

IBM Netezza에서 고급 분석 기능 활용

어플라이언스 접근 방식을 통한 고급 분석 간소화 방법



컨텐츠

- 2 소개**
 - 3 IBM Netezza Analytics의 유용한 기능**
 - 4 IBM Netezza Analytics 사용 방법**
 - 데이터 준비 및 변환
 - 모델 작성 및 게시
 - 모델 평가
 - 알고리즘 개발
 - 7 요약**
 - 7 Netezza 정보, IBM Company**
-

소개

IBM Netezza 데이터 웨어하우스 어플라이언스를 IBM Netezza Analytics와 함께 사용하면 대용량 데이터를 처리할 수 있는 확장 가능한 고성능 분석 기능으로 IBM의 혁신적인 데이터 웨어하우스 어플라이언스를 강화하여 분석의 한계를 극복할 수 있으므로 일반 솔루션에 비해 복잡한 문제를 훨씬 더 빠르게 해결할 수 있습니다. IBM Netezza Analytics는 모든 IBM Netezza 어플라이언스와 함께 제공되며 내부 데이터베이스 분석 기술을 개발하여 배포함으로써 상황을 반전시키는 결과를 도출할 수 있도록 해주는 유연성이 뛰어난 개방형 고급 분석 플랫폼입니다.

IBM 어플라이언스에 분석 기능을 내장하여 IBM은 Netezza 기본 아키텍처의 처리 속도를 극대화하면서 분석 기술을 쉽게 배포할 수 있도록 간소화했습니다.

IBM Netezza Analytics의 데이터 준비, 데이터 마이닝, 예측 모델링 및 최적화를 위한 병렬 분석 기술을 사용하면 IBM Netezza AMPP(Asymmetric Massively Parallel Processing) 아키텍처를 통해 대용량의 데이터를 빠른 속도로 고급 분석할 수 있습니다. IBM Netezza 데이터 웨어하우스 어플라이언스를 쉽게 확장하여 강력한 고급 분석 기술을 개발한 후 어플라이언스에 결합할 수 있습니다. Analytic 애플리케이션, 시각화 툴 및 비즈니스 인텔리전스 툴은 SQL, Java, MapReduce, Python, R, C, C++, Fortran 등과 같은 다양한 툴을 통해 병렬화된 고급 분석 기능을 활용하여 강력하고 통찰력이 뛰어난 분석 기술을 제공합니다. 포괄적인 고급 분석 환경에서 원하는 대로 쉽게 작업할 수 있습니다. 즉, 기본 툴을 사용하여 특별 분석하고 고급 분석 프로토타입을 작성한 후 프로덕션 환경에서 전개할 수 있습니다.

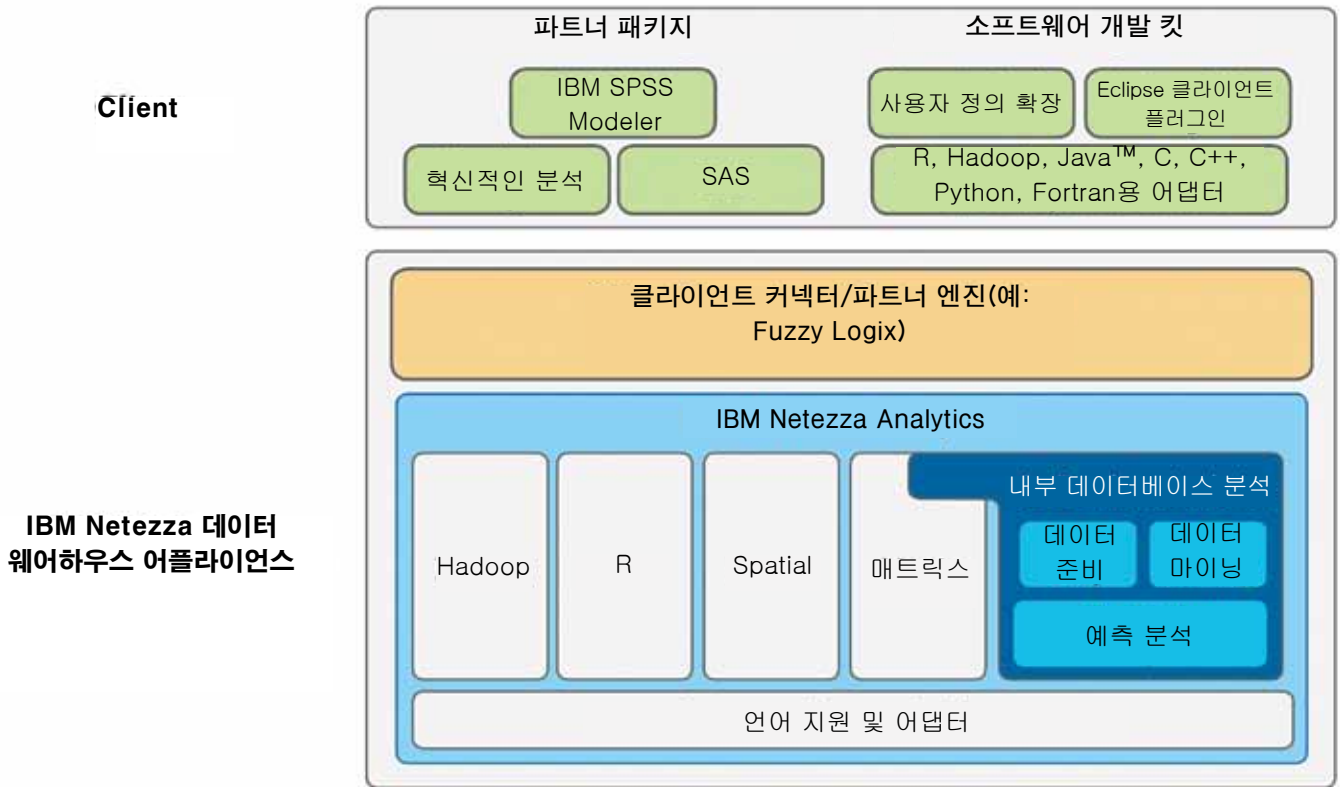


그림 1: IBM Netezza Analytics 아키텍처

IBM Netezza Analytics의 유용한 기능

IBM Netezza Analytics는 대용량 데이터와 병렬 처리 기능을 통합하여 고성능의 확장성이 뛰어난 분석 애플리케이션을 활성화합니다. 분석에서 IBM Netezza 어플라이언스의 AMPP™(Asymmetric Massively Parallel Processing™) 아키텍처를 사용하면 데이터 및 태스크 병렬화를 극대화하여 매우 빠르고 통찰력 있는 결과를

제공할 수 있습니다. IBM은 분석 기술을 IBM의 고유한 AMPP 아키텍처와 긴밀하게 통합하여 기술적 한계를 극복함으로써 이전에는 불가능했던 대용량의 데이터를 처리할 수 있는 우수한 성능을 제공합니다. 데이터 볼륨과 복잡한 연산에 대한 성능이 모두 향상됨에 따라 통계학자, 모델러 및 개발자는 올바른 툴을 적재적소에 활용하여 가치가 높은 영업 기회를 식별할 수 있습니다.

대부분의 경우 데이터 웨어하우스에서 데이터를 복사한 후 분석에 활용할 수 있도록 처리해야 하기 때문에 개발 주기가 길어지고 소프트웨어 비용이 상승하게 됩니다. 여기에는 데이터 웨어하우스와 다른 플랫폼 간에 데이터를 추출, 변환 및 로드하는 프로그램이 사용됩니다. IBM Netezza 데이터 웨어하우스 어플라이언스를 IBM Netezza Analytics와 함께 사용하면 데이터 웨어하우스에서 데이터를 바로 분석하여 부가가치가 창출되지 않는 필요 없는 단계가 제거되거나 대폭 축소되므로 프로세스를 간소화할 수 있습니다. 분석가는 환경을 간소화하여 조직에 많은 가치를 제공할 수 있는 모델링, 시뮬레이션, 최적화 등과 같은 부가가치가 높은 활동에 집중할 수 있습니다.

IBM Netezza Analytics는 IBM, IBM 파트너 및 공개 소스 커뮤니티의 풍부한 고급 분석 기술을 통합합니다. 애플리케이션 개발자와 모델러는 고급 분석 기술을 상황에 맞게 조합하여 당면한 과제를 해결하는 포괄적인 솔루션을 설계 및 개발할 수 있습니다. 기업에서는 고성능 병렬 분석을 기반으로 우수한 성능의 분석 기술을 구축하거나 분석 애플리케이션을 빠르게 개발함으로써 "영업 기밀"을 개발하거나 경쟁 우위를 확보할 수 있습니다.

예를 들어, 엄청나게 많은 트랜잭션과 고객 간의 가변성이 높은 CRM에서는 일반적으로 고객 분류를 통해 수익성이 높은 고객을 식별하고 예측 모델링을 통해 수익성이 낮은 고객을 수익성이 높은 고객으로 전환하는 방법을 모색할 수 있습니다. 모델링과 최적화를 결합하여 고객의 반응을 극대화하는 데 가장 적합한 설계, 메시지 및 판촉 방안을 결정할 수 있습니다. 데이터 마이닝과 예측 모델링은 우량 고객을 식별하여 유치하고 적절한 인센티브를 제공하여 고객의 충성도를 높이기 위한 접근 방안으로 활용할 수 있습니다.

IBM Netezza Analytics 사용 방법

IBM Netezza Analytics는 신중한 분석을 위해 개발된 고급 임베디드 분석 플랫폼입니다. IBM Netezza Analytics의 용도는 다음과 같습니다.

- 고급 분석 애플리케이션 작성 및 배포
- 시각화 또는 비즈니스 인텔리전스 툴을 통해 병렬 분석 활용
- 대용량 데이터 또는 컴퓨팅 집약적인 문제에 대한 특별 분석 수행

시각화 및 비즈니스 인텔리전스 툴에서는 SQL을 통해 IBM Netezza Analytics의 분석 기능을 활용하여 대상을 조회합니다. 하지만 애플리케이션 개발자와 모델러는 다음과 같은 애플리케이션 개발 또는 특별 분석을 위해 보다 유연한 기능이 필요합니다.

- 데이터 준비 및 변환
- 모델 작성 및 게시
- 모델 평가
- 알고리즘 개발

데이터 준비 및 변환

데이터 준비 및 변환 활동을 수행하는 분석가는 IBM Netezza Analytics 툴 및 유틸리티를 활용하여 프로세스를 지원할 수 있습니다. 사용되는 툴은 다음과 같습니다.

- 기술 통계
- 데이터 정리
- 데이터 변환
- 기능 선택 및 차원 축소
- 통계 검증 및 유틸리티

데이터 준비와 변환이 필요한 경우 IBM은 어플라이언스에 저장할 수 있는 사용자 데이터를 페타바이트 단위로 확장하는 동시에 고성능 패키지를 통해 프로세스를 가속화합니다.

분석 모델 작성 및 게시

분석 모델 작성은 일반적으로 대화식 프로세스입니다. 모델을 작성하는 데 사용되는 데이터는 모델러를 위한 데이터를 준비하는 모델러(데이터 마이너, 통계학자, 정량 분석가 등)가 준비 및 변환할 수 있습니다. 모델러는 경영진의 요청에 따라 특별 분석에 착수할 수 있습니다. 결과를 검토한 후 모델러는 최종 프로덕션 전개를 위해 원형에 가깝게 개선할 수 있습니다. 모델러는 일반적으로 이 프로세스 중에 여러 모델을 개발하여 각 모델의 성능을 비교한 후 최종 배치할 모델을 선택합니다. 반복 개발 주기의 각 단계에서 모델러는 변화하는 요구사항과 성능 목표에 따라 다른 개발 툴과 분석 기술을 사용할 수 있습니다.

IBM Netezza Analytics는 이 프로세스의 각 단계에 필요한 유연성을 제공합니다. 모델러는 Eclipse, R GUI 등과 같은 여러 지원되는 개발 환경 중 하나를 사용하거나 다음을 직접 사용하여 모델을 작성할 수 있습니다.

- Eclipse용 IBM 플러그인
- R GUI용 IBM 플러그인
- 파트너의 통합 개발 환경(IDE)
- 분석 워크플로우 작성을 위한 파트너의 분석 개발 환경(ADE)
- IBM Netezza 지원 프로그래밍 언어(SQL, Java, MapReduce, Python, R, Fortran, C 또는 C++)

모델에서는 다음과 같은 고급 분석 기술을 조합하여 사용할 수 있습니다.

- **개방형 소스 분석** – R CRAN 패키지(IBM Netezza Analytics를 통해 쉽게 설치 가능)
- **파트너 분석** – 애플리케이션 개발에 활용되는 업계에서 검증된 분석 기술
- **내부 데이터베이스 분석** – 데이터 준비, 데이터 마이닝, 예측 및 지리 공간 분석을 위한 IBM의 고성능 병렬 내부 데이터베이스 분석 기술

모델을 사용하여 미래의 결과를 예측한 후 모델에서 예측 모델링과 최적화를 결합하여 최적의 결정을 내릴 수 있도록 자동화할 수 있습니다. 애플리케이션에 최적화 기능을 포함시켜서 다음과 같은 어렵고 복잡한 비즈니스 문제를 해결할 수 있습니다.

- 재고 및 배송 비용 최소화
- 자본 투자 및 전체 포트폴리오 위험 최소화
- 매출 및 고객 충성도 극대화
- 마케팅 예산 최소화 및 잠재고객 발굴(lead generation) 최대화

이러한 작업은 예산, 직원 현황, 고객의 선호도 및 행동, 회사의 선호도 등 다양한 실제 제약 조건을 고려하여 수행됩니다.

모델 평가

모델을 작성한 후 예측 모델과 일반적으로 연관되는 대용량의 데이터에 모델을 적용할 수 있습니다. 이 프로세스를 평가라고 합니다. 예를 들면, 수백만 개의 레코드에 모델을 적용하여 다음을 수행할 수 있습니다.

- 사기 탐지
- 신용 점수 계산
- 고객 계정 가격 조정
- 미래의 고객 구매 예측
- 다양한 마케팅 프로모션 대상 결정
- 신용 한도 증액/감소 여부 결정
- 연체 계정을 매각할지 여부 결정
- 현장 지원 호출을 기반으로 제품 보증 전략 결정
- 판매 증대를 위한 적절한 고객 판촉 결정

알고리즘 개발

알고리즘 개발자는 개발 또는 프로그래밍 언어에 대한 GUI 기반 접근 방식으로 통합 개발 환경(IDE)을 비롯한 다양한 개발 툴을 선택할 수 있습니다. IBM은 개방형 소스 IDE인 Eclipse용 플러그인을 제공합니다. 또한 IBM 파트너는 알고리즘 개발 주기를 단축할 수 있는 IDE를 제공합니다. IBM은 R 언어를 기반으로 알고리즘을 개발하는 데 필요한 R GUI를 지원합니다. IBM은 Java, MapReduce, Python, R, Fortran, C, C++ 및 사용자 정의 확장을 비롯하여 알고리즘 개발을 위한 여러 프로그래밍 언어를 지원합니다. 사용자 정의 확장은 다음과 같습니다.

- **UDF** – User Defined Function의 약어이며 행을 처리합니다.
- **UDA** – User Defined Aggregates의 약어이며 행 그룹을 처리합니다.
- **UDTF** – User Defined Table Function의 약어이며 모든 크기의 테이블을 반환합니다.
- **UDAP** – User Defined Analytic Process의 약어이며 하나 이상의 처리 단계를 수행할 수 있도록 완벽한 유연성을 제공합니다.
- **Stored Procedures** – 함수, 집계 및 기타 논리를 프로시저로 그룹화하는 데 사용됩니다.

알고리즘을 개발하여 테스트한 후 지원되는 프로그래밍 언어를 통해 다른 내부 데이터베이스 분석 기능처럼 액세스할 수 있도록 어플라이언스에 등록할 수 있습니다.

요약

IBM Netezza Analytics의 IBM Netezza 데이터 웨어하우스 어플라이언스는 최고의 성능과 확장성을 제공하면서 고급 분석 기술의 개발, 전개 및 사용을 간소화하는 확장성이 뛰어난 강력한 고급 분석 플랫폼입니다. IBM Netezza, 파트너, 개방형 소스 및 사용자의 고급 분석 기술을 결합한 후 SQL, Java, MapReduce, Python, R, Fortran, C 또는 C++를 통해 IBM Netezza 데이터 웨어하우스 어플라이언스의 고성능 병렬 아키텍처를 활용하여 상황을 반전시킬 수 있는 비즈니스 결과를 제공할 수 있습니다.

IBM Netezza

IBM Netezza는 데이터베이스, 서버와 스토리지를 관리하기 편리한 하나의 어플라이언스로 통합함으로써 데이터 웨어하우스 어플라이언스 분야를 개척했습니다. 설치 및 지속적인 관리의 필요성을 최소화하면서 더 신속하고 일관성 있는 분석 성능을 제공합니다. 최근에 선보인 데이터 웨어하우스 어플라이언스인 IBM Netezza 1000은 모든 분석 작업을 어플라이언스, 즉 데이터가 저장된 곳에 통합함으로써 비즈니스 분석을 크게 간소화하고 매우 빠른 성능을 제공합니다. IBM 데이터 웨어하우스 어플라이언스로 각 단계의 복잡성을 최소화하고 진정한 비즈니스 가치를 실현할 수 있는 방법을 netezza.com 또는 thinking.netezza.com 에서 확인하십시오.

IBM 데이터 웨어하우징 및 분석 솔루션 정보

IBM은 매우 포괄적이며 폭넓은 데이터 웨어하우징, 비즈니스 분석 및 정보 관리 소프트웨어, 하드웨어 및 솔루션 포트폴리오를 제공하여 고객이 정보 자산의 가치를 극대화하고 새로운 통찰력을 개발하여 더 효율적이면서 신속한 결정을 내리고 비즈니스 결과를 최적화할 수 있도록 돕습니다. IBM의 데이터 웨어하우징 및 분석 솔루션 모두가 비즈니스 분석 통찰력을 제공하는 과정을 단순화하고 가속화하기 위해 고안되었습니다.

IBM의 포트폴리오는 데이터 웨어하우스 어플라이언스를 통해 최소화된 설치 및 지속적인 관리가 요구되는, 관리하기 쉬운 단일 어플라이언스로 데이터베이스, 서버 및 스토리지를 통합하며 더 신속하고 일관적인 분석 성능을 제공합니다. 또한 IBM에서는 운영 지성을 위해 미리 작성되고 통합된 워크로드 최적화 데이터 웨어하우징 및 분석 플랫폼 그리고 데이터 웨어하우스 소프트웨어를 제공합니다. 이러한 오퍼링은 대용량의 DIM(Data-In-Motion) 볼륨을 지속적으로 신속하게 분석하는 기능 등 분석 워크로드의 새로운 유형 및 대규모 데이터에 대한 추가 지원을 통해 더 향상되었습니다.

자세한 정보

IBM 데이터 웨어하우징 및 분석 솔루션에 대해 자세히 알아보려면 해당 IBM 영업 담당자에게 문의하거나 ibm.com/software/data/infosphere/data-warehousing/ 을 방문하십시오.



☎135-270☎
서울특별시 강남구 도곡동 467-12☎
군인공제회관 빌딩
한국아이비엠주식회사 ☎
고객만족센터 ☎

IBM 홈 페이지는 다음 URL에서 확인할 수 있습니다.

ibm.com/kr

IBM, IBM 로고, ibm.com 및 Netezza는 미국 또는 기타 국가에서 사용되는 International Business Machines Corporation의 상표입니다. 이러한 상표 및 기타 IBM 상표가 상표 기호(® 또는 ™)와 함께 이 정보에서 처음 표시되어 있는 경우 이 기호는 이 정보가 출판되었을 때 IBM이 보유한 미국 등록 상표 또는 보통법상 상표임을 나타냅니다. 또한 이러한 상표는 기타 국가에서 등록상표 또는 일반 법적 상표입니다. 현재 IBM의 상표 목록은 "저작권 및 상표 정보"

ibm.com/legal/copytrade.shtml

Microsoft 및 SQL Server는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

기타 회사, 제품 및 서비스 이름은 해당 회사의 상표 또는 서비스표입니다.

© Copyright IBM Corporation 2012



재활용 하십시오.