

## 데이터 허브: 스토리지의 새로운 비전

### 변화하는 데이터의 본질

데이터가 기업을 견인하는 가장 중요한 전략적 자원으로 부상하며 새롭게 주목을 받고 있습니다. 이러한 새로운 데이터 시대는 몇 가지 트렌드를 기반으로 합니다. IT 인프라의 진화와 함께 데이터가 급증하고, 애널리틱스가 엔터프라이즈 데이터센터의 핵심 역량으로 부상함에 따라 컴퓨팅 자체의 본질도 변화하고 있습니다. IT 업체들은 제품을 발전시켜야만 새로운 데이터 시대의 요구에 부응할 수 있을 것이며, 이는 스토리지 업계에 특히 더 해당된다 할 수 있습니다.

데이터의 양, 속도 그리고 그 본질 자체가 변화하고 있습니다. 기존의 데이터센터 외부에서 데이터 수집과 프로세싱이 이뤄지는 엣지 컴퓨팅(edge computing)은 엔터프라이즈 스토리지 시스템에 저장 및 처리되어야 하는 데이터를 논스톱으로 전송합니다. 엣지 컴퓨팅은 원거리, 기업의 지사, 운송수단이나 공장 등에 설치된 센서, 스마트 빌딩, 감시 시스템, 상점 내에서 고객들을 추적하는 리테일 비콘 등에서부터 데이터를 전송합니다. 이외에도 엣지 컴퓨팅으로 인한 새로운 데이터 타입과 소스를 많이 찾아볼 수 있습니다.

데이터 본질의 변화 중심에는 엔터프라이즈 워크플로우에 도입된 고급 애널리틱스가 있습니다. 머신러닝의 근본 기술이 발전했고, 엔터프라이즈 IT가 기술을 도입하는 것이 더 실용적이게 되었으며, 기술이 더 이상 과학적인 컴퓨팅 환경에 국한되어 있지 않습니다. 적절한 가격대의 GPU, 인공지능(AI)과 머신러닝 소프트웨어의 가용성 확대로 기업들은 이제 데이터를 새롭고 혁신적인 방식으로 볼 수 있게 되었습니다.

머신러닝은 복수의 데이터 소스의 빠른 상호 연결을 도모해 비즈니스 크리티컬한 패턴을 파악합니다. 예를 들어, 소매업체와 금융기관이 거래 및 재고 데이터에 딥 러닝을 활용하게 되면 거래 사기를 발견할 수도, 손실을 막을 수도 있습니다. 물류업체는 인공지능을 활용해 일기예보, 실시간 교통정보와 화물열차나 트럭에 장착된 센서를 연결해 물류 노선과 스케줄링을 결정할 수 있습니다. 이와 같이 고급 애널리틱스가 기업들의 더 나은 의사결정을 이끌고, 인사이트를 제공하는 사례는 쉽게 찾아볼 수 있습니다.

엣지 컴퓨팅과 애널리틱스가 기업에 새로운 역량을 제공하는 가운데, 데이터센터의 근본적인 인프라도 진화하고 있습니다. 퍼블릭, 프라이빗 클라우드 서비스 가용성 증대와 온-프레미스 서비스형 인프라(IaaS)로 인해 컴퓨트(Compute)를 쉽게 이동할 수 있는 시대가 열렸습니다.

컴퓨팅 워크로드의 서버 간 이동, 로컬 서버와 클라우드 간 이동은 비교적 쉽습니다. 그러나 워크로드의 데이터 종속성이 높은 경우 문제가 야기될 수 있습니다. 데이터 그래비티(data gravity)가 실제로 존재하며, 서로 다른 인프라스트럭처 간의 큰 데이터 블록 이동은 비용이 많이 들 뿐만 아니라 시간도 많이 소모됩니다. 특히 워크로드가 로컬 및 접근 가능한 데이터 모두를 필요로 하는 경우에 더 문제시 됩니다. 즉, 데이터와 데이터를 저장하는 시스템 모두 하이브리드 클라우드 시대에 맞게 새롭게 설계되어야만 하는 것입니다.

수많은 혁신이 데이터를 전략적인 차별화 요소로 만들고 있습니다. 진화하는 데이터 세계의 새로운 가능성을 실현하기 위해서는 데이터 저장 방식과 위치, 데이터의 가용성에 의존하는 워크로드 간의 데이터 공유 방식을 새로운 시각으로 살펴봐야 합니다.

## 스토리지 아키텍처의 한계

일반적인 데이터센터는 별개의 사일로를 이용해 엔터프라이즈 기업 내에 데이터를 퍼뜨립니다. 따라서 데이터가 데이터 웨어하우스 시스템, 특수 용도의 서버에 연결된 스토리지 어레이, 혹은 심지어 가상화 혹은 컨버지드 인프라를 지원하는 소프트웨어 정의 데이터 풀에 위치하게 될 수 있습니다. 이러한 데이터 사일로는 현대의 애널리틱스 기반 워크플로우의 방해 요소입니다.

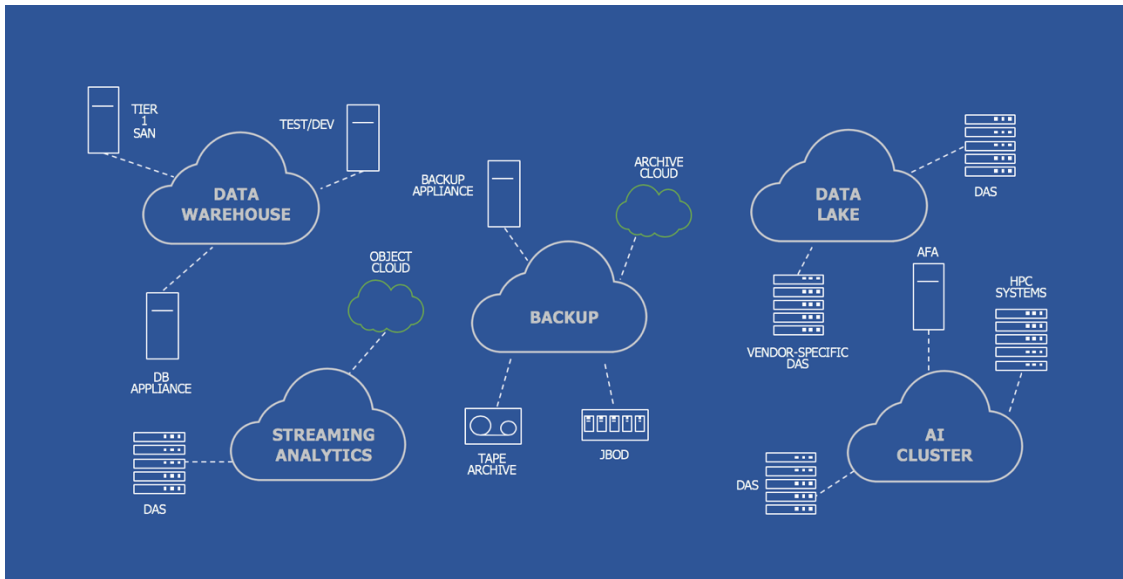
비즈니스 인텔리전스 애플리케이션, 애널리틱스, 엣지 스토리지, 가상 데스크톱 등 각각의 컴퓨팅 모델이 연계된 스토리지 인프라스트럭처로부터 필요로 하는 기능은 상이합니다. 그러나 각 시스템의 요구사항이 다름에도 종종 같은 데이터 세트가 활용되곤 합니다. 이러한 불균형이 존재하기 때문에 엔터프라이즈 스토리지 시스템을 새로운 시각으로 봐야 하는 것입니다.

예를 들어, 인공지능을 주로 활용하는 애널리틱스는 수집된 데이터의 사용과 애널리틱스 워크로드를 수행하는 서버로의 데이터 제공, 이 두 가지에 대한 니즈를 가지고 있습니다. 애널리틱스 특성 상 데이터는 사일로에 분산되어서는 안되며, 하나의 통합체로서 사용이 가능해야 하고, 대용량의 데이터를 빠른 응답시간으로 컴퓨팅 엔진에 전달되어야 합니다.

데이터의 비정형적 특성, 서로 다른 데이터 세트의 공존 필요성을 고려했을 때, 이제는 데이터 자체의 저장방식을 재고해봐야 합니다. 기업에서의 데이터 사일로의 문제점은 애널리틱스에 필요한 데이터를 하나의 통합체로 구성하는 것을 저해한다는 점입니다. 또한, 사일로로 인해 엔터프라이즈 전반에서 데이터 무결성을 유지하고, 데이터 보호와 거버넌스에 일관된 방향을 제시하는 것이 어렵습니다.

애널리틱스가 점점 대세가 됨에 따라 데이터 레이크가 데이터 사일로 일부로 대체하게 될 것입니다. 데이터 레이크는 정형, 비정형 데이터 전체를 전사적인 관점에서 볼 수 있게 해줍니다. 데이터 레이크에서는 모든 애플리케이션이 지속적, 통합적으로 데이터에 접근하는 것이 가능하지만, 기술적 한계로 인해 이를 기업들이 활용할 수 없었습니다. 기존의 아키텍처에서는 데이터의 실시간, 다차원적 제공이 불가능하기 때문입니다.

### 그림 1 : 사일로 환경에 갇힌 오늘날의 데이터



출처: 퓨어스토리지

엔터프라이즈 데이터 레이크 도입을 막는 방해요인으로 데이터를 보유하고 있는 스토리지 시스템의 한계를 꼽을 수 있습니다. 스토리지 기술은 전혀 없는 수준의 성능을 제공할 능력을 충분히 보유하고 있습니다. 핫 데이터(hot data)와 웜 데이터(warm data)에서는 SSD가 기존 하드 드라이브를 대체하고 있으며, NVMe와 같은 인터커넥트는 스토리지 경로에서 응답시간 지연을 완전히 제거해줍니다. 이더넷 상에서 연결되는 고속 스토리지는 초대용량의 처리량을 제공합니다. 스토리지는 그 어느 때보다 빠르고, 집적도가 높으며, 경제적입니다.

또한, 오늘날의 스토리지 시스템은 데이터 풀을 공유하는 다양한 워크로드에 동시에 서로 다른 기능을 제공하기보다는 특정 애플리케이션의 요구사항에 적합하게끔 구축되어 있습니다.



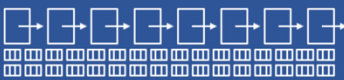

즉, 스토리지 시스템의 한계는 기술의 한계가 아니라 사람이 만든 것이며, 관행적인 잘못된 구조로 인한 결과입니다. 스토리지 시스템의 기초 요소들은 새로운 아키텍처에 통합되어, 빠르게 변화하고 있는 엔터프라이즈 컴퓨팅과 데이터 분석의 요구사항을 들어줄 수 있기만을 기다리고 있습니다.

## 데이터 허브: 데이터와 스토리지 연계

애널리틱스의 부상으로 데이터는 전략적 요소가 되었습니다. 애널리틱스의 가능성, 엣지 컴퓨팅과 같은 새로운 데이터 소스의 등장으로 데이터는 더 풍부해졌습니다. 기업들은 데이터 레이크를 활용해 데이터를 필요로 하는 워크플로우에 데이터에 대한 일관적 관점을 제공하고자 하지만 기존 솔루션의 복잡성과 낮은 성능으로 어려움을 겪고 있습니다.

기존 스토리지 시스템 상에 새로운 비전을 실현하는 것은 쉽지 않습니다. 왜냐하면 기존 스토리지 시스템에서는 같은 데이터를 각각의 워크플로우가 요구하는 성능 수준에 맞게, 여러 스트림으로 나눠서 전달할 수가 없으며, 시스템이 하나의 애플리케이션 세트의 요구사항에만 부합하게끔 설계되어 있기 때문입니다.

그림 2: 사일로 별 스토리지 요구사항

	BACKUP & DATA WAREHOUSE	DATA LAKE	STREAMING ANALYTICS	AI CLUSTER
				
	APPLIANCES	DAS	DISAGGREGATED	HPC SYSTEM W/ GPU
Data Type	structured	unstructured	unstructured	unstructured
Processing Mode	batch	batch	micro-batch / real-time	real-time
I/O type	random read	sequential	random	sequential to random
Architecture	scale-up	scale-out	multi-dimensional	massively parallel

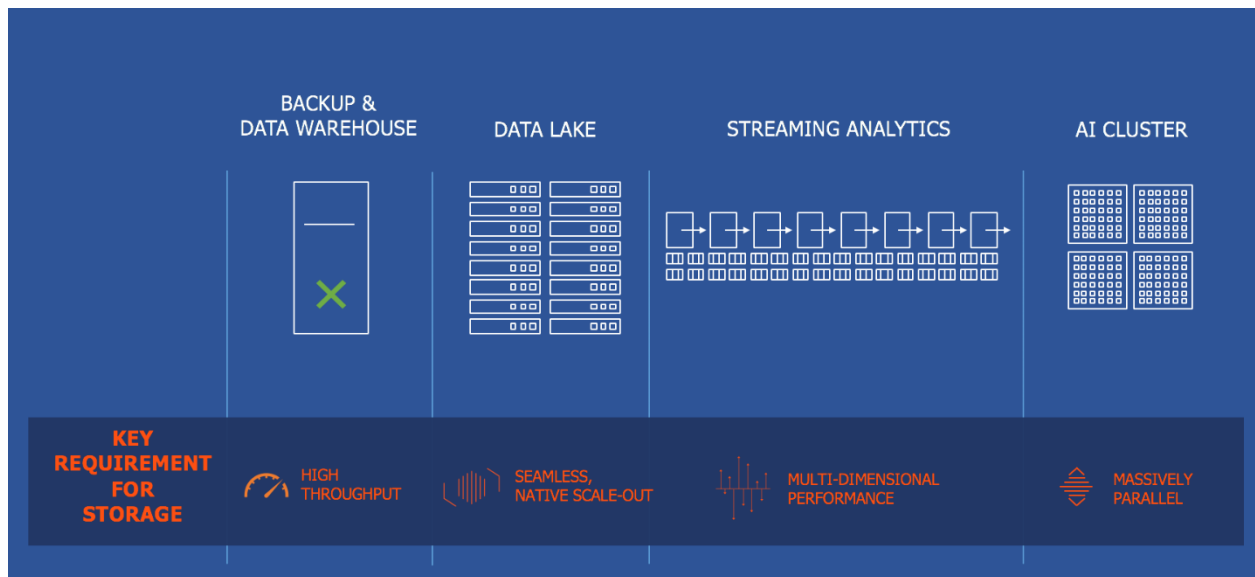
출처: 퓨어스토리지

엔터프라이즈 컴퓨팅의 변화하는 니즈를 수용하고, 앞서 언급한 한계를 극복할 수 있는 새로운 스토리지와 데이터 아키텍처가 필요합니다. 그리고 우리는 이 새로운 아키텍처를 “데이터 허브”라 부릅니다.

데이터 허브는 데이터 관점에서의 엔터프라이즈의 비전을 실현하기 위해 몇 가지 가능성을 제공할 수 있어야 합니다. 대용량의 파일과 오브젝트 스토리지를 생성하고, 애플리케이션에 맞는 성능 (레이턴시, 처리량, IOPS)을 제공함과 동시에 애플리케이션 맞춤형 기능을 다수의 클라이언트에 동시에 제공할 수 있어야 합니다. 또한 데이터 허브는 진정한 스케일 -아웃을 가능하게 해야 합니다.

데이터 허브 위에 복제, 스냅샷, 압축, 암호화 등 기업이 요구하는 일반적인 스토리지 기능을 얻을 수도 있습니다. 데이터 허브를 기존의 스토리지 기술과 차별화시키는 요소는 다차원적 성능의 제공입니다.

### 그림 3: 데이터를 통합하는 데이터 허브



출처: 퓨어스토리지

데이터 허브의 가능성을 실현하기 위해서는 스토리지 시스템의 재해석이 필요합니다. 스토리지 어레이는 이제 더 이상 서버 마더보드, 백엔드 드라이브 어레이 위에 구축되지 않을 것입니다. 그 대신, 데이터 허브는 복수의 병렬 컴퓨팅과 개별적으로 분할되고, 결합된 다양한 워크로드에 맞춤화될 수 있는 스토리지 요소를 필요로 합니다. 기초적인 하드웨어 구성보다 소프트웨어 구성이 더 중요해집니다.

업계는 컴퓨팅, 스토리지, 네트워킹을 하나의 긴밀히 연결된 자원 풀에 통합하는 등의 컨버전스를 통해 위와 같은 요구사항에 부응하려 했습니다. 하이퍼 컨버지드 인프라스트럭처(hyper-converged

infrastructure)가 대다수의 워크로드에 적합하고, 또 데이터 허브 내의 데이터를 사용하기도 하지만, 다수의 애플리케이션 도메인에 일관된 데이터를 제공하기엔 어렵습니다.

데이터를 공유하는 애플리케이션이 조직 내에서 분산되어 있거나 온-프레미스와 클라우드 기반 서비스에 걸쳐 있는 경우, 컨버전스는 무너집니다. 하이브리드 클라우드가 새로운 표준이 되고, 엣지 컴퓨팅과 같은 새로운 아키텍처가 자원을 분산시킴에 따라 데이터와 컴퓨팅의 코로케이션(co-location)이 더 복잡해집니다. 그 결과로 데이터에 대한 액세스를 조정하는 것이 단기적인 문제가 아닌 장기적인 문제가 됩니다.

데이터 허브 아키텍처는 유연하고, 확장성 있게 엔터프라이즈 데이터를 종합적으로 활용할 수 있게 합니다. 데이터 허브 소프트웨어는 데이터를 필요로 하는 애플리케이션으로의 데이터 전달을 조정하고, 필요한 성능 요건도 함께 제공합니다. 그리고 이는 소프트웨어 정의 데이터센터(SDDC)를 지원해 IT 실무자들이 애플리케이션 요구사항에 의해 정의된 데이터 중심 역량을 제공할 수 있도록 도와줍니다. 실무자들은 컨버전스에 내재된 경직성과 긴밀히 연결된 자원 풀에 의존하는 다른 아키텍처의 제약을 받지 않게 됩니다.

#### 그림 4: 데이터 허브의 4 가지 주요 특성



출처: 퓨어스토리지

## 결론

기업은 데이터의 속도에 따라 움직입니다. 그런데 그 데이터의 속도가 가속화되고 있습니다. 그 어느 때보다 많은 소스에서 데이터가 전달되고 있으며, 데이터는 애널리틱스 엔진을 통해 유용한 인사이트를 기업에 제공합니다. 새로운 데이터 세계를 완전히 포용하기 위해 기업들은 데이터 아키텍처를 재점검해야 하고, 기술 업체들은 스토리지 시스템 설계방식을 재고해야 합니다.

새로운 데이터의 현실에 맞는 새로운 스토리지 아키텍처를 정의할 기회가 IT 업계에 주어졌습니다. 오늘날의 IT 세계에서 스토리지 사일로가 설 자리는 없으며, 기존 스토리지 시스템과 아키텍처의 한계로 오늘날 환경에 데이터 레이크를 효과적으로 구현하는 것이 어렵습니다.

스토리지 업체는 애플리케이션 맞춤형 성능, 대용량 파일 및 오브젝트 스토리지, 네이티브 스케일-아웃 등의 뛰어난 성능과 다양한 기능을 제공하는 솔루션을 제공해야 합니다.

데이터 허브는 스토리지 아키텍처를 생각하는 새로운 사고방식입니다. 데이터 허브를 가능케 하는 모든 요소들이 오늘날 이미 존재합니다. 결국 새로운 비전을 제공하는 업체가 기업들이 데이터 관점의 애널리틱스를 완전히 아우르는 데에 필요한 인프라를 제공하게 될 것입니다.

새로운 데이터 세계에 대응하기 위해 기업들은 데이터 자체 및 데이터의 저장과 전달 방식에 대한 새로운 접근 방식을 요구할 것입니다.

## 중요정보

### 공헌

[Steve McDowell](#), 책임 애널리스트, [Moor Insights & Strategy](#)

[Patrick Moorhead](#), 창립자, 회장, 수석 애널리스트, [Moor Insights & Strategy](#)

### 출판

[Patrick Moorhead](#), 창립자, 회장, 수석 애널리스트, [Moor Insights & Strategy](#)

### 문의사항

이 보고서에 대한 문의사항이 있을 시 [연락주세요](#). Moor Insights & Strategy 가 신속히 답변 드리겠습니다.

### 인용

이 문서는 공인된 언론, 애널리스트에 의해 인용될 수 있으며, 맥락에 맞게 인용되어야 하며, 저자명, 제목, "Moor Insights & Strategy"를 함께 표기해야 합니다. 언론, 애널리스트가 아닌 경우, Moor Insights & Strategy 의 사전 서면 동의 후에 인용이 가능합니다.

### 라이선스

이 문서 및 근거자료는 Moor Insights & Strategy 의 소유며, 이 문서는 Moor Insights & Strategy 의 사전 서면 허락 없이 재생산, 배포 혹은 공유될 수 없습니다.

### 정보공개

이 보고서는 Pure Storage 가 의뢰한 것입니다. Moor Insights & Strategy 는 보고서에 포함된 다수의 기술업체에 연구, 분석, 조언 및 컨설팅을 제공하고 있습니다. Moor Insights & Strategy 의 직원 모두 이 문서에 언급된 기업의 지분을 소유하고 있지 않습니다.

### 고지사항

이 문서에 제시된 정보는 정보전달 목적만을 가지고 있으며, 본 정보에는 기술적 오류 또는 인쇄 상의 오류가 있을 수 있습니다. Moor Insights & Strategy 는 본 정보의 정확성, 완전성 혹은 적합성에 대해 책임지지 않으며, 본 정보 내의 오류, 누락 혹은 불충분함에 대해 법적 책임을 지지 않습니다. 이 문서는 Moor Insights & Strategy 의 견해만을 담고 있으며, 사실의 진술이라 해석되어서는 안됩니다. 이 문서 내의 견해는 사전 고지 없이 변경 가능합니다.

Moor Insights & Strategy 는 미래 사건의 정확한 예측이 아닌 방향 제시의 지표로 예측 및 전향적 서술을 합니다. 예측 및 전향적 서술은 Moor Insights & Strategy 의 미래에 대한 현재 판단을 대변하지만 실질적으로 다른 결과를 낼 수 있는 리스크와 불확실성의 대상이 됩니다. 따라서 이 문서의 발간일 당시의 의견을 반영한, 예측 및 전향적 서술에 지나치게 의존하지 않기를 주의 드립니다. Moor Insights & Strategy 는 새로운 정보 혹은 미래 사건에 따른 예측 및 전향적 서술의 수정 혹은 수정의 결과를 공개적으로 발표할 법적인 의무가 없음을 유의하시길 바랍니다.

©2018 Moor Insights & Strategy. Company and product names are used for informational purposes only and may be trademarks of their respective owners.