

## 실시간 분석·AI로 확장되는 빅데이터 기술 공인된 솔루션 플랫폼 ‘Hitachi UCP’로 통하다

Writer 임경훈 / Senior Consultant, Hitachi Vantara Korea

메인프레임, UNIX-오픈 시스템, 서버-클라이언트, x86, 서버 가상화 그리고 클라우드.

시대별 사람들의 유행이 유행에 따라 변화하듯, 어쩌면 IT도 시대별 유행을 따른다. ‘신념’이라 불릴 정도로 굳건했던 기술들이 새로운 기술의 출현으로 급격히 소멸해가는 것을 우리는 수없이 경험했다. 이런 IT 변화의 흐름에 역행하거나 동떨어져 있는 상황에서 기업이 살아 남기란 매우 어렵다.

새로운 기술이 과거의 문제를 해결한다. 시장은 새로운 기술의 전문가를 찾게 되며, 이는 주류 인력 풀의 이동으로 이어지고, 동시에 이를 통해 다양한 신규 솔루션이 지속적으로 시장에 공급되는 선순환 구조가 형성된다. 과거에는 불가능했던 문제를 해결할 수 있는데 굳이 과거에 집착할 필요가 없기 때문이다. 우리는 이것을 혁신이라고 부르며 IT가 이 혁신을 주도하고 있다는 것을 아무도 의심하지 않을 것이다.

### 혁신을 위한 플랫폼이 필요할 때

근래 IT 업계에서 모든 이들의 관심사를 세 가지 꼽자면 클라우드, 빅데이터 그리고 AI(인공지능)를 들 수 있겠다. 세 가지 모두, 과거의 기술들이 풀지 못한 숙제를 진일보한 하드웨어와 소프트웨어의 결합을 통해 해결해 낸 기술적 쾌거다.

사실 클라우드 기술은 과거에도 수많은 데이터센터 내에서 부분적으로 조금씩 사용된 것이다. 하지만 현재에 이르러서야 전체 기술들이 하나로 통합된, 충분히 성숙해진 플랫폼을 볼 수 있게 됐다. 이로 인해 기업의 혁신과 폭발적 성장이 가능해졌다는 것에 아무도 이견을 내지 않을 것이다. ‘더 적은 비용’으로 ‘더 많은 일’을 ‘더 견고하게’ 해내야 하는 것이 당연하게 여겨지는 분야가 IT 산업이기에, 솔루션을 위한 최적의 플랫폼을 선정하고 안정적으로 운영하는 것이 더욱 더 중요해지고 있다.

초창기 빅데이터 분석 프로젝트의 대다수는 기존에 분석하지 못했던 데이터에서 인사이트를 얻기 위해 저렴한 일상재 하드웨어를 오픈소스 소프트웨어로 플랫폼화해 수행되었다. 하지만 각기 다른 하드웨어 구성, 잦은 장애, 소프트웨어 구성 오류, 확장 불가, 업그레이드 및 유지보수의 어려움 등 이슈들이 불거지면서 처음 도입한 상태 그대로 수년간 운영되는 것이 현실이었다.

도입 부서(마케팅 부서)나 실무 부서의 경우 IT 관련 이슈에 생소하다는 점은 작은 문제에 불과하다. 엔터프라이즈급 기능 부족이라는 오픈소스 소프트웨어의 태생적 한계로 인해 시각화와 자동화가 어렵다는 문제를 항상 안고 가야만 했다. 분석 시스템의 결과는 임원들이 자주 확인하기 때문에 장애를 최대한 피해야 한다. 이런 요구 조건을 충족시키지 못해 결국은 아무런 변화를 주지 않는 것이 가장 최선의 방법이었다. 이 때문에 아파치 스파크(Apache Spark)의 인메모리를 기반으로 빠른 맵리듀스(MapReduce) 및 다양한 라이브러리를 활용하는 것이 주류가 된 지 수년이 지났음에도 불구하고 여전히 디스크 기반의 일반 아파치 하둡(Apache Hadoop)으로 빅데이터 분석 체계를 유지하는 사례가 적지 않다.

### 빅데이터 솔루션의 확장: 실시간, BI 그리고 AI

초기 하둡 기반의 분석에만 집중했던 빅데이터 솔루션들이 이제는 실시간 분석, 비즈니스 인텔리전스(Business Intelligence, 이하 BI), 인공지능(Artificial Intelligence, 이하 AI) 영역까지 확장되고 있다. 이에 필요한 요구사항을 간략히 정리하면 다음과 같다.

실시간 분석	비즈니스 인텔리전스	인공지능
 <p>기존과 같이 데이터를 축적한 후 분석하는 것이 아닌 스트리밍 데이터를 정해진 시간 내에 처리 완료할 수 있어야 한다.</p>	 <p>비 IT 관련자라도 쉽게 데이터를 다루고 인사이트를 추출할 수 있어야 하며 풍부한 시각화를 제공해야 한다.</p>	 <p>다양한 오픈소스 머신러닝 라이브러리(R, Python, WEKA, Spark MLlib 등)를 지원해야 하며 필요한 경우 GPU를 통한 딥러닝까지 제공해야 한다.</p>

이들 요구 조건은 아래 공통 사항으로 정리가 가능하다.

- ✓ 견고한 운영과 자동화가 가능한 엔터프라이즈급 솔루션
- ✓ 오픈소스 및 타 기술들과 유연한 연계가 가능한 소프트웨어
- ✓ 성능, 안정성, 확장성을 두루 갖춘 검증된 하드웨어
- ✓ 현재는 물론 미래의 변화까지 대비한 플랫폼

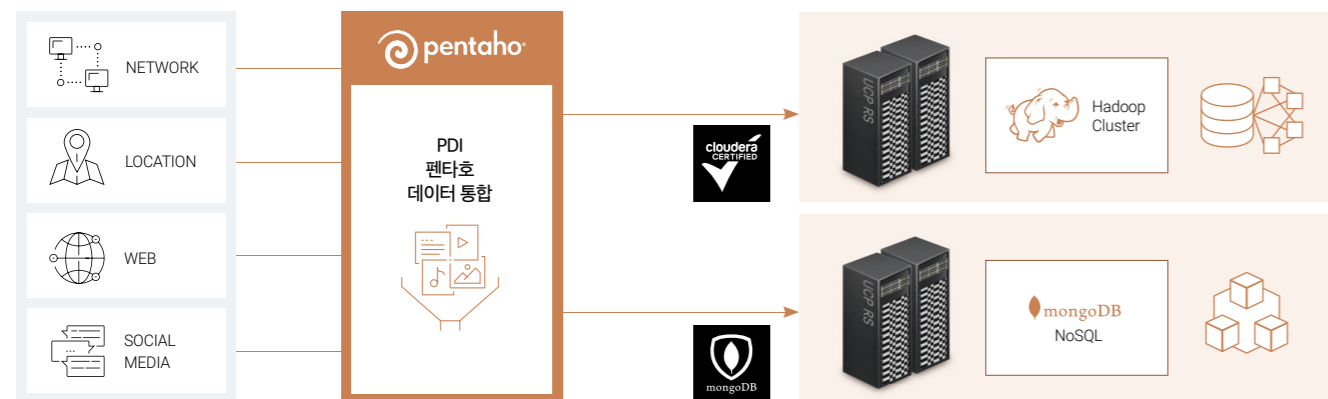
### 각광받는 공인 솔루션 플랫폼

위 문제를 해결하고자 최근 많은 고객들이 Certified Solution Platform을 도입하고 있다. 이 플랫폼이 탄생하게 된 배경을 보면 온전히 고객의 요구에 의한 것임을 확인할 수 있다.

우선 하드웨어 제조사와 소프트웨어 제조사가 함께 특정 솔루션에 가장 최적화된 서버, 스토리지, 네트워크를 하나의 제품으로 통합한다. 충분한 검증을 통해 안정성을 강화한 레퍼런스 아키텍처로 Pre-engineered 하드웨어를 플랫폼화 한다. 소프트웨어 제조사에서는 하드웨어와의 공합과 관련된 문제를 사전에 제거할 수 있고 낮은 장애율, 유연한 구성 및 확장, 업그레이드와 같은 요구를 충족시킬 수 있다. 게다가 최적의 성능을 위한 튜닝 작업도 필요없다. 고객은 '공인' 된 플랫폼이 제공하는 장점을 활용해 솔루션 활용의 유연함, 안정적인 서비스, 투자 보호 등의 목표를 모두 달성할 수 있다.

히타치 벤틀라는 빅데이터 환경을 위한 Certified Solution Platform으로 컨버지드 인프라인 Hitachi UCP RS를 제공한다. 일반적인 서버 인프라와 다른 점은 서버, 스토리지, 네트워크가 모두 히타치 벤틀라를 통해 공급된다는 것이다. 모든 하드웨어는 발주를 할 때마다 공장 출하 전 전담 엔지니어가 직접 하드웨어와 소프트웨어를 공인된 형태로 구성한 후 충분한 테스트를 통해 하드웨어 오류를 사전에 제거한 후 출하한다. 특히 메모리가 대량으로 들어가는 빅데이터 솔루션 특성 상 실제 운영 시 발견될 수 있는 장애를 줄여준다는 데 큰 의미가 있다. 고객은 장비 수령 후 즉시 솔루션을 가동할 수 있기 때문에 프로젝트 기간 단축 효과도 기대할 수 있다.

그림 공인된 엔터프라이즈 솔루션 플랫폼



히타치 벤틀라의 빅데이터 분석 솔루션인 Pentaho 뿐만 아니라 현재 Cloudera의 Enterprise Data Hub, MongoDB Enterprise가 공인된 솔루션 플랫폼으로 등재되어 있으며, 앞으로도 다른 엔터프라이즈급 빅데이터 솔루션들이 지속적으로 추가될 예정이다.

### 혁신을 위한 플랫폼, Hitachi UCP

히타치 벤틀라는 스토리지 시장에서 최상의 안정성, 뛰어난 성능 그리고 시대를 앞서가는 선도적 기술로 수십 년간 시장에서 사랑받아 왔다. 이 고집을 그대로 x86 서버에 접목시킨 것이 바로 Hitachi UCP(Unified Compute Platform)다. 주요 기업들이 원하는 솔루션 구성에 따라 다음의 세 가지 모델을 제공한다.

#### UCP CI Converged Infrastructure

서버와 스토리지를 결합해 일반적인 서버 가상화뿐만 아니라 SAP HANA, Oracle DB, Pentaho Business Analytics와 같은 미션 크리티컬 업무까지 최적화된 플랫폼이다. 2U 서버에 4개의 GPU가 탑재되는 DS225G 모델이 있다.

#### UCP HC Hyper-Converged Infrastructure

복수 서버의 내장 디스크 전체를 분산 파일시스템으로 묶음으로써, 별도의 외장 스토리지 없이 가상화 환경을 제공하는 VMware vSAN 기반 하이퍼컨버지드 플랫폼이다. 소규모 프라이빗 클라우드를 위해 작게 시작해(최소 2대) 1대 단위로 무중단 상태에서 서버와 디스크 공간을 개별적으로 늘려나갈 수 있다. 전통적인 IT 인프라가 갖지 못한 '사용한 만큼 지불하는 방식(Pay as you grow)'과 '타임-투-마켓(Time to Market)'에 최적화되어 있다.

#### UCP RS Rack Scale Infrastructure

랙 단위로 증설 가능한 대용량 IT 인프라 서비스 플랫폼으로, 두 가지 세부 모델로 나뉜다. VMware Cloud Foundation 기반 프라이빗 클라우드를 위한 UCP RS는 서버, 스토리지, 네트워크 전 영역에 걸쳐 소프트웨어 정의 데이터센터(SDDC)를 제공한다. 다른 하나는 이 글에서 다뤘듯이 UCP RS 하드웨어 구성 위에 빅데이터, BI, SI와 같은 대용량 처리 솔루션들에 특화된 공인(Certified) 플랫폼이다.