

Cannon Kalra

Data Platform Engineer — Distributed Systems & Streaming

📍 Delhi, India

cannonkalra@gmail.com ✉

+91 99144 33993 📞

cannonkalra 🌐

cannonkalra in

cannonkalra.github.io 🌐

SUMMARY

Engineer who designs, builds, and operates large-scale distributed data platforms end to end — event-driven ingestion, lakehouse storage, and analytical serving. Owns the hard guarantees: exactly-once correctness, fault tolerance under partition, and predictable latency and cost at **billions of events/month**. Deep in streaming systems, analytical-database internals, and the architectural tradeoffs behind reliable, performant platforms at scale.

EXPERIENCE

Lead Data Engineer - Teleparty

Jun 2024 - Present

- Designed and own a **multi-tenant** data platform processing **2B+ events/month** with exactly-once semantics, architected for backpressure and graceful degradation at scale
- Architected a tiered storage model — an **Apache Iceberg** lakehouse as system of record and **Apache Doris** for low-latency analytical serving — resolving the batch-consistency vs. real-time-query tradeoff behind one source of truth
- Engineered the **Doris** serving layer for sub-second analytics on high-cardinality event data: **Merge-on-Write** tables, partitioning and bucketing, materialized views, and compaction tuned for freshness vs. query cost
- Designed distributed batch as a **medallion architecture** with partition-aware scheduling and adaptive query execution for reliable replay and incremental compute: **10x** throughput, **35%** lower cost
- Own platform-wide **SLOs**, observability, and data contracts across engineering, data science, and product; mentor engineers on streaming and analytical-database design

Fractional Cloud & Data Architect - Independent / Freelance

Sep 2023 - Jun 2024

- Migrated **300+ Kubernetes services** across clouds (AWS to Azure) via GitOps (ArgoCD, Helm, Kustomize) at **zero downtime**; led incident response on a compromised cluster, re-platforming to hardened AWS ECS
- Built repeatable **Terraform** infrastructure platforms enabling self-service deployment and operational consistency (provisioning **-60%**, spend **-30%**); designed streaming->lakehouse->serving reference architectures for clients

Platform Lead - Solulever

Feb 2023 - Sep 2023 / Jan 2020 - Mar 2022

- Designed an **event-driven** real-time IoT platform ingesting high-frequency telemetry from globally distributed factories, built for low-latency delivery and horizontal scale on Kafka and Spark Streaming
- Built a **metadata-driven ingestion framework** with Kappa-style replay and outage-resilient buffering, guaranteeing **100% capture** during factory network partitions
- Removed processing bottlenecks via async queue redesign and horizontal scaling, improving **equipment uptime 20%** and operational efficiency **15%**

Technical Lead - Elucidata

Mar 2022 - Jan 2023

- Led a **300TB+ biomedical data platform** on Delta Lake powering AI-driven discovery over genomics and clinical datasets
- Re-architected analytics from **OpenSearch to a Delta Lake lakehouse** on Athena — trading document search for SQL + multi-schema analytical access — and consolidated fragmented ETL into Airflow DAGs, cutting runtime **40%**
- Hardened pipelines with validation checkpoints, automated backfills, and lineage/cataloging for **SOC2/GDPR** compliance, eliminating silent data corruption

Software Engineer - FutureSoft

Jul 2019 - Jan 2020

- Re-architected backend for **3,000+ IoT devices**, decoupling ingestion from processing to cut latency and lift release velocity **30%**; reduced OTA update size **150x** via edge-gateway design patterns

EDUCATION

B.Tech, Information Technology

Punjab Engineering College

Dec 2018

Chandigarh, IN

INTERNSHIP

Research Intern

LSDMA, KIT Campus Nord

Feb 2015 - Jul 2015

Karlsruhe, DE

SKILLS

Programming

Python SQL Scala Go
PySpark Pandas Terraform
Docker Git CI/CD

Technologies

Spark Kafka Airflow Iceberg
Delta Lake Apache Doris AWS
Kubernetes ArgoCD PostgreSQL
Athena Glue gRPC

Familiar

Flink GCP BigQuery
Trino dbt Kinesis Java
OpenTelemetry Prometheus
Grafana

IMPACT

- 2B+** events/month, exactly-once
- 10x** throughput, **-35%** cost
- Sub-second serving on **Doris**
- 300+** K8s services migrated

ACTIVITIES

- Homelab – Self-hosted infra
- FOSS – Contributor
- Handball – National level