ADMIN-280

Administering Cloudera on cloud

Course Overview

Course Type
Instructor-Led Training

Level
Intermediate

Duration
4 days

Platform
Cloudera on cloud

Topics Covered
- Cloudera Architecture
- CDP CLI
- Cloudera Data Hubs
- Cloudera Data Services
- Cloudera SDX
- Observability

About This Training

This training provides a comprehensive understanding of the steps required to configure, operate, and maintain Cloudera on cloud instances. This four-day instructor-led course covers everything from setup to configuring various data services to executing workloads on AWS, Azure, and GCP using Cloudera Management Console. The course covers configuration options using the web interface, as well as automation scenarios using Ansible. On the optimization side, it covers load balancing and tuning Cloudera on cloud instances. This course equips participants to effectively address the real-world challenges faced by administrators running Cloudera on cloud.

What You Will Learn

Through instructor-led discussion and interactive, hands-on exercises, you will learn how to:

- Evaluate and select the appropriate deployment option
- Setup Cloudera using Cloudera Management Console
- Setup and configure various data services
- Configure and monitor instances using Cloudera Manager
- Optimize cluster performance and security
- Detect, troubleshoot, and repair problems with the cluster
- Auto scale Data Hub clusters and Data Services

What to Expect

This course is best suited to cloud systems administrators and operators who have at least basic Linux and AWS/Azure/GCP experience. Prior knowledge of Cloudera is not required, but helpful.

Preparation

Students are highly recommended to go through the free OnDemand courses to make the best of the instructor-led classroom learning experience:

- AWS for Cloudera Cloud Fundamentals (FREE!)
- Quickstart: Azure for Cloudera (FREE!)
- Quickstart: AWS for Cloudera (FREE!)

ADMIN-280

Administering Cloudera on cloud

Installation Overview (Quick Start)

- Cloudera Management Console
- Cloudera Credentials
- Cloudera Control Plane Regions
- Register a Cloudera environment

Cloudera Data Platform

- Industry Trends for Big Data
- The Challenge to Become Data-Driven
- The Enterprise Data Cloud
- Cloudera Overview
- Cloudera Form Factors

Cloudera Architecture

- Overview
- Key Concepts & Components
- Cloudera Runtime Overview
- Minimum Hardware
- Outbound Connections

Control Plane Overview

- Accessing and Managing an Environment
- Data Management Overview
- Management Console
- Dashboard
- Environments
- Data Lakes
- User Management
- Classic Clusters
- Data Hubs
- Data Catalog
- Replication Manager
- Observability

CDP CLI (Command Line Interface)

- CDP CLI Command Line Interface
- Installing CDP CLI / CLI Client Setup
- CLI Modules
- Generating an API access key / Configuring Cloudera client
- Logging into the CDP CLI/SDK
- Configuring CLI autocomplete / CLI reference / Accessing CLI help
- Cloudera API overview / Cloudera SDK for Java overview / CDP curl overview

Managing Cloudera Access

- Management Console
- User Management
- Create Machine User
- User Permissions
- Sync Users
- Configure Groups
- Identity Providers
- Roles and Resource Roles
- Global Settings
- Audit Data Storage Credential

Data Hubs Overview

- Data Hubs
- Planning / Creating your Data Hub Cluster
- General Planning Considerations
- Configuring Nodes
- Managing Data Hub
- Choosing the Right Hardware
- Advanced Cluster Configuration
- Data Hub Types
- DataFlow
- Data Engineering
- Troubleshooting

ADMIN-280

Administering Cloudera on cloud

Managing Data Hubs

- Best Practices on Data Hubs
- Sizing Data Hubs
- Cloudera Manager
- Data Hub Services
- Autoscaling/Data Hub Info
- Checking Cluster Health Status / Events and Alerts
- Host Maintenance
- Upgrading a Data Hub Cluster
- Monitoring / Monitoring Features

Data Services Overview

- Data Services Overview
- Data Services
- Planning Your Data Service Cluster
- Choosing the Right Hardware / Network Considerations
- Creating Data Services
- DataFlow
- Data Engineering
- Data Warehouse
- Operational Database
- Machine Learning
- Troubleshooting

DataFlow

- DataFlow Service Overview
- Data Ingest Overview
- Ingesting Data using File Transfer or REST Interfaces
- Ingesting Data Using NiFi
- Autoscaling

Cloudera Data Engineering

- Data Engineering Service Overview
- Apache Spark/Flink/Kafka streams Overview
- Autoscaling

Cloudera Data Warehouse

- Data Warehouse Service Overview
- Adding and Managing a Database Catalog
- Adding and Tuning a Virtual Warehouse
- Querying a Data Warehouse
- Data Visualization
- Monitoring & Troubleshooting

Cloudera Operational Database

- Operational Database Service Overview
- Apache HBase/Search Overview
- Autoscaling

Cloudera AI

- Cloudera AI Service Overview
- Cloudera Engines
- Requirements for Cloudera AI Workspaces
- Provisioning a Cloudera AI Workspace
- Cloudera AI Auto-Scaling
- Monitoring

ADMIN-280

Administering Cloudera on cloud

Monitoring and Management

- Monitoring and Management in Cloudera on cloud
- Data Lake Cluster Monitoring and Cloudera Auditing
- Getting Started with Monitoring in Cloudera
- Monitoring with Cloudera Manager: Health Tests and Dashboards
- Monitoring Clusters, Services, Hosts, Roles, and Activities
- Troubleshooting Cluster Configuration and Operation

Data Management

- SDX – Security and Governance
- Security Concepts
- Access Cloud Storage
- Data Lake Security: SDX
- Apache Ranger
- Cloudera Authorization / Authentication
- Data Governance
- Apache Atlas
- Data Catalog

Observability

- Overview
- Support
- Observability deployment architecture
- Monitoring capabilities
- Working with alerts, costs, and reports

VERSION 2.0.0