Christopher Collins

PRINCIPAL ENGINEER | ENTERPRISE CONTENT, SEARCH & AI PLATFORMS | MULTI-TENANT SAAS AT SCALE

chris@hitorro.com | (650) 868-6129 | linkedin.com/in/chrisjcollins | github.com/geekychris

SF Bay Area

EXECUTIVE PROFILE

Principal-tier engineer with a three-decade arc across **enterprise content management, search & IR, and applied AI on top of customer content** — from the Documentum core-server organization and a petabyte-scale compliance / ILM indexing appliance at Kazeon, through shared semantic-search substrates at Intuit and SAP Ariba, a company-wide data-platform direction at SAP SuccessFactors, to high-blast-radius event / webhook infrastructure and AI code-assist evaluation at Intuit. Signature: turn fragmented or opinion-driven capabilities into measured, reusable platforms — using migration discipline, evidence in place of preference, and executive-through-support influence at the 200+-engineer scale expected of a Principal Engineer.

CORE COMPETENCIES

- **Enterprise Content Platforms:** Documentum core-server engineering; typed content models & metadata; content transformation, rendition & viewer pipelines; federated administration of clustered content servers; regulated content (life-sciences GxP, financial ecosystems, privacy-remediated ad measurement).
- **Multi-Tenant SaaS & Search:** Semantic Discovery over typed catalogs (tens of thousands of tenants, millions of items each); Scout Labs analytics on a 7B-record rolling corpus; shared search-as-a-service at ~30M queries/tax day; Lucene / Solr / inverted-index craft; semantic + lexical + vector search; RAG; relevance measurement (NDCG / MRR / MAP).
- **Applied AI:** LLM integration; MCP tool surfaces; agentic + human-in-the-loop workflows; NLP enrichment (NER, sentiment, classification); production ML runtimes with offline + online evaluation.
- **Data & Event Platforms:** Kafka, Flink, Spark, Iceberg, S3; durable ordered store-and-forward webhook delivery; high-cardinality DynamoDB queues; structured-logging paved roads; CQRS with KV / search / graph serving tiers.
- **Migration Discipline:** Parallel run, duplicate feeds, shadow / parity validation, per-subscriber cutover + rollback, sender-edge compatibility shims — migration as a platform capability, not a deployment afterthought.
- **Languages & Runtimes:** Java (deep), Python, Go, TypeScript, Rust, C/C++, Swift, Kotlin. Adopts a team's stack.
- **Leadership at Scale:** Multi-year technical direction; executive persuasion without formal authority; security / vendor / procurement navigation; brown-bags, Innovation Days, durable mentorship.

SELECTED IMPACT

- **~200M events/day, ~99.99% uptime, sub-second in-order delivery** — Intuit External Event Gateway; ~10K prod + ~12K sandbox third-party subscribers migrated with no major migration incident; ~half legacy cost; supports ecosystem continuity for 12M+ QuickBooks Online small businesses.
- **Company-direction-changing data-platform strategy (SAP SuccessFactors)** — reversed a "no data plan" posture; sponsorship through the President of SuccessFactors; direction still followed today.
- **~4× search capacity, ~$80M avoided expansion (SAP Ariba — recollected)** — SSD modernization proved via a bounded, inexpensive experiment.
- **CEO Innovation Award (Intuit)** — shared semantic-search substrate across TurboTax, QuickBooks, Mint; ~30M queries on a single tax day; grew the search org 1 → 18.
- **AI code-review platform rollout** — CodeRabbit + Cursor Bugbot to 8,000+ engineers across Intuit + Credit Karma in ~2

weeks; ~30–40% reported lift in acceptance / time-to-merge; tens of millions/year projected savings.

PROFESSIONAL EXPERIENCE

Intuit | Staff Engineer / Technical Lead — External Event Gateway, then Code Assist *(May 2022 – Present)*

External Event Gateway (May 2022 – Nov 2025)

- Replaced a ~9-year-old webhook infrastructure that could lose events, deliver out of order, offer no replay, and take up to several minutes — a system underpinning third-party integrations that support business continuity for **12M+ QuickBooks Online small-business accounts**. High blast radius, effectively zero tolerance for data loss.
- Architected four horizontally scalable multi-AZ services on Kafka
  DynamoDB delivering durable ordered store-and-forward, replay, and HA; modeled the high-cardinality set of mostly-sparse ordered subscriber queues on DynamoDB composite keys. Turned Kafka-vs-Pulsar peer disagreement into evidence via a Pulsar POC and retained DynamoDB on cost + missing-capability grounds.
- **Made migration a platform feature.** Cut over ~10K prod + ~12K sandbox subscribers (most uncontactable) via duplicate feeds, shadow / parity validation, per-subscriber dial-up + rollback, sender-edge compatibility shims when internal deprecations slipped — **no major migration incident, no data loss reported**.
- Outcomes: sub-second delivery (from up to 6 min), ordering + replay, ~half legacy cost, upstream sources ~8 → mid-20s, ~200M events/day / ~10M/day to third parties, ~99.99% uptime. Also shipped a Flink / Spark / S3 event-metrics platform and a RAG-based API discovery POC that seeded Intuit's internal developer API-discovery offering.
- Coached a college-level intern/co-op through a multi-service stretch project sized for learning with mentorship; the mentorship relationship continued through interview prep and career discussions afterward.

Code Assist (Nov 2025 – Present)

- Owns **Code Inbox** in Intuit's broadly installed Mac developer desktop app — a unified triage surface for created, assigned, watched, and review queues by criticality, age, and size, directed at human oversight of end-to-end cloud-agent workflows.
- Built a reusable **AI code-review evaluation harness** over golden repos and replayable commit sequences; measured relevance, coverage, severity alignment, precision, recall, noise. Extended a **PR analytics platform** across hundreds of repos to attribute AI-review impact against a defensible baseline. Prototyped **CodeCraft**, injecting Intuit code-guidelines and internal API patterns into agentic loops.
- Led the preceding evidence-based vendor selection (CodeRabbit + Cursor Bugbot) that displaced a favored high-profile vendor — coordinating security, procurement, international enablement, vendors, and alpha testers; ~30–40% reported lift in acceptance / time-to-merge; ~2-week rollout to 8,000+ engineers across Intuit + Credit Karma; tens of millions/year projected savings via negotiated usage-cost pricing.

SAP SuccessFactors | Distinguished Engineer & Chief Architect, Data + Search *(Dec 2019 – May 2022)*

- Recruited by the SuccessFactors CTO (formerly my Chief Architect at Intuit) to reverse a "no data plan" posture: proposed and won sponsorship for a company-wide **Data Platform Direction** — the primary body of work for the tenure — designed as incremental, not big-bang, transformation.
- Framed the platform as a six-stage maturity path so value could be delivered before every layer existed: structured capture & instrumentation via a "paved road" into Kafka-style pub/sub; analytics + insight; cohort / A-B / hypothesis testing; reusable batch + stream processing; model experimentation, training & serving; CQRS precomputation into fit-for-purpose serving tiers (KV, search, graph). Distinguished lightweight frontier-model AI (adoptable early) from heavier custom-training work — AI didn't gate on platform completion.
- Sold the vision organization-wide to the CTO, Chief Architect, President of SuccessFactors, and peer architects using business-specific POCs, executive pitch decks, and an internal architecture roadshow that taught vocabulary alongside direction. Secured funding, seeded the platform team (~8), and left a roadmap that continues today after multiple technology swaps.
- **Semantic Discovery Platform** (secondary): forked SAP Ariba's semantic-search codebase; stripped Ariba-specific runtime

wrapping and unfamiliar stream-processing dependency; reached production polish for résumé and job-description search but did not launch before departure — honest separation of technical readiness from launch. Held a principled security dissent on "search indexes are inherently secure" and advanced an encrypted-index position.

SAP Labs | Software Architect — Conversational AI *(Sep 2018 – Dec 2019)*

- Architecture owner for question routing and pre-LLM curated-knowledge answering inside SAP's Conversational AI bot platform — distinguishing informational questions from commands, matching to curated FAQ knowledge.
- Established the evaluation rigor needed to move from prototype to production: real customer / partner FAQs, held-out queries, translation/back-translation paraphrases, out-of-domain probes, embedding-distance retrieval, MRR / MAP / NDCG-family measures, and A/B diagnostic instrumentation. Subsystem shipped.

Facebook | Software Engineer — Ads Identity & Measurement *(Oct 2017 – Jul 2018)*

- Owned and stabilized a sprawling cross-company data pipeline in an identity-prediction org after experienced engineers were reassigned; ran with limited access to prior experts at extraordinary scale.
- Executed **privacy-driven decommissioning and constraint work** during the Cambridge Analytica period — retiring or constraining data uses and sources under executive / legal urgency, even where it reduced system efficacy. Treated legal and societal constraints as first-order design requirements. Specifics remain sanitized.

SAP Ariba | Software Architect — Search Capacity + Semantic Discovery *(Mar 2016 – Oct 2017)*

- **SSD Modernization (~4× capacity, ~$80M avoided expansion — recollected):** data-center floor space had run out and continued procurement-catalog growth required a different approach. Proposed SSDs against skepticism, borrowed representative ops hardware plus a small SSD investment, built a performance test bench and new runtime configurations, and produced decision-grade evidence that overturned the debate. Canonical example of turning strategic infrastructure disagreement into a bounded, inexpensive experiment.
- **Semantic Discovery** (distinct chapter): architected the multi-tenant enterprise-procurement search platform over typed customer catalogs — millions of line items per customer, tens of thousands of tenants, rapid updates, NLP enrichment, configurable ranking, relevance measurement, and recommendations. Reached production after departure — later-reported context, not a personal launch claim.
- Personally seeded and sponsored a hypothesis-driven **Innovation Days** week for ~a couple hundred engineers (imported from Intuit's practice, framed by *The Lean Startup*) — engineering NPS showed at least a temporary lift and some experiments influenced teams' day jobs.

Intuit | Search Architect — Semantic Technologies & Search *(May 2012 – Mar 2016)*

- Recruited by Intuit's CTO to establish a **company-wide shared search capability** across TurboTax, QuickBooks, Mint, and others, in place of duplicated engineer-built search solutions with uneven relevance and weak measurement. Named and built "Semantic Technologies and Search" as a common substrate covering the complete search experience — not indexing alone.
- Designed a typed semantic document model driving repeatable multilingual NLP enrichment (segmentation, NER, classification), with the same semantic concepts informing both indexing and query processing. A query-coordination layer applied configurable recipes over query text + user context, retrieved candidates, and post-processed with enriched data from fast stores.
- Built the **relevance-improvement loop**: authored ranking recipes, evaluated offline against labeled judgments (NDCG-style), promoted promising changes into controlled online experiments. Established the platform / product split — platform owns reusable capabilities; product teams own tuning and product-specific decisions.
- Scale: ~30M queries on a single TurboTax tax day; grew the search org **1 → 18** over three years — inbound hiring only, personally trained every hire outside the two data scientists. Earned Intuit's **CEO Innovation Award**. Established regular brown-bag share-outs so engineers taught one another and junior engineers presented as peers.

Kazeon Systems | Senior MTS — Compliance / ILM Indexing Appliance *(2004 – Aug 2005)*

- **Initial sole architect-implementer** of the search platform for a clustered compliance / information-lifecycle-management appliance — sat beside enterprise storage, performed shallow or deep indexing, and made results actionable for locking, moving, shredding, or routing content in regulatory and legal workflows.
- Target requirement was **petabyte scale**; re-index throughput on fixed commodity hardware mattered more than interactive query latency. Partitioned content into small chunks and indexes, pipelined stages, saturated every node and thread, hid latency, profiled each emerging bottleneck alongside a hardware performance engineer. Reportedly led a public indexing benchmark that year (identity requires verification).

Documentum | Software Engineer — Core Server *(1996 – 1999)*

- Joined as a developer-level specialist alongside the traditional support organization (early forward-deployed engineering); moved into the **core server** organization and was given high-profile customer problems directly by the CTO / co-founder.
- **AutoRender turnaround**: rebuilt a fragile, support-heavy content-transformation product (Office → text + PDF for search highlighting, viewing, annotation) to run unattended, recover from hung Office apps, support multiple repositories, and use an extensible plug-in model. Pricing reportedly moved from a few thousand to ~$10K per copy — became one of the most profitable SKUs relative to engineering input for a period (recollected).
- Architected a **federated administration suite** for clustered Documentum server instances, replacing the increasingly inadequate CLI model.

Sandoz / Novartis | Lead Engineer — Pre-Clinical Document Platform *(1993 – Dec 1995)*

- Reconstructed requirements from a fragile Uniface / Oracle / DEC-script system used for strictly regulated pre-clinical reports across the drug-application lifecycle; proposed and built a **Documentum-based replacement**, involving the strongest user critics in iterative testing. Reportedly used by thousands; won the merger-era technical comparison to become **Novartis's pre-clinical document-management platform** after the Sandoz / Ciba-Geigy merger. Origin of Chris's Documentum customer perspective.

EARLIER CAREER HIGHLIGHTS

- **Scout Labs / Lithium** | Chief Architect *(Jan 2008 – Nov 2011)* — End-to-end architect of a rebuilt near-real-time multilingual social-media analytics SaaS (Disney, Coca-Cola, US-government customers). Partition-first pipeline over persistent logs / HDFS with hot/cold search tiers; rolling 6-month corpus reached ~7B records on ~20 search nodes. Introduced organization-scoped human overrides for ML sentiment — early principle for keeping ML inference trustable at multi-tenant scale. Acquired by Lithium.
- **Turn** | Senior Technical Leader / MTS *(Aug 2005 – Aug 2006)* — Original relevance team; built specialized text-feature indexing for staged ML ad ranking under a tight latency budget. Defining contribution: instrumented, production-safe execution paths for data-science experiments — a lasting belief that algorithmic systems need built-in measurement, tuning, traceability, and safe experimentation.
- **Ariba** | Senior Software Engineer — Catalog / Search *(Mar 1999 – Mar 2004)* — Independently built shared standards-data infrastructure (UNSPSC, ANSI X12, ISO 4217, hazmat, locations, languages) with mappings between competing standards, used across products. Extended the type system, search, and UI for configurable products/services; delivered end-to-end catalog-image support (spec → ingestion → validation → resize / thumbnails → CDN delivery → image-aware search UI).

ADVISORY, COMMUNITY & EDUCATION

Advisory: Aira.io — applied AI for assistive tech serving the blind and low-vision community; facilitated NIH / NIST grant work. Personal.AI — personal AI models, search, and data architecture. Additional early-stage startups in search, data, and applied ML.

Community, Standards & Education: Represented Ariba in ECCMA / ISO working groups on classification and product-data standards. Primary organizer / hardware fabricator for the Amiga 30th & 40th anniversary events at the Computer History Museum. Conference talks on "Agentic development with the Commodore Amiga." **BSc (Hons) Computing**, University of Staffordshire, UK (1989–1993) — microprocessors and distributed systems.