NIKHIL GHIND

San Jose, CA | +1 (408) 304-8342 | nikhilghind19@gmail.com | www.linkedin.com/in/nikhil-ghind | www.github.com/nikhil-ghind

Incoming Software Engineer Intern, AI and Systems Software @ NVIDIA (Fall 2026)

Summary

M.S. Computer Science student and **Software Engineer** with five years building **high-throughput backend systems** in **Java** and **C++**. Shipped **event-driven** trading and payments infrastructure processing **$550M+ daily**, **low-latency** caching services, and **distributed** storage tooling across **billion-file** datasets.

Education

Master of Science | San Jose State University **August 2025 - Expected May 2027**
Computer Science - GPA: 3.7/4.0 | Distributed Systems | Software Design | Machine Learning

Bachelor of Engineering | University of Mumbai
Computer Engineering - GPA: 3.74/4.0 | Data Structures & Algorithms | Operating Systems | Probability and Statistics

Experience

NetApp **June 2026 - Present**
Software Engineer Intern

- Cut metadata scans from **8--10 hours to 45 minutes** across **10B+ files** by rebuilding NetApp's metadata engine as **GUFI**, a **multithreaded C++** indexer with a **custom SQL query engine**.
- Sustained near-linear **multi-core** scaling on **billion-file** traversals with a **lock-free work-stealing** queue, per-thread database connections, and careful **cache-aware memory** layout in **modern C++**.
- Accelerated model checkpointing on the **ONTAP distributed storage** platform via **delta checkpointing** in **PyTorch**, persisting only changed tensor state to cut write time and storage I/O.

Interactive Brokers **January 2023 - June 2025**
Software Engineer II

- Generated **$200M+** in new business with an **event-driven** Open Banking gateway in **Java** and **Spring Boot** integrating **Plaid**, fronted by **React**, serving **500k+** users.
- Halved read latency from **2.8s to 1.4s** with a **low-latency Java** caching microservice on **Kubernetes** fronting **Elasticsearch**, **Cassandra**, and a **distributed SQL** store, serving **~100k requests** per 10 minutes.
- Cut transaction processing time **90%** by refactoring legacy workflows into an FHSA regulatory-compliance **Spring Boot** microservice with **REST** APIs and **SQL**-backed audit trails shipped **into production**.
- Reduced **mean time to resolve** incidents **60%** with **Java** log aggregation into **Cassandra** and **Elasticsearch** dashboards across **distributed** trading infrastructure.
- Hardened bank-linking reliability to **99.9%** with idempotent retries, circuit breakers, and dead-letter replay on **Kafka** consumers, eliminating duplicate ledger postings during outages.
- Owned services end to end across **code review**, **JUnit**/integration testing, and **CI/CD** release, raising coverage on critical funding paths to **85%+**.

Software Engineer I **July 2020 - December 2022**

- Enabled processing of **$550M+** in daily deposits across **23 currencies** by architecting an **event-driven Java**/**Spring** pipeline on **Kafka** over a **distributed SQL** ledger, cutting unidentified deposits **50%**.
- Automated reconciliation with a rules-based matching engine over **PostgreSQL**, tuning **SQL** joins, window functions, and indexes to remove **thousands of monthly** manual touches on wire and ACH deposits.
- Mentored **5** new engineers through an onboarding program on **Java** services, **Kafka** messaging, and deployment workflows, cutting ramp-up to first production commit.

Debian **May 2019 - August 2019**
Google Summer of Code -- Student Developer

- Built a **Python** web crawler for **open-source** security tracking that normalized vulnerability patches across upstream repositories into structured records.

Projects

- **CloudPlay --- GPU Game-Streaming Orchestration (Java, Spring Boot):** Built four **microservices** --- API gateway, session orchestrator, **WebRTC** SDP/ICE signaling, and a distributed node registry --- placing sessions onto GPU hosts with a best-fit **bin-packing** allocator. Made reservations atomic via **Redis**/**Lua** with heartbeat leases so crashed sessions self-release, and drove **Kubernetes** Deployments/HPAs from a demand-based autoscaler over an **EKS** GPU pool provisioned by **Terraform**.
- **BlitzQueue --- Lock-Free MPMC Message Queue (C++17):** Implemented a **lock-free** multi-producer/multi-consumer queue over a bounded ring buffer using **atomics**, per-slot sequence counters, and explicit **memory ordering**, with **cache-line padding** to eliminate false sharing. Exposed the broker over **gRPC** and **Boost.Asio** with backpressure, and benchmarked **2M+ messages/sec** against a mutex baseline under **Google Test** and **ThreadSanitizer**.
- **Pythia --- Agentic RAG over 10k+ Documents (Python):** Built a stateful **LangGraph** agent (route → retrieve → reflect → generate) with **MCP** tool-calling over **500k+** chunks in **Pinecone**, re-querying on low-confidence context with inline citations. Served from **FastAPI** on **AWS** with caching and async fan-out that cut response latency **40%**, and tracked retrieval quality and regressions through **LangFuse** evaluations.

Achievements & Leadership

- **1st Place**, **Smart India Hackathon 2020** -- led a 4-member team; expanded **Open Banking** to **1M+ European customers**. **Published research** in the *Int'l Journal of Future Generation Communication and Networking*; presented at **ICCTAW** and **ICRACT**.

Skills

Languages: Java, C++ (11/14/17/20), Python, Go, C, SQL, JavaScript/TypeScript
Fundamentals: Data Structures & Algorithms, Concurrency, Lock-Free Programming, Operating Systems, Distributed Systems, Design
Backend & Data: Spring Boot, REST, gRPC, Node.js, React, Kafka, Elasticsearch, Redis, Spark, PostgreSQL, Cassandra, MongoDB, PyTorch
Cloud & Tools: Linux, Docker, Kubernetes, AWS, Terraform, Git, CMake, Google Test, CI/CD