Prithaj Nath

Burlington, VT prithajnath@gmail.com (518) 645-1487 www.prithaj.dev
iamprithaj prithajnath prithajnath

Summary

Data Engineer with 7+ years of experience designing ETL pipeline architectures, building multi-source data integrations, and optimizing data infrastructure in cloud environments. Skilled in Python, SQL, SQLAlchemy, PostgreSQL, and AWS with a strong academic focus on complex systems and applied machine learning.

Technologies

Programming & Scripting: Python, SQL, Bash, Ruby, JavaScript

Frameworks & Libraries: Flask, SQLAlchemy, Alembic, Pandas, NumPy, scikit-learn

Cloud & DevOps: Docker, Terraform, Kubernetes, Redis/RQ, AWS (EC2, Lambda, ELB, EKS, SNS, SQS, S3, Secrets Manager)

Data Infrastructure: PostgreSQL, BigQuery, Snowflake, Redshift

Education

University Of Vermont *Aug 2024 – Aug 2026*
M.S. in Computer Science with a specialization in complex systems and data science

SUNY Plattsburgh *Aug 2013 – May 2017*
B.S. in Computer Science

Projects

Analysis Of Superuser Churn On StackOverflow (https://bit.ly/41uz0yI) *Aug 2024 – Dec 2024*

- Set up ETL pipelines using Python, SQL, and Bash to transfer 90+ GB of Stack Overflow user data in XML to BigQuery
- Used BigQuery SQL to identify a superuser cohort (~22k users) and generated time-series data by calculating their rolling six-month averages of posts
- Detected churn patterns by fitting exponential decay models on user activity time-series

Experience

Software Engineer *New York, New York*
Singuli, Inc *May 2022 – May 2024*

- Built end-to-end ETL integrations with 12 third-party logistics and fulfillment platforms (Flexport, ShipMonk, Deposco, Expeditors, and others) using Python, SQLAlchemy, PostgreSQL, and AWS SQS/S3, with OAuth 2.0 authentication, paginated ingestion, and watermark-based incremental imports
- Designed a shared message-based import pipeline with typed enum dispatch, PostgreSQL advisory locks for concurrent deduplication, and per-stream watermark tracking used across all integrations
- Built a competitor intelligence ETL system from scratch — S3/Redshift-backed crawlers for 10+ brands, a purpose-built time-series schema, and aggregation APIs tracking pricing, in-stock rates, and discount metrics
- Extended the analytics report builder with 6 multi-dimensional actual metric types (sales, returns, net sales, inventory, unique SKU count, bundles) using SQLAlchemy Core and materialized CTEs
- Reduced median client onboarding time by 80% via daily ETL pipelines ingesting sales, inventory, and revenue data; enforced data integrity via recursive PII scrubbing and hash-based deduplication across all ingestion pipelines

Integration Engineer *Burlington, VT*

Faraday, Inc *June 2019 – May 2022*

- Implemented ETL pipelines on Kubernetes (EKS) to ingest customer and geospatial data from 50+ sources using Python, Ruby, SQL, and BigQuery
- Built micro-services in Python and Ruby to seed Facebook lookalike audiences with AI-powered cohorts, increasing client ROAS by 50%
- Implemented a micro-service in Python to deploy AI-powered bid modifiers to Google Ads at the zip code level, increasing ad campaign conversion rates by 38%

Jr Data Scientist *Burlington, VT*

IXIS Digital *Aug 2017 – Sept 2019*

- Analyzed e-commerce data from Google Analytics and Adobe Analytics, generating ad-hoc reports using R, SQL, and Python
- Implemented CI/CD pipelines for bare metal servers using Docker and GitLab CI
- Built a Python tool to concurrently scrape 100+ sites and track changes in critical DOM elements