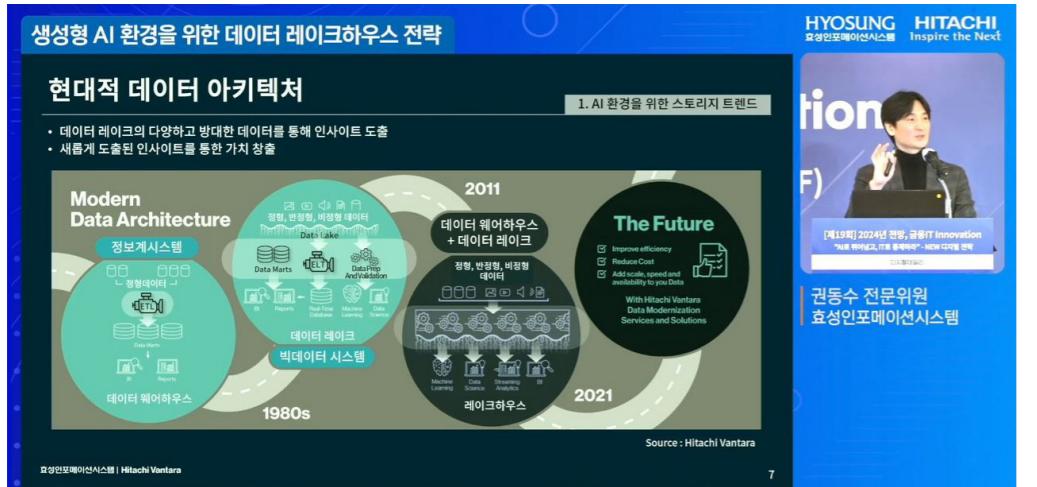


Part I

2024년 전망, 금융 IT Innovation 버추얼 컨퍼런스 생성형 AI 환경을 위한 데이터 레이크하우스 전략



지난해 말, 효성인포메이션시스템은 ‘2024년 전망, 금융 IT Innovation 버추얼 컨퍼런스’에서 ‘생성형 AI 환경을 위한 데이터 레이크하우스 전략’을 주제로 발표했다. 컨퍼런스에서는 AI 환경을 위한 스토리지 트렌드, AI를 위한 최적의 스토리지 및 AI 활용을 위한 데이터레이크 사례에 대해 소개했다.

다크 데이터도 문제 없는 데이터 레이크하우스

기업들이 업무에 AI를 접목하기 시작하면서, 데이터 양이 급속히 증가하는 한편, 다크 데이터의 양 역시 이에 비례하며 증가하고 있다. 다크 데이터는 다양한 컴퓨터 네트워크 운영을 통해 얻어지는 데이터이지만, 의사결정이나 이해를 위한 수단으로 사용되지 않는 불완전한 데이터를 말한다.

이전까지는 다크 데이터를 분석 및 처리할 수 있는 도구가 없었지만, 소프트웨어와 하드웨어 기술이 발전하면서 대량의 데이터를 분석 및 활용할 수 있는 시대가 열렸다.

데이터레이크는 ‘다양한 형태의 원형(raw) 데이터들을 모은 저장소의 집합’을 의미한다. 스토리지에 데이터 노드를 저장하는 전통적인 방식에서 탈피해 다양한 데이터를 저장할 수 있는 오브젝트 스토리지를 기반으로 진정한 데이터레이크가 구축되었으며, 현재는 오브젝트 스토리지와

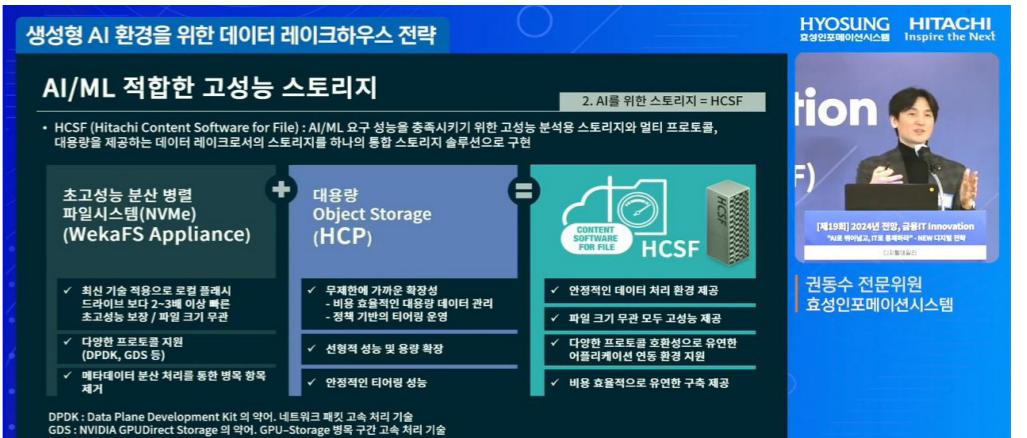
초고성능 병렬파일시스템 기반의 데이터레이크 즉 데이터 레이크하우스 시대에 접어들었다.

데이터 레이크하우스는 데이터레이크가 가지고 있는 유연성, 비용 효율성, 대용량 지원 기능과 데이터 웨어하우스의 데이터 관리 기능과 ACID¹⁾ 트랜잭션을 통합한 새로운 형태의 오픈 데이터 관리 아키텍처이다.

기업들은 데이터레이크의 다양하고 방대한 데이터를 통해 인사이트를 도출할 수 있게 되었으며, 이를 기반으로 비즈니스 성장을 위한 가치 창출 역시 가능해졌다.

AI/ML에 적합한 솔루션 구현의 핵심

효성인포메이션시스템은 초고성능 병렬 파일시스템과 오브젝트 스토리지 기반의 HCSF를 통해 완벽한 데이터 레이크하우스 구현을 지원한다.



기업들이 데이터 레이크하우스를 구축하기 원한다면, 자사의 환경과 요구사항을 사전에 점검하고 전문 벤더의 PoC 및 BMT를 통해 솔루션의 성능을 필수적으로 점검해야 한다. 특히 대용량의 파일을 처리하고 안정적으로 확장하기 위해서는 데이터레이크 구축 경험이 충분한 벤더와 협업해야 한다.

효성인포메이션시스템은 풍부한 경험과 기술력을 기반으로 데이터 레이크하우스의 구축과 운영, 확장 및 증설에 대한 필요 사항을 지원한다. 특히 완벽한 국내 기술지원 체계를 마련해 장애 지원 및 신규 AI 솔루션까지 연계할 수 있도록 다양한 로드맵을 제공하고 있다.

1) ACID : 데이터베이스 트랜잭션이 안전하게 수행된다는 것을 보장하기 위한 성질을 가리키는 약어로 원자성(Atomicity), 일관성(Consistency), 고립성(Isolation) 그리고 지속성(Durability)을 나타냄.

Part II

CIO Summit 2024

스스로 탐구하는 AGI와 GPU 인프라 이야기



2월 22일, 효성인포메이션시스템은 전자신문 CIO Summit 2024에서 ‘스스로 탐구하는 AGI와 GPU인프라 이야기’를 주제로 AI를 넘어선 AGI 시대에 더욱 중요해진 GPU 시스템, 그리고 AI 업무에 적합한 고속 데이터 처리를 위한 인프라에 대해 소개했다.

AGI 시대, 최적의 인프라 도입 요건

생성형 AI로 인해 기업들의 AI 비즈니스가 활발해지고 있으며, 거대언어 모델(LLM) 개발 열풍까지 불고 있다. 그리고 AI를 넘어 AGI가 등장하며 AI 비즈니스에도 변화가 있을 것으로 예상되고 있다.

AGI는 인간과 유사한 지능과 스스로 학습할 수 있는 능력을 갖춘 AI로, 학습이 가속화될수록 더욱 많은 데이터가 생성될 것으로 예상된다. AGI의 등장은 더 큰 데이터 세트와 복잡한 AI 모델이 필요함을 의미하며, GPU시스템 및 데이터 처리 효율이 매우 중요해질 수밖에 없다는 것을 시사한다.

기업들이 자사에 최적화된 AI 인프라를 도입하기 위해서는 어떤 요건이 필요할까?

첫째, 최신 기술이 적용된 GPU 인프라를 사용해야 한다. 모든 기업이 고비용에 최신 성능의 GPU를 도입할 수는 없다. 효성인포메이션시스템은 글로벌 전문 파트너인 Supermicro와 함께 기업별 환경에 적합한 GPU를 제안 및 공급하고 있다.

둘째, GPU 시스템 사용 효율을 고려해야 한다. 컨테이너 환경에서 GPU 자원의 분할 가상화 솔루션을 도입해야 하며, 데이터 과학자가 AI 개발 업무에 집중하도록 손쉬운 AI 개발 환경을 제공해야 한다.

마지막으로 AI 최적화 시스템을 설계해야 한다. GPU의 고성능을 최대로 발휘하기 위해 서버와 네트워크단의 최적의 설계가 전제되어야 한다. 또한 업무 단위로 적절한 GPU 인프라를 효율적으로 배치해주는 컨설팅 및 기술지원이 뒷받침되어야 한다.

AI 데이터 처리를 위한 완벽한 솔루션

효성인포메이션시스템은 AI 데이터의 고속 처리를 위해 초고성능 병렬 파일 스토리지 HCSF를 제공하고 있다.

AI 업무는 랜덤 입출력과 빈번한 Small I/O, 메타데이터 오버헤드가 발생하기 때문에 IOPS 성능이 매우 중요하다. HCSF는 타사 대비 7.5~225배의 높은 IOPS 성능을 나타내, AI 업무를 위한 저장장치로서 완벽한 성능을 증명하고 있다. 또한, HCSF는 단일 통합 파일시스템으로 다양한 멀티 프로토콜을 지원해 AI 업무 모든 단계의 애플리케이션 용도에 맞춰 저장자원으로 활용된다.

HCSF는 증설 및 확장할수록 성능이 선형적으로 증가하는 AI 데이터 인프라이다. 메타데이터를 처리하는 시스템이 노드별로 분산되어 있기 때문에 데이터 처리에 있어 병목이 발생하지 않는다. 따라서 확장 시 성능저하 이슈가 전혀 없는 솔루션으로, AI 인프라가 충족해야 할 모든 요구 사항을 갖춘 고성능 솔루션이다.

효성인포메이션시스템은 고성능 GPU 서버 및 HPC 서버를 기반으로 기업들이 AI 업무를 위한 연산 인프라를 구축할 수 있도록 지원하며, 고성능 데이터 처리에 탁월한 AI 스토리지를 제공한다. 특히 AI/ML Ops, GPU DB 등 전문 파트너와 협업을 통해 기업 환경에 최적화된 AI 솔루션을 제공한다.

Part III

효성인포메이션시스템 & 에이플랫폼(SingleStore)

최적의 데이터 레이크하우스 구현을 원한다면!

실시간 데이터 분석과 고성능 데이터 처리의 중요성이 더해지고 있는 이때, AI 시대에 걸맞은 데이터 레이크하우스 전략을 위해 효성인포메이션시스템과 에이플랫폼이 힘을 합했다.

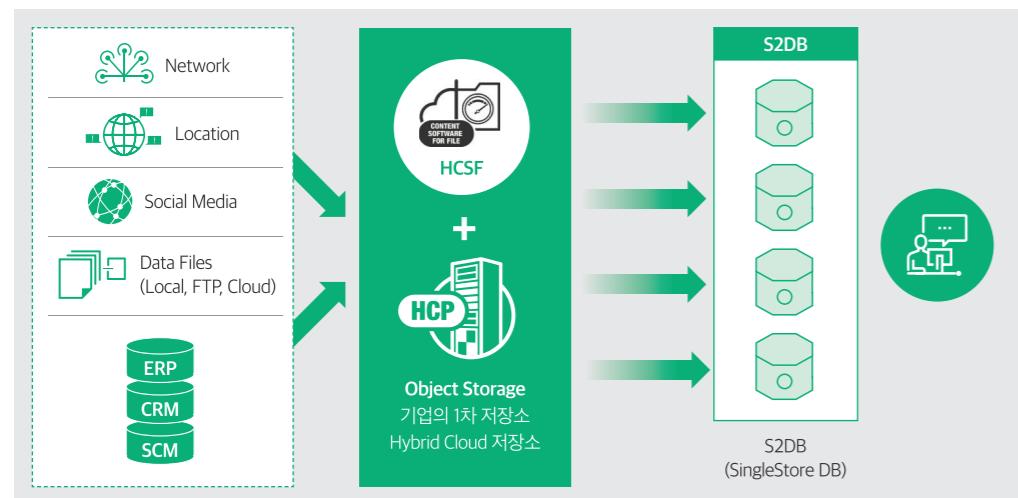
최고성능 병렬 파일시스템과 클라우드 네이티브 데이터베이스의 만남

기업들은 대용량의 다양한 데이터를 실시간으로 수집 및 분석하고 효율적으로 관리할 수 있는 데이터레이크 솔루션을 구축하고자 한다. 그리고, 다양한 장치에서 발생하는 방대한 양의 데이터를 저렴한 비용으로 구축할 수 있는 오브젝트 스토리지가 각광받고 있다.

오브젝트 스토리지가 통합 구성된 효성인포메이션시스템의 HCSF 기반 데이터레이크는 HPC 환경의 분석을 위한 방대한 양의 데이터를 비용효율적으로 저장할 수 있다.

최근, 효성인포메이션시스템은 실시간 데이터 처리와 대용량 DB의 안정적인 저장을 지원하고 생성형 AI 워크로드를 위한 고성능 데이터 처리와 최적의 데이터 레이크하우스 전략을 제시하기 위해 에이플랫폼과 손을 잡았다.

↓ 데이터 웨어하우스 + 데이터레이크 = 데이터 레이크하우스



에이플랫폼은 실시간 빅데이터 솔루션 전문기업으로, 싱글스토어DB의 국내 총판이다. 싱글스토어DB(SingleStoreDB)는 ANSI SQL 표준기반으로 인메모리와 디스크를 활용해 높은 동시성과 빠른 성능으로 모든 형태의 데이터 처리가 가능한 클라우드 네이티브 데이터베이스이다. 싱글스토어DB는 2017년부터 벡터(VECTOR) 기능을 제공해 왔으며, 최근 ChatGPT로 부각된 생성형 AI 또는 LLM 서비스를 구축 및 도입하려는 기업들의 벡터DB로 도입 검토가 확대되고 있다.

AI 워크로드를 위한 최고의 성능과 인프라 환경 제공

기업들은 데이터 수집부터 분석, 관리, 적용까지 데이터를 실시간으로 활용할 수 있는 시스템을 구축하고, 이를 통해 인사이트를 도출하고자 한다.

이때 데이터 웨어하우스(싱글스토어DB)와 데이터 레이크(HCSF)를 연계하면 최적의 데이터 레이크하우스를 구현할 수 있다.

싱글스토어DB는 인메모리(Rowstore)와 디스크(Columnstore)를 활용한 실시간 분산병렬처리 지원을 통해 실시간 데이터 레이크하우스를 구축한다. 또한 단일 데이터 플랫폼에서 표준 SQL 기반으로 다양한 워크로드를 실시간 처리할 수 있으며, NVMe 기반의 최고성능 병렬 파일시스템과 오브젝트 스토리지가 결합된 HCSF를 통해 다양한 데이터 타입 저장과 실시간 데이터 수집 및 고속 처리 등 유연한 데이터 활용이 가능하다.

효성인포메이션시스템은 기존 고객을 대상으로 데이터레이크 시장을 공략하고, HCSF와 싱글스토어DB를 연계해 온프레미스 데이터 레이크하우스 신규 시장을 공략할 계획이다.

특히 최근 국내 주요 금융기관의 데이터레이크 구축 사업을 수주하며 시너지 효과를 입증한 양사는, 앞으로 실시간 데이터 분석 환경과 대용량 DB의 안정적이고 비용효율적인 저장이 필요한 AI 워크로드를 중심으로, 고객들의 AI 혁신을 위해 적극적인 활동을 펼칠 계획이다.

