

KOENIG SOLUTIONS · COURSE OUTLINE

Azure Databricks for R-Based Data Analysis & Engineering (Microsoft)

Learn to run R workflows in Azure Databricks, from notebooks and Delta tables to reproducible, governed analytics pipelines.

3 Days

8 HOURS A DAY

6

MODULES

35

TOPICS

COURSE OVERVIEW

This course is designed for R users who want to work with large-scale data using Azure Databricks without moving away from familiar R-based workflows. It introduces Databricks from an R user's perspective, focusing on how distributed computing and Delta Lake can be leveraged for data preparation, analytics, and research use cases, especially in healthcare and scientific domains. Participants will learn how to run R code in Databricks notebooks, prepare and manage large datasets, build analysis-ready data pipelines, and perform advanced analytics efficiently at scale. The course also emphasizes reproducibility, governance, and best practices to ensure that research and analytics workflows remain reliable, auditable, and easy to maintain in collaborative environments.

WHO THIS COURSE IS FOR

R Programmers

Data Analysts

Data Engineers

Research Scientists

Healthcare Data Analysts

Analytics Teams

PREREQUISITES

- Participants should have a working knowledge of R, including experience with data frames or tibbles, data cleaning, and basic data analysis.
- Familiarity with common R packages such as dplyr, tidyr, and ggplot2 is recommended.
- Basic understanding of data concepts such as tables, files, and datasets will be helpful.
- Prior experience with SQL, cloud platforms, or distributed systems is beneficial as minimal and practical exposure to these topics will be covered where necessary for data access and workflow integration.

TOOLS AND REQUIREMENTS

Labs Requirement

- Participants must use their own Azure subscription / Databricks environment.

What You Will Learn

- Explain Azure Databricks architecture and concepts for R users
- Write and run R code within Databricks notebooks
- Read data frames and tibbles from tables and files at scale
- Clean, transform and structure data into analysis-ready datasets
- Apply Delta Lake features such as versioning and time travel
- Perform advanced aggregations and feature engineering in R
- Prepare datasets for modelling, inference and downstream reporting
- Plan reproducible, governed R workflows for collaborative teams

COURSE SYLLABUS

6 Modules, 35 Topics

01

Azure Databricks Through an R User's Lens

TOPICS

- What Azure Databricks is — explained for R users
- How Databricks complements R & RStudio workflows

WHEN TO USE:

- R
- Minimal SQL
- Distributed compute
- High-level Databricks architecture (only what R users need)
- Typical healthcare & research analytics workflows
- What R users must know — and what they can safely ignore

02

Working with R in Databricks Notebooks

TOPICS

Databricks notebooks for R users · Writing and running R code in Databricks · Working with data frames and tibbles at scale · Reading data from tables and files into R · Minimal SQL exposure (only for data access)

03

Data Preparation & Engineering Using R

TOPICS

- Data ingestion patterns from an R perspective
- Cleaning and transforming data using R
- Handling missing values, inconsistencies, and outliers
- Creating reusable, analysis-ready datasets
- Structuring datasets for long-term use

04

Advanced Dataset Management with Delta

TOPICS

- Delta tables explained for R users
- Reading and writing Delta data from R
- Versioning and time travel for research use cases
- Updating datasets safely without breaking analyses
- Performance considerations for large datasets in R

05

Advanced Analytics in R on Databricks

TOPICS

- Advanced aggregations and summaries in R
- Feature engineering for statistical analysis
- Preparing datasets for modelling and inference
- Working with large datasets efficiently in R
- Exporting results for downstream reporting

06

Reproducible R Workflows, Governance & Best Practices

TOPICS

- Designing reproducible R workflows in Databricks
- Refreshing datasets without breaking analyses
- Organising R projects for team use
- Common failure patterns and how to avoid them
- Dataset documentation and data dictionaries
- Practical checklist for “production-ready” research datasets
- Version control concepts for data & analysis (practical, light)

NOTES

- Azure provides a free subscription valid for 30 days.

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