

Renato Perez

Data Engineer

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

[LinkedIn](#) | [GitHub](#)

Summary

Data Engineer with a Master's in Business Analytics and extensive experience in data pipeline development, distributed systems and big data processing. Skilled in Python, SQL and Spark for scientific data analysis and tool development. Experienced in collaborating with cross-functional teams to understand product goals and data needs, transforming raw data into actionable insights.

Skills

- **Data Science & Analytics:** Multivariate Analysis, Machine Learning, Statistical Data Analysis, Data Mining, Data Modeling, Data Visualization (Python, Excel, Power BI, Tableau)
- **Big Data & Engineering:** Apache Spark, AWS Glue, Redshift, PL/SQL, SparkSQL, PySpark, Scripting Languages, Unit Testing, Django, Data Architecture, Automated Testing
- **Cloud & DevOps:** AWS (EC2, RDS, Redshift, Kinesis, EMR, Lambda), Azure, Design Patterns
- **Version Control & Collaboration:** Git, GitHub, Slack
- **Professional Skills:** Communication, Collaboration, Leadership, Research, Product Development, Fast-paced Workflow, Self-Driven

Professional Experience

Founder & Lead Data Engineer | Saturn Labs

May 2024 – Present

- Architect and implement data models and ETL processes for clients in South America, improving data processing efficiency by 80%.
- Analyzed and automated business processes using Python, resulting in a reduction from 6 hours of manual workload to 2 minutes, on average reduced workload by 99%.
- Design and execute SQL queries for data extraction and analysis, supporting data-driven decision making.
- Lead data migration projects from Excel to robust database systems, enhancing data integrity and accessibility.
- Collaborate with client teams to understand business requirements and deliver tailored data solutions.
- Manage end-to-end data warehouse implementations, ensuring data quality and performance.

Data Scientist | SFJAZZ

August 2023 – December 2023

- Designed and developed ETL pipeline to ingest data from Google Ads and Meta Ads using Python and SQL.
- Conducted a comprehensive analysis of factors influencing live show ticket sales, utilizing OLS regression modeling to identify key drivers of revenue.
- Engineered features from diverse data sources including ticket sales, membership history, and digital marketing campaigns to create a unified dataset for analysis.

- Developed data-driven recommendations to optimize ticket sales strategies, including insights on subscription models and multi-seat purchase discounts.
- Utilized pandas, NumPy, and scikit-learn libraries in Python to perform data manipulation, statistical analysis, and predictive modeling.
- Created visualizations using matplotlib and seaborn to effectively communicate insights to stakeholders.

Data Engineer | Amazon

May 2023 – August 2023

- Engineered global data metric ETL pipelines using AWS Glue, Spark, and Python, improving data processing efficiency by 30%.
- Designed and implemented a data pipeline to process semi-structured data by integrating 500 million raw records from 30+ data sources using AWS Glue and PySpark and stored processed data in Redshift.
- Implemented carbon emission tracking for in-house products, supporting the company's clean energy initiatives.
- Collaborated with cross-functional teams to understand product goals and data needs, resulting in dashboards for KPI visualization, improving stakeholder decision-making processes.

Data Engineer | The Home Depot

June 2021 – December 2021

- Implemented new data models and processes, reducing data retrieval time by 40%.
- Performed data analysis to generate business insights, leading to a 15% increase in targeted marketing effectiveness.
- Built data expertise and owned data quality for allocated areas, reducing data errors by 25%.
- Supported critical data processes in production, ensuring 99.9% uptime.

Data Engineer | Ernst & Young

April 2019 – June 2021

- Architected and deployed new data models for 20+ market analysis projects.
- Managed data warehouse plans for multiple client products, optimizing storage and query performance.
- Built and maintained data pipelines in fast-paced environments, meeting tight deadlines.
- Performed complex data analysis to generate actionable business insights.

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)

Certifications

- Microsoft Certified: Azure Data Engineer Associate
- Microsoft Certified: Azure Fundamentals