Muneeb Kamran

muneebkamran2507@gmail.com | linkedin.com/in/muneeb-kamran-0a345722a | github.com/muneebkk

Education

Simon Fraser University Expected May 2028
Bachelor of Science, Computing Science, Minor in Statistics
Awards & Leadership: 1st of 220+ in Huawei Research Challenge, 2x MLH Winner, CS Peer Tutor, Residence Advisor

Experience

Royal Bank of Canada (RBC) – Amplify | Software Developer Intern May 2026 – Aug 2026

- Built a patent-pending fraud detection platform with **Python/FastAPI** and **React** to flag 20+ threat indicators in call transcripts and support analyst review; presented the system to 1,000+ attendees at RBC's internal expo
- Cut projected LLM costs by $1.5M/yr with a **scikit-learn** gate filtering 65% of calls; tracked evaluations in **MLflow**
- Secured **FastAPI** endpoints with **OIDC/JWT** validation and role-based access control through **Keycloak**
- Made **S3** ingestion on **Kubernetes** idempotent with retries and a dead-letter queue for failed transcripts
- Added **OpenTelemetry** tracing and **Grafana** alerts, catching a scoring regression before production

SFU Blueprint | Technical Lead sfublueprint.org ↗ Jan 2026 – Apr 2026

- Led 6 developers building a **Java/Spring Boot** platform for a poverty-reduction nonprofit serving 1,000+ MAU
- Cut lookups from 820ms to 310ms with a composite **PostgreSQL** index and batch fetching to fix N+1 queries
- Automated **AWS** deploys with **Terraform/GitHub Actions**, adding health checks and automatic rollback
- Wrote **JUnit** unit and API integration tests for membership approvals, access controls, and invalid requests

OpenQQuantify (AI Design Automation) | Backend Software Engineer Intern Aug 2025 – Dec 2025

- Built an AI design backend in **Java/Kafka**, ordering 50K+ events/day at 10x peaks with tenant-keyed partitions
- Profiled **Java** threads with **VisualVM**, cutting API p95 from 380ms to 270ms with a fixed thread pool
- Built a per-tenant **Redis** sliding-window rate limiter, reducing 5xx errors by 45% during traffic spikes
- Added **Prometheus** metrics and correlation IDs across four services, cutting debugging time from 25 to 5 minutes

CJSF Radio | Backend Developer Apr 2025 – Jul 2025

- Built a **FastAPI/Elasticsearch** API with fuzzy search over 80K+ audio records, supporting 500+ concurrent users
- Built archive notifications with **Redis Streams and WebSockets**, delivering updates in under 200ms end to end
- Fixed a race condition behind 15% of scheduled broadcast failures with **Redis** distributed locks on time-slot claims

Projects

CaseComp.ca casecomp.ca

- Launched a **React/TypeScript/Node.js** case competition platform for 2,000+ users across 10+ universities
- Made **Stripe** webhooks idempotent by deduplicating event IDs, preventing duplicate payment processing
- Cut page load from 1.8s to 240ms with **PostgreSQL** query tuning; added **OpenTelemetry** tracing

Coalesce, Batch Inference Server

- Built a **Go/gRPC** inference server with size- and time-based batching, reducing per-request overhead 3x
- Added backpressure to reject excess requests under overload; exported batch latency to **Prometheus**
- Compared dynamic and fixed batching under traffic spikes using **k6**; profiled goroutine contention with **pprof**

Huawei Challenge, Memory-Constrained DAG Scheduler

- Designed a **C++** beam-search DAG scheduler for strict memory limits, solving 10K+ node graphs in under 2s
- Cut peak memory 40% vs baseline; validated small-graph solutions against **Gurobi** MILP optima
- Used critical-path priorities in **C++** to reduce peak memory versus greedy and runtime versus branch-and-bound

Technical Skills

Languages: Java, Python, Go, C++, JavaScript, TypeScript, SQL
Frameworks & APIs: React, Node.js, Spring Boot, FastAPI, gRPC, REST, OpenAPI
Data & Messaging: PostgreSQL, Redis, Kafka, Elasticsearch
Cloud & Infrastructure: AWS, Kubernetes, Docker, Terraform, GitHub Actions, Linux
Observability & Tools: OpenTelemetry, Prometheus, Grafana, k6, pprof, VisualVM, scikit-learn, MLflow