

Big Data Certification with Hadoop & Java

Course brief

How big is BIG? Become a big data expert through an intensive training program customised across various levels designed specifically for you. It will make participants solve real-time problems with huge datasets. Through this intensive program we aim to train the participants in a way that they are prepared to appear for International Certifications. The main aim of this course is to introduce hadoop and map-reduce with hands-on exercises. It is to make the student familiar with hadoop and map-reduce environment. Successful completion of this course will provide a platform to clear HDP Certified Developer: Java (HDPCD: Java), an internationally acclaimed certification program.

Syllabus

Subjects

Big Data Certification with Hadoop & Java

Duration

6 Weeks (M-W-T)

Big Data Certification with Hadoop & Java

Curriculum

Bigdata Concepts

Introduction Data, Storage, Bigdata, Distributed environment, Hadoop introduction History, Environment, Benefits - Subprojects HDFS, Map-Reduce, PIC, Hbase, Hive, Zoo-Keeper, SQOOP, Mahout, MongoDB, Hadoop DB

Learning Outcomes:

- Understand big data, challenges, distributed environment.
- Know hadoop and sub projects.

HDFS

Hadoop Architecture : Overall Architecture-NameNode - Datanode Fault Tolerance - Read&Write operations - Interfaces(Command line interface, JSP, API) - HDFS Shell - FS Shell Commands - Java API Programs

Learning Outcomes:

- Acquire knowledge of HDFS components , Namenode, Datanode.
- Acquire knowledge of storing and maintaining data in cluster, reading and writing data to/from cluster.
- Be able to maintain files in HDFS
- Be able to access data from HDFS through java program

Basic Map-Reduce

Map-Reduce Introduction - Map-Reduce Architecture - Yarn Architecture - Basic M-R Programs - Detailed description of M-R Methods and exercises -

Learning Outcomes:

- Understand Map-Reduce paradigm and Yarn Architecture.
- Analyze a given problem in map-reduce pattern.
- Be able to write Basic Map-Reduce Programs.

Customize Key/Value from Map to Reduce

Rkey/value pairs - Different types of values from a mapper - GenericWritable - Custom values from mapper - Writable - Custom keys from Mapper - WritableComparable - Exercises

Learning Outcomes:

- Understand the key-value pairs from map to reduce
- Be able to design applications with custom value types
- Be able to design applications with custom key types
- Applications with Generic writable

Custom Input/Output files for Map-Reduce

Input format - FileInputFormat - Steps for Input - RecordReader - Custom FileInputFormat - Custom RecordReader - Exercise
Output format - FileOutputFormat - RecordWriter - Custom FileOutputFormat - Custom RecordWriter

Learning Outcomes:

- Understand the input and output formats of map-reduce application.
- Be able to read different formats of files into map-reduce application.
- Be able to produce different formats of files from map-reduce application.

Process through Map to Reduce

Combiners - Partitioners - Secondary Sorting - Exercises

Learning Outcomes:

- Understand the process between map and reduce phases.
- Be able to optimize the performance of Map-Reduce application.
- Be able to classify the output of map-reduce application.
- Be able to use combiners in Map-Reduce application
- Be able to use Partitioners
- Be able to sort an additional key

Joins

Joins- various types - Reduce Side joins - Distributed Cache - Map-Side Join - Exercises

Learning Outcomes:

- Be able to take data from multiple data sets and join them.
- Be able to implement various joins in Map-Reduce.
- Be able to design applications with map-side joins.
- Be able to design application with reduce side join.

Instructors

Mr. P.V.N.Balarama Murthy

Hadoop Map Reduce

Mr. P.V.N.Balarama Murthy, is an M.Tech(CSE) having over 10 years of teaching and technical training experience. He is specialist in Data Science and Bigdata. He has experience in deploying hadoop clusters. As technical trainer, he has trained a number of people in C,C++, Java, Oracle, Hadoop (Administration, Development with MR, Pig, Hive, Flume, Sqoop) and Data Science with R. He has guided to his credit 15+ students to get Hortonworks certifications for Hadoop.

A dedicated, resourceful and result oriented instructor that he is, it is helping shape up careers of students.

Ms. Jyothi SanjeevaMani

Hadoop Map Reduce

Ms. Jyothi SanjeevaMani has over 15 years of satisfying teaching and technical training experience. She is a Research Scholar of Big Data Analytics from a reputed university. As a technical trainer she trained many students in industry oriented subjects like C, C++, Java, MySQL, Oracle (SQL, PL/SQL), Python, Linux, Openstack, BigData - Hadoop(MapReduce, Pig, Hive, Sqoop, Flume), Data Science with both Python and R.

She is an Asst.Professor with the Department of IT at The Keshav Memorial Institute of Technology (KMIT).

She is a dedicated, resourceful and a result oriented instructor, who strives to help students change marginal grades into good grades.