Tanay Mehendale

San Jose, CA (can relocate) | tanay.mehendale@tamu.edu | 979-344-3679 | LinkedIn/tanay-mehendale | Portfolio | GitHub

SUMMARY

Data Engineer with 2.5+ years of experience in SQL, Python, ETL and automation with exposure to SAP ERP and AWS based data pipelines. Improved data accuracy and system reliability through extraction, validation and RCA. AWS certified Solutions Architect - Associate, focused on scalable infrastructure for analytics and AI workloads

EXPERIENCE

Data Engineer Intern – *Texas A&M University | Remote, San Jose, CA*   **Aug 2025 – Present**
(Python, ETL, AWS, Data Pipelines, Data Modeling, Pandas, Medallion Architecture)

- Architected a Python based ETL pipeline to consolidate 3 data sources into a multi-layer data lake on Amazon S3, reducing data preparation time by 40% and aiding time series analysis
- Developed scalable logic to ingest 2M+ unstructured raw text via Reddit API, implementing regex-based alias matching across 25+ subreddits to expand data coverage by 300%
- Optimized performance by resolving API failures, resulting in 100% data capture reliability and a 34% increase in speed
- Integrated LLM based sentiment analysis to derive 12 distinct themes, creating high-value categorical features to improve segmentation and trend modeling

Software Engineer – *LTIMindtree | Mumbai, India*   **Jan 2021 – Jun 2023**
(SQL, Python, ETL, AWS, S3, Glue, Data Modeling, RCA)

- Developed 45+ custom ERP transactions using ABAP with SQL, improving data retrieval speed by 95% for 400+ users
- Supported SAP to AWS ETL migration by extracting ERP data, validating S3 and Glue outputs, improving Redshift load consistency across weekly cycles
- Diagnosed and resolved 30+ technical support issues via root cause analysis, cutting average resolution time by 35%
- Collaborated with cross-functional global teams to optimize SDLC audits and change management practices, reducing SOX compliance issues by 84% across 52+ releases

PROJECTS

Real Time Sign-Up Health Monitor | Link | *Airflow, Kafka, Spark, Cassandra, Docker, Git, Real-Time Streaming*

- Developed a fault-tolerant real-time streaming pipeline using Kafka, Spark Streaming, and Cassandra to monitor user sign-up health; Implemented anomaly detection logic to alert failures beyond traditional batch monitoring limits

Data Warehouse for Retail Analytics | Link | *SQL Server, SSIS, SSRS, Tableau, Data Warehousing*

- Engineered a retail Data Warehouse using SQL Server, ETL with SSIS, and dashboards with SSRS/Tableau to visualize $600M+ in sales across 6 years; Built data marts to get actionable insights into inventory and seasonal retail patterns

ApartmentFinder – AI Agent for Relocations | Link | *AI Agents, RAG, LLM, MCP, GCP, Agentic AI, Google ADK*

- Designed a multi-agent RAG based AI assistant using Google ADK, reducing insight turnaround time by 95%
- Built MCP client-server architecture for real-time distance and time calculations using Google Maps APIs

AWS Point of Sale (POS) System | Link | *AWS, EC2, Data Modeling, Clustering, Replication, MariaDB, MongoDB*

- Deployed a POS system on AWS EC2 with MariaDB, implemented clustering and sharding to enhance horizontal scaling
- Boosted read efficiency by 45% through optimized NoSQL modeling into MongoDB

ResQVision - Emergency Analytics | Link | *React, D3.js, Node.js, TypeScript, JavaScript, CI/CD, Vercel*

- Built an EMS dispatch insights dashboard using React and D3.js to reduce information delays by 60%
- Integrated CI/CD with GitHub Actions for real-time Vercel deployment error checks

EDUCATION

Texas A&M University - *M.S. Management Information Systems | **College Station, TX***   **May 2025**

University of Mumbai - *B.Tech. Electronics & Communications | **Mumbai, India***   **May 2021**

SKILLS

Certifications: AWS Solutions Architect, Assoc. Data Engineer, **Hackathon:** Google AI Agents Intensive

Programming: SQL, Python, Java, C, C++, HTML, CSS, JavaScript, TypeScript, PostgreSQL, Git, R, Shell, Bash

Database: MySQL, MariaDB, MongoDB, Postgres, SAP HANA, IBM DB2, Cassandra, OLTP, OLAP

Big Data & BI: ETL, ELT, SQL Server, SSIS, SSAS, Airflow, Snowflake, Spark, Kafka, Tableau, SSRS, Power BI, Grafana

Cloud: AWS, EC2, Athena, RDS, Redshift, S3, Docker; **AI:** LLM, RAG, LangChain, Google ADK, GenAI, MCP, Agentic AI