DevOpsSchool

Course curriculum · November 2026

LEVEL 02 OF 04 · ESSENTIAL

# Apache Spark Essential Training

Build reliable batch data pipelines with DataFrames and Spark SQL, from joins and aggregations to partitioned outputs.

| | | |
|---|---|---|
| LIVE ONLINE<br>**Max 8-10 hours** | CORPORATE (DAY-WISE)<br>**2 days · Days 1-2** | SELF-PACED VIDEO<br>**1-year LMS access** |
| BATCH SIZE<br>**Max 10 learners** | NEXT BATCHES<br>**1st of every month · weekends: first Saturday** | CERTIFICATE<br>**Verifiable certificate of completion** |

| Foundation | Essential | Intermediate | Advanced |
|---|---|---|---|
| 1 day · Included | 2 days · This curriculum | 3 days · Next step | 5 days · Later |

## INSIDE THIS CURRICULUM

| | |
|---|---|
| What you will achieve | Outcomes and level focus |
| Who should attend | Audience and prerequisites |
| Agenda | 22 topics · Foundation to Essential |
| Hands-on labs | 2 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/spark/spark-and-essential-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 Spark course path, you will be able to:

1

### Process data at scale

Write PySpark and Spark SQL jobs that transform large datasets reliably.

2

### Tune and stream confidently

Diagnose slow jobs, tune shuffles and joins and build Structured Streaming pipelines.

3

### Run Spark in production

Operate Spark on Kubernetes or managed platforms with orchestration, security and monitoring.

WHAT THE ESSENTIAL LEVEL ADDS

### DataFrames, Spark SQL and data engineering

**Hands-on:** Batch ETL pipeline with Spark SQL. Build a spark-submit job that joins orders and customers, applies window functions and data quality checks and writes partitioned Parquet tables.

WHO SHOULD ATTEND

- Learners who know the basics of Apache Spark and want to use it confidently every day.
- Professionals taking over Apache Spark work on a live project.
- Teams agreeing on a common way of working with Apache Spark.

BEFORE YOU START

- The Foundation level, or equivalent hands-on experience.
- You can already: Understand how Apache Spark distributes work and run your first PySpark and Spark SQL jobs on a local setup.
- 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 ESSENTIAL

# Apache Spark Essential Training agenda

22 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: Spark architecture, setup and first jobs

Included from Foundation

1. Big data processing basics: batch vs streaming, distributed storage and why Spark replaced classic MapReduce
2. Spark architecture: driver, executors, cluster managers, jobs, stages and tasks
3. Spark components: Spark Core, Spark SQL, Structured Streaming, MLlib and GraphX
4. Installation: Java, Spark binaries, PySpark via pip, environment variables and spark-shell or pyspark
5. SparkSession: creating a session, configuration options and the local[*] master
6. RDD basics: parallelize, map, filter, reduceByKey and collect
7. Lazy evaluation: transformations vs actions and how the DAG is built
8. DataFrame basics: creating DataFrames, schemas, select, filter, withColumn and show
9. Reading and writing data: CSV, JSON and Parquet with read and write options
10. Spark SQL basics: temporary views and running SQL queries against DataFrames
11. Spark UI: jobs, stages, storage and executors tabs for a first job

## DAY 2 — Essential: DataFrames, Spark SQL and data engineering

**New at this level**

12. Schemas and data types: StructType, explicit schemas vs inference and handling nulls
13. Column expressions: built-in functions, when and otherwise, string, date and timestamp functions
14. Aggregations: groupBy, agg, pivot and rollup and cube
15. Joins: inner, outer, semi and anti joins and avoiding duplicate column problems
16. Window functions: ranking, running totals and lag and lead over partitions
17. User-defined functions: Python UDFs vs pandas UDFs vs built-in functions
18. File formats: Parquet, ORC, Avro and compression codecs
19. Partitioned writes: partitionBy, save modes, bucketing and small-file problems
20. Catalog and tables: managed vs external tables and the Hive metastore
21. Data quality checks: deduplication, validation rules and quarantine of bad records
22. Notebooks and scripts: Jupyter, packaging jobs and spark-submit options

PRACTICE

# Hands-on labs

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

FOUNDATION

**First PySpark data job**

Install PySpark locally, load a CSV dataset into a DataFrame, clean and aggregate it with DataFrame and SQL queries and write the result as Parquet.

ESSENTIAL

**Batch ETL pipeline with Spark SQL**

Build a spark-submit job that joins orders and customers, applies window functions and data quality checks and writes partitioned Parquet tables.

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 Spark levels, plus 20+ more courses and 50+ tools, at your own pace. | **INR 833/month** billed yearly at INR 9,996 |
| **Corporate (day-wise)** | 2 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 SPARK 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/spark/

ENROL OR RESERVE A SEAT
**Apache Spark Essential Training**

WHATSAPP · INDIA
**+91 99057 40781**

WHATSAPP · USA
**+1 (469) 756-6329**

EMAIL
**contact@DevOpsSchool.com**

**Next step after this level: Apache Spark Intermediate Training** - Tune Spark jobs, process streams with Structured Streaming and manage lakehouse tables with Delta Lake or Iceberg.