

Hadoop and Big Data Analytics



The Course

Bigdata is the most popular and most widely used analytics technology platform in the world. This course, is designed to prepare administrators, engineers, architects, and consultants for designing workload deployments on the Hadoop and Bigdata platform.

When you're finished with this course, you'll have the skills and knowledge needed to securely design highly available and scalable systems, and you'll also be highly prepared for the Hadoop Certifications.

The Eligibility

Passionate Technology Enthusiasts with a minimal knowledge on IT and Operating Systems.

Good to have basic knowledge on Windows, OS & IT Infrastructure.

The Rulepaper Promise

Our training methodologies promises to give the students hands on art enterprise skills to delve deeper into the technologies from a practical and enterprise point of view. Extreme Hands-on-Lab with a self doable on the fly practical based training approaches makes transformation of the student from a no vice to a capable experienced cloud computing engineer.



The Instructor

Enterprise Architect with huge experience on Private and Public Cloud Technologies. The trainers are advisors and members of larger Cloud Computing & Big Data Forums and seasoned integrators of IT Cloud Computing & Big Data technologies with more than 12+ years in global large enterprise giants.

Course Contents

Module 1

Introduction

1. Hadoop
2. Spark
3. Data analytics with Python & R*
4. Machine Learning

Module 2

Hadoop

1. Introduction to Big Data
2. Python Essentials -1
3. Python Essentials -2
4. Intro to Hadoop
5. Hadoop Architecture & Ecosystem
6. Hadoop Distributed File System (HDFS)
7. Introduction to MapReduce
8. MapReduce Programming (Python)
9. Apache Pig

Module 3

Scala

1. Programming with Scala
2. Programming with Scala Advanced
3. Introduction to Apache Spark architecture
4. RDD, transformations and actions
5. Spark Programming
6. Spark SQL
7. Spark Streaming – I
8. Spark Streaming – II
9. Machine Learning with Spark MLlib

Module 4

Machine Learning

1. ML Introduction & Use Cases
2. Statistics 2 – Inferential Statistics
3. Linear Regression
4. Logistic Regression
5. Decision Trees, Random Forest
6. Modelling Techniques(PCA, Feature Engineering)
7. KNN, Naive Bayes
8. Support Vector Machines(SVM)
9. Clustering, K-means
10. Time Series Modelling

The Duration

Duration of the Course is 30 hours.

The Lab Requirements

Students must bring their own laptops with basic configuration

The cost of the Training

Please send an email or contact us at enquiry@rulepaper.com to know more about the cost and next batch schedules.

The certifications

Once the training is completed the student have to enroll with GCP for getting certification and the cost is exclusive of this training . Certification is an optional or good to have future.