DIVESH KUMAR CHORDIA

Bengaluru, Karnataka, India • (+91) 72046 06867 • diveshkumarchordia@gmail.com
linkedin.com/in/divesh-kumar-chordia-84471720b · github.com/Divesh-Kumar-Chordia · divesh-kumar-chordia.github.io/mario-portfolio

SUMMARY

Software engineer on Contentstack's data and analytics platform, building the systems every customer relies on to audit, trust and act on their data: the Go audit-logging platform that keeps billions of events searchable at low cost, and the pipelines and analytics API behind 2B+ monthly API calls, now refreshed about every 20 minutes instead of daily. Owns features from ideation to production, built the first observability for production data pipelines, and cleared 5,000+ long-standing CVEs across 10 services. Also builds production-grade AI systems: a fully local RAG and agent platform with evals, grounding and guardrails.

SKILLS

Languages: Go, TypeScript, JavaScript, Python, Java, C++, SQL
Backend & Design: NestJS, Node.js, Express, FastAPI, gRPC & Protobuf, REST/OpenAPI, Distributed Systems, Microservices, Multi-tenant Systems, OOP, Design Patterns (Factory), System/API Design
Data: Kafka, Apache Spark (ETL), Elasticsearch (ILM), MongoDB, NoSQL, AWS Athena (SQL), Redis, ChromaDB/Pinecone
AI/ML: RAG (hybrid retrieval, re-ranking, Corrective/Self-RAG), LangGraph, LangChain, Agents & Tool-calling, LLM Evals, Guardrails, AI Gateway (model routing, rate limiting), Ollama, vLLM, OpenAI/Anthropic APIs, MCP, Claude Skills
Frontend: React, Redux, Highcharts, Webpack, Vite, Storybook, Playwright
Platform: Docker, Kubernetes, CI/CD (GoCD, GitHub Actions), Observe (monitoring, alerting), AWS/GCP/Azure, Snyk, Trivy, Git

EXPERIENCE

Software Engineer, Contentstack — Aug 2024 - Present · Bengaluru, India (On-site)

- Engineered **distributed, multi-tenant audit logging** in Go, Kafka and Elasticsearch as Contentstack went multi-product, keeping **billions of logs** searchable at low cost.
- Onboarded **6 products and every new customer**, adding multi-topic Kafka support and a direct Elasticsearch query path for complex searches.
- Migrated audit search from **AWS OpenSearch to Elastic Cloud** for security and cross-cloud consistency, using transfer scripts, dual writes and a clean cutover.
- Built the **first data-pipeline observability**: Observe dashboards and **Slack alerts** showing management whether every production data job ran.
- **Primary point of contact** for the audit platform in production, leading root-cause analysis and verification after upgrades and security patches.
- Cut analytics **data freshness from daily to about 20 minutes** via a 4-hour stage, tuning **Spark** windows and partitions to lower compute cost.
- Engineered **Spark ETL pipelines** turning about 1B Cloudflare and Kafka records into the usage analytics customers see.
- Own the **Analytics API** serving **2B+ monthly calls** for all about 87 enterprise customers, onboarding new products with **zero contract changes**.
- Delivered **3 workflow analytics dashboards** exposing stuck content across **100K+ entries per org**, despite legacy data lacking reliable timestamps.
- Built a **Playwright Evaluator** covering **67 operation-sequence cases**, re-run across dev and production to sign off every deployment and release.
- Spearheaded **Feature Adoption** end to end, showing customers **paid vs. used features** and giving CSMs upsell signals; reused by 3 production apps.
- Extended webhooks with **75+ trigger combinations** for the Variants and Taxonomy launches, shipping with **zero breaking changes** to customer integrations.
- Scaled Top URLs exports **10x (1,000 to 10,000+ records)** using Athena SQL pagination and a pre-run row-limit check.

- Cleared **5,000+ year-old CVEs** across 10 services, adding **Snyk and Trivy** scans that block new vulnerabilities before merge.
- Authored **cs-highcharts**, the shared library behind every analytics dashboard, eliminating recurring duplicate axis labels with a log10 tick algorithm.
- Automated **release testing** for biweekly releases: daily runs generate realistic data and clear old data, replacing manual environment setup.
- Orchestrated **Agile release cycles** and daily standups for 1+ year, reviewing code across 35 repositories.
- Working **AI-first** (Cursor, Claude Skills, MCP), delivered an epic solo from **ideation to production in 2 sprints**, and to GA in two more.

Software Engineer Intern, Contentstack Jan 2024 - Jul 2024 · Bengaluru, India (On-site)

- Built the internal CRM's **first scrollable data-table module** (search, pagination, column sorting), later reused as-is across admin tools.
- Shipped **mail templates, logs and rollout UI**, plus a tech-event website on Contentstack CMS.

PROJECTS

LocalAI Forge: Private RAG & Agent Platform (Python, FastAPI, LangGraph, Ollama, Chroma) GitHub

- Raised retrieval **hit@1 from 0.884 to 0.977** using hybrid BM25 + dense search, RRF fusion and cross-encoder re-ranking.
- Lifted golden-set pass rate **from 85.1% to 95.7%** with an eval-gated prompt registry; LLM judge calibrated to humans at kappa 0.80.
- Achieved **0.972 faithfulness** with per-sentence grounding checks that flag unsupported claims before answers reach users.
- Hardened **LangGraph agents** with guard nodes, circuit breakers and human approval, resisting a planted prompt-injection attack.

JevStream (Python, FastAPI, Redis): A Redis priority queue that routes the most urgent support tickets first; atomic Lua claims with expiring leases recover tickets from crashed workers.

Analytics Automation Lab (Node.js, React): Generates realistic Contentstack activity (10,000+ entries and 50,000+ meter events per run, every 5 minutes in CI) so analytics dashboards can be validated against known data.

RAG Pipelines (LangChain, ChromaDB/Pinecone): A production-style RAG pipeline and a from-scratch retrieval lab (multi-query, RAG-Fusion, re-ranking, Corrective/Self-RAG) to reduce hallucinations.

EDUCATION

Bachelor of Engineering - Computer Science, JSS Science and Technology University (SJCE), Mysuru 2020 - 2024 · CGPA: 9.05

CERTIFICATIONS

- Claude 101, Anthropic Oct 2026
- Databricks Fundamentals Accreditation, Databricks Mar 2025 - Mar 2026
- Machine Learning Specialization, DeepLearning.AI & Stanford
- Google Data Analytics Professional Certificate · Google UX Design Professional Certificate

LANGUAGES

English (Fluent) · Hindi (Native) · Kannada (Intermediate)