

Oshan Shrestha

Backend Engineer

📍 Sanepa, Lalitpur ✉ work@shresthaoshan.com ☎ 986-932-2348 🔗 oshan-shrestha

🌐 shresthaoshan 🔗 shresthaoshan.com 🗨 He/Him

Backend engineer with 5+ years of experience in distributed systems, data pipelines, and resilient APIs. Designed the platform architecture of pline.ai and works closely with product teams on technical direction. AWS Certified Solutions Architect.

Skills

Languages & Backend – Python, Go, Rust, TypeScript, FastAPI, GraphQL, gRPC, REST, Temporal, Kafka, Apache Flink, SQS, **Data & Cloud** – PostgreSQL, MySQL, Redis, MongoDB, DynamoDB, Supabase, AWS (API Gateway, Lambda, EKS, Kinesis, S3), Kubernetes, ArgoCD, Docker, GitHub Actions, **Other** – Stripe Billing, LLM structured extraction, Playwright, Next.js

Experience

Grepsr Pvt. Ltd.

Backend Engineer pline.ai – web data API (2026-present) <i>Designed and built the platform architecture: API, auth, metering, billing, and Temporal workers on EKS. Serves ~200k requests/day.</i> - Built API key provisioning on AWS API Gateway with a custom Lambda authorizer for plan and concurrency limits, and a warm pool of pre-created keys. - Designed usage-based billing on Stripe (checkout, webhooks, reconciliation, refunds, dunning, prorated upgrades) on a credit ledger with reservations and itemised receipts. - Built the SQS → Temporal metering pipeline in Go, migrated from Flink; prices each request by proxy tier, render mode, and LLM tokens. Added Go workers for credit cycles, email, and key-pool refills. - Contributed an LLM browser-interaction agent and crash-safe batch scraping to the scraping engine (Python, Rust). - Cut team-page authorization from 18 external DB round trips to 2 and halved the cost of every authorization check; enforced row-level security on all tables and an enforcing CSP. Grepsr data platform (2024-present) - Maintained and extended the backend for ~2,500 concurrent production crawlers and 500+ GB of client data delivered daily. - Cut internal API p95 latency from 3–7 s to under 1 s by rewriting inefficient queries, adding caching, and removing unnecessary data loading. - Directed an LLM transformation engine that shapes each row to a customer schema; built its schema locking, ordered output and cost controls. pline.ai's structured JSON output builds on it. - Diagnosed cross-run data delays in the shared Kinesis → Flink pipeline; tested per-run isolation with Kafka and workflow engines. - Contributed to migrating production infrastructure between AWS regions; led incident response for critical production issues. Associate Backend Engineer - Maintained high-traffic Apache Flink pipelines and migrated workloads to event-driven pipelines. - Built fault-tolerant workflows with Temporal on Kubernetes.	06.2024 – present Kathmandu, Nepal
Software Developer, Surgeon Tech Pvt. Ltd. Built React/Next.js product interfaces and reusable component libraries.	12.2020 – 02.2023 Dhulikhel, Nepal

Education

B.Sc. Computer Science, Kathmandu University (Electives: Machine Learning, Information Security)	08.2018 – 07.2023 Dhulikhel, Nepal
--	---------------------------------------

Certifications

AWS Certified Solutions Architect - Associate – Issued 06/2025 • Credential ID: e24333696de043628ee6ef5df062ae7e