

Shivam Gupta

Pune, India | www.shivamai.dev | contact@shivamai.dev
[linkedin.com/in/shivamgupta1812](https://www.linkedin.com/in/shivamgupta1812) | <https://github.com/shivam-gupta18>

PROFESSIONAL SUMMARY

Backend engineer with 7+ years building secure, multi-tenant systems and internal platforms across fintech and logistics. Designed authentication and SQL query-validation layers that removed injection risk and unified tenant access control, and re-architected compute infrastructure to cut costs up to 80% in the most recent migration. Works end-to-end in Python, from system design through deployment on AWS and Kubernetes.

TECHNICAL SKILLS

- **Backend & APIs:** Python, FastAPI, Pydantic, Flask, RESTful APIs, SQLAlchemy, System Design, Unit Testing (unittest)
- **Cloud & Infrastructure:** AWS (EC2, S3, IAM, Secrets Manager, EventBridge, Glue, SageMaker), Kubernetes (EKS), ArgoCD, Docker, Git, CI/CD (GitLab CI), Agentic AI Workflows (Claude Code, Amazon Q, Gemini API, Google ADK)
- **Observability & Monitoring:** CloudWatch, OpenSearch, Grafana, Prometheus
- **Data & Distributed Systems:** PostgreSQL, MySQL, Redis, Apache Kafka, Apache Airflow, Trino, Apache Spark (PySpark), Pandas, Polars
- **Leadership:** Code Reviews, Mentorship, Technical Documentation, Cross-team Collaboration

PROJECTS

LLM Cost & Latency Observability

Personal Project

- Built a FastAPI service with Pydantic-validated endpoints that wraps the Gemini API, tracking token-based cost and latency for every call and routing prompts between two model tiers based on complexity.
- Added retry with exponential backoff and a budget-cap circuit breaker to prevent runaway inference costs, backed by SQLite storage, Prometheus/Grafana metrics, and unit-tested routing logic.

PROFESSIONAL EXPERIENCE

FPL Technologies Pvt Ltd (Onecard)

Pune, India

Senior Software Engineer

Feb 2023 – Present

- Built a Python authentication library integrating OAuth2 and AWS Secrets Manager, replacing per-tenant credential handling with one RBAC-compliant security layer across JupyterHub sessions and scheduled Git-based jobs.
- Designed a SQLAlchemy query-translation layer that converts user parameters into validated SQL, closing a SQL injection vulnerability and removing a recurring source of query-syntax errors.
- Re-architected the team's data-transformation pipeline, migrating dbt from AWS Glue to Apache Spark on Kubernetes (EKS) via Airflow's KubernetesPodOperator, and cut infrastructure costs 80%.
- Pushed compute-heavy operations down to Trino and rewrote core Pandas logic in Polars, scaling workloads down from 8xlarge to xlarge instances and cutting compute costs 40%.
- Replaced SageMaker-hosted notebooks with a self-managed, Docker-isolated JupyterHub environment on EC2; managed-service costs dropped 30%.
- Shipped a Flask/React monitoring dashboard, deployed on Kubernetes (EKS) via ArgoCD, giving the team real-time visibility into distributed job and cron status.
- Led code-review standards for the backend team and introduced resilient error-handling with exponential backoff for service retries.

Bigfoot Retail Solutions Pvt Ltd (Shiprocket)

Delhi, India

Software Engineer

Jun 2021 – Feb 2023

- Automated the fulfillment billing system with a rules-based decision engine, eliminating 95% of manual accounting work.
- Identified and resolved backend performance bottlenecks with targeted hotfixes, keeping the platform stable through high-traffic peak seasons.

Triyama Software Pvt Ltd

Pune, India

Full Stack Developer

Mar 2020 – Jun 2021

- Built an internal web-page builder that automated UI delivery, cutting project turnaround time and manual developer intervention by 90%.
- Built and maintained web applications that processed and visualized data, and automated recurring analytical reports to deliver insights directly to clients.

GAK E Solution Pvt Ltd

Hyderabad, India

Software Engineer

Jun 2019 – Mar 2020

- Integrated third-party REST APIs for social login and payment gateways.
- Led backend, UI, and SEO improvements across the application, raising SEO score from 64% to 97%.
- Optimized MySQL schema design, improving query response times.

EDUCATION

IFHE Hyderabad

Hyderabad, India

Bachelor of Technology

Aug 2015 – Jul 2019