JEFF SANTOS

Senior Data & AI Engineer | Data Ingestion & Integration
+351 911 012 161 | Lisbon, Portugal (Remote)
jmsapps@outlook.com | jmsapps.ca | linkedin.com/in/jmsapps | github.com/jmsapps

SUMMARY

Senior engineer and technical delivery lead specializing in enterprise data ingestion, integration, and LLM-powered retrieval platforms. Designs and owns end-to-end ETL pipelines that extract structured, semi-structured, and unstructured content (PDF, HTML, Word, multimodal documents), normalize it into governed intermediate representations, and load AI-ready datasets into vector and hybrid search indexes. Leads through influence: sets architecture, delegates implementation across engineering teams, mentors developers, and drives scope with product and business stakeholders in regulated, high-stakes environments.

SKILLS

Data Engineering: ETL/ELT Pipeline Design, Multi-Format Document Ingestion, Schema Normalization, Incremental & Idempotent Processing, Data Quality Controls & Validation, Dataset Versioning, Lineage & Rollback, Pipeline Observability, Cost & Throughput Optimization
AI & Retrieval: RAG, Multimodal Document Processing, Semantic Chunking, Vector Embeddings, Hybrid Lexical/Vector/Semantic Retrieval, Retrieval Evaluation, Structured LLM Outputs, Azure AI Search, Azure Document Intelligence
Languages: Python, SQL, TypeScript, JavaScript, Swift, Nim
Data & Integration: PostgreSQL, MySQL, MongoDB, NoSQL, GraphQL, Hasura, ORM, REST, SOAP, OData, JSON, XML
Cloud & DevOps: Azure, Google Cloud (Cloud Run, Cloud Scheduler), AWS (EC2, S3), Docker, Terraform, GitHub Actions, CI/CD, Linux, Nginx
Backend & Frontend: Django, FastAPI, Tornado, NestJS, ExpressJS, Node.js, React.js, React Native, Redux, Svelte, TypeScript

EXPERIENCE

Senior AI Engineer (Data Ingestion & Integration) — August 2025 – Present

Tactable — *Toronto, ON*

- Technical delivery lead for an enterprise **RAG data platform** serving a financial services client; owned architecture and technical outcomes, delegated implementation across a team of engineers, and drove scope and timelines with product leadership and business stakeholders.
- Designed and built a multi-format **ETL pipeline** that extracts **HTML, PDF, and Word** source documents, normalizes them into a common **intermediate representation**, and loads retrieval-ready Markdown datasets into **Azure AI Search**.
- Engineered the transform layer as a hybrid of **deterministic Python** parsing and **multimodal LLM** interpretation, generating **spatially aware figure, chart, and table descriptions** that make non-textual content retrievable by downstream text and vector search.
- Implemented **semantic chunking** on topic boundaries and analyzed chunking, indexing, and retrieval strategies against retrieval quality to tune RAG performance.
- Established **data quality controls** for AI-ready datasets: schema-validated structured LLM outputs, source-evidence grounding, and completeness/consistency checks that gate low-confidence enrichment before load.
- Built **incremental ingestion with cached LLM enrichment**, cutting processing cost and accelerating pipeline runs by up to **1,000**×.
- Delivered **Azure AI Search** index infrastructure supporting lexical search, vector retrieval, semantic ranking, and metadata filtering across English and French datasets.
- Engineered **versioned ingestion** with staged updates, archived snapshots, and transactional rollback on commit failure, providing traceability and recoverability for published knowledge datasets.
- Applied **Spec-Driven Development** to define scope and acceptance criteria up front; documented pipeline architecture and mentored engineers to scale delivery under tight deadlines.

Software Engineer (Compliance & Integrations) — January 2025 – August 2025

Tactable — *Toronto, ON*

- Led system design and delivery for a **regulated compliance product**, translating business requirements into data models, integration strategy, and an implementation roadmap for the development team.

- Built **enterprise system integrations** over SOAP and REST (Yardi Voyager 7S, MSI), landing third-party XML/JSON data into PostgreSQL through a Hasura/NestJS backend with custom SQL.
- Refactored a 3,000+ line monolithic Python **data ingestion script** into a modular, documented pipeline with **100% test coverage**, stabilizing onboarding of new data sources.
- Built a Python-based **MongoDB migration version control system**, enabling repeatable, auditable schema migrations across environments.
- Implemented CI/CD with GitHub Actions and provisioned Cloud Run services via **Terraform** on GCP, plus local test harnesses for scheduled pipeline endpoints.
- Drove Python/Django performance work: reduced a critical endpoint from 16s to 0.5s (**32× faster**; -96.9% latency).

Lead Software Engineer, Full Stack June 2023 – January 2025
Switch Health *Toronto, ON*

- Redesigned the relational database structure of a regulated health platform while preserving data integrity across **4M+ user records**.
- Delivered major performance improvements on an **Azure** cloud solution, including **200% faster** API response times in critical data paths.
- Cut feature implementation time from **two weeks to one day** by building a reusable pipeline for onboarding new test types and internal tooling.
- Enabled B2B integrations through a White Label platform, **doubling profitability**.
- Mentored junior developers, ran code reviews, and set engineering standards across the product lifecycle.

Software Engineer II, Full Stack January 2023 – June 2023
Switch Health *Toronto, ON*

- Transformed the platform into a White Label solution, enabling B2B integrations and improving business scalability.
- Migrated state management to Redux Toolkit, reducing bundle size by 60% and development time by 25%.
- Led a UI redesign from Figma to production, collaborating closely with Product and QA to ship quality releases.

Software Engineer I, Full Stack April 2022 – December 2022
Switch Health *Toronto, ON*

- Designed and developed RESTful APIs for complex diagnostic test types, optimizing back-end integration for high-throughput data.
- Restructured the project into a well-organized module hierarchy and reduced technical debt, accelerating development by 20%.
- Diagnosed and resolved critical production issues, improving system stability and performance.

PROJECTS

Deterministic ETL Pipeline Framework | *Caravel*

- Built a Python framework for declaring and running **deterministic data pipelines** from typed Pipeline/Stage/Step/Branch primitives, with dataset-owned I/O across local and cloud object storage (`abfs://`, `s3://`, `gs://`) via fsspec.
- Implemented partitioned dataset types and a multi-source loader with **branch routing by source** for fan-out/fan-in across heterogeneous inputs, validating the full plan before mutating any output.
- Built a plugin system for production concerns: **schema-versioned checkpointing** with output reuse for incremental runs, cleanup of stale outputs, immutable run-evidence logging for **lineage**, and advisory leases for writer-conflict detection.

Reactive Python UI Framework | *PYSX*

- Built a reactive UI framework for Python 3.14+ where signals, derived values, and effects drive targeted DOM patches from the server, with t-string templates keeping markup and live state together.
- Applied **Spec-Driven Development** to define the template grammar, wire protocol, and recovery semantics ahead of implementation.
- Shipped scoped styling, standalone and ASGI hosting, runnable playgrounds, and a VSCode extension providing syntax highlighting and template diagnostics.