

# PUNAM SHAW

San Francisco, CA | (469) 353-1775 | punam.shaw23@gmail.com |  
<http://in.linkedin.com/in/punamshaw/>

## PROFESSIONAL SUMMARY

---

Experienced Data Engineer/Software Engineer/ Data Analyst with 6+ years of expertise in SQL development, distributed data pipelines, and cloud-native platforms (**AWS, GCP**). Proficient in **Python** for automation of data pipelines, streamlining data processing and integration tasks. Skilled in **Snowflake** for designing performant data models, optimizing queries, and implementing real-time data ingestion using Snowpipe and cloud integrations. Adept at developing **data visualization and reporting solutions** in **Tableau** to translate complex datasets into actionable insights for business stakeholders. Hands-on experience with **Hadoop, Hive, PySpark**, for large-scale data processing. Familiar with **NLP and Generative AI** for integrating machine learning workflows into data pipelines. Experience with **Kafka Connect, BigQuery** for real-time data streaming, analytics, and reporting. Worked with **Airflow** for orchestrating containerized data applications.

## TECHNICAL SKILLS

---

- Programming: Python, SQL, SQLAlchemy, PySpark
- Databases: PostgreSQL, Snowflake, NoSQL, Google BigQuery
- Cloud Platforms: AWS(Lambda,S3,SNS,SQS), GCP(Cloud Functions, Bigquery,Pubsub, Workflow)
- Orchestration & CI/CD: Airflow Jenkins, Terraform, Ansible, CloudBees
- Tools: SQLAlchemy, Tableau, Looker, Microsoft BI, Gainsight, Plotly, Matplotlib
- Version Control & IDE: Git, Bitbucket, Visual Studio, PyCharm

## PROJECT

---

### Hackathon Project – ESG Scoring Dashboard using GenAI

Sep 2025

- Developed an **ESG Scoring Dashboard** using **Python, GenAI** and **Streamlit**. Applied LLM models to extract Environment, Social, and Governance insights with visualized metrics.
- Delivered a working prototype in **1 working Day during hackathon**, showcasing the potential of **Generative AI in ESG analytics** for faster, transparent, and comparable reporting.

### Chicago Crime & Police Station Data Analysis

Aug 2025

- Performed large-scale data analysis on Chicago crime and police station datasets using **Python** and **PySpark DataFrames**, conducted **time-based analysis** by generating a **bar chart showing crime frequency by day of the week**, uncovering trends in reporting patterns.

## WORK EXPERIENCE

---

### Data Engineer

**KOCH Industries, San Jose, CA**

May 2024 – Oct 2024 (Contract)

- Designed and implemented **SQL based** Snowflake data models to build semantic layers enabling fast and reliable analytics for business teams.
- Built end-to-end **pipelines (EventBridge → S3 → Snowflake)** to support real-time reporting, ensuring **flawless performance and reliability**.
- Implemented Snowpipe and streaming ingestion to enable near real-time data loading from AWS S3.
- Worked on **stored procedures, UDFs**, and scheduled tasks in Snowflake to streamline ETL processes and orchestrate complex data pipelines.

- Tuned complex queries and optimized pipeline performance, demonstrating **strong troubleshooting**.
- Used **Power BI** to visualize the performance of **consumer goods**, identifying key sales trends and product performance metrics for business decision-making.

## Data Engineer

### Clearwater Analytics, San Jose, CA

October 2021 – December 2023

- Engineered backend orchestration workflows using **Airflow DAGs in Python** and **event-driven messaging (SNS/SQS)**, enabling robust and automated ETL pipelines; containerized workflows with **Docker**..
- Delivered high-performance ETL systems leveraging **TypeScript, Pub/Sub, BigQuery, Cloud Functions**, and Workflows to provide real-time customer data access, driving \$5M in client revenue.
- Designed and optimized **SQL** in **BigQuery** schemas, partitioning, and clustering strategies, improving query efficiency and reducing storage costs.
- Wrote and maintained **complex SQL queries, views, and user-defined functions** (UDFs) in BigQuery to support advanced analytics and business-critical reporting.
- Built scalable data pipelines by integrating BigQuery with Cloud Storage, Pub/Sub, and Dataflow, supporting both batch and streaming analytics.
- Developed event-driven Cloud Functions triggered by Pub/Sub messages and Cloud Storage events to support real-time processing workflows.
- Created custom backend components (Airflow operators and sensors in **Python**) to address unique business logic and extend orchestration capabilities.
- Integrated **Snowflake** with Airflow and **CI/CD pipelines using Terraform**, ensuring reliable deployments and infrastructure as code. Designed **transformations, and performance-tuned schemas**, enabling faster analytics, cost optimization, and consistent delivery of trusted data to business teams.

## Data Engineer

### Lex Machina, Seattle, WA

August 2017 – October 2021

- Designed a **SQL** and **Python**-based data pipeline that ingested raw data from **PostgreSQL**, created data models via **SQLAlchemy ORM**, wrote **alembic scripts** and loaded data into **AWS and GCP environments**.
- Developed pipeline to measure employees performance metrics using **SQLAlchemy, Google Cloud function, BigQuery, Tableau, Looker Studio**.
- Monitored GCP resource utilization via Cloud Monitoring and Billing reports, implementing rightsizing and scheduling strategies to cut costs.
- Engineered and supported **full-stack applications** using **Flask, Pyramid, and RESTful backend APIs**, integrated with **React** frontends and **JavaScript** for dynamic, user-friendly interfaces.
- Configured IAM roles and service accounts to enforce least-privilege access across GCP projects.

## Data Analyst

### Mitchell International, San Diego, CA

July 2016 – July 2017

- Collaborated with stakeholders and product managers to define requirements and support cross teams.
- Performed data analysis using **Excel, Python, Tableau** to extract insights and improve system performance and business outcomes.

## Software Engineer

### SAP Labs, India

2012 – 2014

- Developed **enterprise backend services**, supporting large-scale e-commerce transactions.
- Applied **best practices in distributed system design**, contributing to system scalability.

## EDUCATION

---

MS in IT & Management | **University of Texas at Dallas, TX** | Graduated: May 2016 | GPA: 3.82/4.00

PS: Dec 2024 – Present | Working on entrepreneurship with Sakhi Foundation| Upskilling