

UDAY VARIKUPPALA

Memphis (open to relocation) • (901) 633-5949 • uday.v3669@gmail.com

PROFESSIONAL SUMMARY

Data Engineer and Analyst with 2+ years of experience building scalable data systems that power everything from live dashboards to machine learning in production. Work sits at the intersection of analytics, engineering, and business enabling teams to shift from reactive reporting to predictive, automated decision-making.

Delivered measurable impact across fintech, healthcare, and enterprise environments by transforming high-volume, unstructured data into trusted, actionable pipelines. Expertise spans full-stack ETL development using Python, Airflow, and Kafka; data modeling in Snowflake and Databricks; and deploying ML-driven scoring systems used in real-time workflows.

Key Highlights:

- Reduced data delivery time by 50%+ through full-stack ETL pipelines built with Python, SQL, Glue, and Kafka
- Enabled real-time credit scoring and risk forecasting by deploying ML models in production with Spark and SageMaker
- Built stakeholder-facing dashboards in Power BI and Tableau to support operational KPIs and strategic visibility
- Automated compliance and reconciliation processes, cutting manual workload by 80% and improving audit accuracy
- Implemented CI/CD pipelines using GitHub Actions and Terraform for secure, scalable, and testable deployments
- Collaborated cross-functionally with product, finance, and compliance teams to ensure alignment with business goals

Focused on building data systems that are fast, explainable, and deeply embedded in how organizations make decisions driving not just insights, but outcomes.

TECHNICAL SKILLS

Languages & Tools: Python (pandas, NumPy, Dash, scikit-learn), SQL, Java, R, PL/SQL, CSS, HTML, JavaScript

Data Engineering: ETL Pipelines (Python, SQL, SSIS), Snowflake, Kafka, Spark, Hive, YAML, Great Expectations

Analytics & Visualization: Power BI, Tableau, Excel, SSRS, Seaborn, Matplotlib

Cloud & Infra: AWS (EC2, Lambda, S3), Oracle Cloud, Databricks, GitHub, Grafana, Kibana

Modeling & ML: XGBoost, SHAP, Logistic Regression, Time Series (ARIMA, Prophet), Causal Inference (CausalForest)

Other: REST APIs, Spring Boot, FastAPI, Feature Engineering, Data Vault, Agile, CI/CD Collaboration

PROFESSIONAL EXPERIENCE

Caliber Home Loans, Hyderabad

May 2021 – Jul 2023

Data Engineering

- Designed and maintained Snowflake-based data pipelines aggregating mortgage data from origination, servicing, and underwriting systems, supporting downstream analytics and regulatory workflows.
- Developed Spring Boot microservices to handle real-time loan application processing, improving throughput and reducing lag in the underwriting pipeline by 35%.

- Implemented pre-processing and validation layers in Python and SQL, reducing failed submissions due to data quality issues by 18% month-over-month.
- Architected and deployed Snowflake Data Vault models to ensure traceability and compliance with internal audit and financial regulations.
- Automated amortization and loan schedule generation using parameterized Python scripts, increasing processing speed and reducing manual errors by over 70%.
- Developed Power BI dashboards for executives to monitor live KPIs such as delinquency rates, repayment trends, and refinancing rates across portfolios.
- Integrated external credit scoring APIs into eligibility workflows, reducing risk misclassification rates and enhancing underwriting confidence.
- Built reconciliation systems comparing Snowflake data to monthly bank settlements, uncovering mismatches and reducing financial discrepancies by 95%.
- Collaborated with risk and finance teams to model prepayment risk, leveraging behavioral and macroeconomic indicators to inform strategic lending decisions.
- Migrated legacy Excel/VBA-based reports to dynamic Python-based workflows, cutting daily reporting time from hours to under 10 minutes.
- Led efforts to optimize Snowflake warehouse sizing and query efficiency, achieving an annual cost reduction exceeding 40 lakh (\$50,000+).
- Automated hub-link-satellite schema generation using YAML-based configs, enabling rapid onboarding of 20+ data sources within 2 months.
- Refactored batch ETL jobs into Kafka-driven microservices for scalable, event-based data ingestion, cutting end-to-end pipeline latency by over 50%.
- Implemented row- and column-level access policies in Snowflake, ensuring role-based data security and audit-compliant data views across departments.
- Deployed data quality checks using Great Expectations to enforce schema compliance and completeness thresholds, reducing downstream ETL failures.
- Trained and mentored junior data analysts on writing optimized SQL queries, building dashboards, and understanding mortgage lifecycle data.
- Partnered with IT, DevOps, and Compliance teams to redesign pipeline observability and incident response systems using Power BI alerts and logs.
- Participated in quarterly roadmap sessions with data leaders and executive stakeholders, aligning infrastructure improvements with strategic business goals.

PROJECTS

Intelligent Scheduling Risk System | Healthcare

Objective: Reduce patient no-show losses across clinic network

- Engineered a no-show risk model using appointment, NOAA weather, ZIP-code census, and public transit data
- Built features including forecasted weather, transit distance, and time-of-day impact
- Delivered XGBoost model with SHAP interpretability; embedded into overbooking simulation engine
- Quantified \$1.2M in potential reclaimed revenue via optimized capacity planning
- Deployed risk dashboard in Power BI for COO and clinical operations team

Strategic LTV Attribution & Segment Playbook | Fintech

Objective: Predict customer lifetime value (LTV) and optimize growth levers

- Clustered customer behavior using KMeans and synthetic usage, support, and NPS data
- Forecasted LTV using survival regression and explained top drivers with SHAP
- Built per-segment growth playbooks (retention, upsell, expansion strategies)
- Simulated financial uplift; identified \$2.7M ROI potential in Segment A via credit builder plan
- Delivered LTV heatmaps and executive-ready playbook slides for product/growth teams

Journey Interruption Diagnostics | E-commerce

Objective: Identify and fix UX drop-off points to recover revenue

- Parsed BigQuery logs from Google Analytics + CRM overlays to reconstruct user journeys
- Modeled abandonment causes across traffic sources, device types, and time-of-day
- Used anomaly detection and CausalForest modeling to estimate lift from UX fixes
- Identified Safari-specific modal issue costing \$430K/month in lost conversions
- Delivered Tableau dashboard and CRO action plan to product leadership

EDUCATION

University of Memphis, Memphis, TN

Master of Science in Data Science, (Top 1% of class)

Relevant Coursework: Advanced Database Management, Machine Learning, Deep Learning, MLOps, Cloud-Based AI Systems, Data Ethics and Governance, Research Methods and Analytics, Web and Social Media Analytics, Predictive Modeling and Forecasting.

CERTIFICATIONS

- Microsoft Certified: Power BI Data Analyst Associate
- Advanced Google Analytics
- Programming Essentials in Python - Cisco Networking Academy
- Mathematics for Machine Learning and Data Science - Coursera

AWARDS & RECOGNITION

- Best AI Infrastructure Optimization Reduced model drift by 40 percent through AI observability improvements
- Above & Beyond Award Delivered strategic projects on time with consistent quality and zero escalations