DevOpsSchool

Course curriculum · November 2026

LEVEL 04 OF 04 · ADVANCED

Apache Hadoop Advanced Training

Run Hadoop in production with high availability, tuning, federation and erasure coding, and plan migrations to cloud data platforms.

LIVE ONLINE **Max 8-10 hours**	CORPORATE (DAY-WISE) **5 days · Days 1-5**	SELF-PACED VIDEO **1-year LMS access**
BATCH SIZE **Max 10 learners**	NEXT BATCHES **1st of every month · weekends: first Saturday**	CERTIFICATE **Verifiable certificate of completion**

Foundation	Essential	Intermediate	Advanced
1 day · Included	2 days · Included	3 days · Included	5 days · This curriculum

INSIDE THIS CURRICULUM

What you will achieve	Outcomes and level focus
Who should attend	Audience and prerequisites
Agenda	45 topics · Foundation to Advanced
Hands-on labs	4 labs
Formats and fees	Live, 1:1, self-paced, corporate
Trainer and reviews	Rajesh Kumar
Enrolment	Batches, payment, contacts

WHY DEVOPSSCHOOL

- Rated 4.6 on Google
- 1,500+ verifiable certificates issued in the public registry
- Trainer Rajesh Kumar: 20 years in industry
- Small batches: max 10 learners, every session recorded

Course page: www.devopsschool.com/courses/hadoop/hadoop-and-advance-training.html
Talk to us: WhatsApp +91 99057 40781 · +1 (469) 756-6329 · contact@DevOpsSchool.com · reply within 1 hour

OUTCOMES

What you will achieve

By the end of the Apache Hadoop course path, you will be able to:

1

Store and process big data

Use HDFS, YARN and MapReduce to store and process large datasets across a cluster.

2

Query with the ecosystem

Ingest, transform and query data with Hive, Pig, HBase, Sqoop, Flume and Spark.

3

Administer secure clusters

Build, secure, tune and scale highly available Hadoop clusters and plan cloud migrations.

WHAT THE ADVANCED LEVEL ADDS

High availability, performance, scale and migration

Hands-on: Highly available production data platform. Configure NameNode and ResourceManager HA, tune a Hive and Spark pipeline, enable erasure coding and replicate data to a second cluster with DistCp.

WHO SHOULD ATTEND

- Experienced Apache Hadoop practitioners responsible for quality, security and scale.
- Senior engineers, architects and leads who design and troubleshoot Apache Hadoop solutions.
- Teams making Apache Hadoop an organisation-wide standard.

BEFORE YOU START

- The Intermediate level, or equivalent hands-on experience.
- You can already: Build and administer a multi-node Hadoop cluster, run Spark on YARN and apply core Hadoop security controls.
- Basic computer skills and a laptop with a stable internet connection for the live sessions and labs.
- This curriculum includes every earlier level, so it is complete on its own.

AGENDA · FOUNDATION TO ADVANCED

Apache Hadoop Advanced Training agenda

45 topics. Each level builds on the one before it, so this curriculum lists every earlier level in full. Days are for corporate (day-wise) delivery; live online batches cover the same agenda in max 8-10 hours.

DAY 1 — Foundation: Big data concepts, HDFS, MapReduce and single-node setup

Included from Foundation

1. Introduction to Hadoop: overview of Hadoop, its history and its ecosystem in modern data platforms
2. Big data concepts: volume, velocity and variety, distributed storage and data locality
3. Hadoop architecture: HDFS, NameNode, DataNode and Secondary NameNode roles
4. HDFS concepts: blocks, replication factor, rack awareness and the write and read path
5. YARN architecture: ResourceManager, NodeManager, ApplicationMaster and containers
6. Hadoop MapReduce: overview of MapReduce, how it works and its key components
7. Hadoop installation and configuration: setting up a single-node cluster on a Linux machine
8. Core configuration files: core-site.xml, hdfs-site.xml, yarn-site.xml and mapred-site.xml
9. HDFS shell: hdfs dfs -put, -ls, -cat, -mkdir and hdfs dfsadmin -report
10. Web interfaces: NameNode UI, ResourceManager UI and the JobHistory server
11. Hands-on lab: writing and running a basic MapReduce job on the Hadoop cluster

DAY 2 — Essential: Ecosystem tools, Hive, Pig, HBase and ingestion

Included from Essential

12. Hadoop ecosystem: overview of other components such as Pig, Hive, HBase, Sqoop and Flume
13. Pig and Hive: overview of Pig and Hive and their uses in Hadoop
14. Apache Hive: metastore, HiveServer2, beeline and HiveQL DDL and DML
15. Hive tables: managed and external tables, partitions, buckets and ORC and Parquet formats
16. Apache Pig: Pig Latin, LOAD, FILTER, GROUP and STORE and running Pig scripts on YARN
17. HBase: overview of HBase, its architecture (HMaster, RegionServers, ZooKeeper) and its use cases
18. HBase data model: tables, column families, row keys and the HBase shell
19. Data ingestion with Sqoop: importing from and exporting to relational databases and its retired status
20. Data ingestion with Flume: sources, channels and sinks for log collection
21. Writing MapReduce in Java: Mapper, Reducer, combiners, partitioners and job configuration
22. Hands-on lab: working with Pig and Hive to process data on the Hadoop cluster

DAY 3 — Intermediate: Multi-node clusters, Spark, administration and security

Included from Intermediate

23. Multi-node cluster setup: planning masters and workers, SSH, workers file and cluster start-up
24. YARN scheduling: Capacity Scheduler and Fair Scheduler queues and resource limits
25. Apache Spark on YARN: spark-submit, executors and Spark SQL over Hive tables
26. Workflow scheduling: Apache Oozie or Apache Airflow for Hadoop job pipelines
27. HDFS administration: balancer, decommissioning DataNodes, quotas, snapshots and fsck
28. Hadoop security: overview of authentication, authorization and data encryption
29. Kerberos: KDC, principals, keytabs and securing HDFS and YARN
30. Authorization and auditing: HDFS ACLs, Apache Ranger policies and audit logs

31. Encryption: HDFS transparent encryption zones with Hadoop KMS and wire encryption
32. Monitoring: JMX metrics, Prometheus exporters and log analysis for daemons
33. Troubleshooting: failed jobs, stuck containers, missing blocks and safe mode

DAYS 4-5 **Advanced: High availability, performance, scale and migration** — New at this level

34. NameNode high availability: active and standby NameNodes, JournalNodes, ZKFC and automatic failover
35. ResourceManager high availability: active and standby ResourceManagers and YARN federation for large clusters
36. HDFS federation: multiple namespaces and router-based federation for namespace scale
37. Erasure coding in Hadoop 3: policies, storage savings and trade-offs versus replication
38. Performance tuning: block size, container memory and vcores, compression codecs and small-file handling
39. Hive performance: Tez engine, vectorization, cost-based optimisation and LLAP
40. HBase at scale: region splitting, compaction, bloom filters and replication
41. Apache Ozone: object store for Hadoop and S3-compatible access
42. Upgrades and rolling upgrades: Hadoop 2 to 3 migration and version compatibility
43. Backup and disaster recovery: DistCp, HDFS snapshots and cross-cluster replication
44. Cloud and lakehouse migration: S3A connector, object storage, Iceberg tables and managed services
45. Capstone: a highly available, secured Hadoop data platform with ingestion, Hive and Spark pipelines

PRACTICE

Hands-on labs

You build these yourself during the course, guided by the trainer.

FOUNDATION

Run a first MapReduce job

Install Hadoop in pseudo-distributed mode on Linux, load files into HDFS and run a word count MapReduce job, then inspect it in the web UIs.

ESSENTIAL

Analyse data with Hive and Pig

Ingest a relational table and log files into HDFS, then clean data with Pig and query partitioned Parquet tables in Hive.

INTERMEDIATE

Build and secure a cluster

Deploy a three-node Hadoop cluster, configure YARN queues, run Spark SQL jobs and enable Kerberos with Ranger policies.

ADVANCED

Highly available production data platform

Configure NameNode and ResourceManager HA, tune a Hive and Spark pipeline, enable erasure coding and replicate data to a second cluster with DistCp.

HOW YOU CAN LEARN

Formats and fees

The same agenda in four formats. Fees are in Indian rupees as of November 2026; the course page shows them in your currency.

FORMAT	DURATION	WHAT YOU GET	FEE
Live & interactive (online)	Max 8-10 hours	Instructor-led public batch of max 10 learners with demos, labs, Q&A and assignment review. Weekday batches start on the 1st of every month, weekend batches on the first Saturday of every month.	**INR 34,999**
1:1 live training (online)	Max 8-10 hours, one-to-one	The same agenda, taught one-to-one by the trainer. Starts immediately, without waiting for the next batch; sessions planned around your working hours, at your pace.	**INR 99,999**
Self-paced video	1-year LMS access	All four Apache Hadoop levels, plus 20+ more courses and 50+ tools, at your own pace.	**INR 833/month** billed yearly at INR 9,996
Corporate (day-wise)	5 days for this level	A private batch for your team, on demand. Foundation 1 day, Essential 2, Intermediate 3, Advanced 5 (cumulative).	**Quote on request**

ASSESSMENT AND CERTIFICATE

Your progress is checked through hands-on labs and assignments at every level. On completion you receive a DevOpsSchool certificate of completion with a unique credential ID that anyone can verify at devopsschool.com/certificates.

INCLUDED WITH EVERY LIVE BATCH

- Every live session is recorded and stays in your LMS for a year.
- Can't attend your batch? Move to next month's batch.
- Free demo class before you pay.
- Free 10-15 minute call with an advisor or the trainer to check the course fits you.

YOUR TRAINER

Learn from Rajesh Kumar

Principal Architect, trainer and mentor in DevOps, SRE, DevSecOps and platform engineering, with 20 years across 8+ software MNCs, including Staff DevOps/SRE Engineer for a leading Japanese mobile-payment platform and Principal DevOps Architect and Manager in Europe.

Profile: www.rajeshkumar.xyz · Videos: youtube.com/@TheDevOpsSchool

20 years in industry	**200+** corporate trainings
10,000+ engineers mentored	**70+** organisations coached

REVIEWS

What learners say

Rated 4.6 on Google · more reviews at devopsschool.com/testimonials and in our video testimonials.

> “Rajesh is very good trainer. Rajesh was able to resolve our queries and question effectively. We really liked the hands-on examples covered during this training program.”
>
> Indrayani, India

> “The training was very useful and interactive. Rajesh helped develop the confidence of all.”
>
> Abhinav Gupta, Pune

START YOUR APACHE HADOOP TRAINING

Join the next batch

Weekday batches start on the 1st of every month and weekend batches on the first Saturday of every month, with max 10 learners each. 1:1 training can start immediately; corporate batches run on demand. We reply within 1 hour.

Pay by UPI, credit and debit cards, net banking or PayPal. Before you pay, read our refund policy. All levels of this course: www.devopsschool.com/courses/hadoop/

ENROL OR RESERVE A SEAT
Apache Hadoop Advanced Training

WHATSAPP · INDIA
+91 99057 40781

WHATSAPP · USA
+1 (469) 756-6329

EMAIL
contact@DevOpsSchool.com