

고객이 Teradata에서 IBM PureData System for Analytics로 마이그레이션하는 이유



Dwaine Snow(데이터 웨어하우스 및 분석 어플라이언스 부문 제품 관리자/기술 전략 책임자)

요점

- PureData System for Analytics는 업계 최고의 가성비와 단순한 어플라이언스를 제공하도록 구축된 플랫폼을 통해 데이터 웨어하우스 및 분석 환경을 획기적으로 변화시켰습니다.
- 빠르고 신속하며 단순하고 확장성이 뛰어난 Netezza 기술은 성능, 용량 또는 예산 측면에서 기존 시스템이 한계에 봉착한 데이터 웨어하우징 및 분석 고객에게 매우 유용합니다.
- 대부분의 PureData System for Analytics 고객과 이전의 Netezza 고객은 가성비, 저렴한 총소유비용 및 간편한 사용성 때문에 최초 시스템을 구매했습니다. 그 이후에는 시스템을 통해 확보할 수 있는 작업 가능한 분석 정보 때문에 추가 시스템을 구매했습니다.
- 독립적인 연구 결과에 따르면, Teradata의 총소유비용은 PureData System for Analytics보다 55% 높은 것으로 나타났습니다.

IBM은 고객이 저렴한 비용으로 데이터에서 작업 가능한 분석 정보를 매우 빠르게 파생시킬 수 있도록 지원하는 데 주력해 온 반면, Teradata는 Netezza®를 통해 작동하는 IBM® PureData™ System for Analytics(Teradata에서는 Netezza로만 지칭함)를 비난하는 데 주력해 왔습니다. 그 이유는, PureData System for Analytics는 업계 최고의 가성비와 단순한 어플라이언스를 제공하도록 구축된 플랫폼을 통해 데이터 웨어하우스 및 분석 환경을 획기적으로 변화시켰기 때문입니다. Teradata는 이와 같은 환경의 동향을 우려하여 문서상에서만 경쟁을 벌이기로 한 것으로 보입니다. Netezza 기술은 대량의 처리 작업을 매우 빠른 속도로 수행하는 기능을 통해 고급 분석 분야에서 신기원을 개척했습니다.

빠르고 신속하며 단순하고 확장성이 뛰어난 Netezza 기술은 성능 또는 예산 측면에서 기존 시스템이 한계에 봉착한 Teradata 고객에게 매우 유용합니다. 이러한 고객은 대부분 데이터와 애플리케이션의 일부 또는 전체를 PureData System for Analytics로 이동했습니다. Teradata에서 우려할 만한 현상이 벌어지고 있는 것입니다.

고객이 Teradata에서 IBM PureData System for Analytics로 이동하는 이유는 다음과 같습니다.

- 가치 실현 시간
- 새 워크로드가 생성되면 빠르게 확장할 수 있는 기능
- 지속적인 유지보수 및 가동 비용 절감
- 쿼리 성능 개선(특히 복잡한 분석 및 ad-hoc 환경에서, 그리고 데이터 볼륨 확장 시)



Teradata는 IBM을 비난하는 데 대부분의 시간을 할애하느라 자사의 솔루션을 개선하는 데는 소홀했습니다. 즉, IBM 솔루션에서 결함을 찾아내기에 바빠 PureData System for Analytics뿐 아니라 자사의 솔루션에 대해서도 몇 가지 주요 요점에 대해 부정확한 평가를 내렸습니다.

Teradata는 최근 발행한 백서에서도 몇 가지 독단적인 주장을 하고 있지만, PureData System for Analytics뿐 아니라 자사 제품에 대한 정보도 정확하게 확인하지 않은 것으로 나타났습니다. 이 백서에서는 Teradata는 “동시 쿼리를 수백만 개¹ 처리할 수 있는 반면 PureData System for Analytics의 “하드 제한은 시스템당 동시 트랜잭션 63개²”라고 주장합니다. 그러나 동시 쿼리 수가 수백만 개임을 뒷받침하는 정보는 없습니다. PureData System for Analytics는 동시 트랜잭션을 63개보다 훨씬 많이 처리할 수 있습니다. 그리고 Teradata 연구소의 선임 컨설턴트에 따르면, Teradata의 2XXXX 시리즈 시스템의 기본 동시성 제한으로 인해 동시 요청 수는 시스템당 불과 52개로 제한되는 것으로 확인되었습니다.³

흥미롭게도 Teradata가 이 백서에서 주장하는 두 번째 요소는 워크로드 관리에 대한 것입니다. 그리고 바로 이 기능을 사용한 결과 Teradata 시스템당 동시 요청 수가 52개로 제한되었습니다. 백서에서는 Teradata의 워크로드 관리 기능이 “합리화된 설정과 제어⁴”를 제공한다고 주장합니다. 그러나 Teradata 시스템에는 서로 다른 네 가지 유형의 제한이 적용되며, 그 중 세 가지는 쿼리를 지연(대기)시키거나 취소할 수 있고 나머지 한 가지는 쿼리를 지연시킬 수만 있습니다. 이러한 제한은 각기 별도로 구성하는 개별 제한입니다. 또한 일부 Teradata 유틸리티는 제한 규칙을 따르지 않으므로 취소만 할 수 있고, 전체 프로시저가 아닌 모든 개별 단계에 제한이 적용되는 저장 프로시저와 같은 기타 유틸리티⁵는 제한 자체가 불가능합니다. 반면 PureData System for Analytics 고객은 대부분의 경우 사용자 또는 그룹 레벨 임계값만 설정하면 되며, 나머지 모든 작업은 시스템에서 처리합니다.

또한 Teradata 백서에서는 사용자에게 “인덱스 생성 옵션이 제공⁶”된다고 되어 있지만, 이러한 옵션은 성능에 전혀 신경을 쓰지 않는 사용자만이 이러한 옵션을 사용할 수 있습니다. Teradata Professional Services 조직에서는 활성 데이터 웨어하우징 성능 관리 분야의 주요 사례를 문서화 및 공개했는데⁷, 여기에는 DBA가 정기적으로 수행해야 하는 다음과 같은 6가지 주요 태스크가 포함되어 있습니다.

1. 성능 데이터 이해
2. 성능 및 용량 모니터링
3. 워크로드 관리 고려
4. 개발 라이프사이클 프로세스 고려
5. 인덱스, 집계 및 기타 여러 튜닝 노브를 사용하여 우선적으로 튜닝 수행
6. 용량 관리

튜닝을 우선적으로 수행해야 하는 플랫폼은 선택할 이유가 없습니다. 또한 지속적인 모니터링 및 튜닝을 수행해야 하는 시스템의 문제는, 성능 문제가 발생한 이후에야 DBA가 문제를 해결할 수 있으며 문제를 해결해도 시스템을 계속 사용할 수 있다는 보장이 없다는 점입니다. 예를 들어 현재는 문제 없이 실행되는 쿼리가 문제로 확인될 수도 있습니다. 그러면 DBA는 쿼리를 수정하기 위해 생성할 적절한 인덱스를 확인하는 데 시간을 할애해야 하며, 해당 인덱스를 생성하는데 더 많은 시간과 시스템 리소스를 투자해야 할 수도 있지만 해당 쿼리는 다시 실행되지 않을 수도 있습니다.

대부분의 PureData System for Analytics 고객과 이전의 Netezza 고객은 가성비, 저렴한 총소유비용 및 간편한 사용성 때문에 최초 시스템을 구매했습니다. 그리고 회사 내의 다른 조직이나 그룹은 PureData System for Analytics를 사용하는 부서에 작업 가능한 분석 정보를 매우 빠르게 제공할 수 있음을 확인한 후에 두 번째, 세 번째, 네 번째의 더 규모가 큰 시스템을 계속 구매했습니다.

Teradata에서는 여러 세대의 노드를 단일 시스템 내에서 사용하도록 허용하지는 않지만, 2XXXX 시리즈의 경우 시스템 기능은 가장 오래된 노드를 기준으로 제한됩니다. 이러한 방식을 공동 상주⁸라고 합니다. 예를 들어 2650, 2690, 2700 노드를 함께 사용하는 경우 압축과 압축 해제에 FPGA를 사용할 수 없습니다. 2650 시리즈에서는 해당 기능을 지원하지 않기 때문입니다. 또한 기존 Teradata 시스템에 노드만 더 추가하더라도 새 시스템을 구성하고(시스템에 새 노드를 구성하는 데 필요한 모든 작업을 중앙 데이터 영역에서 수행) 데이터 재분배를 시작(모든 노드/스토리지 서버로 데이터 이동)하는 과정에서 작동중단시간이 발생합니다. 그리고 데이터 재분배를 수행하는 전체 시간 동안 기타 모든 워크로드의 성능은 계속 저하됩니다. 2013년의 Teradata 고객 대상 Teradata 파트너 컨퍼런스의 세션 1144 “환매 기본 사항”에서는 Teradata 확장 프로세스가 “Teradata 확장: 비싸고 속도가 느리며(계획) 중단 시간이 매우 길다”라고 설명되어 있습니다.⁹

PureData System for Analytics를 사용하는 경우에는 기능도 제한되지 않고 작동중단시간도 발생하지 않는 상태로 동일 노드 세트에 구성된 새 시스템으로 업그레이드할 수 있습니다.

PureData System for Analytics에서는 현장 프로그램 가능 게이트 배열(FPGA) 액셀러레이터 기술을 사용하므로 시스템에서 최신 세대, 고가의 구성요소 및 대량의 RAM을 사용하지 않아도 데이터 웨어하우스 및 분석 플랫폼의 성능을 절대 최대치까지 끌어올릴 수 있습니다. PureData System for Analytics는 대량의 데이터를 처리 영역으로 이동하는 대신 처리 영역을 데이터로 이동합니다. Teradata 시스템의 경우 데이터에 대해 작업을 수행하려면 데이터를 메모리로 이동해야 한다는 단점을 보완하기 위해 고가의 CPU 및 더 많은 메모리를 사용해야 합니다.

ITG의 실제 고객 대상 연구 결과에 따르면, Teradata의 총소유비용은 PureData System for Analytics보다 55% 높은 것으로 나타났습니다.¹⁰ 기능은 부족하고 가격은 더 비싼 시스템을 선택할 이유는 없습니다. 두 기술을 모두 평가해 보려는 신규 고객의 경우 최저 비용에 최고의 성능을 제공하는 시스템을 선택하십시오. 일부 리소스를 환매하기 위해 몇 주에 걸쳐 시스템을 튜닝해야 하거나 성능을 개선하기 위해 시스템을 확장해야 하는 기존 Teradata 고객이라면 타사의 제품을 고려해 보는 것이 좋습니다. IBM의 제품은 작업 속도, 성능, 간편한 사용 및 저렴한 비용을 보증합니다.

많은 고객들이 Teradata 및 기타 데이터베이스의 워크로드를 PureData System for Analytics로 마이그레이션했으며 매우 만족스러운 결과를 얻을 수 있었습니다. 다른 고객들이 PureData System for Analytics를 사용하여 분석 정보를 매우 빠르게 확보한 방법에 대해 알아보려면 ibm.com/software/kr/data/puredata/analytics 웹 사이트를 방문하거나 IBM 팀에 문의하십시오.

IBM PureData System for Analytics 정보

IBM PureData System for Analytics는 데이터베이스, 서버 및 스토리지를 쉽게 관리할 수 있는 단일 어플라이언스로 통합합니다. 이러한 어플라이언스에서는 최소한의 설정 및 지속적인 관리만 수행하면 되며, 보다 일관성 있는 분석 성능을 더 빠르게 달성할 수 있습니다. IBM PureData System for Analytics는 데이터가 상주하는 어플라이언스에서 모든 분석 작업을 통합하여 비즈니스 분석을 크게 간소화함으로써 업계 최고 수준의 성능을 제공합니다.

ibm.com/software/kr/data/puredata/analytics를 방문하면 IBM의 전문가 통합 시스템 제품군을 통해 모든 단계를 간편하게 수행하고 조직의 진정한 비즈니스 가치를 실현할 수 있는 방법을 확인할 수 있습니다.



© Copyright IBM Corporation 2014

Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
2015년 3월

IBM, IBM 로고 및 ibm.com 전 세계 여러 국가에 등록된 International Business Machines Corp.의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보"(www.ibm.com/legal/copytrade.shtml)에 있습니다.

Netezza는 IBM 회사인 IBM International Group B.V.의 등록 상표입니다.

이 문서는 최초 발행일을 기준으로 하며, 통지 없이 언제든지 변경될 수 있습니다. IBM이 영업하는 모든 국가에서 모든 오퍼링이 제공되는 것은 아닙니다.

이 문서의 정보는 상품성, 특정 목적에의 적합성에 대한 보증 및 타인의 권리 침해에 대한 보증 및 타인의 권리 침해에 대한 보증이나 조건을 포함하여(단, 이에 한하지 않음) 명시적이든 묵시적이든 일체의 보증 없이 "현상태대로" 제공됩니다. IBM 제품에 대한 보증은 제품의 준거 계약 조항에 의거하여 제공됩니다.

법률과 규정을 준수하는지 확인해야 할 책임은 고객에게 있습니다. IBM은 법률 자문을 제공하지 않으며 IBM의 서비스나 제품을 통해 관련 법률이나 규정 에 대한 고객의 준수 여부가 확인된다고 진술하거나 보증하지 않습니다.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다. 사용 가능한 실제 스토리지 용량은 압축 및 비압축 데이터 모든 경우에 대해, 언급한 것과 다르거나 더 작을 수 있습니다.

- 1 Netezza가 효율적이지 않은 5가지 이유(2014년 2월 21일).
<http://tinyurl.com/q2hn6dl>에서 검색
- 2 Netezza가 효율적이지 않은 5가지 이유(2014년 2월 21일).
<http://tinyurl.com/q2hn6dl>에서 검색
- 3 제한 작동 방식 - Teradata 동시성 제어 세부 정보 - 파트너, Teradata User Group(2013년 10월 20~24일, 델러스)
- 4 Netezza가 효율적이지 않은 5가지 이유(2014년 2월 21일).
<http://tinyurl.com/q2hn6dl>에서 검색
- 5 제한 작동 방식 - Teradata 동시성 제어 세부 정보 - 파트너, Teradata User Group(2013년 10월 20~24일, 델러스)
- 6 Netezza가 효율적이지 않은 5가지 이유(2014년 2월 21일).
<http://tinyurl.com/q2hn6dl>에서 검색
- 7 최고 성능 - Teradata Magazine
- 8 <http://www.teradatamagazine.com/v10n01/tech2tech/system-evolution/>
- 9 세션 1137, 파트너 - Teradata User Group(2013년 10월 20~24일, 델러스)
- 10 IBM PureData System for Analytics의 비용/이점 사례: Teradata Data Warehouse Appliance와의 비용 및 가치 실현 시간 비교
(<http://tinyurl.com/kzfq28j>)



재활용하십시오.