part of the training process. This program is designed by the industry experts to provide hands-on training with tools that are used to speed up the training process.

Core Java

1. **Getting Started with Java SE**

   ➢ What is Java?
   ➢ Installing Java
   ➢ The jdk Directory Structure
   ➢ Sdk structures
   ➢ OOPS Concept
   ➢ Java Language
   ➢ Java Virtual Machine
   ➢ World Wide Web and Java
   ➢ Java Platforms

2. **First Java Programs**
   ➢ Writing your first Java program "Hello, World", your first small step towards learning Java
   ➢ Program Structure
   ➢ Output in Java
   ➢ Variables and Expressions

3. **Data types and Variables**
   ➢ Primitive Data types
   ➢ Variable Names
   ➢ Numeric Literals
   ➢ Character Literals
   ➢ String
   ➢ String Literals

> Arrays

4. Introduction to Objects

> Object Models
> Classes and Objects
> Abstract methods and Classes
> Input in Java
> Input Wrapper Class
> Packages

5. Data Types and Operators

> Strong Typing
> Integer Data Types
> Floating Point
> Conversions Between Types
> Arithmetic Operators
> Doing Math in Java
> Precedence
> Errors in Integer Arithmetic

6. Control Flow

7. Booleans and Enumerations

8. Loops and Program Flow

9. Object-Oriented Programming

> Classes and Objects
> Fields and Methods
> Encapsulation
> Access Control
> Inheritance
> Polymorphism
> Interface
> Best Practices

10. Methods

- Methods
- Calling Methods
- Defining Methods
- Method Parameters
- Method Overriding
- Method Overloading

11. Characters and Strings

- Char Data Type
- Character Codes
- ASCII and Unicode
- String Class
- String Input and Output
- String Methods

12. Modular Programming

- Monolithic Programs
- Static Variables and Methods
- Functional Modularity
- Object Modularity
- Top-Down and Bottom-Up Development
- Pass-By-Value and Pass-By-Reference
- Nested Classes

13. Exception Handling and More Flow Control

- Exceptions Overview
- Exceptions
- Declaring Exceptions
- Defining and Throwing Exceptions
- Errors and Runtime Exceptions
- Catching Exceptions
- The finally Block
- Exception Methods
- I/O Exceptions vs Runtime Exceptions

14. Input/Output Streams

- Overview of Streams
- Bytes VS Characters
- Converting Byte Streams to Character Streams
- File Object
- Binary Input and Output
- Print Writer Class
- Reading and Writing Objects
- Basic and Filtered Streams

15. Core Collection Classes

- The Collections Framework
- The Set Interface
- Set Implementation Classes
- The List Interface
- List Implementation Classes
- The Queue Interface
- Queue Implementation Classes
- Implementing a Stack
- The Map Interface
- Map Implementation Classes

16. Collection Sorting and Tuning

- New Features in JSE 6
- Changing in I/O(JSE 6)
- Using Java 6.0 Features with Collections
- Sorting with Comparable
- Sorting with Comparator
- Sorting Lists and Arrays
- Collections Utility Methods
- Tuning Array List
- Navigable Map and Navigable Set
- Tuning Hash Map and Hash Set

17. Inner Classes

- ➢ Inner Classes
- ➢ Member Classes
- ➢ Local Classes
- ➢ Anonymous Classes
- ➢ Instance Initializes
- ➢ Static Nested Classes

SQL

1: Introduction to Oracle Database (SQL):

- ➢ What is Database
- ➢ Why Oracle
- ➢ Introduction to SQL and SQL *Plus, More SQL*Plus Commands
- ➢ The Data Dictionary

2: Data Query Language (DQL):

- ➢ Categories of SQL Statements
- ➢ Oracle Data types
- ➢ The SELECT Statement
- ➢ Practical examples

3: Data Definition Language (DDL):

- ➢ DDL Statements
  The CREATE Statement
- ➢ The DROP Command
- ➢ The ALTER Command
- ➢ Practical examples

4: Data Manipulation Language (DML):

- ➢ DML Statements,
- ➢ The INSERT Statement

- ➢ The DELETE Statement
- ➢ The UPDATE Statement

5: Transaction Control Language (TCL):

- ➢ TCL Statements
- ➢ COMMIT Statement
- ➢ ROLLBACK Statement
- ➢ SAVEPOINT Statement

6: Data Control Language (DCL):

- ➢ DCL Statements
- ➢ GRANT Statement
- ➢ REVOKE Statement

7: Integrity Constraints:

- ➢ Introduction to Integrity Constraints
- ➢ Categories of Integrity Constraints
- ➢ NOT NULL Constraints
- ➢ UNIQUE KEY Constraints
- ➢ PRIMARY KEY Constraints
- ➢ FOREIGN KEY or Referential Integrity Constraints
- ➢ CHECK Constraints, Practical examples

8: SQL Operators:

- ➢ Simple Selects
- ➢ Comparison Operators
- ➢ IN and NOT IN Operators
- ➢ BETWEEN Operator
- ➢ The LIKE Operator
- ➢ Logical Operators
- ➢ IS NULL and IS NOT NULL
- ➢ ANY

- ➤ ALL
- ➤ EXISTS

9: Set Operators:

- ➤ Introduction
- ➤ Selection Criteria
- ➤ Union
- ➤ Union All
- ➤ Intersect
- ➤ Minus

10: Joining Tables:

- ➤ Joins
- ➤ Table Aliases
- ➤ Cartesian Product
- ➤ Inner Joins
- ➤ Equi- Join
- ➤ Non-Equi Join
- ➤ Non-Key Join
- ➤ Reflexive / Self Join
- ➤ Natural Join
- ➤ Outer Joins
- ➤ Right Outer Join
- ➤ Left Outer Join
- ➤ Full Outer Join

11: SQL Sub queries:

- ➤ Introduction
- ➤ Using a Sub query with a DML Statement
- ➤ Typical Sub queries

> - Sub query Operators

> - Standard vs. Correlated Sub queries

> - Correlated Sub query

12: Groups:

> - SQL Statements

> - GROUP BY Clause

> - HAVING Clause

> - Order of a SELECT Statement

13: SQL BUILT-IN FUNCTIONS:

> - Introduction, Pseudo Columns
> - GROUP Functions
> - MATHEMATICAL / NUMERIC Functions
> - STRING / CHARACTER Functions
> - DATE / TIME Functions
> - CONVERSION Functions
> - MISCELLANEOUS Functions

14: More Database Objects:

> - More Database Objects
> - VIEWS
> - SEQUENCE
> - SYNONYMS

Bigdata Hadoop Development

✚ **Course Outline:**

Introduction to BigData, Hadoop

- ➤ Big Data Introduction
- ➤ Hadoop Introduction
- ➤ What is Hadoop? Why Hadoop?
- ➤ Hadoop History?
- ➤ Different types of Components in Hadoop?
- ➤ HDFS, MapReduce, PIG, Hive, SQOOP, HBASE, OOZIE, Flume, Zookeeper and so on…
- ➤ What is the scope of Hadoop?

Deep Drive in HDFS (for Storing the Data)

- ➤ Introduction of HDFS
- ➤ HDFS Design
- ➤ HDFS role in Hadoop
- ➤ Features of HDFS
- ➤ Daemons of Hadoop and its functionality
- ➤ Name Node
- ➤ Secondary Name Node
- ➤ Job Tracker
- ➤ Data Node
- ➤ Task Tracker
- ➤ Anatomy of File Wright
- ➤ Anatomy of File Read
- ➤ Network Topology
- ➤ Nodes
- ➤ Racks
- ➤ Data Center
- ➤ Parallel Copying using DistCp
- ➤ Basic Configuration for HDFS
- ➤ Data Organization
- ➤ Blocks and
- ➤ Replication
- ➤ Rack Awareness
- ➤ Heartbeat Signal
- ➤ How to Store the Data into HDFS
- ➤ How to Read the Data from HDFS
- ➤ Accessing HDFS (Introduction of Basic UNIX commands)
- ➤ CLI commands

MapReduce using Java (Processing the Data)

- The introduction of MapReduce.
- MapReduce Architecture
- Data flow in MapReduce
- Splits
- Mapper
- Portioning
- Sort and shuffle
- Combiner
- Reducer
- Understand Difference Between Block and InputSplit
- Role of RecordReader
- Basic Configuration of MapReduce
- MapReduce life cycle
- Driver Code
- Mapper
- and Reducer
- How MapReduce Works
- Writing and Executing the Basic MapReduce Program using Java
- Submission & Initialization of MapReduce Job.
- File Input/Output Formats in MapReduce Jobs
- Text Input Format
- Key Value Input Format
- Sequence File Input Format
- NLine Input Format
- Joins
- Map-side Joins
- Reducer-side Joins
- Word Count Example
- Partition MapReduce Program
- Side Data Distribution
- Distributed Cache (with Program)
- Counters (with Program)
- Types of Counters
- Task Counters
- Job Counters
- User Defined Counters

- ➢ Propagation of Counters
- ➢ Job Scheduling

PIG

- ➢ Introduction tApache PIG
- ➢ Introduction tPIG Data Flow Engine
- ➢ MapReduce vs. PIG in detail
- ➢ When should PIG use?
- ➢ Data Types in PIG
- ➢ Basic PIG programming
- ➢ Modes of Execution in PIG
- ➢ Local Mode and
- ➢ MapReduce Mode
- ➢ Execution Mechanisms
- ➢ Grunt Shell
- ➢ Script
- ➢ Embedded
- ➢ Operators/Transformations in PIG
- ➢ PIG UDF's with Program
- ➢ Word Count Example in PIG
- ➢ The difference between the Map
- ➢ Reduce and PIG

SQOOP

- ➢ Introduction tSQOOP
- ➢ Use of SQOOP
- ➢ Connect tmySql database
- ➢ SQOOP commands
- ➢ Import
- ➢ Export
- ➢ Eval
- ➢ Codegen etc…
- ➢ Joins in SQOOP
- ➢ Export tMySQL
- ➢ Export tHBase

HIVE

- ➢ Introduction tHIVE
- ➢ HIVE Meta Store
- ➢ HIVE Architecture
- ➢ Tables in HIVE
- ➢ Managed Tables
- ➢ External Tables
- ➢ Hive Data Types
- ➢ Primitive Types
- ➢ Complex Types
- ➢ Partition
- ➢ Joins in HIVE
- ➢ HIVE UDF's and UADF's with Programs
- ➢ Word Count Example

HBASE

- ➢ Introduction tHBASE
- ➢ Basic Configurations of HBASE
- ➢ Fundamentals of HBase
- ➢ What is NoSQL?
- ➢ HBase Data Model
- ➢ Table and Row
- ➢ Column Family and Column Qualifier
- ➢ Cell and its Versioning
- ➢ Categories of NoSQL Data Bases
- ➢ Key-Value Database
- ➢ Document Database
- ➢ Column Family Database
- ➢ HBASE Architecture
- ➢ HMaster
- ➢ Region Servers
- ➢ Regions
- ➢ MemStore
- ➢ Store
- ➢ SQL vs. NOSQL
- ➢ How HBASE is differed from RDBMS
- ➢ HDFS vs. HBase
- ➢ Client-side buffering or bulk uploads

- HBase Designing Tables
- HBase Operations
- Get
- Scan
- Put
- Delete

HCatalog

- HCatalog Installation
- Introduction to HCatalog
- About Hcatalog with PIG,HIVE and MR
- Hands on Exercises

MongoDB

- What is MongoDB?
- Where tUse?
- Configuration On Windows
- Inserting the data intMongoDB?
- Reading the MongoDB data.

Cluster Setup

- Downloading and installing the Ubuntu12.x
- Installing Java
- Installing Hadoop
- Creating Cluster
- Increasing Decreasing the Cluster size
- Monitoring the Cluster Health
- Starting and Stopping the Nodes

Zookeeper

- Introduction Zookeeper
- Data Modal
- Operations

OOZIE

- ➢ Introduction tOOZIE
- ➢ Use of OOZIE
- ➢ Where tuse?

Flume

- ➢ Introduction tFlume
- ➢ Uses of Flume
- ➢ Flume Architecture
- ➢ Flume Master
- ➢ Flume Collectors
- ➢ Flume Agents

SPARK

- ➢ Spark Overview
- ➢ Linking with Spark, Initializing Spark
- ➢ Using the Shell
- ➢ Resilient Distributed Datasets (RDDs)
- ➢ Parallelized Collections
- ➢ External Datasets
- ➢ RDD Operations
- ➢ Basics, Passing Functions to Spark
- ➢ Working with Key-Value Pairs
- ➢ Transformations
- ➢ Actions
- ➢ RDD Persistence
- ➢ Which Storage Level to Choose?
- ➢ Removing Data
- ➢ Shared Variables
- ➢ Broadcast Variables
- ➢ Accumulators
- ➢ Deploying to a Cluster
- ➢ Unit Testing
- ➢ Migrating from pre-1.0 Versions of Spark
- ➢ Where to Go from Here

Prerequisites:
- There are no such prerequisites for Big Data & Hadoop Course. However, prior knowledge of Core Java and SQL will be helpful but is not mandatory. Further, to brush up your skills, we offers a complimentary self-paced course on "Java essentials for Hadoop" when you enroll for the Big Data and Hadoop Course.

Who can Attend:
- Be a graduate (Engineering or Equivalent)

Number of Hours: 100hrs

Certification: None

Key Features:
➢ One to One Training
➢ Online Training
➢ Fastrack & Normal Track
➢ Resume Modification
➢ Mock Interviews
➢ Video Tutorials
➢ Materials
➢ Real Time Projects
➢ Virtual Live Experience
➢ Preparing for Certification