

Faizan Tariq

Full Stack Engineering | Scalable Architecture | Agentic AI Transformation & Strategy | Cloud-Native
(479)430-6050 • faizaan.tariq@icloud.com • <https://www.linkedin.com/in/faizaantariq> • <https://www.faizantariq.com> • Bentonville, AR

SUMMARY

13+ years architecting scalable, cloud-native platforms and agentic conversational AI systems, driving multi-million revenue impact through AI-native transformation at scale, cost optimization and enterprise-scale enablement. Skilled in business alignment, stakeholder communication and building AI-native engineering cultures.

EXPERIENCE

Confiz LLC Bentonville, AR
Principal Engineer - Agentic AI and Data 05/2024 - Present

- Architected a scalable multi-agent conversational AI system via LangGraph and AI Foundry, for store associates scaled from **2 pilot stores to 500+ stores**.
- Influenced leadership and stakeholders by **reducing inference cost 20x by advocating for Nano model** and preserving response quality through prompt engineering.
- Built multimodal RAG with semantic search index of **500+ documents** and another index of **50K products** on Azure AI Search, backed by an ETL pipeline that **refined 170K products to 50K** by eliminating duplication and cleansing data.
- Improved latency for 10,000+ concurrent users** by implementing a pgVector-based semantic cache (evaluated against RedisVL), hitting a **10s streaming SLA (avg 6.5s, P90 8s, P95 9.5s)** by semantically caching intent routing, soft guardrails, hard guardrails and responses as measured by Locust.
- Designed and reviewed PostgreSQL-based text-to-SQL agent for store-specific inventory, pricing, and product location queries.
- Provisioned infrastructure as code via Terraform for consistent, repeatable deployments across environments.
- Researched and implemented **LLM-as-Judge, achieving 85% evaluation score** across correctness, faithfulness, relevance, and completeness with detailed reasoning.
- Established and validated a continuous feedback loop with telemetry for observability and measured **80% user adoption** among store associates.
- Addressed **multi-turn conversation with topic anchor**, preserving relevant context without full chat history, reducing hallucination and context window overflow.
- Architected a **scalable multi-tenant, multi-agent framework** on LangGraph and AWS Bedrock Agent Core with predefined templates, enabling teams to provision conversational AI agents, tools and knowledge-bases on demand via parameterized configuration and deploying onto Agent Core Runtime.
- Enforced RBAC/ABAC via JWT-based identity propagation and secured RAG retrieval** with shared OpenSearch collection and siloed indices, enabling fine-grained, document-level access control and data isolation.
- Led an AI incubation program enabling AI native SDLC and collaborative teamwork culture, mentoring engineering teams to become AI-Native, experimenting and architecting **8 enterprise-scale AI/ML solutions in 12 months** on Databricks and Azure AI Foundry around programmatic media ads, campaigns, recommendation engines, forecasting, segmentation, customer journey and loyalty by relying on Log Level Data (LLD) and behavioral vector embeddings to build Agentic AI solutions to fuel pre-sales and growth across different accounts. This helped **acquired 4 new accounts** which generated **10M+ in revenue in 12 months**.
- Automated data discovery via scheduled GCP query jobs to inspect data flow through Airflow DAGs across Assortment, RDC, and Replenishment in supply chain system to uncover critical anomalies **impacting 300K+ EDW records, 170K+ missed intake records, 70K+ stuck demand requests**, and render on Tableau dashboard.

Confiz LLC - Contractor (Sam's Club - Sam's Garage) Bentonville, AR
Sr. Staff Architect 01/2021 - 04/2024

- Resolved data and latency issues** on a next-gen automobile retail and repair platform by optimizing performance and scalability across **1TB+ of production data**.
- Resolved Azure CosmosDB RUs by optimizing queries into point reads, **reducing RU consumption by 60%** (10,000 to 4,000 RUs), reducing latency and cost at scale.
- Remodeled high-latency search module by advocating and deploying self-managed Elasticsearch and Kibana via Helm on Kubernetes to resolve memory issues, enabling distributed pod-level replication for DR with automatic cluster-manager-node failover and meeting **P90/P95 SLAs of 800ms/1000ms**.

Confiz LLC - Contractor (Walmart - Cloud Powered Checkout) Bentonville, AR
Software Architect 08/2019 - 01/2021

- Modernized legacy POS system with Cloud Powered Checkout (CPC), stateless microservices for Cart, Checkout and Payment based on event-driven architecture.
- Rolled out CPC to 6,000+ stores** across the US and Canada market in a phased canary rollout including both cloud and edge-based offline deployments.
- Executed observability via Dynatrace and Splunk to enable monitoring, alerting, real-time telemetry and centralized logging.
- Implemented multi-store velocity controls for fraud mitigation, capping gift card purchases within a rolling 24-hour window across stores.
- Architected migration from legacy ETL to a real-time Kafka pipeline to publish store-specific tare weights to per-store topics and serving them via a MeghaCache backed microservice within **20ms SLA**.
- Integrated BigQuery data with Looker to build transactional and business KPI dashboards for CPC.

Confiz Limited - Contractor (Walmart - Cloud Powered Checkout) Lahore, Pakistan
Principal Software Engineer 07/2018 - 08/2019

- Modernized a legacy POS discounts and coupons system into a scalable, cloud-native microservices-based promotions engine, **ensuring 40ms SLA** and high availability via cloud and edge-based offline deployments, **rolled out to 6,000+ stores** across the US and Canada market.
- Automated ZSD** with blue-green deployment to ensure **high availability (HA)** and release tags based systematized rollbacks to ensure **disaster recovery (DR)**.
- Executed observability via Dynatrace and Splunk to enable monitoring, alerting, real-time telemetry and centralized logging.
- Setup DevOps and CI/CD pipelines for unit tests, functional tests, contract tests and automated performance tests.

Confiz Limited Lahore, Pakistan
Sr. Software Engineer 06/2013 - 07/2018

- Built 4 iOS/Android native apps for a subscription-based **multi-tenant** enterprise suite covering digital gifting, enterprise messaging, prospecting, and live-streaming.
- Scaled across 5 tenants, each with 5-10K active users**, on a shared multi-tenant architecture.
- Saved 50% dev effort** by migrating 8 native apps across iOS/Android into 4 React Native apps, centralizing shared logic and bridging native modules where needed.
- Set up true CI/CD pipelines on Jenkins to **reduce time to release** and enabled **4 deployments/week**.
- Setup test automation with **85% unit test coverage on Sonar, 80% functional test coverage on Cucumber**, and **80% instrumentation test coverage with Appium**.

EDUCATION

FAST - NUCES Lahore 2009 - 2013
BS (Computer Science)

Government College University (GCU), Lahore 2007 - 2009
ICS (Computer Science)