Sameer Raghunath Deshpande

Senior Solutions Architect – Cloud & Solutions Architecture (Azure)

Pune, India • LinkedIn: linkedin.com/in/sameer-deshpande-95180429 • Contact via portfolio website

Summary

Enterprise Solutions & Cloud Architect with **17+ years** delivering cloud-native platforms across **retail, fintech, and industrial** domains. Blends deep technical architecture with delivery leadership and business alignment — pairing **architecture strategy** with hands-on implementation. Known for establishing **architecture standards**, leading **governance** across multi-team programs, and building **high-availability systems** that hold up in production. Strong hands-on background in **Azure**, **distributed systems**, and **DevOps transformation**, with practical experience designing **AI-enabled platforms** using **vector search, retrieval, and semantic matching**.

- **Core: Azure**, **AWS**, **Distributed Systems**, **Event-Driven & Microservices**, **CQRS**, **Domain-Driven Design**, **Node.js**, **Java / Spring Boot**, **C# / .NET**, REST, **GraphQL**, PostgreSQL, Azure SQL, Cosmos DB, Redis, Elasticsearch.
- **AI / Agentic: Azure AI Foundry**, **Azure OpenAI**, **RAG pipelines**, **Vector Embeddings**, **MCP (Model Context Protocol)**, **AI Agents / Agentic Workflows**, Human-in-the-Loop, LLM Observability, **OWASP LLM Top 10**, privacy-safe retrieval.

Selected Architecture Leadership Highlights

- **Parallel Enterprise Architecture:** Architected multiple enterprise applications **in parallel** across retail and fintech, owning **end-to-end design decisions** and staying hands-on through implementation and delivery.
- **Reusable Azure Blueprints:** Applied **Azure reference architectures** and reusable **IaC patterns** across programs, accelerating solution design by building on proven blueprints rather than starting from scratch.
- **Resiliency & DR:** Implemented resiliency, **disaster recovery**, and observability practices for critical platforms — maintained **RTO under 4 hours** and **RPO under 2 hours** through tested failover and backup strategies.
- **Shared Platform Foundations:** Built reusable backend modules and **shared libraries** across applications, reducing duplicated effort and keeping a consistent, maintainable stack across teams.
- **Stakeholder & Roadmap Ownership:** Served as the **technical point of contact** for stakeholders — contributing to roadmap discussions, effort estimation, and long-term program continuity.
- **Cost Optimization (~$50K/yr):** Drove infrastructure cost optimization of **~$50K annually** through CDN strategy, **architecture consolidation**, and building in-house applications instead of licensed COTS tools where custom builds were the better fit.
- **Demanding NFRs at Scale:** Engineered platforms to **API p99 < 200ms**, **≥ 1000 req/s** sustained, **99.95% availability SLO**, **RTO 4h / RPO 2h**, **MTTR 30 min** — using zone-redundant multi-region **Azure PaaS**, Redis caching, Cosmos auto scale, Front Door failover, and App Insights observability with SLO-based alerting.
- **DevSecOps & Supply-Chain Security:** Established a software supply-chain security baseline across programs — **SBOM** generation, software-composition scanning (**Snyk**, **Docker Scout**) for container and dependency BOM vulnerabilities, and **SonarQube** static-analysis quality gates — enforced as automated **CI/CD** release gates to shift vulnerability detection left and reduce production risk.

Technical Skills

– **Cloud:** Azure (App Service, Functions, Event Hub, APIM, Cosmos DB, Data Factory, CDN, Front Door), **AWS**.

– **Architecture:** Distributed Systems, Event-Driven, Microservices, **CQRS**, **Domain-Driven Design**, C4 Modeling, API Design (REST, **GraphQL**).

– **Backend: Node.js**, **Java / Spring Boot**, **C# / .NET**, REST APIs, **GraphQL**.
– **Data: PostgreSQL**, Azure SQL, **Cosmos DB**, **Redis**, Elasticsearch, MongoDB, **ltree** (hierarchical tree queries), Sparse Vectors.
– **DevOps / IaC: Terraform**, Azure DevOps, GitHub Actions, **Docker**, Jenkins, CI/CD Automation, **Spec-Driven Development (GitHub Copilot)**.
– **DevSecOps / Supply Chain: SBOM** (CycloneDX / SPDX), **Snyk**, **Docker Scout**, **SonarQube** (static analysis / quality gates), dependency & container scanning, secrets scanning, OWASP.
– **AI / ML: Azure AI Foundry**, **Azure OpenAI**, Vector Embeddings, Sentiment Analysis.
– **GenAI / Agentic: RAG Pipelines**, **Vectorless RAG**, Privacy-Safe Retrieval, **MCP (Model Context Protocol)**, **AI Agents / Agentic Workflows**, Human-in-the-Loop, LLM Observability, **OWASP LLM Top 10**.
– **Frontend: React**, **Angular**, JavaScript (ES6+), HTML5 / CSS3.
– **Governance:** Architecture Reviews, Technical Roadmaps, **Agile / Scrum**, TDD, Cost Optimization.

Professional Experience

Vivicta, Pune
May 2015 – Present

Senior Solutions Architect *2023 – Present*

> *Design authority and principal architect for enterprise **Azure** programs across retail verticals.*

- **Architecture Governance:** Defined and governed **end-to-end solution architecture** across active delivery portfolios; chaired **architecture review boards** and led design approvals.
- **AI-Assisted Resume Ranking:** Built a **hybrid retrieval architecture** using **Azure AI Foundry** embeddings and **pgvector** similarity search — reducing recruiter screening effort noticeably in early rollout.
- **SolutionHub Vectorless RAG:** Architected **SolutionHub**, an internal documentation Q&A platform indexing ~3,500 files, using a **vectorless RAG** design — Postgres **ltree**-backed document trees navigated by pure LLM reasoning, complemented by **sparse vector** search — removing the need for a dedicated vector database.
- **HA Feedback Platform:** Architected a highly available customer feedback platform on Azure (**Front Door + Static Web Apps**) handling sustained daily load; integrated **AI-based content moderation** that cut time-to-action for moderators.
- **Responsible-AI Roadmap (Kesko):** Authored a **four-phase responsible-AI** feedback modernization roadmap — sentiment automation with **human-in-the-loop**, **privacy-safe RAG** over an Azure AI Search vector store, and an **MCP agent layer** — with PII-vault, **OWASP LLM Top 10**, and per-feedback token-cost governance applied across every phase.
- **Cloud Platform Governance:** Defined **Azure reference patterns**, **Terraform** IaC standards, and security baselines adopted across programs.
- **Delivery & FinOps:** Maintained strong budget efficiency and capacity utilization across programs through delivery governance and cost-control cadences.
- **Spec-Driven Development:** Introduced **spec-driven development** using **GitHub Copilot and Cursor** — turning architecture specifications into scaffolding, tests, and consistent implementation, tightening the loop between design intent and delivered code across teams.

Lead Architect *2020 – 2023*

> *Hands-on architecture and delivery leadership for **event-driven** and **data-heavy** platforms.*

- **Event-Driven Loyalty:** Architected an **event-driven microservices** ecosystem on **Azure Event Hub** for high-throughput loyalty transaction processing; designed secure partner onboarding and **data isolation** frameworks.
- **Enterprise Data Pipeline:** Designed a pipeline integrating **Stibo Step PIM**, **Hasura GraphQL**, and **PostgreSQL**; Redis caching reduced API latency by **65%**; **Azure Data Factory** processed large-scale product data across **three regions**.
- **CI/CD Transformation:** Transformed CI/CD through **Azure DevOps** pipelines, cutting deployment cycles from several hours to **under 15 minutes**.

Technical Architect *2017 – 2020*

> *Solution design and hands-on development for large-scale **e-commerce** and **digital asset** platforms.*

- **Plussa Loyalty Ecosystem:** Served as **v1 architect** and **v2 core developer** for a production loyalty platform supporting a large active user base and generating meaningful commercial revenue.
- **Frontend Modernization:** Designed a modular frontend architecture for **K-Rauta.fi** and modernized the **Onnshop** platform on **AWS**, integrating SAP, Informatica, and Contentful CMS.
- **Digital Asset Delivery:** Architected a **cloud-native** digital asset delivery platform managing a large media library across four enterprise applications; CDN strategy and automated cache policies delivered significant infrastructure cost savings.
- **Industrial IoT Pipelines:** Built **Python-based IoT** data processing pipelines for industrial telemetry ingestion with strong operational reliability.
- **IaC Foundations:** Standardized cloud delivery foundations through **Terraform** IaC, establishing patterns later scaled across **10+ programs**.

Senior Developer / Solution Designer *2015 – 2017*

> *Full-stack development and early architecture involvement on **Azure cloud migration** and e-commerce programs.*

- **Full-Stack Delivery:** Delivered backend and frontend components for retail and loyalty platforms; contributed to early **cloud migration** design on Azure.

Accenture, Pune *Nov 2012 – May 2015*

Software Engineering Senior Analyst

> *Enterprise UI and backend development for a global **sales order management** platform (ETR 3.0).*

- **Enterprise Interfaces:** Built cross-browser enterprise interfaces handling significant daily transaction volume using **Java/J2EE**, JSP, and jQuery.
- **Real-Time Validation:** Introduced **AJAX-driven** real-time validations, measurably reducing order-entry error rates.
- **Recognition:** Recognized as **CAPG** (second highest performer among **200+ Senior Analysts**); received the **ACE Award** for delivery excellence.

Technologies: Java/J2EE, JSP, jQuery, AJAX, Oracle

Infosys Ltd., Pune *Sep 2008 – Oct 2012*

Technology Analyst

> *Full-stack development for enterprise **sales, service, and e-learning** portals for global clients.*

- **Enterprise Portals:** Engineered large-scale portals using **Java/J2EE**, Struts, JSP, and AJAX; resolved complex cross-browser compatibility issues across legacy and modern browsers.
- **Performance Optimization:** Improved frontend performance notably — reduced page load from **4 seconds to 1.2 seconds** on an e-learning platform.
- **Team Leadership:** Led a small development team; conducted technical design reviews and mentored junior developers.
- **Recognition:** Maintained **CRR 1+** for two consecutive years; received the **E&R Excellence Award** and **MVP Runner-Up** recognition.

Technologies: Java/J2EE, Struts, JSP, AJAX, JavaScript, CSS

Certifications

- **Microsoft Azure Solutions Architect Expert** — Microsoft — AZ-305 / AZ-300
- **Microsoft Azure Developer Associate** — Microsoft — 70-532
- **TOGAF 9 Foundation** — The Open Group — Enterprise Architecture
- **DevOps Practitioner** — Simplilearn — CI/CD & Automation
- **Oracle Certified Associate** — Oracle — Java Programmer

Education

Sikkim Manipal University
MBA – IT Project Management

Shivaji University
Bachelor of Engineering – Civil Engineering

Awards & Recognition

- **Tieto:** You Dunnit Award, Spot Award, Pat on the Back Award, Highest Annual Rating.
- **Accenture:** CAPG (Second Highest among 200+ Senior Analysts), ACE Award.
- **Infosys:** CRR 1+ for two consecutive years, E&R Excellence Award, MVP Runner-Up.

Appendix — Selected Project Case Studies

Sameer Raghunath Deshpande | Senior Solutions Architect

AI-Powered Resume Scoring System

Client: eWork • **Role:** Senior Solutions Architect

Tech: *Azure AI Foundry • PostgreSQL / pgvector • Node.js • Azure Functions • Cosmos DB*

Context: eWork needed to move beyond keyword-based candidate filtering to intelligent, context-aware **resume-to-role matching** at enterprise scale — processing a large, growing resume corpus with low recruiter effort and high match quality.

Architecture Focus

- Designed a **hybrid retrieval architecture** using **Azure AI Foundry** to generate high-dimensional text embeddings from resume and job-description content.
- Stored and indexed vectors in **Azure PostgreSQL (pgvector)**; applied **cosine similarity** scoring for ranked candidate retrieval.
- Architected a **serverless processing pipeline** via **Azure Functions** for scalable, event-triggered ingestion and scoring.
- Defined retrieval and indexing strategy balancing **search quality** and response performance.
- Established an **evaluation approach** to validate match quality and inform embedding-model selection.

Outcome

- Reduced recruiter screening effort noticeably in early rollout by surfacing better-ranked shortlists automatically.
- Architecture is designed to **scale horizontally** as the resume corpus grows, without re-engineering the core retrieval layer.

Plussa Loyalty & E-Commerce Platform

Client: Kesko (Finland) • **Role:** Technical Architect → Lead Architect

Tech: *Azure Web Apps • React • Node.js • Terraform • PostgreSQL • Azure DevOps*

Context: Kesko required a modern, scalable **loyalty and e-commerce** platform to replace ageing systems, support a large active user base, and generate measurable commercial revenue — while maintaining high availability across Finnish retail operations.

Architecture Focus

- Served as **v1 architect** defining platform foundations: multi-tier Azure architecture, data models, and integration patterns with Kesko backend systems.
- Returned as **v2 core developer**, extending the platform with new loyalty mechanics, storefront features, and performance improvements.
- Designed a **modular React frontend** architecture enabling parallel feature development across squads without coupling.
- Standardised infrastructure delivery through **Terraform IaC**, enabling consistent provisioning across dev, staging, and production.
- Introduced **blue/green release** patterns and proactive monitoring to support zero-downtime deployments.

Outcome

- Platform supported a large active user base and contributed meaningfully to Kesko's digital revenue program.
- IaC and CI/CD foundations established here were later adopted as the **standard pattern** across additional Kesko programs.

Customer Feedback Platform

Client: Kesko • **Role:** Senior Solutions Architect

Tech: *Azure Front Door • Static Web Apps • Azure Functions • Azure AI Foundry • Cosmos DB • Terraform*

Context: Kesko needed a production-grade, feedback platform to handle sustained daily submission volumes across Finnish retail stores — with built-in moderation, SLA-based triage, and audit-ready workflows.

Architecture Focus

- Architected an end-to-end platform on Azure: **Front Door + Static Web Apps** for low-latency UX, **Functions** backend for resilient processing, **Cosmos DB** for flexible feedback storage.
- Integrated **Azure AI Foundry** for automated content classification — detecting abuse, bot patterns, spam, and priority routing — reducing manual moderation burden.
- Designed a **zero-trust** access model with strong tenancy boundaries between store, regional, and admin roles.
- Implemented multi-layer resiliency: **autoscaling, circuit breakers, proactive alerting**, and blue/green-style releases.
- Built a full **operations console** for moderators and admins with SLA tracking, triage queues, and audit logs.

Outcome

- Platform achieved strong **RTO/RPO** targets and earned stakeholder recognition for operational stability in the Finnish market.
- AI-assisted moderation measurably improved moderator throughput and reduced average time-to-action on flagged content.

Feedback Management — AI Modernization Platform

Client: Kesko • **Role:** Senior Solutions Architect

Tech: *Azure AI Foundry • Vector Embeddings • MCP Servers • Azure Functions • Terraform*

Context: Kesko's feedback was collected but triaged manually and in silos — slow response times, no sentiment insight, unguarded PII, and no link to loyalty, marketing, offer, or reporting systems. A phased, **responsible-AI** modernization was needed to turn feedback into a real-time competitive-intelligence asset without compromising **GDPR** obligations.

Architecture Focus

- Authored a **four-phase, incremental and reversible** roadmap — AI readiness, smart automation, RAG chat, MCP ecosystem — each phase delivering standalone value under strict cost governance.
- Designed **Phase 1** data foundations: normalized feedback schema, customer-to-opaque-token **PII mapping**, privacy vault, indexation strategy, and observability baseline before any model go-live.
- Architected **Phase 2** sentiment and auto-reply pipeline on Azure AI Foundry / Azure OpenAI with a **human-in-the-loop** approval gate, escalation routing, and per-feedback token-cost monitoring.
- Designed **Phase 3 privacy-safe RAG**: feedback vector store on Azure AI Search, **PII stripped before retrieval** and re-hydrated after generation, intent-based routing, and a retailer self-service chat UI.
- Defined **Phase 4 MCP ecosystem** giving governed AI agents secure, cross-system access — loyalty, marketing, offers, rewards, reporting — for agentic, customer-journey-level resolution.
- Embedded cross-cutting controls in every phase: **PII-vault architecture**, responsible & observable AI, **OWASP LLM Top 10**, and a retrieval-quality framework (MRR, precision, recall, NDCG, faithfulness).

Outcome

- Roadmap adopted Kesko's **modernization blueprint**, reframing feedback from a manual cost centre into a governed, AI-assisted competitive-intelligence capability.
- **Privacy-by-design** (personal data never reaches the model) and value-first cost governance positioned the program for GDPR-compliant, budget-conscious rollout.

SolutionHub — Vectorless RAG Documentation Platform

Client: Vivicta (Internal) • **Role:** Senior Solutions Architect

Tech: *PostgreSQL (ltree) • Sparse Vector Search • Node.js • Gemini API*

Context: Vivicta's internal engineering and product documentation had grown to roughly **3,500 files**, making keyword search unreliable. A Q&A layer was needed that scaled with the corpus without standing up and maintaining a dedicated vector database.

Architecture Focus

- Designed a **vectorless RAG architecture** in which the LLM navigates a category → document → section hierarchy through pure structural reasoning, rather than embedding similarity search.
- Modeled the full document tree in Postgres using **ltree** paths with a GiST index, enabling exact-section subtree fetches in a single indexed query with no recursive CTEs.

- Layered **sparse vector** search alongside tree navigation as a lexical-match signal for keyword-heavy queries, keeping the recall benefits of term-based search without a dense embedding index.
- Implemented **SHA-256 hash-based incremental ingestion**, generating and caching LLM section summaries only for new or changed files — unchanged documents cost zero LLM calls on re-sync.
- Kept all state — documents, tree structure, sparse vectors — in a **single Postgres instance**, eliminating a separate vector-database dependency.

Outcome

- Scaled cleanly to the full **~3,500-file corpus** with re-sync cost of zero LLM calls for unchanged content.
- Removed the operational overhead and cost of a dedicated vector database while delivering sub-second navigation to the exact relevant section.

K-PIM Product Information & Data Platform

Client: Kesko • **Role:** Lead Architect

Tech: *Stibo Step PIM • Hasura GraphQL • PostgreSQL • Redis • Azure Data Factory • Node.js*

Context: Kesko's product data was fragmented across legacy systems, creating inconsistency across markets and slowing e-commerce performance. A unified, **high-performance product data platform** was needed to serve multiple downstream applications across three regions.

Architecture Focus

- Designed an enterprise-grade **data pipeline** integrating **Stibo Step PIM** as the master source with **Hasura GraphQL** over **PostgreSQL**.
- Implemented **Redis-based caching** with TTL and invalidation strategies tuned per data type — significantly reducing API latency for storefront consumers.
- Orchestrated **Azure Data Factory** workflows for large-scale nightly and incremental product-data sync across **three regional markets**.
- Defined API contracts and **versioning strategy** enabling multiple e-commerce platforms to consume product data independently.
- Led a four-member delivery team; conducted architecture reviews and guided technical decisions across the program.

Outcome

- API latency reduced by **65%** through caching strategy, directly improving storefront responsiveness for end users.
- Pipeline established a **single source of truth** for product information, reducing inconsistency across Kesko's digital channels.

Digital Asset Management & CDN Optimisation

Client: Kesko • **Role:** Technical Architect

Tech: *Azure CDN • Azure Blob Storage • Azure Functions • Terraform • Node.js*

Context: Kesko's media assets were served through an unoptimized, costly delivery setup causing performance inconsistencies across four enterprise applications and multiple geographies. A rationalized, **cloud-native asset delivery** platform was needed to reduce cost and improve global performance.

Architecture Focus

- Architected a centralized **cloud-native asset delivery** platform on **Azure Blob Storage** with CDN-layer optimization across geographies.
- Designed automated **cache policies**, TTL strategies, and failover mechanisms reducing origin requests and improving cache-hit rates.
- Implemented an **asset processing pipeline** via Azure Functions for format conversion, resizing, and metadata tagging on ingest.
- Standardized asset delivery contracts across four consuming applications, eliminating per-app workarounds.
- Delivered infrastructure via **Terraform** for consistent, auditable CDN configuration across environments.

Outcome

- Infrastructure cost reduced by **~50% (~$50K annual savings)** through CDN consolidation and cache optimization.
- Bandwidth usage reduced by **~45%**; global content-delivery performance improved across all four enterprise applications.