

## 초고성능 데이터 서비스 플랫폼 ‘HCSF’, 협업의 활주로 타고 비상하다

히타치 벤틀라의 글로벌 기술 협력 파트너인 데이터 플랫폼 전문기업 웨카IO가 올해 한국 지사를 설립하고 국내 영업에 박차를 가하고 있다. 이에 효성인포메이션시스템의 데이터 솔루션 비즈니스도 더욱 탄력을 받고 있다. 효성인포메이션시스템은 웨카IO와 협력해 최신 GPU 워크로드, 고성능 데이터 분석, AI/ML 등을 지원함으로써 기업의 데이터 비즈니스의 성공을 위한 초석인 인프라 환경을 제공한다.

최근에는 데이터 기반의 경영을 지원하기 위해 GPU DBMS 전문기업 스크림(SQream)과의 협업도 강화했다. 이를 통해 대용량 데이터의 신속한 분석과 엔터프라이즈 AI/ML 혁신까지 지원한다는 게 목표다. 그리고 이 모든 협업의 중심에는 효성인포메이션시스템의 HCSF 솔루션이 있다.

### 데이터 레이크를 위한 완벽한 시작

데이터가 기업의 핵심 IT 자산으로 등극하면서, 데이터의 효과적인 저장과 관리 방안에 대한 관심이 높아지고 있다. 특히 데이터 분석과 AI 활용 등을 위한 핵심 자원으로 비정형 데이터의 중요성이 커지면서, 데이터 관련 솔루션에 대한 니즈 역시 증가하고 있다.

기업들이 데이터 관리를 하는 목적은 여러 가지다. 구체적으로, 새로운 비즈니스 모델 적용, 시장 경쟁력 향상, 상품 또는 서비스 출시 기간 단축, 고객 경험 향상, 기업의 비용 절감 등을 위해서다. 이러한 목표를 달성하기 위해서는 데이터 관리 워크플로우에 적합한 컴퓨팅, 스토리지, 네트워킹 구성요소가 필요하다.

고성능 스토리지 시장 규모는 2020년 이후로 매년 확대되고 있다. 특히 제조 및 통신 산업을 중심으로 모든 유형의 데이터를 한 곳에 통합, 저장, 분석할 수 있는 솔루션에 대한 요구가 커지면서 앞으로 데이터 레이크 시장이 더 확대될 것으로 전망된다.

올해 IT 산업의 대표적 키워드 중 하나인 ‘데이터 레이크’는 기업 데이터의 중요한 혁신 수단으로 손꼽힌다. 이는 데이터 레이크가 데이터 사일로(Silo) 현상을 방지하고, 데이터 분석을 위한 모든 종류의 데이터를 취합하는 데 데이터 중앙 공급소와 같은 탁월한 역할을 하기 때문이다.

그러나 진정한 데이터 레이크를 구현하기 위해서는 단순한 대규모 저장소 뿐 아니라 고성능 데이터 처리가 반드시 필요하다. 이 때문에 기존의 데이터 레이크 전략을 개선하면서도 비용 효율적으로 구현할 수 있는 솔루션이 주목받고 있다.

### 진정한 데이터 레이크 솔루션

데이터가 폭증하면서 분석해야 할 데이터양 역시 기하급수적으로 늘고 있다. NVMe 기반의 병렬 처리 시스템은 수십 기가바이트(GB) 이상의 데이터 처리를 위해 분석 환경을 최적화하고 폭증한 데이터를 처리하는 데는 효과적이지만, 큰 걸림돌이 하나 있다. NVMe 스토리지는 PB(페타바이트) 이상을 저장하기에는 비용 부담이 큰, 고가의 솔루션이라는 점이다.

또한, 대부분의 데이터가 초기에는 많은 IO를 요구하지만, 시간이 지날수록 액세스 빈도가 급격히 낮아진다. 따라서 액세스 빈도가 낮아지는 데이터를 애플리케이션 성능에 영향을 주지 않고 티어링을 통해 더 저렴한 스토리지로 이동시킨다면 비용을 꽤 절감할 수 있다.

지난해 출시한 효성인포메이션시스템의 HCSF는 초고성능 병렬 파일시스템과 오브젝트 스토리지가 통합된, 초고성능과 대용량 확장성을 모두 제공하는 파일 스토리지다. 오브젝트 스토리지의 가용량과 클라우드 기능을 갖추면서도 분산 파일 시스템 속도를 제공하며, 파일과 오브젝트 프로토콜을 지원해 데이터 수집이 용이하다. 또한 정책 기반 티어링 파일 시스템 운영으로 비용 효율적인 데이터 관리가 가능하다.

Why HCSF?				
<p>✓</p> <p>슈퍼컴퓨터, HPC를 위한 상용 파일시스템 중 가장 빠름</p>	<p>✓</p> <p>AI/ML처럼 IO 집약적인 워크로드에 적합</p>	<p>✓</p> <p>파일 크기에 따른 성능 저하 없음</p>	<p>✓</p> <p>데이터 카피 불필요 (진정한 데이터 레이크 솔루션)</p>	<p>✓</p> <p>혼합 워크로드에서 초고성능 분석 환경을 위한 스토리지 통합 솔루션</p>

### HPC 환경을 위한 고성능 솔루션

HCSF는 웨카IO의 WekaFS를 기반으로 NVMe 전용 병렬 파일시스템을 통한 고성능 IO 처리가 가능하며, 오브젝트 스토리지인 HCP와 결합해 워م/콜드 데이터를 티어링하고 파일 데이터를 무제한급으로 저장할 수 있다. HCP에 보관된 데이터는 퍼블릭 클라우드의 S3 호환 오브젝트 스토리지와 연동해 데이터 저장 영역을 타 클라우드로도 확장할 수 있다.

파일이 티어링 되더라도 백엔드에서만 저장 위치가 바뀌기 때문에 스토리지를 사용하는 분석 애플리케이션에서는 이를 신경 쓸 필요가 없다. 만약 분석 애플리케이션이 워م/콜드 티어인 HCP의 데이터를 액세스하고자 하면, 해당 데이터는 다시 핫 티어인 NVMe 스토리지로 즉시 이동해 서비스된다. 이 모든 과정은 사용자 입장에서 별도 작업 없이 자동으로 지원되어 빅데이터를 효과적으로 관리할 수 있다.

웨카IO의 WekaFS는 전 세계 IT 및 기술 분야 전문가들이 제품을 검토·평가하는 전문 플랫폼 ‘Gartner Peer Insight’에서 경쟁제품 대비 높은 평가를 받아 전문성을 인정받았다. 웨카IO는 클라우드, AI를 위한 최고의 데이터 플랫폼을 기반으로 기업들이 데이터를 중심으로 한 혁신 경쟁력을 높일 수 있도록 효성인포메이션시스템과 협업을 강화하고 있다.

HCSF는 초고성능이면서, 동시에 비용 효율적인 대용량 데이터 처리가 필요한 HPC(고성능 컴퓨팅) 환경에 매우 적합한 솔루션이다.

현재 HCSF는 슈퍼컴퓨터 및 HPC를 위한 상용 파일시스템 중 가장 빠른 솔루션이다. 특히 높은 처리량, 높은 IOPS, 매우 짧은 대기 시간이 동시에 필요한 혼합 워크로드에 강하기 때문에 EB(엑사바이트) 규모의 확장 및 성능 향상을 제공할 수 있다.

또한 HCSF는 국내외 기술지원 프로세스가 완벽하게 수립되어 있어, 다양한 구축 사례와 경험을 기반으로 원활한 운영을 지원한다. 특히 장애 발생 시에는 국내 전문가와 해외 전문가의 지원을 동시에 받을 수 있어 최단기간 안에 장애 조치가 가능하다.

↓ HCSF 도입 효과



### 다양한 레퍼런스 확보하며 고공행진

HCSF는 출시 이후 성과와 확장성, 관리 용이성 등에서 두각을 나타내며 국내외에서 구축 사례가 눈에 띄게 증가하고 있다.

먼저, 해외의 제조기업 A사는 AI 모델 트레이닝을 위해 수천만 개의 다양한 크기의 파일을 처리할 수 있는 솔루션이 필요했다. 기존 올플래시 NAS와 GPU 서버의 로컬 NVMe로는 AI 트레이닝 시스템의 요구 성능을 충족할 수 없었기 때문이다.

A사는 HCSF를 도입했고, 이를 통해 기존의 올플래시 NAS 대비 10배, 로컬 NVMe 대비 3배 이상의 성능 개선 효과를 거두었다. 그뿐만 아니라 비용 절감 및 GPU 활용도 역시 기존과 대비해 3배 이상 향상되는 효과를 얻을 수 있었다.

국내의 제조기업 B사는 데이터 인사이트 확보를 위해 실시간 대용량 데이터 분석 시스템을 구축하고자 했다. 구체적인 목표로는 데이터 운영 인프라 개선, 데이터 분석/활용 고도화, 거버넌스 도입 등 세 가지였다.

B사는 그간 자사 내외에 데이터 운영 시스템과 저장소가 산재해 있었고, 고비용 및 단단계 구조의 레거시 인프라를 보유한 환경이었다. 특히 수집된 데이터를 제대로 활용하지 못했고, 데이터 관리 주체나 체계가 결여된 상태였다. 데이터 품질 관리나 지원이 미흡하다는 점도 문제였다.

B사는 이와 같은 문제점을 해결하고, AI/ML(머신러닝) 기반의 분석 기법을 도입하기 위해 효성인포메이션시스템의 HCSF를 선택했다. B사가 여러 벤더의 솔루션을 비교 검증한 후 HCSF를 선택한 이유는 다음과 같다.

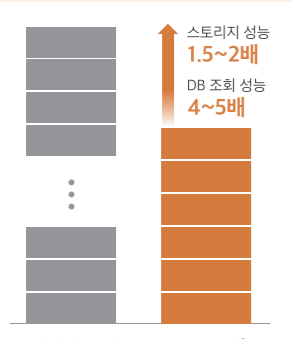
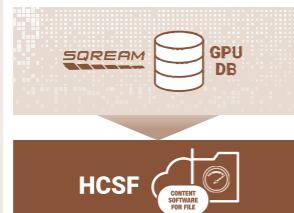
먼저, 경쟁사 대비 절반의 노드 수 구성에도 불구하고 3배 이상의 높은 성과와 데이터 증가에 대응할 수 있는 확장성을 보장했다. 두 번째로, 오브젝트 스토리지와 결합해 대용량 데이터 운영의 비용 절감 효과와 타사 대비 최대 5배 이상의 높은 스토리지 티어링 성능을 동시에 만족시켰다. 세 번째는, HPC 환경에 최적화된 대량 데이터 분석과 AI 학습 환경에 적합한 파일시스템 환경 제공, 엔비디아의 GPUDirect 및 DGX 호환성 등을 보장하는 최신 트렌드를 반영한 스토리지 아키텍처였다는 점이다. 마지막으로, 분석용 데이터베이스와 AI 분석 솔루션에서 취급하는 파일 규모가 수백억 개에서 수천억 개 단위로, 이는 HCSF만이 처리 가능한 규모였다. 특히 데이터 크기에

상관없이 분석 애플리케이션 요구 성능을 만족시키는 면에서도 HCSF는 B사가 원하는 환경에 최적의 솔루션이었다.

B사는 HCSF 도입을 통해 고성능 스토리지 인프라를 구축함과 동시에 다양한 클라이언트 운영 환경을 쉽게 구성했으며, 정책 기반 데이터 티어링 스토리지를 쉽게 구축했다. 이를 통해 매일 수백 TB 규모의 데이터 분석 능력을 확보했고 동시에 데이터 증가나 분석 요건 변화에 유연하게 대응할 수 있었다. 무엇보다 비용 효율적인 스토리지 인프라를 확보했다는 점에서 B사는 매우 만족하고 있다.

타사 대비 HCSF의 두드러진 강점과 매력으로 R&D 기업, AI 서비스 기업, 공공기관 등에서 HCSF를 도입하려는 움직임이 활발하다.

↓ 실제 사용 고객 및 다수의 경쟁, BMT 등을 통해 완벽히 검증된 'HCSF'

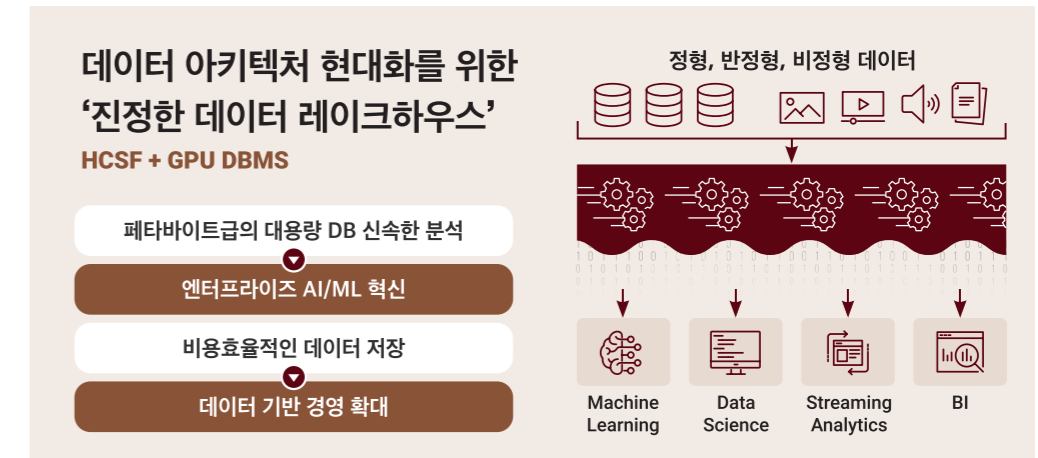
BMT 성능	GPU 환경 최적화	레퍼런스
<b>BMT로 검증된 HCSF 성능</b>  <p>스토리지 성능 1.5~2배 DB 조회 성능 4~5배</p> <p>경쟁사 12대    HCSF 6대</p>	<b>GPU 환경 최적화</b> HCSF는 슈퍼/고성능 컴퓨팅 환경에 최적화  <p>HCSF CONTENT OPTIMIZATION FOR FILE</p> <p>AI, ML, 빅데이터 등 수십 PB 단위 대용량 데이터 분석에 최적화된 스토리지</p>	<b>국내외 다수의 레퍼런스</b> <b>해외</b> 생산 프로세서 파운드리 수율이 계획보다 다소 낮음 인지 HCSF 100노드 클러스터 구성, 전체 파운드리 데이터 적재 및 고속의 분석·지원 <b>국내</b> 제조, 리서치, 교육 및 개발, AI서비스 기업 등에서 다수의 레퍼런스 확보

### 데이터 기반 경영의 초석, '고성능 데이터 레이크하우스'

효성인포메이션시스템은 최근 GPU를 활용하는 데이터베이스 관리시스템(DBMS) 업체 스크림(SQream)과의 협업을 통해 대용량 데이터의 신속한 분석과 엔터프라이즈 AI 및 머신러닝(ML) 혁신을 위한 '데이터 레이크하우스' 구현으로 고객의 데이터 기반 경영을 지원한다.

데이터 레이크하우스는 확장할 수 있는 '데이터 레이크'의 속성과 구조화된 '데이터 웨어하우스(DW)' 기능을 결합한 새로운 개방형 데이터 관리 아키텍처다. 모든 데이터 소스에서 분석 및 인텔리전스가 가능하며 데이터의 확장성과 민첩성, 유연성을 높여준다. 효성인포메이션시스템은 주요 DW 업체들과 협업을 통해 시너지를 높이며 굵직굵직한 데이터 레이크하우스 레퍼런스를 확보하고 있다.

고객들은 효성인포메이션시스템의 HCSF와 스크림의 GPU 기반 DBMS를 이용해 가속 데이터 웨어하우스를 활용할 수 있다. PB급 대용량 DB를 안정적으로 저장하고, 복잡한 쿼리 처리와 신속한 분석 업무도 가능하다. 특히 하둡, 클라우드라, 테라데이터와 같은 기존 분석 시스템을 그대로 활용하면서 대용량·고성능의 분석 요건이 필요하면 GPU DBMS인 스크림을 이용해 추가로 분석 시스템을 구축하고 활용할 수 있다.



효성인포메이션시스템은 국내에서 HCSF 도입이 활발해짐에 따라 웨카IO, 스크림(SQream)과 같은 전문 벤더와의 협업을 늘리고 있다. 앞으로 대규모 제조 기업, 플랫폼, 통신 분야의 기업을 비롯해 대용량 데이터 처리와 저장이 필요한 공공기관, 금융사 등 다양한 산업에 HCSF가 활용될 수 있도록 비즈니스를 적극적으로 펼칠 계획이다.

“다수의 구축 경험과 최고의 전문 인력을 지닌 데이터 전문 벤더와의 협업을 통한 시너지로 고객들의 데이터 혁신을 앞당길 수 있을 것으로 기대한다.”  
- 효성인포메이션시스템

“웨카IO는 고성능 데이터 플랫폼을 기반으로 기업의 데이터센터 투자 가치를 극대화하고, 데이터 중심의 비즈니스 성과를 최대로 창출할 수 있도록 효성인포메이션 시스템과 적극적으로 협업할 것이다.”  
- 웨카IO

“데이터를 비용 효율적으로 비즈니스 가치로 전환하고, 데이터의 잠재력을 극대화할 수 있도록 지원을 아끼지 않을 것이다.”  
- 스크림(SQream)