

Introduction to Hadoop Development

Description

This Apache Hadoop Developer Certification Training will help you get a detailed idea about Big Data and Hadoop. Some of the topics included are introduction to the Hadoop ecosystem, understanding of HDFS and MapReduce including MapReduce abstraction. Learn to install, implement various components of Hadoop like Pig, Hive, Flume, Sqoop and YARN.

Meta Fields

Contenu Cours : Module 1: Introduction to Big Data & Hadoop and its Ecosystem, Map Reduce and HDFS
Module 2: Hadoop Installation & setup
Module 3: Deep Dive in Mapreduce
Module 4: Graph Problem Solving
Module 5: Detailed understanding of Pig
Module 6: Detailed understanding of Hive
Module 7: (AVRO) Data Formats
Module 8: Introduction to Hbase architecture
Module 9: Hadoop Cluster Setup and Running Map Reduce Jobs
Module 10: Advance Mapreduce

Lab Exercises :

- Working with HDFS
- Writing WordCount Program
- Writing custom partitioner
- Mapreduce with Combiner
- Map Side Join
- Reduce Side Joins
- Unit Testing Mapreduce
- Running Mapreduce in Local Job Runner Mode

Documentation :

- Digital courseware included

Profils Participants :

- Software developers, analytics, BI, ETL, and data warehousing professionals
- Big Data Hadoop developers, architects and testing personnel

Connaissances :

- No prerequisites

Objectifs :

- Learn the Hadoop Architecture and Hadoop basics for beginners
- Learn what is Hadoop, HDFS and MapReduce framework
- Write MapReduce programs and deploy Hadoop clusters
- Develop applications for Big Data using Hadoop Technology
- Develop YARN programs on the Hadoop 2.X version
- Work on Big Data analytics using Hive, Pig and YARN
- Integrate MapReduce and HBase to do advanced usage and Indexing
- Learn fundamentals of Spark framework and its working
- Understand RDD in Apache Spark

-
- Learn Hadoop development best practices
 - Job scheduling using Oozie

Examen Inclus : Non certifiant

Duree Unite : jour

Duree Nombre : 3

Reference : HDP-01