

DENG-236



Optimizing Storage with Apache Ozone

Course Overview

Course Type
Instructor-led training

Level
Intermediate

Duration
2 days

Platform
Cloudera on Premises

Topics Covered

- Ozone Concepts
- Working with Ozone
- Integrating Ozone with Hive, Impala, Spark and NiFi
- Migrating Data from HDFS to Ozone
- Performance Tuning

About This Training

This 2-day instructor-led course provides a comprehensive, hands-on introduction to Apache Ozone, the next-generation hybrid storage service offering versatility and out-of-the-box compatibility.

Participants will learn the architecture and internal operations of Ozone, gaining the skills to manage files and objects, control replication, handle failover and recovery, using a service overcoming the limitations of traditional HDFS.

The course blends foundational storage concepts with practical usage to help teams seamlessly move from traditional HDFS architectures to modern, scalable object storage strategies, including complete data migration.

Participants will learn about integrating Ozone seamlessly into the broader data ecosystem, securely pushing data with Apache NiFi and running powerful analytics using Apache Hive, Impala, and Spark while tuning the system to maximize performance.

What Skills You Will Gain

Participants will develop the following skills:

- Understanding the Benefits of Using Ozone
- Managing Files and Objects in Ozone
- Controlling Replication and Understanding Failover and Recovery
- Integrating Hive, Impala, and Spark with Ozone
- Pushing Data to Ozone Using Apache NiFi
- Moving Data between HDFS and Ozone
- Tuning Ozone to Maximize Performance

Who Should Take This Course?

This advanced course is for data engineers and applications developers who are migrating data and applications to Apache Ozone. DevOps engineers who want to ensure optimal performance would also find this course useful. Prior experience with the Cloudera platform, including HDFS, YARN, and Hive, is expected. Students must have access to the Internet to reach the classroom environments, which are located on Amazon Web Services.

DENG-236

Optimizing Data Storage with Apache Ozone

Ozone Introduction and Concepts

- Introduction
- Why Ozone?
- Ozone Concepts

Working with Ozone

- File System and Admin Commands
- Transferring Data between Ozone and HDFS
- Setting Quotas and ACLs
- Controlling Replication
- DataNode Layout and Failover

Application Integration

- Ozone Data Locality
- Integrating Ozone with Apache Hive and Impala
- Integrating Ozone with Apache Spark
- Integrating Ozone with Apache NiFi

Data Migration

- HDFS and Ozone
- Ozone Configuration
- Mapping the HDFS Namespace to Ozone
- Data Migration using DistCp

Performance Tuning

- Ozone Pipelines
- Baseline Load Testing with Freon
- Upgrading Ozone