TEJAS PATTABHI

Staff Engineer — Cloud · Distributed Systems · Applied AI

San Diego, CA · tejas.pattabhi@gmail.com · +1 (469) 570-1988 · tejaspattabhi.com · linkedin.com/in/tejas-pattabhi

SUMMARY

Staff Engineer with 10+ years turning large-scale data and compute problems into reliable production systems. At Qualcomm, works on the critical path of the autonomous-driving program — the infrastructure and applied AI that validate the software running in production vehicles. Named co-inventor on two granted U.S. patents, with deep expertise in Kubernetes-scale orchestration, multi-cloud architecture, and building AI agents that operate mission-critical infrastructure.

CORE SKILLS

Languages	Python · Java · C# · C · Bash	**Cloud**	AWS · Google Cloud (Dataflow) · Microsoft Azure
Orchestration	Kubernetes · Argo Workflows · Argo Events · Docker · Helm	**Applied AI**	LLM agents (Claude Agent SDK · AWS Bedrock) · agentic coding (Claude Code · Codex) · custom Claude skills & plugins · MCP tool integration
ML & vectors	Self-hosted embedding models (Qwen3, ONNX int8) · pgvector cosine similarity · semantic deduplication	**IaC & Ops**	AWS CDK · Prometheus · Grafana · CI/CD · IPMI/BMC hardware mgmt
Architecture	Distributed & event-driven systems · petabyte-scale data pipelines	**AWS services**	ECS Fargate · SQS · EventBridge · RDS/PostgreSQL · S3 · AppSync/GraphQL · IAM
Databases	PostgreSQL/pgvector · MySQL · MS-SQL		

EXPERIENCE

Qualcomm

Sep 2021 – Present

Staff Engineer (previously Senior Engineer) · San Diego, CA

On the critical path of Qualcomm's ADAS/autonomous-driving program: build the large-scale infrastructure that replays recorded real-world drive data at massive scale to validate automated-driving software — perception, radar, vehicle self-motion/positioning, ground-truth labeling, and automatic emergency braking (AEB).

- **Automotive-grade reprocessing clusters.** Designed, built, and operate two independent, massive on-prem Kubernetes clusters of Qualcomm **QCR-100** cards — the same Snapdragon Ride silicon that ships in vehicles — dedicated to bit-accurate ADAS/AD reprocessing, so validation runs on the exact silicon the software executes on in the car rather than an approximation from generic hardware. Purpose-built silicon runs this reprocessing **~100–300× faster than general-purpose compute**, fast enough to replay recorded drives at close to real time and meet customer-delivery timelines.
- **Reprocessing-as-a-Service platform.** Architected a pipeline that runs identically in the cloud or on-prem: job orchestration that selects input files (AppSync/GraphQL + IAM role assumption), resolves inter-step dependencies into a DAG, and submits event-driven **Argo Workflows** (Argo Events); plus parallel data transfer, input population, cleanup, and retryable/infra/application failure classification. Also assisted in integrating the QCR-100 into custom hosted cloud environments so partners and customers can reach this specialized compute on demand.
- **Key collaborator, Qualcomm–BMW Group program.** Helped deliver the reprocessing pipeline behind the jointly developed Snapdragon Ride Pilot automated-driving system (publicly launched Sept. 2025, debuting in the BMW iX3), turning recorded driving data into the validated results behind the platform.
- **Production AI pipeline-triage agent.** As project lead, built and run an agent that autonomously triages tens of thousands of job failures per day. A deduplication service embeds each failure (self-hosted embedding model) and uses **pgvector cosine-similarity** search to suppress recurring failures, so only novel ones reach a **Claude Agent SDK + AWS Bedrock** agent that gathers diagnostics, classifies root cause, and takes autonomous action — files Jira tickets, proposes GitHub config-fix PRs — escalating ambiguous cases to humans. Event-driven, cross-account AWS (ECS Fargate, SQS FIFO, EventBridge Pipes, RDS/pgvector, Managed Grafana); IaC via AWS CDK with full CI/CD and observability.

- **Fleet-health & accuracy monitoring.** Built a Prometheus + Flask monitoring stack — a per-node DaemonSet that discovers each accelerator's health and firmware every 30 seconds, and tracks **Real-Time Factor (RTF)**, the measure of how closely reprocessing output matches ground truth — with a live dashboard giving operators available compute and processing accuracy at a glance.
- **Low-level hardware-maintenance tooling.** Built the cluster maintenance layer — IPMI/BMC server power control and firmware-flash (EDL) mode control, plus safe Kubernetes node isolate/integrate (drain-cordon / uncordon) — enabling hardware servicing without disrupting running jobs.
- **Massively parallel cloud reprocessing (GCP Dataflow).** Re-architected the cloud reprocessing pipeline onto GCP Dataflow to run at massive parallelism, redesigning how work is partitioned and scheduled to raise concurrency and cut end-to-end turnaround while improving output accuracy.
- **Engineering leadership.** Lead and coordinate engineering across teams spanning three time zones, reviewing code and setting technical direction daily to keep distributed teams aligned on a continuous-delivery cadence.

OpenText (formerly Webroot)

Sep 2015 – Sep 2021

Senior Software Engineer (previously System Software Engineer, Software Engineer) · San Diego, CA

Led internet-scale cybersecurity and threat-intelligence systems.

- **Web crawling, classification & threat detection at internet scale.** Led distributed crawling clusters that safely crawled and classified threats across a **peak of 17M+ URLs/day** in real time to protect end users. Co-invented "DeepCrawl," which pivots from a single known malicious link to uncover related malicious infrastructure and grades its threat level with a machine-learning model — **the basis of two granted U.S. patents**.
- **IP threat analysis.** Subject-matter expert and lead; built a system that categorized malicious hosts by activity (phishing, botnet, malware, proxy, spam) to publish threat lists. Owned AWS hosting/DevOps.
- **Cloud security intelligence & reputation.** Designed a reputation-scoring system for cloud applications and their hosting organizations, gathering data vectors dynamically to surface security posture to customers.
- **Multi-cloud honeypots.** Architected honeypots across AWS, GCP, and Azure that captured live attacks (scanners, malware, injections, DoS) to continuously feed the threat-analysis systems.

Earlier Experience

- **Award Solutions** — Student Volunteer, OpenStack (2015). Built a webhook-driven app monitoring VM state transitions in Ceilometer; project work published open-source by UT Dallas.
- **TradeStation Technologies** — Software Developer Intern (2014). Built an AWS EMR application analyzing daily IIS logs; visualized request/response volumes with D3.
- **Société Générale** — Windows Server Administrator, Bangalore (2012–2013). Automated production-server health checks across ESXi hosts in two data centers.

PATENTS & PUBLICATIONS

U.S. Patent 12,452,259 B2 (Oct 2025) & **U.S. Patent 11,201,875 B2** (Dec 2021) — "*Web Threat Investigation Using Advanced Web Crawling,*" named co-inventor, assignee Open Text Inc.

Publication (first author): "Implementing Delaunay Triangles and Bezier Curves to Identify Suitable Business Locations in the Presence of Obstacles," *Int'l Journal of Information Technology & Computer Science (IJITCS)*, vol. 5 no. 3, 2013 — a data-mining and spatial-clustering algorithm.

EDUCATION

M.S. Computer Science — The University of Texas at Dallas · Conferred August 2015.

B.E. Information Science — PES School of Engineering (Visvesvaraya Technological University), India · 2012 · First Class with Distinction in all eight semesters; medal for highest first-year academic performance.