

EDUARDO MONTILVA

Senior Software Engineer | Backend, Distributed Systems & Applied AI

Latam · Remote · UTC-4

eduardomontilvag@gmail.com · linkedin.com/in/ejmontilvag · github.com/edumntg · eduardomontilva.com

EXPERIENCE

Software Engineer II, Beast Industries (@MrBeast)

Remote · July 2024 - Present

- Senior-scope engineer on a team of about fifteen, reporting to the Director of Engineering and the Director of Data, AI and ML. Most days start with other people's pull requests. New backend hires usually get me as the person who walks them through the codebase, and questions about Python, Java or the ML side tend to end up with me.
- The company's Python codebase started as an empty repository I was handed. I chose the architecture, how services are laid out, the CI and how they deploy. Today it holds twelve microservices that process thousands of videos, posts and accounts a day, the embedding jobs, and the collection services that feed them through rotating proxy pools. Databricks takes the heavier batch work.
- The embedding pipelines are mine. They used to call third-party model APIs one record at a time and topped out near 60,000 records a day. Moving inference onto our own GPUs in batches got that to about 2 million. Transcript analysis, topic classification, sponsor detection and brand-safety scoring all sit on top of those embeddings.
- Took an AI image and thumbnail generation service from an empty repository to production: model integration, the async job queue, storage and the public API. It handles hundreds of jobs a day and has been stress-tested at a thousand. I got per-job latency down from around 90 seconds to 30.
- Built the services that aggregate and estimate metrics over millions of videos and channels, and a retrieval-augmented generation service that lets more than 2,000 users ask questions of the analytics data in plain language.
- Wrote the design documents behind several of the team's infrastructure choices and saw them through to adoption: Qdrant for vector search, Manticore and Quickwit for full-text search, and the schemas and data contracts for ClickHouse, MongoDB and Cloud Spanner.
- Design event-driven systems on RabbitMQ, backed by PostgreSQL, Cloud Spanner or ClickHouse depending on whether the workload is transactional or analytical. Some of the consumers I reworked now handle up to 100x more messages per second than before, and a set of queries that used to run for over ten seconds come back in 300 ms or less.
- Maintain roughly 70 Helm-deployed services on GKE through Argo CD. About twenty of them run on a deployment pattern and architecture I designed. Moving service-to-service traffic onto private VPC networking cut the network bill sharply. The infrastructure-as-code, ingress, certificate and DNS automation and the OpenTelemetry tracing are mine as well.
- Built the observability around the payments integration: status-code tracking for the payment provider, exception capture, and alerts to email, Slack and Discord. I am on the on-call rotation and respond to production incidents.
- Write the Kotlin authentication services (token issue and refresh, session handling) that the web and mobile clients sign in through, and work across a large Java and Kotlin monorepo covering analytics, data collection, ranking, outlier detection, search, payments and alerting. Currently adding NestJS and TypeScript services alongside it.
- Work on the Next.js frontends and the React Native app when a feature needs someone to take it end to end.

Backend Developer & Project Manager, Talamo Iberica

Remote · November 2022 - July 2024

- Designed and built a blockchain platform with services in TypeScript, JavaScript, Python and C++, including the security and peer-to-peer communication layers.
- Wrote the REST APIs and WebSocket protocols (Node.js, Express, PostgreSQL) carrying peer-to-peer traffic and client-to-backend communication.
- Split the system into microservices to handle several thousand requests a day from the platform's users.
- Planned sprints, task breakdown and workflow for a team of more than 10 developers.
- Set up the CI/CD pipelines used to deploy blockchain nodes and APIs.

Software Engineer, Freelance / Private Contract Work

Remote · February 2019 - November 2022

- Delivered backend systems for e-commerce, blockchain and SaaS clients using Python (FastAPI, Flask), Node.js with TypeScript, Java, C++ and some Go.
- Designed relational and document schemas across PostgreSQL, MySQL and MongoDB.
- Built the data acquisition pipelines that fed downstream machine learning models.
- Trained and shipped models in production: CNNs for image classification, RNNs for text generation, regression models, and transformer-based language models.
- Set up CI/CD pipelines on AWS.
- Wrote React hooks and components for data fetching and state, working with the clients' frontend developers.

Electrical & Power Systems Engineer, Independent Consultant

Contract · February 2019 - July 2022

- Modeled and simulated the SIEPAC Central American transmission network (100+ busbars, 15 substations) in DigSILENT PowerFactory and ETAP, running AC and DC load flow studies.
- Wrote optimal power flow algorithms in Python and MATLAB, including nonlinear and MINLP formulations with equality and inequality constraints.
- Developed stability control strategies for synchronous generators, motors, wind turbines, photovoltaic systems and EV charging stations.
- Designed the low-voltage distribution for a residential building inside the TACOA substation in La Guaira: short-circuit calculations, panel sizing and cable sizing.
- Built transmission line and tower models in PLS-CADD and PLS-Tower, with protective device coordination.
- Designed photovoltaic systems and simulated them in Simulink under realistic operating conditions.

EDUCATION

B.Sc. Electrical Engineering, Universidad Simón Bolívar

December 2018

- Thesis: Analysis of the Electrical Interconnection System of the Central American Countries (SIEPAC Network)

PRODUCTS

OriginStat (originstat.com): Cookieless web analytics that ties every visit to the revenue it produced.

SocialEar (socialear.app): Social listening across Reddit, Hacker News, Lobsters, YouTube, Instagram, TikTok and X.

bakepdf (bakepdf.app): Merge, split, compress, sign and convert PDFs in the browser, without uploading them.

SELECTED PROJECTS

lazymongo. A lazygit-style terminal UI for MongoDB, written in Rust over ratatui and shipped as a single 3 MB binary. Documents stream through cursor batches under a hard memory cap, so opening a hundred-million-document collection costs what opening a small one does, and every query runs off the render path so the interface never blocks on the network.

Powersys platform. Web-native power system simulation platform written in Rust. Power flow, OPF, short-circuit, transient, EMT and unit-commitment solvers run as independent binaries over a shared domain core, with a sparse-Jacobian Newton-Raphson implementation and event-driven job dispatch.

powersys. Python package for power system analysis, optimization and dynamics. Published to PyPI; the Rust platform above grew out of it.

gnn-opf. Power system state estimation through optimal power flow using graph neural networks. The grid is already a graph, so the inductive bias comes for free.

claude-broker. AMQP-based inter-session messaging for AI coding agents, written in TypeScript over a pseudo-terminal wrapper, so several agents working on the same problem can message each other.

Unit commitment with RL. Unit commitment solved with reinforcement learning on Stable-Baselines3 instead of a mixed-integer solver.

js-mat / wtalkie. TypeScript packages for matrix mathematics and safer WebSocket communication between APIs and backends.

TECHNICAL SKILLS

Languages	Python, Java, Kotlin, TypeScript, JavaScript, Rust, C, C++, Go, SQL, MATLAB, Bash
Backend	Spring, NestJS, FastAPI, Flask, Express, Node.js, REST, WebSockets, gRPC, RabbitMQ, Event-driven, Microservices
Data	PostgreSQL, Cloud Spanner, ClickHouse, MySQL, MongoDB, Redis, Databricks, BigQuery, Supabase, Qdrant, Manticore, Quickwit
ML	PyTorch, TensorFlow, Stable-Baselines3, Embeddings, RAG, NLP, CNNs, RNNs, Transformers, GNNs
Optimization	Newton-Raphson, Interior point, L-BFGS, SGD, Adam, Convex optimization, Nonlinear programming, MINLP, Genetic algorithms, Optimal power flow
Cloud	GCP, AWS, Kubernetes, Helm, Argo CD, Docker, Terraform, CI/CD, OpenTelemetry, Grafana
Frontend	React, Next.js, Vue, React Native
Engineering	DigSILENT PowerFactory, ETAP, PLS-CADD, PLS-Tower, Simulink, AutoCAD
Spoken	Spanish (native), English (professional)