

# S. ABHIJIT RAO

☎ +91 9985727779 ✉ [sar.abhijit2003@gmail.com](mailto:sar.abhijit2003@gmail.com) 🌐 [www.linkedin.com/in/abhijit79](http://www.linkedin.com/in/abhijit79) 📄 <https://github.com/Abhijit7979>  
🌐 [ibm.abhijit-rao.me](http://ibm.abhijit-rao.me)

## Executive Summary

---

Data Engineer with hands-on experience designing and deploying scalable data pipelines, ETL/ELT workflows, and distributed data systems using **Python, PySpark, and SQL**. Proven ability to build production-grade data infrastructure integrating **cloud platforms (AWS)**, databases, and automated ingestion frameworks. Delivered end-to-end data solutions on time across multiple organizations, including enterprise analytics platforms with **PostgreSQL, API integrations, and workflow orchestration using Airflow**. Strong foundation in data quality, performance optimization, and **cross-functional stakeholder collaboration**. B.Tech in AI with coursework in distributed computing, database systems, and cloud computing.

## Professional Experience

---

### Nuevosol Energy Pvt Ltd

Nov 2025 – Dec 2025

*AI Intern*

*Head office Madhapur, Hyderabad, India*

- Designed and built **data ingestion and retrieval pipelines** enabling enterprise-wide access to internal documentation through automated data extraction, transformation, and indexing workflows, delivering the solution **on time** to stakeholders.
- Implemented scalable **data processing workflows** using **Python and SQL**, integrating multiple data sources into a unified searchable knowledge base with vector database (ChromaDB) for semantic indexing.
- Deployed end-to-end data solution with **FastAPI backend** and Next.js frontend on cloud infrastructure (Vercel, Render), ensuring reliable data delivery and real-time query capabilities.

### Prodigal AI

Feb 2025 – June 2025

*Agentic Ai Intern*

*Remote, India*

- Developed automated **ETL pipelines** for **Dhanur AI** video processing platform, transforming raw media data into structured, production-ready outputs using **Python** and distributed processing techniques.
- Built data pipelines integrated with **vector databases** for intelligent data retrieval, content segmentation, and metadata extraction using **ChromaDB**, processing large-scale unstructured data.
- Reduced manual data processing time significantly through **pipeline automation and performance tuning for cost-effectiveness**, improving throughput and operational efficiency across the content production workflow.
- Awarded **Intern of the Month** in April 2025 for exceptional performance and innovation in data pipeline development.

### Astraveda

Jan 2026 – Present

*AI Engineer (Freelance)*

*Remote*

- Architected and deployed production data infrastructure for enterprise petroleum analytics platform, owning end-to-end **database design, ETL pipelines, cloud operations, and data monitoring** serving live business users.
- Engineered scalable **data pipelines** using **Python, FastAPI, and PostgreSQL** with **AWS services (Lambda, S3, Amplify)** for automated data ingestion, transformation, and delivery across multiple data sources.
- Built automated document processing pipeline (“Click Astra” OCR) extracting structured data from unstructured documents, implementing **data validation, quality controls, and compliance checks** with error handling workflows.
- Developed multi-source **data retrieval and aggregation system** (“Ask Astra”) enabling real-time analytics queries across enterprise datasets using LangChain and LangGraph orchestration.
- Production Link : website - [www.petro-astra.in](http://www.petro-astra.in) , App : [app.petro-astra.in](http://app.petro-astra.in)

## Projects

---

### Multi-Agent Financial Chatbot System | [Github Link](#)

- Developed a modular multi-agent system using open-source **LLMs (LLaMA 3.1, DeepSeek)** with the **Phidata framework**, orchestrated via **Groq inference APIs** for high-performance response times.
- Engineered a finance-focused **chatbot** capable of real-time stock data retrieval, web summarization, and analytics by integrating specialized agents (**Finance Agent, Web Agent**) using tools like **YFinance, DuckDuckGo**, and custom utilities.

## Natural Language → SQL Agent | [Github Link](#)

- A natural-language → SQL conversational agent using **LangGraph** and an LLM to generate syntactically-correct SQL from plain-English questions and execute them against a local SQLite database.
- Implemented a 3-node graph (write query → execute query → generate answer), plus a Streamlit UI that shows generated SQL and raw results.

## Technical Skills

---

**Languages:** Python, PySpark, SQL, Java, C++, Node.js.

**Big Data & Processing:** Apache Spark (PySpark), Distributed Computing, ETL/ELT Pipelines, Pandas, NumPy, Data Preprocessing.

**Databases & Storage:** PostgreSQL, MySQL, SQLite, ChromaDB, AWS S3, AWS Redshift, Data Warehousing/Data Mart (Star/Snowflake Schema).

**Cloud Platforms:** AWS (S3, EMR, Glue, Lambda, Redshift, Amplify), Supabase, GCP (BigQuery).

**Orchestration & Streaming:** Apache Airflow, Apache Kafka, Prefect.

**DevOps & Tools:** CI/CD Pipelines, Git, Docker, FastAPI, Streamlit, REST APIs.

**Concepts:** Data Pipeline Architecture, Data Quality & Governance, Data Monitoring & Compliance, Performance Optimization, Cost-Effectiveness Analysis, API/System Integration, Stakeholder Management, Agile/Scrum, SDLC.

**BI & Visualization:** Power BI, Tableau (Basic), Matplotlib, Seaborn.

**Soft Skills:** Collaborative Teamwork, Stakeholder Communication, Adaptability, Problem Solving.

## Education

---

**Mahindra University**

**August 2021 - June 2025**

*Bachelor of Technology in Artificial Intelligence - CGPA of 6.98*

*Hyderabad, India*

**Relevant Coursework:** Data Structures, Database Management Systems, Distributed Computing, Machine Learning, Cloud Computing, Big Data Analytics, Object-Oriented Programming.