Red Hat Summit

Connect

FinOps intelligente: AI e costi ottimizzati con OpenShift

Luigi Ria
Professional Service Director, Kiratech

KIRATECH PLATFORM HEROES | Red Hat

KIRATECH®
PLATFORM HEROES

Professional Services Director – Kiratech
https://www.linkedin.com/in/luigi-ria-a63684b/

I design and bring AI solutions for enterprise companies into production, from conception to value delivery.

I lead teams on Platform Engineering, MLOps/LLMOps, and DevSecOps, integrating models, data, and processes with reproducible pipelines, observability, and model SLOs.

I design innovation roadmaps and translate them into secure, scalable and monitorable operating platforms.

Speaker on AI engineering and enterprise adoption to accelerate experimentation and reduce time to value

KIRATECH®
PLATFORM HEROES

AGENDA

1. FINOPS IN A NUTSHELL
2. FROM DEV TO AI
3. TOOLS & ARCHITECTURE – MLOPS + FINOPS
4. MLOPS + FINOPS – A COMPLETE FLOW
5. MLOPS + FINOPS – USE CASE

© 2025 - Kiratech Spa

KIRATECH®
PLATFORM HEROES

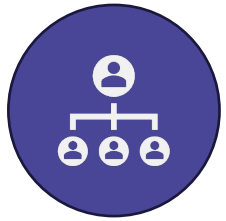
FINOPS IN A NUTSHELL

FinOps (Financial Operations) is a Cloud Financial Management discipline that combines financial and operational expertise to optimize cloud costs and maximize its value for the business

Service objectives: To help customers reduce waste and optimize the utilization of cloud resources, align IT spend with business objectives, and improve operational cost predictability. The perspective is not just savings for its own sake, but the maximum value from the cloud in terms of speed, quality and sustainable innovation

FULL TRANSPARENCY INTO CLOUD COSTS

and on expense drivers, being able to identify "hidden" costs and assign them correctly by team/project



DISTRIBUTED EMPOWERMENT

Not only does finance track costs, but each team (dev, ops, product) becomes aware of its cost impact

COST REDUCTION

(e.g. idle resources, over-provisioning) and improve overall efficiency

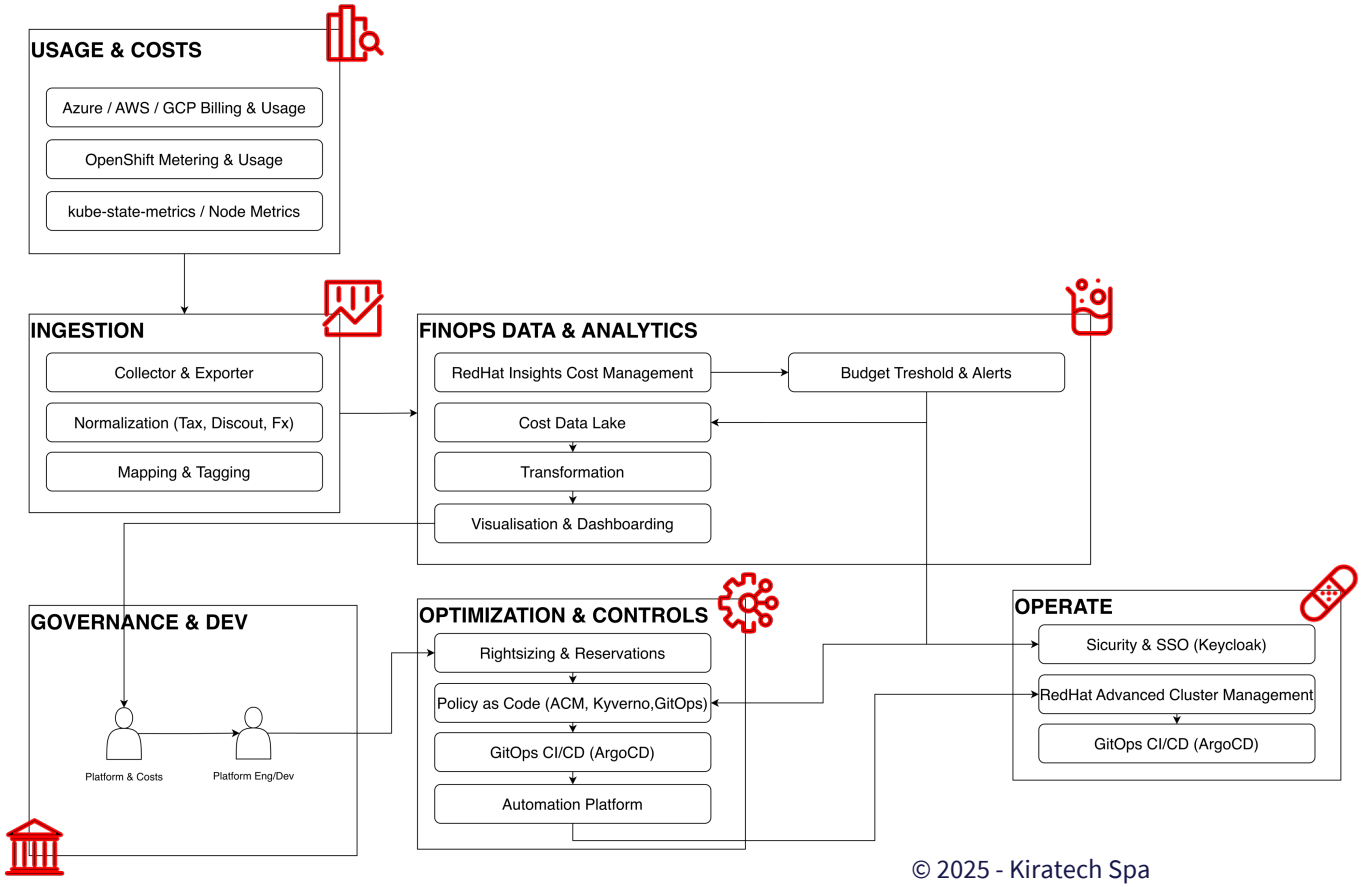
CONTINUOUS IMPROVEMENTS

more accurate budgeting, recurring optimizations and redirecting savings towards innovation.

KIRATECH®
PLATFORM HEROES

FROM DEV TO AI

To effectively integrate the goal of cost optimization, we must go beyond models to be elastic in managing changes in tools, infrastructure, and assets.

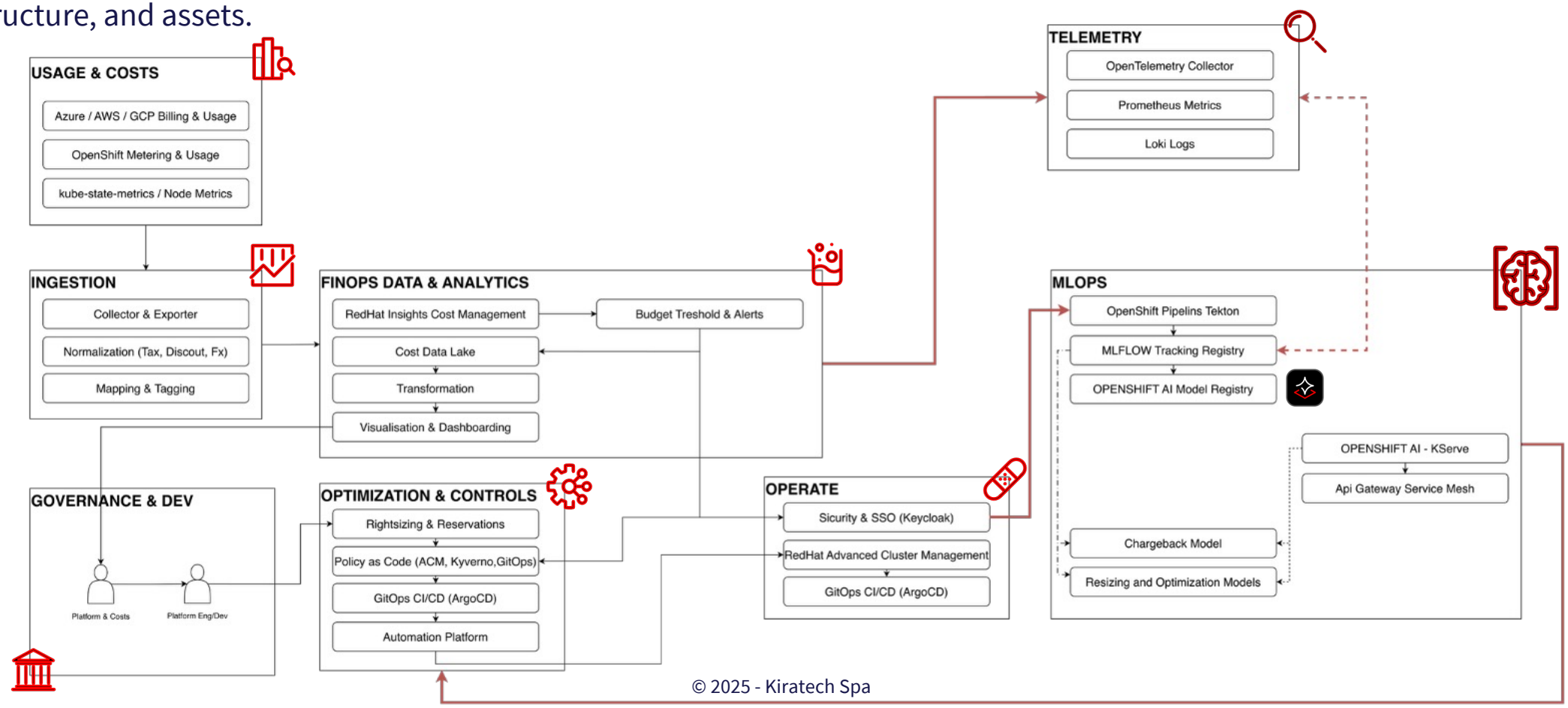


© 2025 - Kiratech Spa

KIRATECH
PLATFORM HEROES

FROM DEV TO AI

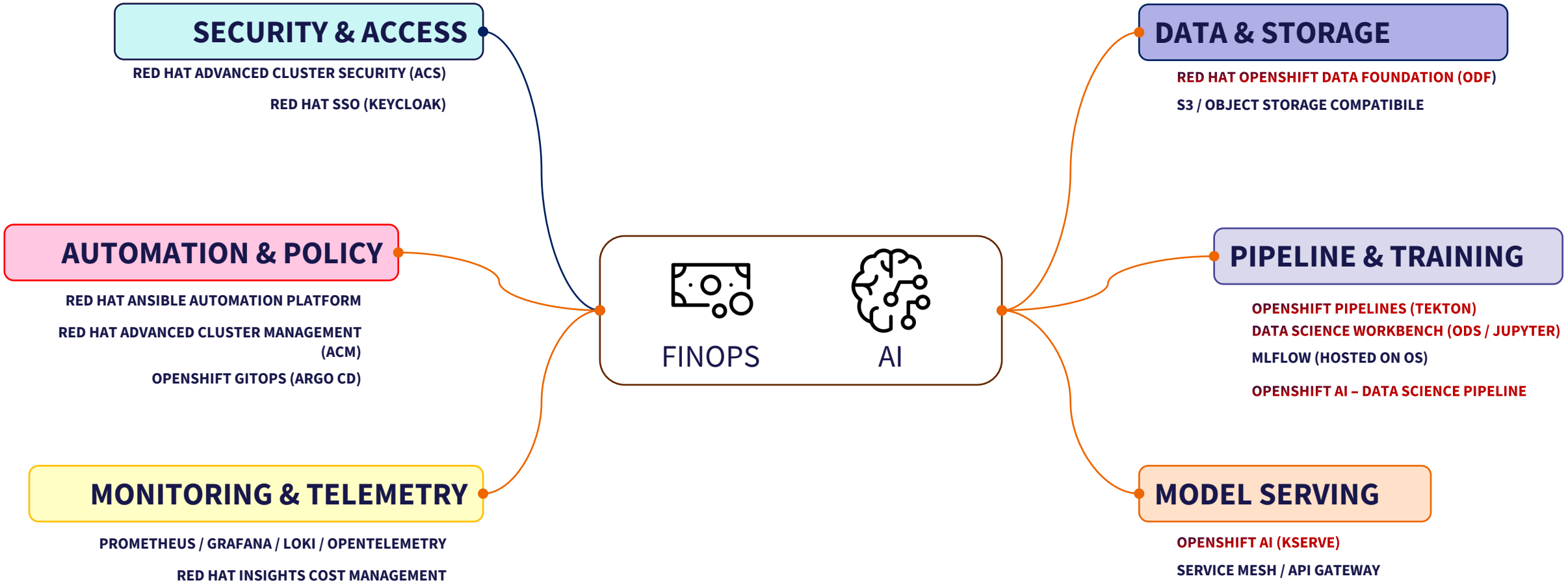
To effectively integrate the goal of cost optimization, we must go beyond models to be elastic in managing changes in tools, infrastructure, and assets.



© 2025 - Kiratech Spa

KIRATECH®
PLATFORM HEROES

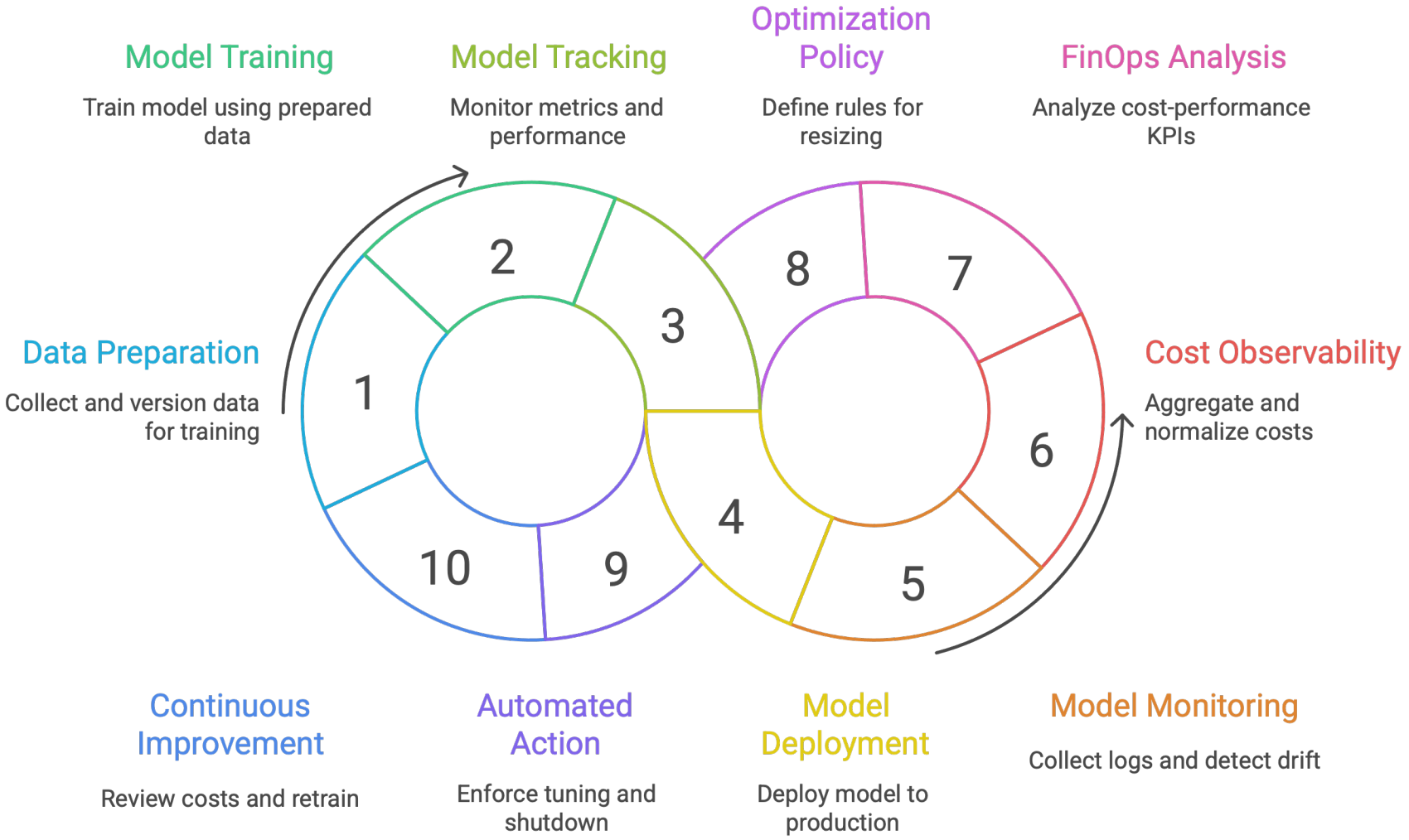
TOOLS & ARCHITECTURE – MLOPS + FINOPS



© 2025 - Kiratech Spa

KIRATECH
PLATFORM HEROES

MLOPS + FINOPS – A COMPLETE FLOW



© 2025 - Kiratech Spa

KIRATECH
PLATFORM HEROES

MLOPS + FINOPS – USE CASE

CENTRALIZE AI COST AND USAGE DATA

Collect information about training costs, GPU resources, and inference costs distributed across multiple OpenShift namespaces.

MEASURE COST PER RUN AND PER INFERENCE

Identify the most expensive and ineffective models in terms of ROI. Integrated Prometheus and MLflow Tracking: each training run records GPU time, energy usage and accuracy.

APPLY AUTOMATIC TUNING POLICIES

Automatically scale and schedule idle or ineffective resources. Use phased release strategies or canary releases to test the optimization scenario

VALIDATE IMPACT ON PERFORMANCE AND ACCURACY

Verify that FinOps optimizations do not reduce predictive quality. Integrated CI/CT pipeline with automatic drift and accuracy tests; monitoring in Grafana and alerts on Cost Dashboards

CONTINUOUS IMPROVEMENT & GOVERNANCE

Automate the cost analysis cycle → action → retraining. Implemented governance loops with ACM, Cost Mgmt, MLflow retrain triggers to adapt spend to the value produced.

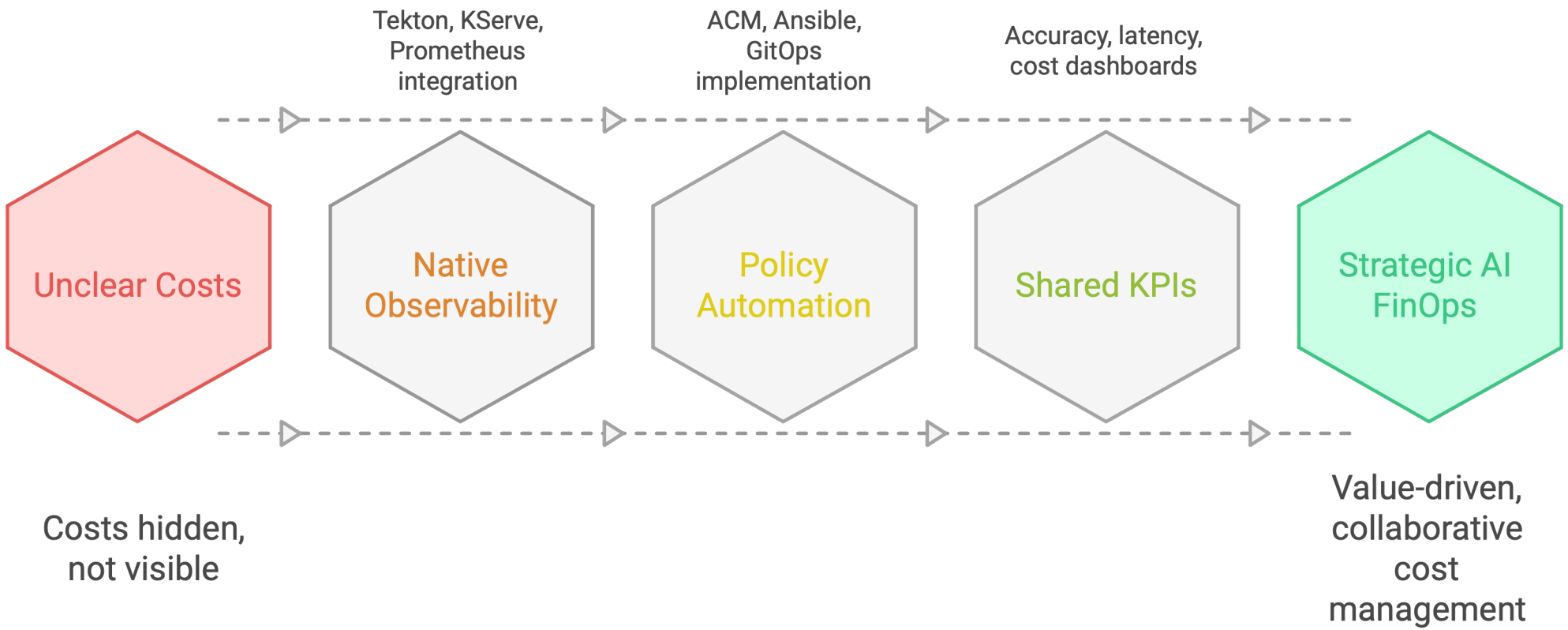
KPI

Training Cost VS. Infrastructure Cost Savings

© 2025 - Kiratech Spa

KIRATECH
PLATFORM HEROES

TAKEAWAY SLIDE



© 2025 - Kiratech Spa

KIRATECH®
PLATFORM HEROES

THANK YOU