Issam Salmi

Engineering Student in Information Systems and Big Data - Data Engineering
LinkedIn: issamsalmi — **Phone:** +212669081511
Email: issamsalmi55@gmail.com

Profile

Engineering student in Information Systems and Big Data specialized in Data Engineering, with proven experience in building production-ready data pipelines and platforms. Skilled in batch data integration, workflow orchestration, data quality management, and cross-functional collaboration with Data Analysts and Scientists. Capable of implementing DevOps practices, managing data access governance, and translating business requirements into scalable technical solutions. Committed to continuous improvement and delivering high-quality data solutions for strategic projects.

Education

- 2021–present : National School of Applied Sciences, Berrechid Engineering Degree in Information Systems and Big Data
- 2019–2020 : Ibno Yassine High School, Sidi Yahia Baccalaureate, Physical Sciences, Honors with High Distinction

Key Skills

- **Cloud Data Platforms**: GCP (BigQuery, Dataflow, Cloud Composer), Azure, AWS, Snowflake
- **Databases & SQL**: SQL Server, PostgreSQL, BigQuery — advanced SQL optimization, indexing, data modeling
- **Data Pipelines & Integration**: ETL/ELT development, batch processing, multi-source data flows, workflow automation
- **Orchestration & DevOps**: Airflow, CI/CD pipelines, Azure DevOps, Git, Docker, monitoring and error handling
- **Data Quality & Governance**: Data validation, data freshness monitoring, integrity checks, access rights management
- **Big Data Technologies**: Spark, Spark SQL, Kafka, Hadoop, Hive, real-time and batch processing
- **Programming**: Python (advanced), SQL (advanced), Java
- **BI & Reporting**: Power BI dashboards, KPI pre-calculation, datamart design
- **Data Modeling**: Star schema, dimensional modeling, Medallion Architecture (Bronze/Silver/Gold)
- **Collaboration**: Cross-team collaboration with Data Analysts, Scientists, and business stakeholders

Professional Experience

AI & Data Engineering Intern *3D SMART FACTORY*   **Hybrid** *07/2025 – 09/2025*

- Designed and deployed automated multi-agent data pipelines for ingestion and transformation of Microsoft Forms
- Implemented data quality checks and structuring workflows with multilingual OCR for text and image data
- Developed indexing solutions in Elasticsearch with workflow automation for rapid data analysis
- Ensured data integrity and freshness through monitoring and validation processes
- **Key Technologies**: Python, LangChain, Ollama, EasyOCR, Elasticsearch, Selenium, JSON, Linux

Data Engineer Intern *Finea CDG*   **On-site** *07/2024 – 09/2024*

- Developed and optimized ETL pipelines for multi-source data ingestion, transformation and batch integration
- Collaborated with business and IT teams to ensure proper data utilization and align technical solutions with requirements
- Optimized Talend/Pentaho workflows and created Power BI dashboards for business reporting
- Designed star schema data models for data warehouse and implemented performance improvements
- **Key Technologies**: Talend, Pentaho, Python, PostgreSQL, SQL, Power BI, Linux

Projects

- **Cloud Data Platform & CI/CD Integration (GCP)** – Design and implementation of a BigQuery-based data platform with automated pipelines using Dataflow and Cloud Composer. CI/CD integration with Git, batch monitoring, and a Medallion architecture ensuring data quality, governance, and scalability.
  Technologies: BigQuery, Dataflow, Cloud Composer, Python, SQL, Git, DevOps

- **Enterprise Pipeline & Performance Optimization** – Development of a robust pipeline for data ingestion and transformation from APIs, SQL databases, and files. Performance optimization through indexing, parallelization, centralized error handling, and data freshness monitoring.
  Technologies: Python, Airflow, Docker, SQL, Spark, PostgreSQL, CI/CD

- **Enterprise Data Warehouse – Star Schema & Governance** – Design and implementation of a CRM/ERP data warehouse using a Bronze/Silver/Gold architecture, automated ETL processes, access control management, and optimized analytical views for BI reporting and business KPIs.
  Technologies: SQL Server, SQL, ETL, Data Modeling, Data Governance

Certifications

- Oracle Cloud Infrastructure 2024 – Generative AI Certified Professional (Oracle)
- AWS Cloud Quest: Cloud Practitioner (AWS)
- Machine Learning Specialization (DeepLearning.AI)
- Associate Data Engineer in SQL (DataCamp)
- Google Cloud Skills Boost — 30+ labs (Qwiklabs) completed
- *Google Cloud Professional Data Engineer (GCP) — In progress*

Languages

- Arabic (Native), French (Fluent), English (Fluent)

Volunteering

- President & Lead, Google Developer Groups On Campus (2024 - Present)