

Claudia Cáceres

Analytics Engineer | Data Engineer

Buenos Aires, Argentina (UTC-3) - Remoto

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

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

RESUMEN

Data engineer especializada en integración de sistemas, modelado de datos y automatización de procesos. Tres años diseñando y operando la plataforma de gestión interna de un club deportivo de 10.000+ socios activos con varias sedes: 10+ módulos Django, 100+ procesos automatizados y bases separadas sincronizadas por REST, SOAP y ODBC. En paralelo, dos sistemas de datos propios en producción sobre BigQuery y dbt, orquestados por un deployment de Airflow compartido que construí y mantengo, dockerizados y con tests en CI.

HABILIDADES TÉCNICAS

Modelado y data warehouse: dbt (capas staging y mart, star schema, modelos incrementales, testing), BigQuery, PostgreSQL, SQL Server, modelado dimensional, SQL, ETL/ELT

Orquestación e infraestructura: Apache Airflow (CeleryExecutor, Redis como broker), Docker, Docker Compose, GitHub Actions (CI), Linux

Lenguajes y frameworks: Python, SQL, Django, pandas

Integración e ingesta: APIs REST, SOAP, ODBC, cargas incrementales y append-only, parseo de PDF (pdfplumber), validación, normalización, deduplicación

Testing y calidad de datos: pytest, pytest-django, tests de dbt (not_null, unique, relationships, accepted_values, freshness), control de duplicados y de vigencia

Reportería y BI: Power BI, sistemas de reportes propios en Django

EXPERIENCIA

Data & Backend Engineer

Náutico Hacoaj - club deportivo y náutico, 10.000+ socios - Buenos Aires, Argentina - 2023 a la actualidad

- Diseñé y construí la plataforma: 10+ módulos Django independientes, cada uno con su propio modelo de datos y reglas de negocio.
- Construí la capa de permisos: un backend de autenticación propio de Django sobre un árbol jerárquico de grupos, con administración delegada por área, donde cada usuario ve solo los datos de su sede y su rol.
- Automaticé 100+ procesos operativos en 160+ management commands, y construí el programador que los ejecuta: programación tipo cron, ejecución en background, historial de corridas, recuperación de ejecuciones colgadas y watchdogs del servidor.
- Construí integraciones por REST, SOAP y ODBC entre PostgreSQL y SQL Server, con control de identidad, remapeo de códigos y bitácora campo por campo, para que ningún registro se pise en silencio.
- Diseñé y optimicé el modelo relacional en PostgreSQL y la ingesta que lo alimenta: sincronización incremental desde una API externa con cursor por fuente, deduplicación por clave de negocio, validación y controles de consistencia.
- Construí reportes y dashboards en Power BI y en un sistema de reportes propio en Django.

Asistente de Enseñanza, Data Science

Henry - Remoto - 2023

- Acompañamiento técnico a estudiantes de un bootcamp de Data Science en ejercicios de Python y SQL.

Analista de Datos

Tippicos - México - Remoto - 2022 a 2023

- Sistema de gestión comercial en Excel: modelo de datos de productos, precios, stock, ventas y compras, con deduplicación y normalización de los datos existentes.

3D Lidar Reviewer

Remotasks - Remoto - 2020 a 2022

- Segmentación y revisión de calidad de nubes de puntos Lidar 3D para entrenamiento de modelos de percepción en vehículos autónomos.

PROYECTOS PROPIOS

Airflow Orchestrator - motor de pipelines compartido

github.com/chipap-dev/airflow-orchestrator

- Un único deployment de Airflow que corre los pipelines de BigQuery y dbt de dos proyectos sin contener lógica de

negocio propia: cada proyecto se monta como volumen y aporta su DAG y sus credenciales.

- Sumar un proyecto toca tres archivos: dependencias de la imagen, volumen del compose y sync de DAGs.
- Airflow 3.3.0 (CeleryExecutor), PostgreSQL 16, Redis 7.2, Docker Compose.

Lanchas - ingesta de PDF a star schema

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

- App Django que publica los horarios oficiales de lanchas del Delta del Tigre: parsea los PDF de tres empresas con pdfplumber (extracción de tablas, detección de sentido por color, matcheo de notas al pie) y normaliza tres fuentes distintas en un único modelo relacional en PostgreSQL.
- Resuelve los horarios de feriados contra el calendario oficial argentino.
- Star schema en dbt sobre BigQuery: cuatro dimensiones y una tabla de hechos con grano de una fila por horario × vía, con 31 tests de dbt.
- DAG mensual en Airflow, pytest en GitHub Actions en cada push, smoke test semanal contra las fuentes vivas para detectar cambios de formato.

Mareas - serie temporal incremental desde APIs públicas

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

- Consume dos APIs oficiales (INA y SMN) sobre tres estaciones del río y las carga en BigQuery en modo append, sumando solo lo que todavía no está, así reejecutar nunca duplica y la historia se acumula en vez de pisar una foto móvil.
- El dashboard en vivo lee un cache JSON local por estación, así la app queda desacoplada del warehouse y no suma latencia.
- Modelos staging de dbt por fuente y un mart que cruza marea y clima por estación, fecha y hora; tests de freshness, duplicados y pronóstico desactualizado.
- DAG diario en Airflow, Django, JS vanilla.

EDUCACIÓN

- Licenciatura en Ciencia de Datos - Universidad Nacional del Delta (UNDELTA) - en curso
- Diplomatura de Posgrado en Business Intelligence - Universidad Tecnológica Nacional (UTN)
- Fundamentos de Bases de Datos - Universidad Tecnológica Nacional (UTN)

IDIOMAS

Español (nativo) · Inglés (nivel profesional de trabajo)