

Implementing a Data Analytics Solution with Azure Databricks (DP-3011)

Duration: 1 day

Audience: Professionnel TI|Analyste de données

Who is this training for?

Data engineers, advanced data analysts, solution architects, developers, and IT professionals.

Training objectives

- Explore and use the Azure Databricks environment (workspaces, clusters, notebooks)
- Ingest, transform, and analyze data at scale with Apache Spark (DataFrames, SQL, PySpark)

Summary

Learn how to harness the power of Apache Spark and high-performance clusters on the Azure Databricks platform to run data engineering workloads at scale in the cloud.

Course outline

- Explore Azure Databricks
- Use Apache Spark in Azure Databricks
- Use Delta Lake in Azure Databricks
- Use SQL warehouses in Azure Databricks
- Run Azure Databricks notebooks with Azure Data Factory

Approach and methodology

Practical and structured approach combining focused theory and guided workshops. Participants gradually implement a Data Analytics solution with Azure Databricks through real-world exercises inspired by business scenarios, promoting immediate application of learning. They learn how to process, transform, and analyze data at scale by leveraging the capabilities of Spark and Databricks' collaborative tools. Led by a Microsoft certified trainer, the training focuses on interactivity and

the development of directly transferable technical skills to design effective analytical solutions in a professional context.

Prerequisites

- Azure Basics (AZ-900 or equivalent content).
- Basic knowledge of Python and data visualization with matplotlib/seaborn.

Recommendations

- Basics in data analysis (cleaning, transformation, visualization)
- Knowledge of Python and/or SQL
- Understanding of data processing concepts (batch, big data)
- Familiarity with Azure and data services Notions of Spark or distributed processing (asset)