

Mohamed Sahnoun

+216 99 520 965 | mohammed.sahnoun@gmail.com | [/in/mohamedsahnoun](https://in/mohamedsahnoun) | github.com/MedNoun | mohamedsahnoun.com

Data Engineer with 3 years across GCP data engineering, data platform work and infrastructure-as-code. Builds production ELT lakehouses, dimensional models and from-scratch protocol implementations, with a bias toward deterministic, testable pipelines.

EXPERIENCE

Data Engineer

Jan. 2025 - Present

Olivesoft (Client: Marc Jacobs)

United States (Remote)

- Built a BigQuery/Dataform ELT lakehouse from scratch (83 SQLX models, ~6,600 lines of GoogleSQL) integrating 9 source systems (Salesforce Commerce Cloud, Cegid Y2, Klaviyo, Anaplan, Dynamics AX) through a Medallion architecture into Kimball star schemas; cut reporting latency from 24 hours to minutes and compute cost by ~40% via partitioning and incremental processing.
- Developed a deterministic Master Data Management engine implementing a Union-Find connected-components algorithm purely in SQL via iterative CTEs, resolving 7M+ fragmented customer records into stable golden IDs.
- Traced silent delta drift to non-deterministic deduplication and made every incremental load reproducible run-to-run; enforced data contracts with 22 Dataform assertion suites and 100+ DAG tags for selective execution.
- Maintained the Terraform IaC codebase (~125 files, ~5,000 lines) provisioning BigQuery datasets, ingestion flows, Pub/Sub and 21 least-privilege IAM bindings across dev/ppd/prd via approval-gated Azure DevOps pipelines; resolved state drift with `moved` blocks and lifecycle rules.
- Internally at Olivesoft: technical mentor for end-of-studies engineering interns through to defence, and built a certification training platform with a Python/Selenium scraping engine extracting 1,500+ structured exam questions from JavaScript-heavy DOMs, served by a responsive SPA on Vercel.

Data Engineer Intern

Feb. 2024 - July 2024

OVHai LLM

Paris, France (On-Site)

- Re-architected the platform preparing training corpora for Transformer-based neural machine translation models into a Message-Driven Architecture of four containerised Python microservices (ActiveMQ, MinIO, EdgeDB, Docker), chosen after a documented monolith-vs-microservices and EDA-vs-MDA trade-off analysis.
- Replaced hard-coded callbacks with a pluggable system (Dependency Inversion, Open/Closed) and introduced batch-level error isolation capturing failing records with tracebacks, eliminating stop-the-world pipeline restarts.

OPEN SOURCE & PROJECTS

edgedb-dto: Open-Source Library | *Python, Jinja2, Pydantic, EdgeDB*

Jun. - Jul. 2024

- Published a code-generation library to PyPI (MIT, 1,000+ downloads, sole author) emitting Pydantic-validated DTOs from EdgeQL queries, combining static AST analysis with runtime `inspect/importlib` introspection and Jinja2 templating to restore ORM-style object composition to EdgeDB.
- Built a recursive resolver persisting nested object graphs depth-first in a single transaction; automated SemVer tag-gated GitLab CI releases dual-publishing to PyPI and a private registry. [Repository](#)

Seedify: BitTorrent Client from Scratch | *TypeScript, Node.js*

2023

- Implemented the full peer wire protocol, rarest-first piece selection, endgame mode, tit-for-tat choking and SHA-1 piece verification; hand-rolled HTTP and BEP 15 UDP tracker transports behind a shared interface (~1,800 LOC, pair-programmed). [Repository](#)

Nesty-kerberos: Kerberos Protocol from Scratch | *NestJS, Redis, PostgreSQL*

2023

- Implemented the three-legged ticket flow across KDC (AS/TGS) and application services: nested encryption, Diffie-Hellman registration, IP binding, clock-skew validation and Redis replay protection TTL'd to ticket lifetime. [Repository](#)

Machine Learning Engineer: Piano Performance Assessment | *Fellowship.ai (Launchpad.ai)*

Apr. - Jul. 2022

- Co-authored a published music information retrieval system scoring student piano performances: neural automatic music transcription with hidden Markov model MIDI-to-MIDI alignment against reference scores, reaching a **95.8% note-classification F-score**, beating Benetos et al. (92.9%) at far lower compute. [Article](#)

EDUCATION

National Institute of Applied Sciences and Technology (INSAT)

Tunis, Tunisia

Master's Degree in Software Engineering & B.S. in Computer Science

Sept. 2019 - Sept. 2024

TECHNICAL SKILLS & CERTIFICATIONS

Languages: Python, SQL (GoogleSQL), JavaScript (ES6+), TypeScript, Bash, EdgeQL

Data & Cloud: BigQuery, Dataform, GCP (Cloud Run, Pub/Sub, Dataflow, Cloud Build, Workflows), Airflow, Terraform, Docker

Practices: ELT, Dimensional Modelling (Kimball), Medallion Lakehouse, CDC, MDM, Data Quality, IaC, CI/CD (Azure DevOps, GitLab CI), Datadog

Certifications: Google Cloud Associate Data Practitioner (2025); Apache Airflow 3 Fundamentals, Astronomer (2025)

Spoken: Arabic (Native), French (Fluent), English (Professional), German (A2)