

빅 데이터 시대의 정보 전략

IBM 정보관리 소프트웨어



빅 데이터 정의

- 모든 종류의 데이터
- Volumes 이 큰 데이터
- 활용 가치는 있지만, 추출이 어려운 대상
- 데이터 발생 이후 최단 시간 내에 분석 필요



1일에 **12 TB** 의
Tweets 데이터 발생



1초에 **2,000** 번 이상의
HD급 동영상 데이터 저장
(다양한 저장 매체)



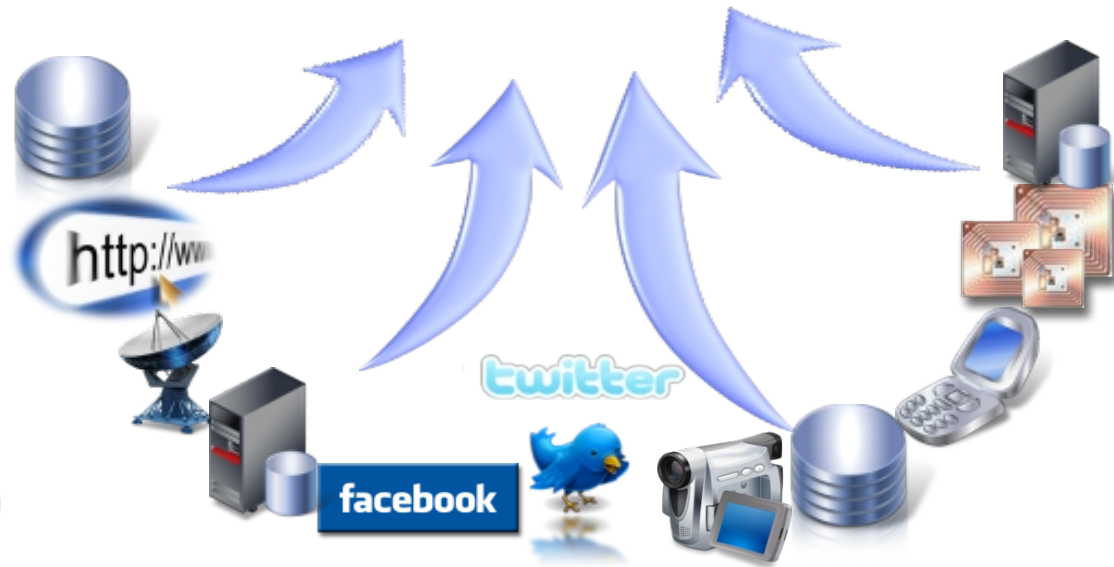
1일 단위 **1TB** 의
주식 거래 정보 생성
(**New York** 기준)



2010년 이후
스마트폰 사용자 급증
(한국 **3천만** 가입자)



네트워크 기술 및 속도 증가
(연간 **30%**)



다양한 분야에서 요구되는 빅 데이터 분석 요건

Big data challenges exist in every organization today

음향학적 분석(acoustic analysis) 1시간 소요 →
70 Milliseconds



전력 오류 방지 (avoids power failures)
분(Minntes)단위 분석 (**10 PB**)



의료 데이터 (vitals to detect illness)
1일전 → 현재 시점



재고 분석 (reduces time to run queries) → **80%** 재고 현황 유지



증권 거래소 (cuts queries)
26시간 → 2 분 (2 PB)



통신 네트워크 (analyses streaming network data)
Hardware 비용 **90%** 절감



금융

- 리스크 & 금융사기 관리
- 360도 고객 view 분석

유통 / 물류

- 결합 채널 마케팅
- 온라인 사용 패턴 분석
- 배송 정체 분석

건강 / 생명공학

- 의료 기록 분석, 유전자 / DNA 분석
- 질병 감시

통신

- CDR (Call detail record) 분석
- 고객 관심사항 분석

에너지 / 장치산업

- 시간대 별 사용량 분석
- 재고 / 자산 관리

사법 기관

- 복합적인 실시간 감시
- 사이버 범죄 탐지

기업내 데이터 전반의 연합된 Discovery and Navigation

Big Data 탐색 (Explorer)

Federate, discover and navigate Big Data sources



Big Data 관리

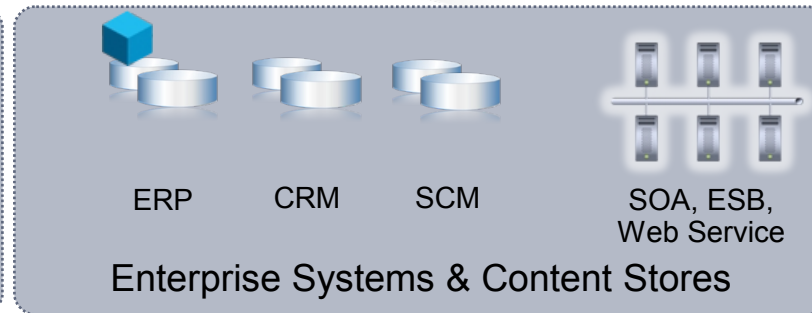
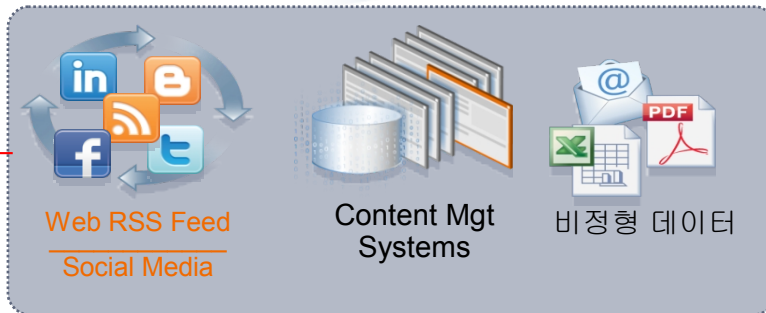
Stream, Process and analyze Big Data



- ✓ 데이터의 생성 양 · 주기 · 형식 등이 기존 데이터 와 상이함
- ✓ 전형적인 접근 방법으로는 수집 · 저장 · 검색 · 분석이 어려움

기존 방법 적용 (EAI, ETL, SOA, 프로그램)

구조화 되지 않은 형태로 존재



여러 시스템 전반에 걸쳐 다양하게 구조화 되어 존재

World's Total Data

80%

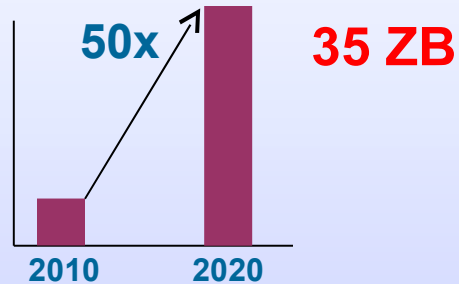
Unstructured (비정형 데이터)

20%

Structured (정형 데이터)

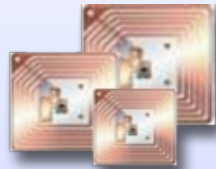
빅 데이터의 특징 (4 dimensions)

비용 효율적인 처리
의 필요성 **Volume**
(크기)



Tera → Peta → Exa → Zeta → Yota

실시간 처리에 대한 요
구사항
Velocity (속도)



30 Billion
RFID sensors
and counting

2분 이내 정보 전달

다양한 형태의 데이터에
대한 분석
Variety (다양)



80% 의 데이
터 비정형 구조

다양한 형태의 빅 데이터



데이터의 신뢰도
구축 필요
Veracity (신뢰)

의사 결정에 활용 하기 위한 요건

1 in 3 비즈니스 리더들은 **decision**를 내
리기 위해 사용되는 정보에 대해 높은 정
확성 / 신뢰성 요구

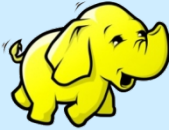
빅 데이터의 플랫폼의 요구사항

빅 데이터 소스에 대한 이해



다양한 데이터 소스에 대한 연동 및 탐색

대용량의 데이터에 대한 저장 및 관리



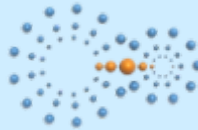
Hadoop File System
MapReduce

정형 및 제어 데이터



데이터 웨어하우스

스트리밍 데이터 관리



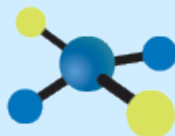
Stream Computing

비정형 데이터 분석



Text Analytics Engine

모든 데이터 소스에 대한 통합 및 통제



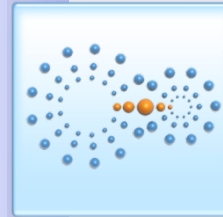
데이터 통합, 데이터 품질, 보안,
Lifecycle Management, MDM

IBM 빅 데이터 플랫폼



다양한 정보 분석

- 다양한 정보를 결합하여 분석 (Novel analytics)
- 예전에 없던 새로운 view 제공



움직이는 정보 분석

- 스트림 (Stream) 데이터 분석
- Large 볼륨 데이터에 대한 ad-hoc 분석



대용량 정보에 대한 분석

- 적은 비용으로 Petabytes 데이터에 대한 분석
- 대용량 데이터의 구조 및 관계 분석



빅 데이터 발견

- 대용량 데이터 Ad-hoc 분석
- 데이터 발견 (discovery & experimentation)



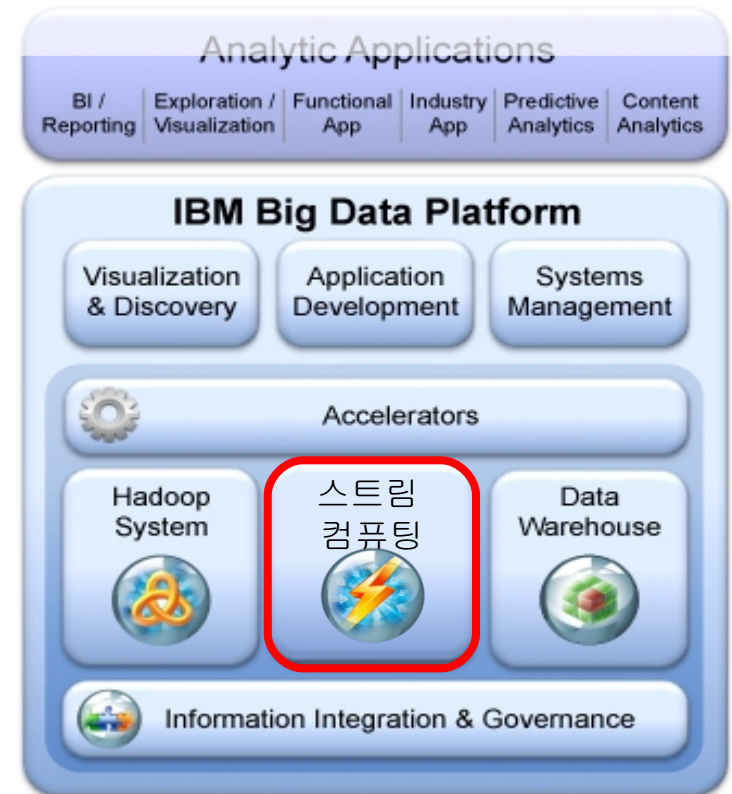
빅 데이터 관리

- 데이터 구조, 통합, 정합성 관리
- 반복적인 조회에 대한 일관성 유지



IBM InfoSphere Streams

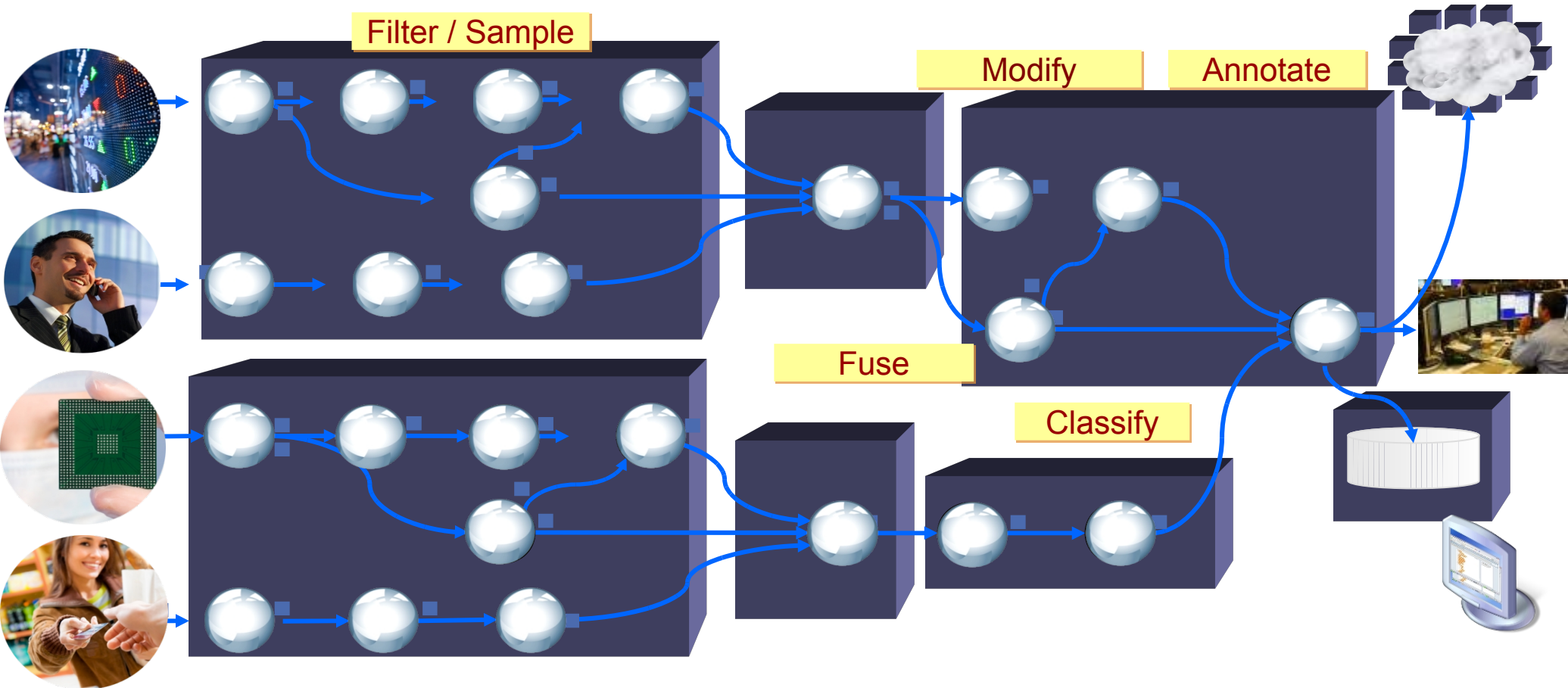
- 실시간으로 정형/비정형 데이터를 분석하기 위한 분석 플랫폼
- 실시간으로 데이터 분석
 - 이동중인 데이터에 대한 분석
 - Microsecond 수준의 응답 시간
 - 뛰어난 확장성
 - 동적 deploy
 - 데이터 마이닝, 텍스트 분석 등의 고급 분석 기능
- 다양한 데이터 소스
 - 정형, 비정형, 비디오, 오디오
 - 내장된 adapters/operators



IBM 빅 데이터 플랫폼 – Stream Computing



기업 전반에 존재하는 다양한 포맷의 데이터에 대한 실시간 추출 및 분석





University of Ontario Institute of Technology (UOIT) Detects Neonatal Patient Symptoms Sooner

Capabilities Utilized:

Stream Computing

- 신생아에서 나오는 생리학적 데이터에 대한 실시간 분석 작업 수행
- 데이터 간의 연관 관계를 분석하여 사소한 변화를 감지하여, 모니터 상에 스태프들에게 경고를 알림
- 조기 경보를 통해 합병증 등에 대하여 빠른 응대를 할 수 있게 됨

Significant benefits:

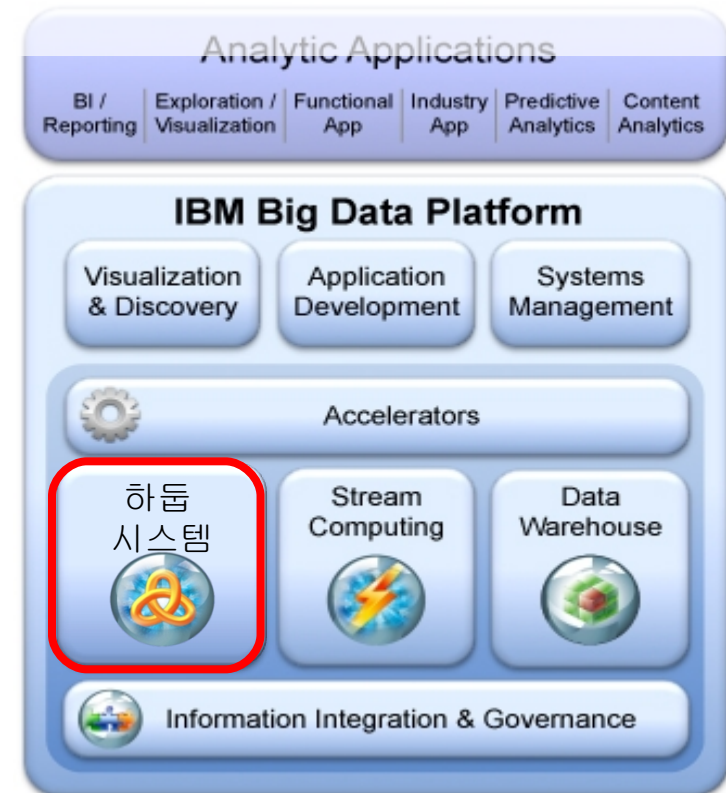
- 생명에 위협을 줄 수 있는 조건에 대한 감지를 24시간 이내로 줄여줌
- 질병률을 낮추고 향상된 환자 보호





IBM InfoSphere Biginsights

- 대용량의 정형/비정형 데이터를 위한 **hadoop** 기반의 분석 플랫폼
- **Open source Hadoop** 지원 역량 제공
 - 성능 최적화
 - 개발 툴링
 - 외부 연계
 - 분석 accelerators
 - 어플리케이션/인더스트리 accelerators
 - 시각화
 - 보안



The value of InfoSphere BigInsights



Hadoop 기반의 빅데이터 시스템 구축 및 분석 작업의 가속화



IBM Innovation

More Than Hadoop

User Interfaces

시각화

개발 툴

관리 콘솔

Accelerators

텍스트 분석

어플리케이션
액셀레이터

BigInsights Engine

Map Reduce +

인덱싱

워크로드 관리

보안

Apache Hadoop

Integration

Databases

Content Management

Information Governance

- 성능 & 워크로드 최적화
- 독보적인 텍스트 분석 엔진
- 데이터 검색 및 탐구를 위한 Spreadsheet 형식의 시각화 툴
- 내장된 IDE 및 관리 콘솔
- 엔터프라이즈 레벨의 보안
- 타 시스템과의 연계를 고 성능의 connector 지원
- 분석 가속기



Vestas optimizes capital investments based on 2.5 Petabytes of information.

Capabilities Utilized:

InfoSphere BigInsights

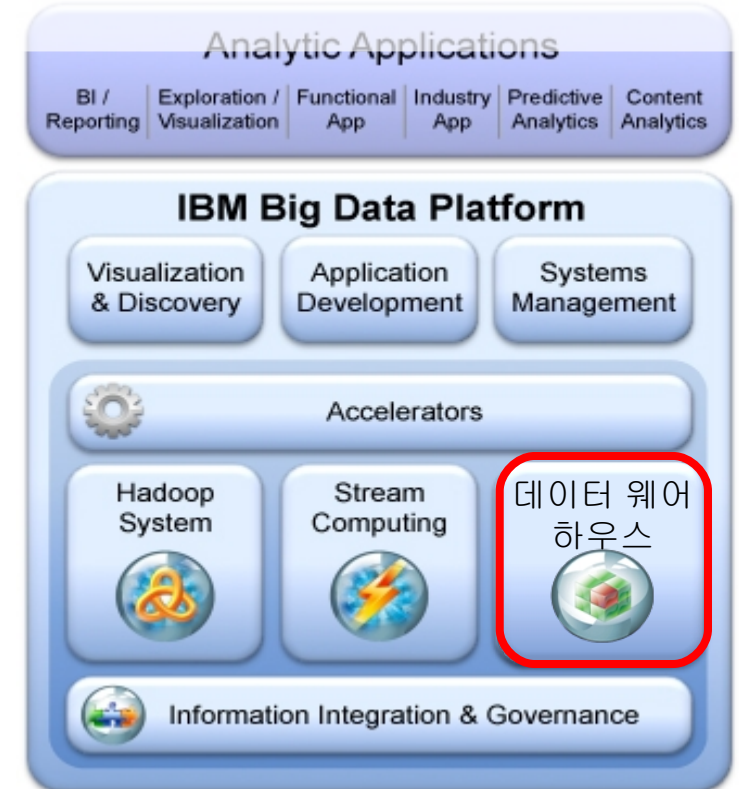
- 터빈 위치, 전력 생산 최대화와 지속성을 높이기 위해 날씨에 대한 모델링 수행
- 터빈의 배치 위치를 찾아내는데 수 주에서 수 시간으로 시간 감소
- 약 2.5PB의 정형과 반-정형 구조의 정보에 대한 연계 분석. 데이터 볼륨이 6PB까지 증가할 것으로 예상됨

Vestas[®]



IBM Data Warehousing

- **Workload** 최적화 시스템
 - 심화분석 어플라이언스
 - 운영 데이터 분석을 위한 어플라이언스 구성
 - 빅 데이터 웨어 하우스 구축
- **기능(Capabilities)**
 - 대용량 병렬처리 engine
 - 고성능 OLAP 분석
 - 운영 과 분석 데이터 가 결합 된 형태의 workload 지원



IBM 빅 데이터 플랫폼 – 데이터웨어하우스



빅 데이터 수용 및 대용량 데이터 확장이 손쉬운 어플라이언스 **DBMS**

IBM Netezza



Deep Analytics
Appliances

Deep Analytics

IBM Smart Analytics System



Operational Analytics
Optimized Systems

Operational Analytics

IBM Warehouse Software



Custom
Solutions

Warehouse Accelerators

Information Management Portfolio

(Information Server, MDM, Streams, etc)

The right mix of simplicity and flexibility

Simplicity, Flexibility, Choice
IBM Data Warehouse & Analytics Solutions

Pacific Northwest Smart Grid Demonstration Project

60k metered customers in 5 states

InfoSphere Streams → Real-time
transactive control system

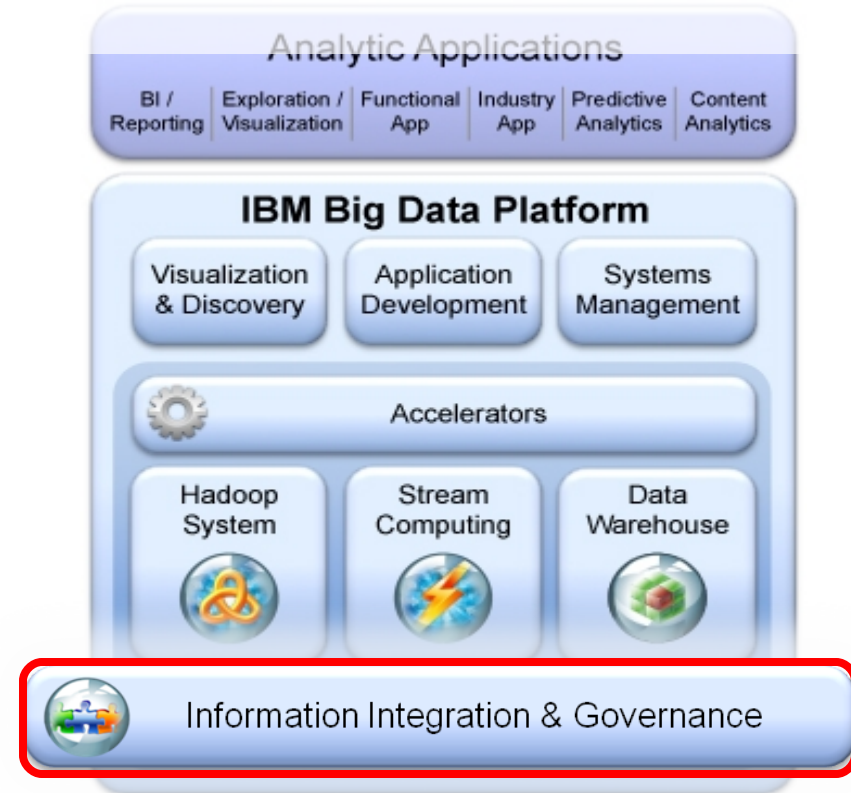
IBM Netezza → Deep analytics over
massive data sets





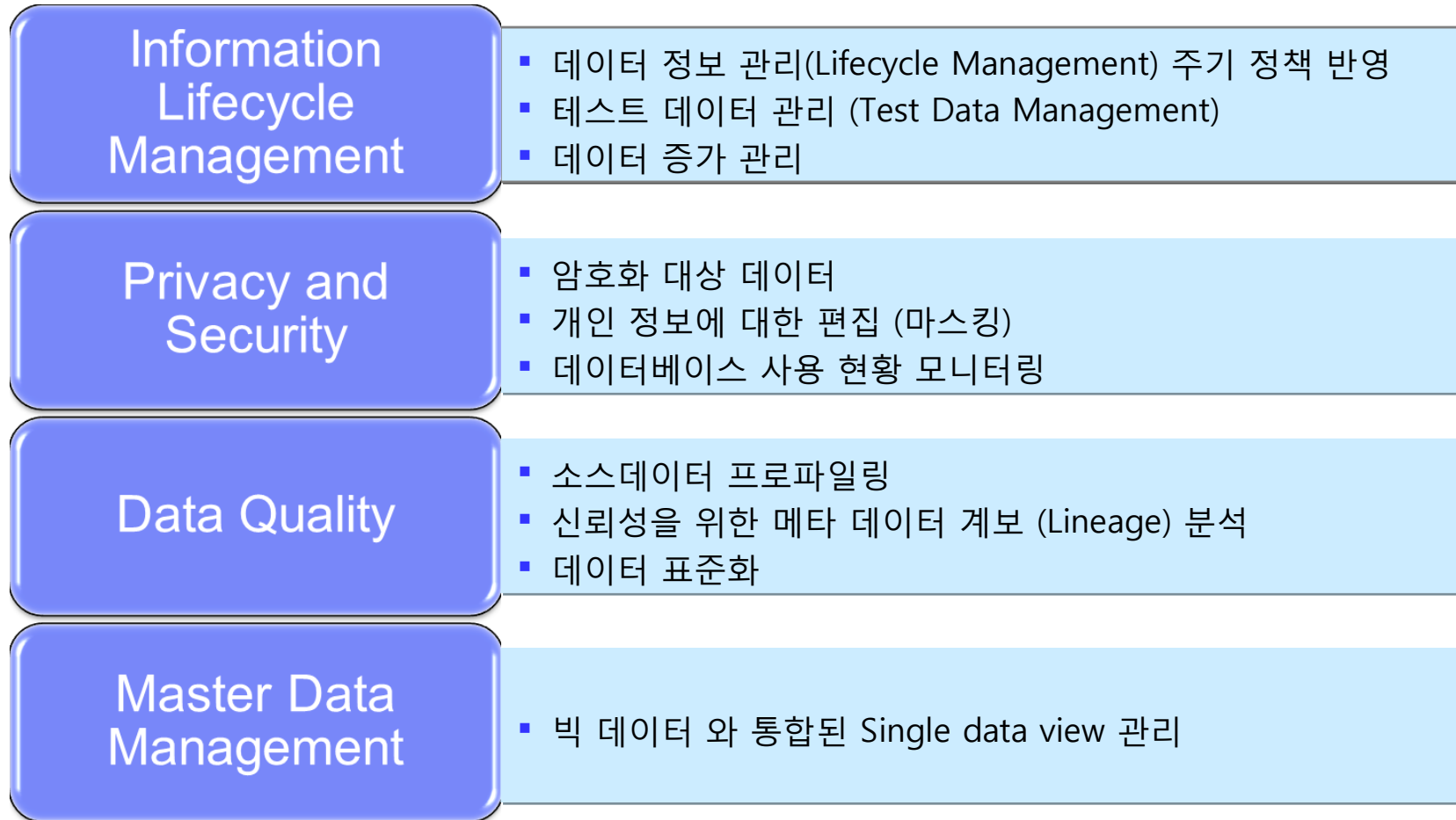
IBM InfoSphere Optim & DQM

- 빅 데이터 플랫폼에 통합된 다양한 데이터 포맷
 - Structured (정형 데이터)
 - Unstructured (비정형 데이터)
 - Streaming (스트림 데이터)
- 빅 데이터 거버넌스
 - 보안에 민감한 데이터
 - 데이터 증가에 따른 Lifecycle 관리가 필요한 데이터
 - 기업내 Single view 유지가 필요한 마스터 데이터





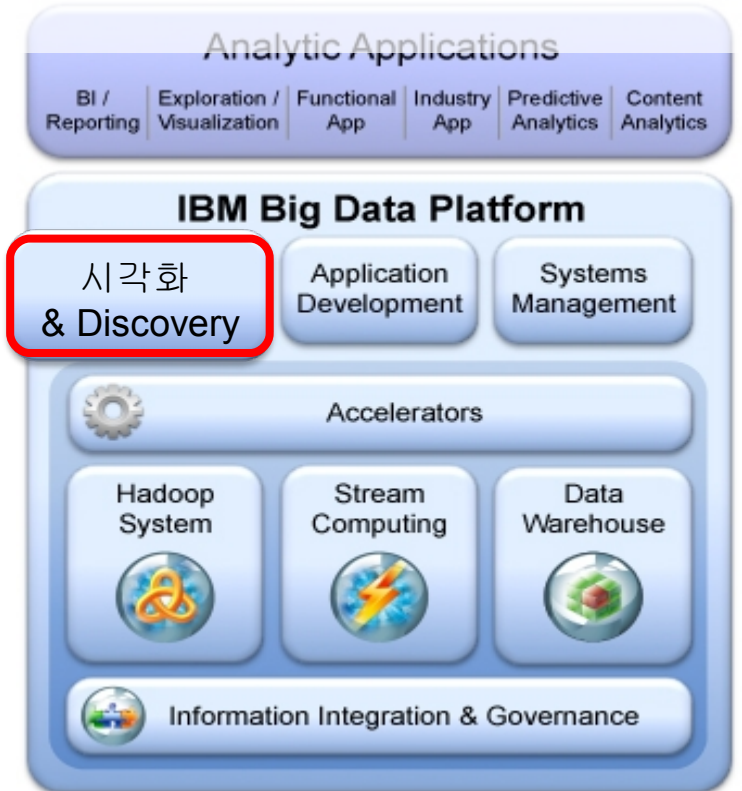
빅 데이터 거버넌스 를 위한 정보통합 거버넌스 솔루션 제공





IBM InfoSphere Data Explorer

- 다양한 빅데이터 소스에 손쉬운 접근
 - 다양한 문서 포맷
 - DBMS , Application
 - Web 문서 등,...
- 정형 / 비정형 데이터를 결합한 가상화된 **Single** 구조 생성
 - 360° view 분석 (고객, 제품, 직원, 프로젝트, . etc.)
 - 기업 콘텐츠와 빅데이터 소스를 연결 (예 : log files, tweets, 지리정보 , 이미지, etc.)
- 다양한 소스 결합을 통하여, 사용자 관심 사항 분석을 통한 새로운 **insights** 발견



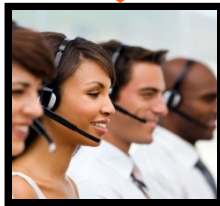
IBM 빅 데이터 플랫폼 – 정보 시각화 & Discovery



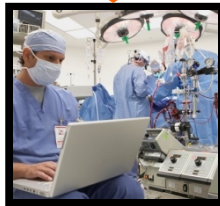
기업 내/외 부의 다양한 데이터 소스를 결합(**Federation**) 하여 분석 하기 위한 **Virtual** 문서 작성



고객 서비스 향상 & 응대 시간 단축



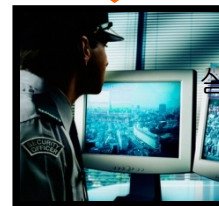
업무 생산성 향상 & 빠른 시장 진입 견인



고객 해석 및 분석 & 고객의 가치 이해



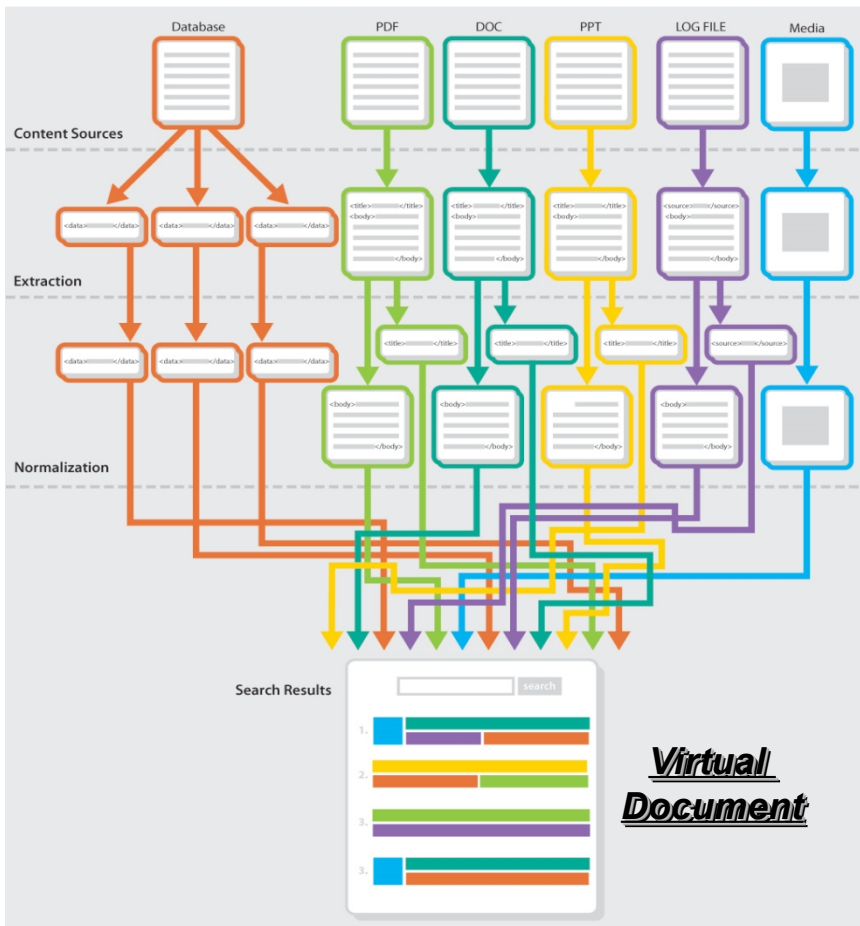
정보 측면의 리스크 분석 및 컴플라이언스 데이터 확인



실시간 감시를 위한 단일화된 View 생성

DEMO (InfoSphere Data Explorer)

다양한 종류의 데이터 소스



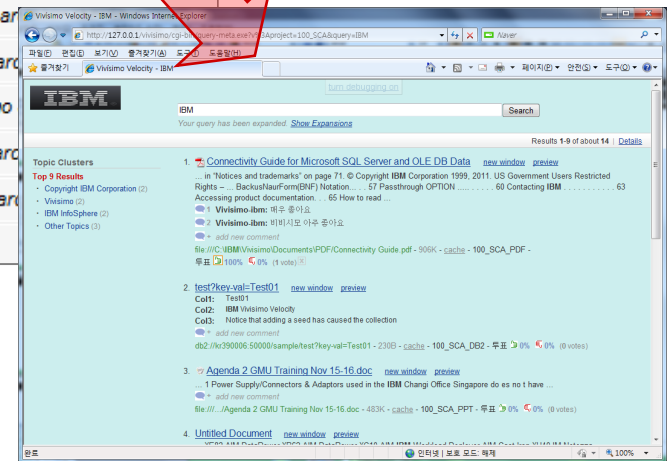
Source 100_SCA_all
last modified by vivisimo-ibm@IBM 9 minutes ago

Overview Declarations Testing XML

This source bundle consists of:
(no components yet)

and the following sources:

- [100_SCA_Cafe_Blog \(Vivisimo Search Engine\)](#)
- [100_SCA_DB2 \(Vivisimo Search Engine\)](#)
- [100_SCA_DOC \(Vivisimo Search Engine\)](#)
- [100_SCA_PDF \(Vivisimo Search Engine\)](#)
- [100_SCA_Symphony \(Vivisimo Search Engine\)](#)
- [100_SCA_PPT \(Vivisimo Search Engine\)](#)
- [100_SCA_XLS \(Vivisimo Search Engine\)](#)
- [100_SCA_BBC_news](#)



Data Explorer 예제 - 연합된 소스 (Federated Sources)

Source 100_SCA_all

last modified by vivisimo-ibm@IBM 9 minutes ago

Overview | Declarations | Testing | XML

This source bundle consists of:
(no components yet)

and the following sources:

- 100_SCA_Cafe_Blog (Vivísimo Search Engine)
- 100_SCA_DB2 (Vivísimo Search Engine)
- 100_SCA_DOC (Vivísimo Search Engine)
- 100_SCA_PDF (Vivísimo Search Engine)
- 100_SCA_Sympony (Vivísimo Search Engine)
- 100_SCA_PPT (Vivísimo Search Engine)
- 100_SCA_XLS (Vivísimo Search Engine)
- 100_SCA_BBC_news

Collection 100_SCA_Cafe_Blog

last modified 33 minutes ago

Overview | Configuration | Live Status

Seeds ([configuration](#))

- <http://blog.daum.net/xml/rss/redstock100>
- <http://blog.rss.naver.com/jhluvmin.xml>

Host	kr390006
Port	50,000
Username	kr390006
Password	*****
Database system	db2
Database name	sample
Table to retrieve	test
Key Column	Col1

인터넷 (카페, 블로그)

R-DBMS

MS엑셀

PDF

IBM실폰니

MS파워포인트

MS엑셀

뉴스 (URL)

Documents > DOC

이름

- Agenda 2 GMU Training N
- Agenda 2 GMU Training N

Documents > PDF

이름

- Connectivity Guide.pdf
- ISDCDC_652_DB2LUW_en.pdf

Documents > XLS

이름

- IBM Korea 2011 BSPL.xls
- ISDCV65ChipsetOSSupportMatrixX

Documents > PPT

이름

- IBM Korea Intro 2010.ppt
- XA & Fast Apply.ppt

Documents > Sympony

이름

- IS Install.odp
- SWG Appliance list 20120919.ods

Method	GET
URL	http://www.bbc.co.uk/search/news/?text=on
'Query' Parameter	q
'Per' Value	20
'Page' Parameter	page
'Page' Offset	1

Data Explorer 예제 – 연결 단어 설정

turn debugging on

IBM Search

Your query has been expanded. [Show Expansions](#)

IBM

<p><u>동의어</u></p> <p><input checked="" type="checkbox"/> Synonyms</p> <hr/> <p><input checked="" type="checkbox"/> 아이비엠</p> <p><input checked="" type="checkbox"/> IBM Korea</p> <p><input checked="" type="checkbox"/> I.B.M</p> <p><input type="checkbox"/> Translations <u>번역</u></p> <hr/> <p><input type="checkbox"/> International Business Machines</p>	<p><u>맞춤법 적용</u></p> <p><input checked="" type="checkbox"/> Spelling Variations</p> <hr/> <p><input checked="" type="checkbox"/> 아이삐엠</p> <p><input type="checkbox"/> 광의어</p> <hr/> <p><input type="checkbox"/> Broader Terms</p> <hr/> <p><input type="checkbox"/> Global IT Company</p>	<p><u>협어</u></p> <p><input checked="" type="checkbox"/> Narrower Terms</p> <hr/> <p><input checked="" type="checkbox"/> IBM IM team</p> <p><input type="checkbox"/> 연관어</p> <hr/> <p><input type="checkbox"/> Related Terms</p> <hr/> <p><input type="checkbox"/> 한국IBM</p> <p><input type="checkbox"/> 한국아이비엠</p>
--	---	--

select all for this term

3. [Untitled Document](#) [new window](#) [preview](#)

```

<word name="아이비엠">↵
<synonym>IBM</synonym>↵
<related>한국IBM</related>↵
<related>한국아이비엠</related>↵
<broader>Global IT Company</broader>↵
<synonym>IBM Korea</synonym>↵
<synonym>I.B.M</synonym>↵
<translation>International Business Machines</translation>
<narrower>IBM IM team</narrower>↵
<spelling>아이삐엠</spelling>↵
    
```

연결 단어
설정

Data Explorer 예제 – Wildcard 검색

turn debugging on

Topic Clusters

Top 11 Results

- + 대학교, 아이비리그 (7)
- IBM, 아이비엠 (2)
- Other Topics (2)

검색 결과를
단어별 자동 그룹핑

1. [아이비 리그 - 위키백과, 우리 모두의 백과사전](#) new window preview
 아이비 리그 위키백과, 우리 모두의 백과사전. (아이
 콜라미나대 라이스대 노트르담대 카네기멜론대 로체
 럽 ...
 + add new comment
 ko.wikipedia.org/wiki/아이비리그 - 43K - cache - 102
2. [미국 - 위키백과, 우리 모두의 백과사전](#) ne
 ... 예일 대학교, 프린스턴 대학교, 펜실베이니아 대
 있다. 이 부분의 본문은 미국의 교육입니다. 퍼블릭
 민지 시절부터 ...
 + add
 ko.wikiped
3. [프린스턴 대학교 - 위키백과, 우리 모두의 백과사전](#) ne
 ... 에 있는 사립 대학으로, 미국 동부 8개 명문 아이비
 수가 다른 아이비리그 학교에 비하여 적은 편이며 학
 는 아이비리그 대학중 하나인 프린스턴 대학교가 지
 + add new comment
 ko.wikipedia.org/wiki/프린스턴_대학교 - 97K - cach

Topic Clusters

Top 11 Results

- + 대학교, 아이비리그 (7)
- IBM, 아이비엠 (2)
- Other Topics (2)

1. [IBM - 위키백과, 우리 모두의 백과사전](#) new window pre
 ... i IBM PC 호환기종 [편집] 주석 ↑ 가 나 아이비엠. 《 오픈토리
 + add new comment
 ko.wikipedia.org/wiki/IBM - 53K - cache - 102_test03 - 투표 0%

Topic Clusters

Top 11 Results

- + 대학교, 아이비리그 (7)
- IBM, 아이비엠 (2)
- Other Topics (2)

1. [인텔 - 위키백과, 우리 모두의 백과사전](#) new
 ... 제스퍼 포레스트 · 볼룸필드 · 걸프타운 · 케인스타운
 레이크 / 스카미몬트 판매 ...
 + add new comment
 ko.wikipedia.org/wiki/인텔 - 104K - cache - 102_test03
2. [버지니아 주 - 위키백과, 우리 모두의 백과사전](#) ne
 ... 강남에 많이 소개된 북버지니아의 페어팩스는 미국
 + add new comment
 ko.wikipedia.org/wiki/버지니아 - 73K - cache - 102_tes

Data Explorer 예제 – Wildcard 검색

