JAY HEMNANI

jayhemnani992000@gmail.com | GitHub: github.com/jayhemnani9910
San Jose, CA

Education

San José State University   Jan 2025 – Present | San Jose, CA
M.S. in Applied Data Intelligence
- **Coursework:** Data Warehousing, Distributed Systems, Big Data Analytics, Data Visualization, Machine Learning.

Pandit Deendayal Energy University (PDEU)   2018 – 2022 | Gandhinagar, India
B.Tech in Computer Engineering | GPA: 8.7/10
- **Coursework:** Database Systems, Operating Systems, Computer Networks, Big Data Analytics.

Technical Skills

Programming: Python, SQL, Java, Go, JavaScript
Data Engineering: Apache Kafka, Apache Airflow, PySpark, Pandas, ETL/ELT Pipelines, Data Modeling
Databases: PostgreSQL, MySQL, MongoDB, TimescaleDB, Redis, Data Warehousing
Cloud & Infrastructure: AWS (S3, EC2, Lambda, Glue), GCP (BigQuery), Docker, Kubernetes
Streaming & Processing: Kafka Streams, Spark Streaming, Batch Processing, Event-Driven Architecture
Tools: Tableau, Power BI, Streamlit, Dash, Git, GitHub Actions (CI/CD)

Experience

Data Analyst, *Elite Hotel Group*   Summer 2025 | San Jose, CA
- Engineered **automated ETL pipelines** using SQL and Python, reducing manual data preparation by **40%** and improving data consistency across multi-property analytics workflows.
- Built **interactive dashboards** in Tableau/Power BI for occupancy, revenue, and KPI tracking; developed **demand forecasting models** using time-series analysis for dynamic pricing.

Independent Technical Consultant,   2022 – 2024 | Remote
- Provided **data analytics and pipeline consulting** for small businesses, building reporting automation and data infrastructure solutions.

AI/ML Intern, *Amnex*   Jan – May 2022 | Gujarat, India
- Built **fraud detection pipeline** using ensemble ML (Random Forest, XGBoost) with SMOTE for class imbalance, achieving **94% precision** on transaction data.
- Developed **automated analytics dashboards** with statistical models for operational reporting, streamlining KPI calculations and anomaly detection workflows.

Software Engineering Intern, *Cactus Creatives*   May – Nov 2019 | Gujarat, India
- Built **cloud-native backend services** on Azure using microservices architecture; configured **CI/CD pipelines** reducing deployment time by **60%**.

Projects

Stock Data Platform — Batch + Streaming Analytics | *Python, Kafka, Airflow, TimescaleDB, Dash*
- Built **distributed streaming architecture** using Kafka topics for real-time tick ingestion, implementing idempotent writes and producer buffering for fault-tolerant data capture.
- Orchestrated **Airflow DAGs** for batch ETL pipelines handling scheduled ingests, incremental backfills, and data validation with reproducible lineage tracking.
- Designed **TimescaleDB hypertables** with star schema (FactPrices + dimensions) for time-series analytics, powering Dash dashboards with rolling indicators and intraday heatmaps.

Kayak — Distributed Travel Metasearch | *Node.js, MySQL, MongoDB, Redis, Kafka, Docker*
- Engineered **3-tier distributed architecture** with API Gateway (JWT auth, rate limiting) routing to independent microservices for horizontal scaling.
- Achieved **sub-100ms search latency** via Redis caching layer; implemented **Kafka event streaming** for real-time inventory synchronization across services.
- Designed **hybrid storage layer** using MySQL for OLTP bookings and MongoDB for analytics/logging with optimized indexing strategies.

BarcaBrain — Football Intelligence Platform | *Python, Pandas, FAISS, Streamlit*
- Built **data pipeline** for feature engineering on 10,000+ player records, transforming match events into dense vector embeddings for similarity search.
- Implemented **FAISS index** for k-NN retrieval achieving <**100ms query latency** with hybrid filters for role/league constraints.