AI-POWERED SERVICE DESK

Three Acquisitions, Three Ticketing Systems, and Nobody Had the Full Picture

A post-acquisition IT services firm had tickets bouncing between teams with zero shared context. KeyDelta built an AI platform that pre-solves tickets before a human touches them — cutting resolution time 62% and escalations 45%.

-62%	-45%	78%	-38%	6 wks
RESOLUTION TIME	ESCALATIONS	AUTO-HYDRATION	AI COST PER TICKET	TIME TO DEPLOY

THE SITUATION

A multi-location IT services company — formed through three acquisitions in 18 months — was drowning in service tickets. Each legacy entity had its own ticketing system, its own triage process, and its own tribal knowledge. Reps were spending more time figuring out where a ticket should go than actually solving it. The result: slow resolution, frustrated customers, and a support team running on heroics instead of systems.

- Three acquired teams with three ticketing systems, three triage processes, and zero shared knowledge
- Reps spending 40%+ of their time on ticket routing and context gathering, not problem-solving
- Escalation rates climbing — tickets bounced between teams because nobody had full context
- No model flexibility — locked into a single AI vendor with rising costs and inconsistent quality
- ServiceNow and Jira running in parallel with no integration — duplicate work everywhere

THE APPROACH

KeyDelta designed and built an AI agent management platform that intercepted every incoming ticket, assembled context from multiple systems, and proposed solutions before routing to the right human:

1. **Platform Architecture**
   Built on Python Django with LiteLLM proxy for model abstraction. This let the platform hot-swap between GPT-4, Claude, and other models without code changes — optimizing for cost, speed, or quality depending on ticket complexity.
2. **System Integration Layer**
   MCP (Model Context Protocol) integration connected the agent swarm to internal knowledge bases, customer records, and configuration data. ServiceNow and Jira integrations unified the fragmented ticketing landscape into a single workflow.
3. **Agent Swarm & Hydration Engine**
   When a ticket arrived, specialized agents executed in a parallel fan-out pattern: context agent, knowledge agent, and similarity agent ran concurrently, then a resolver agent synthesized their outputs into a proposed solution. A confidence score determined auto-route vs. human review. The hydrated ticket landed in the rep's queue ready to close.
4. **Scaling & Optimization**
   AWS Lambda functions handled compute-intensive analysis tasks serverlessly. Continuous learning from every resolution improved hydration accuracy over time — the system got measurably smarter each month.

THE RESULTS — 6 MONTHS

RESOLUTION TIME	ESCALATION RATE	HYDRATION RATE	AI COST/TICKET	TECH SATISFACTION	CUSTOMER SAT
4.2 hrs → 1.6 hrs	34% → 19%	0% → 78%	$2.50 → $1.50	3.0 → 4.4 / 5	3.2 → 4.1 / 5
62% faster, pre-hydrated	Right team, first time	Tickets arrive pre-solved	Model switching optimization	Solving problems, not routing them	Faster + more accurate

WHY IT WORKED — THE KEYDELTA VOOCS LENS

V — VISION
Every ticket arrives at a human already diagnosed, contextualized, and half-solved — reps close issues, not chase them.

O — OUTCOMES
Resolution time, escalation rate, and hydration accuracy measured from day one. Model costs tracked per ticket to prove ROI in real time.

O — OWNERSHIP
Service delivery owned hydration quality. Each agent in the swarm had a defined scope — no overlap, no gaps, no committees.

C — CADENCE
Daily ticket quality reviews + weekly model performance analysis. Bad hydrations caught in hours, not weeks.

S — SCALE
New agents deploy in days, not months. Model abstraction means no vendor lock-in. The platform survived an acquisition without rearchitecting.

> *"We turned our service desk from a routing nightmare into an AI-powered resolution engine. Tickets that used to bounce between three teams now arrive pre-solved. Customer satisfaction jumped from 3.2 to 4.1. Our techs went from frustrated investigators to confident closers — satisfaction scores doubled. And AI cost per ticket dropped 38% because we could swap models based on complexity instead of paying premium prices for password resets."*
>
> — KEYDELTA ADVISORY