

Professional Summary

Senior Software and Data Engineer with 10+ years of experience designing scalable, production-grade systems – spanning event-driven data infrastructure, petabyte-scale ETL pipelines on GCP and AWS, and full-stack backend applications. Skilled in architecting LLM-powered and agentic AI workflows, including RAG pipelines, shared AI infrastructure, and observability systems. Proven track record optimizing mission-critical systems across fintech, energy, and logistics – reducing ML inference latency, automating code reviews with AI-assisted tools, and delivering robust solutions.

Professional Experience

Data Engineer - AMEX | Apr 2024 - Present

- **Architected** and **deployed** enterprise-scale event-driven data infrastructure on **GCP**, leveraging **Pub/Sub** for **real-time** message routing with dead-letter queue strategies that achieved **zero-loss** event processing while reducing data latency from **hours** to sub-minute resolution
- **Established Apache Airflow** as the enterprise **orchestration** standard on **Google Cloud Composer**, managing 70+ production DAGs and successfully migrating legacy batch workflows while optimizing DAG performance through advanced monitoring and bottleneck elimination
- **Designed** and **implemented** end-to-end **ETL** pipelines using **Google Cloud Dataflow** to process petabyte-scale datasets, architecting unified data models that consolidated fragmented source systems into centralized **Big Query** warehouse supporting large scale analytics
- **Built** and maintained a comprehensive **Jenkins-based CI/CD** pipeline encompassing automated unit and integrations testing, data **quality validation**, **code coverage**, enforcement and infrastructure as code-provisioning – **significantly improving**
- **Optimized** infrastructure and resource utilization by automating **GCP** compute instance provisioning and developing custom Python Script for high-performance data processing, while implementing performance **benchmarking** and capacity planning methodologies to reduce costs
- **Accelerate** critical **backup** workflows through **automation**, reducing **processing** time by 98% (from 2h to under 60 sec) via optimized data pipeline design
- **Leverage AI-assisted developer tools** – including **GitHub Copilot**, **ChatGPT** and **Claude** – to accelerate DAG development, automate boilerplate generation for ETL pipelines, and streamline code review and SQL query optimization, improving overall engineering productivity across the data platform team.

Sr. Software Engineer - TRANSIT HARIECHE IMPORT & EXPORT | Dec 2022 – Mar 2023

- **Developed Java-based** data **visualization** and processing **microservices** for handling **large-scale** file ingestion, ensuring data integrity and enabling efficient parallel processing across multiple platform instances
- **Built** scalable backend infrastructure supporting **50K** concurrent users with **zero downtime** using **Python/Django** with robust **authentication**, **order** processing, and **payment** management
- **Designed RESTful APIs** for **inventory**, **cart**, and **order** management with secure **JWT/OAuth2** authentication and integrated payment gateways (**Stripe & PayPal**), enabling multi-method transactions
- **Implemented** real-time **asynchronous** task processing using **Celery** and **Redis** for order **confirmation**, **refunds**, and inventory updates, improving system **responsiveness** and **reliability**
- **Developed ML-powered** product recommendation engine using user browsing and **purchase** history to deliver personalized suggestions, enhancing customer **engagement** and conversion

- **Create** responsive **full-stack** frontend using **HTML**, **CSS**, **JS** and **Bootstrap** with **Ajax** for **real-time** cart operations discount handling, and profile management across all devices
- **Implemented** CI/CD pipelines with **Jenkins & GitHub CI** for automated **testing**, **building**, and **deployment** to ensure smooth and frequent releases into production

Sr. Software Engineer - GAZPROM | Oct 2020 – Nov 2022

- **Built** end-to-end **ML** data platform with **Django backend** supporting user **authentication**, multi-format file uploads (CSV, XLSX, Parquet, JSON...), and **automated** data **cleaning/filtering/transformation**
- **Designed** responsive **full-stack** interface using **HTML**, **CSS**, **JS**, **jQuery** and **Ajax** enabling **real-time** data visualization with **matplotlib**, **Seaborn** and **Plotly** while providing immediate user feedback on uploads and processing
- **Engineered** microservice-oriented architecture with Docker containerization, RESTful APIs for authentication and file processing and Celery/Redis for asynchronous background tasks – reducing data preparation time by 10%
- **Implemented** secure data persistence using PostgreSQL/MySQL for user metadata and file information, with Django session management and dynamic template rendering for seamless data visualization
- **Led** and **Developed** **Java-based** high-performance inventory service handling **real-time** stock **synchronization** across multiple regional **warehouses**, reducing inventory sync latency by **60%** and preventing overselling incidents

Sr. Full-Stack/Backend Engineer - ITMO | Dec 2018 – Sep 2020

- **Architected** **ML-Powered** pipeline monitoring platform using **Flask**, **Python** and **ML** libraries like (**scikit-learn**, **TensorFlow**, **NLTK**...), that improved decision-making accuracy by **20%** over manual sensor monitoring with real-time predictions
- **Built** Scalable backend infrastructure with **RESTful APIs**, **Celery/Redis** for **asynchronous** processing, and **Redis** Caching to handle large-scale datasets while maintaining sub-second response time
- **Implemented** Comprehensive data processing pipeline using **NumPy** and **Pandas** for feature engineering, coupled with **ML** models to classify pipeline conditions (water saturation vs oil readiness) from **real-time** sensor data
- **Established** CI/CD infrastructure using **Jenkins/Gitlab CI** with automated testing and deployment processes, while containerizing the application with **Docker** for consistent dev-to-prod on **AWS (EC2 & S3)**
- **Led** platform adoption and operation and operational efficiency by providing comprehensive training and docs to enable field engineers and technicians to assess pipeline health independently

Software Engineer - TRANSIT HARIECHE IMPORT & EXPORT | Dec 2015 – Nov 2018

- **Built** real-time chat application using **Flask**, **Flask-SocketIO** and **WebSocket** technology enabling low-latency bidirectional communication with secure **JWT**-based user authentication
- **Designed** Scalable backend infrastructure with **AWS RDS MySQL** for message storage, **Redis** for session management and message brokering, and **S3** for media serving – supporting high-throughput chat operations
- **Implemented** **Firebase Cloud Messaging** for **real-time** push notifications on new messages and events, enhancing user engagement and in-app communication responsiveness
- **Secured** the platform with **ClaudeFlare** for **DDoS** protection and **AWS** code pipeline with integrated logging and monitoring via **AWS CloudWatch** for proactive issue detection and system optimization

Education

Master of Science in Machine Learning & Data Engineering

September 2020 – May 2022

Bachelor of Science in Artificial Intelligence & Image Processing

September 2015 – May 2018

Technical Skills

Cloud & Data Infrastructure: GCP (Pub/Sub, Dataflow, Composer), AWS (EC2, S3, RDS, Redshift), Kafka

Data Engineering & Orchestration: Apache Airflow, Google Cloud Dataflow, PySpark, AWS Glue

Databases & Warehouses: BigQuery, PostgreSQL, MongoDB, Cassandra, Amazon S3

Languages & Frameworks: Python, SQL, Java, JavaScript, JQuery, Ajax, Shell Scripting | Flask, Django, FastAPI

Backend & DevOps: REST APIs, Microservices, Docker, Jenkins, Git, CI/CD, AWS CloudFormation

Data Science & ML: scikit-learn, Pandas, NumPy, TensorFlow, PyTorch, HuggingFace, RAG

Certifications

<u>Generative AI Engineering with LLMs Specialization</u>	IBM	2026
Apache Spark With Scala	Udemy	2025
Machine Learning	DeepLearning.ai	2024
Deep Learning	DeepLearning.ai	2024