EPIC EDGE | PRODUCT SHEET

# EdgeForge AI LIVE

Cloud infrastructure deployment in natural language

*Module of the SentinelForge AI ecosystem · on-premise · sovereign · open-source · AI-native*

## The problem

An environment in the cloud is needed: the user describes what they require, and someone has to build it. Manually. Between requirements gathering, sizing, provisioning and application configuration, days go by, not minutes.

The cost is not only time: every manual deployment introduces undocumented variants, configuration errors and different outcomes depending on who carries it out. The knowledge remains in the hands of a few people and onboarding a new engineer takes months.

## The solution

EdgeForge AI is the conversational provisioning module of the SentinelForge AI ecosystem. The user describes in natural language the infrastructure they need; the system interprets the intent, designs the architecture, sizes the resources, estimates the cost and — after approval — carries out the automated deployment on OpenStack.

Under the hood there is no improvisation: Terraform creates the resources and Ansible installs the software, starting from a library of validated and versioned templates. The language model never executes commands directly on the infrastructure — it generates a plan that goes through a separate execution engine, traced and verifiable.

**The model proposes. The automation engine executes.**

## How it works

**1 · DESCRIBE**
The user writes in natural language what they need. No syntax, no template to fill in.

**2 · THE AI PROPOSES**
The model selects the template, sizes the resources and estimates the monthly cost.

**3 · VERIFY AND APPROVE**
Architecture diagram and cost preview before every deployment. It is refined via chat until the plan is convincing.

**4 · AUTOMATIC DEPLOYMENT**
Terraform creates the resources, Ansible installs and configures the software. Complete deployment in under 8 minutes.

> **“I need a VM with AWX installed, HTTPS on port 443 and a 50GB PostgreSQL disk.”**

**EdgeForge AI →**

1x VM (4 vCPU, 16GB RAM) · 50GB disk · HTTPS Let's Encrypt · PostgreSQL on a dedicated volume

**Estimated cost: €45/month | Automatic deployment in a few minutes**

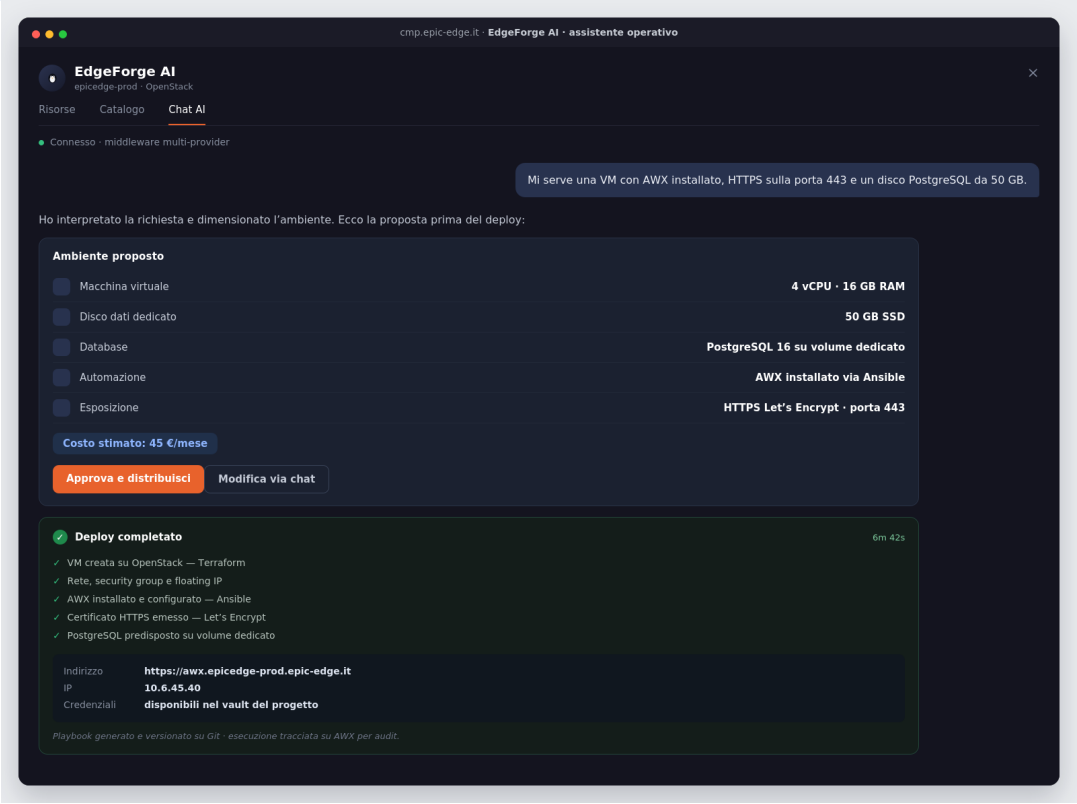


*Figure 1 — EdgeForge AI: from the natural-language request to the proposal with cost estimate, through to the traced execution of the approved plan.*

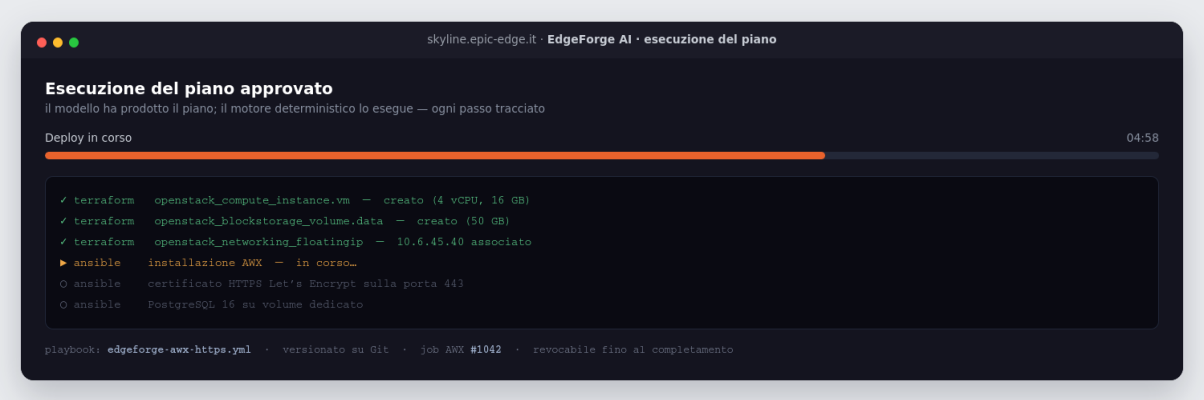


*Figure 2 — Execution of the plan: Terraform creates the resources, Ansible installs and configures — every step traced and revocable until completion.*

## Key features

- **Conversational provisioning in Italian** — the user's intent is interpreted, translated into an architecture and sized automatically.
- **Fully on-premise LLM** — the language engine runs inside the customer's perimeter. No data, no infrastructure description, no credential leaves the network.
- **Library of validated templates** — WordPress HA, MySQL Galera, Kubernetes clusters, VPN Gateway, JupyterHub, PostgreSQL HA. Custom templates generated dynamically for the cases not covered.

- **Pre-deployment approval with cost estimate** — no resource is created without an operator having seen the architecture and the cost.
- **Every deployment is versioned and traceable** — every generated playbook is saved on Git: an inspectable artifact of what was executed and why always remains.
- **Execution decoupled from the model** — the model generates the plan, a separate execution engine (AWX + Ansible) carries it out. It is what makes audit, dry-run and revocation before execution possible.
- **Native integration in OpenStack Skyline** — not a parallel portal: EdgeForge AI lives inside the dashboard that operators already use.
- **Multi-VM orchestration** — complex requests are broken down into a multi-node plan, presented to the operator for confirmation before execution.

## Benefits for the customer

### From days to minutes

The request → operational environment cycle is compressed from days of manual work to an automated deployment of a few minutes.

### Fewer operational errors

Deployment starts from validated and tested templates: fewer manual configurations, fewer undocumented variants.

### Standardized deployments

The same result regardless of who carries out the request. The operational knowledge is in the system, not in a few people.

### Faster onboarding

A new engineer is productive without first having to master the entire provisioning stack.

### Data sovereignty preserved

On-premise LLM: AI automation does not require exposing your own infrastructure to an external service.

### Team productivity

Engineers stop carrying out repetitive deployments and go back to architecture and reliability.

## Stack and technologies

| | |
|---|---|
| **Cloud IaaS** | OpenStack (Nova, Neutron, Cinder, Glance, Keystone, Skyline) |
| **LLM engine** | Open-weight models run on-premise via Ollama — no dependency on external APIs |
| **Infrastructure as Code** | Terraform (resource provisioning) + Ansible (software configuration) |
| **Execution engine** | AWX / Ansible Automation Controller |
| **Versioning** | GitLab — every generated playbook is saved, versioned and inspectable |
| **Interface** | Native integration in the OpenStack Skyline dashboard |

## Who it is for

- **Telco & ISP** — operators with production OpenStack and a high volume of repetitive provisioning requests.
- **Cloud & Service Provider** — those who resell cloud resources and want to reduce the operational cost of each environment delivered.
- **Public administration** — AI automation compatible with data sovereignty requirements: nothing leaves the perimeter.
- **Finance & Banking** — contexts where every infrastructure change must leave a complete audit trail.
- **Enterprise IT** — organizations with a private cloud and an infrastructure team undersized relative to internal demand.

## Deployment model

| | |
|---|---|
| **Mode** | On-premise, within the customer's perimeter. No mandatory SaaS component. |
| **Prerequisites** | Existing OpenStack cluster; GPU resources dedicated to the on-premise inference engine. |
| **Integration** | Native in the existing Skyline dashboard — no parallel portal to adopt. |
| **Activation** | Initial assessment, deployment of the AI engine, Skyline integration, validation of the templates on the customer's workloads. |
| **Customization** | Extension of the template library to the customer's specific workloads. |
| **Support** | Delivered by Epic Edge across the entire stack (OpenStack, Ceph, AI engine), not just the module. |

# The Epic Edge ecosystem


*Every Epic Edge product is part of SentinelForge AI: distinct modules on the same substrate — on-premise LLM, traced execution, autonomy policies, audit, multi-provider middleware.*

## Architectural principles

Building an AI automation platform for infrastructure means holding together things that usually live separately: integrating technologically different providers, maintaining independent adapters that evolve without touching the rest, guaranteeing a complete audit of every action, preserving data sovereignty and — above all — separating the engine that decides from the engine that executes.

This is why EdgeForge AI was created as a module of the SentinelForge AI ecosystem: not a function added to an existing tool, but an architecture designed by people who have lived with those constraints in production.

## Why Epic Edge

**25+ years of operations**
Real data centers, real SLAs, real enterprise customers. Not a startup imagining the cloud.

**Already in production**
EdgeForge AI is live on Skyline today — not a prototype, not a demo.

**Native to the stack**
Built on OpenStack, Ceph and Kubernetes: the stack of European sovereign and telco clouds.

### Want to see EdgeForge AI on your stack?

Let us arrange a demo on your real use case, with a preliminary assessment of your OpenStack environment.

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

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

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