EPIC EDGE | SCHEDA PRODOTTO

EdgeForge AI LIVE

Deploy di infrastruttura cloud in linguaggio naturale

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

Il problema

Serve un ambiente in cloud: l'utente descrive cosa gli occorre, e qualcuno deve costruirlo. Manualmente. Tra raccolta dei requisiti, dimensionamento, provisioning e configurazione applicativa passano giorni, non minuti.

Il costo non è solo il tempo: ogni deploy manuale introduce varianti non documentate, errori di configurazione ed esiti diversi a seconda di chi lo esegue. La conoscenza resta nelle mani di poche persone e l'onboarding di un nuovo ingegnere richiede mesi.

La soluzione

EdgeForge AI è il modulo di provisioning conversazionale dell'ecosistema SentinelForge AI. L'utente descrive in linguaggio naturale l'infrastruttura di cui ha bisogno; il sistema interpreta l'intento, progetta l'architettura, dimensiona le risorse, stima il costo e — dopo l'approvazione — esegue il deploy automatizzato su OpenStack.

Sotto il cofano non c'è improvvisazione: Terraform crea le risorse e Ansible installa il software, a partire da una libreria di template validati e versionati. Il modello linguistico non esegue mai comandi direttamente sull'infrastruttura — genera un piano che passa da un motore di esecuzione separato, tracciato e verificabile.

Il modello propone. Il motore di automazione esegue.

Come funziona

1 · DESCRIVI
L'utente scrive in linguaggio naturale ciò di cui ha bisogno. Nessuna sintassi, nessun template da compilare.

2 · L'AI PROPONE
Il modello seleziona il template, dimensiona le risorse e stima il costo mensile.

3 · VERIFICA E APPROVA
Diagramma dell'architettura e anteprima dei costi prima di ogni deploy. Si raffina via chat finché il piano non convince.

4 · DEPLOY AUTOMATICO
Terraform crea le risorse, Ansible installa e configura il software. Deploy completo in meno di 8 minuti.

> **"Ho bisogno di una VM con AWX installato, HTTPS sulla porta 443 e un disco**

> **PostgreSQL da 50GB.”**
>
> **EdgeForge AI →**
>
> 1x VM (4 vCPU, 16GB RAM) · disco 50GB · HTTPS Let’s Encrypt · PostgreSQL su volume dedicato
>
> **Costo stimato: 45 €/mese | Deploy automatico in pochi minuti**

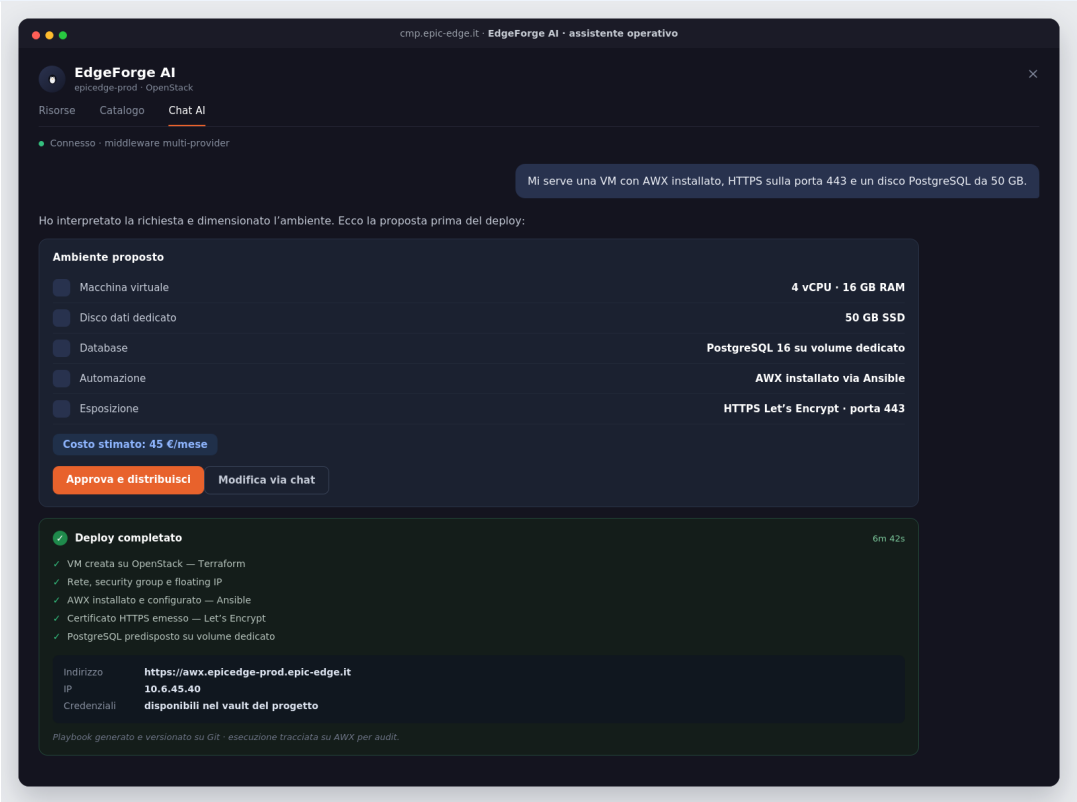

Figura 1 — EdgeForge AI: dalla richiesta in linguaggio naturale alla proposta con stima dei costi, fino all’esecuzione tracciata del piano approvato.

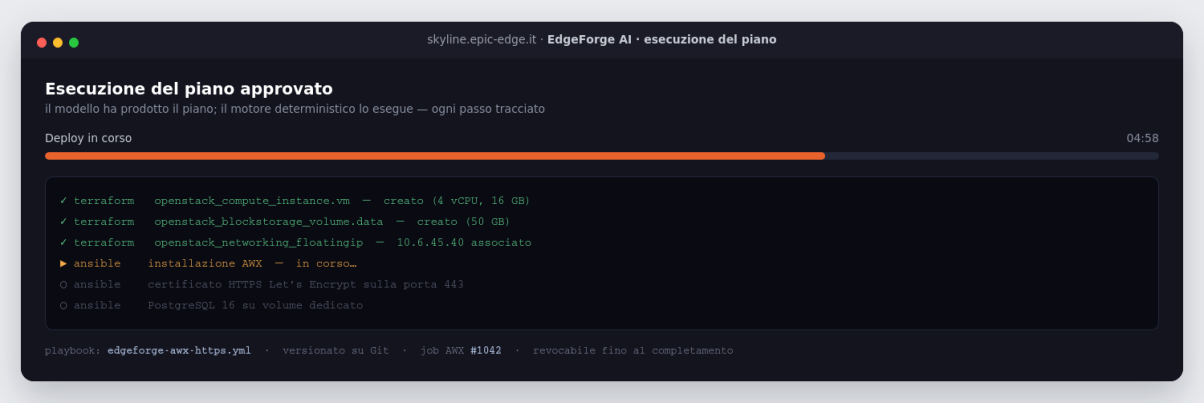

Figura 2 — L’esecuzione del piano: Terraform crea le risorse, Ansible installa e configura — ogni passo tracciato e revocabile fino al completamento.

Caratteristiche chiave

- **Provisioning conversazionale in italiano** — l’intento dell’utente viene interpretato, tradotto in architettura e dimensionato automaticamente.
- **LLM interamente on-premise** — il motore linguistico gira all’interno del perimetro del cliente. Nessun dato, nessuna descrizione di infrastruttura, nessuna credenziale lascia la rete.

- **Libreria di template validati** — WordPress HA, MySQL Galera, cluster Kubernetes, VPN Gateway, JupyterHub, PostgreSQL HA. Template custom generati dinamicamente per i casi non coperti.
- **Approvazione pre-deploy con stima dei costi** — nessuna risorsa viene creata senza che un operatore abbia visto architettura e costo.
- **Ogni deploy è versionato e tracciabile** — ogni playbook generato viene salvato su Git: resta sempre un artefatto ispezionabile di cosa è stato eseguito e perché.
- **Esecuzione disaccoppiata dal modello** — il modello genera il piano, un motore di esecuzione separato (AWX + Ansible) lo esegue. È ciò che rende possibili audit, dry-run e revoca prima dell'esecuzione.
- **Integrazione nativa in OpenStack Skyline** — non un portale parallelo: EdgeForge AI vive dentro la dashboard che gli operatori usano già.
- **Orchestrazione multi-VM** — richieste complesse vengono scomposte in un piano multi-nodo, presentato all'operatore per conferma prima dell'esecuzione.

Benefici per il cliente

Da giorni a minuti
Il ciclo richiesta → ambiente operativo si comprime da giorni di lavoro manuale a un deploy automatizzato di pochi minuti.

Errori operativi ridotti
Il deploy parte da template validati e testati: meno configurazioni manuali, meno varianti non documentate.

Deployment standardizzati
Lo stesso risultato indipendentemente da chi esegue la richiesta. La conoscenza operativa è nel sistema, non in poche persone.

Onboarding più rapido
Un nuovo ingegnere è produttivo senza dover prima padroneggiare l'intero stack di provisioning.

Sovranità del dato preservata
LLM on-premise: l'automazione AI non impone di esporre la propria infrastruttura a un servizio esterno.

Produttività del team
Gli ingegneri smettono di eseguire deploy ripetitivi e tornano su architettura e affidabilità.

Stack e tecnologie

Cloud IaaS	OpenStack (Nova, Neutron, Cinder, Glance, Keystone, Skyline)
Motore LLM	Modelli open-weight eseguiti on-premise via Ollama — nessuna dipendenza da API esterne
Infrastructure as Code	Terraform (provisioning risorse) + Ansible (configurazione software)
Motore di esecuzione	AWX / Ansible Automation Controller
Versionamento	GitLab — ogni playbook generato è salvato, versionato e ispezionabile
Interfaccia	Integrazione nativa nella dashboard OpenStack Skyline

A chi si rivolge

- **Telco & ISP** — operatori con OpenStack di produzione e un volume elevato di richieste di provisioning ripetitive.
- **Cloud & Service Provider** — chi rivende risorse cloud e vuole ridurre il costo operativo per ogni ambiente erogato.
- **Pubblica Amministrazione** — automazione AI compatibile con requisiti di sovranità del dato: nulla esce dal perimetro.
- **Finance & Banking** — contesti dove ogni modifica all'infrastruttura deve lasciare una traccia di audit completa.
- **Enterprise IT** — organizzazioni con cloud privato e un team infrastrutturale sottodimensionato rispetto alla domanda interna.

Modello di deployment

Modalità	On-premise, all'interno del perimetro del cliente. Nessuna componente SaaS obbligatoria.
Prerequisiti	Cluster OpenStack esistente; risorse GPU dedicate al motore di inferenza on-premise.
Integrazione	Nativa nella dashboard Skyline esistente — nessun portale parallelo da adottare.
Attivazione	Assessment iniziale, deployment del motore AI, integrazione Skyline, validazione dei template sui workload del cliente.
Personalizzazione	Estensione della libreria di template sui workload specifici del cliente.
Supporto	Erogato da Epic Edge sull'intero stack (OpenStack, Ceph, motore AI), non solo sul modulo.

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 EdgeForge 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

25+ anni di operations

Data center reali, SLA reali, clienti enterprise reali. Non una startup che immagina il cloud.

Già in produzione

EdgeForge AI è live su Skyline oggi — non un prototipo, non una demo.

Nativo sullo stack

Costruito su OpenStack, Ceph e Kubernetes: lo stack dei cloud sovrani e telco europei.

Vuoi vedere EdgeForge AI sul tuo stack?

Organizziamo una demo sul vostro caso d’uso reale, con un assessment preliminare del vostro ambiente OpenStack.

andrea.martra@epic-edge.it · www.epic-edge.it/sentinelforgeai

EdgeForge AI	CephSentinel AI	CMP	CloudSentinel AI
LIVE	PILOT	IN IMPLEMENTAZIONE	ROADMAP

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