

Mouhammad

DIAKHATE

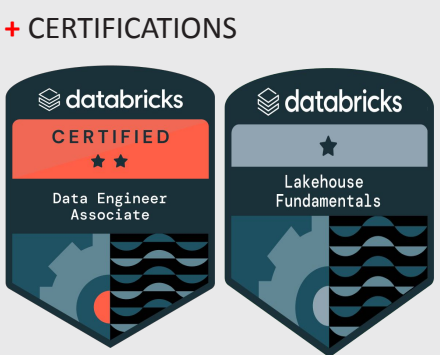
Senior Data Engineer

+ ABOUT ME

Data & Cloud Solutions Architect since 2020, Databricks Certified. I design scalable Lakehouse architectures, from customer discovery to delivery, across AWS, Azure and GCP.

+ CONTACT INFOS

- +33 6 51 92 68 62
- mouhammad.diakhate12@gmail.com
- Paris and Lyon
- French (native), English (fluent)



+ SKILLSET

DATA

CLOUD

LEAD

+ EDUCATION

2022

Master Degree

Data Science at Université Claude Bernard Lyon 1

2021

Engineering degree

Diploma in computer design engineering from Ecole Polytechnique de Thiès

2017

Preparatory classes

Preparatory classes at Ecole Polytechnique de Thiès

PROFESSIONAL EXPERIENCE

Bedrock Streaming - Current Position

Senior Data Engineer - Freelance

01/2026 – Now

Context: Analyze and improve data platform for Europe customers (France, Hungary, Germany and Netherland)

Achievements:

- Collaborate closely with customers to migrate existing data workloads to Databricks
- Leading migration strategy for 300+ batch legacy workloads to Databricks streaming pipelines
- Maintain legacy workloads & Make existing Data Platform IaC's more user friendly

Technical environment and aws services :

Databricks, Lakeflow Declarative Pipelines, Terraform, Spark, AWS (S3, Glue, Athena), Airflow (Legacy)

50Hertz (Elia Group) - 6 months

Senior Data Engineer - Freelance

08/2025 – 01/2026

Context: (POC) Challenged against another team on how to build Elia's settlement application

Led the team rebuilding from scratch using Spark to prove feasibility while ensuring reliability

Achievements:

- Worked closely with tech and business stakeholders to validate accuracy of results, scalability of infra, and performance of computations
- Designed and built from scratch the Spark-based settlement engine
- Built all of it on a low-budget on-prem infrastructure, only 1 node of 4-core/16GB
- Delivered 25 settlement products (5 Levies + 20 ADREA)
- Stress-tested infra on 4M+ records: -15mn total execution time (-6min computation + cold start)
- Demonstrated 78% functional and architectural coverage within PoC scope (even if I met some challenges)

Technical environment and aws services :

Apache Spark, Python, Gitlab CI, Argo CD, Kubernetes, Spark Operator, Docker, Helm, Minio, UV

SOGELINK / Groupe SEB - 1 year 1 month

Data Architect - Freelance

11/2024 – 12/2025

Context: Led Sogelink's architecture improvement and help Groupe SEB maintain and evolve their data platform

Achievements:

- Sogelink
 - Advised stakeholders on migration strategy from a legacy Data Lake to a modern Lakehouse
 - Led the migration of the Data Lake to Apache Iceberg
 - Integrated new CRM/ERP data sources from Dutch and Norwegian subsidiaries (API, OData)
 - Refactored Terraform IaC to streamline developer experience
- Groupe SEB
 - Developed API ingestions for new online and retail marketplaces (Spain, Switzerland, Saudi Arabia, Canada, Romania and Colombia)
 - Maintain and enhance existing Retail pipelines (prod monitoring, incident management, alerts to stakeholders)

Technical environment and aws services :

Apache Spark, Terraform, Python, Pandas, AWS, Apache Iceberg, Gitlab CI, AWS Glue,, Jenkins

CMA-CGM - 4 months

Data Engineer Consultant

05/2024 – 09/2024

Context: Give the direction on how to redesigne the legacy EDI preprocessing to a new robust data processing tool for container stowage optimization at CMA-CGM

Achievements:

- Built an EDI parser and data model, applying SOLID design patterns
- Refactored legacy business rules, cutting code volume by 66%
- Designed a "restow" algorithm to optimize heavy-container placement
- Prepared a clear migration strategy and usage guidelines for internal developers transfer knowledge

Technical environment and aws services :

Apache Spark, Python, Pandas, AWS Lambda, S3, DynamoDB, Gitlab, XLDeploy, Design Patterns

ENGIE - ~1 year and 4 months

Data Engineer Consultant

01/2023 – 04/2024

Context: Migrated legacy BI from SQL Server (AWS RDS) to a Databricks Lakehouse architecture

Roles & Achievements:

- Designed a hybrid data vault and medallion Lakehouse architecture on Databricks
- Built a 3-hop architecture with an SCD2 reliability layer and Change Data Feed for incremental ingestion
- Migrated legacy SQL Server workloads to production-ready gold tables for Power BI and SSRS
- Delivered a data lineage layer and full functional documentation for all migrated jobs
- Optimized production performance through Spark UI monitoring and tuning

Technical environment and aws services :

Databricks, Apache Spark, Delta Lake, Databricks Workflows, Delta Live Tables, S3, Python, SQL Server, Power-BI

PROFESSIONAL EXPERIENCE

Capgemini - ~1 year

Data Engineer Consultant

02/2022 – 11/2022

Context: Advised on new KPI development for the Transilien T-REX project

Achievements:

- Designed cost-effective Azure cloud architecture and secured cloud resources for the project
- Managed Databricks and Data Factory pipeline scheduling, execution and monitoring
- Built a star-schema Data Warehouse on Azure, orchestrated via Data Factory
- Migrated the CD pipeline artifact repository from Nexus to JFrog Artifactory

Technical environment and azure services:

- Scala, **Azure Databricks**, Spark, Azure Data Factory, Azure SQL, Azure Storage,
- Azure Pricing Calculator, Transact-SQL (T-SQL), Power BI, Python

Orange - ~1 year

Junior Data Engineer & DevOps

11/2020 – 08/2021

Context: Automated Cloud infrastructure provisioning and observability of application logs

Achievements :

- Automated Cloud infrastructure provisioning with Terraform, Bash and Nginx scripts
- Built a serverless Kafka pipeline for log ingestion, with source/sink connectors
- Delivered a self-service gateway and template registry for infra provisioning requests

Technical environment and google services:

- Elasticsearch, Knative, OpenFaas, Docker, Kubernetes, React.js, Apache Kafka, Jenkins, Python,
- Google Cloud Platform (GCS, GKE, Cloud DNS, Cloud Run, Cloud Load Balancing, GCE)