# 똑똑한 빅 데이터 분석을 위한 기반, 정보 통합

나상근 팀장, IBM KSSL Information Management Lab

© 2013 IBM Corporation

# 빅 데이터 시대는 새로운 도전과제와 기회...

## 도전과제

**Information from everywhere**

350 million Tweets created daily

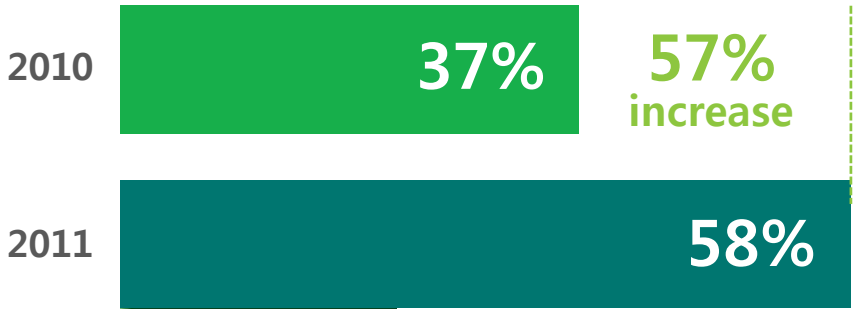
**Radical flexibility**

5 million trade events per second

**Extreme scalability**

## 기회

분석이 경쟁우위 확보를 위해 중요한 요소가 될 것이다라는 응답자




분석을 통해 경쟁우위를 갖게 되는 조직은 동일 산업의 경쟁회사에 비해

**2.2x**

대체로 더 나은 성과를 올리고 있다.



© 2012 IBM Corporation

# 이러한 도전과제들 때문에...

이런 속도로 가면

**IT complexity is doubling**

매 2년 마다.[1]

More than

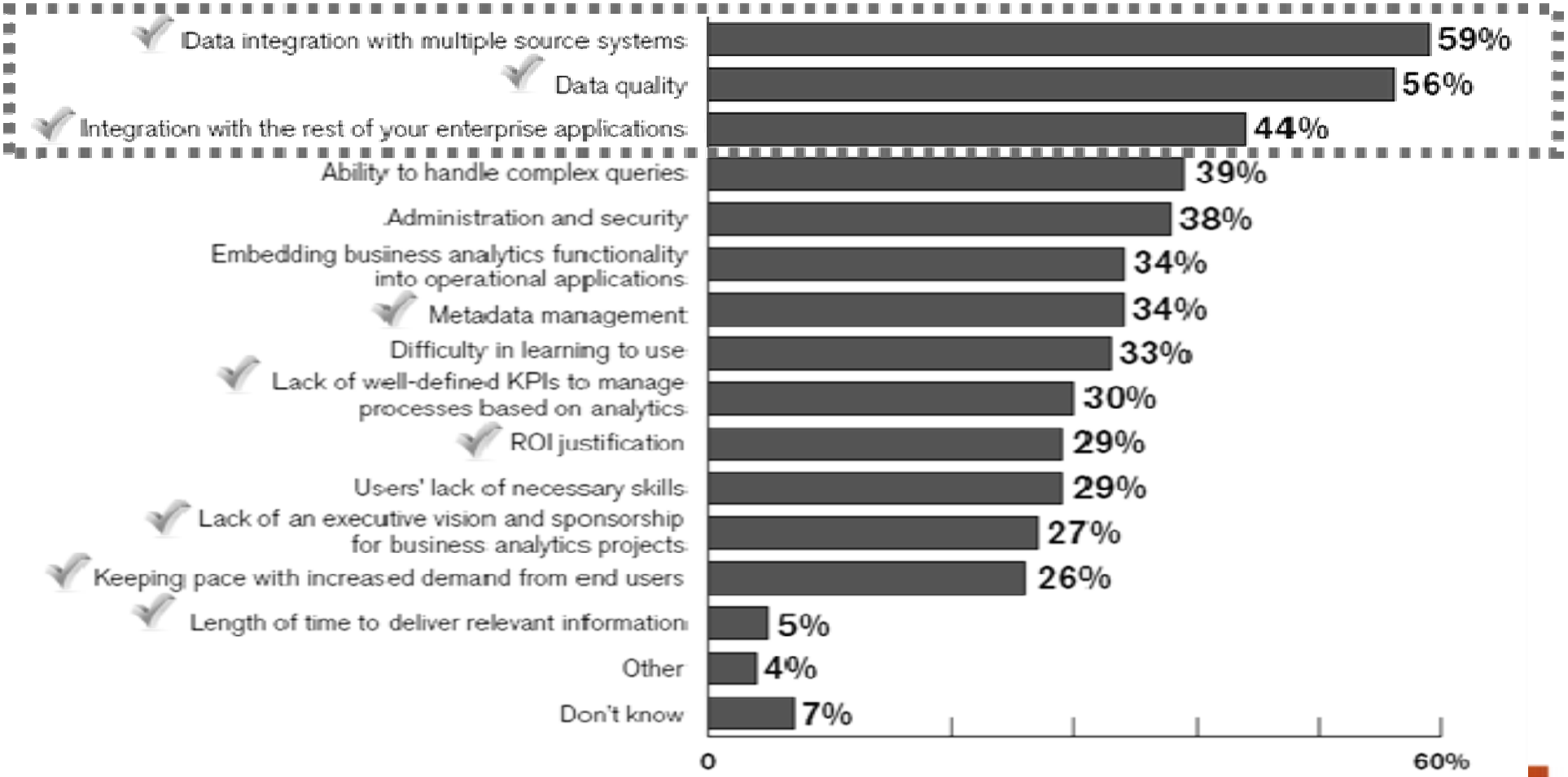
[1] Enterprise Systems Journal, “5 Best Practices for Reducing IT Complexity”)
[2] IBM CIO Study 2010
[3] IBM Market Intelligence Time-To-Value Study, National Analysts, November 2011



© 2012 IBM Corporation

# 분석을 통한 통찰력을 높이기 위해서는...

## 효과적인 비즈니스 분석 프로젝트에 있어서의 주요한 도전과제:

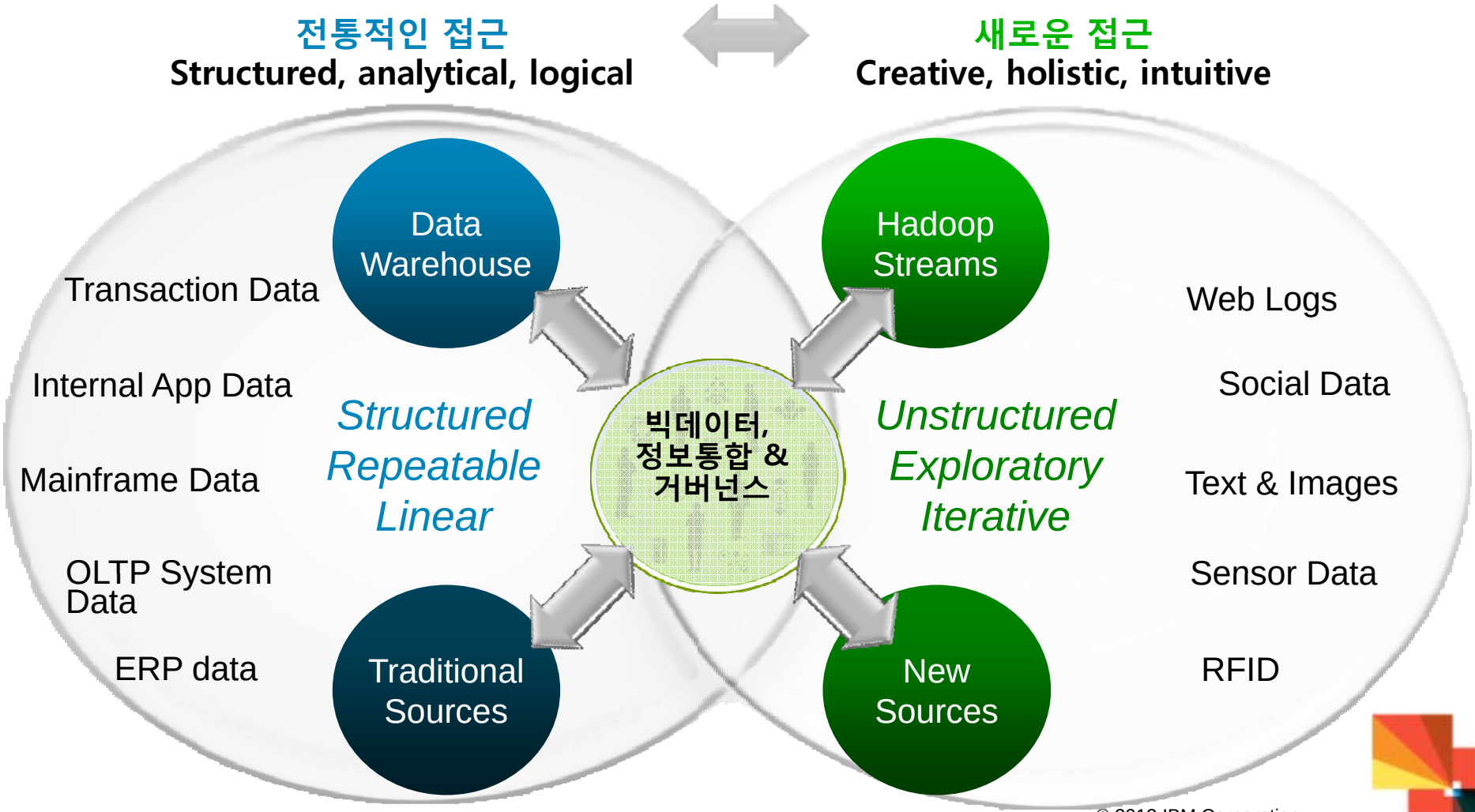



*"Defining Business Analytics and Its Impact on Organizational Decision-Making," ComputerWorld*



© 2012 IBM Corporation

# 빅데이터 시대의 정보통합 요건은 더욱 복잡해지고, 다양한 인프라에 대한 지원 요건이 많아지고 있습니다

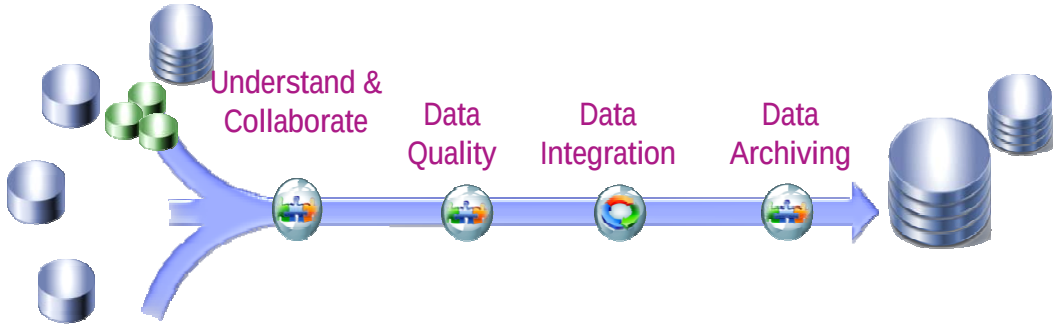




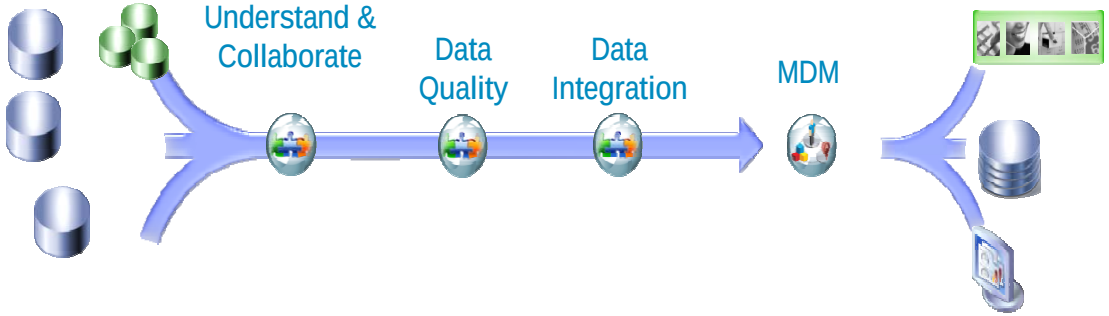

© 2012 IBM Corporation

# 정보통합은 다양한 정보중심의 프로젝트에 매우 중요한 역할을 합니다

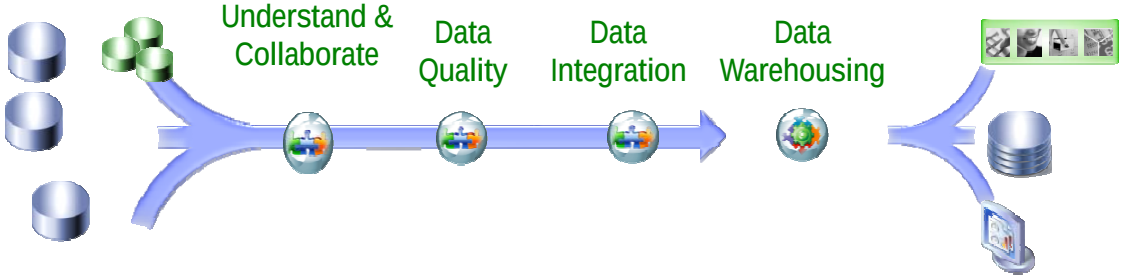
애플리케이션 통합 및 마이그레이션




마스터데이터 (Single View)




데이터 웨어하우징 & 비즈니스 분석






© 2012 IBM Corporation

# 분석은 신뢰할 만한 정보를 기초 해야 합니다

빅데이터 분석에 있어서 도출된 결과를 신뢰할 수 없다면...

**빅데이터 분석**

**통찰력을 얻기 위해 빅 데이터(variety, volume, and velocity)를 처리할 수 있는 역량**

**정보 통합과 거버넌스**

**신뢰할 수 있는 정보를 확보할 수 있는 역량**



© 2012 IBM Corporation

# 빅데이터에서의 큰 오해

## 많은 조직들이 거버넌스가 중요하다는 것을 간과함

- *"일반적으로 빅데이터 프로젝트는 연구를 목적으로 작은 데이터를 갖고 신뢰할 만한 환경에서 시작을 합니다. 그런데 이것이 증명되어 좀더 진행을 하면서 부터, 우리는 이것도 다른 IT 프로젝트와 다르지 않다는 것을 알게 됩니다 – 거버넌스가 필요하게 된다."*

  **•빅데이터도 전통적인 데이터와 마찬가지로 다시 거버넌스를 세우게 되면 비용이 많이 들고 복잡해진다는 것을 생각하지 못한다.**

- *"우리는 빅데이터 프로젝트에 많은 기능들을 구현하게 됩니다. – 분석, 모델링 그리고 등등 – 우리는 필요하다면 우리가 쉽게 거버넌스를 구축할 수 있다고 생각합니다"*

  **•거버넌스는 간단한 이슈가 아닙니다. 단번에 해결하려 한다면 매우 많은 비용이 들게 됩니다.**



© 2012 IBM Corporation

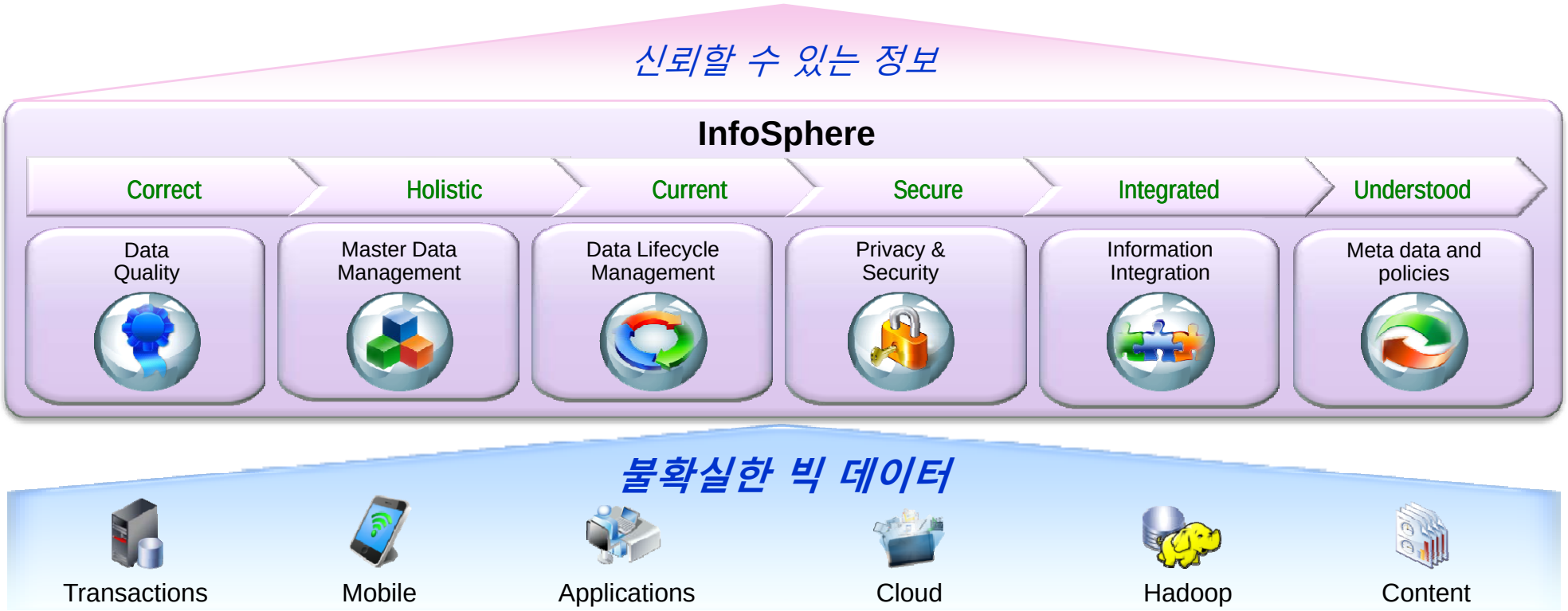
# 어떻게 빅데이터를 신뢰할 수 있도록 할 수 있을까요?

- 빅데이터의 성공은 **신뢰할 수 있는 정보**에 있습니다.
- 정보통합과 거버넌스의 목적은 **신뢰할 수 있는 정보**를 확보하고 제공하는 것입니다.
- **신뢰할 수 있는 정보의 특징**
  - **정확성(Correct)** – 표준화된, 인증된, 검증된, 매칭된
  - **전체성(Holistic)** – 완전한, 동일한 정보에 대해서 상충이 되지 않는
  - **현재성(Current)** – 가장 최근의 관련 데이터만 저장된, 과거 데이터는 아카이브되거나 삭제된
  - **안정성(Secure)** – 거버넌스 정책에 맞는 보호수준이 적용된
  - **통합됨(Integrated)** –소스 정보, 적용된 거버넌스 규칙 및 변환결과 등이 사용자들이 추적, 설명, 확인할 수 있는
  - **이해가 됨(Understood)** – 정보의 출처, 값과 품질 프로파일이 이해가 되는

- → **동일한 프레임워크가 빅데이터에도 적용되어야 합니다**

© 20

# 빅데이터를 신뢰할 수 있는 정보로 제공






© 2012 IBM Corporation

# IBM의 정보통합 및 거버넌스 솔루션- 빅데이터 전략의 차별성

## Only IBM Makes Big Data Actionable

1. **IBM's IIG Strategy with Big Data is Fundamentally Different**
   - *다른 통합업체들은 그들의 제품의 약점을 보완하기 위해서 빅데이터를 사용하려는 반면 (Scale); IBM의 정보통합 솔루션은 빅데이터의 활용에 있어서의 약점을 해결하는 것을 목표로 함 (Data Governance)*

2. **IBM InfoSphere was the First Big Data Integration Solution**
   - IBM DataStage는 2003년부터 MPP환경을 지원하도록 개발됨
   - 이때 출시된 Torrent MPP platform의 ETL툴인 DataStage는 하둡보다 10-15배 이상 빠르게 대용량 데이터를 처리하였음
   - 하둡을 지원하는 BigDataFileStage가 2011년에 GA됨

3. **IBM Data Governance is Uniquely Positioned for Big Data**
   - **Business Glossary & Metadata** → for audit, compliance and lineage
   - **Information Analysis & Data Quality** → for cleansing and validating
   - **Master Data Management** → for golden records and matching
   - **Search & Data Federation** → for navigation, indexing and queries
   - **Information Lifecycle Management** → for archiving and test data


© 2012 IBM Corporation

# IBM의 정보통합솔루션은 신뢰할 수 있는 빅데이터 활용을 위한 중요한 기능을 제공합니다

**비즈니스 중심의 거버넌스**
(Business-driven governance)

더 나은 비즈니스 의사결정을 위한 정보제공에 대한 확신 증가

**지속 가능한 데이터 품질**
(Sustainable data quality)

비즈니스 변화에 더 빨리 대응할 수 있도록 정보의 정확성 증가

**빠르고, 실시간, 페더레이션 데이터 통합**

(Agile, real-time, & federated data integration)

저장장소와 관계없이 기업의 데이터를 전체적으로 볼 수 있는 통합 정보의 제공



© 2012 IBM Corporation

# 비즈니스 중심의 거버넌스 기능

## 비즈니스 효과가 있는 시점에 더 큰 확신을 갖고 의사결정 하도록 지원

**비즈니스 중심의 규칙(Rule) 작성**
누가 고객인가? 우리는 어떻게 고객 데이터를 관리하는가?

**비즈니스 중심의 정책(Policies) 작성**
어떻게 고객데이터가 북미의 거버넌스 정책을 준수한다고 보장합니까? 또 유럽에서는?

**비즈니스 요구 사항에 부합하게 IT를 구현할 수 있도록 규정과 정책을 관리하고 규제함**

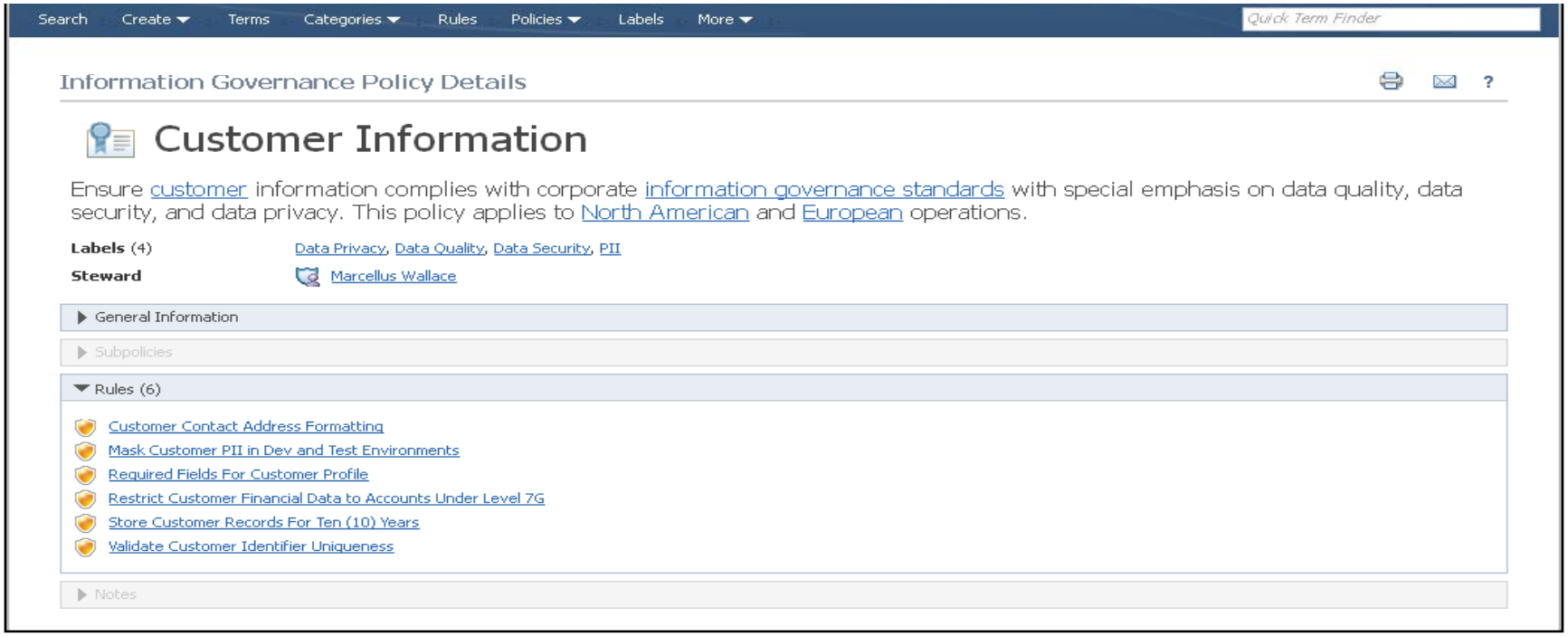


© 2012 IBM Corporation

# 비즈니스 중심의 거버넌스 기능

**비즈니스 가치:** 여러분의 비즈니스에 영향을 줄 수 있는 정보를 더 잘 컨트롤하고 자신감을 가질 수 있도록 함

- IBM은 거버넌스 정책 관리에 혁신을 추구하고 있습니다.
- **정책**은 정보 거버넌스의 영역을 정의합니다.
- **정보 거버넌스 규칙**은 정보에 대한 비즈니스적 기대를 설명합니다.
- 이러한 역량은 정보 거버넌스 전략에 대한 실질적인 출발점을 제공합니다.



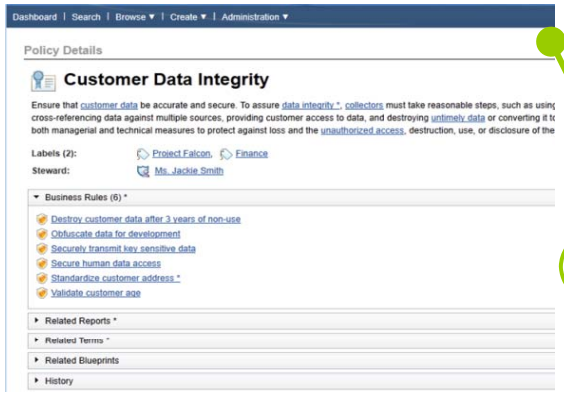



© 2012 IBM Corporation

# 지속 가능한 데이터 품질

## 정보의 정합성을 보장하여 전략적 비즈니스 변화에 더 빨리 대응

새로운 정보 거버너스 규칙(Rules) 및 정책(Policies)




define objectives

확장 플랫폼 지원

discover

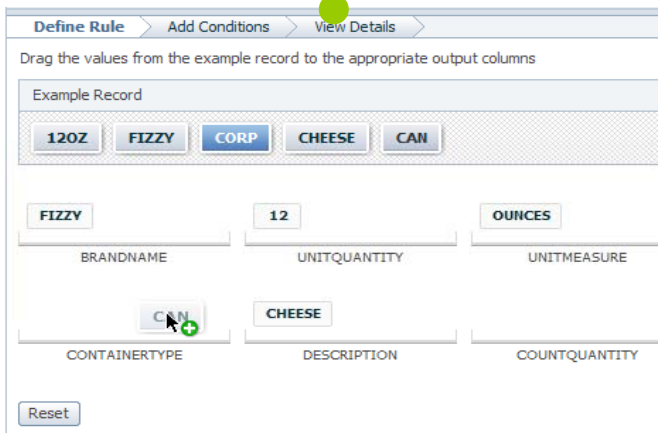
데이터 검증 규칙 영향 분석

validate

데이터 검증 규칙(Rules) 시퀀싱

cleanse & enrich

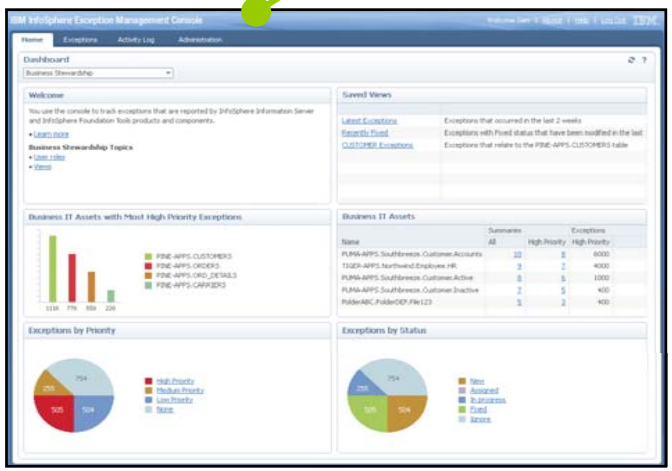
새로운 표준 규칙(Rules) 디자인




monitor / track

"master"

새로운 데이터 품질 콘솔





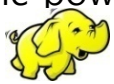

© 2012 IBM Corporation

# 신속한 통합 역량

## Greater agility is necessary to get the most from your environment

**InfoSphere Data Click**
Achieve greater agility by enabling novice users to quickly provision data in with just a few clicks (while still leveraging governance and optimizing the runtime)

**Enhanced support for big data**
Integrate extreme variety and volume of data, regardless of its velocity, and harness the power of Hadoop with InfoSphere Information Server for Data Integration



**Support for IBM PureApplication System (private cloud)**
New InfoSphere Information Server Enterprise Edition system pattern provides unparalleled Private Cloud offering facilitating leaps in time to value

**IBM Operational Decision Management integration**
Provides greater agility by supporting dynamic business changes

**Workload management**
Maximize performance in shared environments

**Balanced optimization for DB2 z/OS**
Exploits the power of Z to improve performance



© 2012 IBM Corporation

# 실시간 통합

## 초고속 양방향 데이터

Low-latency capture of real time information

## 다양한 데이터 분석

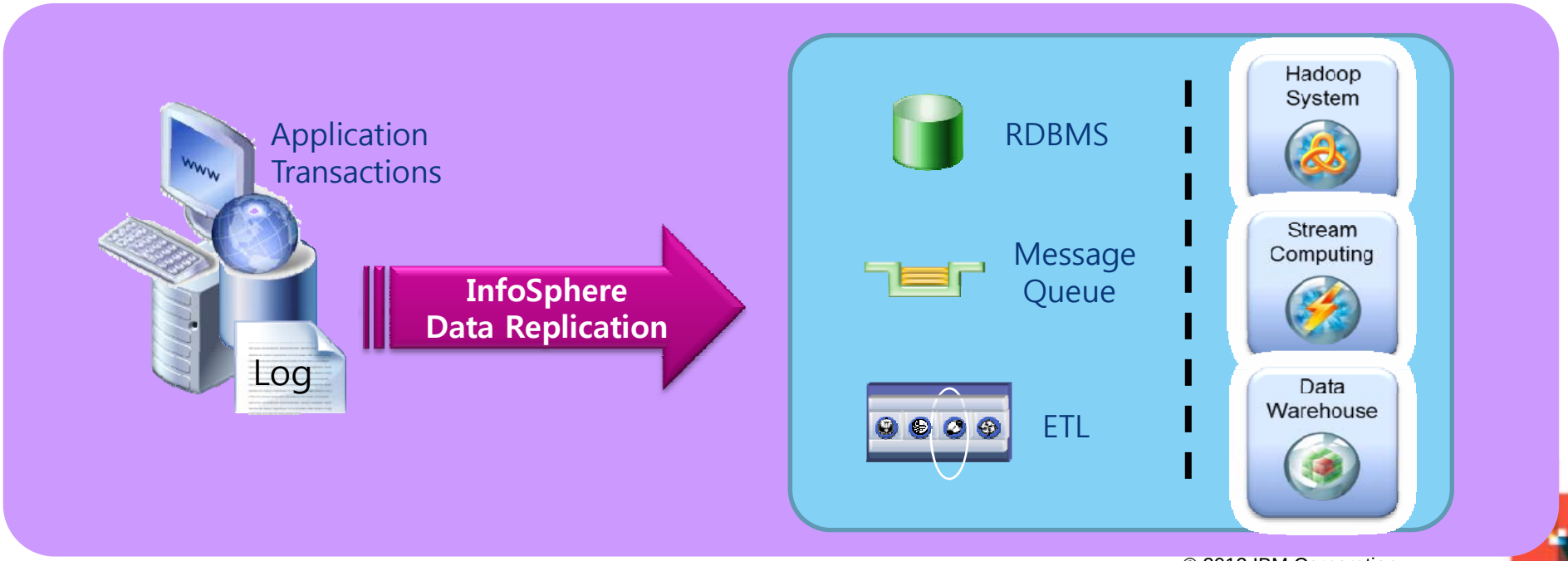
Analyze a variety of data in its native format – streaming audio, video, spatial, among others

## 손쉬운 데이터 캡쳐

Read data from transactional database logs, distribute data to any target – including Big Data Streams, ETL for Warehouses, or Big Insights

## 대용량의 변경되는 정보 분석

Terabytes per second, petabytes per day.

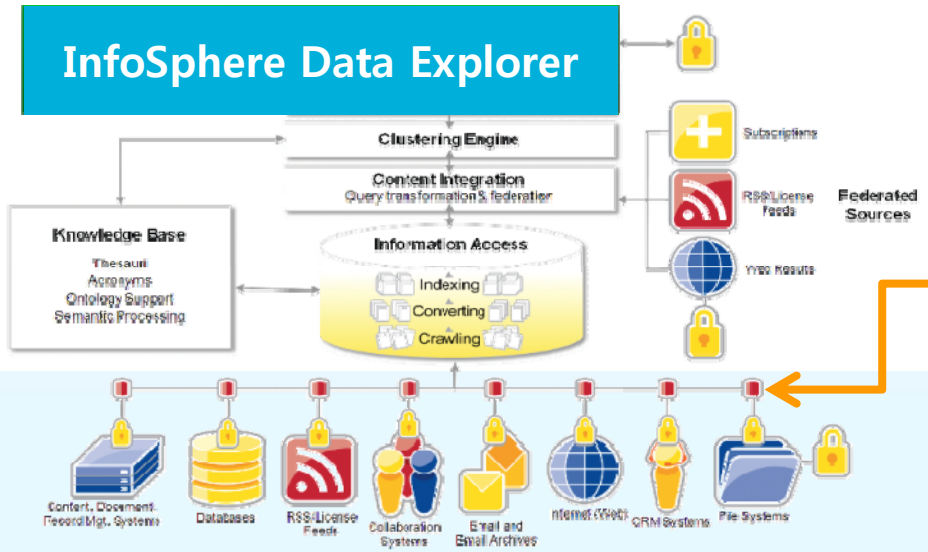




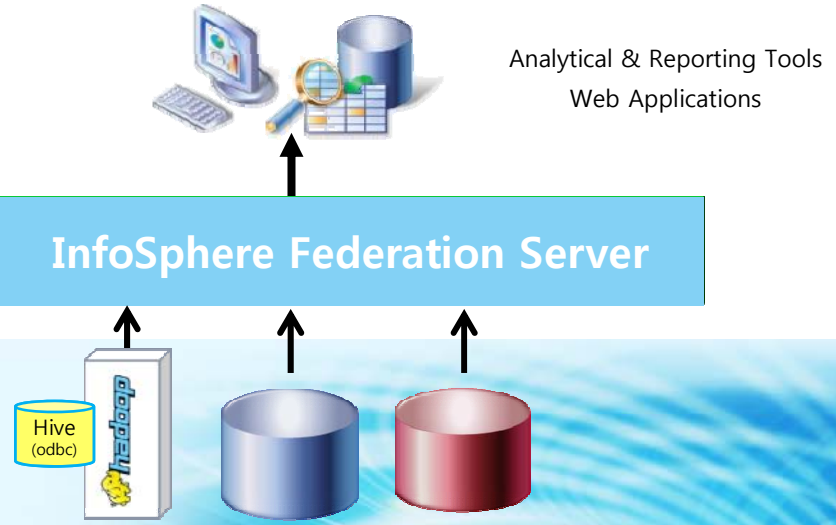

© 2012 IBM Corporation

# Federation을 활용한 빅 데이터 (기존 데이터 포함) 통합

## Search – Analyze – Visualize




## Model – Query – Deliver




*InfoSphere Database Federation can be another source of enterprise data for Big Data Federation solutions like IBM InfoSphere Data Explorer (formerly Vivisimo Velocity)*


© 2012 IBM Corporation

# All Big Things are Built on Trust

- 빅데이터 분석의 성공은 신뢰할 수 있는 정보에 달려 있습니다.
- 정보통합과 거버넌스의 목적은 신뢰할 수 있는 정보를 확보하고 제공하는 것입니다.
- 인포스피어 정보통합 솔루션의 신뢰할 수 있는 빅데이터 제공:
  - 비즈니스 중심의 거버넌스
  - 지속 가능한 데이터 품질
  - 빠르고, 실시간, 페더레이션 데이터 통합

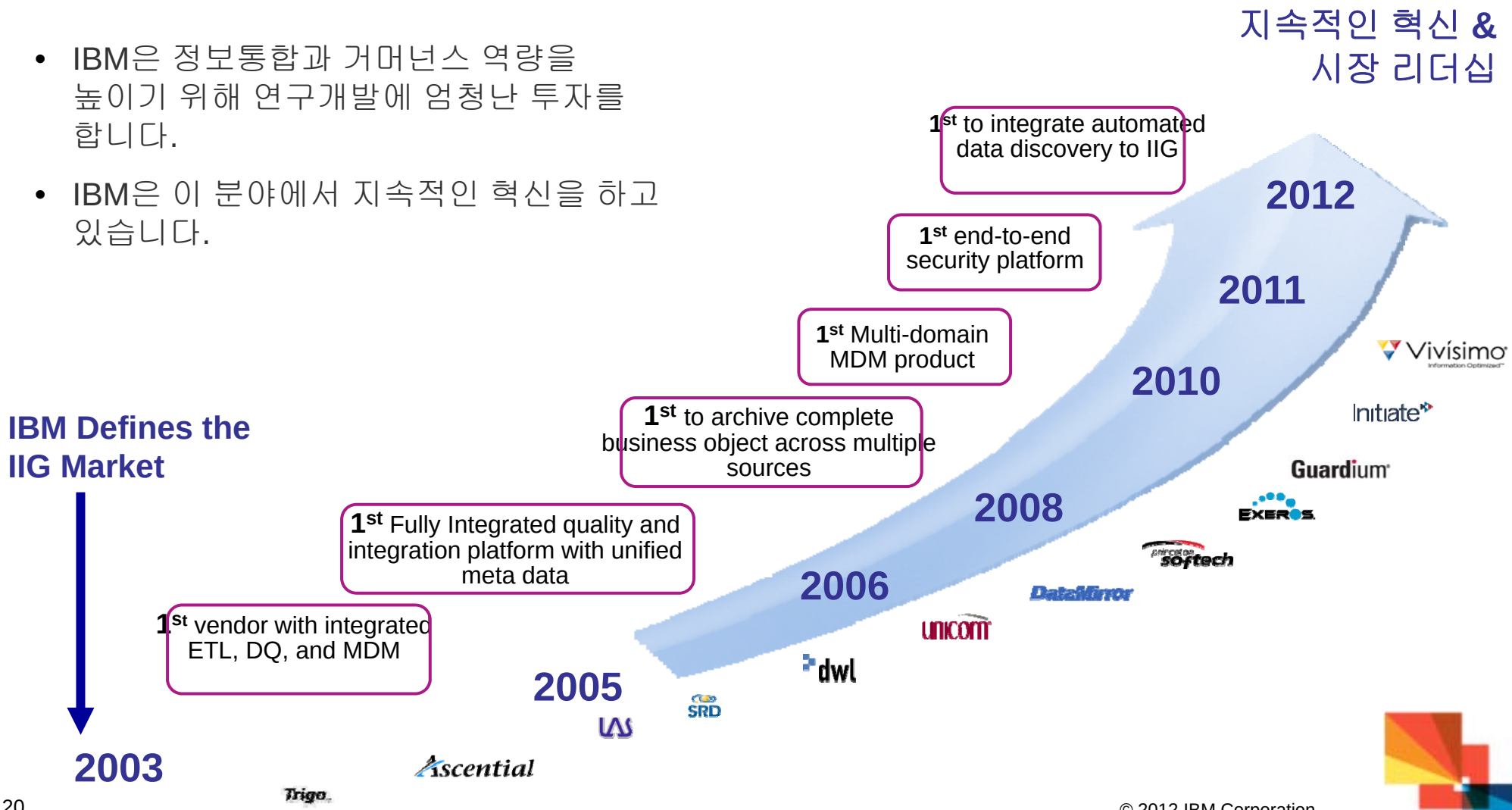
**Information Integration**



© 2012 IBM Corporation

# IBM은 정보통합과 거버넌스에 있어서 리더의 역할을 해왔으며, 이 시장에서 지속적으로 리딩하고 혁신을 할 것입니다.

- IBM은 정보통합과 거머넌스 역량을 높이기 위해 연구개발에 엄청난 투자를 합니다.
- IBM은 이 분야에서 지속적인 혁신을 하고 있습니다.






© 2012 IBM Corporation

# 정보통합에 있어서 IBM InfoSphere의 리더십

**Innovation**

*1st integrated information platform*

*1st ETL platform built for big data*

**Market Leadership**

*#1 in data integration & access market share (IDC)*

*#1 in data quality vision (Gartner)*

*#1 in integration capabilities (Gartner)*

**Customer Satisfaction**

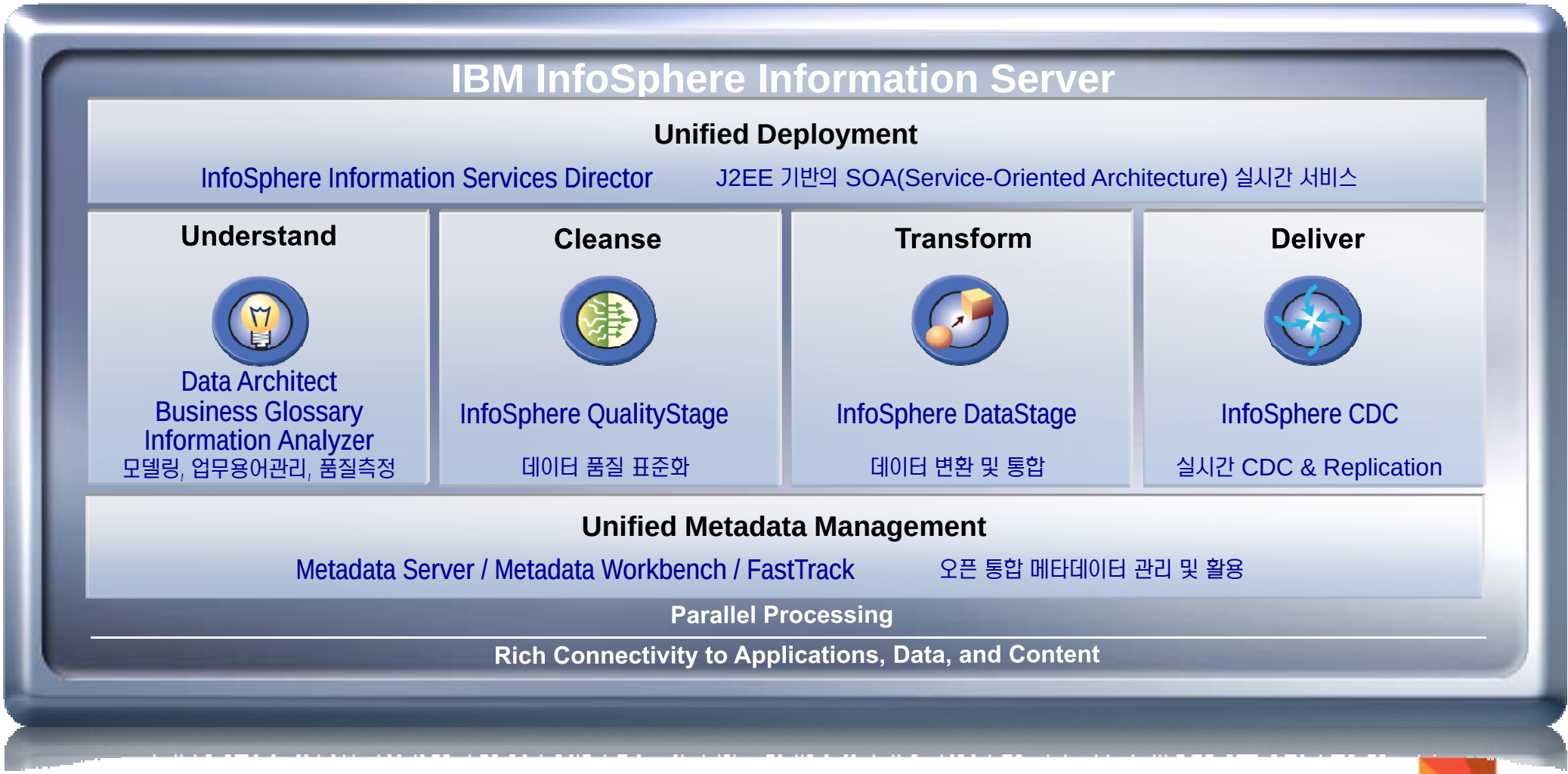
*"On average, InfoSphere Information Server software saves us around six hours per developer in terms of data modeling and ETL job creation"*

*-- University of Arizona*



© 2012 IBM Corporation

# IBM InfoSphere Information Server



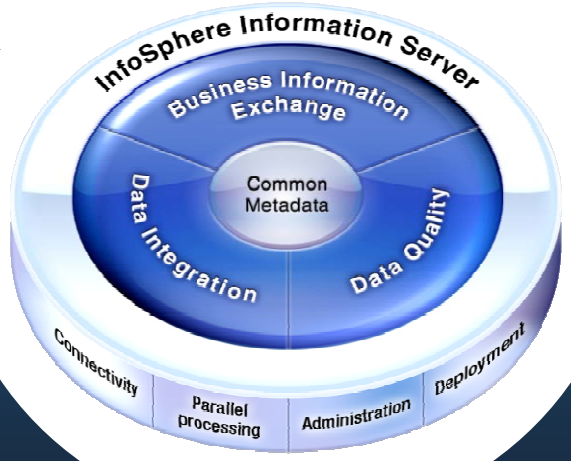



© 2012 IBM Corporation

# 단순화된 정보통합 솔루션 패키지 소개- **Simplified Packaging**

## InfoSphere Information Server Enterprise Edition:

Integrating and transforming data and content to deliver accurate, consistent, timely and complete information on a single platform unified by a common metadata layer




### Business Information Exchange

*Understanding & Collaboration*

- Information blueprints
- Relationship discovery across data sources
- IT-to-business mapping

### Data Quality

*Cleansing & Monitoring*

- Analysis & validation
- Data cleansing
- Data quality rules & management

### Data Integration

*Transformation*

- Massive scalability
- Power for any complexity
- Total traceability

*Delivery*

- Data capture at any time
- Delivery anywhere
- Big data readiness



© 2012 IBM Corporation

THINK
BIG