Hammerspace for Artificial Intelligence

데 이 터 시 트

모든 산업은 이제 방대한 데이터 처리의 부상으로 인해 새로운 데이터 사이클로 진입하고 있으며, 이는 다양한 기업에서 혁신을 촉진하는데 필수적입니다. 기업들은 AI를 활용해 신규 매출 창출과 운영 효율성을 향상하기 위해 데이터 아키텍처를 신속히 재검토하고 있습니다. AI 아키텍트들은 미래를 위한 유연성과 현재 필요한 핵심 기반을 동시에 제공할 수 있는 솔루션을 찾고 있습니다.

조직이 직면한 가장 큰 과제 중 하나는 분산된 비정형 데이터 세트를 AI 전략에 효과적으로 활용하는 동시에 기존 엔터프라이즈 솔루션에서는 제공하기 어려운 성능과 확장성을 제공하는 것입니다.

Hammerspace는 병렬 파일 시스템과 자동화된 데이터 오케스트레이션을 기반으로 단일 글로벌 데이터 플랫폼을 구축하여 조직을 지원합니다. 또한 엔터프라이즈 표준 데이터 서비스를 통합하여 거버넌스 및 데이터 보호 목표를 충족합니다. 기존 스토리지 사일로에서 데이터 세트를 불러오거나 향후 버튼 클릭만으로 새로운 데이터 세트 또는 데이터 소스를 추가할 수 있는 유연성을 제공합니다. AI 모델이 데이터 근처에 있든, 원격 클라우드나 SaaS 도구에 있든 관계없이 Hammerspace는 필요한 시점에 데이터를 손쉽게 접근 · 분석 · 처리 · 이동할 수 있도록 합니다.

Hammerspace는 3,000 노드, 100 Tbit/sec 성능, 사이트당 24,000 GPU 환경에서 세계 최대 규모 LLM 중 일부의 학습을 지원합니다.

Hammerspace가 제공하는 AI 차별화 요소

1. High-Performance (고성능)

대부분의 AI 프로젝트는 의사결정 및 비용 관리 측면에서 비즈니스 핵심으로 자리 잡을 것입니다. 데이터 파이프라인은 사용 가능한 모든 컴퓨팅 파워를 활용하고 Databricks 및 Snowflake와 같은 클라우드 기반 모델에 데이터를 제공할 수 있도록 설계되어야 합니다.

Hammerspace의 고성능 데이터 파이프라인은 단일 클러스터에서 6만 개 이상의 GPU를 보유한 환경 등 세계에서 가장 빠른 컴퓨팅 클러스터를 지원합니다. 데이터 로딩 시 거의 라인 속도(line rate)로 스트리밍할 수 있으며, 체크포인트 시 빠른 결과를 반환하고 애플리케이션이 요구하는 속도로 동시 읽기 요청을 처리합니다.

© 2023 - Hammerspace | www.HAMMERSPACE.com

2. Multiple Data Sources (다양한 데이터 소스)

AI 모델이 단일 스토리지 사일로의 데이터만 사용할 경우, 여러 스토리지 사일로 및 다양한 사이트에 저장된 폭넓은 데이터 세트에 접근할 수 있는 모델에 비해 매우 불리합니다.

AI 데이터 파이프라인은 엣지 디바이스, 데이터센터, 클라우드에서 생성·저장된 데이터를 기반으로 구성되어야 합니다. 이 모든 데이터에 효과적으로 접근하기 위해 Hammerspace는 사일로화된 스토리지들을 통합하고 여러 사이트에 분산 저장된 데이터를 오케스트레이션합니다. 이를 통해 데이터가 어디에 있든, AI모델이 어디에서 실행되든 필요로 하는 데이터의 배치 · 제공 · 저장이 가능해집니다.

3. Burst to Remote GPU Clusters (원격 GPU 클러스터로 버스트)

GPU 자원은 성공적인 AI 아키텍처의 핵심 요소입니다. 최대 용량을 지속적으로 유지하기 위한 과도한 오버헤드 없이, 필요 시 사용할 수 있도록 여러 GPU 클러스터를 활용하는 것이 바람직합니다. AI 분야는 워크로드와 요구사항의 변동성이 크기 때문에 이러한 적응성은 더욱 필수적입니다.

4. Massive Scale (대규모)

AI 모델은 최대한 정확성을 확보하기 위해 대량의 데이터로 학습되어야 합니다. 이용 가능한 데이터가 많을수록 초기부터 더 정확한 결과를 얻을 수 있습니다. Hammerspace는 기존 데이터 세트, 클라우드 인스턴스 및 모든 신규 인프라를 통합하여 AI 워크로드의 발전에 따라 확장 가능한 통합 글로벌 데이터 환경을 구축합니다.

5. Enterprise Standard (엔터프라이즈 표준)

엔터프라이즈 표준 인터페이스와 데이터 서비스는 AI 시스템의 효율성 · 확장성 · 신뢰성을 확보하는데 핵심적인 역할을 하며 컴플라이언스, 거버넌스, 효율성 목표와도 일치합니다. Hammerspace는 자동화된 데이터 서비스로 데이터 접근 · 보호 프로세스를 관리해 프라이버시, 보안, 적절한 데이터 사용을 보장합니다.

또한 모델과 데이터의 통합 속도를 높이기 위해 업계 표준 NFS를 데이터 인터페이스로 제공하여 어플리케이션이 오브젝트 기반으로 재작성할 필요 없이 손쉽게 데이터를 연결할 수 있습니다. NFSv4.2를 활용하면 하이퍼스케일 환경에서 성능을 무제한 확장할 수 있습니다.

6. Data Governance (데이터 거버넌스)

데이터 거버넌스는 AI에서 매우 중요한 역할을 하며, 데이터 소스가 여러 사일로와 사이트에 분산될수록 더욱 복잡해집니다.

Hammerspace는 여러 데이터 소스를 단일 네임스페이스 및 메타데이터 제어 플레인으로 통합하여 데이터 품질을 보장합니다. 이를 통해 편향을 줄이고 데이터의 정확성 · 완전성 · 신뢰성 · 적시성을 확보하여 AI 모델 예측의 정확성을 높이고 의사결정을 위한 데이터의 신뢰성을 보장합니다.

병렬 파일 시스템의 성능과 확장성 — 어디서 생성되든 모든 데이터로 구동

Hammerspace는 데이터를 특정 스토리지 시스템이나 특정 클라우드 서비스에 국한되거나 갇혀 있지 않은, 살아있는 글로벌 공유 자원으로 만듭니다.

AI 시대에는 데이터의 양과 활용 방식이 기하급수적으로 증가할 것입니다. 전 세계적으로 실시간 오케스트레이션되는 데이터에 대한 고성능 로컬 읽기/쓰기 접근은 필수적이고 보편화될 것입니다.

Hammerspace를 사용하면 데이터와 AI 모델이 어디에 위치하든 상관없이 모든 데이터를 손쉽게 즉시 활용할 수 있습니다.

