Enmanuel Cuadros Gomez

Lima, Peru enmanuel.cuadros.ai@gmail.com 925 605 754 enmacuadroswork.com
linkedin.com/in/ecuadrosg36 github.com/ecuadrosg36

Summary

Senior data engineer who works like a scientist, with nearly 6 years diagnosing and fixing messy data problems with dbt and the modern data stack, forming hypotheses, running experiments and evals, and proving each fix with before-and-after results.

Builds through coding agents such as Claude Code and Codex as the main way of working, with project rules, subagents, and hooks that force review, and checks the data instead of trusting the output. Designs LLM-based extraction and enrichment from text and images with traceability and validation, ingests data from APIs and web sources, and orchestrates dbt models with Dagster and Airflow.

Works async with English-speaking US teams, writes high-signal findings with the conclusion first, and reasons from first principles on new problems, owning them end to end without needing to be guided.

Experience

Senior Software Engineer, GlobalLogic – USA (Remote) May 2026 – present
- Diagnose data quality issues for a US-based digital advertising client by tracing data from report back to source, including an 87% loss of impression data, and cross-check the same metric across independent sources before trusting any number.
- Checked whether a table was still being written before prioritizing an 84% error rate, and showed it was 17 months of stale history rather than an ongoing loss.
- Built a Python tool that hashes production and non-production tables and drills down from day to row to cell where they differ, giving before-and-after evidence across 172.9M rows in a 70-pipeline migration.
- Made vendor file ingestion robust to renamed files by matching IDs from a reference sheet instead of hardcoded file prefixes, validating input at the point of parsing.
- Keep a third-party API sync idempotent, pushing 51 audience segments weekly to the Yahoo DSP API and checking existing records before each create or update.
- Use Claude Code as the main way of building, with a spec-driven workflow of project rules, subagents, and hooks that encode review rules and data-safety checks before an agent changes code.

Senior Data Engineer & AI Consultant, Independent Consultant – Lima, Peru (Remote) Oct 2025 – present
- Recovered approximately 96 to 97% of affected records after a change in composite user identifiers broke downstream matching, tracing the wrong mappings back to raw data and correcting the transformations.
- Deduplicated events before dbt merges, keeping the latest record per ID by source update and ingestion timestamps, and resolved identities across sources with entity matching rules.
- Design LLM-based extraction and enrichment from text and images with the Anthropic and Gemini APIs, with structured outputs, source traceability for every field, and validation rules that reject inconsistent values.
- Run evals and experiments to compare approaches, scoring extraction accuracy against labeled samples with DeepEval and reporting before-and-after results.
- Ingest data from APIs and web sources with automated discovery and Python scrapers on Playwright and BeautifulSoup, handling missing fields, inconsistent formats, and rate limits.
- Maintain dbt models with tests and orchestrate them with Dagster software-defined assets, with CI/CD in GitHub Actions and database management on PostgreSQL and Snowflake.
- Design human-in-the-loop review steps where researchers confirm low-confidence records, and use their corrections to automate more of the process.

Big Data Engineer, Indra Peru – Lima, Peru Sep 2024 – Sep 2025
- Implemented automated data quality gates using dbt and Python across 15 critical pipelines.
- Migrated legacy ETL processes to Databricks, reducing execution time by 30%.

Data Engineer, Vooxell Group SAC – Lima, Peru Sep 2023 – Sep 2024
- Built and optimized BigQuery data warehouses with partitioning and clustering, reducing costs and latency by 30%.

Data & ML Engineer, Tenmas SAC – Spain (Remote) Oct 2023 – Jul 2024
- Built and maintained data pipelines in AWS using Python, Spark, and SQL.

Data Engineer, Neoris Peru SAC – Lima, Peru May 2023 – Dec 2023
- Migrated ETL workloads to Snowflake, improving performance and reliability.

Database Developer (Data Engineer Role), Indra Peru – Lima, Peru Jun 2022 – Jul 2023
- Automated ETL workflows processing over 10 million records per month.

Business Intelligence Consultant (Data Engineer Role), Chain Services TI SAC – Lima, Peru — Mar 2022 – Jun 2022
- Migrated databases and pipelines to AWS using Python and AWS Glue.

Business Intelligence Analyst (Data Engineer Role), ManpowerGroup Peru – Lima, Peru — Nov 2020 – Mar 2022
- Automated reporting pipelines, reducing operational effort by 40%.

Projects

RAG Agent for Social Impact, GDG Lima Build with Gemma Competition – Lima, Peru — Aug 2026
- Built a LangGraph agent on Gemma through the Gemini API for primary healthcare and disaster-prevention questions that routes each question, retrieves context from Peru's national open-data platform indexed in FAISS, generates an answer, and runs a grounding check as an anti-hallucination guardrail before responding.
- Used multilingual Hugging Face sentence-transformer embeddings and published a self-contained Kaggle notebook as the competition deliverable.

Education

BS **ESAN University**, Systems Engineering – Lima, Peru — Jan 2021 – Dec 2025

DP **ESAN University / ESAN Graduate School of Business**, International Diploma Corporate Finance – Lima, Peru — Jun 2020 – Oct 2020

BS **National University of San Marcos**, Accounting – Lima, Peru — Jan 2011 – Dec 2016

Certifications

- Databricks Certified Data Engineer Associate
- Professional Data Engineer Certification (Google Cloud)
- AWS Certified Data Engineer Associate
- Snowflake Data Engineering
- Databricks Academy Accreditation – Generative AI Fundamentals

Skills

Data Quality & Experimentation: Root cause analysis, cross-source validation, deduplication, entity resolution, schema validation, evals, before-and-after comparisons of pipeline versions

Modern Data Stack: dbt with tests and incremental models, Dagster, Airflow, Snowflake, BigQuery, PostgreSQL, Databricks

LLM Enrichment: Anthropic API, Gemini API, structured outputs, extraction from text and images, DeepEval, LangGraph

Ingestion: REST APIs, web scraping with Playwright and BeautifulSoup, Airbyte, Python, SQL

Coding Agents: Claude Code with project rules, skills, subagents, and hooks, Codex, Cursor

DevOps: CI/CD with GitHub Actions and Jenkins, Docker, Terraform, Git

Languages: Spanish (Native), English (Advanced)