Saurabh Kumar

Backend Software Engineer

saurabh-kumar-119202144 | saurabhq36@gmail.com | +91-8057981281 / 7895291281

Summary

Backend Software Engineer with 4.5+ years of experience developing scalable microservices, REST APIs, and AWS serverless workflows. Experienced in designing reliable, event-driven architectures with robust error handling, async processing, and fault tolerance. Proven track record of improving distributed systems performance, reliability, compute efficiency, and SLA outcomes.

Technical Skills

Languages and Frameworks: Python, FastAPI, Flask, Django, Node.js, SQL

Backend & Architecture: Microservices, Event-Driven Systems, Serverless, REST APIs, Docker

Messaging & AWS Services: AWS SQS, AWS Step Functions, AWS Lambda, ECS, S3, Kafka-based CDC Event Consumption

Databases: PostgreSQL, MySQL, MongoDB, Redis

Experience

Software Development Engineer Nov 2024 – Present

Spyne *Gurugram, India*

- Engineered a serverless orchestration system for 3D vehicle reconstruction jobs using AWS Step Functions, SQS, Lambda, and S3. Implemented retry handling, failure tracking, and dead-letter queues to ensure long-running jobs, often taking up to 45 minutes, were retried safely and moved to a DLQ after repeated failures, supporting 500+ jobs per day in production.
- **Improved SLA compliance by 70%** through weighted scheduling and round-robin GPU allocation across image processing workloads. Prioritized long-waiting jobs to prevent starvation while maintaining fairness across customer tiers. Introduced demand-based dynamic GPU provisioning to scale capacity up during traffic spikes and down during quiet periods, lowering queue backlogs and idle compute spend.
- **Accelerated media ingestion processing by 40%** by optimizing URL transformation to concurrently migrate expiry-prone client-hosted media URLs to internal S3 bucket URLs, preventing URL-expiry-related processing failures. Added per-URL retries, graceful fallbacks, and fault-tolerant downstream handling, **achieving a 99.6% successful conversion rate** across high-volume ingestion workloads.
- Developed a MediaKit service to asynchronously apply client-specific banner, billboard, and watermark overlays on processed images. Designed customization rules, persistent configuration storage, and serving APIs to support placement-based overlays across image sequences.
- Delivered an event-driven data ingestion and reporting pipeline using Kafka CDC consumers to aggregate image-processing workflow data from MongoDB collections and relational tables into dashboard-ready tables. Enabled visibility into processing status, SLA breaches, delivery updates, and failures.

Software Development Engineer Jun 2023 – Nov 2024

Innovaccer *Noida, India*

- Contributed to Snowflake migration for Innovaccer's DAP platform, leveraging an automated tool to convert ingestion queries and ETL flows from MySQL to Snowflake's columnar model for 3 data services, and performed validation to ensure correctness across large-scale historical datasets before cut-over. Migration contributed to **30% lower analytical query costs for 50+ enterprise clients**.
- Modernized 6 production microservices from Python 3.8 to Python 3.11 by resolving dependency and async compatibility issues, validating critical service flows, and rolling out changes in phases with rollback support. Improved runtime performance and removed end-of-life security risk without user impact.

Software Development Engineer Sep 2021 – Jun 2023

Analytics Vidhya *Gurugram, India*

- **Improved signup quality by lowering junk leads by 70% and increasing signup conversion by 35%** through an OTP verification microservice for new user signups. Integrated SMS provider, rate limiting, retry logic, and verification attempt limits for reliable verification.
- Developed backend services for a Tournament Portal supporting hiring contests and hackathons end-to-end – developed REST APIs for contest creation, user registration, live participation, scoring, and candidate session management, along with integration with third-party proctoring systems.
- Implemented real-time LeadSquared CRM integration for centralized lead capture across multiple platforms by designing ingestion APIs and an asynchronous queue-based processing pipeline to handle high event volume reliably. Applied retries, DLQ-based failure handling, and idempotent lead processing to prevent duplicate entries during retries.

Education

Bachelor of Technology in Computer Science & Engineering 2017 – 2021

GLA University