EPIC EDGE | SCHEDA PRODOTTO

CephSentinel AI PRIMO PILOT

Monitoring continuo e automazione operativa per le flotte Ceph

Modulo dell'ecosistema SentinelForge AI · on-premise · sovrano · open-source · AI-native

Il problema

Ceph è affidabile ma non è semplice. Un cluster in produzione parla per soglie e codici di stato: dice che qualcosa non va, non perché né cosa fare. Chi gestisce Ceph passa il tempo a tradurre warning in decisioni, e quella traduzione vive nella testa di due o tre persone.

Con più di un cluster il problema cambia natura: cluster di clienti diversi, in datacenter diversi, alcuni dietro firewall che non espongono nulla. Nessuna vista d'insieme per sapere dove guardare oggi, e gli errori più costosi non sono i guasti: il flag noout lasciato acceso per settimane, la finestra di manutenzione che nessuno ha chiuso, il warning critico ignorato per ore.

La soluzione

CephSentinel AI è il modulo dell'ecosistema SentinelForge AI dedicato alle flotte Ceph. Non è un sistema di alerting in più: è un demone che osserva in autonomia, correla, deduplica, classifica e prioritizza i segnali invece di aspettare che qualcuno faccia una domanda, e gestisce più cluster contemporaneamente — anche geograficamente distinti e non co-locati con il sistema stesso.

Ogni cluster è un oggetto di primo livello, con credenziali dedicate e policy di autonomia configurabile. Il cuore non è la chat né il modello: è la vista di flotta, un'unica schermata che risponde a «dove devo guardare oggi», con i cluster ordinati per urgenza. È ciò che un operatore con decine di cluster oggi non ha.

Il modello propone. Il motore di automazione esegue.

Come funziona

1 · OSSERVA

Collector per cluster: stato strutturato, topologia, utilizzo per pool, stato dei placement group, metriche SMART. Ciclo continuo con trigger immediato sui warning critici.

2 · INTERPRETA

Livello a regole per le soglie note, livello LLM solo dove serve contesto: una spiegazione in linguaggio naturale calata su quel cluster.

3 · CORRELA E PREVEDE

Collega segnali che le soglie singole non collegherebbero mai — un problema di rete su

4 · PROPONE E AGISCE

Piano d'azione leggibile, classificato per rischio secondo la policy di quel cluster. Esecuzione

un rack e il flapping degli OSD su quel rack sono un incidente solo. Proiezioni di capacità sul trend storico.

tracciata, mai il modello direttamente sul cluster.

> **«Dove devo guardare oggi sui miei 12 cluster Ceph?»**
>
> **CephSentinel AI →**
>
> Rack 2 — Torino: HEALTH_ERR · errore di rete correlato al flapping degli OSD · impatto: alto
>
> **Cluster ordinati per urgenza, azione proposta | esecuzione solo su conferma**

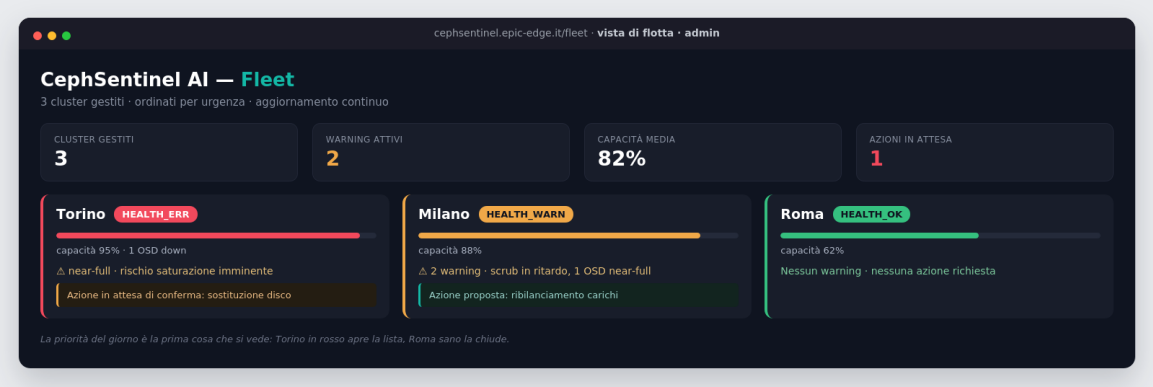


Figura 1 — Vista di flotta: i cluster ordinati per urgenza. La priorità del giorno è la prima cosa che si vede.



Figura 2 — Anatomia di un incidente: segnali diversi nel tempo correlati in un'unica diagnosi, prima che diventino disservizio.

Caratteristiche chiave

- **Vista di flotta, non di cluster** — la home risponde a «dove devo guardare oggi»: cluster ordinati per urgenza, warning attivi e azioni in attesa di conferma, con la connettività persa segnalata diversamente da un cluster sano.
- **Salute reale, non solo il codice Ceph** — un cluster tecnicamente HEALTH_OK ma con una previsione di esaurimento capacità imminente non appare «tutto verde».
- **Correlazione multi-segnale** — il punto in cui l'AI aggiunge valore reale rispetto all'alerting tradizionale: due warning distinti che sono in realtà un unico incidente con una causa radice.
- **Due modelli di connettività** — polling diretto verso l'endpoint del cliente, oppure agente leggero in uscita: nessuna porta in ingresso da aprire, requisito spesso non negoziabile in ambienti storage-critical.

- **Nessun vincolo sul metodo di deploy** — opera su qualunque cluster Ceph via CLI o via API dove disponibili: non solo cluster gestiti da cephadm.
- **Autonomia proporzionale al rischio** — tre livelli configurabili per cluster e per cliente: sola segnalazione, operatività quotidiana previa conferma, oppure consulenza sulle sole modifiche ad alto impatto.
- **Isolamento per cliente** — credenziali dedicate per cluster, mai condivise: un cluster compromesso non dà accesso agli altri.
- **LLM interamente on-premise** — nessun dato di cluster, nessuna topologia, nessuna credenziale esce dal perimetro dell'organizzazione.

Benefici per il cliente

Una flotta, un operatore
Gestire dieci cluster smette di significare dieci sessioni aperte: la priorità del giorno è la prima cosa che si vede.

Meno dimenticanze costose
Flag con timeout, finestre di manutenzione sorvegliate ed escalation sulle conferme mancate: gli errori operativi più comuni diventano strutturalmente difficili.

Competenza Ceph distribuita
La spiegazione contestualizzata rende operativo chi non ha vent'anni di Ceph alle spalle, senza togliere il controllo a chi ce li ha.

Adottabile anche in autonomia
CephSentinel AI può essere adottato anche da solo, su flotte Ceph che non passano dalla piattaforma: pensato per chi gestisce storage distribuito come servizio o infrastruttura critica.

Sovranità del dato preservata
LLM on-premise: monitorare con l'AI non impone di mandare la telemetria dei propri cluster a un servizio esterno.

Ogni azione difendibile in audit
Cosa è stato eseguito, su quale cluster, per quale segnale e con quale approvazione: requisito, non accessorio, nei contesti regolamentati.

Stack e tecnologie

Telemetria	Prometheus + exporter nativo di ceph-mgr — lo standard Ceph, non reinventato
Motore LLM	Modelli open-weight eseguiti on-premise via Ollama — nessuna dipendenza da API esterne
Collector	Servizio Python dedicato per cluster — stato strutturato completo, non solo serie numeriche
Esecuzione	AWX + Ansible, sopra l'orchestratore nativo Ceph dove presente; via CLI o API sui cluster che non lo espongono
Versionamento e audit	GitLab — ogni playbook legato a cluster ed evento di origine
Dashboard e notifiche	React — vista di flotta e drill-down per cluster; notifiche via Email, Slack, Teams, SMS, configurabili per cluster

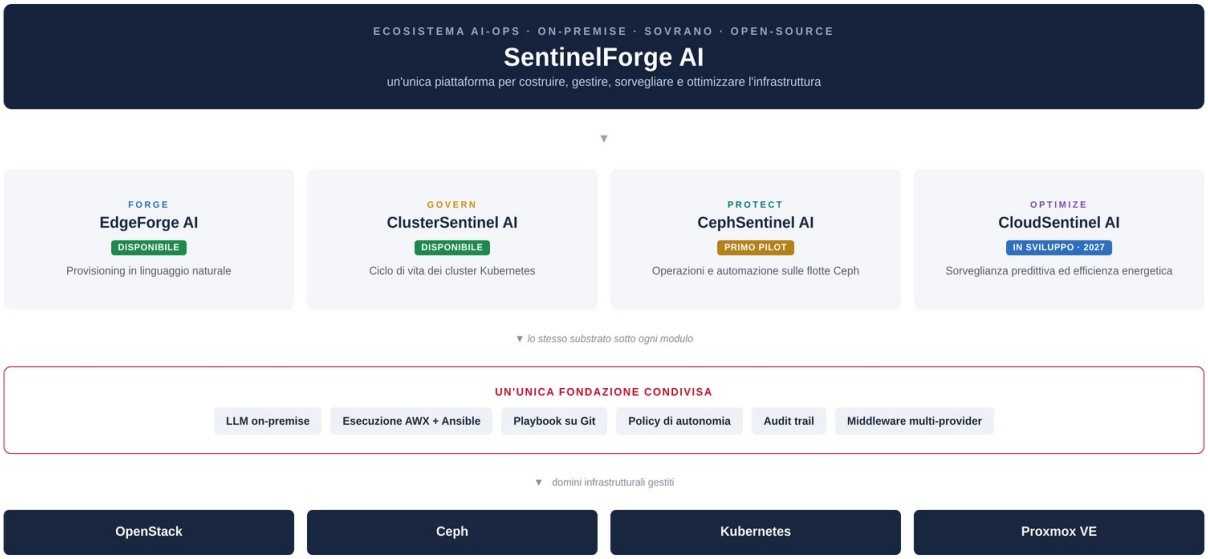
A chi si rivolge

- **Telco & ISP** — più cluster in più datacenter, dove la vista d'insieme oggi non esiste.
- **Service provider e MSP** — chi gestisce Ceph per conto di clienti terzi, con referenti, SLA e livelli di autonomia diversi per ciascuno.
- **Pubblica Amministrazione** — requisiti di sovranità del dato e tracciabilità degli interventi sull'infrastruttura.
- **Finance & Banking** — storage ad alta disponibilità dove ogni modifica deve lasciare una traccia di audit completa.
- **Enterprise multi-datacenter** — storage distribuito su siti diversi, con un team centrale che non può presidiare tutto a mano.

Modello di deployment

Stato	Primo pilot disponibile — osservazione multi-cluster e azioni supervisionate.
Modalità	On-premise, all'interno del perimetro del cliente. Nessuna componente SaaS obbligatoria.
Connettività	Polling diretto verso l'endpoint del cluster, oppure agente leggero in uscita per cluster dietro firewall restrittivi.
Compatibilità	Cluster Ceph in produzione, indipendentemente dal metodo di deploy — accesso via CLI o API.
Avvio	Registrazione del cluster tramite wizard guidato con test di raggiungibilità immediato; policy iniziale prudente, raffinabile dopo.
Autonomia	Configurabile per cluster sui tre livelli di rischio; il perimetro delle automazioni si concorda in fase di attivazione — si parte con tutto supervisionato.

L'ecosistema Epic Edge



Ogni prodotto Epic Edge fa parte di SentinelForge AI: moduli distinti sullo stesso substrato — LLM on-premise, esecuzione tracciata, policy di autonomia, audit, middleware multi-provider.

Principi architetturali

Realizzare una piattaforma di automazione AI per l'infrastruttura significa tenere insieme cose che di solito vivono separate: integrare provider tecnologicamente diversi, mantenere adapter indipendenti che evolvono senza toccare il resto, garantire un audit completo di ogni azione, preservare la sovranità del dato e — soprattutto — separare il motore che decide dal motore che esegue.

È per questo che CephSentinel AI nasce come modulo dell'ecosistema SentinelForge AI: non una funzione aggiunta a uno strumento esistente, ma un'architettura pensata da chi quei vincoli li ha vissuti in produzione.

Perché Epic Edge

Ceph in produzione, non in teoria

Cluster reali, incidenti reali, SLA reali. La competenza viene dal campo, non da una startup che immagina lo storage.

Pattern già collaudati

Motore LLM, esecuzione via AWX + Git, streaming real-time: già validati in produzione con EdgeForge AI e ClusterSentinel AI.

Open-source e sovrano

Nessun lock-in, nessuna telemetria verso terzi: tutto dentro il perimetro del cliente.

Quanti cluster Ceph stai gestendo a mano?

Il primo pilot è disponibile. Registriamo un vostro cluster e vi mostriamo cosa vede — senza toccare nulla, finché non lo decidete voi.

info@epic-edge.it · www.epic-edge.it/epic-edge-platform



Epic Edge S.r.l. · Via Michele Buniva, 26 — 10064 Pinerolo (TO), Italia · info@epic-edge.it · www.epic-edge.it

OPEN-SOURCE • SOVEREIGN • ENTERPRISE-GRADE • AI-DRIVEN