intuit turbotax creditkarma quickbooks mailchimp

CLOUD NATIVE AI + INFERENCE DAY NORTH AMERICA

Ship Fearlessly

Canary Deployments with an AI Quality Gate for Control Plane Releases



Kartikeya Pharasi



Vinay Gonuguntla



Agenda

Admiral

The Problem

Agent Solution

What's Next

Technology @ Intuit



Intuit is leading the way in building an AI-native development platform using cloud native open source technology. We're committed to building tools that scale and giving back to the open source community.

Runs on Kubernetes

turbotax creditkarma quickbooks mailchimp

$3T+

Invoices, bills, payroll

307 Industries

Construction, healthcare, real estate, field services, professional services, etc.

74 Financial domains

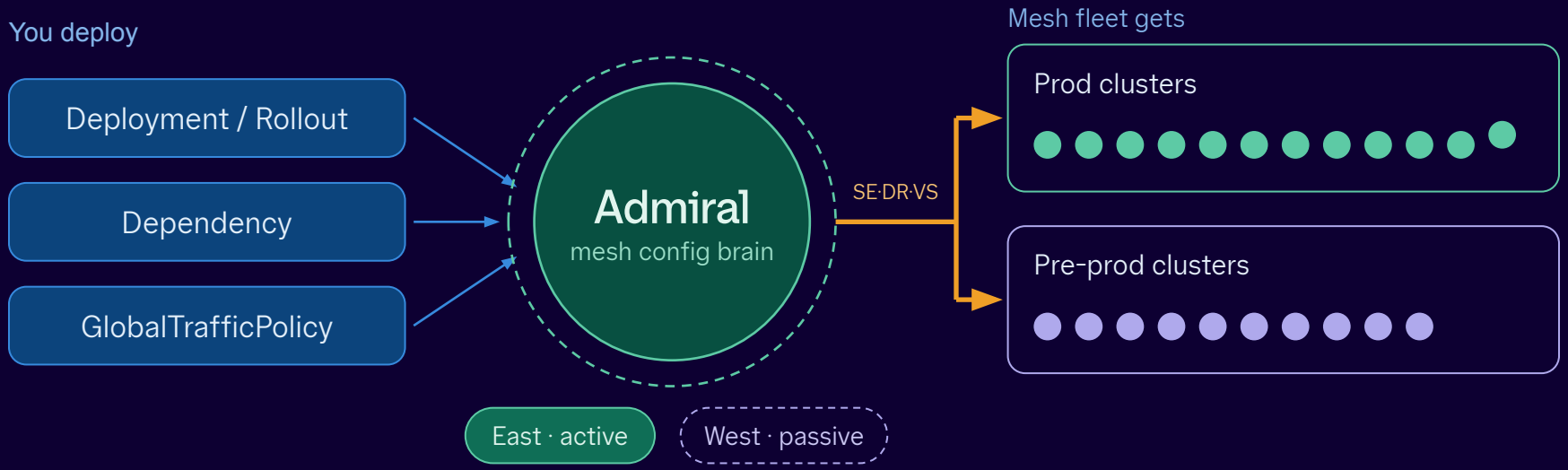
Payments, payroll, bill pay, accounting, taxes, etc.

Admiral



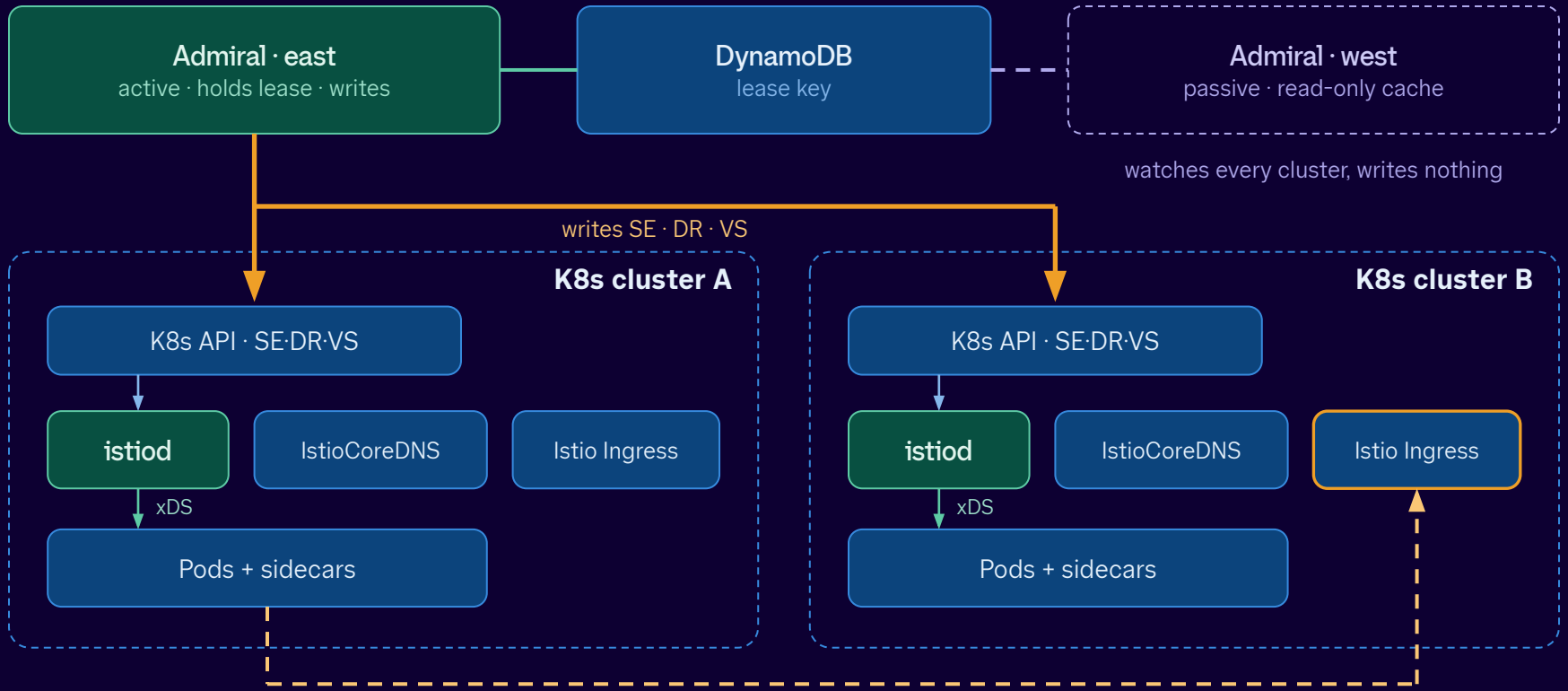
Admiral provides automatic configuration and service discovery for multicluster Istio service mesh.

High Level Overview



One writer, two regions, an istiod in every cluster

CLOUD NATIVE AI + INFERENCE DAY NORTH AMERICA



Scale of Admiral



350+
Prod and Pre-Prod clusters

20k+
Deployments/Rollouts

556k
Peak TPS

1 release
rewires every cross-cluster call

The problem

We find bad releases after the damage is done



Humans are the monitor

Failures surface late

Blast radius is huge

Rollback is a judgment call

Every gap is a job for an agent

Phase 1

We added an LLM to every stage

01
Monitor
canary vs stable at every stage

→

02
Detect
errors, latency, leaks, config drift

→

03
Diagnose
reads the config diff, explains the why

→

04
Decide
a rule, not a judgment call

Same signals an on-call engineer watches, read at every stage.

Phase 2

Roll out features, not releases

one cluster or one service at a time



1 service
in 1 cluster

→

10 services
same cluster

→

1 cluster
all services

→

1 region
pre-prod, then prod

→

Fleet
350+ clusters

Agent gate between every step · blast radius starts at one service

Agent watches each target against its own baseline

- **CPU** — spikes, throttling
- **Memory** — leaks, OOMKills
- **Logs** — new error patterns
- **Service behavior** — latency, traffic shift
- **Status codes** — 5xx, 503 UH / NR
- **Connectivity** — cross-cluster, DNS
- **Sidecar proxy** — xDS sync, NACKs, restarts
- **Your signal** — pluggable checks

Expand to the next step

Hold and extend the soak

intuit

How the agent works

Scores every release and acts on the score



intuit

DEMO



CLOUD NATIVE AI +
INFERENCE DAY
NORTH AMERICA

What's next

Catch it earlier, see more, gate the whole mesh



01

Integration test automation

The agent opens PRs that add integration tests for the runtime gaps it finds.

02

More signals

Traces, Istio telemetry, native histograms and synthetic cross-cluster probes.

03

Whole control plane

istiod, Aviso and the rest of the mesh behind the same gate, as shared A2A agents.

Each release leaves the mesh safer than the last

Stay in the loop

FOLLOW

Intuit Open Source

Don’t miss on exciting OSS events, activities & news

Scan or visit
bit.ly/intuit-oss



Visit our Intuit booth #157 in the Expo Hall to learn more about Admiral and other Intuit Open Source projects!

intuit

CLOUD NATIVE AI + INFERENCE DAY NORTH AMERICA

Thank You



Ship Fearlessly: Canary Deployments with an AI Quality Gate for Control Plane Releases

intuit