

# PureData Systems로 빅 데이터 실현 가속화

2013. 2. 20

김규병 실장 Software Group, IBM Korea



많은 데이터가 생성되고 있습니다.

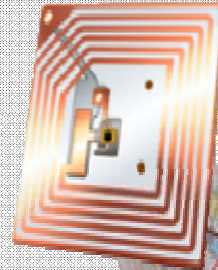
12+ TBs of tweet data every day



25+ TBs of log data every day



30 billion RFID tags today (1.3B in 2005)



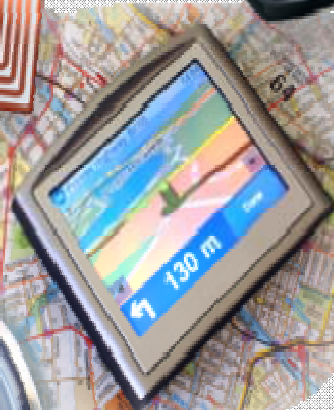
76 million smart meters in 2009... 200M by 2014



4.6 billion camera phones world wide



100s of millions of GPS enabled devices sold annually

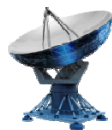


2+ billion people on the Web by end 2011



? TBs of data every day

# 다양한 형태의 대용량 데이터를 신속히 분석하는 것이 관건입니다.



- 방대하고 다양한 데이터 분석
- 스트리밍 데이터에 대한 통찰력
- 대량의 구조화된 데이터 분석

IBM 빅 데이터 플랫폼  
Variety, Velocity, Volume



다중 채널 고객 감성 및 경험 분석  
고객 성향, 행동 분석 활용



신생아실에서 예측 분석을 통해 위험  
상황을 신속하게 감지  
의료기록 분석을 통한 조언



기후/지역 데이터 분석으로 풍력 발전기  
및 풍력 발전소 부지 플래닝과 날씨 패턴  
분석



실시간 트랜잭션 데이터를 기반으로  
리스크 측정 및 이에 대한 의사 결정



산재되어 있는 비디오, 오디오와  
데이터 피드를 통해 범인과 위협 감지

데이터 소스에 대한 통합적 관리와 연계가 필요합니다.

## Transactional and Application Data



- Volume
- Structured
- Throughput

## Machine Data



- Velocity
- Semi-structured
- Ingestion

## Social Data



- Variety
- Highly unstructured
- Veracity

## Enterprise Content



- Variety
- Highly unstructured
- Volume

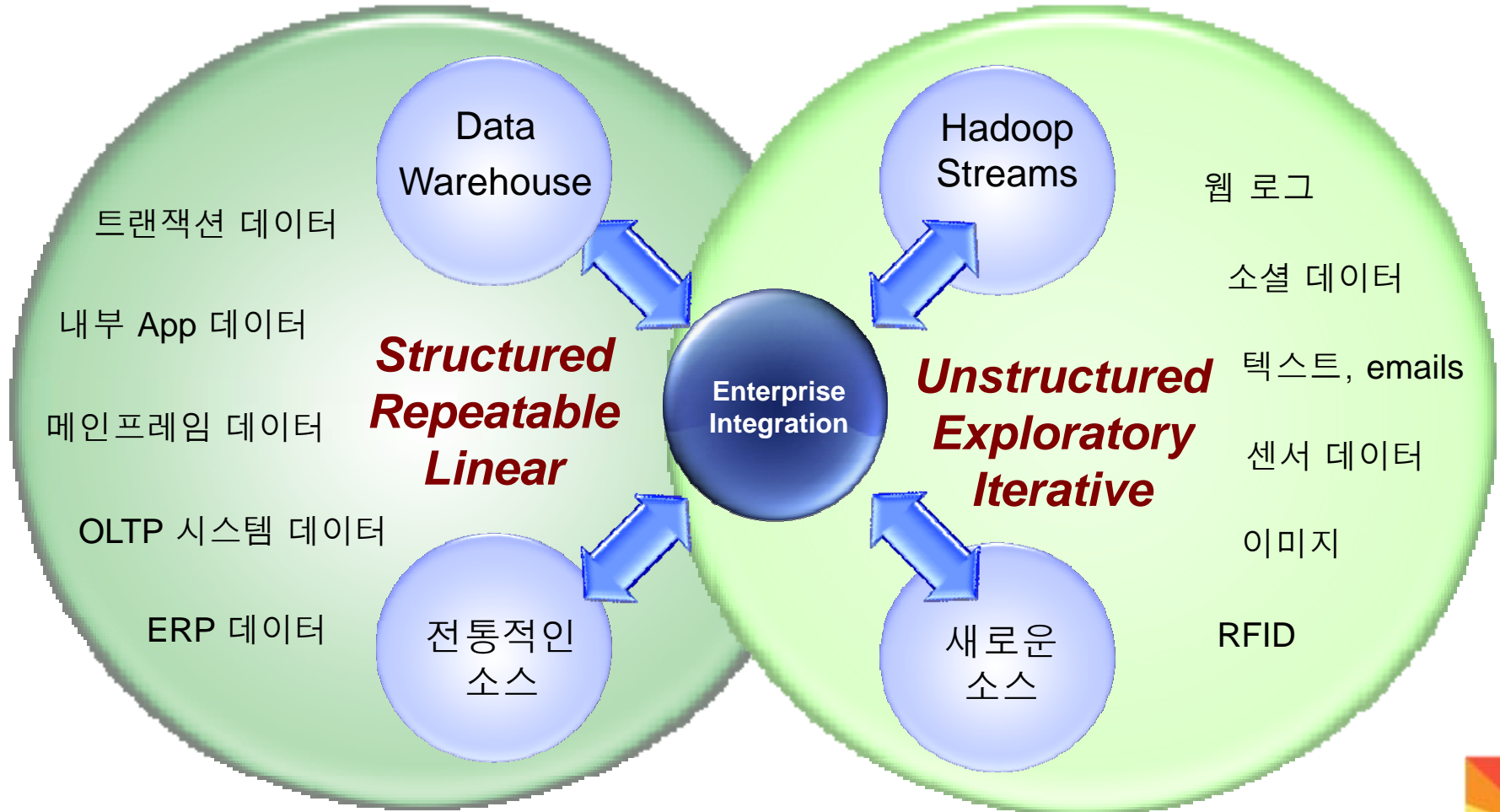
분석 플랫폼간의 통합은 빅 데이터의 가치를 극대화합니다.

### Traditional Approach

구조적, 분석적, 논리적

### New Approach

창조적, holistic thought, 직관적

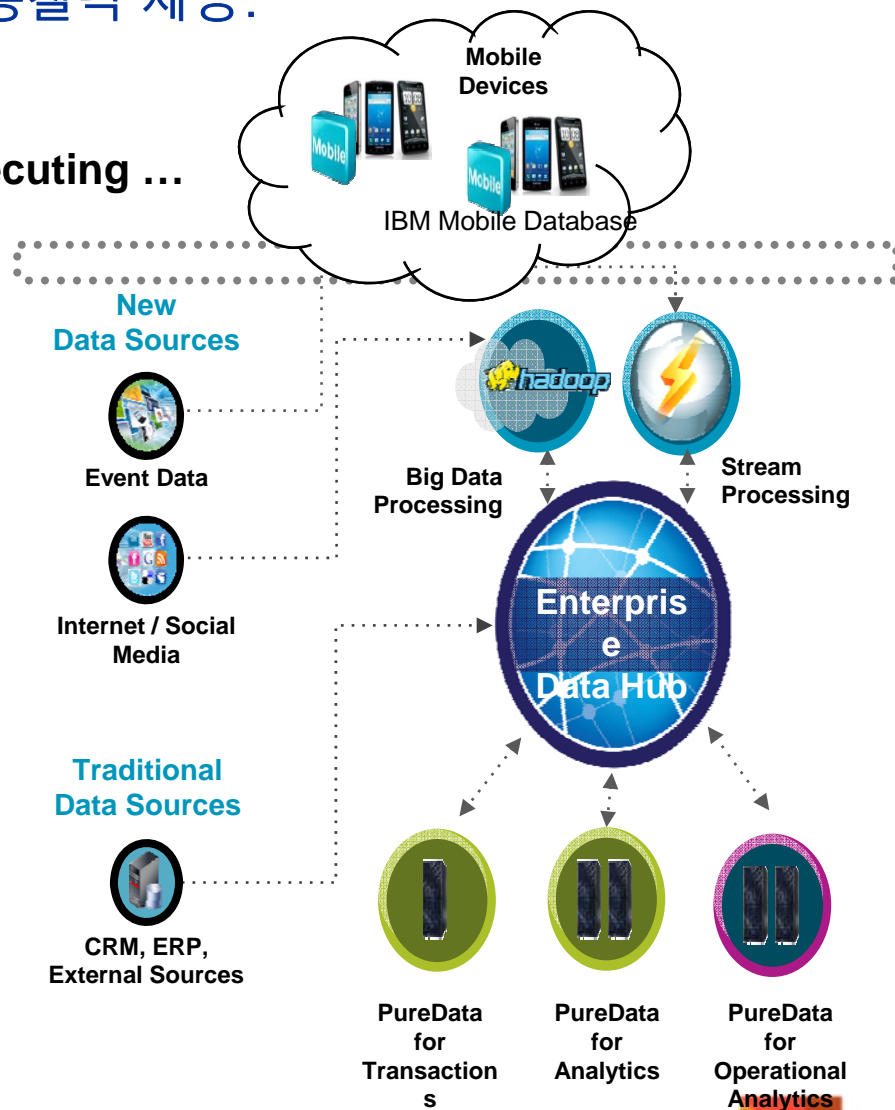
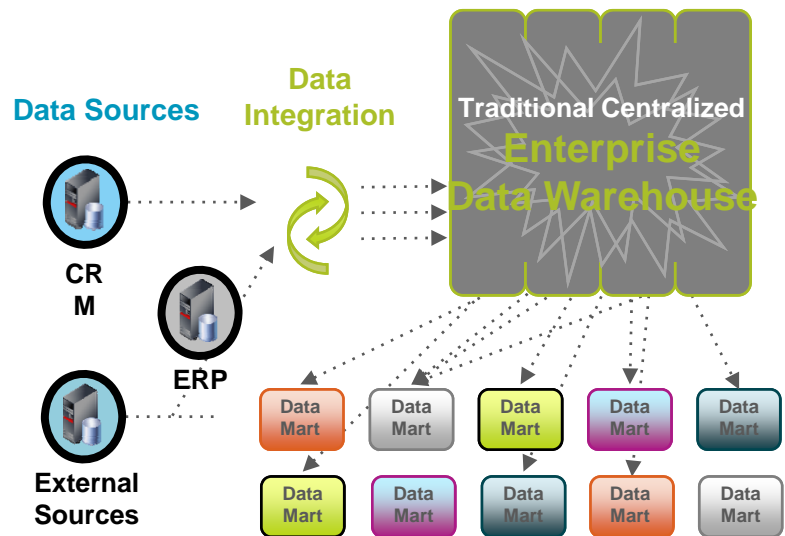




분석 플랫폼간의 통합과 연계를 통한 비즈니스 통찰력 제공.

### Smart Consolidation, Fit for Purpose solution

Accelerate enterprise Information management by executing ...  
the **Right Workload** against the **Right Data**,  
in the **Right Place**, at the **Right Cost** and **Performance**



## 우리가 집중해야 할 새로운 가치는?

- **통합 (Consolidating)** – 애플리케이션과 인프라
- **최적화 (Optimizing)** – 애플리케이션 성능
- **혁신 (Innovating)** – 모바일, 소셜, 분석 기술 활용
- **가속화 (Accelerating)** – 클라우드 구축



# 우리가 직면한 문제들...

## 범용 구성 요소의 튜닝에 상당한 시간과 노력 소요



## 전체 IT 프로젝트 라이프사이클 간소화





# 시스템 라이프사이클의 각 단계를 간소화하는 접근



# IBM의 새로운 전문가 통합 시스템 제품군

## PureSystems

클라우드를 위한 통합형 전문성 시스템

### 집약된 전문성

전문가가 하는 일을 수집하고 자동화  
- 인프라 패턴부터 애플리케이션 까지



### 설계에 의한 통합

하드웨어와 소프트웨어의 심층 통합과 튜닝 -  
즉시 가동할 수 있도록 워크로드에 최적화된 시스템

### 간소화된 환경

IT 라이프사이클의 모든 요소를 더 손쉽게 - 전체 시스템의 통합 관리,  
최적화된 솔루션으로 이루어진 광범위한 개방형 에코시스템

# IT 비용을 절감하고 신속한 가치 창출을 도와 드리는 PureData Systems

## PureData System for Transactions



확장형 트랜잭션 데이터베이스  
트랜잭션 처리

## PureData System for Analytics



분석 데이터 웨어하우스  
보고와 분석

## PureData System for Operational Analytics



운영 데이터 웨어하우스  
운영 분석

## 데이터 서비스를 위해 최적화된 IBM PureData System

### 데이터 서비스에 최적화:

- 트랜잭션
- 분석

### 전문가 통합:

- 데이터 플랫폼
- 인프라
- 통합 플랫폼 관리
- 기본적으로 구현된 전문성

# PureData



데이터 플랫폼

데이터 서비스 제공  
Delivering Data Services

*최고의 성능을 위한 설계*

*워크로드 최적화된 성능*

*몇 시간 내 서비스 가능*

*통합 관리*

*자동화된 유지 보수*

*단일 창구에서 지원*



# 데이터 워크로드의 특성에 맞춰 최적화

## Meeting Big Data Challenges – Fast and Easy!



### **PureData** System for Transactions

*전자 상거래와 같은 애플리케이션...*

매일 수백만 건의 온라인 트랜잭션 처리, 데이터 읽기와 쓰기 작업 혼합

### **PureData** System for Analytics *분석* Netezza 기술 기반

*고객 분석과 같은 애플리케이션...*

최소한의 관리로 페타바이트급 데이터를 손쉽게 로드, 복잡한 분석 및 보고서 실행

### **PureData** System for Operational Analytics

*실시간 사기 탐지와 같은 애플리케이션...*

실시간 의사 결정을 위해 운영 읽기 및 쓰기를 처리하는 동안 복잡한 분석 실행



# IBM PureData System for Transactions

## : 트랜잭션 데이터 워크로드를 위한 최적화

# PureData

## System for Transactions

트랜잭션을 위한 데이터 서비스 제공



### 속도

- 업계 최고의 DB2 성능
- 수초 내에 데이터베이스 노드 복구(\*1)

### 간소화

- 몇 시간이 아닌 몇 분 이내에 데이터베이스 구축(\*1)
- 여러 데이터베이스 소프트웨어 버전 실행
- 하나의 시스템에서 100개 이상의 데이터베이스 처리(\*2)
- 예정된 다운타임 없이 펌웨어/OS 업그레이드(\*1)

### 확장성

- 최대 30배 확장(\*3)
- 예정된 시스템 다운타임 없이 소규모에서 중간, 대규모 구성까지 확장하도록 설계

### 지능성

- 최소한의 변경으로 Oracle Database 애플리케이션 지원. 변경 없이 DB2 애플리케이션 지원
- 적응형 압축을 통해 스토리지 공간을 1/10로 감축한 고객 사례(\*4)

1. 예상되는 일반 워크로드와 정상 운영 조건에서 IBM 내부 테스트 및 시스템 설계 기준. 개별적인 결과는 다를 수 있습니다.  
 2. 하나의 대규모 구성 기준  
 3. 하나의 데이터베이스에 필요한 최소/최대 프로세서 및 메모리 리소스 기준  
 4. DB2 10 Early Access Program 고객 테스트



# IBM PureData System for Analytics : 분석 데이터 워크로드를 위한 최적화

Powered by  
Netezza Technology

## PureData System for Analytics

분석을 위한 데이터 서비스 제공



### 속도

- 기존 맞춤형 시스템 대비 10-100배 빠른 속도<sup>(\*)</sup>
- 전술적 쿼리의 경우 20배 향상된 동시성 및 처리량<sup>(\*)</sup>
- 특허 받은 MPP(Massively Parallel Processing) 하드웨어 가속화 기술

### 간소화

- 몇 시간 내에 데이터 로드 준비
- 데이터베이스 색인 불필요
- 튜닝 불필요
- 스토리지 관리 불필요

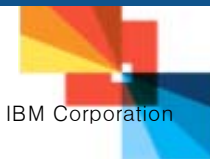
### 확장성

- 페타급 데이터 사용자 용량

### 지능성

- 몇 시간이 아닌 몇 분 이내에 복잡한 분석을 실행할 수 있도록 설계
- 가장 강력한 데이터베이스 내(in-database) 분석

1 IBM 고객이 보고한 결과. "기존의 맞춤형 시스템"은 전문적인 사전 구축, 테스트, 최적화를 거치지 않은 시스템입니다. 개별적인 결과는 다를 수 있습니다.  
2 IBM 내부 성능 벤치마크 기준



# Pre-Built In-Database 분석 기능



# IBM PureData System for Operational Analytics

: 운영 분석 데이터 워크로드를 위한 최적화

## PureData

System for Operational Analytics

운영 분석을 위한 데이터 서비스 제공



### 속도

- 1000+개 이상의 동시 운영 쿼리 처리<sup>(\*)</sup>
- 운영 데이터 연속 수집
- MPP(Massively Parallel Processing) 분석

### 간소화

- 자동 정책 기반 데이터 배치 및 워크로드 관리
- 통합 관리 및 지원

### 확장성

- 여러 크기로 사용 가능, 최대 1PB의 데이터 용량<sup>(\*)</sup>

### 지능성

- 대표적인 애플리케이션을 위한 데이터베이스 내 분석
- 변경 없이 DB2 지원, 최소한의 변경으로 Oracle Database 애플리케이션 지원
- 적응형 압축을 통해 스토리지 공간을 1/10로 감축한 고객 사례<sup>(\*)</sup>

1. 예상되는 일반 워크로드와 정상적인 운영 조건에서 전 세대 시스템에 대한 IBM 내부 테스트 및 시스템 설계 기준. 개별적인 결과는 다를 수 있습니다.  
 2. 하나의 대형 구성, 5개의 풀 렉 데이터 확장 애드온 기준 총 원시 데이터 용량  
 3. DB2 10 Early Access Program 고객 테스트

## 고객이 인식하는 PureData System 가치

### 트랜잭션

지급 처리 분야의 세계적  
선두 주자

동적 트랜잭션 라우팅

DB2 클러스터는 경쟁사 대비 우수 제공

- 더 저렴한 비용으로 지불 라우팅을 수행하여 경쟁력 향상
- 성수기 트랜잭션 볼륨을 수월하게 처리
- 필요한 규모에서 최고의 가용성 실현

### 분석

 NYSE Euronext

 Constant Contact

“NYSE는 ... Oracle 관계형 데이터베이스를 Netezza의 데이터 웨어하우징 어플라이언스로 대체한 결과, 650TB의 데이터를 신속하게 검색할 수 있게 되었습니다.”  
ComputerWeekly.com

“년간 350억개의 이메일을 분석해서 최고 25%의 캠페인 성과를 개선하고  
분석속도를 40배 이상 향상시켰습니다.”

### 운영 분석

세계적인 신용카드사

고객 관리 및 실시간 사기 탐지

운영 데이터 웨어하우스에서 매일 8천만 건의 트랜잭션 처리

- 사기 평가, 고객 관리, SAS 분석 지원
- 평균 동시 사용자 수 2500명 이상
- 일일 5TB 이상 로드 - 처리량 최대 시간 포함

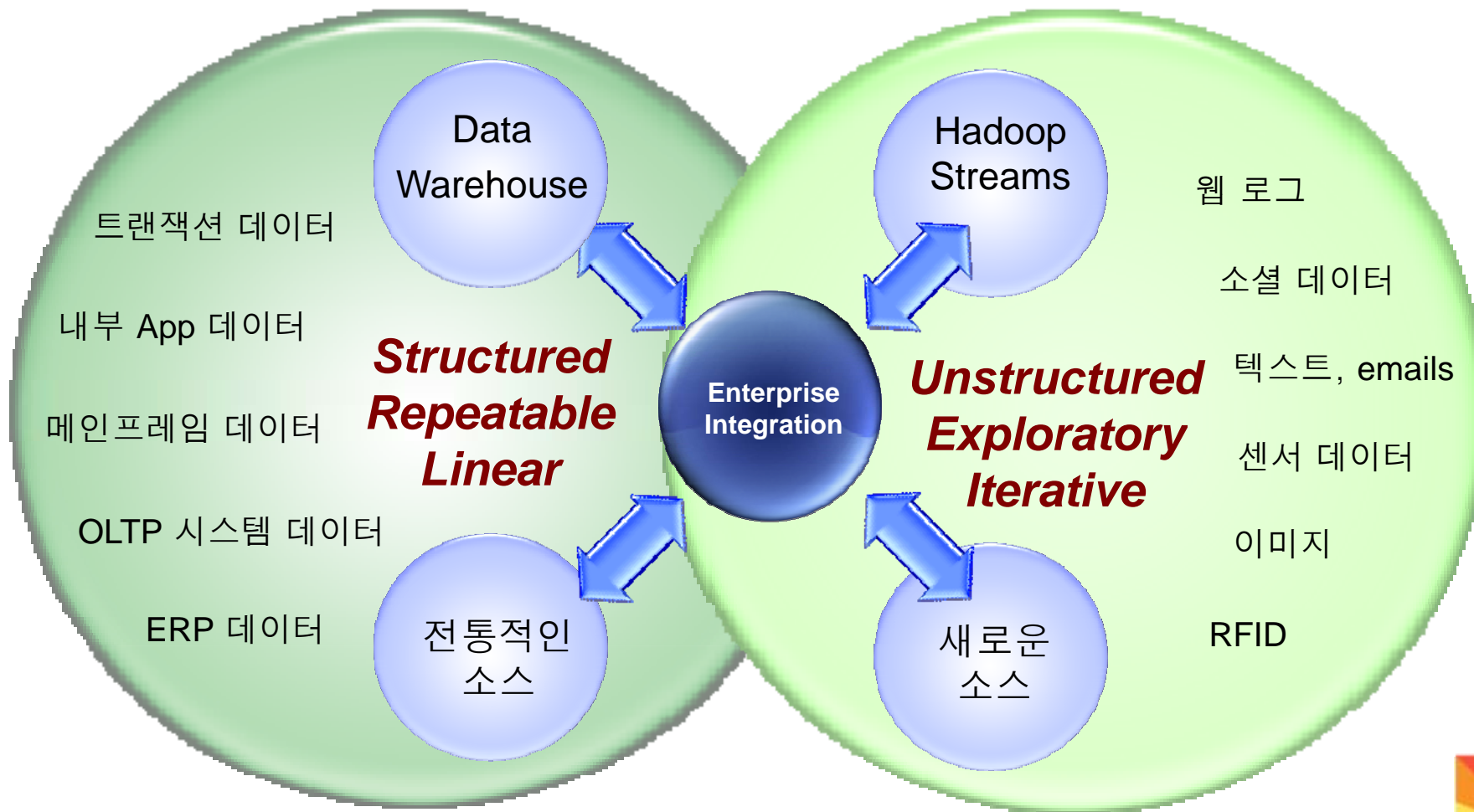
분석 플랫폼간의 통합은 빅 데이터의 가치를 극대화합니다.

### Traditional Approach

구조적, 분석적, 논리적

### New Approach

창조적, holistic thought, 직관적



**THINK**

**BIG**