**Renato Perez**

Hayward, CA 94541 | (925) 356-1720 | resebpego@gmail.com

 [linkedin.com/in/renato-perezg](https://linkedin.com/in/renato-perezg) | <github.com/ren-perez>

# SKILLS

* **Cloud & Big Data:** Azure (Data Factory, Synapse, Databricks, Cosmos DB), AWS (Glue, Redshift, Lambda), Spark, PySpark, Kafka, Apache Airflow, Hadoop, BigQuery
* **Programming:** Python, SQL, NoSQL, Django, DRF, Flask, Databricks CLI, dbt, JavaScript, FastAPI, Docker, Kubernetes
* **Data Engineering:** ETL/ELT, Data Warehousing (Azure SQL, Postgres), Airflow, Serverless, Fivetran, Astronomer
* **Visualization:** Power BI, Tableau, QuickSight, Excel, Metabase, Hex
* **Collaboration:** Agile/Scrum, Data Privacy Compliance (CCPA/GDPR), Metadata Management, Data Storytelling, CI/CD

# **EDUCATION**

* **Master of Science in Business Analytics** | San Francisco State University (2022 - 2023)
* **Bachelor of Science in Computer Science** | Universidad De Lima, Peru (2014 - 2019)

# WORK EXPERIENCE

## Senior Data Engineer | Saturn Laboratories

August 2023 – Present

* Designed and deployed a real-time analytics system for a food e-commerce platform, enabling live tracking of engagement and conversion funnels; improved decision-making speed by 40% using PostHog, Kafka, dbt, DuckDB, and Metabase.
* Engineered a scalable IoT analytics pipeline processing 2TB/day of device data; optimized ETL with Databricks (PySpark) and Snowflake, reducing latency by 35% while ensuring 99.5% uptime for equipment monitoring dashboards in Tableau.
* Engineered and optimized ETL/ELT pipelines on Azure Databricks with Delta Lake, leveraging ACID transactions, schema enforcement, and data compaction to ensure reliability, improved query performance by 30%, and reduced storage costs.
* Developed and maintained automated data catalogs for factory and shared services domains using dbt and Databricks.
* Built supporting tooling, wrote well-documented, standards-compliant code, and stayed current with industry trends in dbt, dbt Labs, and warehouse-as-code practices.
* Developed a financial Data Lakehouse in Databricks, producing predictive models for risk and anomaly detection used to track performance and support portfolio stress testing; this work was crucial for supporting the month-end close process and partnering with audit teams by providing model documentation.

## Data Engineer | Amazon

May 2023 – August 2023

* Architected robust data pipelines using AWS Glue, Spark, and Redshift to create centralized, well-documented datasets from IoT telemetry sources, improving reporting cycle time by 80% through automated validation workflows.
* Built scalable data models with optimized SQL queries for low-latency analytics, establishing performance benchmarks and quality standards that informed cross-functional climate initiative decisions.
* Developed comprehensive KPI-focused visualizations on AWS QuickSight and Power BI, taking ownership of end-to-end dataset delivery from raw data ingestion through final business insights.
* Optimized query performance by 40+% in Redshift through advanced SQL techniques and indexing strategies, establishing best practices for real-time analytics at scale.

## Data Engineer | Ernst & Young

January 2018 – December 2022

* Led design and implementation of golden datasets for fraud detection using Azure Databricks, Spark, and SQL-based data modeling practices, reducing processing times by 50% while maintaining data quality and completeness.
* Integrated Azure Purview with Databricks pipelines for automated metadata cataloging, lineage tracking, and data classification, enabling PII monitoring and ensuring compliance with CCPA/GDPR and enterprise governance standards.
* Engineered centralized data management systems integrating disparate sources through Azure Data Factory orchestration, creating unified, well-documented datasets that powered strategic Power BI dashboards for key client accounts.
* Created 20+ strategic data models using advanced SQL in Azure Synapse with comprehensive documentation, data lineage tracking, and defined quality metrics to ensure long-term dataset maintainability and cross-team usability.
* Established data governance practices including ETL pipeline documentation, transformation logic specifications, and quality assurance protocols to facilitate knowledge sharing and ensure dataset reliability.

# CERTIFICATIONS

* Microsoft Certified: Azure Data Engineer Associate
* Microsoft Certified: Power Platform Developer Associate
* Emacs. UNIX, NumPy, Unix, vim, pandas, data ingestion., statistics, business intelligence, agile, matplotlib, fast paced environment, problem-solving, product analytics, communication skills, BigQuery, JavaScript, mathematics, project management forecasting, web services, cloud computing, development life cycle, MATLAB, R, prototypes, data analytics, attention to detail, organizational skills, effective communication, Slack, data and analysis, key performance indicators, version control, strategic planning, SaaS, attention to detail, business strategy, frontend, backend, PostgreSQL, React, fintech, decision trees, verbal and written communication skills, Google Cloud, MongoDB, MySQL, NoSQL, Oracle, Scikit-Learn, Postgres, software development life cycle, data analytics, data strategy, metadata ,CCPA, Shadow, visual design management, JSON, XML, data pipeline GCP, API, microservices, PySpark, Data Privacy Compliance, Sensor Data Analysis, DSP, Kinematics, Root Cause Analysis, data modeling, data science, database, customer needs, quality assurance, troubleshooting, big data technologies, Excel, clustering, regression, cookies, product features, product management, non-relational , batch data, data solutions, machine learning, kanban, Apache