JOHN LEGASPI

Senior Data & Platform Engineer

London, UK • +44 7916 450026 • jmlegaspi1@outlook.com • github.com/Eco-ex

• https://www.linkedin.com/in/john-mark-legaspi-ba465bb3/

PROFESSIONAL SUMMARY

Senior Data & Platform Architect specializing in high-scale data infrastructure and agentic AI workflows. Pioneer in LLM-orchestrated engineering, having designed multi-agent frameworks for McLaren's Oracle-to-lakehouse migration—reducing a 20-day cycle to 5 days (75% reduction) while slashing compute costs by 90%. Expert in building autonomous MCP-based systems, Python/PySpark orchestration, and Medallion architectures. Proven track record of merging platform engineering rigor with AI governance to drive measurable business impact. Seeking Staff or Lead roles to own AI-infrastructure strategy and next-generation data architecture.

TECHNICAL SKILLS

Languages & Frameworks: Python (PySpark, asyncio, Pandas, Pydantic), SQL (Postgres, T-SQL, PL/SQL), TypeScript

Data Platforms: Databricks, Delta Lake, dbt, Medallion Architecture, Change Data Capture, Star Schema modelling

AI/LLM Engineering: LLM agent orchestration, few-shot prompting, prompt caching, batched & resumable API workflows, RAG systems

Cloud & DevOps: Azure (Fabric, Azure SQL), Azure DevOps, GitHub Actions

Orchestration: Apache Airflow (custom operators, DAGs, SLA enforcement, failure callbacks)

BI & Analytics: Power BI (DAX, Power Query), data visualization and KPI design

PROFESSIONAL EXPERIENCE

Senior Data & Platform Engineer — McLaren Construction | Jul 2024 – Present
Promoted from Data Engineer in Dec 2025

Role Mission: Directing the architecture of McLaren's AI-native data platform. Pioneered a custom multi-agent orchestration framework to automate complex engineering lifecycles, serving as the technical blueprint for the group's legacy-to-cloud transformation and AI governance.

Agentic AI Workflows

- Multi-Agent Orchestration: Architected a framework of specialized agents to translate technical and functional requirements into production Microsoft Fabric workspaces and artifacts.
- Ecosystem Integration: Leveraged Fabric and Azure DevOps MCP servers to automate environment provisioning and full-lifecycle ALM, including context-aware ticket generation (Epics to Tasks).
- Domain-Specific Agents: Engineered dedicated agents for documentation assistance, MCP interface management, and high-performance Python/PySpark notebook generation.
- Engineering Governance: Established "Human-in-the-Loop" protocols, automating PR creation via ADO while maintaining manual review cycles for code and complex Power BI report design.
- Standardized Knowledge Encoding: Embedded team-specific best practices and design principles into agent knowledge bases to ensure high consistency across automated engineering outputs.

- Prompt Optimization: Implemented Token-Oriented Object Notation (TOON) across agent communication layers, significantly reducing token overhead and improving prompt precision.

AI-Augmented Legacy Migration

- Architected and deployed an LLM-agent framework that reduced the end-to-end Oracle-to-lakehouse migration cycle from 20 days to 5 days (**75% reduction**).
- Engineered custom agents to programmatically resolve legacy schema relationships and dependency metadata, automating cross-referencing and type validation previously done by hand.
- Implemented automated translation of legacy PL/SQL API decoders into optimised SQL CASE expressions, preserving functional parity and full data lineage.
- Orchestrated automated generation of production-ready Delta table artifacts, mapping legacy framework columns to modern lakehouse conventions.

Platform Engineering & Cost Optimization

- Cut ingestion times from 1 hour to 2 minutes using asyncio concurrency and optimized PySpark operations—reducing cloud compute spend by 90% through compute/IO separation and first-principles optimization.
- Architected a custom API ingestion framework with circuit breakers and exponential backoff, eliminating cascade failures and throttling errors across upstream integrations.
- Delivered 20+ production pipelines processing millions of records daily, with Pydantic-based schema validation and CDC for incremental loading.
- Integrated JSON-structured logging and the Four Golden Signals (latency, traffic, errors, saturation) for real-time observability of business-critical systems.

Architecture & Leadership

- Established the team's Bronze/Silver/Gold Medallion structure in Delta Lake, standardizing data processing and reducing ad-hoc data requests by 40%.
- Built Azure DevOps CI/CD pipelines that automated deployments and enforced 100% code review coverage, shortening release cycles from weeks to days.
- Developed production-grade Airflow DAGs with custom failure callbacks and SLA enforcement, ensuring 99.9% pipeline reliability.
- Automated Row-Level Security (RLS) and IAM synchronization between Azure SQL and Entra ID.

Business Intelligence & Impact

- Architected analytics solutions and custom KPIs that contributed to McLaren winning 'Digital Contractor of the Year' 2026.
- Built full-stack Canvas Apps with React-based components and T-SQL stored procedure backends for high-performance data manipulation.
- Designed standardized Power BI dashboards using Star Schema and Snowflake modelling, and custom DAX KPIs to monitor subcontractor performance and operational efficiency.

Technical Project Manager — Asante Media | Jan 2024 – Jul 2024

- Bridged business stakeholders and engineering teams, translating client requirements into functional specifications and technical documentation.
- Owned planning, execution, and monitoring of multiple concurrent web projects, ensuring on-time and on-budget delivery.

Front-End Developer — Asante Media | Aug 2021 – Jan 2024

- Engineered responsive components and integrated Algolia search-as-a-service APIs, optimizing for high-performance retrieval and real-time interaction tracking.
- Built reusable, extensible web components using AEM's Java-based component framework in cross-functional Agile teams.

SELECTED PROJECTS

Clinical Decision Support Tool — github.com/Eco-ex/medical-diagnostic-tool
Full-stack web app helping healthcare professionals manage patient records and analyze treatment options with AI assistance. Built with React, TypeScript, Express, and OpenAI API. Demonstrates ability to ship production AI systems from end-to-end.

WhatsApp Data Extraction Pipeline — github.com/Eco-ex/whatsapp-extract
Python CLI that parses WhatsApp chat exports and extracts structured records via the Claude API. Features few-shot prompting, prompt caching, batched & resumable calls, YAML-configurable schemas, and rapidfuzz-based deduplication. Showcases LLM orchestration and production-grade error handling.

EDUCATION & CERTIFICATIONS

- **BSc Environmental Science** — Queen Mary University of London (2:1)
- **Data Analyst in SQL** — DataCamp