IBM SPSS Analytic Server
버전 3.2.1

개요

IBM

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 5 페이지의 『주의사항』에 있는 정보를 확인하십시오.

제품 정보

이 개정판은 새 개정판에 별도로 명시하지 않는 한, IBM SPSS Analytic Server의 버전 3, 릴리스 2, 수정사항 1 및 모든 후속 릴리스와 수정에 적용됩니다.

목차

개요 1
아키텍처 2
Spark 및 Analytic Server 2
버전 3.2.1의 새로운 사항 3
주의사항 5
상표 7

개요

IBM® SPSS® Analytic Server는 IBM SPSS 기술을 빅 데이터 시스템과 결합하고 익숙한 IBM SPSS 사용자 인터페이스로 작업하여 이전에 해결 불가능했던 문제점을 해결할 수 있도록 하는 빅 데이터 분석을 위한 솔루션입니다.

빅 데이터 분석이 중요한 이유

조직에서 수집하는 데이터의 볼륨은 기하급수적으로 증가하고 있습니다. 예를 들어, 금융 및 소매 업계는 모든 고객의 1년치(또는 2년이나 10년치) 거래내역을 가지고 있고 통신 제공업체의 경우 CDR(Call Data Record) 및 디바이스 센서 자료를 보관하며 인터넷 회사는 웹 크롤링 결과를 보관합니다.

다음과 같은 경우 빅 데이터 분석이 필요합니다.

- 특히 구조화된 데이터와 비구조화된 데이터가 혼합된 큰 볼륨의 데이터(테라바이트, 페타바이트, 엑사바이트)
- 빠르게 변경/누적되는 데이터

빅 데이터 분석은 다음과 같은 경우에도 도움이 됩니다.

- 수천 개의 모델을 작성하는 경우
- 모델을 자주 작성/갱신하는 경우

해결 과제

대용량의 데이터를 수집하는 같은 조직에서도 여러 이유로 이러한 데이터를 실제로 활용하지 못하는 경우가 있습니다.

- 기존 분석 제품의 아키텍처는 분산 계산에 적합하지 않습니다. 그리고
- 기존 통계 알고리즘은 빅 데이터 작업에 맞게 설계되어 있지 않습니다(이러한 알고리즘은 데이터를 가져와야 하지만 빅 데이터를 이동하는 데에는 많은 비용이 듦). 따라서
- 빅 데이터에 대한 최신 분석을 수행하려면 빅 데이터 시스템에 대한 새로운 기술 및 상세한 지식이 필요합니다. 이러한 기술을 보유한 분석가는 많지 않습니다.
- 인메모리 솔루션은 중간 규모의 문제점에 대해서는 효과가 있지만 실제 빅 데이터에는 적용하기가 쉽지 않습니다.

솔루션

Analytic Server에서는 다음 기능을 제공합니다.

- 빅 데이터 시스템을 이용하는 데이터 중심 아키텍처(예: HDFS의 데이터를 사용한 Hadoop Map/Reduce).
- 데이터로 이동하도록 설계된 새 통계 알고리즘을 통합하는 정의된 인터페이스.

© Copyright IBM Corp. 2010, 2019

- 빅 데이터 환경의 세부사항을 숨겨서 분석가가 데이터 분석에 집중할 수 있게 하는 친숙한 IBM SPSS 사용자 인터페이스.
- 모든 규모의 문제를 해결할 수 있는 확장 가능한 솔루션.

아키텍처

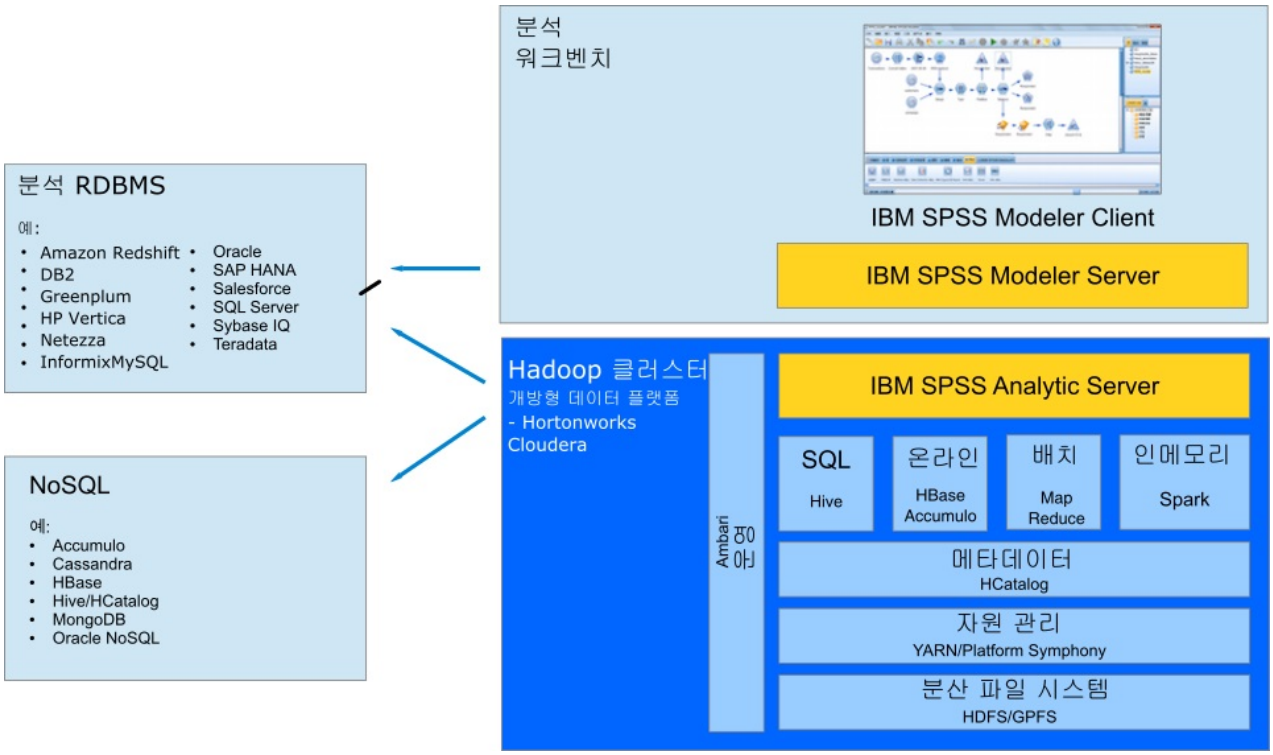

그림 *1.* 아키텍처

Analytic Server는 클라이언트 애플리케이션과 Hadoop 클라우드 사이에 있습니다. 데이터가 클라우드에 상주한다고 가정하면 Analytic Server에 대한 작업의 일반적인 개요는 다음과 같습니다.

1. 클라우드에 있는 데이터에 대한 Analytic Server 데이터 소스를 정의합니다.
2. 클라이언트 애플리케이션에서 수행할 분석을 정의합니다. 현재 릴리스의 경우, 클라이언트 애플리케이션은 IBM SPSS Modeler입니다.
3. 분석을 실행하면 클라이언트 애플리케이션은 Analytic Server 실행 요청을 제출합니다.
4. Analytic Server는 Hadoop 클라우드에서 실행할 작업을 조정하고 결과를 클라이언트 애플리케이션에 보고합니다.
5. 결과를 사용하여 추가로 분석을 정의하고 주기를 반복할 수 있습니다.

Spark 및 Analytic Server

Analytic Server는 성능 향상을 위해 Apache Spark와 통합됩니다.

Spark가 사용되거나 사용되지 않는 경우

Spark가 Hadoop 클러스터에 Ambari 서비스로 설치된 경우 Analytic Server에서는 이를 사용하여 빅 데이터 작업을 처리합니다. Spark가 사용되지 않는 경우를 판별하기 위해 다음과 같은 지침이 적용됩니다.

1. 데이터 세트가 128MB 미만일 경우, Analytic Server가 Analytic Server JVM에서 임베드된 MapReduce 함수를 사용하고, Spark 또는 Hadoop 클러스터를 사용하지 않습니다.
2. Spark가 클러스터에 설치되지 않은 경우, Analytic Server가 MapReduce v2를 사용합니다.
3. Analytic Server는 MapReduce v2를 사용하여 PSM 모델을 작성합니다. PSM 모델 작성으로 작업이 종료된 경우, Analytic Server는 Spark를 사용하여 모델 작성으로 이어지는 모든 단계를 통해 작업을 처리한 다음 디스크에 기록하고 MapReduce를 사용하여 PSM 모델을 작성합니다. 예를 들어, 작업에 결합이 포함되고 결합 후에 PSM 모델 작성이 수행되는 경우, 작업은 Spark에서 실행되고 PSM은 MapReduce의 결합된 데이터에서 수행됩니다.

Spark 사용 방식

Analytic Server 서비스가 시작되어 Spark가 사용 가능한 것으로 검색될 경우, 클러스터 전체에서 분산된 태스크 간 통신을 허용하는 "Spark Hadoop 작업"을 초기화합니다. 이 작업은 Analytic Server 서비스가 실행되는 동안에 실행되며, 모든 Analytic Server 실행에 사용됩니다. 이 접근법을 사용하면 각 Hadoop 작업의 모든 Analytic Server 구성요소를 다시 로드하는 오버헤드가 방지되므로 여러 개의 MapReduce Hadoop 작업을 조정하는 것에 비해 성능이 향상됩니다.

Spark는 MapReduce 작업을 실행할 수 있습니다. 따라서 Analytic Server는 가능한 경우 "네이티브" Spark 알고리즘(예: 결합, 정렬, 조합)을 사용합니다. 동시에 Analytic Server는 Hadoop API를 직접 사용하지 않고 Spark에서 기존 SPSS Map 및 Reduce 알고리즘을 실행할 수 있습니다.

버전 3.2.1의 새로운 사항

버전 3.2.1

플랫폼

Hadoop Data Platform(HDP) 3.0 및 3.1 지원.

Cloudera 6.0 및 6.1 지원.

Rank 함수

`rank` 함수는 입력 데이터 세트를 개별 파티션으로 분할하고 각 파티션 행의 순위를 표시하는 새 필드를 생성하는 데 사용합니다. 이 기능은 Hive 함수 **`rank()`**, **`dense_rank()`** 및 **`row_number()`**와 비슷합니다.

UDF Hive 푸시백

새 Hive UDF 함수가 도입되었습니다. Hive UDF가 HiveDB에 등록되고 나면 Analytic Server는 새 UDF 함수를 사용하여 푸시백을 수행할 수 있습니다.

최신 시스템 요구사항 정보는 IBM 기술 지원 사이트(http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html)에서 제공하는 자세한 시스템 요구사항 보고서를 참조하십시오. 이 페이지에서 다음을 수행하십시오.

1. SPSS Analytic Server를 제품 이름으로 입력하고 **검색**을 클릭하십시오.
2. 원하는 버전과 보고서의 범위를 선택한 다음 **제출**을 클릭하십시오.

주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다. 이 자료는 IBM에서 다른 언어로 제공할 수도 있습니다. 그러나 자료에 접근하기 위해서는 해당 언어로 된 제품 또는 제품 버전의 사본이 필요할 수 있습니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. 이 책에서 IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않는 한, 기능상으로 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수도 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운영에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 책에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 책을 제공한다고 해서 특허에 대한 라이센스까지 부여하는 것은 아닙니다. 라이센스에 대한 의문사항은 다음으로 문의하십시오.

07326
서울특별시 영등포구
국제금융로 *10, 3IFC*
한국 아이.비.엠 주식회사
대표전화서비스*: 02-3781-7114*

2바이트(DBCS) 정보에 관한 라이센스 문의는 한국 IBM에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 책을 "현상태대로" 제공합니다.일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및/또는 프로그램을 사전 통지 없이 언제든지 개선 및/또는 변경할 수 있습니다.

이 정보에서 언급되는 비IBM 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

(i) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함) 간의 정보 교환 및 (ii) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이센스 사용자는 다음 주소로 문의하십시오.

07326
서울특별시 영등포구
국제금융로 *10, 3IFC*
한국 아이.비.엠 주식회사
대표전화서비스*: 02-3781-7114*

이러한 정보는 해당 조건(예를 들면, 사용료 지불 등)하에서 사용될 수 있습니다.

이 정보에 기술된 라이센스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이센스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 프로그램 라이센스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

인용된 성능 데이터와 고객 예제는 예시 용도로만 제공됩니다. 실제 성능 결과는 특정 구성과 운영 조건에 따라 다를 수 있습니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 제품들을 테스트하지 않았으므로, 비IBM 제품과 관련된 성능의 정확성, 호환성 또는 기타 청구에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다.

여기에 나오는 모든 IBM의 가격은 IBM이 제시하는 현 소매가이며 통지 없이 변경될 수 있습니다. 실제 판매가는 다를 수 있습니다.

이 정보는 계획 수립 목적으로만 사용됩니다. 이 정보는 기술된 제품이 GA(General Availability)되기 전에 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 인물 또는 기업의 이름과 유사하더라도 이는 전적으로 우연입니다.

저작권 라이센스:

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 인물 또는 기업의 이름과 유사하더라도 이는 전적으로 우연입니다.

이러한 샘플 프로그램 또는 파생 제품의 각 사본이나 그 일부에는 반드시 다음과 같은 저작권 표시가 포함되어야 합니다.

© IBM 2019. 이 코드의 일부는 IBM Corp.의 샘플 프로그램에서 파생됩니다.

© Copyright IBM Corp. 1989 - 20019. All rights reserved.

상표

IBM, IBM 로고 및 ibm.com은 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.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 및/또는 기타 국가에서 사용되는 Adobe Systems Incorporated의 등록상표 또는 상표입니다.

IT Infrastructure Library는 현재 Office of Government Commerce의 일부인 Central Computer and Telecommunications Agency의 등록상표입니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Intel Centrino, Intel Centrino 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation 또는 그 계열사의 상표 또는 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

ITIL은 미국 특허청(U.S. Patent and Trademark Office)에 등록된 The Minister for the Cabinet Office의 등록상표 및 등록 공동체 상표입니다.

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

Cell Broadband Engine은 미국 또는 기타 국가에서 해당 라이센스에 의거하여 사용되는 Sony Computer Entertainment, Inc.의 상표입니다.

Linear Tape-Open, LTO, LTO 로고, Ultrium 및 Ultrium 로고는 미국 및 기타 국가에서 사용되는 HP, IBM Corp. 및 Quantum의 상표입니다.