

TUSHAR SRIVASTAVA

SOFTWARE ENGINEER

8+ years across backend systems, data engineering, and applied AI, building production systems end-to-end — from APIs and data pipelines to multi-agent LLM applications — through architecture, implementation, and deployment. Combines individual-contributor depth with experience in leading teams and driving projects from requirements to production.

CONTACT

- Mobile: +91-7318196888
- Email: tusharsrivastava162@gmail.com
- Website: tusharsrivastava.com
- LinkedIn: linkedin.com/in/tshrv
- Github: github.com/tshrv

EXPERIENCE

Nagarro, Gurugram — *Staff Software Engineer* — NOV 2024 - PRESENT

Client: **BCG USA**

Technologies: Python, FastAPI, PostgreSQL, Airflow, Snowflake, AWS (S3, Lambda, SQS, Step Functions, DynamoDB), Multi-Agent LLM Pipelines, PydanticAI, OpenAI, Docker, GitHub Actions

- Built an **enterprise knowledge-retrieval platform** that makes consultants' prior research — case studies, market analyses, and decks buried across SharePoint — discoverable across teams. The system crawls workspace-scoped SharePoint sources, parses and enriches documents (LLM-generated summaries and tags), generates embeddings, and indexes them in Elasticsearch for semantic search with page-level result attribution. **Running in production over 300K documents**, turning previously siloed, unsearchable research into a self-serve knowledge base.
- Designed an agentic document-qualification workflow that classifies SharePoint repositories for ingestion using multi-stage LLM reasoning with schema-validated outputs (PydanticAI), **replacing manual triage across ~500K documents**.
- Designed a multi-agent SQL RAG system over a Snowflake warehouse, enabling natural-language audit analytics for non-technical users and **cutting analysis turnaround from hours to minutes**.
- Built large-scale ETL pipelines on Airflow and Snowflake powering audit data analysis across 5 data sources.
- Re-architected two LLM-backed services to fully asynchronous I/O (async OpenAI and AWS calls via aiobotocore), letting each pod handle 10–20 concurrent consumers and collapsing aggressive Kubernetes pod autoscaling into a few high-concurrency workers — cutting compute cost from **~\$28/day to under \$5** for one service, and from **~\$90/day to \$25/day** for another.
- Engineered a pluggable pipeline runtime that dynamically loads configuration-defined Python packages from JFrog, letting downstream teams ship custom processing logic without touching core pipeline code.
- Developed a near-duplicate detection service for post-ingestion SharePoint data using threshold-based similarity scoring, **reducing redundant downstream processing by 30%**.

Neuralix AI, Bengaluru — *Lead Software Engineer* — OCT 2023 - OCT 2024

Client: **Indian Armed Forces**

Technologies: Python, Kafka, Spark, MongoDB, PyIceberg, MinIO, Llama 3, LangChain, Ollama, Scikit-learn, Docker

- Led a team of 5, delivering an end-to-end predictive analytics platform for border operations - automated ingestion, risk inference, and a domain-tuned LLM assistant - used by leadership for operational decision-making.
- Served as the technical face of Neuralix AI to senior defence leadership - driving requirement gathering and solution proposals to secure the engagement against 10+ competing firms.
- Implemented a Spark Structured Streaming pipeline ingesting real-time Kafka streams inside a fully air-gapped, on-premise environment mandated by security compliance.

- Developed a custom PyIceberg plugin enabling ELT into a MinIO-backed Iceberg lakehouse in the isolated environment.
- Deployed Llama 3 with LangChain entirely on-premise, enabling analysts to query a highly sensitive data corpus and generate insights and plots with zero internet exposure.

EPAM Systems, Gurugram — Senior Software Engineer — OCT 2022 - SEP 2023

Client: **Google India**

Technologies: Python, Django, REST, MySQL, React, TypeScript, CI/CD

- Optimized the reporting engine's SQL queries, **cutting report generation time from ~150 seconds to under 20 seconds** for a workforce monitoring platform tracking attendance, assignments, and support tickets.
- Built API integrations that replaced a recurring manual workflow of downloading and re-uploading CSV data, eliminating daily manual effort and a common source of errors.
- Developed and maintained feature modules and the admin dashboard — including React/TypeScript front-end components — and resolved production issues across the platform.

Hitachi Vantara, Pune — Software Development Engineer — APR 2021 - OCT 2022

Client: **Pioneer Natural Resources USA**

Technologies: Python, asyncio, FastAPI, MongoDB, Azure (AD, Service Bus, CI/CD), Event-Driven Microservices, Docker

- Built a real-time monitoring engine for actively drilling wells, running custom statistical models over streaming ETL data to detect 10+ classes of critical drilling issues — **prevented 7 incidents and over \$10M in losses**.
- Cut well-log reconstruction time **from ~1 week to under 30 minutes** by building an ML lifecycle application (Streamlit, scikit-learn) for petrophysicists.
- Designed an asynchronous template and notification engine with recipient management.

Freelance — Software Engineer — APR 2020 - MAR 2021

Clients: **Spetra USA, Solution Umbrella India**

Technologies: Python, Django, Django Channels, PostgreSQL, Redis, Celery, AWS, Docker, NGINX

- Delivered a restaurant-chain business management platform — billing, inventory, staff, sales analytics, and CRM with targeted email/SMS reward campaigns — plus a WebSocket-based internal messaging service.
- Built a services marketplace with an e-agreement enforcement mechanism gating vendor participation on valid agreements, alongside per-vendor and marketplace-wide analytics dashboards.

Amara AI, Gurugram — Associate Software Engineer — APR 2019 - MAR 2020

Technologies: Python, Django, AWS (EC2, RDS, Lambda, SES, SQS), Celery, PostgreSQL, Redis

- Contributed to an HR chatbot campaign engine increasing participation rates **from 35% to 88%**.
- Implemented an asynchronous serverless document generator (AWS Lambda + SQS) with non-blocking delivery, eliminating user wait time on generation-heavy requests.
- Extended Django's role-based Group/Permissions framework with **attribute-based access control (ABAC)**, enhancing authorization over sensitive data; shipped SSO integrations (Okta, OneLogin) and performance upgrades using Celery task queues and Redis caching.

Nucleus Software, Noida — Intern — JAN 2018 - SEP 2018

Technologies: Python, Django, Bootstrap, JavaScript, AMCharts

- Developed an analytics dashboard and the supporting Django REST APIs for an internal monitoring tool.

PROJECTS

Lakehouse Data Platform (*Personal Project*)

- Built an end-to-end lakehouse over **TPC-H SF100** (~100 GB): Spark for ingestion and transformations into Apache Iceberg tables on MinIO, with Nessie for catalog-level versioning, dbt models served through Trino for analytics, and OpenLineage for pipeline lineage tracking.
- Implemented a **multi-engine setup with Spark and Trino** operating over shared **Iceberg tables**, exercising Nessie-backed catalog versioning — branching, merging, snapshot isolation, and time-travel queries — for reproducible, Git-like data workflows.
- Source code: <https://github.com/tshrv/lake-forge>

SKILLS

- **Languages:** Python, JavaScript, TypeScript, SQL
- **Backend & APIs:** FastAPI, Django, REST, asyncio, Microservices, Event-Driven Architecture
- **Generative AI & Agentic Systems:** Multi-Agent LLM Pipelines, RAG, Agentic Workflows, PydanticAI, LangChain, Prompt Engineering, OpenAI APIs, Local LLM Deployment (Ollama, Llama 3), Spec Kit (spec-driven development)
- **Data Engineering:** Airflow, dbt, Snowflake, Trino, Spark (incl. Structured Streaming), ETL/ELT Pipelines, Lakehouse Architecture (Apache Iceberg), OpenLineage, n8n
- **Messaging & Async Processing:** Kafka, Celery, AWS SQS, Azure Service Bus
- **Databases & Caching:** PostgreSQL, MySQL, MongoDB, Redis, DynamoDB
- **Cloud:** AWS (EC2, S3, Lambda, RDS, SQS, SNS, SES, Step Functions, DynamoDB), Azure (Active Directory, Service Bus)
- **DevOps:** Docker, Kubernetes, GitHub Actions, Azure CI/CD Pipelines, NGINX
- **Testing:** Pytest (TDD), Locust (load testing)
- **ML & Data Analysis:** Scikit-learn, Pandas, NumPy, Bokeh, Matplotlib
- **Frontend:** React, Next.js, TailwindCSS
- **Domain Exposure:** Management Consulting, Defence, Oil & Gas, HR Tech, Food & Beverage

AWARDS, ACCOLADES AND CERTIFICATIONS

- 2018 — Nucleus Certified Software Professional — Nucleus Software, Noida, Uttar Pradesh
- 2018 — Finalist — Smart India Hackathon, Ministry of Tribal Tourism, Government of India
- 2017 — Inspiration Award — Smart India Hackathon, Ministry of Railways, Government of India

EDUCATION

- B.Tech CSE 2018 — Abdul Kalam Technical University, Lucknow, Uttar Pradesh — 72%
- Senior School 2013 — Delhi Public School, Prayagraj, Uttar Pradesh — 78%
- High School 2011 — Delhi Public School, Prayagraj, Uttar Pradesh — 8.8 CGPA