

# Informatica Big Data Management Edition

## 이점

- 혁신적인 제품과 서비스를 시작에 빠르게 도입하고 비즈니스 운영 향상
- 증가하는 데이터 볼륨과 복잡성을 처리하면서 빅 데이터 관리 비용 절감
- 프로젝트 전체에서 하둡 채택을 확대하여 성능 향상 및 비용 절감
- 새로운 기술의 복잡성을 해소하고 민감한 데이터를 보호하는 입증된 데이터 통합 소프트웨어에 대한 투자로 위험 최소화

## 페타바이트를 수익으로 전환

비즈니스의 속도가 빨라지고 비즈니스 혁신에 대한 엄청난 경쟁 압박에 직면하는 상황에서 조직은 데이터 아키텍처를 현대화하고 최적화하여 데이터를 조직의 의사 결정에 필요한 전략적 자산으로 만들 수 있습니다.

조직은 데이터를 사용해 고유한 통찰력을 도출하여 수익성을 강화하고, 기회를 발견하고, 제품과 서비스 제공 속도를 높이고, 우수한 고객 환경을 전달함으로써 차별화를 이루고 성장할 수 있습니다. 이와 동시에 데이터를 통해 사기, 범죄, 시스템 다운타임 또는 보안 위반의 위험을 완화하면서 운영 효율성을 높일 수도 있습니다.

Informatica Big Data Management는 업계 최초이자 가장 포괄적인 데이터 관리 솔루션을 제공하며 빅 데이터를 대규모로 수집, 처리, 정제, 관리 및 보호하도록 설계되었습니다. Informatica 솔루션은 빅 데이터 통합, 빅 데이터 품질 및 거버넌스, 빅 데이터 보안을 하나의 통합 솔루션으로 제공합니다.

## 동적이고 최적화된 빅 데이터 통합

Informatica Big Data Management에서는 높은 처리량으로 데이터를 수집하고 대규모 데이터 통합을 처리할 수 있어 비즈니스 분석가가 차세대 분석 툴을 사용해 더 나은 결정을 내릴 수 있습니다. 사전 구축된 수백 개의 고성능 커넥터, 데이터 통합 변환, 파서를 통해 거의 모든 유형의 데이터를 하둡, NoSQL, MPP 어플라이언스 등의 빅 데이터 플랫폼에서 신속하게 수집하여 처리할 수 있습니다. 동적 매핑 및 매개변수화를 통해 데이터 통합 프로세스를 프로그램 방식으로 자동화할 수도 있습니다.

단 몇 개의 템플릿을 사용하여 설계 패턴에 따라 런타임에 모든 유형의 데이터 흐름을 자동으로 생성할 수 있으므로 생산성이 크게 개선되고 유지 보수도 한결 간편해집니다. 동적 스키마가 지원되므로 다양한 데이터 형식을 처리할 수 있습니다. 또한 최적화된 엔진은 대규모로 데이터를 통합할 수 있도록 성능 및 리소스 활용을 최대화해 줍니다. 예를 들어, Informatica는 YARN을 통해 MapReduce와 새로운 Informatica Blaze 엔진 조합을 사용하여 빅 데이터 워크로드를 최적화합니다.

## 엔드 투 엔드 빅 데이터 거버넌스 및 품질

엔드 투 엔드 빅 데이터 거버넌스 및 품질은 현업 부서 사용자와 IT 사용자가 사용 중인 데이터를 신뢰할 수 있다는 것을 의미합니다. Informatica Big Data Management는 직관적인 일반 사용자 환경을 제공하므로 데이터 과학자, 분석가, 현업 부서가 효과적으로 협업할 수 있습니다. 비즈니스 규칙 프로파일링을 포함한 데이터 프로파일링 및 검색 기능을 통해 데이터 품질 문제와 이상을 강조하여 표시하고 모니터링 규칙 및 알림을 손쉽게 생성하여 품질 문제를 추적하고 표시할 수 있습니다.

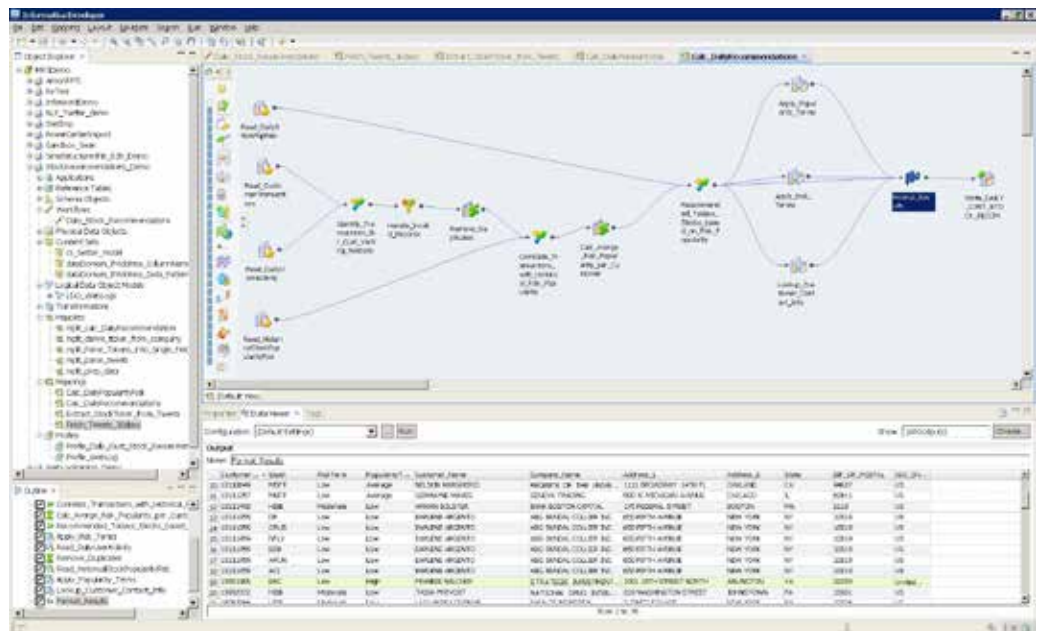
하둡에서 데이터 품질 규칙과 주소 유효성 검사를 구축하여 목적에 맞게 대규모로 데이터를 정제하고 준비할 수 있습니다. 범용 메타데이터 카탈로그 및 지식 그래프가 포함된 Live Data Map을 통해 기업 데이터를 검색하고 이해할 수 있으며 Spark를 통해 대규모의 지식 그래프를 빠르게 생성할 수 있습니다. 하둡을 넘어서는 엔드 투 엔드 데이터 계보(Lineage)는 규정 준수를 지원하며 효과적인 데이터 품질 근본 원인 및 영향 분석을 실현합니다.

## 위험 중심의 빅 데이터 보안

위험 중심의 빅 데이터 보안을 위해 위험 및 주요 취약점을 신속하게 탐지하고 조치를 취할 수 있도록 모든 데이터를 분석합니다. 이 작업을 위해서는 위험 분석 및 위험한 데이터에 대한 정책 기반 보호에서 지원하는, 민감한 데이터에 대한 360도 뷰가 필요합니다. 그러면 빅 데이터 보안에서 기업 정책 및 업계 규정에 따라 통제되는 정보를 식별 불가능하게 만듭니다. Informatica Big Data Management는 하둡 개발 및 운영 환경에서 민감한 데이터를 보호하기 위한 옵션으로 비침입적인 영구 데이터 마스킹을 제공하여 보안 위반의 위험을 최소화합니다.

"우리가 하는 모든 일의 중심에는 고객이 있습니다. Western Union은 독보적인 역량을 통해 데이터 분석 기능을 활용하여 향후 비즈니스를 형성하고 고객의 환경을 개선하는 데 도움이 될 트렌드를 식별할 수 있습니다. 앞으로도 Informatica와 협력하여 전 세계적인 비즈니스 및 고객 기반으로 진정한 잠재력을 실현해 나갈 것입니다."

- Sanjay Saraf, Western Union 수석 부사장 겸 CTO



Informatica Big Data Management는 사전 구축된 변환을 제공하고 재사용을 촉진하는 시각적 개발 환경을 사용하여 핸드 코딩에 비해 최대 5배 높은 생산성을 실현합니다.

## 주요 특징

### 하둡에서의 데이터 통합(ETL)

이 솔루션은 하둡에서 기본적으로 실행되는 사전 구축된 데이터 통합 변환 기능의 광범위한 라이브러리를 제공하므로 테라바이트에서 페타바이트에 이르는 모든 크기의 데이터 유형을 처리할 수 있습니다. IT 팀에서는 핸드 코딩에 비해 최대 5배 높은 생산성을 제공하는 시각적 개발 환경을 사용해 하둡에서 데이터 흐름을 신속하게 개발할 수 있습니다.

### 동적 스키마 및 매핑 템플릿

Informatica Big Data Management에서는 매핑 템플릿을 사용해 적은 수의 설계 패턴으로 수백 개의 런타임 데이터 흐름을 생성할 수 있습니다. 이러한 매핑은 빅 데이터 프로젝트에 일반적으로 포함되는 웹 및 기기 로그파일과 같은 동적 스키마를 처리하기 위해 간편하게 매개변수화할 수 있습니다. 즉, 유지 보수가 쉽고 스키마 변경에 유연하게 대응하는 데이터 흐름을 신속하게 구축할 수 있습니다.

### 하둡에서의 데이터 프로파일링

Informatica 개발자 툴 및 브라우저 기반 분석 툴을 통해 하둡에서 데이터를 프로파일링할 수 있습니다. 따라서 개발자, 분석가 및 데이터 과학자들은 쉽게 데이터를 이해하고, 데이터 품질 문제를 조기에 확인하며, 데이터 흐름 규격에 관해 협업하고, 변환 매핑 및 규칙 로직의 유효성을 검사할 수 있습니다.

### 하둡에서의 데이터 품질

하둡에서 모든 유형과 볼륨의 데이터를 정제하고 매칭하고 표준화하여 인증되고 신뢰할 수 있는 데이터를 제공합니다. 사전 구축된 데이터 품질 규칙의 확장 세트를 사용하거나 시각적 개발 환경에서 자체 규칙을 생성할 수 있습니다. 주소 유효성 검사를 실행해 글로벌 주소 데이터를 파싱하고 정제하고 표준화하고 보완할 수 있습니다.

### 하둡에서의 복합 데이터 파싱

Informatica Big Data Management를 사용하면 웹 로그, JSON, XML, 머신 기기 데이터와 같은 복합적이고 다양한 정형, 비정형, 산업 표준 데이터에 간편하게 액세스하고 파싱할 수 있습니다. SWIFT, ACORD, HL7, HIPAA 및 EDI와 같은 산업 표준 데이터 및 업계 데이터를 위한 사전 구축된 파서를 사용할 수 있습니다.

### 범용 메타데이터 카탈로그

데이터 과학자와 분석가는 이제 엔터프라이즈 데이터를 검색하고 발견하고 이해할 수 있는 Live Data Map, 범용 메타데이터 카탈로그, 지식 그래프를 통한 360도 데이터 뷰를 사용할 수 있습니다. Live Data Map에서는 Spark를 활용하여 대규모 지식 그래프를 빠르게 생성할 수 있습니다.

### 엔드 투 엔드 데이터 계보(Lineage)

신뢰도와 규정 준수를 위해 데이터 분석가와 현업 부서 사용자가 완전한 엔드 투 엔드 데이터 계보(Lineage)를 볼 수 있습니다. 이러한 시각적 데이터 계보(Lineage)에는 모든 타겟 애플리케이션부터 원본 소스 시스템에 이르기까지 하둡 및 기존 시스템에서 일어나는 모든 데이터 이동과 변환에 대한 자세한 이력이 포함됩니다. 데이터 객체 및 해당 데이터 계보(Lineage)와 관련된 공통 비즈니스 용어가 정리된 비즈니스 글로서리를 통해 현업 부서와 IT 부서의 협업 및 검색이 향상되었습니다.

### 하둡에서의 Persistent Data Masking

Persistent Data Masking 옵션을 사용하면 민감한 데이터의 민감도를 낮추고 식별 불가능하게 만들어 데이터 보안과 기밀 유지를 강화합니다. 모호한 데이터의 컨텍스트를 보존하고 참조 무결성을 일관되게 유지하여 테스트, 분석 또는 지원 환경에서 마스킹된 데이터를 사용할 수 있습니다.

### 범용 데이터 액세스

IT 팀이 RDBMS, OLTP, OLAP, ERP, CRM, 메인프레임, 클라우드 등 모든 유형의 트랜잭션 빅 데이터에 액세스할 수 있습니다. 소셜 미디어 데이터, 로그 파일, 머신 센서 데이터, 하둡, NoSQL 형식, 문서, 이메일, 기타 비정형 또는 다양한 정형 데이터 유형 및 데이터 저장소에도 액세스할 수 있습니다.

### 고속 데이터 수집 및 추출

소스와 타겟 시스템 간에 또는 하둡, NoSQL 데이터 저장소, 데이터 웨어하우스에서 직접 빅 데이터에 액세스, 적재, 변환 및 추출할 수 있습니다. 병렬 처리를 사용하여 기본 API를 통해 소스 및 타겟 시스템으로의 고성능 연결을 제공함으로써 고속 데이터 통합 및 추출이 가능합니다.

### 하둡에서의 데이터 검색

하둡에서 데이터 도메인 및 관계 검색을 자동화합니다. 예를 들어, 규정 준수를 위해 마스킹할 수 있도록 고객 및 제품 관련 데이터 세트 또는 주민등록번호나 신용 카드 번호와 같이 민감한 데이터를 검색할 수 있습니다.

### 하둡에서의 자연어 처리

키워드나 구문을 사용하여 이메일, 소셜 미디어 데이터, 보험금 청구, 서비스 레코드 및 텍스트 파싱과 NER(Named Entity Recognition)을 사용하는 문서 등의 비정형 데이터로부터 고객과 제품에 관련된 엔티티를 간편하게 추출하고 분류할 수 있습니다. 경쟁력 있는 가격과 같은 고객 행동이나 제품 정보에 대한 통찰을 통해 마스터 데이터를 강화할 수 있습니다.

### 단 한 번의 설계로 신속하게 구축

하둡 에코시스템은 오픈 소스 커뮤니티에서 끊임없이 부상하는 새로운 혁신을 통해 빠르게 변화하고 있습니다. Big Data Management는 오픈 소스 하둡 프레임워크상에 구축되며 모든 변환 로직을 데이터 파이프라인에 보관합니다. 따라서 개발자는 하둡의 개념과 언어에 대한 전문 지식 없이도 한 번 설계하여 하둡에 변경 사항이 있을 때마다 재구축하지 않고 데이터 파이프라인을 간편하게 구축할 수 있습니다. 그 결과 운영계 시스템에 영향이나 위험을 적게 미치면서 하둡 혁신을 빠르게 구현할 수 있습니다.

다음 Informatica Big Data Management 에디션을 비교하여 적합한 에디션을 선택하십시오.		
	Enterprise	Advanced
데이터베이스, 하둡, 소셜 미디어, 앱에 대한 연결	●	●
고속 데이터 수집	●	●
하둡에서의 데이터 통합	●	●
데이터베이스 푸시다운 최적화	●	●
하둡에서의 데이터 프로파일링 및 도메인 검색	●	●
복합 데이터 파싱	●	●
텍스트 파싱 - NER(Named Entity Recognizer)	●	●
동적 스키마 및 매핑 템플릿	●	●
범용 메타데이터 카탈로그	●	●
엔드 투 엔드 데이터 계보(Lineage)	●	●
하둡에서의 데이터 품질		●
하둡에서의 주소 유효성 검사		●
비즈니스 글로서리		빅 데이터 환경으로 제한
100명의 Informatica 데이터 분석가 기명 사용자	●	●
지원(서브스크립션 라이선스가 있는 경우에만 포함)	8 x 5	24 x 7

## Informatica 소개

Informatica는 모든 데이터의 미래를 위한 변화의 혁신을 제공하는 선도적인 독립 소프트웨어 공급업체입니다. 전 세계 기업들이 Informatica와 함께 정보의 잠재력을 실현하고 중요한 비즈니스 과제를 해결하고 있습니다. 5,800개 이상의 기업이 Informatica를 통해 소셜 네트워크를 포함한 인터넷, 클라우드 및 기업 내에 있는 정보 자산을 최대한 활용하고 있습니다.

## 주요 이점

### 혁신적인 제품과 서비스를 시장에 빠르게 도입

데이터 과학자와 분석가는 Informatica Big Data Management를 사용해 데이터 통합이 아닌 새로운 데이터 통찰력에 초점을 맞출 수 있으며, 기업은 이를 통해 혁신적인 제품과 서비스를 개발할 수 있습니다. 이는 기존 PowerCenter® 매핑을 가져와 Informatica 리소스 기술을 재사용할 수 있는 데이터 통합 개발 환경을 제공합니다. 또한, 사용이 편리한 시각적 개발 환경 및 반복 가능한 개발 프로세스 패러다임을 제공합니다. 핸드 코딩 없이 소셜 미디어 데이터, 웹 로그, 센서 기기 데이터, 문서, 이메일 및 기타 다양한 정형 또는 비정형 데이터로 데이터 통합 및 데이터 품질 관리를 확장합니다. 따라서 이 솔루션은 IT 팀의 생산성을 획기적으로 증대하고 구축 시간을 단축합니다.

### 빅 데이터 프로젝트 비용 절감

Informatica Big Data Management를 사용하면 빅 데이터 비용이 줄어듭니다. 웨어하우스 및 소스 시스템의 데이터와 처리를 저렴한 하드웨어로 분담할 수 있습니다. 핸드 코딩이 아니라 입증된 데이터 통합 생산성 툴을 사용함으로써 생산성이 최대 5배 높아집니다. 한 번 구축해 어디에서든 재활용할 수 있어 기존 데이터 플랫폼 및 하둡과 같은 새로운 기술 모두에서 데이터 처리 활용을 최적화하여 비용을 절감할 수 있습니다.

### 기업 전체에 하둡 채택 확대

Informatica Big Data Management를 통해 여러 프로젝트에서 하둡을 광범위하게 채택할 수 있습니다. 하둡의 성능 및 비용 이점을 실현하기 위해 Java MapReduce, Python 또는 스크립팅 언어에서 핸드 코딩을 할 필요가 없으므로 빅 데이터 프로젝트 업무 수행을 위해 Informatica 교육을 이수한 전 세계 100,000명 이상의 개발자를 손조롭게 고용할 수 있습니다. 이 솔루션은 관리가 용이하며 다수의 빅 데이터 프로젝트를 지원합니다. 또한, 최적의 엔드 투 엔드 성능을 제공하고 물론 엔드 투 엔드 데이터 계보(Lineage)로 신뢰도를 보장하고 보안을 통해 위험을 최소화하며 구축을 가속화합니다.

### 새로운 기술 채택의 위험 최소화

Informatica Big Data Management는 하둡과 같은 최신 기술과 관련된 위험을 최소화합니다. IT 조직은 한 번의 투자로 새로운 기술 및 기존의 데이터 인프라 모두에서 업무 수행이 가능하도록 할 수 있습니다. 이 솔루션은 동적 스키마의 영향 및 민감한 데이터를 보호하는 기능과 관련된 위험을 최소화합니다. 또한, 모든 유형의 데이터에 대한 일관된 단일 데이터 통합 접근 방식을 제공합니다. 하둡과 같은 새로운 기술이 계속 등장하는 상황에서 Informatica Big Data Management는 매핑과 데이터 흐름을 재구축하지 않고도 새로운 혁신을 쉽게 채택할 수 있도록 합니다.

