

모든 데이터를 효과적으로 통합

InfoSphere Information Server

우수한 확장성으로 데이터 이동 및 통합

Service Oriented Finance의 새로운 데이터 통합 과제

네, 이제 움직이는 데이터와 움직이지 않는 데이터를 어떻게 분석하는지 알겠습니다.

그런데, 이러한 데이터를 어떻게 기존 데이터베이스와 통합합니까?



Service Oriented Finance CTO

네, 좋은 질문입니다.

이제 모든 빅데이터를 통합할 최적의 방법을 보여 드리겠습니다.



IBM

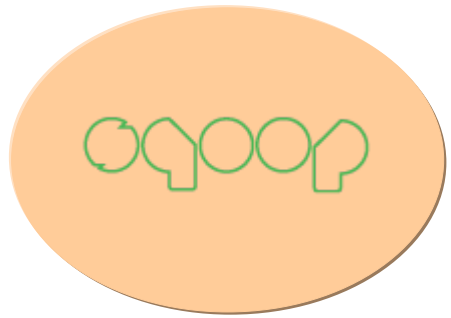
빅데이터는 통합이라는 큰 과제를 제시



- 새로운 데이터 저장
 - ▶ Hadoop HDFS
 - ▶ NoSQL
- 새로운 데이터 유형 및 형식
 - ▶ 비정형 데이터, 반정형 데이터
 - ▶ JSON, Avro 등
 - ▶ 비디오, 문서
- 단편적인 오픈 소스 솔루션
 - ▶ 표준을 충족하지 못함
- 대량의 데이터
 - ▶ 엄청난 양의 데이터를 이동하고, 지우고, 변환해야 함
 - ▶ 빅데이터에는 뛰어난 확장성이 필요함

통합 방법 1 - 오픈 소스 사용

- Oozie - 데이터 플로우 오케스트레이션
- Flume - 외부 데이터 소스 가져오기
- Sqoop - HDFS와 관계형 데이터베이스 통합



완전하지 않은 세 가지 툴을 사용할 때의 우려사항

- 생산성 및 비용
 - ▶ 팀이 세 가지의 툴을 배우는 것이 얼마나 생산적일까요?
- 리스크
 - ▶ 이 툴들은 완전하지 않습니다. 완전하지 않은 툴이 나중에 큰 문제를 일으킬까요?
 - ▶ 이 툴들은 빅데이터 요구사항에 맞게 확장됩니까?
 - ▶ 이 툴들이 모든 요구 조건을 만족하지 못하면 다른 툴들을 추가해야 합니까?

05 - 모든 데이터를 효과적으로 통합

엔터프라이즈급 통합을 제공하는 Information Server

- 데이터 소스에 상관 없이 동일한 그래픽 설계 도구 제공
- 다년간 광범위하게 사용되어 검증된 도구
- 검증된 선형적 확장성
- 생산성 향상, 비용 절감, 리스크 감소
 - ▶ 공유 메타데이터가 협업 촉진
 - ▶ 재사용 최대화 - 한 번 빌드한 후 공유
 - ▶ 수백 가지의 사전 빌드된 구성요소
- 모든 데이터 소스에 연결
 - ▶ 데이터베이스, 애플리케이션, 파일, 메시지 큐, Hadoop, NoSQL



IBM 빅데이터 플랫폼간의 연계를 위한 Information Server



고객에 대한 전방위적 시각을 구축하려 하는 Service Oriented Finance

관계형 데이터베이스로 분석한
부정적 감정의 **Twitter** 데이터를
통합하려면 어떻게 해야 합니까?



Service Oriented Finance CTO

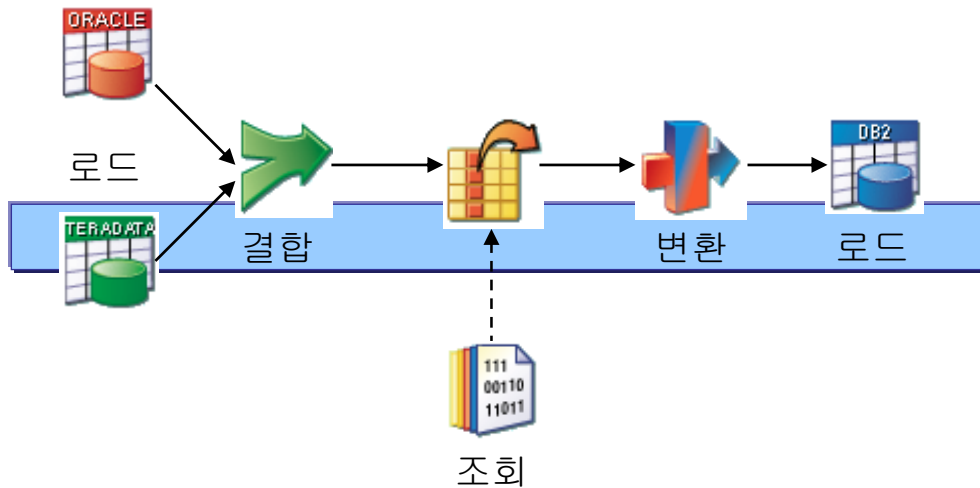
Information Server 제품군의 주요
도구 중 하나인 **DataStage**를 사용하면
됩니다.



IBM

DataStage를 통해 모든 데이터 소스를 손쉽게 통합

끌기, 놓기, 구성

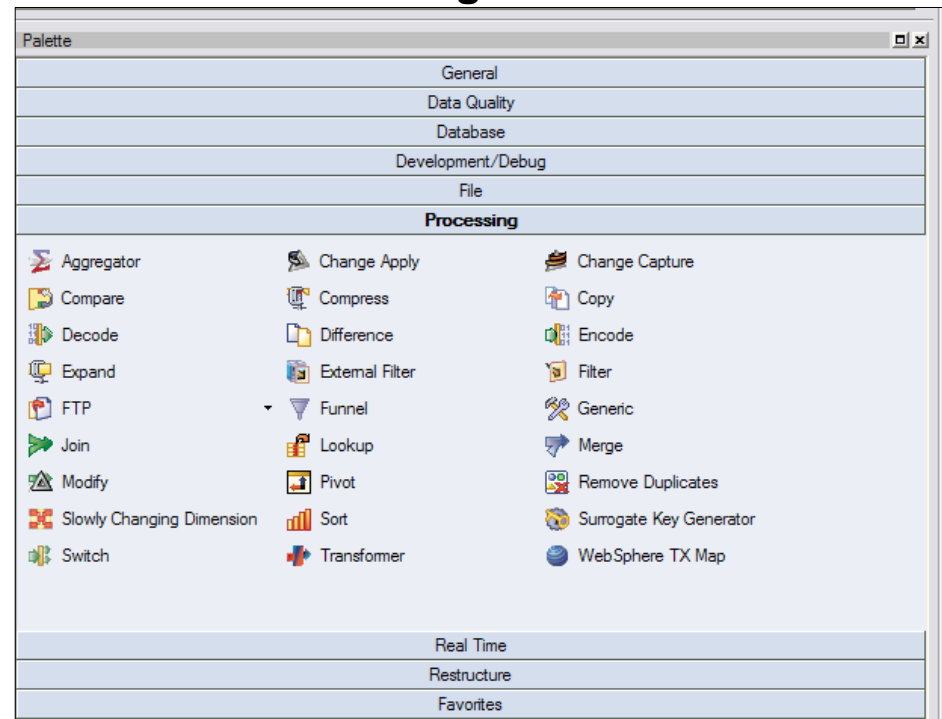


- 생산성
 - ▶ 데이터 플로우에 대한 그래픽 설계
 - 하위 레벨 구현이 아닌 플로우에 포커스를 맞춤
 - ▶ 사전 빌드된 여러 구성요소
- 확장성
 - ▶ 병렬 처리 기능 제공
- 리스크 감소, 비용 절감
 - ▶ 모듈형 방식으로 구성요소 재사용 최대화
 - ▶ 공유 메타데이터로 협업 개선
- 하나의 틀에서 모든 데이터 소스 통합

사전 빌드된 여러 가지 변환 구성요소가 빠른 생산성 향상을 촉진

- 사전 빌드된 50개 이상의 변환 구성요소로 빠른 작업 생성과 생산성 향상이 지원됨
- 여러 내장 함수를 통해 고급 변환 로직을 손쉽게 작성
- 사용하기 쉬운 편집기를 사용하여 빠르게 구성

DataStage 팔레트



공통 메타데이터 계층은 최적의 생산성과 협업의 핵심



- 더 효과적이고 빠른 커뮤니케이션
- 메타데이터 중심의 강력한 설계 도구
 - ▶ 고급 검색
 - ▶ 영향 분석
 - 변경할 경우 어떤 일이 발생합니까?
 - ▶ 데이터 계보 보고서
 - 데이터가 어디에서 오고, 어디로 갑니까?
- 표준 컴플라이언스 향상
- 정보 신뢰도 향상

모든 시스템과 애플리케이션을 연결하는 Information Server

관계형 데이터베이스

DB2(IBM System z®, System i®, System p® 또는 System x® 기반)

Oracle

IBM Informix® Dynamic Server 및 Informix Extended Parallel Server

Ingres

Netezza

Progress

RDB

RedBrick

SQL/DS

SQL Server

Sybase(ASE & IQ)

Teradata

Universe

IBM UniData®

NonStop SQL

InfoSphere Federation Server

InfoSphere Classic Federation

기타



일반 액세스

순차 파일

복합 플랫폼 파일

파일 세트

데이터 세트

네임드 파이프

iWay

FTP

SFTP

압축/인코딩된 데이터

외부 명령 호출

병렬/랩핑된 비IBM 앱

EMC InfoMover

웹 로그

이메일



엔터프라이즈

애플리케이션

JDE/PeopleSoft OneWorld

Oracle Applications

PeopleSoft

SAS

SAP BW

SAP R/3

Siebel

Ariba

Manugistics

I2

기타



표준 및 실시간

WebSphere MQ

JMS(Java Message Service)

Java

XML 및 XSL-T

EBXML

Web Services(SOAF)

EJB(Enterprise Java Beans)

EDI

FIX

SWIFT

HIPAA



변경 데이터 캡처

DB2(System z, System i, System p 또는 System x 기반)

Oracle

SQL Server

Sybase

Informix

IMS

VSAM

ADABAS

IDMS

Datacom

레거시

Allbase/SQL

C-ISAM

D-ISAM

Datacom/DB

DS Mumps

Enscribe

Essbase

FOCUS

IDMS/SQL

ImageSQL

Infoman

KSAM

M204

MS Analysis

Nomad

Nucleus

RMS S2000

Supra

TOTAL

TurboImage

Unify

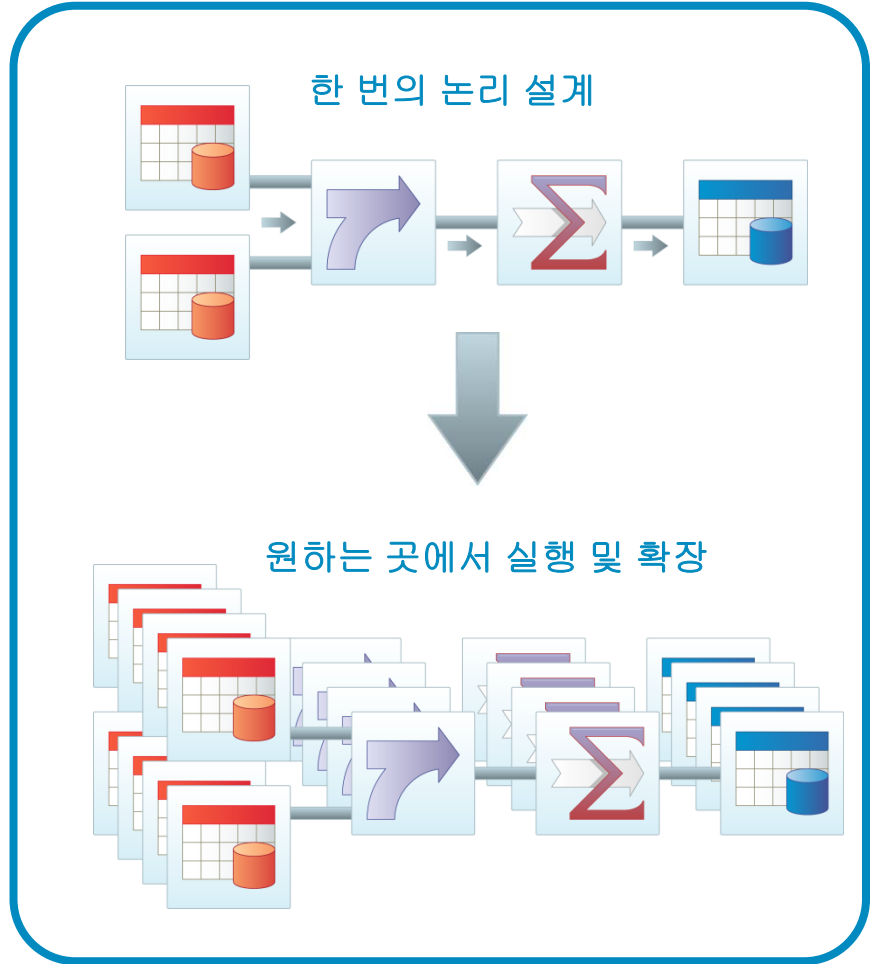
기타

새로운 빅데이터 소스도 모두 연결



한 번 설계한 후 원하는 곳에서 실행 및 확장

- 작업 설계는 배치/확장 방식과 상관 없이 동일
- 단순한 텍스트 파일로 확장 구성
- Hadoop과 마찬가지로, 데이터 및 코드가 병렬 처리를 위해 노드 클러스터로 파티셔닝됨
- 하드웨어 리소스가 추가되면 즉각적으로 성능 향상
- 원하는 규모의 정보를 추출, 변환, 로드



지금까지 볼 수 없었던 뛰어난 확장성!

엔터프라이즈급 통합 허브인 Information Server

- 업계 최고의 확장성을 가지는 통합 런타임
- 가장 광범위한 데이터 소스에 연결
 - ▶ 기존: 데이터베이스, 애플리케이션, 메시지 큐
 - ▶ 신규: Hadoop, 스트림, MongoDB, JSON 등
- 생산성 향상
- 개발 비용 절감
- 개발 리스크 감소

통합 도구에 대한 Gartner Magic Quadrant



Gartner 보고서: <http://www.gartner.com/technology/reprints.do?id=1-1HBEFSF&ct=130717&st=sb>