

# JUGAL SHETH

Senior AI Data Engineer — AI/LLM Platform Architecture — Cloud Data Warehousing Specialist

📞 +1 (857) 869-8235 📩 jugal.sheth63@gmail.com 💬 LinkedIn 🤖 GitHub 📊 Tableau

## Professional Summary

Data Engineer with 5+ years architecting scalable data platforms and AI-powered analytics solutions for enterprise environments. Proven expertise in AWS cloud infrastructure, data warehouse design, and building fault-tolerant ETL/ELT pipelines processing millions of records daily. Deep technical proficiency in Snowflake, PostgreSQL, Python, and SQL with hands-on experience implementing RAG architectures and LLM integrations. Track record of translating complex business requirements into robust data models that enable BI, analytics, and data science initiatives while maintaining data quality, security compliance, and 99.9% system reliability across production environments serving 600+ users.

## Professional Experience

### VaynerX

Senior AI Data Engineer

Oct 2025 – Present

New York, NY

- Architected and scaled cloud data warehouse on Snowflake serving 600+ enterprise users, designing canonical datasets tracking user growth, engagement, and revenue metrics while implementing data governance frameworks and SLA management to ensure 99.9% data availability and sub-second query performance for executive dashboards
- Led cross-functional partnerships with Data Analysts, Product Managers, and Software Engineers to understand business needs and translate requirements into scalable data models, reducing reporting cycle time by 35% through optimized data ingestion pipelines and automated ETL orchestration using Airflow
- Defined and managed production SLAs for 50+ critical datasets, establishing data lineage tracking and auditability frameworks that improved data quality monitoring and enabled proactive incident response, reducing data quality issues by 40% while ensuring security and compliance with industry standards

### VaynerX

Data Engineer

Dec 2023 – Oct 2025

New York, NY

- Designed and deployed AI-powered conversational data assistant leveraging OpenAI LLMs and RAG architecture to enable natural language querying of business metrics, implementing prompt engineering techniques and ensuring high-quality data inputs that reduced time-to-insight by 50% for 200+ non-technical users and democratized access to real-time analytics
- Engineered robust and fault-tolerant data ingestion and processing systems across AWS cloud environment, optimizing ETL/ELT pipelines on Snowflake and PostgreSQL that delivered 40% reduction in pipeline cycle time and compute costs while processing 10M+ records daily with zero data loss and maintaining 99.9% uptime
- Built and maintained data infrastructure using best practices including automated testing, CI/CD pipelines via GitHub Actions, and distributed processing frameworks (PySpark), while developing canonical datasets for product analytics that enabled Data Science teams to build ML models and improved forecasting accuracy by 12%

### VaynerMedia

Data Analyst

Oct 2021 – Dec 2023

New York, NY

- Collaborated with Product, Growth, and Marketing teams to design A/B testing frameworks and implement data models supporting \$200K in budget allocation decisions, leveraging SQL and Python to develop ROI tracking systems that optimized campaign effectiveness and improved media spend efficiency by 18% across 10+ digital channels
- Developed BI dashboards in Tableau tracking granular ROAS and CPA metrics with real-time data refresh, establishing standardized KPI definitions and metric layer that enhanced cross-functional data trust and enabled self-service analytics capabilities for marketing and operations teams
- Engineered data pipelines integrating multiple marketing platforms via APIs, creating unified data models for media measurement and content performance tracking that reduced manual reporting effort by 60% and improved decision-making speed for streaming and content distribution strategies

### Optima Global (Client: The Hanover Insurance Group)

Claims Data Analyst

Jun 2021 – Oct 2021

Remote

- Optimized PostgreSQL queries processing 100K+ insurance claims records, achieving 60% reduction in fraud detection reporting latency through advanced SQL techniques including indexing strategies and query plan optimization, while building Power BI dashboards tracking SLA adherence that improved operational efficiency and reduced claim cycle time by 5 days

# Data Engineering & AI Solutions Projects

---

## LLM-Powered Conversational Analytics Platform – Production AI Agent Ongoing

- Building production-grade conversational AI agent for customer self-service analytics, integrating OpenAI GPT-4 with RAG pipeline architecture to enable natural language interactions with BI data, implementing prompt engineering best practices and ensuring high-quality data inputs for LLM performance optimization, targeting 5K+ monthly conversations
- Architecting multi-turn dialogue management system with context retention and intent classification capabilities, deploying on AWS cloud infrastructure with fault-tolerant design patterns to deliver adaptive conversational experiences and 40% reduction in support tickets while maintaining data security and compliance

## SKU Master Ware – Enterprise E-Commerce Data Platform

Oct 2025

- Architected and deployed production-scale inventory analytics platform for 3 e-commerce clients managing 10K+ SKUs across multiple channels, designing PostgreSQL database schema with 15+ normalized tables, Row-Level Security for data governance, and implementing distributed processing for real-time multi-channel data reconciliation
- Built real-time analytics engine computing 40+ KPIs with event-driven ETL orchestration using Airflow, automated reorder workflows, and barcode-driven GRN systems, reducing manual intervention by 80% while engineering serverless DevOps pipeline on AWS achieving 99.9% uptime and sub-second page loads for BI dashboards

## GeoPulse AI – LLM-Powered Financial Analytics Application

Feb 2025

- Deployed production OpenAI-powered Streamlit analytics application analyzing global financial news to generate market forecasts, demonstrating LLM integration expertise with RAG architecture, dbt-based modular ETL pipelines, and explainable AI metrics framework ensuring interpretable ML-driven insights with data lineage tracking and auditability

## Education

---

### Northeastern University

Boston, MA

*Master of Science in Engineering Management*

- Relevant Coursework: Data Mining and Machine Learning, Database Management and Design, Computation and Visualization, Algorithmic Digital Marketing, Engineering Probability and Statistics

### University of Mumbai

Mumbai, India

*Bachelor of Engineering in Mechanical Engineering*

## Technical Competencies

---

**Cloud & Data Warehousing:** AWS (EC2, S3, Lambda, RDS, Redshift), Snowflake (Expert), PostgreSQL, Data Storage Solutions, Distributed Processing, Cloud Infrastructure Design, Data Security & Compliance

**Data Engineering & ETL:** ETL/ELT Pipeline Development, Data Ingestion Systems, Airflow, Dagster, Prefect, Data Orchestration, Fault-Tolerant Systems, Real-Time Processing, PySpark, Kafka, Data Quality Management

**AI/LLM Integration:** RAG Architecture, OpenAI API, Prompt Engineering, LLM Performance Optimization, Conversational AI, NLP, AI-Driven Analytics, Modern AI Technology Stack, Data Quality for ML Models

**Programming & Data Modeling:** SQL (Advanced), Python, R, API Integrations, Data Architecture, Canonical Dataset Design, Schema Design, Data Governance, SLA Definition & Management, Data Lineage & Auditability

**BI & Analytics Tools:** Tableau, Power BI, Looker, dbt, TensorFlow, Streamlit, Business Intelligence Solutions, Self-Service Analytics, Cross-Functional Collaboration

**DevOps & Best Practices:** GitHub Actions, Docker, CI/CD Pipelines, Automated Testing, System Reliability, Incident Response, Agile/SDLC, Technical Documentation, Jira