

Farhan Ali Khan

Software Engineer and Problem Solver

arriqaaq321@gmail.com · linkedin.com/in/arriqaaq · github.com/arriqaaq · farhan.arriqaaq.com

SOFTWARE ENGINEER: STORAGE, DISTRIBUTED SYSTEMS AND AGENT MEMORY

Software engineer and problem solver with 13 years across startups and large companies. Led the storage effort at SurrealDB, data engineering at GreedyGame and infrastructure at HapPay, owning problems from the customer's need to production. Builds storage engines and distributed systems, and checks them with formal models and simulation testing.

LEADERSHIP AND OWNERSHIP

- Led the storage effort at SurrealDB, the data engineering effort at GreedyGame and the infrastructure side at HapPay.
- Mentored and guided teams: concurrent Go and testing practice at Cisco, code review and end-to-end testing at VMware.
- Worked close to users: end customers at HapPay, business and operations teams at GreedyGame, platform tenants at VMware (SRE on-call).
- Four startups (SurrealDB, Codenotary, GreedyGame, HapPay) and three large companies (VMware, Cisco, Tata Consultancy Services).
- Starts from the problem and first principles, then chooses tools; verification engineering (TLA+, Kani, Verus, property, mutation and simulation testing) for systems that must stay correct.

SKILLS

Storage and databases: LSM tree (log-structured merge tree), B+ tree, key-value stores, RocksDB, write-ahead log, compaction, adaptive radix tree, object storage, lakebase architecture

Distributed systems: distributed transactions (TAPIR, Transactional Application Protocol for Inconsistent Replication), Raft, Viewstamped Replication, SWIM gossip protocol, CRDTs (conflict-free replicated data types), replication

Verification: TLA+, Kani, Verus, property-based testing, mutation testing, deterministic simulation testing

AI: agent memory, ontology design, fact and entity modelling, AI coding agents (Claude Code)

Platform and data: Kubernetes operators, Docker, Terraform, ELK stack (Elasticsearch, Logstash, Kibana), directed acyclic graph (DAG) workflows, enterprise resource planning (ERP) integrations, Amazon Web Services (AWS), Google Cloud Platform (GCP), gRPC, Prometheus, Grafana, Kafka, BigQuery, Redis, Aerospike, PostgreSQL, MySQL

Languages: Rust, Go, Python, SQL · **Leadership:** team lead, mentoring, design and code review, customer and product work

EXPERIENCE

SurrealDB · Senior Software Engineer, Storage Team Lead

Jun 2023 - Oct 2026

Database startup: a multi-model database (document, graph, vector). Led the storage effort.

Storage engines and team

- Led the storage effort, working on optimisations and on what the company needed from storage.
- Built SurrealKV, an LSM-tree key-value store in Rust, from scratch, and made it performant.
- Worked on a copy-on-write B+ tree key-value store, built internally for our use case.
- Worked on performance improvements in RocksDB and at the storage-engine level.
- Did low-level optimisation on in-memory Rust crates such as vart (a versioned adaptive radix trie) and several others.

Distributed and cloud storage

- Added object storage support on top of our LSM engine, to build a lakebase architecture for the company's growing needs.
- Worked on our distributed transactional store, built on the TAPIR paper (Zhang et al., SOSP 2015), and did deterministic simulation testing (DST) across our internal products.

AI and agent memory

- Worked on SurrealDB's agent memory product.
- Worked on understanding facts and entities and the logic around them, and on defining ontologies for the product.
- Used an agentic workflow with TLA+ modelling before writing code, then Kani checks, Verus proofs, and property and mutation testing.

Codenotary · Software Engineer (Consultant), immudb

Aug 2022 - May 2023 · Remote

Startup. immudb: open-source immutable database with cryptographic (Merkle tree) verification; tamper-resistant and auditable.

- Worked on data retention, implementing efficient deletes in the database, where proofs and storage are interlinked.

- Worked on compaction and on improving key-value store performance.
- Developed new features, maintained core components and tested the database.
- Engaged with the immudb community.

VMware · Senior Software Engineer, VMware Developer Platform

Apr 2021 - Jul 2022 · Bangalore

Large company. Internal developer platform running many Kubernetes clusters with thousands of nodes under PCI, HIPAA and FedRAMP High compliance, with separate clusters for development, staging, production and sensitive workloads across data centres.

- Built Kubernetes operators for the platform, including an operator for multi-cluster management across clouds that gave tenants Kubernetes federation and multi-tenancy.
- Designed and built a self-service model for platform-managed services: tenants deploy curated Kubernetes apps (Kafka, Vault, the Prometheus monitoring stack) with the team's deployment practices built in, instead of installing and running each one from scratch.
- Managed platform releases through a Kubernetes-native release operator built by the team.
- Took SRE on-call shifts, which showed where tenants struggled and kept me hands-on with their issues.
- Introduced an end-to-end (E2E) testing framework and behavioural tests written from the tenant's point of view, so operators were tested before tenants used them.
- Improved the team's peer code review process to raise code quality and collaboration.

Cisco · Senior Software Engineer, Secure Threat Services

Apr 2019 - Apr 2021 · Bangalore

Large company. Containerised firewall service that automates network operations and monitors customer networks, with fast router onboarding and near real-time configuration updates.

- Worked on moving firewall deployments from a centralised architecture to a decentralised peer-to-peer one, with a Go management layer running alongside the Snort firewall on each router.
- Researched group membership algorithms and built a gossip-based membership mesh across routers on SWIM (Scalable Weakly-consistent Infection-style Process Group Membership), using HashiCorp memberlist.
- Built an in-memory shared key-value store for configuration and version updates, gossiped across the cluster and ordered with Lamport clocks.
- Designed a more robust successor that models the data as CRDTs (conflict-free replicated data types) to remove clock-skew problems.
- Built core STS services: designed the gRPC APIs, coordinated with other teams, planned deployment strategies for testing the distributed setup, and guided the team on concurrent Go, design patterns and testing strategy.

GreedyGame · Senior Software Engineer, Data Engineering Lead

Jun 2017 - May 2019 · Bangalore

Ad-tech startup: an ad platform for mobile games serving more than 500 million ad requests a day. Led the data engineering effort and worked closely with business and operations teams.

Ad serving

- Designed the real-time bidding (RTB) protocol: each request broadcasts bid requests to bidders (AdMob, Kraken, Facebook and others), evaluates responses, notifies the winner and serves the ad markup.
- Researched RTB papers and designed an asynchronous flow that prepares ads for subsequent requests, keeping every response under 500 ms.
- Set up the microservice infrastructure on Google Cloud (Kubernetes, Terraform, Consul) with a caching layer on Redis and Aerospike.
- Reduced infrastructure spend through algorithm and caching optimisations while holding the latency target.
- Built the monitoring stack (Prometheus, Grafana) with real-time cache dashboards the business used for scaling decisions. Slides: Ad serving platform using Redis (RedisConf 2018).

Data engineering (lead)

- Led the data engineering effort: built the company's first data pipelines on Kafka and BigQuery, ingesting more than 2 TB a day.
- Converted raw gzipped data to Parquet and Cap'n Proto to make transfers cost-effective, and optimised BigQuery queries with the data team.
- Cached query results in Redis data structures for faster queries; optimised Aerospike queries and cost, giving the operations team data to choose which campaigns to run during the day.
- Built BigQuery dashboards that surfaced errors in the Android SDK.
- Set up a highly available MySQL cluster on GCP with replication and backups for critical data, and a data lake on Amazon Glacier with pipelines for querying historical data.

HapPay · Senior Software Engineer, Infrastructure Lead

Oct 2015 - May 2017 · Bangalore

Fintech startup: a corporate expense management product. Led infrastructure and worked directly with customers.

- Worked directly with end customers.
- Designed and built role-based access control (RBAC), a core feature that manages the hierarchy of users in each customer organisation.

- Built and managed integrations with third-party APIs, and the public API HapPay exposed to partners (Dunzo integrated with it in 2016).
- Built accounting integrations that sync customers' expenses into SAP, Oracle, QuickBooks, Tally and other ERP tools.
- Built an engagement workflow engine (DAG / state machine) to track the customer lifecycle, scaling the in-house streaming pipeline, Elasticsearch and scheduler.
- Led infrastructure for the product: Docker and AWS, log management and analysis across services on the ELK stack, and scaling, monitoring and migrating the critical MySQL and PostgreSQL databases.

Tata Consultancy Services · Associate Software Engineer

Jul 2013 - Aug 2014 · Bangalore

- Built custom SharePoint 2013 and 2010 solutions in Visual Studio for enterprise clients: web parts, site definitions, designer workflows, event receivers and timer jobs, with HTML5, CSS3 and JavaScript.
- Customised client applications with the server and client object models, delivered enhancements and change requests, and debugged critical defects.

OPEN SOURCE

flashdb (embedded key-value database, 370+ GitHub stars), chord (Chord DHT), art (adaptive radix tree), merkle tree (RFC 6962 proofs), vsr (Viewstamped Replication). Engineering blog: farhan.arriqaaq.com/blog.

EDUCATION

B.Tech, SRM University, Chennai · 2009 - 2013