

UDAYAN SAWANT

San Francisco, CA 94111 | udayansawant7@gmail.com | (201) 496 9231 | [LinkedIn](#)

PROFESSIONAL SUMMARY

Experienced Senior Data Engineer with over 5+ years of expertise in building scalable data platforms, data engineering, and data architecture. Proven ability to deliver high-quality data solutions using Snowflake, Databricks, Python, and SQL. Skilled in designing data models, optimizing data pipelines, ensuring data quality, and implementing data governance frameworks. Adept at collaborating with cross-functional teams to drive technical and architectural decisions, and committed to fostering innovation and best practices in data engineering.

WORK EXPERIENCE

J. P. Morgan Chase

San Francisco, CA

Data Engineer

July 2023 – Present

- Led the migration of FINRA's data collection infrastructure from a legacy XML-based relational database to Amazon DocumentDB, resulting in a modern, scalable, and efficient data management system.
- Applied advanced optimization techniques and performance tuning strategies to improve query execution, data retrieval speed, and resource utilization in Amazon DocumentDB.
- Leveraged DocumentDB's features to enhance system performance and scalability, achieving a 50% reduction in development cycles and a 50% cost saving with AWS Graviton2 instances.
- Developed and maintained robust data pipelines using AWS Glue and Amazon Kinesis for data ingestion, transformation, and streaming, ensuring reliable and timely processing of large data volumes while maintaining data integrity and consistency across multiple sources.
- Established proactive monitoring and maintenance protocols utilizing AWS CloudWatch and custom tools, implementing real-time issue detection and mitigation strategies to ensure ongoing system health, data availability, and compliance with regulatory requirements.

First Republic Bank

San Francisco, CA

Data Engineer

November 2021 – July 2023

- Designed and implemented an ETL strategy utilizing PySpark and Snowflake, reducing processing time by 70% and saving \$3.2M annually.
- Led the migration from on-premise data warehouses to Amazon Redshift, optimizing data management for a petabyte-scale environment, employing AWS Redshift for high-performance querying and data analytics.
- Managed end-to-end data migration using AWS Schema Conversion Tool (SCT), converting and deploying Snowflake schemas to Amazon Redshift.
- Leveraged Redshift's MPP architecture and columnar storage for efficient data querying, reducing load times and implementing best practices for data integrity and cost-efficiency.
- Developed SQL queries and utilized Redshift Spectrum for advanced data analysis and market trend insights., creating comprehensive reports for business decision-making and strategic planning.
- Ensured regulatory compliance with FINRA and SEC requirements through Redshift's encryption, auditing, and security features.

Netflix, Inc

Los Gatos, CA

Data Engineer

August 2020 – November 2021

- Developed Metaflow, a Python library designed to streamline data science workflows by integrating various layers of the data science stack, including modeling, deployment, versioning, orchestration, compute, and data management.
- Defined and scheduled complex data workflows with Metaflow, managing up to 20,000 tasks and ensuring high availability and scalability.
- Developed and tested workflows using Metaflow's local scheduler for rapid prototyping and transitioned successful prototypes to production-grade schedulers like AWS Step Functions.
- Built a forecasting model for Netflix subscription numbers using advanced time series analysis methods including ARIMA, ETS, and STL.

- Conducted EDA to uncover trends, patterns, and correlations in the subscription data, utilizing visualization techniques to present insights into historical data and identify key features for model improvement.
- Employed evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) for model assessment. Implemented cross-validation techniques to ensure model robustness and generalizability.

Pace University, Seidenberg School of Computer Science and Information Systems
Data Modeler - Graduate Research Assistant

New York, NY

September 2019 – July 2020

- Created a Python scraper for structuring data on 5M+ publications and used Natural Language Toolkit for abstract searches, evaluating performance using RMSE.
- Provided ongoing support for application issues, resulting in a 47% increase in user satisfaction through bug fixes and feature improvements.

United Nations Children's Fund – UNICEF

New York, NY

Data Analytics Associate

May 2019 – August 2019

- Led data-driven initiatives and worked closely with Senior Leaders across product, sales, and operations to guide product development, creating strategic decisions and launching new business development projects.
- Automated historical data persistence in DynamoDB using Python, and analyzed market data to support strategic initiatives in the Asia Pacific and US regions.

Citi Bank - India
Software Engineer

Mumbai, India

July 2017 – August 2018

- Designed and implemented ETL pipelines using a combination of SQL, Python, AWS Lambda and AWS Glue, achieving 50% speedup in data processing times and facilitating seamless data integration from diverse sources.
- Implemented data quality checks and monitoring mechanisms to ensure the reliability and accuracy of data pipelines, achieving a 95% data accuracy rate.

SKILLS & INTERESTS

- **Programming Languages:** Python (pandas, scikit-learn, numpy, matplotlib), SQL, C++
- **Cloud Technologies:** AWS (Elastic Beanstalk, CloudFormation, Kinesis, Lambda, Step Functions, CloudWatch)
- **Big Data Frameworks:** Hadoop, Spark, HDFS, EMR
- **ETL Tools:** AWS Glue, Informatica PowerCenter, DBT
- **Data Warehousing:** Snowflake, Redshift, Databricks
- **Database Services:** MySQL, MongoDB, Amazon (Aurora, DynamoDB, ElastiCache, DocumentDB)
- **DevOps and CI/CD:** Git, Azure DevOps, Jenkins, Docker, Kubernetes, Redhat Openshift
- **Orchestration & Workflow:** Apache Airflow, Luigi
- **Data Visualization:** Tableau, AWS QuickSight
- **Machine Learning:** PCA, t-SNE, ML Algorithms (Supervised, Reinforcement), MEA, MSE, R-squared, Confusion Matrix, Random Forest, AdaBoost, XGBoost, SMOTE, Class Weighting, Time & Forecasting Series.
- **Quantitative Finance & Algorithmic Trading:** Markowitz, Black-Scholes, Capital Asset Pricing, Vasicek, Monte-Carlo Simulation, Brownian Motion, CDO's, CLO's, Sharpe Ratio.

CERTIFICATIONS

- **Amazon Web Services (AWS),** Certified Cloud Partitioner (CLF-C02)
- **Amazon Web Services (AWS),** Certified Solutions Architect – Associate (SAA-C03)
- **Microsoft,** Microsoft Certified: Azure Fundamentals (AZ-900)
- **Cornell University,** Product Management Certificate

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems
 Master of Science (MS) in Information Systems

New York, NY

University of Mumbai, SSPM's College of Engineering
 Bachelor of Engineering (BE) in Computer Engineering

Mumbai, India