



IBM Integrated Analytics System

데이터 사이언스의 속도 향상

주요 특징

- 클라우드 환경으로 배포가 용이한 고성능의 하드웨어/소프트웨어 통합 플랫폼으로 사용자 지원
- 함께 제공되는 IBM® Data Science Experience Local로 첨단 분석 모델 개발
- 향상된 데이터 로드 속도, 분석 워크로드에서 쿼리 응답 시간 단축
- IBM이 제공하는 Common SQL 엔진 및 내장된 데이터 가상화 기능을 사용하여 하이브리드 데이터 관리 인프라의 전 범위에서 원활한 상호 운용성 실현

효과적인 데이터 분석으로 더 신속한 의사 결정이 가능하며 그에 따라 업계 선두 주자와 경쟁 열위 그룹이 결정되기도 합니다. 그러나 기업에 분석 기능을 제공하는 일은 다음과 같은 여러 요인으로 인해 복잡해집니다.

- 머신 러닝을 비롯하여 복잡한 분석 요구 사항을 지원하기 위해 데이터 사이언티스트를 비즈니스에 참여시키는 것의 중요성
- 비정형 데이터 및 반정형 데이터를 포함하여 여러 데이터 타입을 지원할 필요성
- 실시간 및 스트리밍 분석을 요구하는 사물 인터넷(Internet of Things, IoT) 등 새롭게 등장하는 활용 사례
- 온프레미스 및 클라우드 환경에서 관리해야 하는 데이터의 폭발적인 증가 및 그에 따른 데이터 사일로화(Data Silo)
- 데이터웨어하우스 인프라 내에서 SLA(service level agreement)에 부합하기 위한 성능 및 확장성 필요

기존의 인프라에 대한 투자를 유지하면서 위 요구사항을 수용하기 위해서는 조직의 분석 및 데이터 웨어하우스 솔루션이 다음 조건을 충족해야 합니다.

- 고급 분석 모델의 개발 및 배포에 관련된 어려움도 최소화할 수 있는 솔루션을 데이터 사이언티스트에게 제공합니다.
- 다양한 데이터 타입과 위치, 기존 배치 처리 및 준 실시간 처리 또는 혼합 워크로드를 지원합니다.
- 클라우드에 대한 연결 및 미아그레이션을 지원하는 하이브리드 데이터 전략을 수용합니다.
- 워크로드의 규모와 유형에 상관없이 우수한 성능 및 확장성을 제공합니다.
- 클라우드와 유사한 유연성과 확장성을 지원하되 데이터 센터 내에 위치합니다.

조직의 분석 요구 사항이 늘어남에 따라 올바른 솔루션에 투자하는 것이 무엇보다 중요합니다. 어떤 선택을 하느냐에 따라 SLA를 달성할 수도, 성공의 장애물을 만날 수도 있습니다.



IBM Analytics

솔루션 개요

IBM Integrated Analytics System은 클라우드 환경에 최적화된 고성능 데이터 플랫폼으로 데이터 사이언티스트의 개발 및 배포 작업을 가속화하여 귀사의 경쟁력을 강화하는데 필요한 인사이트를 발굴합니다.

데이터 사이언티스트는 내장되어 있는 통합 데이터 사이언스 솔루션인 IBM Data Science Experience를 활용하여 귀사에 저장된 데이터를 최신의 데이터 사이언스 툴에 적용할 수 있습니다. 이러한 연결을 통해 성능 최적화된 Common SQL 엔진과 임베디드 Apache Spark 프로세싱 기능을 활용하여 데이터 사이언티스트가 머신 러닝 분석을 개발하는 것이 가능합니다.

IBM Integrated Analytics System은 처음부터 튜닝 및 유지 보수를 거의 또는 전혀 필요로 하지 않으면서 고성능 및 페타바이트급 확장성을 요하는 까다로운 워크로드도 관리하고 운영할 수 있습니다. IBM Integrated Analytics System은 임베디드 Apache Spark 프로세싱 엔진을 머신 러닝에 활용하면서 더 우수한 성능의 분석을 지원합니다. 뿐만 아니라 이 기능으로 다른 환경에 분석 및 데이터를 이동하는 작업이 불필요해지므로 복잡성을 줄일 수 있습니다. IBM 하이브리드 데이터 관리 전체 제품군의 기반이 되는 Common SQL 엔진을 통해 기존 온프레미스 및 클라우드 애플리케이션과 연동할 수 있습니다. 이러한 유연성 덕분에 각 작업에 적합한 처리 환경을 선택하는 것이 가능합니다.

이 통합 아키텍처는 AMPP(asymmetric massively parallel processing)와 같은 고급 소프트웨어 기능을 IBM Power® 기술 및 플래시 메모리 스토리지 하드웨어와 통합하여 제공합니다. IBM Integrated Analytics System은 기존 배치 워크로드와 혼합 워크로드를 처리합니다. 이 워크로드는 대개 방대한 데이터 볼륨을 대상으로 한 쿼리, 소규모 데이터 세트에 대한 빠른 질의, 다수의 동시 처리를 필요로 합니다. 이를 위해 IBM Integrated Analytics System은 단일 솔루션에서 각종 데이터 타입 및 위치를 처리하는 다양한 분석 활용 사례를 지원합니다. 이러한 유연성 덕분에 데이터 사이언티스트에게 무한한 가능성을 제공할 수 있습니다.



Spark를 기반으로 한 임베디드 분석 및 머신 러닝

데이터 사이언티스트는 IBM Integrated Analytics System의 임베디드 Apache Spark 프로세싱 기능으로 예측 모델에 대한 학습과 평가 작업을 단순화할 수 있습니다. 이 기능으로 많은 시간을 들여 다른 시스템으로 데이터를 이동하고 변환할 필요가 없습니다. 데이터 사이언티스트가 선호하는 툴을 사용하여 모델을 개발한 다음 데이터가 있는 곳에서 테스트, 배포, 학습을 수행할 수 있습니다. 노드마다 Spark 실행(executor) 프로세스가 있어 지연이 최소화되므로 독립형 Spark 클러스터보다 빠른 속도로 데이터 액세스 및 연산이 이루어집니다. 데이터 사이언티스트가 시스템에서 워크로드의 부담을 덜어야 하는 경우 표준 도구 및 Common SQL 엔진을 활용하여 분석 모델을 Spark 클러스터에 손쉽게 이동할 수도 있습니다.

IBM Analytics

솔루션 개요

뿐만 아니라 간소화된 프로세스로 머신 러닝 기능을 포함한 분석 작업에 더 우수한 성능 및 유연성을 제공합니다. IBM Data Science Experience Local에 포함된 5 Authorized User 라이선스를 통해 데이터 사이언티스트가 곧바로 시스템의 데이터에 연결한 다음 모델 개발을 시작할 수 있습니다. 이 대화형 협업 클라우드 기반 환경은 데이터 사이언티스트가 인사이트를 발굴할 수 있도록 다양한 툴을 제공합니다. 또한 Jupyter Notebook 컨테이너가 시스템에 포함되어 있으므로 Jupyter Notebook에서 Python, R, Scala를 사용할 수 있습니다. Jupyter는 클릭 한 번으로 컴파일 및 Spark 애플리케이션으로 배포할 수 있어, 대화형 코드 실행에 사용할 수 있습니다.

데이터 마이닝, 예측, 변환, 통계, 지리 정보 데이터, 데이터 준비를 위한 함수들이 내장되어 있을 뿐 아니라 IBM Integrated Analytics System의 임베디드 Spark에서 오픈 소스 R 및 기타 프로그래밍 언어(Python, Java, C, C++, Lua 등)를 지원합니다.

IBM Integrated Analytics System은 선형 회귀, 의사 결정 트리 클러스터링, k-means 클러스터링, Esri 호환 지리 정보 확장 등의 다양한 알고리즘이 포함된 IBM Netezza® Analytics 기술을 내장하고 있습니다. 이 시스템은 다양한 비즈니스 분석 및 시각화 툴, 즉 IBM Cognos®, SAP BusinessObjects, Kognitio, Microsoft Excel, QlikView, SAS, Microsoft SQL Server Reporting Services(SSRS), Tableau 등과 함께 사용할 수 있습니다. IBM SPSS®, Fuzzy Logix, 오픈 소스 R 및 SAS 등과 같은 모델 개발 및 평가 툴도 지원합니다.

통합과 성능

IBM은 20년 넘게 다양한 업종 및 지역에서 수천여 고객과 함께 일한 경험을 바탕으로 한 설계로 분석 시스템의 배포 및 관리를 간소화합니다. 소프트웨어와 하드웨어는 성능 최적화된 단일 솔루션의 형태로 귀사의 데이터 센터에 공급됩니다. 시스템 기동후 수시간내에 데이터베이스 인덱스나 튜닝없이 데이터를 로드할 수 있습니다.

IBM Integrated Analytics System 및 IBM Db2® Warehouse 기술을 이용한 클라이언트는 IBM 하이브리드 데이터 관리 솔루션 포트폴리오 전 영역의 기반이 되는 Common SQL 엔진을 즉시 인식합니다. IBM Db2 Warehouse는 데이터 웨어하우스 및 분석 워크로드를 위해 설계되었습니다. Common SQL 엔진은 IBM Db2 및 IBM BLU Acceleration® 기술을 기반으로 한 동적 인메모리(in-memory) 열 기반(columnar) 기술로 다양한 워크로드를 지원합니다. BLU Acceleration의 MPP(massively parallel processing) 아키텍처는 고속 심층 데이터 분석을 위해 설계되었으며 페타바이트 단위까지 확장 가능합니다. 쿼리 응답 속도가 이전 시스템보다 최대 100배 빠른¹ BLU Acceleration 열 기반 테이블은 기존 행 기반 테이블과 동일한 스키마, 스토리지, 메모리에 위치할 수 있으므로 행 및 BLU Acceleration 열 기반 테이블 모두 동시에 쿼리하는 것이 가능합니다. 기존 인메모리 기능에 BLU Acceleration 기술을 추가함으로써 데이터 세트가 메모리 크기를 초과하더라도 성능을 향상시킬 수 있습니다. BLU Acceleration의 동적 인메모리 열 기반 기술과 데이터 스키핑 기능으로 데이터가 압축된 상태에서도 효율적으로 탐색하여 관련 데이터를 찾을 수 있습니다.

IBM Integrated Analytics System은 IBM Power Systems™ 및 IBM FlashSystem® 기술을 활용하여 하드웨어 수준에서 안정성 및 성능을 높입니다. 현재의 IBM Power 아키텍처에서는 이전 솔루션보다 적은 수의 노드에서 동급의 성능을 제공하는 고밀도 시스템이 가능합니다. 시스템의 기본 스토리지로 IBM FlashSystem을 사용하여 초저지연(ultra-low latency)과 빠른 인메모리 I/O(near-in-memory I/O) 성능을 탁월한 안정성과 함께 제공합니다.

분석 애플리케이션이 최고 성능으로 실행되더라도 IBM Integrated Analytics System은 SLA를 만족하거나 초과달성하기 위해 새로운 차원의 안정성을 제공합니다. Power Systems와 IBM FlashSystem 스토리지는 단일 장애 지점(single point of failure)을 차단하는 내장애성 설계로 더 우수한 가동 시간을 인정받았습니다. 2017 ITIC(Information Technology Intelligence Consulting) 조사에서 IBM Power Systems는 예기치 않은 다운타임이 연간 서버당 2.5분으로 주요 Linux 서버 플랫폼 중 가장 낮았습니다.²

시스템 전반의 구성 요소들이 이중화로 구성되어 하드웨어 오류가 발생하더라도 작업을 진행할 수 있도록 도와줍니다. 또한 성능을 유지하기 위해 자동으로 페일오버를 수행하는 고가용성 기능도 내장되어 있습니다. 하드웨어 및 소프트웨어의 모든 구성 요소를 모니터링하고 관리하기 위해 Db2 제품군에서 사용되는 IBM Data Server Manager 기반 콘솔이 내장되어 있습니다.

클라우드 및 데이터를 위한 하이브리드 접근

데이터에 관한 한 하나의 방법으로 모든 요구 사항을 해결할 수는 없습니다. IBM Integrated Analytics System은 Common SQL 엔진을 기반으로 조직의 인프라 전반에 원활한 상호 운용성을 제공합니다.

예를 들어 어떤 팀에서 지금까지 사용하던 데이터 웨어하우스를 시즌별 용량 수요 때문에 클라우드로 이전해야 하는 경우가 있습니다. 이 워크로드를 IBM Db2 Warehouse on Cloud로 마이그레이션하는 작업은 IBM Bluemix® Lift와 같은 툴을 사용하면 간단하게 해결됩니다. Common SQL 엔진 덕분에 고객의 애플리케이션을 재개발할 필요가 없습니다.

Common SQL 엔진은 데이터의 물리적 위치에 상관없이, 비정형 데이터 또는 반정형 데이터까지 포함하여 조회할 수 있습니다. Common SQL 엔진에 내장된 데이터 시각화 서비스로 논리적 데이터 웨어하우스 전 범위에서 데이터 액세스를 통합함으로써 Db2, Hadoop, 서드파티 데이터 소스까지 포괄하는 통합이 가능해집니다.

확장성 및 확장 옵션

소프트웨어 및 하드웨어 아키텍처 모두 원활하게 비즈니스를 지원하기 위해 워크로드 증가에 맞춰 확장 가능하도록 시스템에서 설계되었습니다. 컴퓨팅 및 스토리지 용량은 독자적인 확장이 가능하므로 클라우드와 같은 수준의 유연성 및 탄력성을 누릴 수 있습니다. 하드웨어 확장은 비즈니스에 영향이 없도록, 시스템에서 곧바로 이루어집니다.

IBM Integrated Analytics System은 사용빈도 기반 계층형 스토리지(multi-temperature tiered storage)도 지원하므로 방대한 데이터 볼륨에도 최고 수준의 성능을 보장할 수 있습니다. 최근에 사용되고 활성화된 데이터는 시스템 스토리지 노드에서 직접 관리하는 반면 더 오래되고 사용이 저조한 데이터는 비용 효율적인 고밀도 IBM Storwize® 스토리지 디바이스에 저장합니다.

활용 사례

IBM Integrated Analytics System의 Common SQL 엔진으로 워크로드에 적합한 실행 플랫폼을 선택할 뿐 아니라 데이터 타입, 위치, 크기에 상관없이 데이터 액세스를 보장할 수 있습니다. 이 솔루션을 도입하려는 고객이 참고할 만한 몇 가지 활용 사례를 소개합니다.

- 상호 운용되는 Hadoop, 데이터 마트, 기타 환경을 사용하여 논리적 데이터 웨어하우스를 구축하고 온프레미스 또는 클라우드의 데이터에 대한 통합된 조회 환경을 제공하면서 새로운 차원의 운영 우수성을 실현합니다.
- 고성능 및 간소화된 확장성을 요하는 분석에 내부 데이터를 연계하여 실시간 맞춤형 고객 경험 환경을 만들어냅니다.
- 임베디드 머신 러닝 기술로 더 신속하게 복잡한 분석을 수행하고 사용자에게 대한 인사이트를 확보하여 혁신적인 신제품의 출시 일정을 단축합니다.
- 더욱더 빠르게 데이터를 로드하고 튜닝 또는 광범위한 구성 작업이 필요없는 시스템을 통해 비즈니스 사용자에게 신속하게 워크로드를 제공합니다.
- 복합 운영 워크로드를 추가하여 비즈니스 사용자들의 다양한 워크로드를 처리합니다.
- 데이터에 대한 더 정확한 관점을 제시하여 운영에서 컴플라이언스를 보장합니다.
- 데이터 가상화를 통해 논리적 데이터 웨어하우스 및 기타 기술 전반에서 데이터 액세스를 통합함으로써 다양한 데이터 타입 및 데이터 세트를 포괄하는 분석을 구현하여 비즈니스 성장 및 확장을 위한 새로운 차원의 인사이트를 제공합니다.

사양

IBM Integrated Analytics System은 모든 컴퓨팅, 스토리지, 네트워킹 리소스를 분석 및 데이터 웨어하우스 소프트웨어와 통합하고 최적화합니다. 다음 표에서 보여주는 것처럼 랙 구성 형태로 제공됩니다.

| 사양 - 단일 랙 시스템 | | | |
|---------------------------------------|--|-----------|-----------|
| IBM Integrated Analytics System 모델 번호 | M4001-003 | M4001-006 | M4001-010 |
| 활성 노드 | 3 | 5 | 7 |
| CPU 노드 | 72 | 120 | 168 |
| 메모리(TB) | 1.5 | 2.5 | 3.5 |
| 사용자 데이터(TB)(4배 압축 가정) | 64 | 128 | 192 |
| 랙당 최대 전력(와트) | 4,400 | 6,900 | 9,400 |
| 냉각(BTU/시간) | 15,000 | 23,500 | 32,000 |
| 랙 중량(Kg) | 479.9 | 563.8 | 647.7 |
| 높이(cm) | 202 | | |
| 깊이(cm) | 110 | | |
| 너비(cm) | 64.8 | | |
| 전원 코드 | 북미 및 일본—200 – 240V, 단상, 30 A EMEA—220 – 240V, 단상, 32 A | | |
| 드롭/랙 | 4 | | |
| 안전성 | UL/CSA 60950-1, EN 60950-1 및 IEC 60950-1 | | |
| 방출 | FCC part 15, ICES-003, AUS/NZ CISPR 22, VCCI, EN55022/55032 Class A, European immunity: EN 55024, EN 61000-3-2, EN 61000-3-3 | | |
| 참고 | 멀티랙 시스템도 이용 가능합니다. 자세한 내용은 영업 대표 또는 IBM 비즈니스 파트너에게 문의하십시오. | | |

IBM 하이브리드 데이터 관리 소개

IBM은 가장 포괄적인 하이브리드 데이터 관리 솔루션 포트폴리오를 제공하면서 고객의 정보 자산 가치를 극대화하고 새로운 인사이트를 발굴하여 더 신속하고 현명한 의사 결정 및 최적의 비즈니스 성과 달성을 지원합니다.

추가 정보

IBM Integrated Analytics System에 대한 자세한 내용은 IBM 영업 대표 또는 IBM 비즈니스 파트너에게 문의하거나, 다음 웹 사이트를 참조하십시오. ibm.biz/BdjfvP.

또한 IBM Global Financing은 비용 효율적이고 전략적인 방법으로 비즈니스에 필요한 소프트웨어 역량을 신속하게 갖추 수 있도록 지원합니다. 일정 신용 등급 이상의 고객을 대상으로 비즈니스 및 개발 목표에 적합하고, 현금 흐름을 효과적으로 관리하며, 총 소유 비용을 낮추어 주는 맞춤형 금융 솔루션을 제공합니다. IBM Global Financing으로 중요한 IT 투자를 지원하고 비즈니스 발전을 촉진하십시오. 자세한 내용은 ibm.com/financing/kr에서 확인하실 수 있습니다.



© Copyright IBM Corporation 2017

IBM Corporation
IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
2017년 9월

IBM, IBM 로고, ibm.com, BLU Acceleration, Bluemix, Cognos, Db2, IBM FlashSystem, Power, Power Systems, SPSS, Storwize는 전세계 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.

Linux 는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

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

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

본 문서에 포함된 성능 데이터는 특정 운영 조건에서 산출된 것입니다. 실제 결과는 다를 수 있습니다. 그러나 IBM 제품 및 프로그램과 함께 사용한 기타 다른 제품이나 프로그램의 운영에 대한 평가와 검증은 사용자의 책임입니다.

이 문서의 정보는 상품성, 특정 목적에의 적합성에 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 규정에 대한 고객의 준수 여부가 확인된다고 진술하거나 보증하지 않습니다.

우수 보안 관리제도에 대한 설명: IT 시스템 보안은 귀사 내/외부로부터의 부적절한 접근을 방지, 감지, 대응함으로써 시스템과 정보를 보호하는 일을 포함합니다. 부적절한 접근은 정보의 변경, 파괴 또는 유출을 초래하거나, 타 시스템에 대한 공격을 포함한 귀사 시스템에 대한 피해나 오용을 초래할 수 있습니다. 어떠한 IT 시스템이나 제품도 완벽하게 안전할 수 없으며, 단 하나의 제품이나 보안 조치만으로는 부적절한 접근을 완벽하게 방지하는 데 효과적이지 않을 수 있습니다. IBM 시스템과 제품은 합법적이며 종합적인 보안 접근방법의 일부로서 고안되며, 이러한 접근방법은 필연적으로 추가적인 실행절차를 수반하며 가장 효과적이기 위해서는 다른 시스템, 제품 또는 서비스가 필요할 수도 있습니다. BM은 시스템과 제품이 임의의 당사자의 악의적 또는 불법적 행위로부터 영향을 받지 않는다는 것을 보장하지는 않습니다.

- 1 <http://www.redbooks.ibm.com/technotes/tips1204.pdf>
- 2 <http://itic-corp.com/blog/2017/06/ibm-lenovo-servers-deliver-top-reliability-cisco-ucs-hpe-integrity-gain/>



재활용하십시오