

Vincent DiRenzo

Philadelphia, PA | [linkedin.com/in/vincedirenzo](https://www.linkedin.com/in/vincedirenzo) | github.com/VinceDiR | vincentdirenzo.com

Senior Data Engineer with 10+ years in data engineering, analytics, and cloud. Builds scalable ETL/ELT pipelines and data platforms on GCP and AWS (Python, SQL, Airflow, BigQuery, Docker, Terraform) that deliver privacy-compliant subscription and marketing data.

EXPERIENCE

Hearst Magazines - Senior Data Engineer

2024 - Present

- Built a GCP subscription data platform (Airflow, BigQuery, Chargebee/CDS APIs) unifying digital and print subscribers for 27 brands.
- Refactored 50+ production Airflow DAGs to retire deprecated libraries, add structured logging, and enforce black/flake8 standards.
- Built Dockerized API clients (Chargebee, Sailthru, Meta, Vimeo) run via KubernetesPodOperator for reuse across production DAGs.
- Delivered lead-ingestion and audience-activation pipelines for Meta and Google Ads with privacy-compliant (CCPA) PII handling.
- Built an ML personalization pipeline (AWS Personalize batch inference to Sailthru) that tailors newsletters to each subscriber.

Bespoke Analytics - Data Analytics Engineer

2022 - 2024

- Designed and maintained cloud data warehouses (Snowflake, dbt) for client analytics, tuning for performance, cost, and scalability.
- Built self-serve BI dashboards and automated KPI reporting, coaching stakeholders on metric definitions and data interpretation.
- Translated complex datasets into actionable insights and presented findings to executives and non-technical stakeholders.
- Advised clients on long-term data strategy, including metric development, semantic-layer design, and data governance.

Annalect - Data Engineer

2022 - 2022

- Owned the aggregate media data infrastructure for Omni, Omnicom's analytics platform used by 1,500+ OMG agencies worldwide.
- Re-architected a disorganized data swamp into a governed S3 data lake with logical schemas, partitioning, and lifecycle policies.
- Managed DSP API integrations, preempting upstream schema changes and shipping code fixes to preserve data consistency.
- Built AWS ETL pipelines (Airflow, PySpark, S3, Athena) ingesting 1M+ records per day from 8 distinct media data sources.

Metis - Data Engineering Consultant

2021 - 2022

- Built and supported cloud data infrastructure on AWS and GCP for 10+ clients, serving as primary technical point of contact.
- Migrated Airflow codebases from 1.x to 2.x and optimized DAG performance, reducing pipeline runtimes by more than 40%.
- Engineered ETL pipelines extracting from REST APIs in Python, transforming with PySpark, and loading to cloud data warehouses.
- Modeled star-schema data warehouses and dimensional data marts that made key business metrics accessible to client teams.

Brandwatch - Developer Relations Engineer, API

2020 - 2021

- Managed enterprise API customer relationships and led technical enablement and integration demos with Postman and Python.
- Monitored upstream data-provider changes and adapted API endpoints and integrations to maintain uninterrupted data access.
- Liaised between customers and Product/Engineering to shape the API roadmap, beta-test features, and report adoption metrics.

Remesh - Team Lead, Public Opinion Research

2019 - 2020

- Founded the Public Opinion Research team, scaling it to 3 analysts and securing early clients for the 2020 U.S. presidential election.
- Grew presidential-campaign client relationships into \$320K in annual revenue from platform licenses and professional services.
- Partnered with Engineering to build a pipeline for processing on-platform user input and training an NLP sentiment-analysis model.

Google Cloud - Customer Engineer

2016 - 2019

- Managed 100+ enterprise accounts and a \$1.5M annual run rate as a Google Cloud solutions expert, driving adoption and growth.
- Designed and prototyped cloud data analytics and machine learning architectures on Google Cloud with client engineering teams.
- Advised clients on cloud migration, change management, enterprise architecture, and application infrastructure.

POLITICO - Marketing Data Analyst

2015 - 2016

- Analyzed platform traffic against marketing data to pinpoint high-ROI acquisition channels and campaigns.
 - Partnered with Engineering to build an ETL pipeline and data warehouse schema, establishing data marts for marketing analysis.
 - Delivered Power BI dashboards for senior leadership and analyses supporting the POLITICO Europe expansion investment decision.
-

PERSONAL PROJECTS

War in Ukraine ETL Pipeline & Dashboard (Python, PySpark, Airflow, AWS S3, Streamlit)

[GitHub](#), [Medium Post](#)

Automatic Target Recognition in Satellite Imagery (TensorFlow, Keras, deep learning)

[GitHub](#), [Medium Post](#)

Asteroid Diameter Predictor Web Application (PySpark, Google Cloud Storage, Docker, Streamlit)

[GitHub](#), [Medium Post](#)

Agile Machine Learning Prototyping with Streamlit (Python, pandas, Streamlit)

[GitHub](#), [Medium Post](#)

SKILLS

- **Languages & Data:** Python (pandas, PySpark), SQL (BigQuery, Snowflake, PostgreSQL), Bash, REST APIs, JSON/YAML
 - **Pipelines & Modeling:** Apache Airflow (Cloud Composer), dbt, ETL/ELT design, dimensional modeling, data quality, Monte Carlo
 - **Cloud & Infrastructure:** GCP (BigQuery, GCS, GKE, Cloud Functions), AWS (S3, Athena, Personalize), Docker, Kubernetes, Terraform
 - **Delivery & Analytics:** GitHub Actions (CI/CD), Git, Jira/Agile, code review, data governance, privacy compliance (CCPA), Power BI
-

EDUCATION

New York University - Bachelor's, Individualized Study in Computer Science & Statistics

2010 - 2014