



IBM InfoSphere DataStage V11.5를 사용한 유연한 통합

주요 특징

- 데이터 볼륨이나 복잡성과 관계없이 모든 크기의 데이터에 맞춰 확장
- 다양한 소스 전체에서 민첩하고 재사용 가능한 통합 기능 제공
- 사용자가 비즈니스 변화에 신속하게 대응하도록 지원
- 빠른 처리와 확장성 향상을 위해 Apache Hadoop에서 기본적으로 실행되도록 기능 제공
- 기존 및 빅 데이터 프로젝트에 대한 셀프 서비스 데이터 통합 지원
- 클라우드에서 사용 가능 - 기존 워크로드 이전 및 신규 배포의 가치 실현을 가속화
- 클라우드 환경에 대한 통합 제공
- 강화된 보안 제공

대량 병렬 처리(MPP) 아키텍처 상에 구축된 IBM® InfoSphere® DataStage® V11.5는 기업이 대량의 이기종 데이터를 손쉽게 변환 및 통합할 수 있도록 설계되었습니다. InfoSphere DataStage를 사용하면 한 번의 작업 설계로 어디서든 배포할 수 있으므로 성능이 향상되고 좀 더 민첩하게 통합할 수 있으며 비용이 절감됩니다.

새로운 컴퓨팅 시대가 열리면서 데이터의 양, 다양성 및 속도가 폭발적으로 증가하고 있습니다. 기업이 비즈니스 인텔리전스 목표를 달성하고 경쟁 우위를 확보하려면 환경 어디에서나 데이터를 통합하고 이를 신뢰할 수 있는 정보로 변환할 수 있어야 합니다. 계속해서 요구 사항이 변하고 데이터 볼륨이 증가하는 이러한 동적 환경에서는 정보를 빠르고 효율적으로 통합하는 능력이 매우 중요합니다.

InfoSphere Information Server의 엔드 투 엔드 정보 기능을 통해 데이터를 이해하고 정리, 모니터링, 변환 및 전달할 수 있을 뿐만 아니라 비즈니스와 IT의 간극을 메울 수 있습니다. InfoSphere Information Server가 제공하는 “어디서나 통합” 기능을 활용하면 빅 데이터 및 영향도 분석에서부터 마스터 데이터 관리(MDM) 및 데이터 웨어하우징까지 비즈니스와 전략적 이니셔티브의 동력이 되는 정보가 일관되고, 실시간으로 관리되며, 신뢰할 수 있는 정보가 되도록 보장할 수 있습니다.



DataStage 기능으로 복잡한 통합 문제 해결

InfoSphere DataStage V11.5는 정보가 풍부한 오늘날의 환경에 적합한 새로운 강력한 기능을 제공합니다.

빠르게 액세스하여 데이터를 정리할 수 있는 간편한 확장성 및 뛰어난 성능

데이터가 상주하는 하둡(Hadoop)에서 기본적으로 실행되는 대량 병렬 처리 엔진은 풍부한 통합 및 관리 기능을 지원합니다. 기업에서는 이러한 기능을 사용하여 데이터 연결, 변환, 정리, 보강 및 전달 기능 등과 같은 데이터 통합 및 관리 활동의 성능과 속도를 향상할 수 있습니다. 예를 들면, 다음과 같습니다.

- 모든 연결, 변환 및 데이터 전달 기능이 이제 하둡에서 기본적으로 실행될 수 있습니다.
- Kerberos나 보안 게이트웨이와 같은 중요한 보안 기능을 비롯하여 다양한 형식의 파일과 문자 세트로 이루어진 하둡 분산 파일 시스템(HDFS)에 대해 확장되고 간소화된 액세스를 제공합니다.

생산성 향상으로 저렴한 통합 비용

InfoSphere Information Server 엔진에서는 개발자가 내장된 그래픽 디자인 도구 기능을 사용할 수 있습니다. 이를 사용하는 고객의 경험에 비추어 볼 때 수동 핸드 코딩과 비교하여 상당한 비용을 절감할 수 있습니다.

클라우드에서 사용 가능

새로운 작업 방식 - 모든 유형의 데이터를 통합하는 하이브리드 환경을 위한 유연성

전체 InfoSphere DataStage 인프라를 클라우드로 이전하길 원하든 개발이나 테스트 환경을 개발하여 작게 시작하길 원하든, IBM DataStage on Cloud는 빠른 배포와 유연한 구독 요금을 통해 고객의 데이터 통합 요구 사항을 지원할 준비가 되어 있습니다.

IBM DataStage가 처음이라면 사내 구축형 하드웨어 구매와 부서 및 예산 승인 과정을 거칠 필요 없이 위험도가 낮은 구독 요금의 DataStage on Cloud로 클라우드에서 시작하고 비즈니스에 필요한 전체 ETL 인프라를 신속하게 배포할 수 있습니다.

추가 비용 없이 제공되는 부가 기능

기업에서는 널리 사용되는 엔터프라이즈 애플리케이션에 액세스하는 등 다음과 같은 부가 기능을 활용할 수 있습니다.

- Salesforce.com
- Hyperion Essbase
- Siebel
- JD Edwards EnterpriseOne
- PeopleSoft Enterprise
- Oracle 애플리케이션

또한, InfoSphere DataStage Balanced Optimization 기능을 통해 푸시다운 최적화를 사용할 수 있습니다. 즉, 관계형 데이터베이스와 InfoSphere DataStage의 가용 용량과 컴퓨팅 능력을 충분히 활용할 수 있습니다.

클라우드 환경을 위한 데이터 통합

InfoSphere DataStage V11.5는 클라우드 환경을 위한 신속하고 간편한 데이터 통합 기능을 제공합니다. 사내 구축형에서 실행하는 경우, 클라우드로 데이터를 송수신할 수 있도록 Amazon Simple Storage System(S3)과의 직접 통합을 지원합니다. 데이터가 S3로 통합된 후에는 다른 클라우드 데이터베이스 기술과도 통합될 수 있습니다.

또한, InfoSphere DataStage V11.5에는 REST API(Application Programming Interface)와의 통합을 지원하는 계층적 단계("XML" 단계에서 확장 및 이름 변경됨)가 포함되어 있으므로 XML 및 JSON(JavaScript Object Notation) 메시지에 대한 통합이 지원됩니다.

InfoSphere Information Server는 REST 기반 연결을 통해 IBM Cloudant®와 같은 분산 DBaaS(서비스형 데이터베이스) 오퍼링뿐만 아니라 REST 기반 상호 작용을 제공하는 기타 온프레미스 및 오프프레미스 솔루션을 지원할 수 있습니다.

마스터 데이터 관리(MDM) 프로젝트를 위한 긴밀한 통합

InfoSphere DataStage는 MDM 통합 단계를 통해 InfoSphere MDM과 직접 통합될 수 있으므로 사용자는 손쉽게 InfoSphere DataStage를 사용하여 InfoSphere MDM으로 데이터를 로드하고 InfoSphere MDM에서 데이터를 추출할 수 있습니다. MDM 통합 단계를 활용함으로써 고객은 MDM 데이터를 자체 데이터 통합 흐름에 포함하고 도메인 데이터(고객, 파트너, 공급자, 제품 및 기타 데이터)를 직접 MDM 시스템에 로드할 수 있습니다. 또한, 이제 사용자는 대용량 로드를 활용하여 최고의 성능을 실현함으로써 최대 규모 및 성능의 혜택을 누릴 수도 있습니다.

InfoSphere Information Server 데이터 변환 및 전달 방식의 이점을 활용하는 것 외에도 MDM 통합 단계에서는 InfoSphere Information Server 데이터 품질 기능도 활용할 수 있습니다. 예를 들어 InfoSphere Information Server는 어떤 데이터든 데이터가 MDM 시스템으로 로드되기 전에 이를 표준화할 수 있습니다. 이러한 방법으로 데이터 품질을 사전에 관리함으로써 프로젝트는 매칭 정확도 향상과 다양한 항목에 대한 360도 전방위 관점의 지원이라는 혜택을 누릴 수 있습니다.

빅 데이터 통합: 빅 데이터 성공의 열쇠

오늘날과 같이 빠르게 움직이는 비즈니스 환경에서 빅 데이터 분산으로 데이터를 빠르고 간편하게 전송하는 기능은 반드시 필요한 기능입니다. MPP 아키텍처 덕분에 InfoSphere DataStage는 하둡 환경의 까다로운 데이터 통합 요구 사항에 맞춰 손쉽게 확장될 수 있습니다. InfoSphere DataStage의 "한 번 설계로 어디서나 실행"이라는 접근 방식은 사용자가 단일

시스템과 저상용성(low-commodity) 서버 클러스터 간에 데이터 통합 작업을 손쉽게 이동할 수 있게 해줍니다. 작업 설계가 런타임으로부터 분리되어 있으므로 개발자는 해당 아키텍처를 위한 코딩이 아니라 현재 비즈니스 요구 사항에 집중할 수 있습니다.

InfoSphere DataStage는 처음부터 MPP상에 구축되었으며, "빅 데이터"라는 용어가 생기기 전부터 대규모 데이터 볼륨을 지원해 온 통합 플랫폼입니다. InfoSphere DataStage는 다음과 같이 빅 데이터 프로젝트를 신속하고 비용 효율적으로 진행하는 데 도움이 되는 강력한 빅 데이터 관련 기능을 제공합니다.

- **하둡용 Balanced Optimization:** 이 기능을 사용하면 통합 설계자가 기존 ETL(Extract, Transform and Load)에서 사용하는 것과 동일한 설계 패러다임을 사용하여 작업을 구축하고, 필요할 때 생성된 MapReduce를 통해 해당 로직을 실행할 수 있습니다. 따라서 MapReduce 언어 또는 특정 환경에만 적용되는 다른 보조 도구 세트로 교체하고 팀을 다시 교육하는 비용이 들지 않습니다.
- **빅 데이터 작업 시퀀스:** InfoSphere DataStage를 사용하면 Oozie가 포함된 모든 MapReduce 작업을 작업 시퀀스에 포함할 수 있습니다. 기업에서는 단일 그래픽 워크플로우 구문으로 데이터를 하둡으로 로드하고, 자체 개발된 MapReduce 분석 프로그램을 실행한 후, 분석 결과를 데이터 웨어하우스로 로드할 수 있습니다. 이는 InfoSphere Information Server와 하둡 양쪽에서 실행되는 이기종 토폴로지 전체에서 엔드 투 엔드 워크플로우를 제공합니다.
- **빅 데이터 거버넌스:** InfoSphere DataStage는 통합 지점에 대한 영향 분석 및 데이터 계보와 같은 빅 데이터 관련 거버넌스 기능을 지원하므로 기업의 통찰력을 유지하면서 분석을 확장할 수 있습니다.

- **IBM InfoSphere Streams 통합:** 실시간 분석 처리를 주로 다루는 빅 데이터 프로젝트를 위해 이제 IBM에서는 InfoSphere DataStage와 InfoSphere Streams 간에 데이터 흐름을 직접 통합하는 기능을 제공합니다. 기업에서는 표준 데이터 통합 규칙을 사용하여 실시간 분석 프로세스로 정보를 수집 및 전달할 수 있습니다.
- **빅 데이터 액셀러레이터:** 이 오픈 소스 구성 요소는 ibm.com/developerworks에서 사용할 수 있습니다. 이 액셀러레이터는 InfoSphere DataStage 환경에 바로 플러그인으로 연결되며 다른 단계와 마찬가지로 작동합니다. MongoDB, Hive, Cassandra, HBase, Avro 등을 위한 액셀러레이터가 제공됩니다.
- **계층적 데이터 지원으로 확장:** 버전 11.5에서는 InfoSphere DataStage가 REST API에 대한 지원을 제공하도록 확장되므로 XML이나 JSON 메시지와 같은 계층적 데이터에 쉽게 액세스하고 이러한 데이터를 통합할 수 있습니다.

민첩한 셀프 서비스 데이터 프로비저닝 및 거버넌스

InfoSphere DataStage V11.5에 포함된 InfoSphere Data Click을 사용하면 작업을 완료하는 데 걸리는 시간을 며칠 또는 몇 주에서 몇 분 또는 몇 시간으로 단축함으로써 가치 실현을 가속화하고, 비즈니스 민첩성을 향상하며, 비용을 낮출 수 있습니다. 폭넓은 연결 기능을 제공하여 빅 데이터 랜딩 존(landing zone), 데이터 웨어하우스, 클라우드 환경을 비롯하여 다양한 유형의 환경의 적시성을 개선하는 데 도움이 됩니다. InfoSphere Data Click은 JDBC 및 ODBC를 통해 빠른 성능을 위한 특정 DBMS에 대한 기본 연결과 광범위한 데이터 소스를 위한 범용에 가까운 연결을 모두 제공합니다(그림 1).

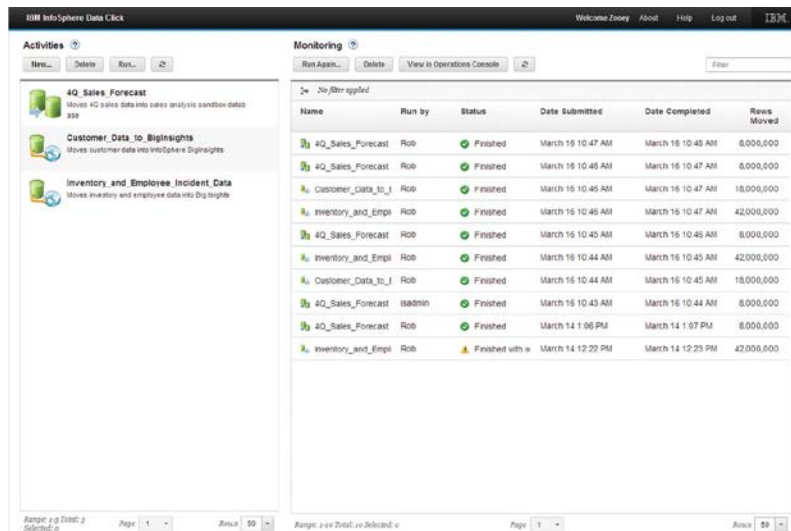


그림 1. 사용자는 간소화된 InfoSphere Data Click UI로 직접 로그인하여 활동을 빠르고 간편하게 생성, 수정, 실행 및 모니터링할 수 있습니다.

데이터는 IBM DB2®, Oracle, EMC Greenplum, IBM Informix®, IBM PureData® for Analytics(IBM Netezza® 기반), Salesforce.com, Microsoft SQL Server, Teradata, Amazon S3 등을 비롯한 기존 및 빅 데이터 환경에서 가져오거나 이러한 환경으로 전송될 수 있습니다. InfoSphere Data Click을 사용하면 누구나 IBM Biginsights® 환경(하둡) 내에서 데이터를 프로비저닝할 수 있습니다.

InfoSphere Information Server의 강화된 보안

InfoSphere Information Server는 데이터 통합, 데이터 품질 및 데이터 거버넌스 기능을 지원하는 일반적인 플랫폼 서비스(연결 서비스, 관리 서비스, 배포 서비스 등)를 제공합니다. 클라우드 기능 향상과 대규모 데이터 볼륨 지원과 더불어 V11.5에서는 다음을 지원합니다.

- **SSO(Single Sign-On):** 이제 브라우저 기반 클라이언트에서 SSO를 지원하므로 하나의 인터페이스로 인증하는 고객이 다른 사용자 인터페이스와 좀 더 원활하게 상호 작용할 수 있습니다.
- **SSL(Secure Sockets Layer) 통신:** InfoSphere Information Server에서는 SSL을 사용하여 모든 클라이언트 인터페이스에 대한 통신 보안을 제공합니다.
- **더욱 강력한 암호화:** RSA-2048 및 SHA-512가 기본 암호화 메커니즘으로 제공됩니다.
- **셀 공유:** InfoSphere Information Server는 IBM WebSphere® Application Server 표준 보안 도메인을 충분히 활용합니다. 따라서 기존 셀에 이미 배포된 프로파일과 애플리케이션을 방해하지 않고 InfoSphere Information Server를 보안 배포 관리자가 관리하는 기존 셀에 배포할 수 있습니다.

다음 단계: InfoSphere Information Server for Data Integration

InfoSphere DataStage와 InfoSphere Information Server를 함께 사용하면 기업이 새롭게 등장하는 데이터 출처에서 나오는 빅 데이터를 유연하고 강력하게 관리할 수 있습니다.

데이터 통합 요구 사항이 더욱 다양한 InfoSphere DataStage 사용자라면 InfoSphere Information Server for Data Integration V11.5를 고려해보는 것도 좋습니다. 이 솔루션에는 변경 데이터 전달, 실시간 데이터 통합, 데이터 모델링, 블루프린트, 데이터 검색, 통제된 메타데이터 관리와 같이 자주 요청되는 데이터 통합 기능과 더불어 InfoSphere DataStage가 포함되어 있습니다. 또한, 설계 시간을 가속화하고, 소스와 대상 매핑을 생성하며, 작업을 자동으로 생성하는 기능도 포함되어 있습니다.

또한, IBM은 다른 정보 통합 요구 사항에 부합하기 위해 InfoSphere Information Server for Data Quality, InfoSphere Information Governance Catalog 및 InfoSphere Information Server Enterprise Edition을 제공합니다. 마지막으로 데이터 통합, 데이터 품질 및 정보 거버넌스라는 세 가지 기능 세트가 모두 포함된 포괄적인 오퍼링도 제공합니다.

왜 IBM인가?

IBM Watson™ Foundations의 핵심 요소인 InfoSphere Information Integration and Governance(IIG)는 빅 데이터 문제를 해결할 수 있는 업계 최고의 기능을 제공합니다. InfoSphere IIG는 대규모 데이터 볼륨을 위한 최적의 확장성과 성능, 증가하는 데이터 속도를 위한 적절한 크기의 통합 및 거버넌스, 사내 구축형 또는 클라우드의 다양한 데이터 유형 및 빅 데이터 시스템을 위한 지원 및 보호 기능을 제공합니다. InfoSphere IIG를 사용하면 비즈니스 사용자가 통찰력에 근거하여 자신 있게 일을 추진할 수 있으므로 빅 데이터 및 분석 프로젝트를 성공적으로 수행하는 데 도움이 됩니다.

InfoSphere 기능은 다음과 같습니다.

- **메타데이터, 비즈니스 용어 및 정책 관리:**
InfoSphere Information Governance Catalog로 메타데이터, 비즈니스 용어 및 거버넌스 정책을 정의합니다.
- **데이터 통합:** 배치 데이터 변환 및 이동(InfoSphere Information Server), 실시간 복제(InfoSphere Data Replication) 및 데이터 연동(InfoSphere Federation Server)을 비롯하여 모든 통합 요구 사항을 처리합니다.
- **데이터 품질:** 기업 데이터를 InfoSphere Information Server for Data Quality로 구분 분석, 표준화, 유효성 검증 및 매칭합니다.
- **MDM:** InfoSphere MDM을 통해 고객, 제품, 공급자, 위치 및 계정의 신뢰할 수 있는 뷰를 기반으로 수행합니다.
- **데이터 수명 주기 관리:** IBM InfoSphere Optim™을 통해 테스트 데이터 생성에서부터 폐기 및 아카이브에 이르기까지 데이터 수명 주기를 관리합니다.
- **데이터 프라이버시:** InfoSphere Optim을 사용하여 민감한 데이터를 마스킹하고 보호하도록 지원합니다.

자세히 알아보기

InfoSphere Information Server, InfoSphere DataStage for Data Integration, InfoSphere Information Server 또는 DataStage on Cloud에 대한 자세한 내용은 IBM 영업 대표 또는 IBM 비즈니스 파트너에 문의하시거나, 다음 웹 사이트를 참조하십시오.

- ibm.com/software/products/en/ibminfodata
- ibm.com/software/products/en/infoinfoservfordatainte

또한, IBM Global Financing은 비즈니스에 필요한 기술 역량을 갖출 수 있도록 다양한 지불 옵션을 지원합니다. 구매에서 폐기까지 IT 제품 및 서비스의 전체 수명 주기에 대한 관리를 제공합니다. 자세한 내용은 ibm.com/financing/kr에서 확인하실 수 있습니다.



© Copyright IBM Corporation 2016

IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
2016년 6월

IBM, IBM 로고, ibm.com, BigInsights, Cloudant, DataStage, DB2, IBM Watson, Informix, InfoSphere, Optim, PureData 및 WebSphere는 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보"(ibm.com/legal/copytrade.shtml)에 있습니다.

Netezza는 IBM Company인 IBM International Group B.V.의 상표입니다.

Microsoft 및 SQL은 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다. Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.

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 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.



재활용하십시오.