

WORK EXPERIENCE

Data Engineer**Roxom****March 2026 – October 2026**

- In two months, migrated a serverless setup to a medallion lakehouse (bronze, silver, gold) on Athena and Iceberg with dbt, covering several areas of the company. Moved 90+ raw tables from JSON to typed Iceberg.
- Owned 100+ Airflow 3 DAGs (Astronomer Cosmos) on EKS, with KubernetesExecutor and deploys that rebuild only the changed dbt models. Cut the core pipeline from 32 to 19 minutes.
- Wrote ingestion for 15+ exchange, custody, and banking APIs, writing directly to Iceberg. Managed Airbyte as code with Terraform, with schema-drift and sync-recovery alerts.
- Built gold KPI models (growth funnel, cohorts, sales fee attribution, reconciliations) and a semantic layer. Added data-quality checks and alerts when something goes wrong.
- Reduced cloud cost and made a core report about three times faster, from about 300 seconds to 90. Built Airflow plugins the team uses for data health, cost, and reports, and automated the recurring business reports.

Lead Teacher**Le Wagon****June 2025 – Present**

- Lead teacher for the data engineering courses in Latin America. I teach the classes, run the demos, and lead student projects on GCP that integrate tools such as Airflow, dbt, and BigQuery.

Data Engineer**Mutt Data****December 2019 – February 2026**

- Helped build ETLs to migrate data from Teradata to Athena, with data-validation checks, and added Google Sheets reports to the pipeline.
- Taught Apache Airflow and Apache Spark seminars at Digital House.
- Refactored Sqoop and Airflow ETL jobs, applying best practices, to migrate data from Oracle to Impala.
- Built an analytics dashboard that produces an automatic diagnosis from geographic data, using analytics and unsupervised learning.
- Built an automated data product that computes a customer credit score from telecom and banking data, and deployed the services, databases, and dashboards behind the models.
- Built the codebase for a fraud project on virtual-wallet transactions.
- Took part in onboarding new team members through mentoring and follow-up.
- Contributed to data pipelines on GCP (Composer, Cloud Storage, Speech, Artifact Registry, BigQuery, Dataproc, and serverless Spark) and refactored them for idempotency and better data quality.
- Reduced technical debt on an ad-tech project: a modern Python stack, idempotent pipelines, data validation, and process refactors.
- Built a modern data stack on AWS (Airflow, Airbyte, dbt, DataHub, and MLflow; EC2, ECR, RDS, S3, Athena) as a machine-learning platform for credit scoring.
- Built a modern data stack (MongoDB, Airflow, dbt, and Preset on AWS: EC2, ECR, DMS, RDS, S3, Athena, SSM) to give a ticketing company analytics. Wrote the news-feed ingestion design (SQS, SNS, Firehose, Lambda, and S3), mentored junior members, and wrote the architecture decisions and the platform roadmap.
- Wrote a technical review of performance issues on operational Aurora MySQL databases, with short- and long-term improvements.
- Migrated AutoSys, Control-M, ESP, JAMS, Automic, and Matillion to Airflow with Astronomer Orbiter, each project in one to two months, and deployed the result on Astro. github.com/astronomer/orbiter

Data Engineer**Nubimetrics****May 2019 – December 2019**

- Treated outliers and built algorithms to detect anomalous values.

- Refactored Nubimetrics' core from Spark SQL to native Spark, with unit and integration tests.
- Automated ETL and reporting on Azure.

Data Mining Analyst

dataEvo

August 2018 – May 2019

- Analyzed Twitter data and built visualizations for political campaigns.
- Built Shiny dashboards for devoBooster, an in-house model, and improved it with unsupervised learning.

Independent consultant

January 2020 – Present

- Implemented a low-cost modern data stack for a pet shop, ingesting CRM and API data to report on sales and customers.
- Sped up a school analytics system by refactoring Kestra and a multi-project dbt setup.
- Put a web scraper in production for storage and later transformation, using AWS Batch, CloudWatch Events, Lambda, ECS, and ECR, with CI/CD.
- With a front-end team, built a proof of concept to collect health news by topic. It runs on an Azure VM, lands raw data in Azure Blob Storage, and stores it in MongoDB for a dashboard.

EDUCATION

- **Information Systems Engineer**, Universidad Tecnológica Nacional, Tucumán (2013 – 2018).
- **Master in Management + Analytics**, Universidad Torcuato Di Tella (2018 – 2020).

CERTIFICATIONS AND AWARDS

- **Astronomer Champion** (2026).
- **Seventh-highest GPA, UTN FRT** (2017), 8.53/10.
- **Apache Airflow Fundamentals** (March 2025) and **DAG Authoring** (August 2025), Astronomer Academy.
- **Databricks Certified Data Engineer Professional** (January 2025).
- **Machine learning pipelines with Spotify data**, UTN FRT (June 2021). youtu.be/QQZ_nkEcj1I

OPEN SOURCE

- **Apache Airflow**. Made OAuth credentials optional in the Airbyte provider, so self-hosted Airbyte OSS runs without OAuth. [#64986](#), merged.
- **Astronomer Cosmos**. Fixed the dbt-on-EKS operator for newer Amazon provider versions. [#2481](#), merged.
- **Apache Airflow**. S3 DAG bundles stage from a single archive. [#73015](#), open.

TECHNICAL STACK

- Python, SQL, Scala, Bash, and R.
- Apache Airflow, dbt, Apache Spark, Airbyte, Apache Iceberg, Apache Superset, DataHub, and MLflow.
- AWS, GCP, Azure, and Databricks. Kubernetes, Terraform, and CI/CD.

LANGUAGES

- Spanish (native), English (professional working proficiency), and Italian (professional working proficiency).