Updated 06/14/2024

Sign up

Cloudera Production Administrator Training
3 days (21 hours)

Presentation

Cloudera Administrator is one of the world's most powerful open source data management platforms. Based on Hadoop, it offers a range of integrated, multi-functional tools for analyzing and centralizing data. Cloudera manages hybrid and multi-cloud data that is fully integrated with data engineering and Machine Learning, enabling easy, fast and secure business analysis. This platform will offer you security, compliance, migration and metadata management across all environments. Our Cloudera Administrator training will teach you how to manage the data lifecycle on private and public clouds. You'll also learn how to apply your data analysis and business intelligence skills to leading tools such as Apache Impala and Apache Hive. At the end of the course, you'll know how to manage and maintain a Hadoop cluster using Cloudera Manager, modify your company's software infrastructure and use various data services such as Data Warehouse, Machine Learning (ML) or Data Visualization.

Objectives

- Master the basics of the Hadoop, Spark and MapReduce environment
- Master HDFS and Yarn/MapReduce operation
- Configure and manage services with Cloudera Manager
- Developing Cloudera Hadoop solutions
- Deploy, configure and manage a cluster with Cloudera Manager
- Integrating the cluster into the IT system
- Guarantee users a high level of security and traceability

Target audience

- System administrators
- IT managers
- Systems architects
- Database specialists
- Developers

Prerequisites

- Knowledge of unix/linux commands
- Basic knowledge of SQL

Cloudera Administrator training program

Introduction to Hadoop Cloudera

- What is Cloudera?
- Advantages of Cloudera
- Installing and configuring MySQL Database for Cloudera Manager
- Introduction to cluster installation
- Installing and configuring Cloudera Manager

Cloudera features

- Creating cloud-based data lakes
- Access to data catalogs
- Real-time data analysis
- Operating in private/public clouds
- Modify software infrastructure

Setting up a Cloudera cluster

- Installation and configuration parameters
- Configuration files
- Managing role instances
- Adding new services
- Deleting hosts

Hadoop Distributed File System

- Introduction to HDFS
- Parameter settings
- Modification of logs and checkpointing
- Fault tolerance and HDFS performance
- HDFS command lines and tools

Cluster management and maintenance

- Cluster sizing parameters
- Network considerations
- Cloud deployment options
- Cluster monitoring
- Copy data between clusters
- Directory snapshots

Advanced cluster configuration

- Service gate configuration
- HDFS parameters
- MapReduce parameters
- enable HDFS High Availability

Hive and Impala

- Apache Hive
- Apache Impala

Apache Spark

- Introduction to Spark
- Apache Spark applications
- Running applications on YARN
- Monitor applications

Data Warehousing

- SQL engines
- Deploying data warehouses
- Transfer workloads from site to cloud
- Large-scale data engineering
- Visual data restitution with Data Visualisation

MapReduce and YARn

- Introduction to YARN
- Running and exploring applications
- Main YARN application logs
- MapReduce applications

Security and governance

- Security on Hadoop
- Securing a cluster
- Ensuring data security and compliance
- Kerberos overview and operation
- Metadata migration
- Metadata management in any environment

Further information

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.