

Aman Sayyad

aman2003sayyad@gmail.com | linkedin | github

EDUCATION

Pimpri Chinchwad College of Engineering

Bachelor of Technology

June 2020 – June 2024

Pune, Maharashtra

EXPERIENCE

Graduate Trainee Engineer (Database Engineer)

May 2024 – Present

Pune, Maharashtra

- Automated SQL data sync pipelines across PostgreSQL, MySQL, and IBM DB2 using Bash and Python scripts on Linux systems, reducing manual effort.
- Led end-to-end MySQL database migration from on-prem to AWS RDS, ensuring high availability, data integrity, and minimal downtime.
- Partnered with developers, data engineers, and data scientists to define data requirements, implement DDL/DML changes, and design scalable schemas and data models.
- Built and maintained robust data ingestion workflows for internal and external sources, adhering to standard ETL best practices.
- Tuned and optimized complex SQL queries on large-scale datasets in PostgreSQL and MySQL, significantly improving query performance and resource efficiency.
- Executed schema migration and reconciliation across staging, pre-prod, and prod environments to maintain relational integrity and version control.
- Automated internal database access request workflows by integrating Python with the Jira API, streamlining access provisioning and reducing manual DBA intervention.
- Administered Neo4j graph database access policies and automated housekeeping tasks to enhance performance and ensure security compliance.

PROJECTS

CoinSight | Kafka, Spark, Airflow, PostgreSQL, Streamlit, AWS

April 2025 – Present

- Engineered real-time data pipelines to stream crypto data from CoinGecko to Kafka and process with Spark Structured Streaming.
- Designed and built ETL pipelines following Medallion architecture for structured, ML-ready data.
- Automated hourly data aggregation workflows using Apache Airflow for reliable data delivery.
- Deployed anomaly detection (Isolation Forest) and time series forecasting (Prophet) models for live monitoring and insights.
- Integrated external sentiment signals (e.g., Fear-Greed Index) into the data pipeline and stored results in PostgreSQL.
- Developed interactive dashboards with Streamlit and Dash for visualizing KPIs, anomalies, and forecasts.

BharatLawAI | React, FastAPI, ChromaDB, Ollama, Python, TypeScript

June 2025 – Present

- Engineered a Retrieval-Augmented Generation (RAG) pipeline, leveraging ChromaDB for efficient vector search and contextual retrieval of legal statutes to ground LLM responses.
- Designed and implemented robust RESTful APIs (FastAPI) to serve a searchable, paginated legal acts library with structured metadata, incorporating in-memory caching for optimized data fetching.
- Managed data persistence for users, conversations, and chat history using SQLAlchemy and SQLite, ensuring data integrity and schema validation for structured data.
- Developed an AI-powered legal assistant featuring an interactive chat interface and a responsive user experience built with React, TypeScript, and Tailwind CSS.

TECHNICAL SKILLS

Languages: Python, SQL (PostgreSQL, MySQL), Java, Cypher (Neo4j)

Cloud Platforms: AWS, Azure

Big Data Technologies: Apache Spark (PySpark), Kafka, Hadoop, Hive, AWS Glue, Amazon Redshift, Google BigQuery

Databases: PostgreSQL, MySQL, Neo4j, ClickHouse, IBM DB2

Orchestration & Tools: Apache Airflow, Git, Docker, VS Code

Data & ML Libraries: pandas, NumPy, scikit-learn, Matplotlib, Prophet, Streamlit, Dash

CERTIFICATIONS

Operating System Fundamentals – NPTEL Online Certification, IIT Kharagpur

2023