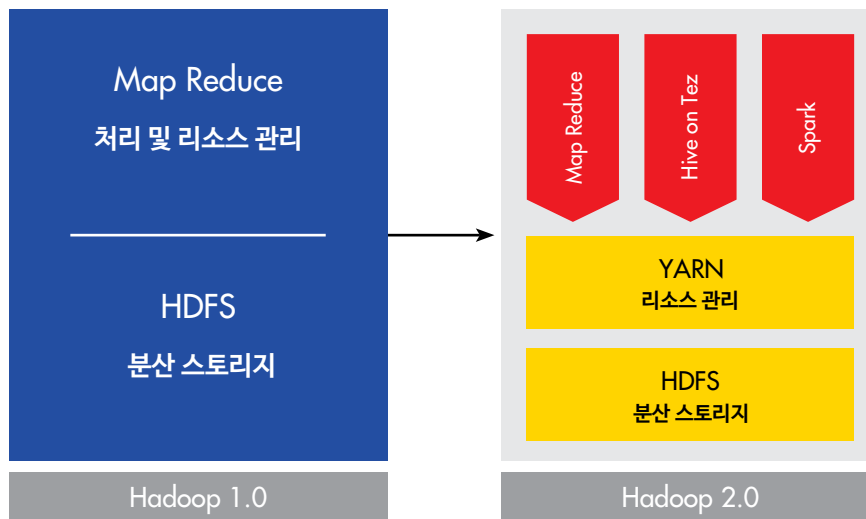


Blaze Data Management Engine

하둡에 대한 요약

하둡 에코시스템은 빅 데이터를 수집하고 조직하며 보관 및 처리하기 위한 플랫폼입니다. 빅 데이터 처리 초기에는 데이터 세트의 분산 처리와 리소스 관리 기능 제공을 위해 MapReduce를 프로그래밍 프레임워크로 사용하고 하둡 1.0은 HDFS(하둡 분산 파일 시스템)에서 빅 데이터를 저장하는 용도였습니다. 개발자는 데이터를 처리하기 위해 일련의 MapReduce 작업으로 변환된 Hive 또는 Pig 스크립트를 만들거나, Java를 사용해 MapReduce 프로그램을 써야 했습니다.



하둡 1.0에서 통합된 리소스 관리 및 프로그래밍 프레임워크의 제약은 YARN(Yet Another Resource Negotiator)의 개발로 이어졌습니다. 차세대 하둡 2.0인 YARN은 리소스 관리자로서 MapReduce의 리소스 관리 기능과 처리 엔진을 분리합니다. 하둡의 데이터 운영 체제로 종종 불리는 YARN은 복잡한 여러 분산 애플리케이션이 하나의 멀티 테넌트 플랫폼에서 실행될 수 있도록 설계되어 워크로드를 관리 및 모니터링하고, 보안 통제를 구현하며, 고가용성 기능을 관리할 수 있습니다.

데이터 엔지니어는 Informatica의 Big Data Management 솔루션을 통해 기존 데이터 흐름에 아무 변화를 주지 않고 다양한 데이터 관리 사용 사례에 맞게 최고의 하둡 처리 엔진을 위한 최적화된 지원을 특별히 제공받아 Tez 또는 Spark와 같은 새로운 하둡 혁신을 빠르게 도입할 수 있습니다.

Informatica는 비 컴퓨팅 집약적인 ETL/ELT 배치 처리를 위해 MapReduce 처리 프레임워크를 활용합니다. 또한 Informatica는 Apache Tez를 통해 성능 및 확장성의 이점을 활용해 거의 실시간으로 마이크로 배치 처리 워크로드를 지원합니다.

Informatica는 Apache Spark를 활용함으로써 예약되거나 이벤트가 적용된 전달뿐 아니라 거의 실시간으로 완전히 자동화된 데이터 전달을 지원하는 Publish/Subscribe 상호 작용 패턴을 제공합니다. 또한 Apache Spark는 모든 데이터 자산에 대한 심도 있는 정보를 지속적으로 검색 및 학습하고, 제공하는 Informatica Live Data Map과 통합되어 완벽한 포괄적 데이터 거버넌스 솔루션을 전달할 수 있습니다.

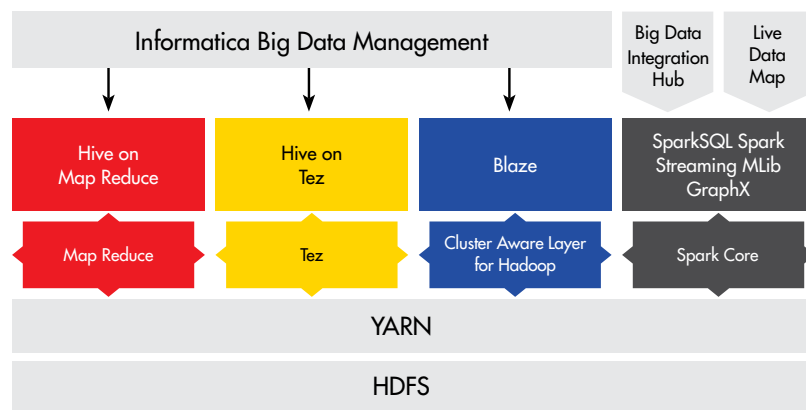
Informatica 소개

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.

Informatica Blaze

Blaze는 지능형 데이터 파이프라이닝, 작업 파티셔닝, 작업 복구 및 규모 조정 기능을 제공하도록 YARN과 통합된 업계 고유의 데이터 처리 엔진으로, Informatica의 상용 클러스터 인식 데이터 통합 기술을 활용하는 고성능 데이터 처리 기능을 제공하도록 최적화되어 있습니다.

여러 처리 모델을 지원하기 위해 하둡 외에 맞춤형 애플리케이션 프레임워크를 생성할 수 있는 기능을 제공하는 YARN은 Informatica의 데이터 변환 엔진이 기본적으로 하둡과 통합될 수 있게 했습니다. Informatica Blaze는 MapReduce 또는 Hive에 의존하지 않고 기본적으로 YARN에서 실행되는 메모리 기반의 데이터 교환 프레임워크를 사용해 구축되고 MapReduce의 기능적 차이에 주의를 기울입니다. 그리고 Informatica의 Big Data Management 솔루션을 보완하여 하둡에서의 데이터 처리 기능을 확대하고, MapReduce, Hive on Tez, Informatica Blaze 및 Spark와 같은 여러 처리 패러다임을 지원함으로써 가능한 최고의 처리 엔진에서 각 워크로드를 실행합니다.



주요 이점

- Blaze 분산 처리 엔진은 하둡에서 기본적으로 포함된 Informatica 데이터 변환 엔진을 통해 대용량의 복잡한 배치 워크로드에 대한 고속 데이터 처리 작업을 수행하고 규모를 조정할 수 있습니다.
- Cluster Aware Optimizer가 하둡 또는 기본 Informatica 서버를 사용해 Informatica Blaze, Hive on Tez 및 Hive on MapReduce와 같은 다양한 실행 엔진 전체에 걸쳐 런타임에 최적의 배치 처리 레이어를 자동으로 결정합니다.

최고의 오픈 소스 기술, YARN 그리고 20년에 걸친 데이터 관리 경험을 결합함으로써 Informatica Blaze의 도입은 하둡에서 유연하고 확장 가능한 고성능 데이터 처리 기능을 전달하는 방식으로 여러 데이터 처리 프레임워크를 위한 Informatica의 최적화된 지원을 강화합니다. 조직은 이러한 지원을 받아 궁극적으로 최적화된 빅 데이터 통합용 엔드 투 엔드 플랫폼을 보유할 수 있습니다.

Informatica는 포괄적인 Big Data Management 플랫폼을 제공하여 동적이고 최적화된 빅 데이터 통합, 엔드 투 엔드 빅 데이터 거버넌스, 위험 중심 데이터 보안을 구현해줍니다. 빅 데이터 통합은 고속 처리 데이터 수집 및 규모에 따른 처리 기능을 제공하여 비즈니스 분석가가 차세대 분석 툴을 활용하여 보다 나은 의사 결정을 내릴 수 있도록 합니다.



한국인포매티카 06611 서울시 서초구 서초동 강남대로 465 교보타워 B동 13층, 대표 전화: +82 2 6293 5000
informatica.com/kr linkedin.com/company/informatica twitter.com/InformaticaKR

© 2015 Informatica Corporation. All rights reserved. Informatica® 및 Put potential to work™는 미국 및 전 세계 관할 국가 내에서 Informatica Corporation의 상표 또는 등록 상표입니다. 다른 모든 회사 및 제품 이름은 상품명 또는 등록 상표일 수 있습니다.