

POWERED BY ENEY CLOUD

Private Cloud Proposal

AX를 위한 클라우드 운영관리 솔루션

에네이 프라이빗 클라우드 · AI-Native Serving Platform



제안 개요

네트워크·보안 인프라 구축 경험과 ENEY CLOUD의 AI-Ops 운영 기술을 결합하여, 귀사 전용의 안전하고 효율적인 AX 인프라 환경을 제안합니다.

제안 배경

- 온프레미스·레거시 인프라의 운영 복잡도 증가
- 생성형 AI·LLM 도입에 따른 GPU 자원 운영 부담
- 보안·백업·모니터링 개별 구축에 따른 비용 부담

제안 목적

- 귀사 전용 프라이빗 클라우드 환경 구축
- 컨테이너 기반 LLM 배포 자동화로 AX 전환 가속
- GPU 자원의 효율적 배분과 운영 최적화 체계 확립

제공 범위

- 프라이빗 클라우드 인프라(IaaS·PaaS Node) 구축
- 컨테이너 기반 AI·LLM 서비스 배포 플랫폼
- GPU 모니터링·최적화 및 통합 운영 포털

단계적 구축

레거시 장비 환경 기반으로 초기 투자 부담 최소화

독립 운영

자원은 공유, 정책·네트워크·보안은 테넌트별 분리

병렬 확장

서비스 규모와 GPU 워크로드에 따라 확장 가능한 구조

서비스 특장점

컨테이너·인스턴스·보안·백업·모니터링을 표준화된 정책과 승인 체계로 관리하고, LLM 배포와 GPU 운영까지 아우르는 AX 통합 운영 체계를 제공합니다.

01

시큐어 엣지 기반 통합 관리

서비스 영역별 접근 정책 운영과 컨테이너·인스턴스 통합 관리

02

클라우드 운영관리 자동화

프로비저닝부터 모니터링·장애 대응까지 표준화·자동화

03 · AX

컨테이너 기반 LLM 배포

검증된 모델과 서빙 프레임워크를 즉시 배포, 멀티테넌시 운영

04

통합 시스템 관리 포털

중앙 대시보드와 알림, 백업·보안 정책의 통합 관리

05

Dedicated Tenant Policy

테넌트 간 자원·보안 완전 분리로 독립성과 효율 동시 확보

06 · AX

GPU 운영관리 최적화

유연한 GPU 할당과 실시간 사용률 모니터링, 시간당 과금

AX 차별화

① 컨테이너 기반 LLM 배포

② GPU 운영관리 최적화

컨테이너 기반 LLM 배포

3-Click

Pod 생성으로 LLM 서비스 즉시 배포

사전 검증된 모델과 서빙 프레임워크를 컨테이너로 배포하여, GPU 인프라 구성 없이 AI 서비스를 실행합니다.

서빙 프레임워크 즉시 배포

Node.js · Spring Boot · nginx · Ollama · vLLM

검증된 LLM 6종

Qwen · Gemma · Mistral — 선택만으로 설정 자동화

GPU 스택 최적 매칭

모델 크기별 L4 · 3090Ti · A6000 즉시 프로비저닝

Namespace 멀티테넌시

고객사·서비스별 워크로드 독립 운영

GPU 운영관리 최적화

GPU 인스턴스를 워크로드 규모에 맞춰 유연하게 할당하고, 실시간 모니터링으로 자원 낭비 없는 운영이 가능합니다.

컨테이너 단위 할당

모델 크기와 워크로드 규모에 맞춘 탄력적 확장

실시간 리소스 모니터링

GPU·CPU·메모리 사용률 가시화로 유휴 자원
최소화

통합 대시보드

자원 현황을 한눈에 파악하고 병목 구간에 신속
대응

GPU / CPU / MEM

사용률 실시간 추적

BANDWIDTH

병목 감지 및 알림

시간당 과금

PoC부터 상시 운영까지 비용 효율화

프라이빗 클라우드 상품 구성



기존 레거시 장비 환경 기반 구축 가능 · GPU-AI 워크로드에 따라 병렬 확장 · 구축 범위와 보안 정책 수준은 최종 협의

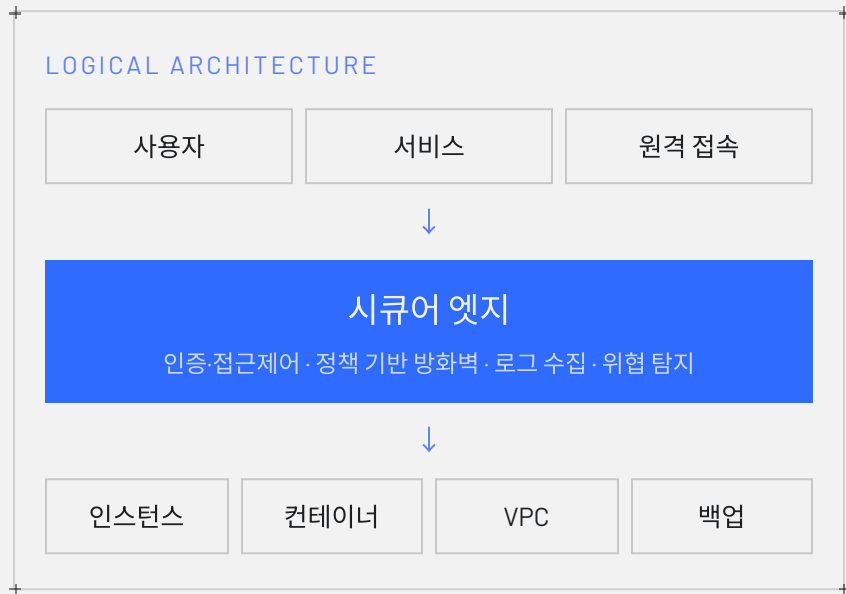
목표 고객

전용 인프라 기반의 독립 운영 환경이 필요하며, 자체 보안·네트워크 정책과 AI-GPU 워크로드를 독립적으로 관리하려는 조직

- 중견·대기업
- 공공기관
- 금융·의료·제조
- SaaS 사업자
- AI 고도화 기업

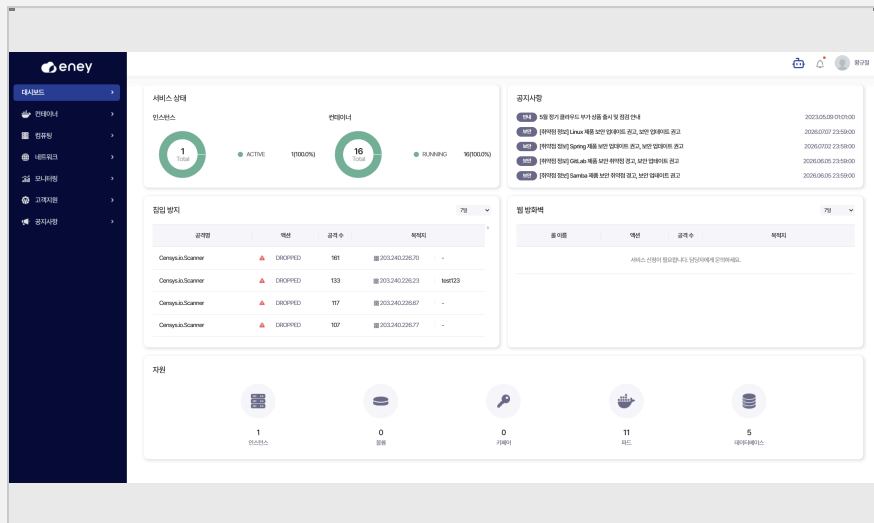
시큐어 엣지 서비스

사용자와 서비스가 접속하는 가장 가까운 지점에서 인증, 접근제어, 방화벽, 로그 수집, 위협 탐지를 수행하는 통합 보안 아키텍처입니다.



Network	공인 IP · IPsec VPN · SSL-VPN · IP/Port 기반 정책
방화벽	IPS · 웹 공격 탐지(기본) · 트래픽 로깅
웹 방화벽	웹 공격 탐지 (Expert)
백업	인스턴스 백업 (매일)
보안관제	실시간 모니터링 (24×365) · 침입 탐지 대응 · 정기 보고서

통합 운영 포털



인스턴스 · 컨테이너 운영 상태

가상 서버와 LLM 컨테이너의 실행 상태 확인

GPU · 컴퓨팅 자원 현황

GPU 할당량과 사용률 등 AI 워크로드 자원 모니터링

보안 이벤트 및 침입 탐지

IPS·웹 방화벽 로그 기반 위협 징후 탐지

네트워크 · VPC 현황

VPC 연결 구조와 트래픽 흐름 점검

임계값 및 알림 설정

Global·Custom 임계값 초과 시 이메일 알림 발송

제안사 소개 및 도입 효과

2014

설립 이후 11년

314+

기업·공공 고객 레퍼런스

ISO 27001

정보보호 관리체계 심사원보 보유

BBB

기술평가등급 T4·신용평가

운영 효율성 강화

통합 대시보드 기반 실시간 가시성으로 신속한 의사결정

AI·GPU 운영 최적화

LLM 배포 자동화와 GPU 자원 최적 배분으로 AX 가속

보안 수준 향상

시큐어 엣지 기반 테넌트별 독립 보안 체계

비용 효율화

레거시 인프라 활용과 시간당 GPU 과금으로 투자 부담 절감

유연한 확장성

서비스 규모와 AI 워크로드에 따른 병렬 확장 구조

레퍼런스 고객

 창원산업단지공단	 주식회사 해인	 (주)유니씨앤	 (주)이아이씨앤	 송실대학교	 이화여자대학교
 한양대학교	 충북대학교	 서울시립대학교	 주식회사 인프로	 충남대학교	 한국자산신탁
 경기테크노파크 GYEONGGI TECHNOPARK	 세종텔레콤	 SK브로드밴드	 스테이지파이브	 핀인사이트	 와이비즈

함께 확인할 사항

현재 인프라 구성과 활용 가능한
레거시 장비 범위

도입 예정 AI-LLM 서비스와
GPU 워크로드 규모

요구 보안 정책 수준과
연동 대상 시스템

“Everything and Everyone for You”

네트워크·보안 구축 노하우 + AI-Ops 운영 기술을 갖춘 AX 전환 파트너

www.enecloud.com

sales@enecloud.com

070-7933-1616