Rohan Gupta

+1 (302) 363-8352 rohangupta883@gmail.com LinkedIn GitHub

Education

University of Pennsylvania, *Philadelphia, PA* **Sep 2020 – May 2024**
B.S.E. in Computer Science (NETS program) GPA: 3.96
Summa Cum Laude, E. Stuart Eichert, Jr. Memorial Prize, Course Instructor of CIS 1880, Eta Kappa Nu
***Technical Skills**: Software Infrastructure, Distributed Systems, Cloud, DevOps, Linux, Compilers/Software Standards*

Experience

Citadel Securities, *New York City, NY* **Sep 2023 – now**
Lead Software Engineer

- Worked on SDLC: maintaining and optimizing a **cloud-based build system** / CI (Bazel, Kubernetes, Nix) and developer tools (Nix, AI SWE agent). Promoted to tech lead in 6 months into joining full-time with explicit ownership over cloud infrastructure and developer tooling.
- Solved **system design, performance, and reliability problems** leading to over 70% faster CI/CD and 50% less cloud spend. Accelerate developer productivity through faster/better UX for internal tools, observability and test coverage - act as a solutions engineer to distribute tools and practices to all lines of business across US, EU and APAC operations.
- Wrote a **generative AI coding agent to act as a DevOps engineer** tasked with automating solutions to operational overhead/tech debt: migrations, testing, configuration, and runbooks. Highly leveraged SoTA research (context heavy, feedback driven search) for agent design.

Five Rings, *New York City, NY* **June 2023 – Aug 2023**
Software Developer Intern

- Optimised critical path function using **SIMD hashing** techniques (**C++, x86**), leading to ≈ 60% speedup (ns order).
- Worked on **thread-safe, efficient batch write** techniques (**C++**) using UDP, UDS sockets and shared memory.
- Optimised large-scale symmetric **matrix operations** using open-source tooling, parallel computation, and vectorised instructions (**C++**), resulting in a ≈ 50% speedup (ms order)

Stripe, *Seattle, WA* **May 2022 – August 2022**
Software Engineering Intern

- Wrote **Scala** to integrate new, **Memcached**-backed K/V store for ML features, created to replace AWS-hosted **Redis** (Elasticache). This project saves Stripe **$9.4M** annually.
- Took ownership and architected large-scale code rewrites and optimisations leading to **9x end-to-end** latency improvement to hit SLA targets.
- Worked closely with caching team to reproducibly test latency and consistency at **20k+ RPS** and **millisecond SLA**, devise various locking mechanisms to improve write throughput, as well as proactively find and fix bugs in their **Java SDK**.

Penn Labs, *Philadelphia, PA* **Oct 2020 – May 2024**
Co-Director/Team Lead/Backend Engineer

- Led student organisation of **30+** engineers, designers, and business developers to maintain and develop new products **(100k+ unique users)** for the Penn community.
- Oversaw management and all levels of the tech stack for Penn's official club repository (Penn Clubs), including developing the backend API (**Django**), optimising database queries, and managing K8s infrastructure.

Projects

Second Brain | *second-brain* **Nov 2023 - Mar 2024**

- Created an LLM-powered filesystem web-app with smart search, Q/A, automatic file categorisation, and summarization over user PDFs. **Self-hosted** everything but the LLM on an optimized K3s cluster.
- Utilized **RAG with Vespa** and **Mixtral 8x7b-Instruct** to deliver high-quality (better than ChatGPT) results on document Q/A and summarisation.
- Used the project to autonomously complete 2 homework assignments which received a full grade :)

Spruce Programming Language | *Spruce* **Nov 2022 – Dec 2022**

- Wrote a parser and interpreter in **Haskell** for a **custom, functional** programming language (named *Spruce*) with type hints, builtins, first-class functions, and lexically-scoped closures.
- Used a Monad transformer stack to replicate C-style **low-level concurrency** primitives (fork, wait) in the language, with shared memory (STM) and atomic blocks to support transactions over memory via a simple interface