

Netezza 데이터 어플라이언스 아키텍처:

고성능 데이터 웨어하우징 및 분석 플랫폼

서론

기업의 성공 여부는 올바른 의사결정을 내릴 수 있도록 적시에 최적의 정보를 보유하고 있는지에 달려 있습니다. 기회를 효과적으로 활용하고 시간이나 자원을 덜 사용하는 정보라 할지라도 조직에 위기를 초래할 수 있습니다. 그러나 가장 적합한 조치를 판단하는 데 중요한 정보를 찾는 것은 불확실성에도 불구하고, 결과를 예측하고 최신 경향을 파악하거나 최선책을 찾기 위해 수십억 개의 데이터로 연결된 페타바이트 규모의 데이터를 분석하는 과정을 의미합니다. 필요 시 이러한 유형의 인텔리전스를 확보할 수 있는 기업은 경쟁사보다 상황에 신속하게 대처하고 올바른 의사결정을 내릴 수 있습니다.

분석 업무를 끊임없이 개선하는 기업은 비즈니스의 모든 영역에 긍정적인 결과를 가져오는 인텔리전스라는 뜻밖의 혜택을 누릴 수 있습니다. 하지만, 중요한 정보를 긴급하게 찾아야 하는 경우 이 정보를 제공하는 플랫폼은 마지막에 고려해야 합니다. 이러한 플랫폼은 방해되는 엄청난 볼륨의 워크로드를 간단하게 처리할 수 있도록 조명 스위치처럼 단순하고 신뢰할만하며 즉각적으로 작동해야 합니다. 또한 점차 복잡해지는 워크로드를 수행하는 사용자가 늘어나고 데이터 볼륨이 급속하게 증가하는 경우에도, 성능을 그대로 유지하는 기술에 기초하여 오랜 기간 사용할 수 있고 최소한의 총 소유 비용(TCO)만 소비해야 합니다.

어플라이언스의 단순성에 비해 우수한 성능

Netezza는 향후 몇 년 동안 데이터 웨어하우스와 분석 업무를 단순한 어플라이언스를 갖춘 플랫폼으로 대체할 것이며 이 플랫폼은 업계 최고의 가격 대비 성능을 제공할 수 있도록 구축되었습니다. Netezza는 고급 분석의 새로운 기준이며, 어떤 한계나 상황에서도 매우 빠른 속도로 엄청난 양의 처리 과제를 수행할 수 있습니다. 사용자와 조직에게 있어 Netezza는 요구사항에 관계없이 이를 필요로 하는 모든 대상에게 최고의 인텔리전스를 의미합니다.

Netezza 데이터 웨어하우스와 분석 어플라이언스는 업계 최고의 가격 대비 성능을 제공한다는 원칙에 따라 혁신적으로 설계되었습니다. Netezza는 신속한 분석을 위한 목적 기반 어플라이언스로, 뛰어난 성능을 발휘할 수 있는 이유는 가장 강력하고 값비싼 구성요소가 아니라 적절하게 결합된 구성요소가 상호 협력을 통해 성능을 극대화하는 방식으로 구현되기 때문입니다. MPP(Massively Parallel Processing) 스트림은 멀티코어 CPU와 Netezza 고유의 FAST™(FPGA Accelerated Streaming Technology) 엔진을 결합하여 훨씬 고가의 시스템과 비교할 수 없는 성능을 제공합니다. 또한 이 시스템은 사용이 쉬운 어플라이언스로, 색인화 또는 튜닝 작업을 수행하지 않고도 사용하는 즉시 놀라운 결과를 제공합니다. 어플라이언스의 단순성은 애플리케이션 개발 영역까지도 적용되며 빠른 혁신과 기능을 통해 가장 광범위한 사용자와 프로세스를 대상으로 고성능 분석을 실현할 수 있게 해줍니다.

이 백서는 Netezza의 AMP™(Asymmetric Massively Parallel Processin™) 아키텍처를 소개하고 이 시스템이 어떻게 쿼리와 분석을 조율하여 전례 없이 신속하게 업무 수행이 가능한지에 대해 설명합니다. Netezza 소프트웨어와 하드웨어가 어떤 방식으로 결합되어 모든 주요 구성요소의 이용률을 극대화하는지, 그리고 수만 명의 사용자가 대용량의 데이터 쿼리 작업을 수행할 수 있도록 최적화된 시스템이 실제로 어떻게 작동하는지 알아보겠습니다. Netezza는 비교할 수 없는 가격 대비 성능을 지닌 고유한 데이터 웨어하우스 및 분석 플랫폼으로, 오늘날의 요구사항과 미래의 해결 과제 모두를 충족시켜주는 제품입니다.

아키텍처 원칙

Netezza 어플라이언스는 분석 프로세스에 맞게 최적화되고 유연한 확장을 위해 설계된 컴팩트 시스템에서 데이터베이스, 프로세스 및 스토리지를 통합합니다. 이러한 시스템 아키텍처는 다음과 같이 업계 최고의 가격 대비 성능을 유지하는 Netezza의 핵심 원칙에 기초합니다.

데이터 소스와 밀접한 처리 방식

Netezza 아키텍처는 대용량 데이터 세트에 작업을 수행할 경우, 반드시 필요 시에만 데이터의 위치를 옮긴다는 컴퓨터 공학의 기본 원칙에 근거하고 있습니다. Netezza는 가능한 한 데이터 스트림 초기 단계에, 데이터를 디스크에서 최대한 빨리 읽어올 수 있도록 관련 없는 데이터를 필터링하는 데 FPGA(Field Programmable Gate Array)라는 제품 구성요소를 활용함으로써 이러한 원칙을 완벽하게 적용합니다. 데이터 소스와 밀접한 데이터 제거 프로세스는 IO 병목현상을 제거하고 대용량 데이터를 처리하는 CPU, 메모리 및 네트워크와 같은 다운스트림 구성요소의 공간을 확보해 줌으로써 시스템 성능을 엄청나게 증가시킵니다.

안정된 대량 병렬식 아키텍처

Netezza 아키텍처는 페타바이트급 데이터를 단시간에 분석할 수 있도록 특별히 설계된 어플라이언스를 구축하기 위해 SMP와 MPP라는 최고의 요소를 결합합니다. 프로세서, FPGA, 메모리 및 네트워크를 포함한 아키텍처의 모든 구성요소는 디스크처럼 신속하게 데이터를 제공하고 비용과 전력 소비를 줄이기 위해 신중하게 선별되고 최적화되어 있습니다. Netezza 소프트웨어는 이러한 구성요소를 데이터 스트림에서 파이프라인 방식으로 동시에 작동하도록 조율함으로써 MPP 노드를 통해 이용률을 극대화하고 처리량을 최대화합니다. 기본적인 성능 외에도 안정된 아키텍처는 선형 확장성을 제공하므로 천여 개의 처리 스트림을 동시에 실행할 수 있으며 TCO 부담이 적습니다.

고급 분석용 플랫폼

소스와 밀접한 데이터 처리 및 MPP 원칙은 대용량 데이터 세트의 고급 분석에도 동일하게 적용할 수 있습니다. Netezza는 조잡한 병렬식 또는 그리드 프로그래밍 과정을 거치지 않고도 복잡한 비SQL 알고리즘을 MPP 스트림의 처리 요소에 쉽게 적용할 수 있게 해줍니다. 대용량의 데이터를 복잡도에 관계없이 "원활하게" 분석할 수 있는 기능은 데이터를 각 하드웨어로 이동시키는 시간과 비용을 절감시켜 줍니다. 또한, Netezza는 데이터 크기에 따라 성능을 가속화할 수 있어 데이터 웨어하우스와 고급 분석을 통합하는 이상적인 플랫폼입니다.

단순한 어플라이언스

Netezza는 정규 업무를 자동화 및 간소화하므로 사용자가 플랫폼의 복잡성을 고려할 필요가 없습니다. 다른 어플라이언스의 특징을 디자인에 적용할 경우에도 항상 단순성이 우선시됩니다. 타사 솔루션과 달리 Netezza는 요구사항이 많은 쿼리와 단시간에 완료해야 하는 복잡한 워크로드를 처리하며 다른 시스템에 필요한 튜닝 작업 없이 설치 즉시 작동됩니다. 설치, 업그레이드, 고가용성 및 비즈니스 연속성 확보 등 일반적으로 시간이 많이 걸리는 업무도 매우 단순화되므로 시간과 자원을 절약할 수 있습니다.

혁신의 가속화와 성능 향상

Netezza 아키텍처의 핵심 목표 중 하나는 장기적인 가격 대비 성능 향상과 경쟁사에 앞서 혁신적인 기능을 제공하는 것입니다. 개방형 블레이드 기반의 구성요소를 통해 Netezza는 강화된 기술을 단시간에 통합할 수 있는 동시에 FPGA의 뛰어난 효과, 안정적인 하드웨어 구성 및 긴밀하게 연결된 지능형 소프트웨어를 결합하여 각 요소를 결합한 성능보다 더 우수한 성능을 제공합니다. 실제로, Netezza 제품은 출시 이래 타사 공급업체 제품에 비해 2년마다(무어의 법칙의 2배¹) 성능이 4배 이상 향상되었습니다.

유연한 구성과 최고의 확장성

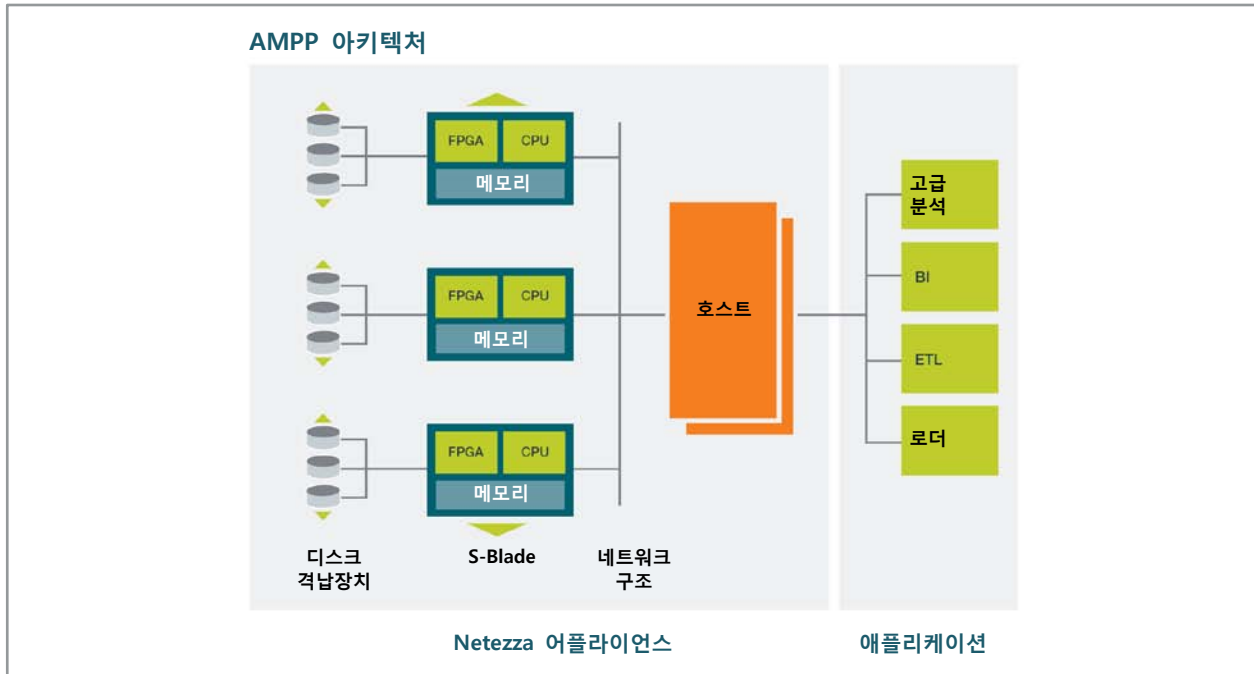
Netezza는 쿼리 가능한 사용자 데이터의 경우 수백 기가바이트에서 수십 페타바이트 수준까지 모듈식으로 확장할 수 있습니다. 이 시스템 아키텍처는 적응성이 뛰어나 데이터 웨어하우스와 분석 업무의 다양한 요구사항에 맞게 서비스를 제공할 수 있습니다. 개방형 블레이드 기반의 구성요소를 활용하여 성능 또는 스토리지 중심의 요구사항에 맞게 구성에서 디스크 프로세서와 메모리 속도를 쉽게 변경할 수 있습니다. 또한 이 아키텍처는 업무 위주의 애플리케이션에 빠른 실시간 분석을 제공하는 메모리 기반 시스템도 지원할 수 있습니다.

다음 페이지에서는 Netezza가 이러한 원칙을 어떻게 실제로 적용하는지 알아보겠습니다.

시스템의 기본 구성요소

Netezza의 성능 이점에 있어 핵심적인 요소는 쿼리를 처리하기 위해 SMP 프론트엔드와 SN(shared-nothing) MPP 백엔드가 결합된 독특한 AMPP 아키텍처입니다. 이 아키텍처의 각 구성요소는 안정된 전체 시스템을 구축하기 위해 신중하게 선별된 후 통합되었습니다. 각각의 처리 요소는 다중 데이터 스트림에서 작동하고 관련 없는 데이터는 신속하게 필터링합니다. 이렇게 사용자 정의된 천여 개의 MPP 스트림은 워크로드를 "분산하고 해결"하기 위해 결합됩니다.

1. Intel의 공동 창립자인 Gordon Moore씨는 1965년도에 칩에 내장된 트랜지스터의 수는 2년마다 2배로 늘어날 것이라고 예측했습니다. 일반적으로 소프트웨어 애플리케이션의 성능이 오랜 시간 동안 향상되려면 프로세서의 성능 또한 개선되어야 합니다.



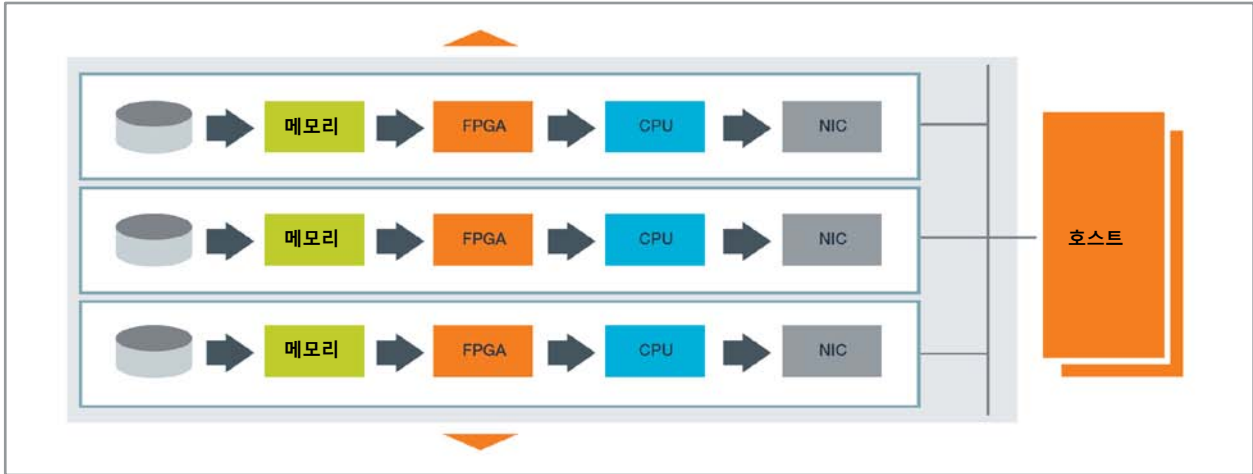
어플라이언스의 주요 구성요소

- Netezza 호스트: SMP 호스트는 고성능 Linux 서버로, 고가용성을 위한 활성-수동 구성 두 가지로 설정되어 있습니다. 활성 호스트는 외부 도구와 애플리케이션에 대한 표준 인터페이스를 나타냅니다. 또한 SQL 쿼리를 스니펫이라는 실행 가능한 코드 세그먼트로 컴파일하고 최적화된 쿼리 계획을 수립하며 실행하기 위해 이 스니펫을 MPP 노드로 전달합니다.
- S-Blade(Snippet Blade): S-Blade는 지능형 처리 노드로, 어플라이언스의 터보형 MPP 엔진을 구성합니다. 각 S-Blade는 독립 서버로, 강력한 멀티코어 CPU, 멀티엔진 FPGA 및 기가바이트급 RAM을 보유하고 있으며 이 모든 요소가 균형 있게 결합되어 동시에 작동하므로 최고의 성능을 제공합니다. CPU 코어는 풍부한 헤드룸으로 설계되어 있어 고급 분석 애플리케이션의 대용량 데이터를 대상으로 복잡한 알고리즘도 실행합니다.
- 디스크 격납장치: 디스크 격납장치에는 RAID로 보호되는 고밀도의 고성능 디스크가 들어 있습니다. 각 디스크에는 데이터베이스 테이블 형태로 된 데이터 조각이 들어 있습니다. 디스크 격납장치는 고속의 상호 연결을 통해 S-Blades에 연결되며 이 상호 연결을 통해 Netezza의 모든 디스크에서 데이터를 최대한 신속하게, S-Blade로 동시에 전송할 수 있습니다.
- 네트워크 구조: 모든 시스템 구성요소는 고속의 네트워크 구조로 연결되어 있습니다. Netezza는 사용자 정의된 IP 기반 프로토콜을 실행하는데 이 프로토콜을 통해 네트워크 구조의 모든 광대역을 완벽하게 활용하고 네트워크 트래픽이 정체되거나 폭발적으로 증가한 상황에서도 혼잡을 줄일 수 있습니다. 네트워크는 천 개 이상의 노드까지 확장하도록 최적화되어 있으며 각 노드는 대용량 데이터를 다른 노드에 동시에 전송할 수 있습니다.

참고: 모든 시스템 구성요소는 중복될 수 있습니다. 호스트가 활성-수동형이라도 어플라이언스의 기타 모든 구성요소는 핫스왑(hot-swappable)이 가능합니다. 사용자 데이터는 99.99% 이상의 가용성을 확보함과 동시에 완전한 미러링이 가능합니다.

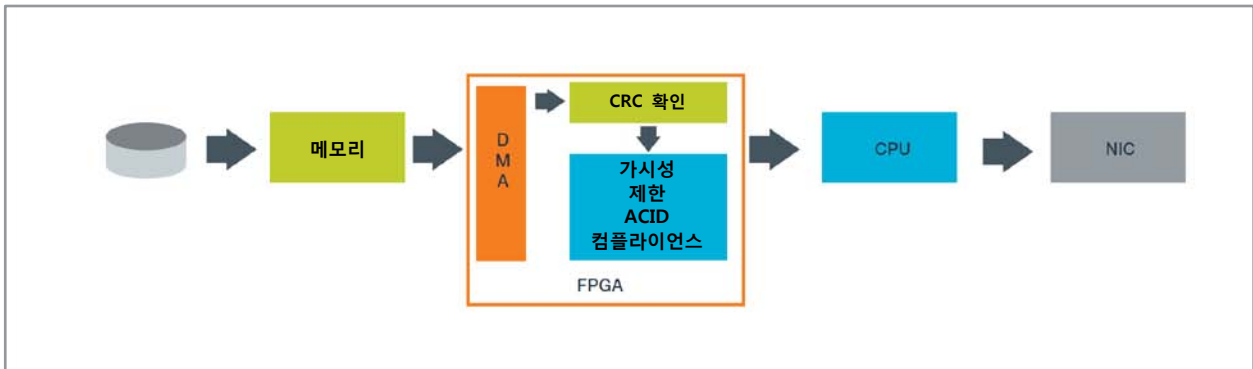
최고의 성능을 발휘하는 S-Blade

스니펫 프로세서(다수 중 하나): 제품 구성요소와 Netezza 소프트웨어는 결합을 통해 각 MPP 노드에서 최고의 처리량 수준에 도달합니다. 스토리지 배열에 연결된 전용의 고속 인터커넥터는 데이터를 디스크에서 읽을 때처럼 메모리에 신속하게 전달해줍니다. 메모리에 압축 데이터를 캐싱할 때 사용되는 스마트 알고리즘은 디스크에 액세스하지 않고도 가장 많이 액세스되는 데이터를 메모리에서 즉시 가져올 수 있도록 해줍니다. FPGA에서 병렬 방식으로 실행되는 FAST 엔진은 물리적인 속도로 테이블 데이터의 95-98%를 압축해제하고 필터링하며 쿼리 응답과 관련 있는 데이터만 보관합니다. 스트림에 남아있는 데이터는 CPU 코어에 의해 동시에 처리되며 병렬 방식으로 실행됩니다. 이 프로세스는 Netezza에서 실행되는 천 개 이상의 병렬 스니펫 프로세서를 통해 반복됩니다. 그 결과 크기별로 고가의 시스템보다 우수한 성능을 발휘합니다.



S-Blade 성능 극대화: Netezza FAST 엔진의 강력한 성능

FPGA는 Netezza의 가격 대비 성능의 이점을 극대화하는 주요 요소입니다. 각 FPGA에는 데이터 스트림에서 필터링과 변환 기능을 수행하는 임베디드 엔진이 탑재되어 있습니다. 동적 재구성 기능을 갖춘 FAST 엔진은 소프트웨어 전체에서 수정 또는 확장이 가능합니다. 이 엔진은 쿼리 실행 시 제공되는 매개변수를 통해 *모든 스니펫*을 대상으로 사용자 정의할 수 있으며 고속의 DMA(Direct Memory Access) 모듈이 제공하는 데이터 스트림에서 작동합니다.



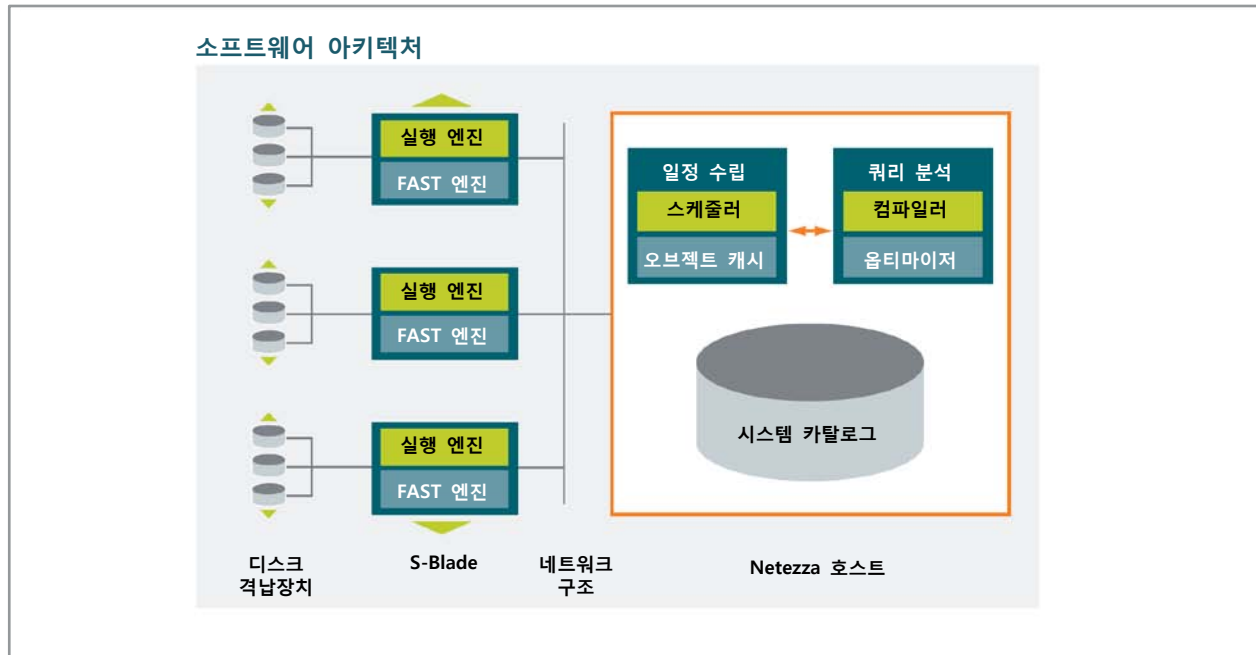
FAST 엔진 구성요소

- 압축 엔진: Netezza의 혁신 기능으로, 시스템 성능을 4 ~ 8배속으로 증가시킵니다. 이 엔진은 데이터를 연결 속도로 압축 해제하여 디스크의 각 블록을 메모리의 4 ~ 8 블록으로 변환시켜줍니다. 따라서 디스크와 같이 데이터 웨어하우스에서 가장 속도가 느린 구성요소라도 속도가 엄청나게 향상될 수 있습니다.
- 프로젝트 및 제한 엔진: SQL 쿼리의 SELECT 및 WHERE 절에 있는 매개변수를 기준으로 열과 행을 각각 필터링하여 성능을 더욱 강화시켜 줍니다.
- 가시성 엔진: Netezza에서 ACID(Atomicity, Consistency, Isolation 및 Durability) 컴플라이언스를 스트리밍 속도로 유지하는 중요한 역할을 담당합니다. 이 엔진은 쿼리에서 "표시"되지 않아야 하는 행을 필터링합니다. (예: 아직 커밋되지 않은 트랜잭션에 속한 행)

Netezza FAST 엔진은 Netezza 소프트웨어의 기능을 개선하여, 향후 추가될 혁신적인 새로운 기능을 위한 확장형 프레임워크를 제공합니다. 새로운 기능은 시스템 성능, 보안 및 신뢰성을 더욱 강화시켜 줍니다.

Netezza의 쿼리 조율 기능

Netezza 제품의 하드웨어 구성요소와 지능형 시스템 소프트웨어는 밀접하게 연결되어 있습니다. 소프트웨어는 단순 쿼리, 복잡한 임시 쿼리 또는 심층적인 분석에 관계없이 어플라이언스의 하드웨어 기능을 모두 활용할 수 있게 설계되었고 다양한 혁신 기능을 통합해 뛰어난 성능을 발휘합니다. 이 절에서는 시스템에 구축되는 인텔리전스에 대해 단계적으로 알아보겠습니다.



Netezza 소프트웨어 구성요소 항목

- 고급 병렬 옵티마이저: 쿼리를 보다 효율적으로 실행하도록 변환하고 모든 처리 노드에서 각각의 구성요소를 최대한 활용할 수 있게 해줍니다.
- 지능형 스케줄러: 워크로드에 상관없이 처리량이 가장 많은 경우에도 시스템이 지속적으로 실행되도록 합니다.
- 강력한 스니펫 프로세서: 다중 쿼리를 효율적으로 실행하고 복잡한 분석 기능을 동시에 실행합니다.
- 스마트 네트워크: Netezza를 통해 대용량 데이터를 원활하게 전송합니다.

사용자가 쿼리를 제출하는 시점부터 시작하여 이러한 요소들이 어떻게 결합하고 작동하는지 알아보겠습니다. 기술 지식이 많은 사용자는 Netezza가 타사 데이터 웨어하우스 시스템과는 매우 다른 방식으로 쿼리를 처리한다는 점을 알 수 있습니다.

최적화된 쿼리 계획 수립...

호스트가 쿼리를 컴파일하고 Netezza의 AMPP 아키텍처에 맞게 최적화된 쿼리 실행 계획을 수립합니다. Netezza 옵티마이저의 인텔리전스는 시스템의 최대 강점 중 하나입니다. 옵티마이저는 시스템에 있는 모든 MPP 노드를 이용해 쿼리에서 참조한 모든 데이터베이스 테이블의 세분화된 최신 통계를 수집합니다. 이 통계치의 대부분은 낮은 오버헤드에서 쿼리 실행 시 수집되며 쿼리별로 개별화된 통계치를 적시에 생성합니다. Netezza 어플라이언스의 특징은 상호 커뮤니케이션이 가능한 통합 구성요소를 통해 비용 기반의 옵티마이저가 작업과 관련된 디스크, 처리 및 네트워크 비용을 보다 정확하게 측정할 수 있게 해준다는 점입니다. 옵티마이저는 대략적인 데이터보다는 정확한 데이터에 근거하여 모든 구성요소를 효율적으로 이용하는 쿼리 계획을 수립할 수 있습니다.

옵티마이저 인텔리전스: Join 순서 계산

옵티마이저 인텔리전스에는 복합 Join에서 최적의 Join 순서를 결정하는 기술이 있습니다. 예를 들어 여러 개의 작은 테이블을 큰 팩트 테이블에 Join할 경우, 옵티마이저는 각각의 S-Blade에 작은 테이블을 전달하도록 선택할 수 있지만 큰 테이블인 경우 모든 스니펫 프로세서 전체에 분배하도록 설정할 수 있습니다. 이 방식은 데이터 이동을 최소화하는 동시에 AMPP 아키텍처를 이용하여 Join을 병렬화합니다.

옵티마이저는 이 통계 수치를 활용하여 프로세스 시작 전에 쿼리를 변환하여 데이터 웨어하우스 시스템의 성능을 저하시키는 디스크 I/O와 데이터 이동을 최소화합니다. 옵티마이저가 수행하는 변환 작업에는 다음이 포함됩니다.

- 올바른 Join 순서 결정
- 표현식 다시 작성
- SQL 조작의 중복성 제거

스니펫으로 변환...

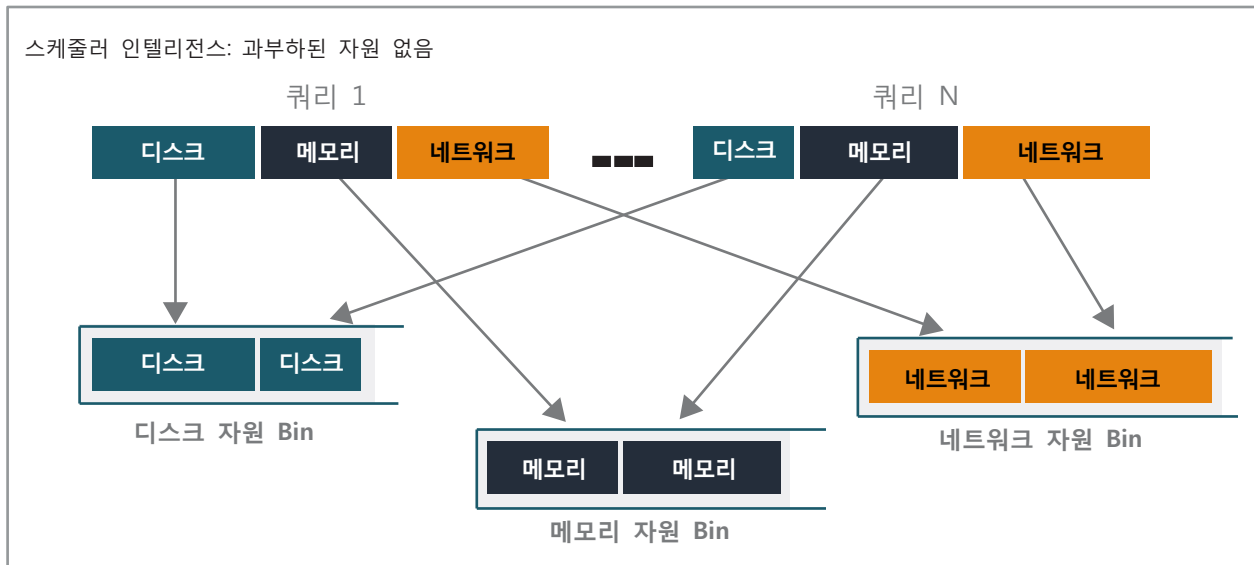
컴파일러는 쿼리 계획을 스니펫이라는 실행 가능한 코드 세그먼트로 전환하는데 스니펫이란, 어플라이언스의 데이터 스트림 전체에서 스니펫 프로세서가 병렬 방식으로 실행하는 쿼리 세그먼트를 말합니다. 각 스니펫에는 2가지 요소가 있는데 하나는 개별 CPU 코어로 실행되는 컴파일된 코드이며 다른 하나는 FAST 엔진을 사용자 정의하여 특정 스니펫을 필터링하는 FPGA 매개변수 세트입니다. 스니펫별(snippet-by-snippet) 사용자 정의를 통해 Netezza는 실제로, 개별 쿼리에 맞춰 신속하게 최적화된 하드웨어 구성을 제공할 수 있습니다.

컴파일러 인텔리전스: 오브젝트 캐시

이 호스트는 오브젝트 캐시라는 기능을 통해 쿼리 성능을 더욱 가속화합니다. 이것은 매개변수의 변경을 지원하는 기존에 컴파일된 스니펫 코드의 대형 캐시입니다. 예를 들어, "where name='bob'" 절이 포함된 스니펫이 "where name='jim'" 절이 포함된 스니펫과 동일한 컴파일 코드를 활용하지만 다른 이름으로 된 설정을 사용합니다. 이러한 방식을 통해 스니펫의 99% 이상에 해당하는 컴파일 단계를 생략할 수 있습니다.

적시에 실행하도록 일정 수립...

Netezza 스케줄러는 다양한 사용자의 목적에 맞게 복잡한 워크로드 간의 작업을 균등하게 조정하는 동시에 최대의 이용률과 처리량을 유지합니다. 스케줄러는 S-Blades에서 스니펫을 언제 실행할지 결정해야 할 경우 쿼리 우선순위, 크기 및 자원의 가용성과 같은 여러 요인을 고려합니다. 어플라이언스 아키텍처는 스케줄러가 시스템의 각 구성요소에서 가져온 자원의 가용성에 대해 보다 많은 정확한 최신 정보를 수집할 수 있게 해줍니다. 스케줄러는 고급 알고리즘을 통해 100%에 가까운 디스크 광대역을 활용하고 메모리와 네트워크 자원으로 인해 과부하가 발생하지 않으며 시스템이 효율성을 유지하도록 보장함으로써 시스템의 처리량을 극대화합니다. 이것은 Netezza 제품이 지닌 가장 중요한 특징으로, 시스템이 과부하 상태에서도 최상의 처리 능력으로 성능을 유지할 수 있도록 지원합니다.



스케줄러에 녹색불이 켜지면 스니펫이 지능형 네트워크 구조를 통해 모든 스니펫 프로세서에 전달됩니다.

병렬 방식으로 실행...

모든 S-Blade의 각 스니펫 프로세서는 명령에 따라 스니펫의 일부를 실행해야 합니다. 호스트 스케줄러 외에도 스니펫 프로세서마다 우선적으로 사용할 수 있는 고유한 스마트 스케줄러가 있는데 이것은 여러 쿼리에서 가져온 스니펫을 동시에 실행시킵니다. 이 스케줄러는 사용자 또는 그룹을 위해 준비된 쿼리와 자원의 우선순위를 고려하여 언제 그리고 얼마나 오래 동안 특정 스니펫을 실행하기 위해 계획해야 하는지 결정합니다. 스케줄러는 다음과 같이 실행됩니다.

1. 각 스니펫 프로세서의 프로세서 코어는 쿼리 스니펫에 포함된 매개변수로 FAST 엔진을 구성하고 데이터 스트림을 설정합니다.
2. 스니펫 프로세서는 디스크 스캔 시간을 단축시키는 Netezza의 혁신 기능인 ZoneMap™ 을 활용하여 디스크 배열에서 테이블 데이터를 메모리로 읽어옵니다. 스니펫 프로세서는 또한 데이터 블록 디스크에 액세스하기 전에 캐시로부터 정보를 가져와 해당 데이터가 메모리에 있는 경우 스캔을 중지합니다.
3. FPGA는 데이터 스트림에서 작동합니다. 처음에는 연결 속도로 데이터 스트림의 압축을 해제하여 최대 4 ~ 8배속으로 데이터 스트림을 가속화합니다.
4. FAST 엔진은 쿼리와 관계없는 데이터를 필터링합니다. CPU 코어에서 동시에 처리할 수 있도록 나머지 데이터는 메모리에 다시 저장됩니다. 이 데이터는 원래 스트림보다 2 ~ 5% 작은 조각으로, 프로세서 코어의 실행 시간을 크게 단축시켜 줍니다.
5. 프로세서 코어는 데이터 스트림을 선택하고 정렬, 조인(Join) 및 집계와 같은 주요 데이터베이스 작업을 수행합니다. 또한 고급 분석 처리를 위해 스니펫 프로세서에 임베드된 복합 알고리즘을 적용합니다.
6. 각 스니펫 프로세서에서 가져온 결과는 메모리에 모인 다음 전체 스니펫에 대한 2차 결과가 생성됩니다. 이 프로세스는 병렬 방식으로 실행되는 수백 또는 수천 개의 쿼리 스니펫을 활용하여 천 개 이상의 스니펫 프로세서 전체에서 동시에 반복됩니다.

ZoneMap의 가속화 기능 - Netezza의 색인 방지 기능

ZoneMap의 가속화 기능은 데이터 웨어하우스에 있는 원래의 행 순서를 이용하여 크기 순으로 성능을 가속화합니다.

이 기술은 쿼리의 시작과 끝을 벗어난 열 값을 가진 행을 스캐닝하지 못하도록 방지합니다. 예를 들어 테이블에 2년 치 주간 레코드(최대 100주)가 있고 쿼리는 1주에 해당하는 데이터만 검색할 경우, ZoneMap의 가속화 기능을 통해 최대 100배까지 성능이 향상될 수 있습니다. 색인과 달리, ZoneMap은 관리 부담이 없고 각 데이터베이스 테이블에 자동으로 생성되고 업데이트됩니다.

결과 제출!

이제 모든 스니펫 프로세서에 조합해야 할 스니펫에 대한 결과가 생성되었습니다. 스니펫 프로세서는 지능형 네트워크 구조를 통해 호스트와 프로세서 간에 유연한 통신이 가능하므로 중간에 계산과 집계를 수행할 수 있습니다.

네트워크 인텔리전스:

예측 가능한 성능과 확장성

Netezza의 맞춤형 네트워크 프로토콜은 대용량 데이터 웨어하우징과 연관된 데이터 볼륨과 트래픽 패턴에 맞게 특수 설계되었습니다. Netezza 프로토콜은 네트워크 대역폭에 과부하를 일으키지 않으면서 대역폭을 최대한 활용할 수 있도록 지원하며 회선 속도와 비슷한 예측 가능한 성능을 제공합니다.

트래픽은 다음과 같은 3가지 영역에서 활발합니다.

- 호스트에서 스니펫 프로세서(1 ~ 1000+, 브로드캐스트 모드)로 이동
- 스니펫 프로세서에서 호스트(1000+ ~ 1, S-Blades에서 통합된 시스템 랙 레벨인 경우)로 이동
- 스니펫 프로세서(1000+ ~ 1000+, 중간 처리 시 대용량 데이터가 원활하게 이동하는 경우) 간 이동

호스트는 스니펫 프로세서에서 가져온 중간 결과를 조합하여, 최종 결과를 컴파일한 후 사용자 애플리케이션에 전송합니다. 한편, 기타 쿼리는 여러 완료 단계를 거쳐 시스템을 통해 전송됩니다.

정보를 필요로 하는 누구에게나 제공되는 정보

최고의 솔루션은 때로 가장 규모가 크고 값비싼 제품이 아니라 가장 똑똑한 설계를 기초로 합니다. Netezza는 스트리밍 처리 방식이 다른 분석 및 데이터 웨어하우징 시스템이 활용하는 모든 기존의 컴퓨팅 아키텍처보다 더 고유한 장점을 지니고 있다는 사실을 이미 인지하고 있었습니다.

그 결과 대형 시스템보다 작지만, 수천 명의 사용자가 동시에 이용하는 대용량 데이터와 다양한 워크로드로 구성된 복잡한 알고리즘을 매우 빠른 속도로 실행하는 성능을 갖춘 컴팩트한 어플라이언스가 출시되었습니다. 이 어플라이언스의 처리 성능은 기업의 성공을 돕는 Netezza의 고유한 플랫폼과 같은 기타 기능으로 보완할 수 있습니다.

- **사용의 용이성:** Netezza 제품은 어플라이언스와 같이 자체 관리되며 항상 최고의 처리량을 달성합니다. 시스템 소프트웨어는 수작업이 필요 없으므로 사용하기 쉽습니다.
- **기업 전체에서 개선된 의사결정:** 임베디드 기능은 최소한의 개발 노력만 필요로 하므로 데이터베이스 분석의 새로운 장을 열어줍니다. 대용량 데이터 전송 시 서버 하드웨어를 따로 준비할 필요가 없고 시간을 절약할 수 있습니다. 매우 신속하게 결과를 전달하고 주요 비즈니스 인텔리전스를 필요로 하는 조직 내 모든 계층의 직원들에게 제공할 수 있습니다.
- **미래를 대비한 민첩성:** 이 시스템은 오늘날의 과제 해결만을 위한 것이 아니라 다가올 미래를 대비하기 위한 것입니다. 수십 배의 페타바이트급 데이터까지 선형 확장이 가능하며 무어의 법칙의 영향을 받는 기존 가속화에 비해 훨씬 더 빨리 성능을 가속화할 수 있습니다.

Netezza 제품은 사용자와 기업이 뛰어난 성능은 그대로 유지한 상태에서 가장 명확한 의사결정을 내릴 수 있게 해줍니다. 하지만, 설명으로만 이해하는 것보다는 Netezza를 제대로 평가하려면 실제로 어떻게 작동하는지 확인하는 것이 좋습니다. Netezza보다 데이터를 더 효율적으로 활용하는 제품은 없다고 자부하기 때문입니다. **(N)**

면책 범위

"면책 범위(Safe Harbor)"는 1995년 미 증권민사소송개혁법(US Private Securities Litigation Reform Act)에 다음과 같이 명시되어 있습니다. 이 문서의 성격상 명시된 일부 정보는 미래지향적인 것일 수 있습니다. 미래지향적인 내용에 근거한 예측은 리스크와 불확실성을 비롯한 중요한 요인의 영향을 받습니다. 여기에 수록된 여러 기타 요소로 인해 제공되는 제품, 기능 또는 개선사항이 미래지향적인 내용에 기초한 예측과는 실제로 다를 수 있습니다. Netezza는 향후 사건이나 상황을 반영하기 위해 미래지향적인 내용을 업데이트할 책임을 지지 않습니다.



Netezza Corporation

26 Forest Street, Marlborough, MA 01752

전화+1 508 382 8200

팩스+1 508 382 8300

Netezza 정보

Netezza(NYSE: NZ)는 기업 전반에 걸쳐 고성능 분석을 단순화하는 데이터 웨어하우스 및 분석 어플라이언스 부문의 글로벌 리더입니다. Netezza의 기술은 조직이 디지털 미디어, 에너지, 금융 서비스, 정부기관, 의료 건강, 유통 및 통신 등 오늘날의 데이터 집약적인 산업에서 빠른 속도로 대용량 데이터를 처리하여 우수한 경쟁력 및 경영 우위를 확보할 수 있게 해줍니다.

Netezza는 Massachusetts의 Marlborough에 본사를 두고 있으며 북 버지니아, 캐나다, 영국, 독일, 프랑스, 일본, 한국, 호주 및 싱가포르에 지사를 두고 있습니다.

Netezza에 대한 자세한 내용을 알아보려면 www.netezza.com을 방문하십시오.