ProTech protechtraining.com *"Charting the Course ... ... to Your Success!"*

Big Data and Hadoop Administrator

Course Summary

Description

This is an ideal course package for every aspiring professional who wants to make his/her career in the Big Data sector. Hadoop administrator will be able to build and maintain the infrastructure which is needed to store and process big data.

This course offers a hands-on experience to install, configure and manage the Apache Hadoop platform. The course covers topics to deploy, manage, monitor and secure a Hadoop Cluster. In addition, it also focuses on the whole ecosystem of Big Data and Hadoop.

By the end of the training participant will have the knowledge and skills to become a successful Hadoop Architect.

Objectives
By the end of this course, participants will be able to:

- Create and maintain Big Data and Hadoop ecosystem
- Develop advanced cluster configuration features
- Understand the Hadoop Distributed File System
- Understand and work on MapReduce and YARN
- Understand important Hadoop components and ecosystem components like Pig, Hive, Impala, Ganglia, Nagios, Sqoop and others
- Understand Hadoop's Security system

Topics

- Course Introduction
- Introduction to Big Data and Hadoop
- Planning Hadoop Cluster
- Hadoop Installation and Configuration
- Advanced Cluster Configuration Features
- Hadoop Distributed File System
- Overview of MapReduce and YARN
- Important Hadoop Components
- Hadoop Administration and Maintenance
- Hadoop Ecosystem Components

Audience

This course is suitable for professionals aspiring for a career in Big Data using Apache Hadoop, and individuals who intend to design, deploy and maintain Hadoop clusters. System administrators, Developers, Architects, IT professionals, Analytics professionals and experts are also the key beneficiaries.

Prerequisites

- Fundamental knowledge of any programming language and Linux environment
- Participants should know how to navigate and modify files within a Linux environment
- Existing knowledge of Hadoop & Java is not required

Duration

Six days

ProTech "Charting the Course ...
protechtraining.com
... to Your Success!"

Big Data and Hadoop Administrator

Course Outline

I. Course Overview
- A. About Big Data and Hadoop Administrator course

II. Introduction to Big Data and Hadoop
- A. Introduction to Big Data
- B. Introduction to Hadoop
- C. Why Hadoop
- D. Hadoop & Traditional RDBMS
- E. Components of Hadoop & Hadoop Architecture
- F. History and uses of Hadoop

III. Planning Hadoop Cluster
- A. Overview of Hadoop Clusters
- B. Planning your Hadoop Cluster
- C. Overview of Hardware and other Network configurations
- D. Network Topology for Hadoop Clusters
- E. Overview of Cluster Management

IV. Hadoop Installation and Configuration
- A. Overview of various deployment types
- B. Installing and configuring Hadoop
- C. Configuring a single node Hadoop Cluster
- D. Configuring a multi node Hadoop Cluster
- E. Checking the correctness of Hadoop installation
- F. Demos:
  1. Install Ubuntu Server 12.04
  2. Hadoop 1.0 in Ubuntu Server 12.04
  3. Create a Clone of Hadoop Virtual Machine
  4. Perform Clustering of the Hadoop Environment
  5. Install Hadoop 2.0 in Ubuntu Server 12.0

V. Advanced Cluster Configuration Features
- A. Hadoop configuration overview and important configuration file
- B. Configuration parameters and values
- C. HDFS parameters MapReduce parameters
- D. Hadoop environment setup
- E. 'Include' and 'Exclude' configuration files
- F. Demo: Configuration Settings of Hadoop
- G. Lab Exercise

VI. Hadoop Distributed File System
- A. Introduction to HDFS
- B. Overview of HDFS Architecture
- C. Overview of HDFS Storage mechanisms
- D. Overview of HDFS Rack
- E. Writing and reading files from HDFS
- F. Understanding the important commands of HDFS
- G. Introduction to Sqoop
- H. Installing and configuring Sqoop
- I. Demos:
  1. Install Sqoop
  2. HDFS Demo
- J. Lab Exercise

VII. Overview of MapReduce and YARN
- A. Introduction to MapReduce
- B. MapReduce Architecture and working with MapReduce
- C. Development and Libraries of Map Reduce
- D. MapReduce components failures and recoveries
- E. Introduction to YARN
- F. YARN Architecture
- G. Installing and configuring YARN
- H. Working with YARN & YARN Web UI

VIII. Important Hadoop Components
- A. Understanding Hive
- B. Installing and configuring Hive
- C. Understanding Pig
- D. Installing and configuring Pig
- E. Understanding Impala
- F. Installing and configuring Impala
- G. Demos:
  1. Install Hive
  2. Install Pig
- H. Lab Exercises

IX. Hadoop Administration and Maintenance
- A. Namenode/Datanode directory structures and files
- B. File system image and Edit log
- C. The Checkpoint Procedure
- D. Namenode failure and recovery procedure
- E. Safe Mode
- F. Metadata and Data backup
- G. Potential problems and solutions / what to look for
- H. Adding and removing nodes
- I. Lab Exercise

Big Data and Hadoop Administrator

Course Outline (con't)

X. Hadoop Ecosystem Components

A. Eco system Component: Ganglia
B. Install and Configure Ganglia on a Cluster
C. Configure and Use Ganglia
D. Use Ganglia for Graphs
E. Eco system Component: Nagios
F. Nagios Concepts
G. Install and Configure Nagios on Cluster
H. Use Nagios for Sample Alerts And Monitoring
I. Eco system Component: Sqoop
J. Install and Configure Sqoop on Cluster
K. Import Data from Oracle/MySQL to Hive
L. Overview of Other Eco system Components:
M. Oozie
N. Avro
O. Thrift
P. Rest
Q. Mahout
R. Cassandra
S. YARN
T. MR2
U. Hadoop Security
V. Kerberos and Hadoop
W. Why Hadoop Security is Important?
X. Hadoop's Security System Concepts
Y. What Kerberos is and How it Works?
Z. Configuring Kerberos Security
AA. Securing a Hadoop Cluster with Kerberos
BB. Lab Exercise