

RAKHI AGRAWAL

Toronto, ON | rakhiagr2020@gmail.com | +1 602-503-9868 | linkedin.com/in/rakhi-agr

SUMMARY

Software Engineer with 4+ years of experience building and scaling large-scale distributed systems and backend infrastructure at LinkedIn. Specialized in high-availability backend services, streaming data platforms, and reliability engineering, with expertise in Kafka/Flink architectures, MySQL scalability, and metadata systems processing millions of daily events.

EXPERIENCE

Software Engineer, LinkedIn (Mountain View, CA & Toronto, ON) Jan 2023 – Present

- Improved reliability of a metadata ingestion platform processing **170M+ daily events** on Flink/Kafka by building automated failure handling, DLQ replay tooling, observability, and alerting systems. Defined OpenTelemetry-based health metrics and operational tooling that helped achieve a **99.99% ingestion success rate**.
- Re-architected a compliance-critical metadata platform from single-primary to **multi-primary MySQL** architecture under live production traffic serving **15K read QPS** and **10K write QPS**. Built dual-read/write migration and backfill systems enabling seamless production cutover without downtime or data loss.
- Built a centralized cross-system retention-policy and enforcement platform for HDFS and Iceberg-based data lake datasets as part of LinkedIn's **HDFS Corpus Reduction** initiative projected to save **\$55M+ annually**. Developed lineage-based policy resolution, retention synchronization, and validation workflows across **30K+ datasets**, establishing a unified retention enforcement and verification layer across distributed storage systems.
- Reduced multi-hop graph query latency from **18s to 3.9s** average and high-volume query execution time from **4–5s to 0.05s**, while eliminating schema-evolution table lock times from **1000+ seconds to near zero** through query optimization, indexing, and application-level performance tuning across large-scale MySQL backed systems.
- Led migration of **5 critical metadata ingestion pipelines** from Samza to Flink, processing **250M+ records/day** across **500+ Kafka topics**, and drove seamless production rollout across large-scale streaming infrastructure.
- Designed and implemented a **reusable metadata ingestion framework** to standardize workflow execution across REST and gRPC systems, simplifying integration of new ingestion workflows and reducing onboarding effort from **2 weeks to 3 days**.
- Led reliability engineering initiatives across **10+ services** and **100+ gRPC endpoints** by defining uptime targets, observability standards, alerting, and production readiness practices.
- Built **AI-powered developer tooling** to automate on-call summarization and local testing workflows, reducing manual operational effort and improving testing reliability across distributed services.
- Mentored **3 engineers** through onboarding and design reviews, accelerating ownership of large-scale distributed systems.

Software Engineer Intern, LinkedIn (Mountain View, CA) May 2022 – Aug 2022

- Developed APIs & validation frameworks to enforce data-quality standards across LinkedIn's offline data lake, preventing malformed assertions from triggering production failures and reducing operational on-call burden caused by invalid assertions.

Back-End Developer, Inspectinity Private Limited (India) Aug 2020 – Dec 2020

- Developed **20+ REST APIs** using Python and Flask for high-volume backend workflows, optimized PostgreSQL queries for high-traffic components, and built reusable modules for authentication, request validation, and data-processing workflows, improving backend performance, developer productivity, and **reducing infrastructure costs by 20%**.

PROJECTS

LINGO: Visual Analytics for NLP Dataset Bias Aug 2021 – Dec 2021

- Built an interactive visualization system for analyzing NLP instruction bias using GPT-3 outputs and semantic similarity analysis, featuring Beeswarm plots and chord diagrams; accepted to **EuroVis 2023**.

Auto-Scalable Facial Recognition System Jan 2022 – Mar 2022

- Built a distributed facial-recognition system using AWS EC2, SQS, S3, and Python with **auto-scaling** across EC2 instances, reducing total computation time by **50%** through parallel workload distribution.

TECHNICAL SKILLS

Languages: Java, C++, SQL, Python, Scala, JavaScript, Shell

Backend/Data Infrastructure: Distributed Systems, Microservices, gRPC, REST, Protobuf, Kafka, Flink, Samza, MySQL, PostgreSQL, MongoDB, Couchbase, Elasticsearch, HDFS, Iceberg

Cloud, Monitoring & Developer Tooling: AWS (EC2, S3, Lambda, DynamoDB, SQS), Airflow, Docker, Kubernetes, Linux, Git, DataHub, Grafana, Kusto/KQL, Claude Code, Codex, Cursor, Windsurf

EDUCATION

M.S. in Computer Science, Arizona State University

GPA: 4.0/4.0

Jan 2021 – Dec 2022