



분석에 대한 실제 비용

*Hadoop은 저렴해 보일 수 있지만
숨겨진 비용이 있습니다.
귀사의 조직에 적합한 플랫폼을 선택하는 방법을 알려드립니다.*

RICHARD WINTER 저

최 근 몇 년 동안 급증한 데이터 볼륨과 다양성 요구를
충족하기 위해 조직은 주요 비즈니스 부서에 대한 새로운
정보를 수집하고, 이미 수집된 데이터의 가치를 극대화하고,
비즈니스에 대한 포괄적인 뷰를 제공할 수 있는 분석 솔루션을 필요로
해왔습니다. >>

방에 있는 코끼리는 말하자면, Apache™ Hadoop®과 같은 존재로 엔터프라이즈 데이터 웨어하우스(EDW)에 대해 비용이 더 저렴한 대안으로 관심을 얻고 있는 빅 데이터 저장 및 처리를 위한 오픈소스 플랫폼입니다. 테라바이트(TB)당 1,000달러 이하로 거래되는 Hadoop 클러스터 가격은 데이터 웨어하우스 플랫폼의 평균 정가보다 몇 배나 더 저렴합니다. 그러나, 비즈니스 가치에 대한 진정한 기준을 얻기 위해 꾸준한 데이터 분석을 수행할 수 있는 소프트웨어를 개발 또는 구매하는 데 필요한 투자도 고려해야 합니다.

비즈니스 비용

조직이 데이터 웨어하우스, Hadoop 또는 두 가지 모두의 조합 중 어느 것으로 갈 것인지에 대한 결정을 내릴 수 있도록 Winter Corp.는 총 데이터 비용(TCOD)을 추정할 수 있는 프레임워크를 개발했습니다. 시스템 정가 외에도, TCOD는 일정 기간 동안의 데이터 사용 비용을 고려합니다. 여기에는 시스템 상단에 구축되는 복잡한 쿼리, 분석 및 분석 애플리케이션 등 비즈니스 솔루션의 개발 및 유지보수가 포함됩니다.

TCOD의 주요 구성 요소는 다음과 같습니다.

- > 하드웨어와 시스템 소프트웨어의 구매, 유지보수, 지원 및 업그레이드(공간, 전력 및 냉각 포함), 데이터 웨어하우스와 함께 데이터베이스 관리 소프트웨어의 비용도 고려
- > 시스템 및 데이터를 관리하는 전문 인력의 채용 및 유지
- > 분석을 위한 데이터를 준비하기 위해 ETL(Extract, Transform, Load) 또는 이와 유사한 솔루션 구매
- > SQL로 표현할 수 있는 쿼리 작성
- > SQL로 표현하기에 너무 복잡한 데이터 분석을 수행하는 절차 프로그램 작성
- > 금융 기관에서의 캠페인 관리 또는 대출 승인 등 반복적 프로세스를 지원하기 위해 데이터를 사용하는 실질적인 애플리케이션 개발 >>

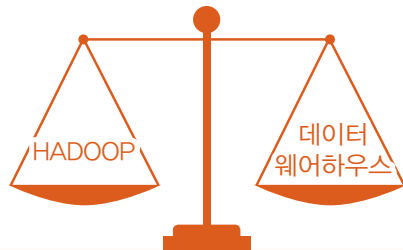
선택은 여러분의 몫입니다

둘 중 어느 것을 선택할 것인지 또는 데이터 웨어하우스와 Apache™ Hadoop® 플랫폼 모두를 결합할 것인지는 시스템 요구 사항에 따라 달라집니다. 총 데이터 비용(TCOD) 분석은 다음과 같이 적절한 빅 데이터 분석 솔루션을 선택하기 위한 몇 가지 지침을 강조합니다.

- > 플랫폼을 선택하기 전에 각 애플리케이션에 대한 분석 및 데이터 관리 요구 사항을 **확인 및 평가**합니다. 데이터 웨어하우스 기술과 Hadoop은 각각 특정 유형의 분석 애플리케이션에 효과적이며, 이들의 비용 절감 효과는 많은 요인에 따라 달라집니다.
- > 빅 데이터 솔루션의 장기적인 실제 비용을 **과소 평가**하거나 (심한 경우) 무시하지 **않도록** 주의해야 합니다. 명백한 시스템 가격 외에도, 진행 중인 요구 사항을 평가합니다. 몇 년 후에는 이것이 시스템 가격보다 높아질 수 있습니다. 이를 무시하는 조직은 수백만 또는 심지어 수십억 달러에 이르는 천문학적 비용을 쏟아 붓는 엄청난 실수를 쉽게 범할 수 있습니다.
- > 데이터 웨어하우스 기술과 Hadoop 모두의 강점을 이용하는 **유연한 분석 아키텍처를 사용**합니다. 성공한 많은 조직들은 가장 적합한 상황에서 각각 배포할 수 있는 기회와 더불어 두 개의 플랫폼을 모두 포함하는 통합 구조를 구축할 수 있습니다.

THE TOTAL COST OF DATA: 총 데이터 비용

총 데이터 비용(TCOD)은 요구 사항의 조합에 따라 데이터 웨어하우스 또는 Apache™ Hadoop® 중 하나를 선호할 수 있습니다.



솔루션을 결정할 경우, 조직은 대략적인 총 비용을 비교하고 각각의 아키텍처가 가장 적합한 곳이 어디인지 파악해야 합니다.

Hadoop을 사용하면, 데이터 스토리지를 저렴한 비용으로 사용할 수 있습니다.

데이터 웨어하우스를 사용하면, 솔루션 개발, 특히 복잡한 쿼리 및 분석 비용을 최소화할 수 있습니다.

중앙 솔루션에 대한 요구

빅 데이터에 대한 각각의 접근 방식은 Apache™ Hadoop® 솔루션이든 엔터프라이즈 데이터 웨어하우스(EDW)든 상관 없이 조직의 상황에 따라 혜택을 안겨줍니다. 각각의 방식은 고유한 이점을 제공합니다. 다음은 주요 이점 및 각 솔루션을 선호하는 간단한 사례입니다.

엔터프라이즈 데이터 웨어하우스

분석을 위한 데이터 웨어하우스 플랫폼은 다음과 같은 이점을 제공합니다.

- > 비즈니스에 대한 데이터를 설명하고, 데이터 무결성을 유지하고, 정보에 대한 액세스를 제어 및 최적화하는 데이터 관리 인프라
- > 성능을 극대화하기 위해 시스템에 대해서뿐만 아니라 간단하고 복잡한 쿼리 및 분석 모두에 대해 SQL을 사용할 수 있는 능력
- > 데이터에 액세스할 수 있는 애플리케이션 또는 사용자 뷰를 변경하지 않고, 새로운 비즈니스 요구 사항과 데이터 소스를 수용하기 위해 데이터 정의를 쉽게 변경할 수 있는 유연성

EDW 선호

주요 금융 기관이 가진 분석 요구 사항으로는 엄청난 양의 데이터 소스, 사용자, 복잡한 쿼리, 분석 및 분석 애플리케이션, 그리고 데이터 통합 및 무결성에 대한 요구, 급변하는 비즈니스 요구 사항과 장기적 데이터 수명을 수용할 수 있는 재사용 및 민첩성에 대한 요구가 있습니다. EDW의 5년 간 총 데이터 비용(TCOD)을 살펴 보면, 데이터 웨어하우스에 대한 시스템 비용이 Hadoop 보다 훨씬 높다 하더라도 데이터 웨어하우스 플랫폼 비용이 2억 6,500만 달러로 7억 4,000만 달러에 달하는 Hadoop 솔루션보다 비용면에서 훨씬 더 효율적인 것으로 나타났습니다.

HADOOP

Hadoop 기술의 주요 이점은 다음과 같습니다.

- > 시스템 구매 비용 절감
- > 많은 양의 데이터와 여러 개의 다양한 데이터 유형을 빠른 속도로 병렬 처리
- > 폭넓은 선행 데이터 모델링 및 데이터베이스 설계 없이 데이터 액세스 가능
- > 자동 시스템 기반 데이터 복제
- > 단일 클러스터로 여러 대의 서버를 지원할 수 있는 능력

HADOOP 선호

한 기업은 대형 산업 디젤 엔진의 센서 출력을 세분화하기를 원했습니다. 요구 사항은 대용량 데이터 처리는 물론, 밀접하게 관련된 소량의 데이터 세트를 빠르고 집중적으로 처리하는 것을 포함합니다. 또한, 분석을 통해 전체 데이터 세트를 읽을 수 있고, 원시 데이터의 수명은 상대적으로 짧고, 소규모의 전문가 그룹은 분석에 대한 협업을 수행할 수 있어야 합니다. 5년 간 수집한 500TB의 데이터에 대한 TCOD는 930만 달러의 Hadoop이 3,000만 달러의 데이터 웨어하우스 어플라이언스보다 더 우수한 솔루션이라는 사실을 입증합니다.

TCOD 프레임워크의 주요 목표는 조직이 이 과정에서 많은 비용 손실을 초래하는 실수 가능성을 최소화하는 동시에 비용 효율적인 빅 데이터 분석 솔루션을 개발 및 제공할 수 있도록 지원하는 것입니다.

중요한 의사 결정

효과적인 분석을 수행하는 것은 조직 전체에 대한 중요한 성공 요인으로 빠르게 자리매김하고 있습니다. Hadoop과 같은 신기술이 업계에 발표되면 일반적으로, 모든 문제를 해결할 수 있는 더 나은 솔루션이 출시된다는 확신으로 이어집니다. Winter Corp.는 Hadoop이 모든 대규모 분석 환경에서 중요한 역할을 담당하겠지만 일부 분석 문제에 대해서는 그렇지 않을 것으로 확신하고 있습니다.

TCOD 프레임워크에서 볼 수 있듯이, 각각의 분석 분야에 적합한 솔루션을 선택하는 기업은 주어진 분석 목적을 달성하기 위해 노력한다는 점이 매우 중요한 요소로 작용해 더욱 비용 효과적인 조직이 될 수 있습니다. 궁극적으로, 각각의 빅 데이터 요구 사항에 가장 적합한 플랫폼을 선택하는 것은 조직의 성패를 결정한다는 것을 뜻합니다.

Richard Winter는 Winter Corp.의 CEO로 대용량 데이터베이스 기술과 빅 데이터 구현 전문가로 활동하고 있습니다. 그는 20년 간 이 분야에 종사해 온 전문가입니다.

온라인

총 데이터 비용에 대한 자세한 내용은 **Teradata.com**에서 "Big Data: What does it really cost?" 를 다운로드하십시오.