
Carlos D. Escobar-Valbuena

Principal AI Data Architect & Engineering Lead

carlosdavidescobar@gmail.com +57 301 7758620 Bogotá, Colombia

broomva.tech linkedin.com/in/broomva github.com/Broomva

SUMMARY

AI Engineering Lead and data platform architect with 7+ years building **AI-native, multi-tenant data platforms** in production — across regulated and high-stakes domains: fintech / cross-border payments, enterprise procurement (FIFA 2026 + Olympics Committee scale), and oil & gas SCADA. Concurrent senior leadership across three roles for the last 18 months: AI Lead at Stimulus (Databricks-backed agentic procurement), Data Architect contractor at TEAM International, and Co-founder & CTO at Wedi Pay (agentic payments on Kafka + Databricks, through Jun 2026).

Strong on the AI-Ready stack: production **RAG** (pgvector, BM25 hybrid retrieval, LLM-as-judge), **governed agent orchestration** (LangGraph, MCP, policy-bounded tool adapters with task memory), **lakehouse** architecture (Databricks Unity Catalog, Snowflake, Iceberg), **audit-grade governance** (RBAC, Postgres RLS, KMS-per-tenant, lineage via OpenTelemetry, full audit trails), **CDC and event sourcing** (Dagster, Kafka/Redpanda, append-only journals), and multi-tenant isolation at enterprise scale. Open-source author of `vigil` (OTel-native Rust observability crate). MSc in AI at Universidad de los Andes (2026).

SKILLS

Lakehouse & Data Platforms: Databricks (deep) — Unity Catalog RBAC, Zerobus streaming, medallion architecture (bronze / silver / gold), Delta Lake, PySpark; Snowflake, Iceberg, Trino, Kafka / Redpanda, Dagster, Azure Data Factory, MLflow; AWS Athena / Glue / Lake Formation / EMR Serverless / OpenSearch (architectural)

AI / RAG / Agentic: Anthropic Claude, OpenAI, AWS Bedrock, LangChain, LangGraph, LangSmith, Langfuse, MCP, pgvector, FAISS, BM25 hybrid retrieval, LLM-as-judge evaluation, Vercel AI SDK

Databases: PostgreSQL (incl. pgvector + RLS), Delta Lake, Snowflake, AWS RDS, MySQL, MongoDB, Redis, NeonDB, MSSQL / T-SQL

Observability & Governance: OpenTelemetry (authored `vigil` Rust crate), CloudWatch, Datadog, data contracts, lineage tracking, Unity Catalog RBAC, Postgres RLS, KMS-per-tenant, multi-tenant isolation, JWT auth hardening

Cloud & Infra: AWS (S3, EC2, RDS, Lambda, IAM/KMS, IoT Core, Greengrass), Azure (AKS, ML, Digital Twins, DevOps), GCP, Kubernetes, Terraform, GitHub Actions

Languages: Python, Rust, TypeScript / JavaScript, SQL; Java (read & review proficiency, no production ship)

ML / MLOps: PyTorch, PyTorch Lightning, TensorFlow, Hugging Face Transformers (fine-tuning), MLflow, SageMaker, Azure ML, Prophet, TDSP-aligned MLOps

Spoken Languages: English (C1, full professional), Spanish (native)

EXPERIENCE

AI Lead *Stimulus — Supplier-Intelligence Procurement SaaS*

Nov 2024 – Present

Leading the architecture redesign of a Databricks-backed agentic procurement platform serving customers including FIFA 2026 and the Olympics Committee.

- Architected the real-time data substrate around **Databricks Zerobus** for streaming ingestion into a **Unity Catalog**-governed **medallion architecture** (bronze / silver / gold) — a lakehouse-native pattern that eliminates external bus dependencies while preserving lineage and per-tenant column-level access controls.
- Redesigned the product foundation around **best-in-class data governance and managed access to artifacts**, with state sharing and agentic flows across the system. Delivered enterprise-grade features at production scale.
- Owned LLM deployment and tuning across Claude and OpenAI variants: auto-failover, AKS/EKS comparisons, cost-latency trade-off decisions, hardened rate limiting and **JWT authentication for multi-tenant clients**.
- Delivered agent flows combining **multi-document retrieval with predefined policies, task memory, and deterministic tool adapters** for supplier research, risk checks, and sourcing recommendations.
- Mentored engineers on prompt patterns, guardrails, and evaluation harnesses; coordinated AI brownbag sessions.

Co-founder & CTO *Wedi Pay — Agentic Cross-Border Payment Orchestration*

Oct 2024 – Jun 2026

Led platform architecture and AI/ML engineering for an AI-native, multi-tenant B2B payment platform unifying banking infrastructure with blockchain rails.

- Designed a modular Turborepo monorepo: Next.js + FastAPI + LangGraph agentic services, with a **NeonDB / Postgres data layer using Row-Level Security** for stringent multi-tenant isolation.
- Architected a **Kafka / Redpanda event-driven backbone** integrated with Databricks asset bundles for gold-table ETLs feeding both application models and analytics — producing real-time insights into payment flows.
- Defined Wedi Agents: an in-product assistant that interprets historical activity and, under strict policy / consent / scoped credentials, executes payment actions (e.g., USDC transfers) with full audit trails.
- Validated the revenue model against LATAM benchmarks (Global66, DolarApp, Wise), mapping it to CFO/finops reporting needs.
- Signed and shipped payment-rail integrations on a provider-agnostic layer: **Prometeo** (open-banking B2B account-to-account), **TruBit / VelaFi** (USDC on/off-ramp, multi-country), **Thirdweb** (web3-native flows).

TEAM International

Nov 2020 – Present

Microsoft Gold Partner; named among IAOP Global Outsourcing 100 and Software 500 companies. Concurrent contracts since Nov 2024.

Data Architect (Contract)

Nov 2024 – Present

- Architectural oversight on big-data pipelines I previously designed; supporting and enhancing them as workloads scale.
- Architecting next-generation iterations to meet new business and technical requirements (higher throughput, new data sources, advanced analytics).
- Overseeing platform migrations with minimal downtime and full data integrity.

Sr. Machine Learning Engineering Tech Lead — Digital O&G Solutions

Sep 2022 – Nov 2024

- Owned the Time Series Engine product end-to-end — a multi-tenant streaming + batch telemetry engine on **Databricks + MLflow + PyTorch + Prophet** under TDSP-aligned MLOps. Anchor customer: a **NYSE-listed US natural-gas operator with ~70,000 wells** across multiple US basins. Production lakehouse with **medallion-style bronze / silver / gold tables since 2021**.
- Led the development team: backlog, estimation, risk and feasibility, architectural specification across multi-quarter roadmaps.
- Built a specialized ML package for forecasting and anomaly detection (PySpark + MLflow distributed learning), productized as the SCADALite Well Manager offering with bilingual capability decks and tiered pricing.
- Shipped and operated a multi-tenant telemetry solution handling **10,000+ records/sec** with concurrency and workload distribution across stream / batch pipelines and customizable ETL logic.
- Implemented a digital-twin framework via Azure Digital Twins for asset management and operational insights across thousands of wellsites.

Machine Learning Engineering Tech Lead

Aug 2021 – Sep 2022

- Led big data streaming solutions: distributed processing/storage, real-time inference for forecasting and anomaly detection (PySpark, Prophet) for O&G safety and production.
- Owned the Time Series Engine product backlog and architectural roadmap; designed and implemented an orchestration engine for dynamic task creation and execution.

Data Engineer

Nov 2020 – Aug 2021

- Designed, tested, and shipped a **Snowflake** data warehouse for streaming IoT device data — a unified source of truth for O&G analytics. Defined the migration from legacy DWH using Talend + Azure Data Factory.
- Led an R&D team building Anomaly Detection, Alarm Threshold Estimation, and Time Series Inference engines (Azure ML, scikit-learn, TensorFlow, PyTorch, PowerBI).
- Engineered a Snowflake pipeline transforming semi-structured IoT data into structured DWH models.

AI Engineer *Independent Engagements & Selected Project Deliverables*

Aug 2023 – Present

Anonymized engagements; case studies on request.

- **AWS Premier Consulting Partner (US-based AI advisory):** agentic RAG over an internal SageMaker engineering knowledge base — pgvector + FAISS hybrid retrieval, three switchable interaction modes (graph-based agentic RAG, tools agent, conversational RAG), LangServe + Chainlit. Delivery accepted at evaluation stage.
- **Tier-1 LATAM e-commerce + fintech platform:** Polars-based ML ranking pipeline for the recommendation carousel (prints / taps / payments feature engineering, GridSearchCV-tuned classifiers + regressors, MLflow autolog). Class-based, reproducible, batch + incremental.
- **Public-sector procurement intelligence platform:** TF-IDF baseline + LangGraph agentic RAG (FAISS + retriever tools) to match upcoming government project tenders against subcontractor capabilities and certifications.
- **LLM-enhanced CDC pipeline** (Dagster + Supabase + LangChain + Twilio) with WhatsApp interaction surface; **Spark SQL LLM Agent** on Databricks Delta under **Unity Catalog RBAC** for role-scoped agent-driven analytics.

Digitalization Leader *INSA Ingeniería S.A.S*

May 2019 – Nov 2021

Siemens Solution Partner. End-to-end digitalization, automation, and instrumentation across O&G, energy, and discrete manufacturing.

- Deployed a digitalization strategy on **AWS IoT Core, Greengrass, EC2**, ThingsBoard, and the Siemens IoT suite, with ISO 27001-aligned cybersecurity — robust IT/OT integration end-to-end.
- Completed digitalization of an analog O&G field with 50+ wells across 3 sites: reduced transportation costs for monitoring by 50% and avoided 2 spills.
- Achieved cost reductions up to 35% across industrial process improvements.

OPEN SOURCE & RESEARCH

Continuously evolving substrate at broomva.tech. Highlights:

- **Life Agent OS** — Rust agent runtime kernel. Multi-tenant gateway (TLS 1.3, JWKS, KMS abstraction across Vault Transit / AWS KMS / GCP KMS), token-bucket rate limiting, event-sourced persistence (Lago, on redb v2), and **OpenTelemetry-native observability (Vigil — crates.io)**. 15+ crates published.
- **Haima** — Agentic finance engine: x402 metered M2M payments with USDC settlement on Base. 6-crate Rust workspace.
- **bstack + broomva/skills** — open-source agentic development harness: 20 irreducible automation primitives (policy gates, knowledge bookkeeping, CI productive-wait, persistent cross-context loops, cross-model adversarial review) + **48 portable agent skills** (agentskills.io spec) compatible with Claude Code, Codex, Cursor, Gemini CLI, Goose, and Copilot.
- **Agentic control architecture** — control-native governance for autonomous workflows: typed schemas, hook-enforced safety shields, verification causally independent of the acting model; ongoing autonomy / search / intelligence research on the RCS substrate.
- **RCS (Recursive Controlled Systems)** — 5-paper series formalizing LLM-as-controller agents as a 7-tuple Σ with hierarchical stability budgets. Foundations paper complete.
- **Spanish/Wayuu Translation Model** — Fine-tuned BART/T5/MT5/Llama-2-7B on HuggingFace for low-resource indigenous-language translation.
- **Hyperopt Prophet, MLFlowOps, Databricks Session** — open-source Python packages for distributed hyperparameter optimization, streamlined MLflow ops, and Databricks session bootstrapping.

EDUCATION

MSc, Artificial Intelligence *Universidad de los Andes*

2026

BSc, Mechatronics Engineering *Universidad de San Buenaventura*

2013 – 2018

GPA 4.1/5.0. Best GPA Academic Distinction (2016). Speaker, IV International Congress on Advanced Mechatronics Technologies (2018) — *On Cascade ADRC for Time-Delay Systems with Subsystems Faults*.

Selected training: Building Transformer-Based NLP Applications (Nvidia); Diplomado ML & Data Science Avanzado (Universidad Nacional); GCP Big Data & ML (Google); AWS Machine Learning (AWS); TensorFlow 2.0 & Deep Learning (SuperDataScience).