

About Jumping Rivers

Established in 2016, Jumping Rivers is a leading provider of data science, analytics, and programming training. Our trainers and consultants bring over fifty years of combined experience across R, Python, SQL, and a wide range of modern data technologies. We have worked with some of the largest organisations in the world, spanning industries such as finance, healthcare, government, environmental science, technology, and higher education.

At Jumping Rivers, we pride ourselves on delivering high-quality, expert-led training that is practical, engaging, and immediately applicable to real-world workflows. All courses are designed and delivered by experienced data scientists who bring both industry knowledge and teaching expertise.

Jumping Rivers Delivery Style

We offer training in two flexible formats to suit your team's needs:

Onsite Training

- Delivered at your offices
- 1 full day (7 hours) per course
- Interactive, immersive, perfect for team collaboration

Online Training

- Delivered via secure virtual meeting platform
- Two half-day sessions (3.5 hours each) per course
- Ideal for distributed teams or remote-first organisations

Training Material

- Slides
- Demo scripts and exercise files
- A live trainer guides the session, and attendees can download all course content—including slides, demos, and exercises.

Option 1: Databricks Foundations for Data Science & ML Workflows

1. Introduction to Databricks

- What Databricks is and where it fits in the data platform ecosystem
- Workspace overview: clusters, notebooks, repos, workflows
- Databricks Lakehouse concept and how it supports DS/ML teams
- Typical use cases & architecture patterns

2. Working with Data Sources

- Connecting to **Snowflake** and **S3**
- Reading/writing data in notebooks (**Python**)
- Data exploration workflows
- Using Databricks File System (DBFS)
- Hands-on: Load a dataset, explore, and create basic transformations

3. Feature Creation & Experiment / Model Development

- Creating experimental features using notebooks + Databricks UI
- Using internal Python libraries inside Databricks
- Managing dependencies in clusters
- Delta Lake features for versioning & schema evolution
- Introduction to MLflow tracking
- Hands-on: Build simple features & track experiments

4. Collaboration, Promotion, and CI/CD

Light, introductory level – as requested

- Notebook collaboration: repos, versioning, and Git integration
- How promotion typically works from dev → test → prod
- CI/CD basics using **GitHub Actions** with Databricks Repos
- Overview of Databricks Workflows (jobs, scheduling)

5. Governance, Auditability, and Traceability

- Feature lineage: what is available in Databricks today
- Unity Catalog overview
- Data access controls, permissions, and table-level security

- Tracking transformations, model versions, and experiments
Limitations and what may require external tooling

Optional: Day 2 Extension

A hands-on session, including:

- Building an end-to-end mini pipeline
- Setting up a repo + simple CI/CD run
- Working with Delta Live Tables
- **Using MLflow to train, log, and register a model (Hands-on only)**

Option 2: Intro to LLM Ops in Databricks

Outline

- Overview of Databricks LLM capabilities
- Using DBRX, MosaicML, and model serving
- Creating and deploying LLM endpoints
- Prompt engineering basics in notebooks
- Managing fine-tuning and evaluating LLMs
- Understanding governance, security, and responsible AI
- Example mini-project: using an LLM endpoint for internal automation

This course is suitable even if the team is "just learning" Databricks — we keep it conceptual and practical.

Pricing

Workshop Development + Delivery Costs

1. Slides
2. Exercises
3. Hands-on demos and interaction
4. Based on up to 16 attendees
5. Cost Excl VAT

Training Options & Costs

Option	Description	Time	Cost (Excl VAT)
1: Databricks Foundations for Data Science & ML Workflows	1-day delivery covering core Databricks workflows	1 day	£3,200
Optional Add-On (Hands-On Session)	Full practical session including: • Building an end-to-end mini pipeline • Setting up a repo + simple CI/CD • Working with Delta Live Tables • Using MLflow to train, log & register a model	1/2 day	£2,000
2: Introduction to LLM Ops in Databricks	Overview of LLM operationalisation within Databricks workflows	1 day	£3,200