C. VINCE WELKE, Ph.D.

San Diego, CA | (505) 269-4232 | vince.welke@gmail.com | linkedin.com/in/vince-welke

Technical Profile

Data engineering lead with 10+ years building data pipelines, ML systems, and analytics platforms at scale. Deep expertise in Python, SQL, cloud data architecture, and LLM integration. Led delivery of ad measurement and marketing analytics platforms generating $10M+ annual revenue for Fortune 500 clients.

Technical Skills

Data Engineering: ETL/ELT Pipelines, Airflow, Data Modeling (bronze/silver/gold), Data Warehousing, Streaming Data, RESTful APIs

Cloud & Infrastructure: AWS (EMR, Lambda, S3, Redshift, Athena, Glue, Step Functions), Spark, PySpark, EC2, PostgreSQL

AI & ML: LLM Integration (Claude API), Prompt Engineering, scikit-learn, Regression, Classification, Feature Engineering

Languages & Tools: Python, SQL, Flask, Git, CI/CD, Trino/Presto, SQLAlchemy, Pandas, NumPy

Experience

Lead Data Engineer

DeepSync, Remote | August 2024 to Present

- Architected end-to-end ETL pipeline from ingestion through bronze/silver/gold tiers using Python, SQL, Spark, EMR, Lambda, S3, and Airflow for B2B identity solutions
- Enhanced data model by implementing CTV ID tracking and IP match fidelity across 500M+ row datasets, enabling cross-device marketing measurement
- Built self-service deliverables platform (Flask, PostgreSQL, EC2) enabling product team to configure and generate customer datasets directly, reducing delivery staging from 2 days to under an hour
- Led data engineering team of 3 engineers through quarterly planning using Agile methodology, delivering features for enterprise clients generating millions in new revenue
- Mentored junior engineers through pair programming and code reviews, reducing PR cycle time by 40%

Senior Data Scientist III

LexisNexis Risk Solutions, San Diego, CA | 2023 to August 2024

- Built end-to-end testing framework in Python for data product, enabling efficient onboarding of new data sources with validation across multiple datasets
- Developed commercial driver risk model achieving 50%+ relative improvement over legacy models using ML techniques including regression and classification
- Led cloud migration planning for on-premise infrastructure, coordinating with cross-functional teams

Staff Data Scientist, Technical Lead Manager

LiveRamp, Remote | 2019 to 2023

- Architected and built ad measurement platform in Python (SQLAlchemy, Pandas, NumPy, scikit-learn) for Fortune 500 clients, measuring campaign performance across customer data and marketing channels
- Developed ETL pipelines processing digital and TV advertising data in AWS, enabling cross-channel attribution for 30+ recurring customers generating $10M+ annual revenue
- Upgraded data ingestion pipeline using Trino and AWS services, reducing infrastructure costs by 60% while maintaining data quality
- Built ML pipeline using regression and kNN for synthetic control methodology, reducing bias in marketing campaign measurement
- Managed team of 6 engineers and data scientists, driving technical roadmap through collaborative problem-solving

Lead Scientist

FICO, San Diego, CA | 2017 to 2019

- Developed real-time predictive ML models for FICO Falcon Fraud Manager deployed 3x yearly, representing 20% of FICO revenue
- Implemented model upgrade improving card-not-present fraud detection by 30%+ using neural networks and feature engineering
- Built Python toolbox for standardized data comparison adopted across product and data science teams

Graduate and Post-Doctoral Researcher

University of California, San Diego | 2009 to 2017

- Led particle physics research at CERN, overseeing C++ ETL pipeline processing petabytes of data across distributed data centers
- Built modeling toolbox used by 20+ researchers, enabling 17 peer-reviewed publications

Projects

Autonomous Tournament Bot

- Built Discord bot using Claude API to run gaming tournaments autonomously, handling team registration, scheduling, and notifications for 100+ users across 32 teams. Cloud deployment with GitHub CI/CD.

AI Inventory System (Contract)

- Developed inventory management system for retail client using Claude to parse delivery manifests and sync product catalog to Shopify. Completed in 2 weeks vs estimated 2 months manual effort. Cloud deployment with GitHub CI/CD.

Portfolio Website (vince-welke.com)

- Built personal portfolio site with integrated AI assistant using OpenAI API. Next.js, TypeScript, Tailwind CSS deployed on Vercel with GitHub CI/CD.

Education

Ph.D., Particle Physics | University of California, San Diego (2016)
M.S., Physics | University of California, San Diego (2011)
B.S., Applied Mathematics, Minor in Physics | University of New Mexico (2009) | Magna Cum Laude