

KOENIG SOLUTIONS • COURSE OUTLINE

Cloudera Administrator Training: CDP Private Cloud Base

Learn to install, configure, secure and maintain a CDP Private Cloud Base cluster, from data ingest to monitoring and troubleshooting.

16

MODULES

101

TOPICS

COURSE OVERVIEW

This course covers the Cloudera Data Platform and the installation of CDP Private Cloud Base with Cloudera Manager, followed by cluster configuration, data storage with HDFS, HBase and Kudu, and data ingest using Sqoop and NiFi. It continues through data flow with NiFi and Kafka, data access and discovery with Hive, Impala, Hue and CDSW, and data compute using YARN, Tez, Spark and Phoenix. Later modules address resource management, cluster planning, advanced configuration, maintenance, monitoring and troubleshooting, concluding with security through Kerberos, Ranger and Atlas, and a comparison of private and public cloud capabilities.

WHO THIS COURSE IS FOR

Cluster Administrators

System Administrators

Data Engineers

Hadoop Operators

Platform Engineers

LEARNING OBJECTIVES

- Describe the Cloudera Data Platform and its form factors
- Install CDP Private Cloud Base using Cloudera Manager
- Configure cluster services, roles and host membership
- Explain HDFS, HBase and Kudu storage architectures
- Ingest data using Sqoop, NiFi and file transfer interfaces
- Apply Kafka and NiFi for data flow management
- Analyse YARN, Spark and Tez application behaviour
- Implement security using Kerberos, Ranger and Atlas

16 Modules, 101 Topics

01

Cloudera Data Platform

TOPICS

Industry Trends for Big Data · The Challenge to Become Data Driven · The Enterprise Data Cloud · CDP Overview · CDP Form Factors

02

CDP Private Cloud Base Installation

TOPICS

Installation Overview · Cloudera Manager Installation · CDP Runtime Overview · Cloudera Manager Introduction

03

Cluster Configuration

TOPICS

Overview · Configuration Settings · Modifying Service Configurations · Configuration Files · Managing Role Instances · Adding New Services · Adding and Removing Hosts

04

Data Storage

TOPICS

Overview · HDFS Topology and Roles · HDFS Performance and Fault Tolerance · HDFS and Hadoop Security Overview · Working with HDFS · HBase Overview · Kudu Overview · Cloud Storage Overview

05

Data Ingest

TOPICS

- Data Ingest Overview
- File Formats
- Ingesting Data using File Transfer or REST Interfaces
- Importing Data from Relational Databases with Apache Sqoop
- Ingesting Data Using NiFi
- Best Practices for Importing Data

06

Data Flow

TOPICS

- Overview of Cloudera Flow Management and NiFi
- NiFi Architecture
- Cloudera Edge Flow Management and MiNiFi
- Controller Services
- Apache Kafka Overview
- Apache Kafka Cluster Architecture
- Apache Kafka Command Line Tools

07

Data Access and Discovery

TOPICS

Apache Hive · Apache Impala · Apache Impala Tuning · Search Overview · Hue Overview · Managing and Configuring Hue · Hue Authentication and Authorization · CDSW Overview

08

Data Compute

TOPICS

- YARN Overview
- Running Applications on YARN
- Viewing YARN Applications
- YARN Application Logs
- MapReduce Applications
- YARN Memory and CPU Settings
- Tez Overview
- Hive on Tez
- ACID for Hive
- Spark Overview
- How Spark Applications Run on YARN
- Monitoring Spark Applications
- Phoenix Overview

09

Managing Resources

TOPICS

Configuring cgroups with CPU Scheduling · The Capacity Scheduler · Managing Queues · Impala Query Scheduling

10

Planning Your Cluster

TOPICS

General Planning Considerations · Choosing the Right Hardware · Network Considerations · CDP Private Cloud Considerations · Configuring Nodes

11

Advanced Cluster Configuration

TOPICS

Configuring Service Ports · Tuning HDFS and MapReduce · Managing Cluster Growth · Erasure Coding · Enabling HDFS High Availability

12

Cluster Maintenance

TOPICS

Checking HDFS Status · Copying Data Between Clusters · Rebalancing Data in HDFS · HDFS Directory Snapshots · Host Maintenance · Upgrading a Cluster

13

Cluster Monitoring

TOPICS

Cloudera Manager Monitoring Features · Health Tests · Events and Alerts · Charts and Reports · Monitoring Recommendations

14

Cluster Troubleshooting

TOPICS

Overview · Troubleshooting Tools · Misconfiguration Examples

15

Security

TOPICS

- Data Governance with SDX
- Hadoop Security Concepts
- Hadoop Authentication Using Kerberos
- Hadoop Authorization
- Hadoop Encryption
- Securing a Hadoop Cluster
- Apache Ranger
- Apache Atlas
- Backup and Recovery

16

Private Cloud / Public Cloud

TOPICS

CDP Overview · Private Cloud Capabilities · Public Cloud Capabilities · What is Kubernetes? · WXM Overview · Auto-scaling

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training

Live Online Classes (ILO)

Fly-Me-A-Trainer (FMAT)

Flexi (Self-Paced)

1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com

✉ info@koenig-solutions.com