



AI 서비스의 여정을 위한 믿을 수 있는 우주선, Kosmos

복잡한 AI 인프라 환경을 하나의 유기적 시스템으로 통합한 인프라 우주선.
QKS, AIStor, MCP, AI Agent로 구성되어 있으며,
AI 개발자와 플랫폼이 안심하고 '탑승'할 수 있는 완전한 항해체계입니다.

About Kosmos

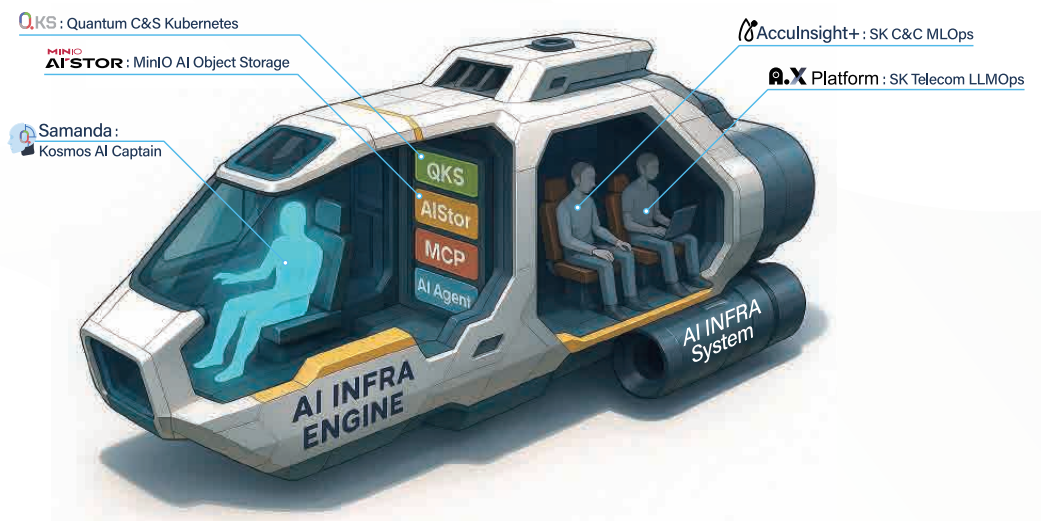


통합형 AI 인프라

우주를 향해하는 AI 인프라, Kosmos
Kosmos는 복잡한 AI 인프라 환경을 하나의 유기적 시스템으로 통합한 AI 인프라 전용 우주선입니다.
AI 선장 Samanda가 조종하는 이 우주선에는 다양한 AI 플랫폼, 예를 들어 AccuInsight+, A.X Platform 같은 승객들이 탑승합니다. 이들은 Kosmos의 핵심 시스템(QKS, AIStor, MCP, AI Agent)을 통해 목적지인 인사이트와 지능에 도달하게 됩니다.
우리는 클라우드, 데이터, GPU 자원을 정교하게 조율하여 AI 개발자와 서비스가 신뢰하고 탑승할 수 있는 완전한 항해체계를 만들었습니다.

“AI 서비스의 여정을 위한 믿을 수 있는 우주선.
Kosmos, 이것이 AI 인프라의 미래다.”

Kosmos Technical configuration



AI Applications

AccuInsight+와 A.X Platform은 Kosmos 위에서 실행되며, 복잡한 AI 서비스를 유연하게 처리

AI Infra Engine

- QKS: Kubernetes 및 백오피스 시스템 자동 설치 및 운영
- AIStor: 대용량 학습 데이터를 위한 MinIO 기반 오브젝트 스토리지
- MCP: 다양한 LLM을 제어 및 서빙하는 Model Context Protocol
- AI Agent: 자연어 기반 파이프라인 자동화 및 제어 시스템

AI Infra System

Kosmos는 고성능 인프라 기술로 추진

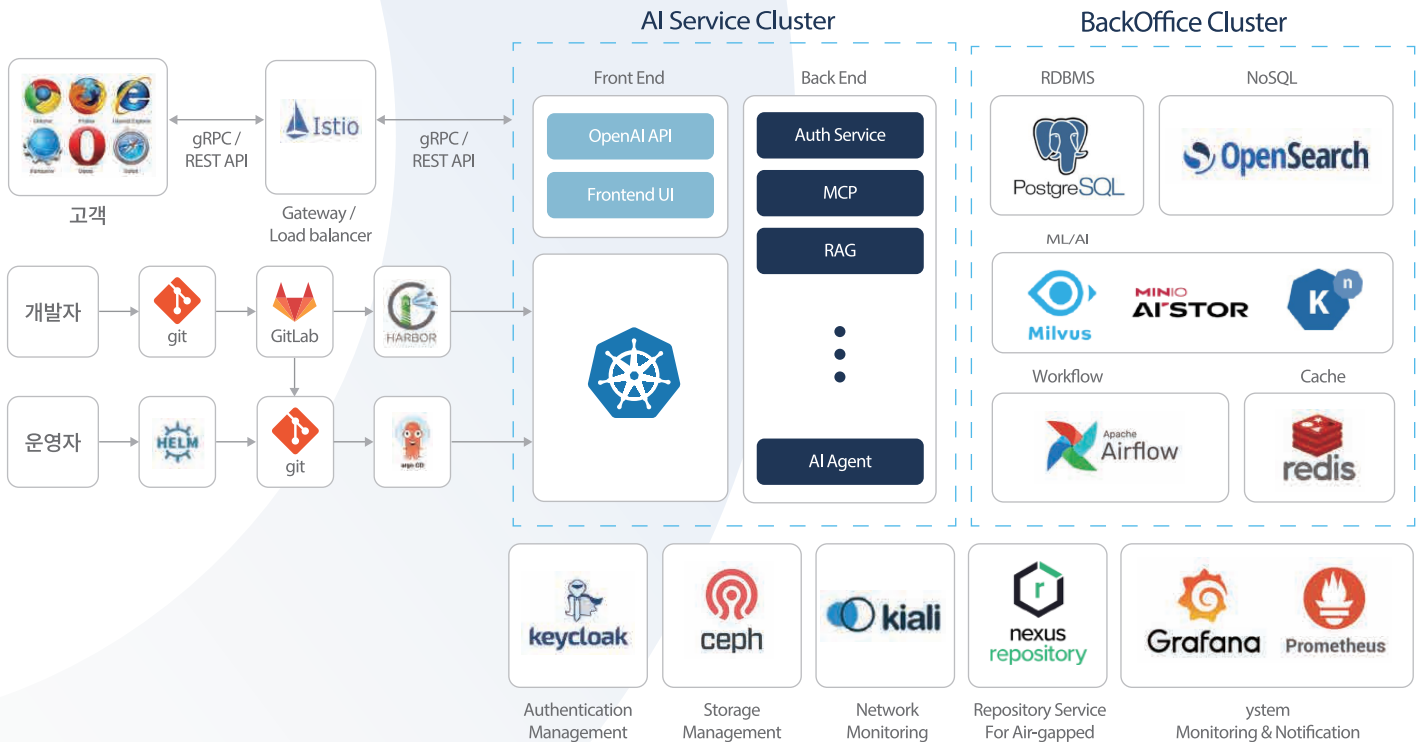
- GPU Acceleration: Mixed Precision, 병렬 연산
- High-Speed Networking: RDMA, GPUDirect, NVLink, RoCE, SR-IOV, eBPF
- High-Performance Storage: GDS, DirectPV, S3 over RDMA 등 고속 I/O 기술



AI 인프라 자동화의 시작

QKS는 Quantum C&S가 자체 개발한 Kubernetes 설치 자동화 틀체인입니다.

- Bare-metal 환경에서 자동 OS 설치 및 설정
- 네트워크 및 스토리지 구성 자동화
- Kubernetes 클러스터 구성 및 Helm 기반 앱 설치
- GPU 스케줄링, 모니터링, 인증 설정까지 원클릭 자동화



AI 워크로드 중심의 쿠버네티스 인프라

AI 연산에 최적화된 GPU 노드 관리, RDMA, GDR, NUMA 설정, MIG 지원 등 고성능 환경제공주거기적인 K8s 및 백오피스 업그레이드 지원긴급 보안 패치 서비스

통합된 CI/CD 파이프라인

GitLab, ArgoCD, Harbor, Helm 을 기반으로 코드부터 클러스터 배포까지 완전 자동화된 파이프라인 구성

인증 및 접근 제어 자동화

Keycloak 기반의 OIDC 연동으로 인증, 권한, 멀티클러스터, 멀티 테넌시까지 통합관리

통합 모니터링 및 로깅 환경

OpenTelemetry, Prometheus, Grafana, OpenSearch 기반으로 시스템 성능 및 이벤트 흐름 실시간 분석

Air-gapped 환경을 위한레지스트리 프록시 서비스

Nexus, Harbor를 통한 내부전용 패키지 및 이미지 저장소 제공으로 보안 강화

스토리지 및 데이터 관리 통합

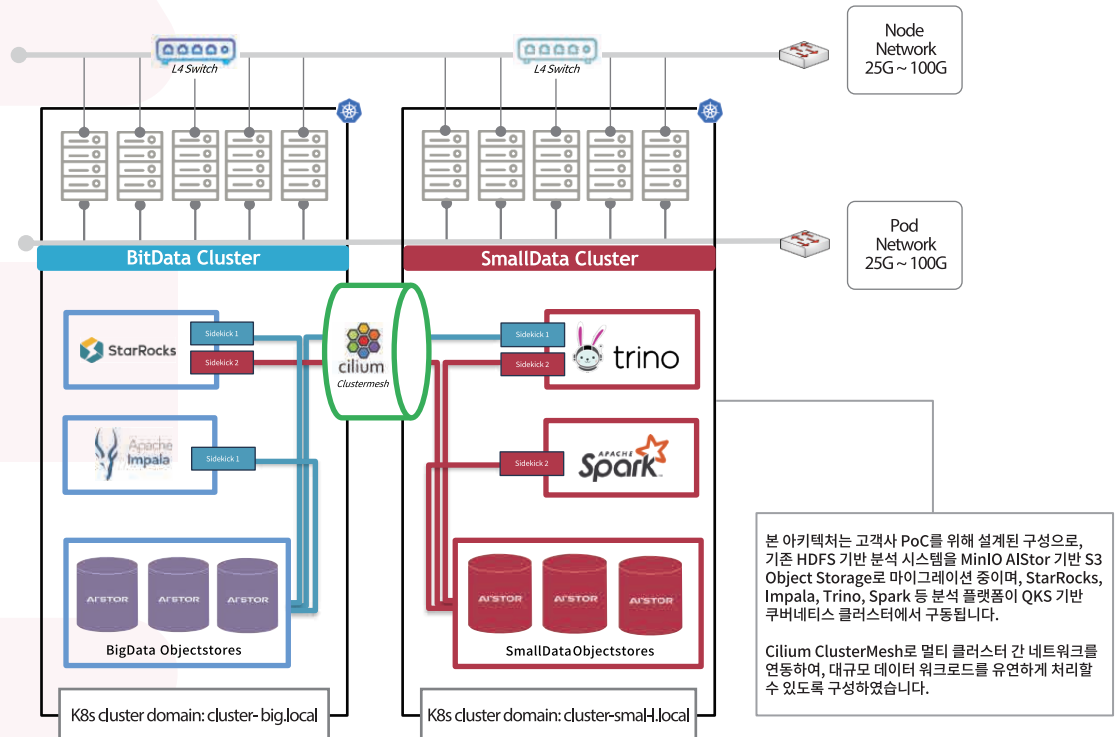
Ceph 기반의 퍼시스턴트 스토리지와 MinIO AI*Stor 연동으로 AI특화 대용량 데이터 처리 최적화



고속, 대용량, AI 특화 오브젝트 저장소

MinIO 기반으로 설계된 AIStor는 AI 학습/추론/검색 데이터를 위한 고성능 오브젝트 스토리지입니다.

- GDS, RDMA 기반 S3-over-NVMe 성능
- DirectPV로 Kubernetes에 Native 통합
- 대용량 학습 데이터셋에 최적화된 고가용성 설계
- 고속 네트워크(RoCE, NVLink)와의 연계로 AI 파이프라인 속도 극대화



하둡 대체 가능한 고성능 S3 스토리지

HDFS 대비 구축과 운영이 간편하고, AI/ML 분석에 최적화된 S3 오브젝트 기반의 고속 스토리지 제공

클라우드 네이티브 완전 대응

쿠버네티스 환경에 최적화된 DirectPV 기반 운영, StatefulSet 및 Helm 배포 지원

멀티 클러스터 · 멀티 리전 확장

클러스터 간 동기화, 글로벌 Namespace 구성으로 이기종 환경에서도 원활한 데이터 연동 지원

S3 over RDMA로 모델 업로드 속도 극대화

MinIO는 RDMA 전송을 지원하여, AI 모델 및 데이터셋을 최대 수백 GiB/s 속도로 업로드할 수 있는 초고속 환경을 제공합니다.

AIHub 및 MCP 연동으로 AI 워크로드 자동화

모델 학습/추론용 데이터를 AI-Hub에서 관리하고, MCP와의 통합을 통해 GPU 리소스 및 데이터 흐름을 자동 제어하여 AI 워크로드를 최적화합니다.

탁월한 성능

단일노드: 220 GiB/s Throughput
- 1,000,000 IOPS (25G NIC + RDMA + NVMe SSD / MinIO warp 기준)
멀티노드: 1.6 TiB/s Throughput (16노드)
- 10,000,000 IOPS 이상 (100G NIC + 멀티노드 / MinIO 공식 블로그 기준)

About Quantum C&S

AI가 신뢰하는 인프라, Quantum C&S

Cloud Native와 DevOps에서 출발한 Quantum C&S는, 이제 AI 시대에 맞는 새로운 기준, ‘AI Native’ 인프라를 만듭니다. 복잡한 AI 서비스가 단순하게 작동하고, 개발자와 운영자가 더 빠르게 움직일 수 있도록 우리는 AIOps로 제어되는 지능형 인프라를 설계합니다. AI가 잘 돌아가는 시스템, 그것이 우리의 목표입니다.

History of Quantum C&S



Our Partners



Our Clients





Kubernetes Certified Service Provider



Proprietary AI Infrastructure



Official MinIO Partner

Quantum C&S 는 Kubernetes 공식 인증 기업(KCSP) 이자 MinIO의 전략적 파트너이며, 자체 플랫폼 Kosmos를 통해 완성도 높은 AI 인프라를 제공합니다.



Q.QUANTUM C&S

우리는 기술로 신뢰를 증명합니다.

서울특별시 구로구 디지털로31길 62 에이스아티스포럼 707호
business@quantumcns.ai
02.6951.5151