William (Guillem) Gonzalez

New York City, USA | +1 267-241-6872 / +34 610 328 457 | william.gc1320@gmail.com | www.williamgc.com

SUMMARY

Cloud Architect & Data Engineer with a track record of transforming **legacy systems** into **scalable**, **automated GCP infrastructures**. Expert in **BigQuery**, **Airflow**, and **PySpark**, with strong experience in **CI/CD**, **DAG orchestration**, and **cost-aware cloud design**. Skilled in building **metadata-driven platforms**, **modeling automation**, and internal tools, with a focus on **speed**, **reliability**, and **performance optimization**.

CORE SKILLS

- **Cloud & Infrastructure:** GCP (BigQuery, Cloud Composer/Airflow, GCS, Dataproc, Cloud SQL, Cloud Functions, Cloud Scheduler, Vertex AI, VMs, VPCs, VPNs, IAM), AWS (S3)
- **Programming & Scripting Languages:** Python, BigQuery SQL, PySpark, Bash, VBA, Java, R, C/C++, SAS, Matlab, HTML, JavaScript, PHP, SQL Server, PostgreSQL
- **Infrastructure as Code & DevOps Tools:** Terraform, Docker, Jenkins, GitHub CLI, GitHub Actions, CI/CD pipelines, Linux/Unix/Windows environments
- **Data Engineering & Automation:** Dynamic Airflow DAGs, metadata-driven pipelines, ETL frameworks, PySpark processing, sync systems, cloud-based automation
- **Libraries & Frameworks:** Pandas, NumPy, scikit-learn, PyTorch, TensorFlow, Keras, Jupyter, Django, Seaborn
- **Data Analytics & Visualization:** BigQuery, Excel, Tableau, Looker Studio, Time Series, Financial Econometrics
- **Business Systems & Internal Tools:** Custom dashboards (Django), real-time ROI/KPI tools, automated reporting (Excel/PPT), monitoring systems, self-serve analytics platforms
- **AI & Modeling Automation:** Scalable model training/scoring, imbalanced data handling, metadata-controlled ML ops, CASS/NCOA pipelines
- **Security & Monitoring:** Spend Tracking Dashboards, Role-Based Access, Query Auditing, Usage Limits, Backup Automation
- **ETL & Data Flow:** Multi-source ingestion (SFTP, GCS, MySQL, Alteryx, S3), Automated File Watchers, Real-time Table Sync, External Vendor Data Pipelines
- **Interests & Focus Areas:** Cloud Architecture, Scalable Data Systems, Machine Learning, MLOps, Forecasting Models, Workflow Automation

EXPERIENCE

LS Direct Marketing, New York   June 2023 – Present
Cloud Architect Lead, Engineering Department

- Led the full transition from Alteryx/MySQL to a modern GCP-based architecture using **BigQuery, Cloud Composer (Airflow), Dataproc (PySpark), GCS, Cloud Functions, Terraform, and Vertex AI Workbench**, improving scalability and reducing costs.
- Built metadata-driven DAGs in **Airflow** that dynamically processed, ingested, and scored data across **200+ clients**, using control tables stored in **BigQuery** and managed through a **Django + HTML/JS/AJAX UI**, enabling business users to launch and control workflows without code.
- Architected and deployed BigQuery environments with **partitioning, clustering, and IAM-based access control**, drastically **reducing query cost by 60% and latency by 70%** across terabyte-scale datasets.
- Developed **CI/CD pipelines** using **Jenkins, GitHub Actions, and GitHub CLI** to deploy code and environments across GCP VMs and Dataproc clusters, ensuring reproducibility and seamless team collaboration.
- Automated parallel model training and scoring using **PySpark on Dataproc**, supporting **100+ client models weekly**, with runtime improvements from hours to under 20 minutes for multi-client groups.
- Migrated legacy workflows from **Alteryx + MySQL** to dynamic, metadata-driven DAGs in **Apache Airflow**, running on **Cloud Composer**, with all intermediate and final data persisted in **BigQuery**.
- Built robust file ingestion pipelines using **GCS buckets**, **Dataproc (PySpark)**, and **Cloud Functions**, validating user-uploaded files and auto-correcting schema issues for ingestion.
- Synced over **1TB/day** of data across **800+ MySQL tables and proprietary databases**, using **Airflow-triggered Dataproc clusters** and PK-based logic for incremental loads.
- Automated daily marketing data pipelines across **GCS, SFTP, and AWS S3**, improving consistency and turnaround for external campaigns.
- Implemented report automation with **openpyxl, Python-pptx**, and **pandas**, cutting report generation times from **4h+ to under 10min**.
- Created internal Django-based web tools for ROI/KPI reporting, data quality checks, BQ usage monitoring, and trigger-based campaign/vendor management—reducing processing time from 30+ mins to seconds.

Data Scientist, Data Insights Department
- Built a successful **GCP Proof of Concept** using **BigQuery, PySpark, Cloud Composer, and GitHub** that led to a full-stack transformation—resulting in the hiring of consultants and team expansion to scale this new cloud-native approach.
- Developed a **customer targeting meta-model** using **Random Forests and Neural Networks**, improving campaign response rates by **10+ basis points**.
- Built automated data enrichment pipelines with **Python, Selenium, and BeautifulSoup**, integrating scraped web data to enhance model inputs and improve targeting precision.
- Designed and deployed early-stage **ML workflows** for automated model **retraining and scoring** using **Python, Jupyter Notebooks, and Google Colab**—paving the way for production-scale orchestration with **Airflow and Dataproc.**
- Developed automated systems to extract structured data from unstructured PDF documents using **Python, PDFMiner, and Regex**, later used in marketing reports.

IronHold Capital, New York   September 2020 – June 2021
Analyst, Investment Department
- Built and deployed **automated statistical models** in **Python** to forecast corporate performance across multiple industries, enhancing investment decision-making.
- Conducted deep-dive data gathering and analysis using **SEC filings (10-Ks), investor presentations, Bloomberg, FactSet**, and industry reports to support financial modeling and valuation.
- **Authored** equity reports combining **financial analysis** with insights from **statistical models** and **market research**.
- Leveraged proprietary models to deliver data-driven answers to strategic investment questions, supporting portfolio management and sector research.

Temple University, Philadelphia   November 2019 – May 2020
Research Assistantship in Machine Learning, Finance Department
- **Researched** ML approaches to complex financial problems and **developed** a **Deep Hedging model** using **RNNs**; **built Python pipelines** to process and **validate structured and unstructured financial data**.

SOLVENTIS A.V. SA, Barcelona, Spain   January 2018 – June 2018
Risk IT Developer, Financial Consulting Department
- **Optimized** risk calculations using **Excel VBA**, **Mathematica**, and **Java**; built an **automated Reuters data pipeline** with **Python**, **Jupyter**, **Slack API**, and **SQL Server**, saving ~**€60K/year**.
- **Led automation efforts** with custom **Python solutions**, cutting manual workload and boosting team efficiency.

ACCENTURE, Barcelona, Spain   July 2016 – February 2017
Infrastructure Analyst, Accenture Consulting Group
- **Managed** the **GTR monitoring platform** for SLA compliance, serving as liaison between **bank managers** and **IT teams**; **led developers** to maintain and enhance reporting systems, delivering results to the **Executive Board**.
- **Streamlined data workflows** across **8 departments** by standardizing inputs from **14 sources**; **cut project cycles** from 3 weeks to 1 day with **custom Excel VBA automations**.

EDUCATION
TEMPLE UNIVERSITY, The Fox School of Business, Philadelphia   August 2018 – May 2020
Master of Science, Quantitative Finance and Risk Management, GPA: 3.79/4.0
- *Core Courses: Quantitative Portfolios, Derivative Markets, Financial Time Series, Asset Pricing, Stochastic Volatility*

UNIVERSITY OF LA SALLE, Barcelona, Spain   September 2009 – November 2015
Bachelor of Science, Telecommunications Engineering (Equivalent to Computer Science)
- *Core Courses: Algebra, Programming, Calculus, Physics, Statistics, Mathematical Analysis*