Aashish Kushwaha

Kushwahaashu444@gmail.com| 8630728329 | linkedin.com/in/aashish-kushwaha-731163232

github.com/Aashish-kushwaha

Summary

Results-driven Software Engineer with expertise in Python-based data engineering, backend APIs (FastAPI, Flask), and LLM-powered automation. Experienced in building ETL pipelines, AI research agents, and ML monitoring platforms with scalable and optimized PySpark workflows.

Experience

System Engineer, TCS Nov 2024 – Present

- Designed and developed Python-based ETL pipelines for large-scale data migration projects, ensuring efficient data transformation and validation between Hive and Databricks platforms.
- Engineered an accelerator framework that automated the conversion of legacy SQL (Synapse, Oracle, MySQL, Teradata) and XML(Informatica) logic into optimized PySpark workflows, cutting manual re-engineering time by over 70%.
- Designed and deployed Azure Data Factory pipelines to automate data extraction from Synapse to ADLS, improving data availability and reducing manual intervention.
- Improved data reliability and pipeline scalability through robust transformation logic, performance tuning, and modular design practices.
- Partnered with cross-functional teams to streamline CI/CD adoption, establish coding standards, and ensure seamless integration within enterprise data workflows.

Project

Web-Based ML Model Monitoring & Reliability Framework

- Conducted faculty-guided research on ML model reliability, data drift, and performance degradation in deployed systems.
- Designed and implemented a Flask-based full-stack platform for model inference, versioning, logging, and statistical drift detection, containerized using Docker.

Autonomous Deep Research Agent

- Built a Python-based multi-agent LLM system using LangGraph to autonomously ingest, analyze, and synthesize large-scale technical and academic data with iterative reasoning.
- Designed cyclic agent orchestration with memory and reflection loops, enabling structured research outputs while reducing hallucinations and improving synthesis quality.

ML-Powered Data Quality & Lineage

- Built an interactive deep learning visualization system to inspect internal training behavior, including gradient flow, activation distributions, and convergence across CNNs, LSTMs, and Transformers.
- Implemented custom forward/backward pass tracking and real-time dashboards using PyTorch/TensorFlow internals to diagnose vanishing/exploding gradients and optimization issues.

Technical Skills

Languages: Python, SQL, Java

Frameworks & Tools: PySpark, Databricks, FastAPI, Flask, REST APIs, scikit-learn, NumPy, Pandas, Git, Jupyter

Cloud & Data: Azure, Azure Data Factory, ADLS Gen2, Microsoft Fabric; MySQL, Oracle, Synapse

Concepts: Data Structures & Algorithms, OOP, Data Engineering, DBMS

Education

B.Tech(Graphic Era University), Dehradun 2020 – 2024

- **CGPA:** 8.64
- **Coursework:** Computer Science

Professional Certifications and Achievements

- Azure Fundamentals
- Azure Data Fundamentals
- Databricks Data Engineer Associate certification