Jeferson Stiv Argueta Hernandez

Guatemala - +(502) 3312-6786 - ajefersonstiv@gmail.com
srjefers.github.io

Senior Data Engineer with 7+ years across entertainment and media research, supply chain, marketing, telecommunications, pharmaceutical, and healthcare. Currently operating a Snowflake, dbt and Dagster platform on GCP (GKE) with ~1,600 dbt models and ~290 Dagster assets: platform upgrades and security fixes, pipeline stabilization, data observability and release management. Skilled in relational and NoSQL databases, Python, C#, Julia, JavaScript, C++ and Linux servers. Information technology professional graduated from Mariano Gálvez University.

Professional Experience

Senior Data Engineer - TeamInternational Medellin, Colombia — September 2022 - Present

- Designed and maintained ETL and ELT data pipelines in Postgres and Snowflake using dbt, Snowpark, and PySpark, introducing custom macros and incremental strategies to automate DDL/DML and Slowly Changing Dimension (SCD) loads.
- Built large-scale ingestion and transformation pipelines in Python/Spark; tuned Snowflake warehouses and queries to improve performance and reduce compute spend by 35%.
- Implemented an event-driven data lake on AWS and GCP and exposed curated datasets via FastAPI endpoints to power dashboards and downstream applications.
- Developed robust SQL for data quality and cleaning, with automated checks and retries for resiliency.
- Partnered with Data Science/ML teams in a Scrum setting to productionize features and datasets; contributed to sprint planning and estimations, and communicated roadmaps and results to stakeholders.

Big Data Engineer - MILLICOM Tigo, Guatemala — February 2021 - September 2022

- Migrated the on-premises data warehouse to AWS, rewriting Talend pipelines in Python/PySpark and adopting S3, Glue, Lambda, Athena, Redshift, ECS/Fargate; improved pipeline execution times by 40-50% and simplified operations with containerized runs.
- Built data lake pipelines on AWS S3 with PySpark and Python; automated orchestration with Airflow.
- Integrated Kafka and Kinesis for streaming to enable near real-time analytics and alerts.
- Performed data modeling and data mining on large-volume datasets to support analytics use cases.
- Collaborated with DS/ML and business stakeholders in network analytics within a Scrum team.

Data Analyst/Business Intelligence - IQVIA, Guatemala — February 2020 - January 2021

- Built batch and incremental pipelines with Python and SSIS to integrate healthcare/pharma sources into a SQL Server data warehouse; implemented slowly changing dimension patterns and audit logging to ensure lineage and traceability.
- Designed star/snowflake dimensional models and wrote production-grade T-SQL (CTEs, window functions, stored procedures) to curate KPIs and subject-area data marts for commercial and medical stakeholders.
- Automated reporting with SSRS, Tableau, and Power BI, streamlining recurring deliverables and standardizing refresh schedules.
- Led KPI standardization and report-delivery improvements; refactored critical T-SQL stored procedures used for data quality and data processing/cleaning, improving runtime by 70% and stabilizing refreshes.
- Established data-quality checks and alerting to improve reliability of downstream dashboards and analyses.

Software Engineering National Consultant - CDAG, Guatemala — February 2019 - January 2020

- Developed an internal process-management web application on .NET Framework, increasing stakeholder visibility into process status by 40%.
- Automated data cleaning and reporting with Python and C#; delivered BI with SSRS and Power BI.

Selected Projects

- Operating and hardening a Snowflake/dbt/Dagster platform on GKE. For a two-repo data platform (~1,600 dbt models, ~27K tests, ~290 Dagster assets), led the Dagster, dbt-core and Python major upgrades, closing 8 CVEs and removing an SQL-injection path from a custom Snowflake IO manager; fixed recurring out-of-memory failures by reusing the dbt partial-parse artifact across Dagster runs; built source freshness checks and model-health signals for a client-facing status view, and automated the hand-off from the weekly pipeline run to the client's QA suite with failure alerts.

- Cloud migration: AWS → GCP. Migrated a serverless-first data platform from AWS to GCP, replicating an architecture using Python, PySpark, Dagster, Snowflake, and dbt. On AWS, workloads ran on ECS Fargate with S3 as the data lake; re-platformed to GCP managed services while preserving data models, orchestration patterns, and SLAs. Validated end-to-end with data-quality checks; reduced cloud spend by 45% and, after identifying optimization opportunities during re-platforming, improved pipeline execution times by 35%.

- LLM agent for aircraft repair knowledge. Built a retrieval-augmented agent on GCP to surface historical maintenance insights from Oracle records. Implemented a dual-pipeline architecture: data ingestion (Python on Compute Engine via cron) extracted from Oracle to Cloud Storage; data transformation (Cloud Functions trigger → Airflow DAG) loaded curated layers into Postgres following a medallion architecture (bronze/silver/gold). Indexed entities and documents for embeddings in Postgres and served as a Vertex AI LLM agent for search and Q&A across past repair procedures; enabled end users to resolve maintenance issues 60% faster by leveraging historical repair context.

Education & Certifications

Master's Degree in Information Security - Mariano Galvez University	2026
Snowflake - SnowPro Core Certification	2025
Software and Computer Science Engineering - Mariano Galvez University	2024
Oracle Cloud Infrastructure Foundations 2021 Associate	2022
Data Science and Big Data: Data-Driven Decisions – MIT xPro	2021

Technical Skills

- Cloud & Warehousing: AWS (S3, Lambda, Glue, ECS/Fargate, EC2, Athena, Aurora, SNS/SQS), Snowflake, Redshift; GCP (Cloud Storage, GKE, Pub/Sub); Oracle, Postgres, DB2, MySQL, SQL Server, MongoDB, Apache Cassandra, Hadoop, DuckDB
- Orchestration & Modeling: dbt (incremental models, macros, tests), Dagster (assets, partitions, sensors), Airflow; dimensional (star/snowflake), Slowly Changing Dimension (SCD); Change Data Capture
- Processing & Pipelines: PySpark, Snowpark, SQL; ELT/ETL design; event-driven/serverless architectures
- Streaming & Integrations: Kafka, Kinesis; FastAPI, REST
- BI & Reporting: Tableau, Power BI, SSIS/SSRS, Crystal Reports
- Programming & Tools: Python, SQL, C#, JavaScript, C++, Julia; Docker, Kubernetes (GKE), Git, GitHub Actions CI/CD, Claude Code; Linux servers; Agile/Scrum
- Data Quality & Governance: dbt tests, source freshness checks, data observability, lineage, audit logging, role-based access control