Kunwar Pratap Singh

in Linkedin ☎ +91 7042816009 ✉ me@kunwar.dev ⌂ Github

Technical Skills

Languages: Golang, Typescript, Python
Tools: Docker, Kubernetes, RabbitMQ, gRPC, Redis, Postgres, WASM, FFmpeg, Git, Cloudflare Products, NodeJS, Helix, Tmux, Fish

Work Experience

Cloudflare | *Systems Engineer – RealtimeKit / AI Gateway*   *May 2025 – Present*
Cloudflare Workers, Durable Objects, ClickHouse, WASM, WebRTC, Stripe, R2, Observability, Workers AI, TypeScript.

- Led the end-to-end **RealtimeKit backend migration** from Dyte's legacy services to **Cloudflare Workers** and **Durable Objects** (distributed database) and socket layers, scaling infrastructure to support 4000+ concurrent sessions with minimal latency.
- Architected and implemented distributed **track-level audio recording** using Cloudflare Workers, Durable Objects, and **Thunderclap SFU**; processed PCM audio streams to Opus/WebM utilizing WASM, multi-destination uploads (R2/S3/GCS), queue-based idempotent finalization, and handled 100% of live sessions reliably.
- Spearheaded the development of **AI Gateway**, optimizing streaming responses for **LLM models**; led the implementation of **Unified Billing**, integrating multiple providers under a single billing system to enable accurate **usage tracking**, **alerts**, and **dashboards** for 500+ paying customers.
- Migrated and optimized **realtime transcription** pipelines, integrating with Cloudflare worker AI catalog of models to deliver live transcripts with <150ms latency, while ensuring full backward compatibility with legacy clients.
- Designed and delivered the **Post-Meeting Transcription** feature as a low-cost, reliable alternative, enabling end-of-meeting transcript generation with full ownership of the pipeline, storage, and durable object orchestration.
- Mentored an intern and wrote detailed technical specs for new features and migrations, enabling cross-team development and adoption of best practices across RealtimeKit and AI systems.
- Led API-layer migrations for critical services, implemented socket-layer improvements, and created detailed technical specs for cross-team development; mentored an intern to assist with migrations and feature rollout.
- Enhanced observability and monitoring for RealtimeKit using ClickHouse and Grafana dashboards; integrated metrics, alerts, and logging across Workers, APIs, and transcription pipelines, improving incident response and reliability for 1M+ daily events.

Dyte (YC W21) | *Software Engineer – now Cloudflare*   *Dec 2023 – May 2025*
WebRTC, Websockets, Kubernetes, Golang, Typescript, Redis, RabbitMQ, PostgreSQL, Stripe, AI

- Joined **Dyte**, a YC-backed startup, taking ownership of multiple core systems and wearing multiple hats across backend, frontend, realtime AI pipelines, billing and WebRTC.
- Redesigned and scaled the **Realtime Transcription Service** from a Python/K8s system handling 100–200 concurrent participants to a distributed Golang master/worker architecture supporting 2,500+ concurrent users with **sub-150ms latency**; built an end-to-end **summarization pipeline** processing **2M+ transcripts monthly**, driving **revenue growth by 35%.**
- Led the design and implementation of the **automated billing system** with Stripe APIs and internal metrics, capturing **14% more hidden usage** and optimizing financial reporting.
- Fixed critical **breakout room bugs**, added scalable API features and database enhancements; optimized previously unusable webhook logs API to respond in **sub 500ms**, Redesigned developer dashboard to improve UX for webhook and billing pages; implemented cost-saving changes to reduce service resource usage.
- Led migration initiatives, including **BOM → IAD** infrastructure transitions; developed tooling to replicate events across regions within microseconds with **zero downtime**, collaborating closely with SRE.

Folioplay | *Lead Backend and DevOps Developer*   *May 2022 – Jan 2023*
Nestjs, Redis, Web3.0, MongoDB, AWS, GitHub Actions

- Developed backend infrastructure for an online fantasy gaming platform, enabling users to create teams and participate in tournaments based on crypto coins. Leveraged Nestjs and Redis to build a robust and scalable backend architecture.
- Developed a real-time cryptocurrency data pipeline using Chainlink ABI and ether.js, enabling seamless and accurate price updates for the platform.
- Implemented authentication and integration with Web3.0 technologies like **Moralis** and **Magic Link**.
- Designed and implemented a tournament system that operates on scheduled events, ensuring timely updates of team data to the blockchain-based leaderboard scores.

Droot / The Downtown Club | *Software Development Engineer / Contractor*   *Dec 2021 – Jul 2022*
FastAPI, Django, Redis, PostGIS, MySQL, MQTT, Docker, AWS (EC2, IoT Core, SES), Azure, Flask, PostgreSQL

- Led the design and development of backend systems for multiple projects (**Downtown Club** e-commerce platform for D2C brands, **Cliste-Life**, **Ralikson**), architecting RESTful APIs and relational database schemas with maintainability and scalability in mind.
- Implemented robust geospatial search functionality using **GeoDjango/PostGIS** and real-time IoT pipelines with **MQTT/AWS IoT Core**, applying design patterns and best practices to ensure reliability and extensibility.
- Integrated **Shopify APIs** and AWS services (**SES, S3**) to build core product management and email workflows, and developed an admin Brand Dashboard for D2C e-commerce platform, enabling efficient management and discovery of brand products within the mobile application.
- Engineered containerized environments with **Docker/Docker Compose** and CI/CD pipelines on cloud platforms (**AWS, Azure**), enabling reproducible deployments and consistent development workflows.
- Served as a key technical contributor across projects, collaborating with cross-functional teams to define requirements, plan architecture, and guide development efforts following industry-standard best practices.