ADMIN-434

Troubleshooting HDFS on Cloudera On-Premises Base

Course Overview

Course Type
Instructor-led training

Level
Intermediate

Duration
2 days

Platform
Cloudera on Premises Base

Topics Covered

- Identifying Problem Areas
- HDFS Components & Role Failures
- RPC Latency & Performance
- Small Files & Audit Logs
- Security & Encryption
- Upgrade Issues
- Metadata & Block Corruption

About This Training

This training offers a focused, hands-on introduction to troubleshooting Apache HDFS in Cloudera on-premises environments. We begin by identifying problem areas, then review key HDFS components and common failures across NameNode, JournalNode, DataNode, HttpFS, and the Failover Controller. Participants will analyze RPC latency, address performance issues such as slow responses, decommissioning, balancing, and GC pauses, and work through labs on small-files problems and audit log analysis. The course also covers security topics, including ACLs, Kerberos/SPNEGO, KMS, and encryption at rest, along with upgrade challenges, metadata corruption, client write/block issues, and essential tools like distcp.

What Skills You Will Gain

During this course, you will learn to:

- Identify and analyze HDFS problem areas in Cloudera on-premises environments
- Troubleshoot failures across NameNode, JournalNode, DataNode, HttpFS, and Failover Controller
- Diagnose performance issues such as slow responses, decommission slowness, balancing, and GC pauses
- Interpret RPC latency metrics and analyze HDFS audit logs
- Resolve security issues involving ACLs, Kerberos/SPNEGO, KMS, and encryption inconsistencies
- Address upgrade-related challenges during CDH/HDP-to-CDP migrations
- Detect and troubleshoot metadata corruption and HDFS client write/block errors
- Use tools like distcp effectively in operational workflows

Who Should Take This Course?

This course is ideal for data engineers, system administrators, and support teams who manage or troubleshoot HDFS in production environments. A basic understanding of HDFS is required to get the most value from the training.

ADMIN-434

Troubleshooting HDFS on Cloudera On-Premises Base

Before Starting

- Identify the Problem Area(s)

Overview of HDFS Sub-Components

- Overview of HDFS Components

Role Failures

- NameNode and NameNode HA
  - NameNode and DataNode heap issue
  - NameNode crash and performance issue due to contentSummary
  - NameNode performance issue due to LDAP slowness
  - NameNode performance issue due to Ranger
  - NameNode performance issue due to HDFS balancer
  - NameNode performance issue due to logging contention
  - NameNode slowness due to various large operations
  - NameNode slow performance every few weeks due to FoldedTreeSet
  - NameNode checkpoint issue due to Jetty bug
- JournalNode
- DataNode
- HttpFS
- Failover Controller
- Hands-on Exercise: Analyzing RPC latency issues

Performance Issues

- Slow Response
- Slow NameNode
- HDFS block missing and corrupt
- HDFS Erasure Coded data loss
- HDFS EC plus balancer causes data loss
- HDFS snapshot-related issue
- HDFS block placement issue
- Slow DataNode
- Decommission
- Balancing
- GC Pauses
- Hands-on Exercise: Solving small files issues
- Hands-on Exercise: Analyzing HDFS audit logs using Zeppelin notebook

Security

- ACLs
- Kerberos/SPNEGO
- KMS
- Encryption at Rest
- Hands-on Exercise: Solving Kerberos failure due to encryption inconsistency between Kerberos client and server

CDH/HDP to CDP Upgrade

- Troubleshooting Issues Related to Upgrades

Others

- Metadata Corruption
- HDFS Client: Write/Block Corruption
- Hands-on Exercise: Recover corrupt metadata using fsimage

Tools

- distcp

© 2026 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies.
Information is subject to change without notice.

VERSION:
2.0.0