JITESH S. PATIL

AI Data Engineer

jiteshpatil596@gmail.com — linkedin.com/in/jitesh-patil01 — +91 7900002579

Core Competencies

Data Platform Engineering: Apache Airflow, Apache Spark, ETL Pipeline Architecture, Data Lake, Real-time Stream Processing, Apache Kafka, Apache Flink, Data Quality Automation

AI/ML & LLM Technologies: Large Language Models (Qwen2.5, Ollama), RAG Systems, LangChain, Vector Databases (Chroma), Document Embeddings, Self-RAG Implementation, Model Fine-tuning, Semantic Search, Hugging Face Transformers

Cloud Infrastructure: AWS (S3, EC2, Athena, Glue, Lambda, EMR, RDS), Microsoft Azure, Docker, CI/CD Pipelines

Data Technologies: MongoDB, Elasticsearch, SQL Server, MySQL, Amazon S3, NoSQL

Development Stack: Python, SQL, PySpark, Scala, REST APIs, FastAPI, Flask, Git, Shell Scripting, Streamlit

Analytics & Intelligence: Power BI, Advanced Data Visualization, KPI Development, Business Intelligence Solutions

PROFESSIONAL ACHIEVEMENTS

Data Engineer — Alpha Plus Capital Advisors LLP　　　　　　　　　　　　**July 2023 - Present**

AI/ML Innovation Projects:

- **Developed Self-RAG PDF Analysis System** using Qwen2.5:7b LLM with **35% improvement in answer accuracy** over traditional RAG through intelligent document verification and self-reflection mechanisms
- **Implemented custom LLM fine-tuning pipeline** for financial document analysis, achieving **40% better domain-specific understanding** and reducing manual review time by 60%
- **Built production vector database solution** using Chroma and mxbai-embed-large embeddings, enabling semantic search across 10,000+ financial documents with sub-second response times
- **Created intelligent document processing system** with automated relevance filtering, handling multiple file formats and providing research-grade analysis with source verification
- **Generated 50K+ annual savings** by architecting Apache Airflow infrastructure that Streamlined 40+ daily workflows, boosting operational efficiency by 30%
- **Achieved 99.9% uptime** while processing 10TB+ monthly financial data through high-performance ETL pipelines, ensuring zero disruption to trading systems
- **Reduced data latency by 40%** via real-time Apache Kafka and Flink implementation, enabling sub-second decision-making for trading algorithms worth millions
- **Delivered 93.8% data accuracy** through Systematized quality monitoring framework, eliminating 70% of manual validation overhead
- **Cut storage costs by 25%** while implementing AWS S3 data lake with robust governance and security measures

Data Operations Analyst — Accenture Solutions Pvt Ltd　　　　　　　　**November 2020 - July 2022**

- **Lowered error rates by 25%** through Programmed Python and SQL validation frameworks with proactive anomaly detection
- **Processed 1M+ daily records** at 99.8% accuracy, building robust ETL workflows for social media content analysis
- **Improved processing speed by 40%** and resource utilization by 35% through strategic code optimization and indexing
- **Led data governance implementation** across 5+ client projects, establishing best practices for cross-functional teams

EDUCATION & CREDENTIALS

PG Diploma in Data Science — IT Vedant Pvt. Ltd　　　　　　　　　　2022-2023 — **Score: 92%**
Bachelor of Science (B.Sc) — Mumbai University　　　　　　　　　　2016-2019 — **CGPI: 7.6**
Specialized Coursework: Machine Learning, Deep Learning, Statistical Analysis, Database Management, Cloud Computing

PROFESSIONAL CERTIFICATIONS

Python for Data Science — IBM — **Machine Learning Specialization** — IT Vedant Pvt. Ltd — **BCG Strategy Consulting Virtual Experience Program**

SIGNATURE PROJECTS

Self-RAG PDF Assistant with LLM Fine-tuning: Architected advanced document analysis platform using Self-Reflective RAG with Qwen2.5:7b LLM and custom fine-tuning pipeline. **Achieved 40% improvement in answer relevance** through intelligent verification system and vector database optimization.Executed production-grade Streamlit interface with automated quality assessment.

Real-time Financial Data Pipeline: Architected mission-critical streaming infrastructure processing 100K+ transactions per minute using Apache Kafka, Spark Streaming, and AWS Kinesis. **Applied real-time fraud detection** with systematized alerting, protecting millions in financial assets.

Automated ETL Framework: Created reusable Python-based ETL framework with Apache Airflow and dynamic DAG generation. **Decreased pipeline development time by 50%** while implementing comprehensive monitoring, logging, and error handling capabilities.