

Claudia Cáceres

Analytics Engineer | Data Engineer

Buenos Aires, Argentina (UTC-3) - Fully remote

+54 11 6459 1274 - claudiacaceres.info@gmail.com

linkedin.com/in/claudiacaceresv - github.com/chipap-dev - chipap.net

SUMMARY

Data engineer focused on system integration, dimensional modeling and process automation. Three years designing and operating the internal management platform of a sports club with 10,000+ active members across several locations: 10+ Django modules, 100+ automated processes and separate databases kept in sync over REST, SOAP and ODBC. Alongside it, two data systems of my own in production on BigQuery and dbt, orchestrated by a shared Airflow deployment I built and maintain, dockerized and covered by tests in CI.

TECHNICAL SKILLS

Data modeling & warehousing: dbt (staging and mart layers, star schema, incremental models, testing), BigQuery, PostgreSQL, SQL Server, dimensional modeling, SQL, ETL/ELT

Orchestration & infrastructure: Apache Airflow (CeleryExecutor, Redis broker), Docker, Docker Compose, GitHub Actions (CI), Linux

Languages & frameworks: Python, SQL, Django, pandas

Integration & ingestion: REST APIs, SOAP, ODBC, incremental and append-only loads, PDF parsing (pdfplumber), validation, normalization, deduplication

Testing & data quality: pytest, pytest-django, dbt tests (not_null, unique, relationships, accepted_values, freshness), duplicate and staleness monitoring

Reporting & BI: Power BI, custom Django reporting systems

EXPERIENCE

Data & Backend Engineer

Náutico Hacoaj - sports and nautical club, 10,000+ members - Buenos Aires, Argentina - 2023 to Present

- Designed and built the platform: 10+ independent Django modules, each with its own data model and business rules.
- Built the permission layer: a custom Django authentication backend over a hierarchical group tree, so administration is delegated by area and each user only sees the data of their own site and role.
- Automated 100+ operational processes across 160+ management commands, and built the scheduler that runs them: cron-style programming, background execution, run history, recovery of stalled jobs and server watchdogs.
- Built integrations over REST, SOAP and ODBC between PostgreSQL and SQL Server, with identity checks, code remapping and a field-level change log, so no record is overwritten silently.
- Designed and optimized the relational model in PostgreSQL and the ingestion feeding it: incremental sync from an external API with a per-source cursor, deduplication by business key, validation and consistency checks.
- Built reporting and dashboards in Power BI and in a custom Django reporting system.

Teaching Assistant, Data Science

Henry - Remote - 2023

- Technical support for Data Science bootcamp students working through Python and SQL exercises.

Data Analyst

Tippicos - Mexico - Remote - 2022 to 2023

- Commercial management system built in Excel: a data model covering products, prices, stock, sales and purchases, with deduplication and normalization of the existing data.

3D Lidar Reviewer

Remotasks - Remote - 2020 to 2022

- Segmentation and quality review of 3D Lidar point clouds for training perception models in autonomous vehicles.

PROJECTS - BUILT AND RUN INDEPENDENTLY

Airflow Orchestrator - shared pipeline engine

github.com/chipap-dev/airflow-orchestrator

- A single Airflow deployment that runs the BigQuery and dbt pipelines of two projects without holding any business logic of its own: each project is mounted as a volume and contributes its own DAG and credentials.
- Adding a project touches three files: image dependencies, compose volume and DAG sync.

- Airflow 3.3.0 (CeleryExecutor), PostgreSQL 16, Redis 7.2, Docker Compose.

Lanchas - PDF ingestion to star schema

chipap.net/apps/lanchas - github.com/chipap-dev/lanchas

- Django app serving the official Tigre Delta boat timetables: parses government PDFs from three operators with pdfplumber (table extraction, colour-coded direction detection, footnote matching) and normalizes three different sources into a single relational model in PostgreSQL.
- Resolves holiday timetables against Argentina's official calendar.
- Star schema in dbt on BigQuery: four dimensions and one fact table at a grain of one row per timetable entry × route, with 31 dbt tests.
- Monthly Airflow DAG, pytest in GitHub Actions on every push, weekly smoke test against the live sources to catch upstream format changes.

Mareas - incremental time series from public APIs

chipap.net/apps/mareas - github.com/chipap-dev/mareas

- Consumes two official APIs (INA and SMN) across three river stations and loads them into BigQuery append-only, adding only what is not already there, so re-runs never duplicate and history accumulates instead of overwriting a rolling snapshot.
- The live dashboard reads a local per-station JSON cache, keeping the user-facing app decoupled from the warehouse and adding no latency.
- dbt staging models per source and a mart joining tide and weather by station, date and hour; tests for freshness, duplicates and forecast staleness.
- Daily Airflow DAG, Django, vanilla JS.

EDUCATION

- BSc in Data Science - Universidad Nacional del Delta (UNDELTA) - in progress
- Postgraduate Diploma in Business Intelligence - Universidad Tecnológica Nacional (UTN)
- Database Fundamentals - Universidad Tecnológica Nacional (UTN)

LANGUAGES

Spanish (native) · English (professional working proficiency)