Saurabh Kumar

Backend Software Engineer

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

Summary

Backend engineer experienced in building scalable microservices and serverless architectures. Skilled in designing APIs, working with message queues, and deploying production-ready services. Focused on building reliable, clean, and scalable backend systems.

Technical Skills

- **Languages:** Python, 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 — Spyne Nov 2024 – Present

- Designed and built 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.
- **Reduced SLA breaches by 70%** by weighted scheduling and round-robin GPU allocation across image processing workloads. Prioritised long-waiting jobs to prevent starvation while maintaining fairness across customer tiers. Added demand-based dynamic GPU provisioning to scale capacity up during traffic spikes and down during quiet periods, reducing queue backlogs and idle compute spend.
- **Reduced media ingestion processing time 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.
- Built 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.
- Built an event-driven reporting pipeline using Kafka CDC consumers to aggregate image-processing workflow data from MongoDB collections and relational tables into dashboard-ready tables, enabling visibility into processing status, SLA breaches, delivery status updates, and failures.

Software Development Engineer — Innovaccer Jun 2023 – Nov 2024

- Core contributor 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% reduction in analytical query costs** for 50+ enterprise clients.
- Upgraded 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 — Analytics Vidhya Sep 2021 – Jun 2023

- **Reduced junk leads by 70% and improved signup conversion by 35%** by building an OTP verification microservice for new user signups. Integrated SMS provider, rate limiting, retry logic, and verification attempt limits for reliable verification.
- Built 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.
- Built 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. Implemented retries, DLQ-based failure handling, and idempotent lead processing to prevent duplicate entries during retries.

Education

2017 – 2021 Bachelor of Technology (Computer Science & Engineering), GLA University