

Hadoop Administration Fundamentals

Learn to plan, install and administer Hadoop clusters, covering HDFS, MapReduce, YARN, monitoring, security and maintenance.

5

MODULES

40

TOPICS

COURSE OVERVIEW

This course introduces Hadoop history, concepts, ecosystem, distributions and architecture before covering planning and installation, including hardware and network selection, cluster sizing, rack topology and benchmarking. It then addresses HDFS operations such as nodes and daemons, health monitoring and storage management, followed by MapReduce operations covering JobTracker and TaskTracker, job configuration, schedulers, optimisation and YARN architecture. Finally, it examines advanced topics including hardware and cluster monitoring, backup and recovery planning, cluster configuration, Oozie scheduling and securing a cluster with Kerberos.

WHO THIS COURSE IS FOR

Hadoop Administrators

System Administrators

Cluster Operators

Infrastructure Engineers

IT Operations Staff

LEARNING OBJECTIVES

- Explain Hadoop history, concepts and ecosystem components
- Identify Hadoop distributions and high-level architecture
- Plan hardware, network and rack topology for a cluster
- Configure directory structure, logs and multi-tenancy settings
- Describe HDFS nodes, daemons and health monitoring
- Apply command-line and browser-based HDFS administration
- Analyse MapReduce job configuration, scheduling and optimisation
- Implement cluster security using Kerberos and Oozie scheduling

5 Modules, 40 Topics

01

Introduction

TOPICS

Hadoop history and concepts · Ecosystem · Distributions · High level architecture · Hadoop myths · Hadoop challenges (hardware/software)

02

Planning and Installation

TOPICS

Selecting software and Hadoop distributions · Sizing the cluster and planning for growth · Selecting hardware and network · Rack topology · Installation · Multi-tenancy · Directory structure and logs · Benchmarking

03

HDFS Operations

TOPICS

- Concepts (horizontal scaling, replication, data locality, rack awareness)
- Nodes and daemons (NameNode, Secondary NameNode, HA Standby NameNode, and DataNode)
- Health monitoring
- Command-line and browser-based administration
- Adding storage and replacing defective drives

04

MapReduce Operations

TOPICS

- **Parallel computing before MapReduce:** compare HPC versus Hadoop administration
- MapReduce cluster loads
- Nodes and Daemons (JobTracker and TaskTracker)
- MapReduce UI walk through
- MapReduce configuration
- Job config
- Job schedulers
- Administrator view of MapReduce best practices
- Optimizing MapReduce
- **Fool proofing MR:** what to tell your programmers
- **YARN:** architecture and use

05

Advanced Topics

TOPICS

- Hardware monitoring
- System software monitoring
- Hadoop cluster monitoring
- Adding and removing servers and upgrading Hadoop
- Backup, recovery, and business continuity planning
- Cluster configuration tweaks
- Hardware maintenance schedule
- Oozie scheduling for administrators
- Securing your cluster with Kerberos
- The future of Hadoop

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