

Abiy Dema

📍 Lille, France

<https://www.linkedin.com/in/abiy-dema> |  <https://www.abiydema.com>

+33 7 68 38 75 09 | <https://github.com/Haabiy> | abiy.dema@gmail.com

🎓 EDUCATION



École Centrale de Nantes

Group of the Centrale Schools

Nantes, France

Top 3 French school of Engineering

October | 2018 - 2022

BSc, Engineering

💼 PROFESSIONAL EXPERIENCE

Decathlon

📍 Lille, France

September 2024 – Present

Data Engineer

- Developed dktbricks, Python library to enable the team to launch JAR files on Databricks via a terminal-based UI, with configurable options for cluster size and number of worker nodes.
- Optimised Scala-based data pipelines for the core marketplace data product, efficiently handling large datasets with techniques such as Z-Ordering and Liquid Clustering for faster queries, ensuring smooth data flow, improving performance, enhancing system scalability, and reducing Databricks costs by around **15%**.
- Built a Spark-based ETL pipeline following the Medallion Architecture: implemented API data extraction via AWS Lambda into the Bronze layer, developed Scala-based transformations for Silver and Gold layers with Delta Tables on Databricks, and orchestrated workflows using Airflow for scheduling and triggering jobs across the cluster.

Mangabey

📍 Paris, France

January 2023 – August 2024

Data Engineer | Backend Engineer

- Designed Python Django-based backend system for MINE application, a data marketplace, featuring APIs for three key components namely notification centre, update centre, and delivery system, tailored to meet end-user needs.
- Engineered the backend architecture and core functionalities of the MINE application, optimising Spark-based ETL and non-Spark pipelines, which enhanced data processing speed by **40%** on the Amazon EMR cluster.
- Implemented Apache NiFi workflow for real-time data ingestion, seamlessly transferring data from S3 to Redshift.
- Constructed Apache Spark ETL based data pipeline solution utilized by **AstraZeneca** for executing a comprehensive data quality assessment workflow. This included extracting data from Amazon S3, conducting in-depth analysis, storing the derived insights in Redshift, and establishing seamless real-time visualisation in POWER BI through direct query integration.
- Completed a KPI analysis framework using Pandas, later optimising with Polars to achieve a remarkable ~ **90%** reduction in execution time, while also orchestrating multiple data pipelines using Apache Airflow within Docker containers for streamlined and efficient data processing.

🛠 PERSONAL PROJECTS

- Developed a PyPI package emrrunner that lets users trigger both Spark and non-Spark jobs directly from the terminal on Amazon EMR. [Link ↗](#)
- Developed a Python terminal-based remote access tool that creates a shell on the target machine with reverse tunneling, enabling access from another machine and supporting commands like file management, google search, screen recording, and camera access. [Link ↗](#)

⚙ SKILLS, INTERESTS & LANGUAGES

Soft skills : Analytical Problem Solving, Intellectual Curiosity, Collaborative Teamwork, Strong Work Ethic, and Clear Communication.

Hard skills: Python, SQL, Apache Spark (PySpark, Scala, & SQL), Apache Airflow, Apache NiFi, AWS EC2, Amazon Redshift, Amazon EMR, AWS Lambda, Databricks, Power BI, HTML, CSS (Tailwind), Bootstrap, React js, Flask, Django, C++, MATLAB, and Excel VBA.

Interests & Hobbies: Football, cooking, and open-source projects.

Languages: Amharic – Native, English – Proficient, French – Intermediate (B1)

