비즈니스를 가속화하는

IBM 빅데이타 솔루션

한국 IBM 소프트웨어 그룹 안명주

21.May 2013

지금 비즈니스 모델들은 위협받고 있습니다.

고객들은
점점 더 긴밀하게 연결되어
점점 더 많은 것을 요구 하고 있으며

며칠만에
새로운 브랜드가 만들어졌다가
없어지기도 하며

좋은 관계는 좋은 제품을 만듭니다.

지금 비즈니스 모델들은 위협받고 있습니다.

변화하지 않는 것은 잃어버리는 것과 같다.

통신사

네트웍 데이타를 이용한 새로운 서비스 제공으로 비즈니스 창출을 위한 변화

소매점

소셜 미디어 및 네트웍, 그리고 모바일 커머스를 이용한 서비스의 변화

정치 캠페인

유권자 개개인에 대한 타케팅 및 모집으로의 변화

빅 데이타 분석이 왜 필요한가?

All information

All information
Transaction data
Application data
Machine data
Social data
Enterprise content

All perspectives

Past (historical, aggregated)
Present (real-time)
Future (predictive)

All people

All departments
Experts and non-experts
Executives and employees
Partners and customers

All decisions

Major and minor
Strategic and tactical
Routine and exceptions
Manual and automated

65%

비즈니스가 아직까지 비즈니스를 위해 빅데이타를 이용하고 있지 않다고 합니다.

빅 데이타는 혁신을 주도하는 차세대 천연 자원입니다.

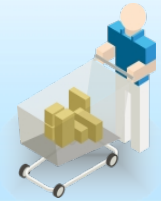
빅 데이타는 다양한 분야에서 비즈니스 혁신을 꾀합니다.

음향학적 분석
(acoustic analysis)
1시간 ➔ 70 밀리초

전력 오류 방지
(avoids power failures)
10 PB를 수 분내에 처리

조산아 Vital Sign분석
(vitals to detect illness)
1일전 ➔ 즉시

재고 최적화를 위한 분석
쿼리 응답 속도 80% 향상



증권 거래소
거래 심층 분석
26시간 ➔ 2 분 (2 PB)

통신사 네트워크 분석
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

IBM이 빅데이타 분석을 위한 솔루션을 제공합니다.

New

DB2 10.5 with BLU Acceleration 기술

강화

Big Data Platform 강화

New

PureData System for Hadoop 출시

BLU Acceeleration 은 생각의 속도로 분석을 처리합니다.

DB2 10.5 *with BLU Acceleration*

- **8배에서 최대 25배까지 더 빨라진 분석 및 레포팅 기능** 주1
- **베타 테스트 결과 10배의 압축율** 주2
- **NO 인덱스 ,aggregates , 튜닝 및 SQL 과 스키마 변경**

주 1) IBM 내부 테스트 결과 DB2 10.1의 행 기반의 테이블과 10..5의 columnar 테이블에 분석 워크로드를 수행할 IBM 내부 테스트 결과로 워크로드, 구성 및 조건에 따라 성능 결과는 달라질 수 있습니다.

주 2) DB2 10.5의 early release program에 참여한 고객 테스트 결과로 결과는 워크로드, 구성 요소 및 조건에 따라 달라질 수 있습니다

DB2® 10.5 특장점

with BLU Acceleration

빅 데이타의 Multi-workload 를 지원하는 차세대 데이타베이스

1

BLU Acceleration 기술

빅 데이타 시대의 분석 처리를 위해 인메모리 및 Columnar 기술을 통한 성능 및 압축율 극대화

2

DB2 pureScale 기능 강화

무중단 서비스를 지원하기 위한 DB2 pureScale 기능 강화

3

차세대 신 기술을 지원하는 다목적 데이타베이스

mobile 디바이스를 위한 다양한 기능 제공 및 차세대 애플리케이션을 위한 NOSQL 지원

4

오라클 호환성 강화

오라클 마이그레이션시 위험도와 비용을 줄일 수 있도록 오라클 호환성 강화

BLU Acceleration 이란?

1. 데이타를
2. 인메모리 기술로
3. 특정 컬럼 의 분석 질의 에 적합한

더 빠르고, 더 쉽게

생성하여 데이타 로드하고 실행하면 끝!!!

- 인덱스 불필요
- 튜닝 불필요
- aggregation 불필요
- SQL 및 스키마 변경 불필요

BLU Acceleration 기반 기술

Dynamic In-Memory

자주 사용되는 데이타를 동적으로 메모리로 이동 함으로써 효율적인 메모리 활용 및 획기적인 Query Latency 감소

Actionable 압축

순서를 그대로 유지한채 압축하는 업계 최초의 기술로 압축 해제 없이 데이타 이용 가능

병렬 Vector 프로세싱

Multi-core 와 SIMD(Single Instruction Multiple Data) 병렬처리로 성능 극대화

Data Skipping

쿼리 수행시 조건이 맞지 않는 데이타의 불필요한 처리를 자동으로 Skip 함으로써 성능 극대화

BLU Acceleration 특장점

분석을 위한 차세대 데이타베이스

- 성능 극대화
- 스토리지 절감 극대화
- 획기적인 운영 분석 비용 절감

기존 DB2 엔진과의 완벽한 통합

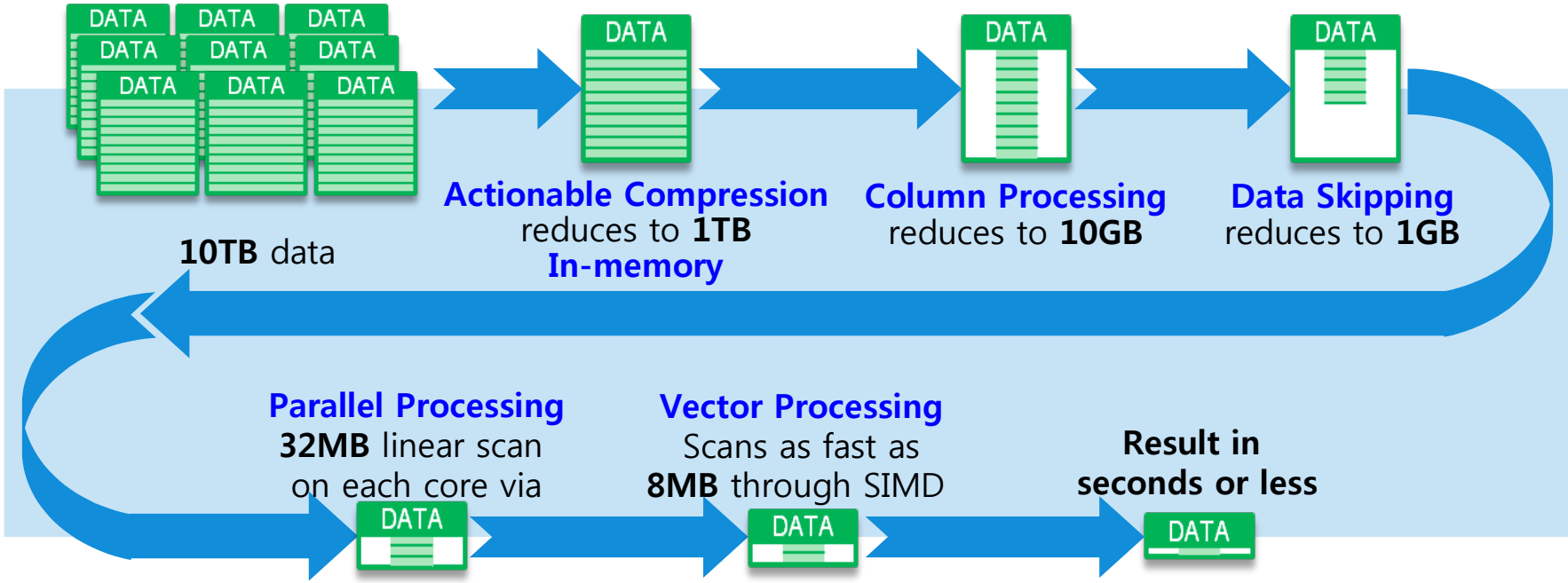
- 동일한 SQL, 인터페이스 및 관리 사용
- 디자인 및 튜닝이 필요 없는 개발 및 운영의 단순화

하드웨어 성능 최적화

- 메모리 압축으로 압축 최적화
- Modern CPU Exploitation
- 필요한 데이타만을 읽음으로써 I/O 최적화

BLU Acceleration 예제

10TB 데이타를 Read 하는데 일초 미만이 소요



- 시스템 사양 : 32 cores / 1TB memory
- 데이타 크기 : 100 컬럼을 가진 10TB 테이블 (10년동안 축적된 데이타)
- Query : 2010년에 얼마나 많은 "SALES"가 일어 났는가 ?
  - `SELECT COUNT(*) FROM MYTABLE WHERE YEAR = '2010'`

BLU Acceleration의 간편성

DATABASE

MICROSOFT
SYBASE
TERADATA
ORACLE

Database Design and Tuning

- Decide on partition strategies
- Select Compression Strategy
- Create Table
- Load data
- Create Auxiliary Performance Structures
  - Materialized views
  - Create indexes
    - B+ indexes
    - Bitmap indexes

Repeat

- Tune memory
- Tune I/O
- Add Optimizer hints
- Statistics collection

DB2® 10.5 특장점

with BLU Acceleration

빅 데이타의 Multi-workload 를 지원하는 차세대 데이타베이스

1. **BLU Acceleration 기술**

   빅 데이타 시대의 분석 처리를 위해 인메모리 및 Columnar 기술을 통한 성능 및 압축율 극대화

2. **DB2 pureScale 기능 강화**

   무중단 서비스를 지원하기 위한 DB2 pureScale 기능 강화

3. **차세대 신 기술을 지원하는 다목적 데이타베이스**

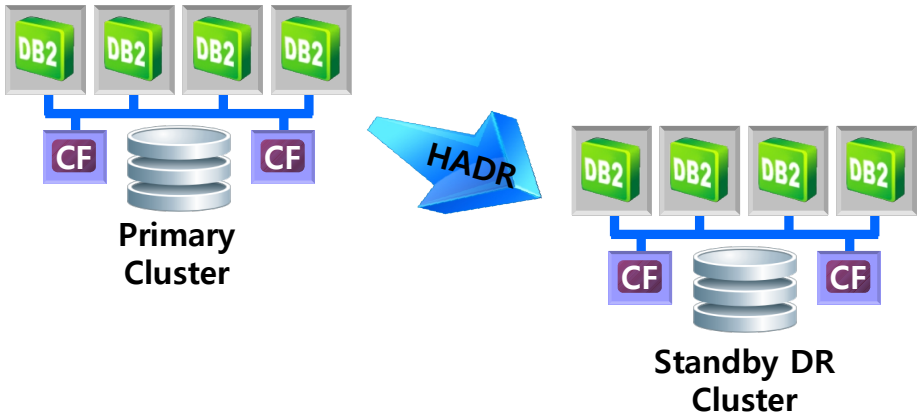
   mobile 디바이스를 위한 다양한 기능 제공 및 차세대 애플리케이션을 위한 NOSQL 지원

4. **오라클 호환성 강화**

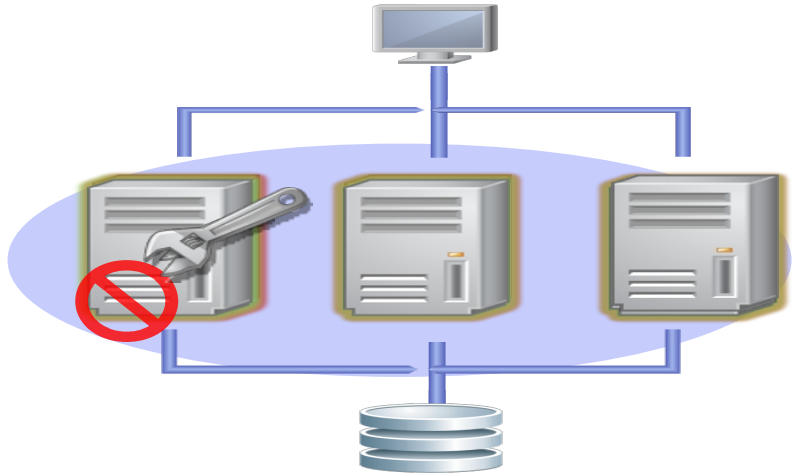
   오라클 마이그레이션시 위험도와 비용을 줄일 수 있도록 오라클 호환성 강화

무중단 서비스를 위한 pureScale 기능 강화

1. pureScale과 HADR를 결합하여 다양한 장애 대처 가능

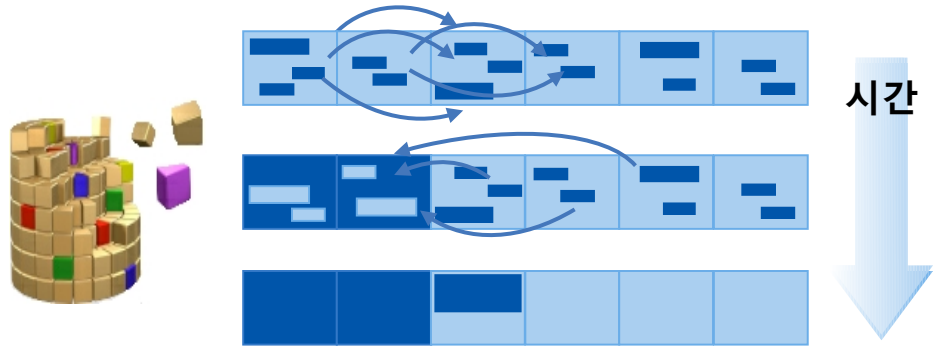


2. Rolling fix Pack Update 로 무중단 운영 관리 제공

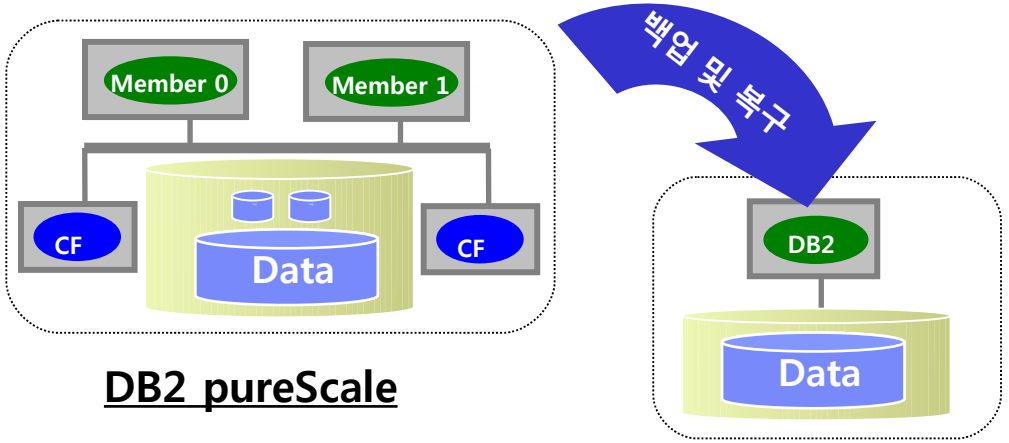


무중단 서비스를 위한 pureScale 기능 강화

3. 온라인 테이블 REORG으로 관리 편의성 제공



4. 멤버수에 관계없이 백업/복구가 가능하므로 DR 및 개발 시스템 구축 용이



DB2 pureScale

DB2 single

IBM이 빅데이타 분석을 위한 솔루션을 제공합니다.

New

DB2 10.5 with BLU Acceleration 기술

강화

Big Data Platform 강화

New

PureData System for Hadoop 출시

빠른 빅데이타 분석으로 비즈니스 활용을 극대화합니다.

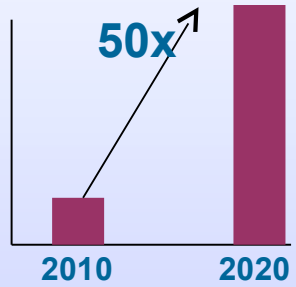
Big Data 플랫폼 기능 강화

- **Big SQL** 표준 ANSI SQL을 이용한 BigInsight내 데이타 접근 가능
- **GPFS-FPO** POSIX 컴플라이언스 및 보안 기능 강화
- bouned list 및 maps3를 이용하여 **2배에서 10배까지 빨라진** 스트림 처리 주3

주 3) InfoSphere Streams 3.0과 비교할 때 bounded List 및 MAP3를 통해 런타임 성능 향상

빅 데이터의 특성 – Volume, Velocity, Variety, Veracity

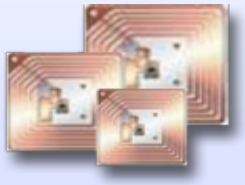
비용 효율적인 처리의 필요성 Volume (크기)



35 ZB

Tera ➔ Peta ➔ Exa ➔ Zeta ➔ Yota

실시간 처리에 대한 요구사항 Velocity (속도)



30 Billion RFID sensors and counting

2분 이내 정보 전달

다양한 형태의 데이터에 대한 분석 Variety (다양)

80% 의 데이터 비정형 구조

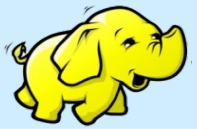
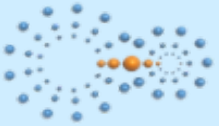
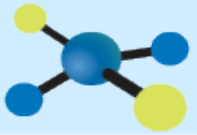
다양한 형태의 빅 데이터

데이터의 신뢰도 구축 필요 Veracity (정확성)

의사 결정에 활용 하기 위한 요건

1 in 3 비즈니스 리더들은 decision를 내리기 위해 사용되는 정보에 대해 높은 정확성 / 신뢰성 요구

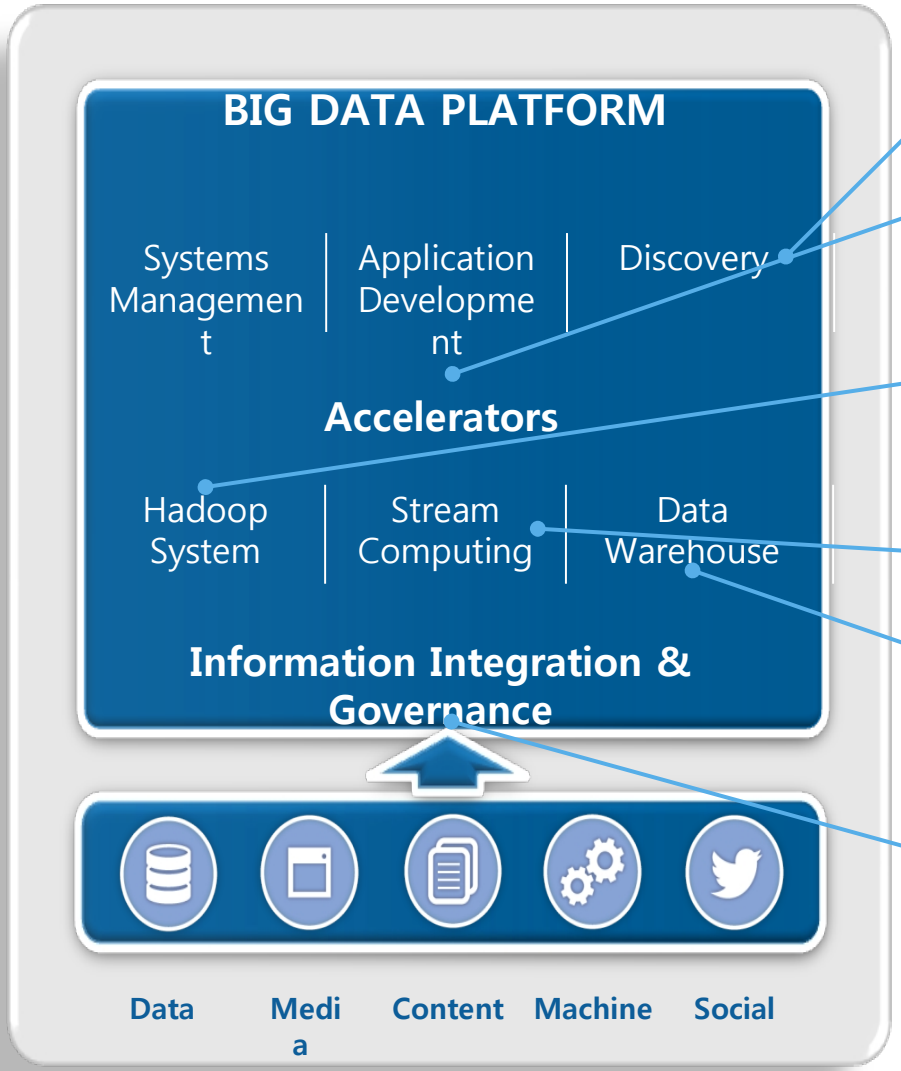
빅 데이터 분석을 위한 기술 요소

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**

비즈니스 중심의 Big Data를 위한 플랫폼 요건

1. **Unlock Big Data**
   (빅 데이터 검색 및 발견)
   빅데이터를 빠르게 이해 할 수 있는 사용자 View 필요
2. **Analyze Raw Data**
   (원시 데이터 분석)
   데이터가 존재하는 포맷 그대로의 분석
3. **Simplify Your Warehouse**
   (웨어하우스 단순화)
   Deep analytics 작업 목적의 warehouse 구축
4. **Reduce Costs with Hadoop**
   (하둡으로 비용 절감)
   비용측면의 효율성을 위한 Hadoop 기반의 시스템 구축
5. **Analyze Streaming Data**
   (스트리밍 데이터 분석)
   연결된 Stream 데이터 에 대한 분석 필요

IBM의 주요 빅데이타 플랫폼



InfoSphere Data Explorer
빅 데이타 발견,탐색 및 시각화

Accelerators
Speed time to value with analytic and application accelerators

InfoSphere BigInsights
오픈 소스인 하둡을 기업내에서 사용가능토록 안정성 및 보안 강화

InfoSphere Streams
실시간으로 움직이는 데이타에 대한 빠른 분석

PureData for Analytics and InfoSphere Warehouse
Delivers deep insight with advanced database analytics & operational analytics

Information Integration and Governance
Govern data quality and manage the information lifecycle

IBM 빅 데이터 플랫폼 – IBM BigInsight 2.1 발표

IBM InfoSphere BigInsights

- **대용량의 정형/비정형 데이터를 위한 hadoop 기반의 분석 플랫폼**

- **Open source Hadoop 기반의 솔루션으로 기업의 빅데이타 활용 극대화 및 보안 강화**
  - 성능 & 워크로드 최적화
  - 독보적인 텍스트 분석 엔진
  - 데이터 검색 및 탐구를 위한 Spreadsheet 형식의 툴 시각화 툴
  - 내장된 IDE 및 관리 콘솔
  - 엔터프라이즈 레벨의 보안
  - 타 시스템과의 연계를 위한 고 성능 connector 지원
  - 분석 가속기

IBM 빅 데이터 플랫폼 – IBM BigInsight 2.1 발표

IBM InfoSphere BigInsights

- **대용량의 정형/비정형 데이터를 위한 hadoop 기반의 분석 플랫폼**

- **Open source Hadoop 기반의 솔루션으로 기업의 빅데이타 활용 극대화 및 보안 강화**
  - 성능 & 워크로드 최적화
  - 독보적인 텍스트 분석 엔진
  - 데이터 검색 및 탐구를 위한 Spreadsheet 형식의 툴 시각화 툴
  - 내장된 IDE 및 관리 콘솔
  - 엔터프라이즈 레벨의 보안
  - 타 시스템과의 연계를 위한 고 성능 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 찾아대는데 수 주에서 수 시간으로 시간 감소
- 약 2.5PB의 정형과 반-정형 구조의 정보에 대한 연계 분석. 데이터 볼륨이 6PB까지 증가할 것으로 예상됨

Vestas

Cisco turns to IBM big data for intelligent infrastructure management

Capabilities Utilized

- ***Streaming Analytics***
- ***Hadoop System***
- ***Business Intelligence***

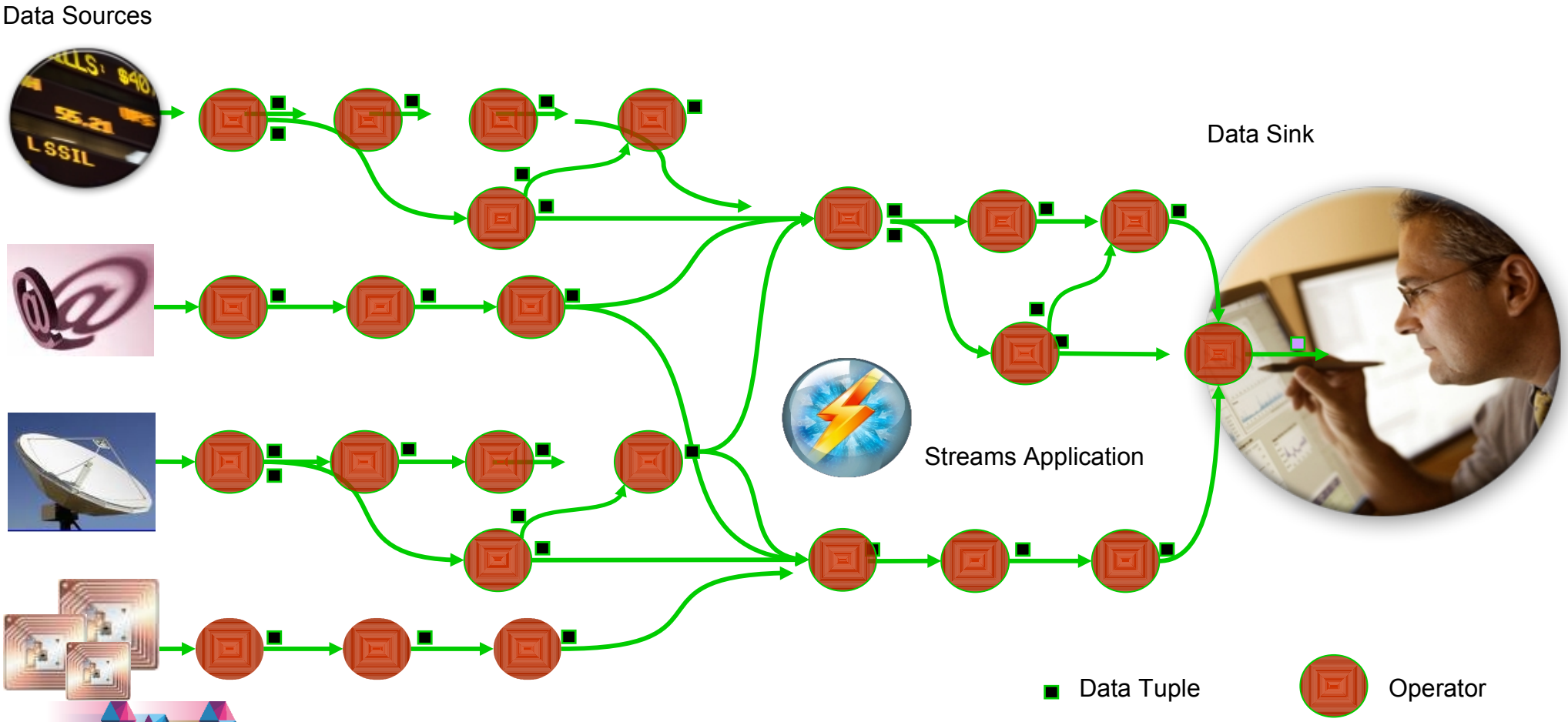
Results

• 중앙 관제식 모니터링을 통해 빌딩 내의 **에너지 소요량을 최적화**

• **자동 예방 및 수정 관리**

IBM 빅 데이터 플랫폼 – Stream Computing

기업 전반에 존재하는 다양한 포맷의 데이터에 대한 실시간 추출 및 분석



IBM 빅 데이터 플랫폼 – InfoSphere Streams 3.1 발표

IBM InfoSphere Streams

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

InfoSphere Streams 3.1 발표

- maps 및 lists를 이용하여 2배에서 최대 10배까지 런타임 성능 향상
- 애플리케이션 개발 편의성을 위해 개발자 를 위한 다양한 기능 제공
- 단순화된 대규모 클러스터 구성 및 통합

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

Capabilities Utilized:

Stream Computing

- 신생아에서 나오는 생리학적 데이터에 대한 실-시간 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

UOIT
CHALLENGE INNOVATE CONNECT

“Identifies and classifies potential security threats – miles away”

TerraEchos Turns to IBM Big Data for Low Latency Surveillance Data Analysis

Capabilities Utilized

Stream Computing

- Highly Sensitive Lab에서 **잠재적인 위협을 감지, 구분, 탐색 및 추적을** 위한 보안 감시 시스템을 구축
- 광섬유 센서들에서 **음향 데이터를 수집 및 분석**
- 분석된 음향 데이터를 TerraEchos intelligence platform에 적재하여 위험 감지, 구분, 예측 및 통신 용도에 활용

Results

- **실 시간으로 음향 데이터를 분석하고 구분할 수 있는 시스템 구축**
- Lab과 보안 스태프들에게 전체적인 뷰 측면에서 잠재적인 위험 또는 일상적인 상황인지를 전달
- 어느 위험에도 보다 빨리 지능적으로 응대할 수 있게 해줌

TerraEchos

어플라이언스의 간편성으로 더 많은 데이타를 분석합니다.

PureData System for Hadoop

- 맞춤 솔루션보다 **8배 빠른** 설치 및 구성 주4
- **분석 aceelerator**가 내장된 최초의 어플라이언스
- **아카이빙 툴**이 내장된 유일한 하둡 시스템

주 4) IBM 내부 테스트 및 고객 feedback을 기반으로 한 자료로 맞춤형 클러스터는 전문적으로 사전에 구성/테스트/최적화가 되지 않은 클러스터를 의미한다.

빅 데이타 분석의 이슈

"빅 데이타는 비즈니스에 많은 이득을 주지만 숨겨진 비용과 복잡성은 조직에서 고군분투해야 할 장애물이 된다. "

- The Cost of Big Data, Eric Savitz, Forbes 5/2012

- 오픈 소스인 아파치 하둡을 기업내에서 사용하기에는 불안하다.
- 하둡 전문가가 부족하다.
- 개별적으로 구축된 솔루션은 통합 클러스터 관리가 부족하다.
- 기존 분석 시스템과 통합을 위해서는 별도의 노력이 요구된다.

빅 데이타 분석의 단순성

여러 개 솔루션을 조립해야 하는 복잡성에서 단순성으로

Designed to...

- 하둡 클러스터의 구축, 전개 및 관리의 단순화
- 하둡과 구조화되지 않은 데이타의 빠른 가치 실현
- 전체 유관 분석 시스템의 효율성 극대화
- 기업 레벨의 보안 및 플랫폼 관리 제공

IBM PureData System for Hadoop의 특장점

내장된 전문성

빅 데이타 가치 실현 시간 단축

- 맞춤형 솔루션보다 **8배 빠른 배치**
- 통찰력을 가속화하기 위한 내장된 **시각화 기능**
- 타사 대비 우수한 성능의 소셜 및 기계, 텍스트 **analytics accelerator 내장**

단순화된 전문가 경험

단순화된 빅 데이타 도입 및 활용

- **단일 콘솔을** 통해 전체 시스템 관리
- **자동화**를 통한 신속한 소프트웨어 업데이트 제공
- 불필요한 조립 및 **수 시간내에 데이타 로드**

설계부터 고려된 통합

빅 데이타의 기업내 가치실현을 위한 구현

- 내장된 아카이빙 툴이 있는 **유일한 통합 하둡 시스템**
- open source software 보다 **한층 더 강화된 보안** 제공
- **고가용성**을 위한 아키텍처

IBM은 준비된 파트너로써 비즈니스 성공을 도와 드립니다

IBM은 모든 데이타를 이해합니다.

- **변화를 주도하는 혁신 :** Watson 컴퓨터, BLU 가속화 기술 스트리밍 분석 및 전문가 통합 시스템, 20년간 특허 취득 1위
- **비즈니스를 준비하는 역량 :** 심도 있는 사용을 위한 광범위한 빅데이타 및 분석 기능의 통합 및 강화, 유연한 배치 옵션

IBM은 데이타를 가치로 바꾸는 방법을 알고 있습니다.

- **고객 경험**– 깊이 있는 indutry 노하우 및 WW에서 사용하는 글로벌 솔루션
- **강력한 Ecosystem** – 375곳 이상의 파트너 및 200 곳이 넘는 대학에 대한 투자 증대
- **신기술에 대한 꾸준한 투자** – 누구도 따라올 수 없는 폭넓고 깊이 있는 새로운 기능을 통해 기존의 분석 및 정보 인프라 강화

IBM은 빅데이타 및 분석에 많은 투자를 해왔고 하고 있습니다.

- **160억 달러 이상의 인수 규모** – 2005년부터 솔루션 인수를 통해 변화를 주도하는 혁신 수행
- **9곳의 분석 솔루션 센타 운영**–4000곳이 넘는 조직이 센타 방문하여 전문 지식 공유
- **200억 달러 규모의 비즈니스 전망** – 2015년경에 데이타 분석 비즈니스를 통한 연간 수입이 200억 달러에 이를 것으로 전망

빅데이타 분석, IBM이 도와 드립니다.

THINK

BIG