

AVIATION SUPPLY CHAIN

On-premise, high-availability AI platform design for an aviation network

A regulated aviation network needed AI agents that run entirely inside its own data center. Wexa delivered the full technical architecture and on-prem deployment design, from microservices to GPU nodes.

100%

on-premise: data, models and audit
inside the boundary

2

H100 GPU nodes in the production
HA design

3 + 5

control-plane and worker nodes,
load-balanced

33 pp

architecture documentation (arc42
style)

CLIENT

**B2B procurement network connecting
airlines and suppliers**

REGION

US / global

WHEN

2025

ENGAGEMENT

Solution design

WEXA COMPONENTS

**Full Wexa platform • on-prem Kubernetes
• GPU inference • 200+ connectors**

The challenge

- Airline and supplier data cannot leave the network
- Agents must integrate with CRM, email, messaging and document stores
- Production needs high availability, observability and a test strategy

What Wexa built

- ✓ Microservices: web, identity, data, workflow and schedule workers, event service, knowledge base, LLM gateway
- ✓ Durable agent execution on Temporal; MongoDB, Qdrant vector store, Redis, object storage and vault-managed secrets
- ✓ Kubernetes with Helm, NGINX ingress, autoscaling; Prometheus, Grafana, Alertmanager and Loki
- ✓ Local GPU inference alongside hosted models through one gateway
- ✓ Production HA: HAProxy and Keepalived virtual IP, three control-plane nodes, five workers, shared SAN/NAS

How it runs on Wexa

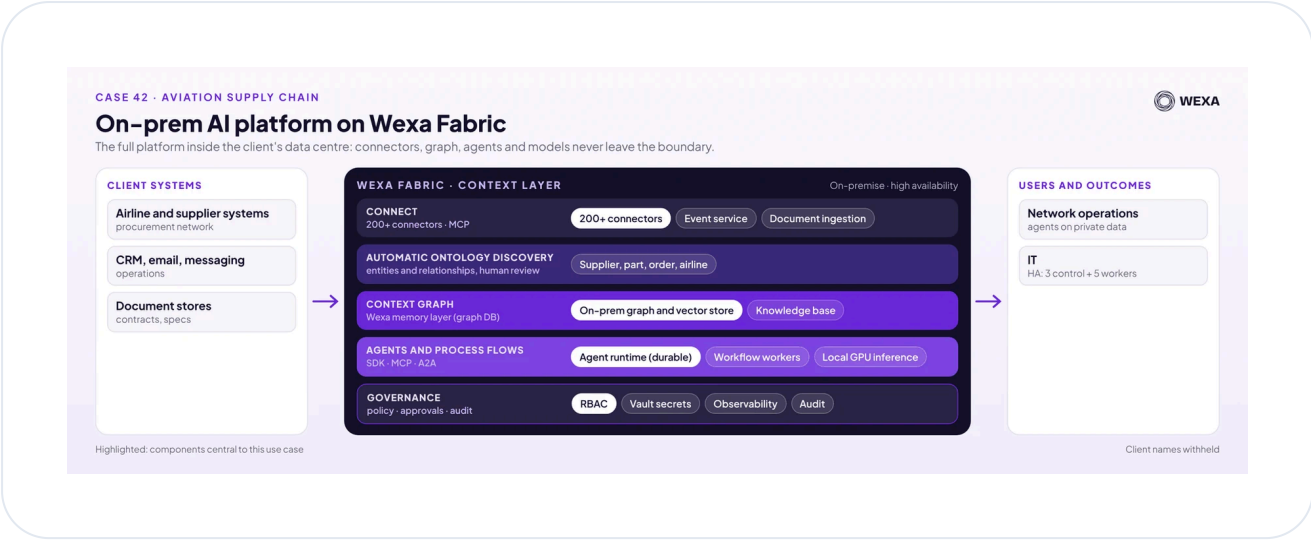


FIGURE How this use case runs on Wexa: client systems connected, ontology discovered, context graph in the Wexa memory layer, agents and process flows, and governance. Highlighted components are central to this use case.

On-prem HA deployment

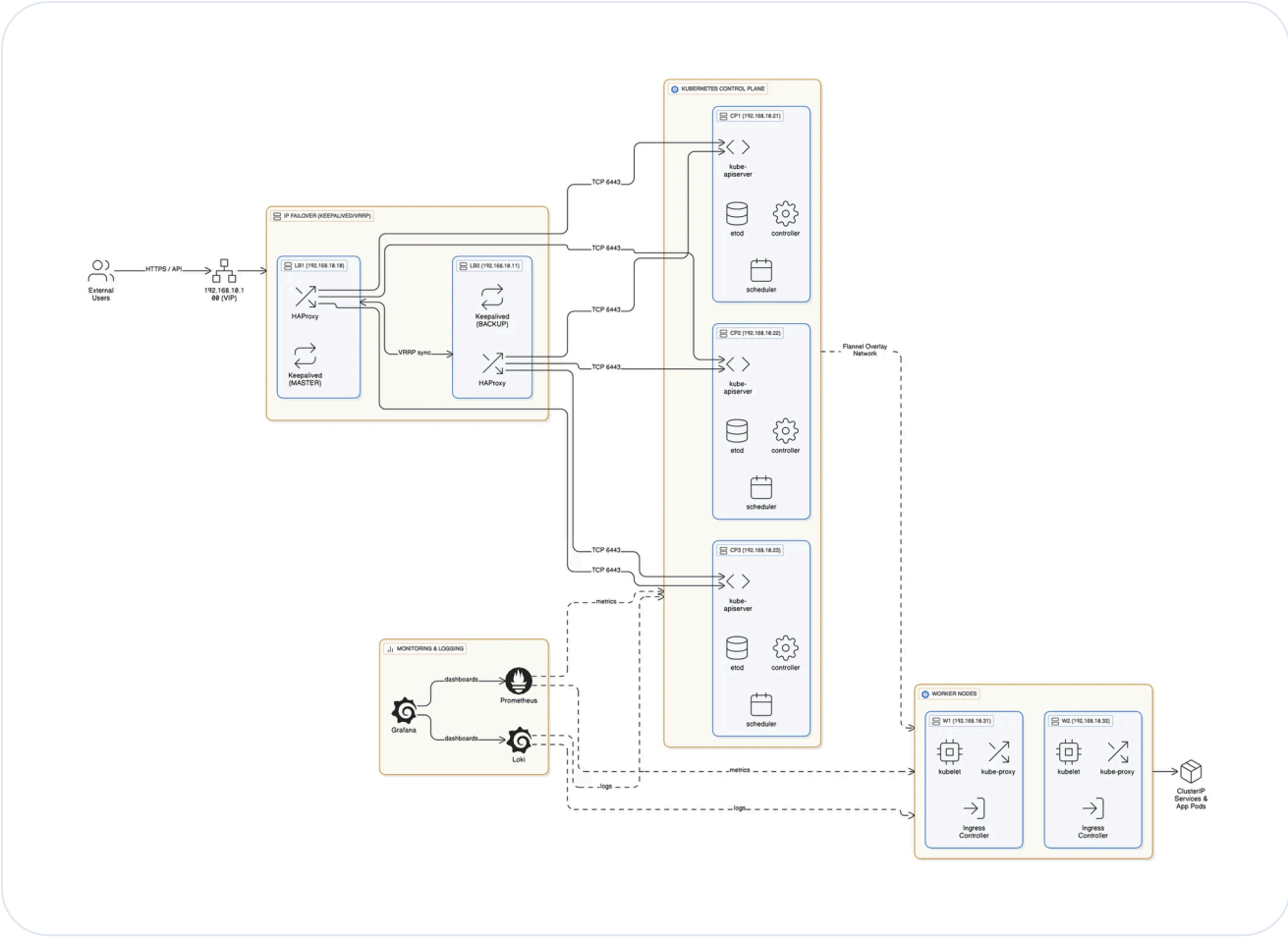


FIGURE Production on-prem high-availability deployment topology.

Results

MEASURE	BEFORE / WITHOUT WEXA	WITH WEXA
Procurement-support staff across the network	40 staff	24 staff
Annual cost of the process	\$2.8M	\$1.7M

MEASURE	BEFORE / WITHOUT WEXA	WITH WEXA
Time to implement the AI solution	9 months	12 weeks
AI build cost (one-time)	\$1.2M	\$133K
AI operating cost per year (models, infrastructure, upkeep, platform)	\$704K	\$336K
Total AI cost in year 1	\$1.9M	\$469K

WITH WEXA

\$784K

net annual value to the client after Wexa's own operating cost

3x

faster to implement

52%

lower AI operating cost

\$1.4M

less AI spend in year 1