

Devesh Kashyap

+91-9667844659 | kashyapdevesh100@gmail.com | [linkedin.com/in/devesh-kashyap/](https://www.linkedin.com/in/devesh-kashyap/) | github.com/Kashyapdevesh

EXPERIENCE

O9 Solutions

Application Engineer 2

April 2025 - Ongoing

- * Achieved **85% reduction** in processing runtime, cutting processing time from **120 minutes to 18 minutes** on a 1TB daily PySpark data validation pipeline through partition-aware checks, streamlined transformations, and parallelized I/O.
- * Reduced data ingestion time for high-volume systems by **50%** by designing and implementing a modular Multi-Tenant data flow framework
- * Introduced a unit testing framework as a pre-commit hook in Azure DevOps (ADO), achieving **75%** coverage and establishing peer reviews to improve system robustness and code quality.

Application Engineer 1

July 2023 - March 2025

- * Reduced pipeline runtime by **80%** through Performance, Stability, and Reliability (PSR) initiatives by developing an automated Spark parameter tuning tool, implementing real-time pipeline monitoring, and introducing pipeline grouping to minimize job initialization overhead.
- * Achieved a **60%** reduction in AWS pipeline runtime by optimizing workflows with parallel execution, multi-threading, and compute tuning across EC2 instances.

Application Engineer Intern

January 2023 - June 2023

- * Increased data acquisition efficiency **3x** by developing and deploying multithreaded Python web scraping pipelines using **Selenium, BeautifulSoup, and Scrapy** to extract structured product data from 50+ global retail sites.
- * Enabled seamless ingestion and transformation of large-scale datasets by designing and deploying event-driven data preprocessing pipelines using **AWS Lambda, S3, and RDS**.

TECHNICAL SKILLS

Languages: Python, SQL (Postgres, MySQL), Java, C, C++

Cloud & Big Data: AWS (EC2, S3, Lambda, RDS, Glue), PySpark, Delta Lake, GCP, Azure (ADF, ADLS, SQL DB)

DevOps & Tools: Git, GitLab, GitHub Actions, Docker, Linux, Nginx, Gunicorn

Technologies & Frameworks: Selenium, BeautifulSoup, PyTorch, Django, Flask, FastAPI, Streamlit

Machine Learning & GenAI: LangChain, LlamaIndex, RAG Pipelines, Hugging Face Transformers, CrewAI, Vector DB (Milvus, ChromaDB, FAISS), Knowledge Graph DB (Neo4j)

PROJECTS

DeepDoc Search Engine - Asynchronous Data Pipeline | *Python, FastAPI, AWS S3, ChromaDB, Elasticsearch*

- Built a production-ready asynchronous ingestion pipeline from AWS S3, reducing processing time by **50%**.
- Implemented scalable, concurrent document processing supporting **up to 50 PDFs simultaneously** from **4 different cloud/SFTP sources** with automated indexing and monitoring.
- Developed a hybrid search system using **ChromaDB** for embeddings and **Elasticsearch** for metadata retrieval.
- Integrated keyword and semantic search, reducing average query response time from **2s to 0.8s**.

EDUCATION

Indian Institute of Information Technology Ranchi

B. Tech.(Hons.) in Computer Science and Engineering (CGPA : 8.34/10)

August 2019 - May 2023

ACHIEVEMENTS

- **Q2 Business Milestone Award 2025:** Launched Multi-Tenant feature driving product scalability
- **1st Place - O9 SPARK Hackathon 2024:** Built AI-powered supply chain log analyzer
- **O9 SPOT Award 2024:** Recognized for engineering innovation and automation
- **Global Rank 3 - Alibaba Cloud AI Challenge 2021:** Top 3 of 2500+ participants
- **Finalist - TDeFi BizThon, GITEX 2021**
- **Top 10 - Adani Jack the Hack 2021**