+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated on 09/10/2025

Sign up

Red Hat Ceph Storage Training

5 days (35 hours)

Overview

Our Red Hat Ceph Storage training course will enable you to understand, deploy and operate a complete Ceph cluster. You'll be able to efficiently manage petabyte-scale storage for mission-critical workloads: datalakes, distributed databases, backups, AI/ML or cloud-native services. You'll also be able to guarantee fault tolerance, automate volume provisioning for your Kubernetes pods, and monitor your entire storage infrastructure. By the end of this course, you'll be able to deploy, secure and scale a Red Hat Ceph Storage infrastructure in production. Like all our training courses, this one is based on the latest stable version of Ceph, with practical workshops covering the most common Red Hat Ceph Storage field use cases.

Objectives

- Understand Ceph's distributed architecture and its inner workings
- Deploy a Ceph cluster with cephadm and manage services (MON, OSD, MGR, RGW...)
- Configure and operate three types of storage: object (S3), block (RBD), file (CephFS)
- Integrate Ceph in a Kubernetes cluster with CSI or Rook
- Supervise and secure a Ceph infrastructure (Prometheus, alerting, self-healing)

Target audience

- Data engineers
- DevOps
- Platform engineers
- System administrators
- Infrastructure managers

Prerequisites

- Basic knowledge of Linux administration
- Notions of networks, disks and storage (RAID, LVM...)
- Knowledge of Kubernetes or OpenStack environments appreciated
- Experience in DevOps, infrastructure or data engineering recommended

Red Hat Ceph Storage training program

Overview of distributed storage

- Storage types: NAS, SAN, object, block, file
- Limitations of traditional architectures
- Why software-defined storage (SDS)?
- Data and cloud use cases where Ceph is relevant
- Positioning Ceph in relation to MinIO, HDFS and Lustre
- Overview of Red Hat Ceph Storage

Fundamental architecture of Ceph

- Ceph = RADOS cluster + interfaces (object, block, file)
- Roles of MON, MGR, OSD and MDS daemons
- The CRUSH system: distribution without a central directory
- Redundancy, replication and erasure coding
- Peer-to-peer operation and absence of SPOF
- Customer relations ? Ceph cluster

Deploying a Ceph cluster with cephadm

- Introduction to cephadm and Ceph containers
- Container architecture (Podman or Docker)
- Automatic deployment of MON, OSD and MGR services
- Dynamic addition/removal of nodes
- Workshop: Setting up a 3-node cluster with cephadm (local or cloud)

Discovering the Ceph interface

- Ceph CLI: main commands
- Web dashboard (MGR Dashboard)
- Cluster status monitoring: health, status, PG
- Performance visualization and alerts
- Workshop: Getting to grips with the dashboard + executing key commands

Object storage with RADOS Gateway (RGW)

- Object storage principles and S3/Swift compatibility
- Configuring an RGW gateway
- Creation of buckets, users and ACLs
- Access via AWS CLI, rclone, Python boto3
- Workshop: Deploy RGW + create an S3 Ceph bucket + upload via AWS CLI

Block storage with RBD

- RBD: presentation and use cases (VMs, databases, volumes)
- Creating and mapping a block volume
- Formatting, mounting, persistence
- Snapshots, clones and rollback
- Workshop: Creating an RBD volume + hosting a PostgreSQL database

The CephFS file system

- POSIX structure, MDS, shared directories
- Filesystem creation + client mounting
- Concurrency, quota, locking
- Use cases: ML processing, project storage
- Workshop: Creating a shared CephFS + accessing multiple ML jobs/containers

Pool management, placement and replication policy

- Pools: principles, creation, properties
- Replication vs. Erasure coding: strategic choice
- Tiering: hot vs. cold storage
- Customized CRUSH mapping
- Rebalancing and dynamic distribution

Kubernetes integration with RBD and CephFS

- Presentation of CSI Ceph drivers
- Creating StorageClass, PVC, PV
- Deploying an app with persistent volume
- How it works in OpenShift, Rancher, K3s
- Workshop: Creating a Ceph PVC + attaching a PostgreSQL pod in Kubernetes

Automation with Ansible and cephadm

- ceph-ansible: complete installation from YAML
- cephadm: orchestration via CLI/API
- Managing roles, labels and services
- Security, variables and secrets
- CI/CD pipeline integration

Resilience and self-healing

- OSD/MON failure detection
- Automatic object recovery
- Simulation of disk or service failure
- Minimum replication rate and tolerance
- Simulation: stop an OSD + observe reconstruction and logs

Supervision and alerting

- Prometheus and Grafana integration
- Ceph exporters: available metrics
- Typical dashboards: IOPS, latency, errors
- Alertmanager: thresholds, rules, notifications
- Customizing a Grafana Ceph dashboard

Security and best practices

- CephX authentication, tokens, TLS
- ACLs, user roles, access control
- Data isolation and multi-tenancy
- Logs, audit logs, rotation
- Port and service hardening

Data & AI/ML scenarios with Ceph

- Ceph as a datalake backend or AI pipeline
- Shared CephFS for Spark jobs, MLflow, JupyterHub
- Dataset archiving via S3 object
- Sharing snapshots between test environments
- SSD/HDD optimization for AI workloads

Large-scale operation

- Geographical distribution (multi-site RGW)

- Hybrid cloud integration
- Growth management: add racks, zones
- Backup, replication, inter-cluster snapshots
- Technical debt and upgrade management

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced IT 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 forthcoming training course, 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 training: 60% hands-on, 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.

Certification

A certificate will be awarded to each trainee who has completed the entire course.