



Container Platform

Capsis



Contents

1.	개요	02
2.	비즈니스 임팩트	05
3.	Capsis 제품 구성 Multi-Cluster Orchestration Helm·애드온 관리 GPU·NPU 스케줄링	07
4.	관측성·보안·운영 Network Policy Workload Isolation	11
5.	Works with	14

1. 개요

엔터프라이즈 쿠버네티스를 AI 워크로드에 맞게 설계한 컨테이너 플랫폼입니다.

Capsis는 멀티 클러스터 오케스트레이션, 통제된 컨테이너 운영, AI 워크로드에 맞춘 스케줄링을 하나의 환경에서 제공합니다. 기업은 쿠버네티스를 직접 구축하고 운영하는 부담 없이 Paxis 에이전트, Metis inference 서비스, Maxis training 파이프라인을 실행할 수 있습니다.

01 하나의 관리 화면에서 운영

여러 환경과 지역에 흩어진 쿠버네티스 클러스터를 한 곳에서 운영합니다. 클러스터마다 다른 도구를 쓰지 않고 같은 절차와 같은 화면 구성을 모든 클러스터에 적용합니다.

02 통제된 컨테이너 운영

컨테이너 이미지 관리, 애드온 배포, 설정 변경이 정해진 절차와 이력 안에서 이루어집니다. 무엇이 언제 어디에 배포되었는지를 확인할 수 있습니다.

03 AI 워크로드를 이해하는 기반

네임스페이스 격리와 자원 할당량, GPU·NPU 구성을 반영한 배치로 training·inference·에이전트 워크로드가 서로 간섭하지 않고 공존합니다.

04 기업이 통제하는 환경 안에서

기업의 데이터센터 안에서 전부 운영합니다. 폐쇄망 환경을 포함하며 컨테이너 워크로드는 기업의 네트워크 정책과 보안 통제 아래에서 동작합니다.



여러 AI 플랫폼이 하나의 컨테이너 기반 위에서 동작하도록 설계했습니다.

쿠버네티스는 엔터프라이즈 AI 워크로드의 사실상 표준 실행 환경입니다. Capsis는 이 실행 기반을 상품으로 제공해 Paxis 에이전트와 Metis inference 서비스, Maxis training 파이프라인이 각자 기반을 만들지 않고 같은 기반 위에서 동작하도록 합니다.

AI 워크로드는 일반 애플리케이션과 요구 조건이 다릅니다. 가속기 구성을 반영한 배치, 워크로드 사이의 격리, 배포 이력의 추적을 처음부터 갖춘 상태로 설계했습니다.

01 하나의 기반을 여러 플랫폼이 공유

Paxis, Metis, Maxis가 각각의 쿠버네티스를 갖지 않고 하나의 실행 기반을 함께 씁니다. 플랫폼을 추가해도 기반을 다시 만들지 않습니다.

02 클러스터 수와 무관한 운영 방식

클러스터가 늘어나도 같은 절차와 같은 관리 화면을 씁니다. 환경과 지역이 달라도 운영 방식은 하나입니다.

03 워크로드 단위로 나눈 실행 경계

training, inference, 에이전트 워크로드가 네임스페이스와 자원 할당량으로 나뉘어 같은 기반 위에서 공존합니다.

04 기본 구성에 들어 있는 통제와 감사

접근 통제, 이미지 출처 확인, 배포 이력 추적이 처음부터 포함되어 별도 도구를 덧붙이지 않아도 됩니다.

2.

비즈니스 임팩트

컨테이너 기반을 표준화하여 AI 플랫폼 도입 속도와 운영 안정성을 함께 높입니다.

Capsis는 Paxis, Metis, Maxis가 공통으로 필요로 하는 실행 기반을 하나로 통합합니다. 팀마다 다르게 구축된 쿠버네티스를 하나의 통제된 플랫폼으로 대체합니다.

기업은 인프라 유지에 쓰던 인력과 시간을 줄이고 같은 운영 기준을 모든 사업부와 환경에 적용할 수 있습니다.

01 AI 플랫폼 도입 기간 단축

컨테이너 실행 기반이 이미 갖춰져 있어 inference·training·에이전트 플랫폼을 올릴 때 인프라 준비 단계를 건너뛸 수 있습니다. 새 플랫폼을 추가할 때마다 기반을 다시 만들지 않습니다.

02 인프라 유지 인력 절감

팀마다 따로 만든 쿠버네티스를 유지보수할 필요가 없어 플랫폼 엔지니어링 인력을 서비스 개발로 돌릴 수 있습니다.

03 운영 표준의 일관성 확보

모든 사업부와 환경에서 같은 통제 기준, 같은 네트워크 정책, 같은 배포 절차를 적용합니다. 환경 차이에서 오는 장애가 줄어듭니다.

04 AI 인프라 활용 효율 개선

GPU·NPU 구성을 반영한 배치와 자원 할당량 관리로 유휴 자원을 줄이고 보유한 가속기가 실제 작업에 쓰이도록 합니다.

05 규제 요건을 갖춘 격리

규제 대상 워크로드를 네트워크와 자원 단위로 분리하고 모든 컨테이너 운영에 감사 이력을 남깁니다. 정부가 통제하는 환경 안에서만 배포하고 운영해야 하는 경우도 같은 플랫폼으로 충족합니다.

3.

Capsis 제품 구성



Multi-Cluster Orchestration : 모든 클러스터를 하나의 관리 화면에서 운영

여러 환경과 지역에 흩어진 쿠버네티스 클러스터를 하나의 관리 화면에서 운영합니다. 워크로드는 자원 여유, 토폴로지, 정책을 기준으로 적합한 클러스터에 배치됩니다.

통합 클러스터 목록

보유한 모든 클러스터의 상태, 버전, 자원 현황을 한 목록에서 확인합니다.

클러스터별 자원 관리

클러스터를 선택하면 워크로드, 서비스, 스토리지, 설정, 정책을 같은 화면 구성으로 관리합니다.

환경에 무관한 일관성

온프레미스와 클라우드 어디에 있는 클러스터든 같은 운영 절차와 같은 관리 화면을 적용합니다.

정책 기반 워크로드 배치

자원 가용량과 토폴로지, 운영 정책을 반영해 워크로드가 배치될 클러스터를 결정합니다.



Helm · 애드온 관리 : 플랫폼 구성요소의 생명주기를 통제

플랫폼 구성요소와 AI 워크로드가 필요로 하는 의존 패키지를 배포하고 설정하고 갱신합니다. 모든 클러스터가 갱신 주기 전체에 걸쳐 같은 구성을 유지하도록 합니다.

표준화된 설치 절차

필요한 앱과 오퍼레이터를 정해진 입력 폼을 통해 설치하고 설치 결과와 생성된 자원을 함께 확인합니다.

설치된 구성요소 관리

어느 클러스터에 무엇이 어떤 버전으로 설치되어 있는지 목록으로 확인하고 설정 변경과 삭제를 수행합니다.

클러스터 간 구성 일관성

여러 클러스터에 같은 구성요소를 같은 버전으로 유지하여 환경 차이에서 오는 장애를 줄입니다.

배포 이력 추적

설치, 갱신, 삭제 이력을 남겨 구성 변경의 원인과 시점을 확인할 수 있도록 합니다.



GPU · NPU 스케줄링 : 적합한 워크로드를 적합한 하드웨어에 배치

가속기의 구성을 이해해 컨테이너를 배치합니다. training 작업, inference 서비스, 에이전트 런타임이 각각에 맞는 하드웨어에 배치되며 운영자가 노드 지정 규칙을 직접 작성할 필요가 없습니다.

토폴로지 인지 배치

가속기의 연결 구조를 반영해 분산 training처럼 노드 간 통신이 많은 작업을 적합한 위치에 배치합니다.

워크로드 특성별 배치

training, inference, 에이전트 워크로드의 자원 요구 특성을 구분해 배치 기준을 다르게 적용합니다.

수동 설정 제거

노드 선택 규칙을 워크로드마다 직접 작성하지 않아도 적합한 하드웨어에 배치됩니다.

가속기 활용 효율

유휴 가속기를 줄이고 보유한 GPU·NPU 자원이 실제 작업에 쓰이도록 합니다.

4.

관측성·보안·운영



Network Policy : AI 환경에 맞춘 통신 제어

클러스터에 걸쳐 네트워크 정책을 정의하고 적용합니다. 외부에서 들어오는 통신, 외부로 나가는 통신, 그리고 에이전트·inference·training 워크로드 사이의 내부 통신을 각각 제어합니다.

인그레스 제어

외부에서 클러스터 안으로 들어오는 통신을 서비스 단위로 허용하고 차단합니다.

이그레스 제어

워크로드가 외부로 나가는 통신을 제한하여 데이터가 허가되지 않은 경로로 나가지 않도록 합니다.

워크로드 간 통신 제어

Paxis 에이전트 환경, Metis inference 서비스, Maxis training 워크로드 사이의 통신을 필요한 범위로만 허용합니다.

정책의 일관 적용

클러스터마다 정책을 따로 만들지 않고 같은 기준을 전 환경에 적용합니다.



Workload Isolation : 여러 플랫폼의 워크로드가 서로 간섭하지 않도록 관리

네임스페이스 격리, 자원 할당량, 테넌트를 인지하는 배치로 Paxos·Metis·Maxis의 워크로드가 자원 경합이나 상호 간섭 없이 같은 기반 위에서 공존하도록 합니다.

네임스페이스 격리

격리 단위를 쿠버네티스 네임스페이스로 두어 플랫폼과 팀별로 자원과 권한을 나눕니다.

자원 할당량 관리

네임스페이스별 CPU·메모리·GPU 사용 한도를 정해 특정 워크로드가 자원을 독점하지 않도록 합니다.

테넌트 인지 배치

워크로드가 속한 테넌트를 반영해 배치하여 서로 다른 조직의 작업이 같은 노드에서 충돌하지 않도록 합니다.

자원 사용 현황 확인

네임스페이스와 클러스터 단위의 자원 사용량을 확인하고 할당 정책을 조정합니다.

5.

Works with



Paxis · Metis · Maxis가 함께 쓰는 하나의 컨테이너 실행 기반입니다.

Capsis는 Infrastructure 계층에서 컨테이너 실행 기반을 말합니다. Metis, Maxis 플랫폼은 Capsis로 제공되는 클러스터 위에서 동작합니다.

CONTAINER PLATFORM (CAPSIS)

Kubernetes

PRIVATE CLOUD PLATFORM (AEGIS)

VM · Bare Metal

01 Aegis + Capsis

인프라 실행 기반 — 같은 인프라 계층에서 Aegis의 컴퓨트·스토리지와 함께 배포하거나 떼어내 단독으로 운영합니다.

AI PLATFORM (METIS)

ML / LLMops Agent Ops

CONTAINER PLATFORM (CAPSIS)

Kubernetes

AI TRAINING PLATFORM (MAXIS)

End-to-end training

CONTAINER PLATFORM (CAPSIS)

Kubernetes

02 Metis + Capsis

inference 실행 기반 — Metis의 모델 serving과 inference endpoint가 Capsis 클러스터 위에서 동작합니다.

03 Maxis + Capsis

training 실행 기반 — Maxis의 모델 개발·training 파이프라인이 Capsis 클러스터 위에서 동작합니다.



Paxis · Metis · Maxis가 함께 쓰는 하나의 컨테이너 실행 기반입니다.

Capsis는 Infrastructure 계층에서 컨테이너 실행 기반을 말합니다. Metis, Maxis 플랫폼은 Capsis로 제공되는 클러스터 위에서 동작합니다.

AGENT PLATFORM (PAXIS)

Thaki Cloud Application

CONTAINER PLATFORM (CAPSIS)

Kubernetes

04 Paxis + Capsis

에이전트 실행 기반 — Paxis의 Agent Builder·RAG·MCP Tool Hub가 Capsis 클러스터 위에서 동작합니다.

SIGNUM

Unified Control Plane

05 Signum

공통 파운데이션 — 접근 통제, 운영 로그·알림·감사, 보안 정책, 인프라 구축을 담당하며 모든 구성에 기본으로 포함됩니다.

CONTAINER PLATFORM (CAPSIS)

Kubernetes

06 Capsis

Capsis는 단독 판매할 수 있는 독립 상품입니다.