DevOpsSchool

Course curriculum · November 2026

LEVEL 04 OF 04 · ADVANCED

Big Data Tools Advanced Training

Design and operate scalable, governed data platforms with tuned performance, high availability and cost control.

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, 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/bigdata/big-data-tools-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 Big Data Tools course path, you will be able to:

1

Process data at scale

Store and process large datasets with HDFS, Hive and Spark using efficient file formats.

2

Build streaming pipelines

Move and transform real-time data with Kafka, Spark Structured Streaming, Flink and Airflow.

3

Run governed data platforms

Design, secure, tune and migrate lakehouse platforms with governance and cost control.

WHAT THE ADVANCED LEVEL ADDS

Architecture, performance, governance and migration

Hands-on: Governed lakehouse platform. Design a lakehouse with streaming and batch ingestion, catalog and access policies, tuned Spark jobs and monitoring dashboards.

WHO SHOULD ATTEND

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

BEFORE YOU START

- The Intermediate level, or equivalent hands-on experience.
- You can already: Build streaming and orchestrated pipelines on lakehouse table formats, and secure and tune big data platforms.
- 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

Big Data Tools 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, Hadoop and file formats

Included from Foundation

1. Big data concepts: volume, velocity, variety and veracity, and when data becomes big
2. Distributed computing basics: horizontal scaling, data locality, partitioning and fault tolerance
3. Ecosystem map: storage, processing, streaming, query, orchestration and governance tools
4. Hadoop architecture: HDFS NameNode and DataNodes, YARN ResourceManager and NodeManagers
5. Lab setup: a single-node Hadoop and Spark environment with Docker
6. HDFS operations: hdfs dfs commands, replication, block size and permissions
7. MapReduce: the map, shuffle and reduce model and a word-count job
8. Data formats: CSV and JSON versus columnar Parquet and ORC and row-based Avro
9. Batch versus streaming: latency, throughput and typical use cases
10. Prerequisite skills: Linux shell, SQL and Python for data engineering

DAY 2 — Essential: Hive, Spark, Kafka and NoSQL basics

Included from Essential

11. Apache Hive: HiveQL, managed and external tables, partitions and the metastore
12. Spark architecture: driver, executors, cluster managers and lazy evaluation
13. Spark DataFrames: reading, transforming, joining and writing data with PySpark
14. Spark SQL: temporary views, SQL queries and built-in functions
15. Data ingestion: Spark JDBC reads, Apache NiFi flows and bulk loading
16. Apache Kafka basics: brokers, topics, partitions, producers, consumers and consumer groups
17. HBase basics: column families, row keys and random read and write access
18. NoSQL choices: Cassandra, MongoDB and HBase compared for different access patterns
19. Notebooks: Jupyter and Zeppelin for interactive analysis
20. Partitioning and file sizing: partition strategies, small files problem and compaction
21. Data modelling for analytics: star schemas, wide tables and denormalisation

DAY 3 — Intermediate: Streaming, orchestration, lakehouse and security

Included from Intermediate

22. Spark Structured Streaming: sources, sinks, triggers, watermarks and checkpoints
23. Kafka Connect and Schema Registry: source and sink connectors and Avro schemas
24. Apache Flink basics: DataStream API, event time and stateful processing
25. Workflow orchestration: Apache Airflow DAGs, operators, sensors and scheduling
26. Interactive SQL engines: Trino for federated queries across data sources
27. Lakehouse table formats: Delta Lake, Apache Iceberg and Apache Hudi, ACID and time travel
28. Data quality: validation rules and checks with Great Expectations
29. Spark tuning basics: partitions, caching, broadcast joins and the Spark UI
30. Security: Kerberos authentication, Apache Ranger policies and encryption

31. Cloud big data services: Amazon EMR, Google Dataproc and Databricks
32. Spark on Kubernetes: running jobs with the Spark operator
33. Troubleshooting: failed jobs, out-of-memory errors, skew and consumer lag

DAYS 4-5 **Advanced: Architecture, performance, governance and migration** — New at this level

34. Architecture patterns: lambda, kappa and lakehouse architectures and when to use each
35. Advanced Spark performance: adaptive query execution, skew handling, shuffle tuning and file layout
36. High availability: HDFS NameNode HA, Kafka replication and in-sync replicas and Flink checkpoints
37. Scaling clusters: autoscaling, workload isolation and resource queues in YARN or Kubernetes
38. Data governance: catalogs such as Unity Catalog, DataHub and Apache Atlas and data lineage
39. Fine-grained security: column and row-level access, masking and auditing
40. Cost optimisation: storage tiering, spot instances, compaction and right-sizing clusters
41. Migration: moving on-premises Hadoop workloads to a cloud lakehouse
42. Platform observability: metrics, logs and alerts for Spark, Kafka and Airflow
43. Real-time analytics at scale: Apache Druid, Pinot and ClickHouse options
44. Machine learning on big data: Spark MLlib and feature pipelines
45. Capstone: end-to-end governed batch and streaming data platform

PRACTICE

Hands-on labs

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

FOUNDATION

First dataset on Hadoop

Start a Dockerised Hadoop cluster, load a public dataset into HDFS and run a MapReduce word count and HDFS commands.

ESSENTIAL

Batch analytics pipeline

Ingest raw files into HDFS, transform them with PySpark into partitioned Parquet and query results with Hive and Spark SQL.

INTERMEDIATE

Real-time clickstream pipeline

Stream events through Kafka into Spark Structured Streaming, write Iceberg or Delta tables and orchestrate jobs with Airflow.

ADVANCED

Governed lakehouse platform

Design a lakehouse with streaming and batch ingestion, catalog and access policies, tuned Spark jobs and monitoring dashboards.

HOW YOU CAN LEARN

Formats and fees

The same agenda in three 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**
Self-paced video	1-year LMS access	All four Big Data Tools 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). 1:1 training is also available on demand.	**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 BIG DATA TOOLS 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. Corporate and 1:1 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/bigdata/

ENROL OR RESERVE A SEAT
Big Data Tools Advanced Training

WHATSAPP · INDIA
+91 99057 40781

WHATSAPP · USA
+1 (469) 756-6329

EMAIL
contact@DevOpsSchool.com