

성공적인 미래를 현실로 - IBM 빅데이터 플랫폼

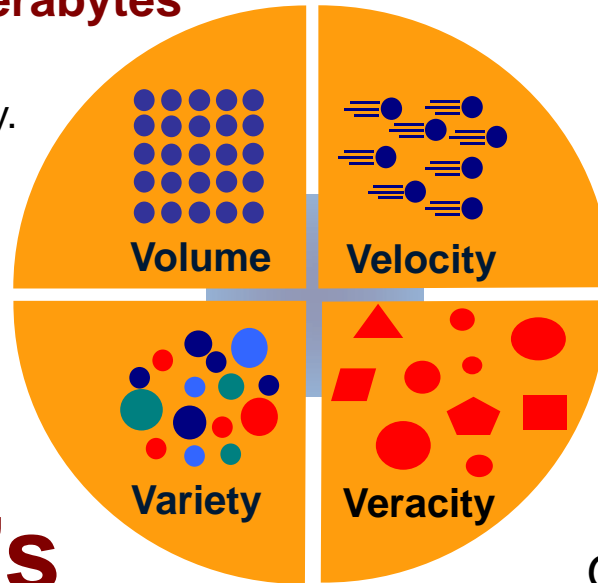
IBM Korea / SWG Information Management
소가영 차장 (gyso@kr.ibm.com)



컴퓨팅의 새로운 시대 > 빅데이터

비용 효율적인
대용량 처리

12+ terabytes
of Tweets
create daily.



실시간 처리

5+ million
trade events
per second.

다양한 형태의
데이터 분석

100's
of different types of data.

Only **1 in 3**
decision makers trust
their information.

데이터의 신뢰도



빅데이터는 비즈니스의 모든 부분에 영향을 미칩니다.



고객의 모든 것을 알 수 있습니다.

개별 고객과 관련된 모든 데이터 소스, 채널 및 소셜 미디어 상호작용을 포함한 데이터를 이용한 고객 분석

실시간 운영이 가능합니다.

가능한 모든 운영 데이터의 실시간 분석 및 효과적 프로세스를 통한 실시간 비즈니스 대응

신속하고 폭넓은 신상품 도입이 가능합니다.

다양한 소스의 피드백을 수집하고 한발 앞서 시장을 분석하여 신상품 도입에 접목

실시간 사기 및 리스크 감지가 가능합니다.

모든 가능한 데이터 소스의 분석으로 좀 더 나은 사기 및 리스크 감지 모델 개발과 스트리밍 트랜잭션 분석을 통한 실시간 감지

기계적 자산을 활용할 수 있습니다.

기계 자산의 실시간 모니터링을 통하여 유지보수 이슈를 방지하고 예측하며, 새로운 제품 및 서비스 개발에 활용





Constant Contact Transforming Email Marketing Campaign Effectiveness with IBM Big Data

Capabilities

- *InfoSphere BigInsights, IBM PureData for Analytics, Cognos BI*


Need

- Analyze **35 billion** annual emails to guide customers on best dates & times to send emails for maximum response

Benefits

- 40 times improvement in analysis performance
- 15-25% performance increase in customer email campaigns
- Analysis time reduced from hours to seconds

Constant Contact 



Ufone reduced churn and kept subscribers happy, helping ensure that campaigns are highly effective and timely

Need

- To ensure that its marketing campaigns targeted the right customers, before they left the network
- To keep its higher usage customers happy with campaigns offering services and plans that were right for them

Benefits

- Predictive analytics is expected to improve the campaign response rate from about 25% to at least 50%
- CDRs can be analyzed within 30 seconds, instead of requiring at least a day
- Expected to reduce churn by approximately 15-20%



빅데이터 활용의 직면 과제

다양한 빅데이터 소스들의 연계 및 이해

정형, 반정형, 비정형 다양한 데이터 타입의 빅데이터 분석

빅데이터 분석의 비용 효율성

실시간 데이터 처리 및 분석

폭발적 데이터량 증가로 인한 기존 정형 데이터 분석의 효율성



빅데이터 활용을 위한 플랫폼 요구사항

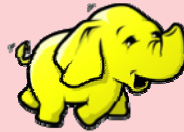
플랫폼
빅 데이터

다양한 빅데이터 소스의
이해와 네비게이션



데이터 소스의 연동과 탐색

다양한 타입의 대용량
데이터에 대한 저장 및 관리



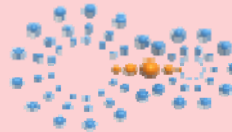
Hadoop File System
MapReduce

정형 데이터 관리 및 분석



Data Warehousing

스트리밍 데이터 관리



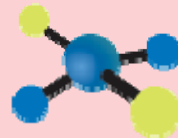
Stream Computing

비정형 데이터 분석



Text Analytics Engine

모든 데이터 소스에 대한
통합 및 통제



데이터 통합, 품질, 보안
ILM, MDM



IBM 빅데이터 플랫폼



다양한 정보 통합 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 소스의 발견

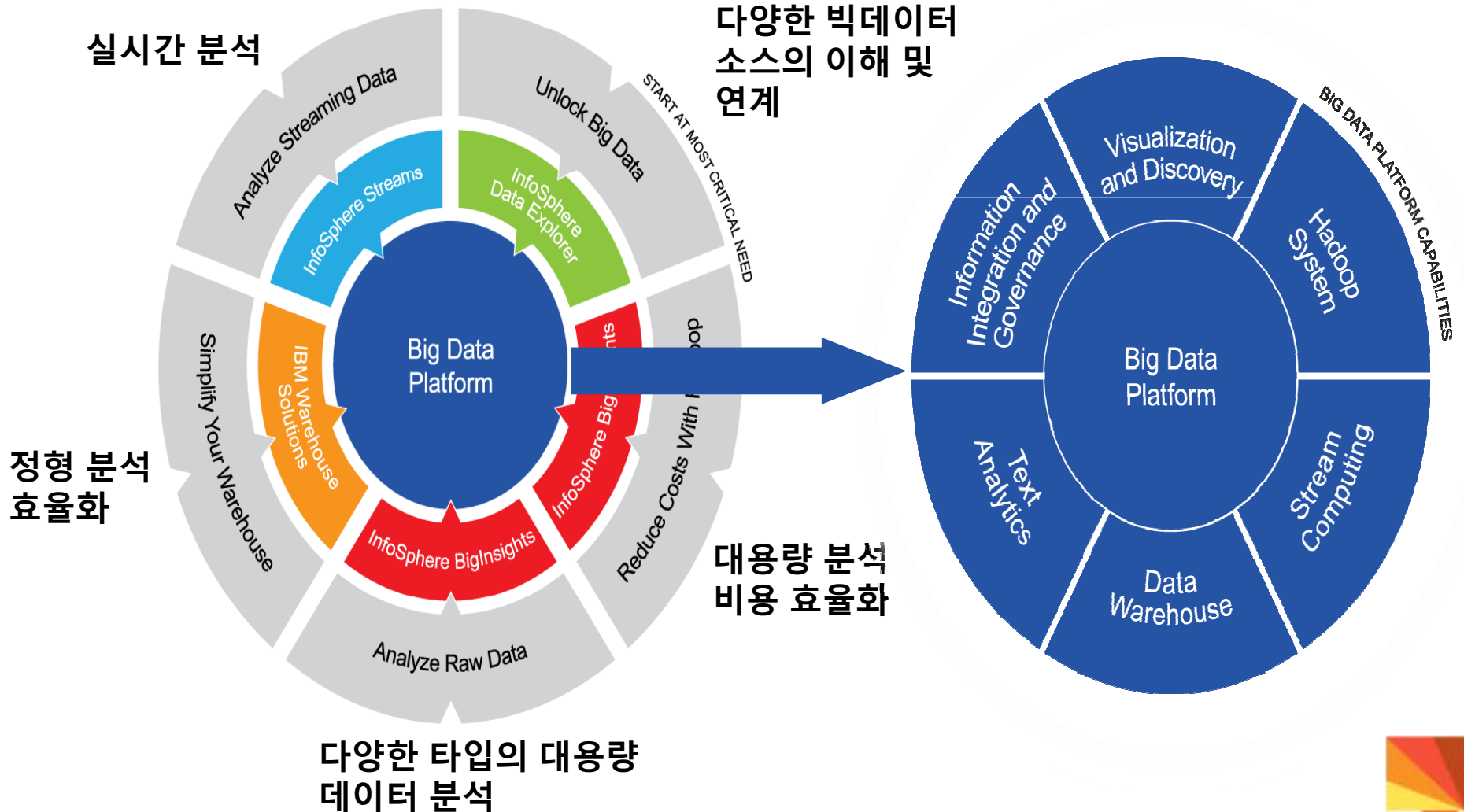
- 대용량 데이터 Ad-hoc 분석
- 데이터 발견(discovery & experimentation)

빅 데이터 관리

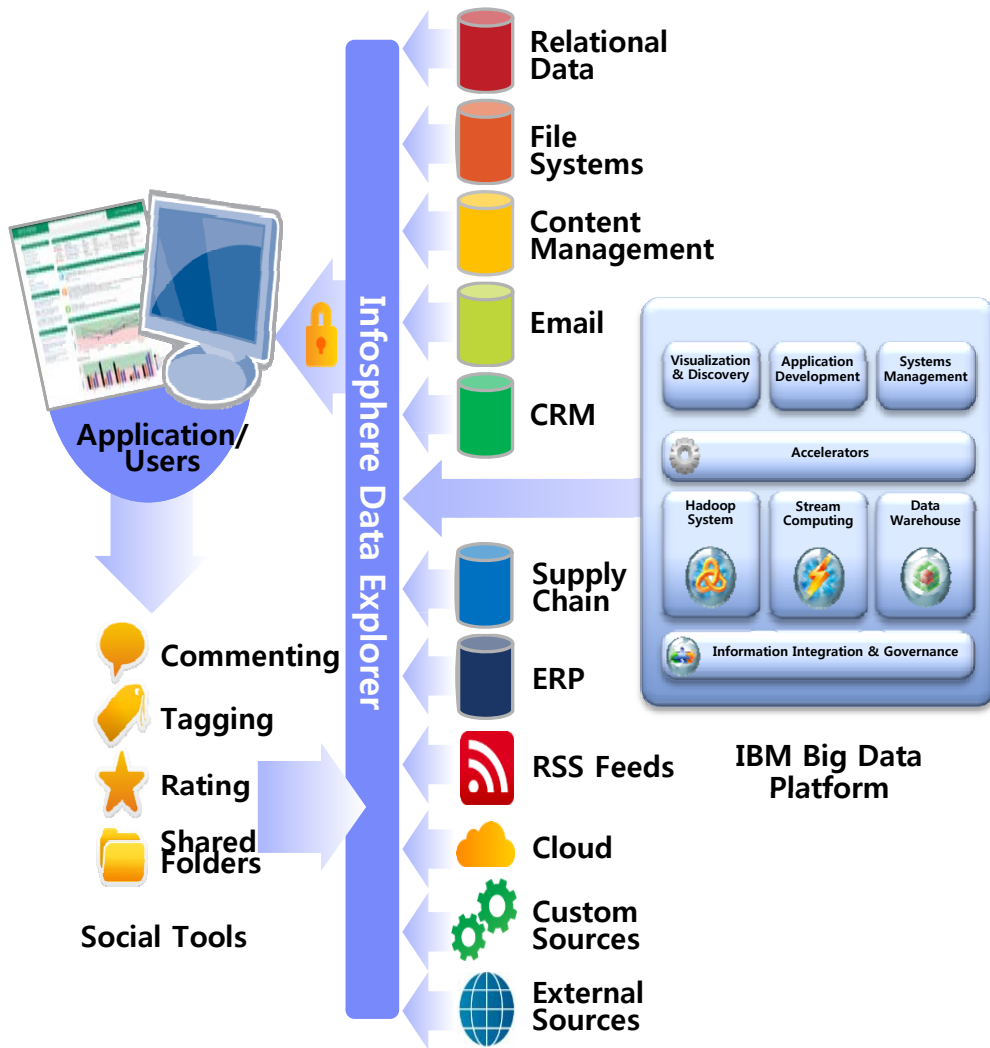
- 데이터 구조, 통합, 정합성 관리
- 반복적인 조회에 대한 일관성 유지



비즈니스 중심의 빅데이터 활용 프로젝트



빅데이터 소스의 이해 및 시각화 > InfoSphere Data Explorer



-다양한 어플리케이션 및 데이터 저장소의 통합

정보 액세스 보장

-기업 전반 데이터 소스 기반의 발견 및 검색,

네비게이션 가능

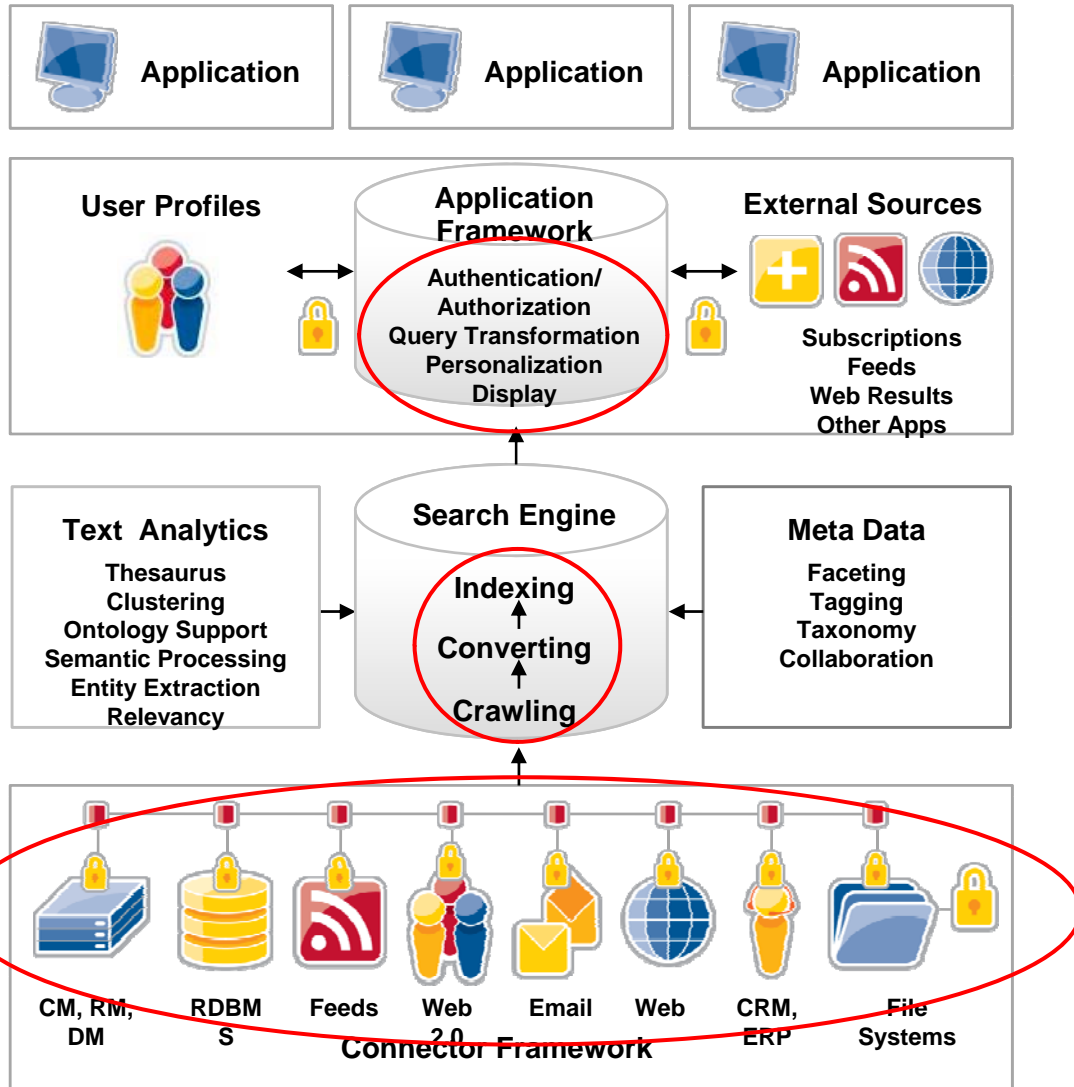
-정형과 비정형의 통합

-협업을 위한 소셜 툴

- ✓ 정보의 사용 및 재사용 효율화
- ✓ 좀 더 좋은 의사결정
- ✓ 효율적 비즈니스 운영
- ✓ 고객 관계 효율화
- ✓ 리스크 감소와 법규 준수 향상



빅데이터 소스의 이해 및 시각화 > InfoSphere Data Explorer



✓ 분석 Scalability

- 방대한 양의 데이터 분석, 기업 인프라 전반의 다각도의 탄력 있는 영향도 분석
- 다양한 빅데이터 자산 분석 : - 정형 데이터, 비정형 데이터, Social Media, Web Content, 기업 어플리케이션 (Siebel, SAP, SharePoint) 등...

✓ Security

- 빅데이터 거버넌스 모델과 연계 (LBAC Security / LDAP Authentication)
- 사용자별 권한 부여 및 분석 결과 제한

✓ Accuracy

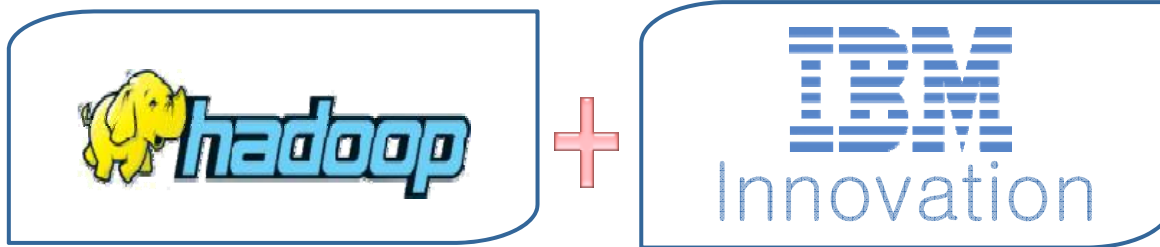
- 빅데이터 에 대한 높은 수준의 정합성 & 정확성 있는 분석 결과 제공
- 타 솔루션 대비 유일한 index technology 기술 적용

✓ Integration

- "virtual" single repository 기반의 빅데이터 분석
- 다양한 종류(1,000 여개) : CRMs, ERPs, ECMs, Web Content, Twitter, Facebook 등...



Hadoop 기반의 대용량 데이터 저장 및 분석 > InfoSphere Biginsights



User Interfaces

- 시각화 (Visualization)
- 개발 툴 (Development Tools)
- 관리 콘솔 (Management Console)

Accelerators

- 텍스트 분석 (Text Analysis)
- 어플리케이션 액셀레이터 (Application Accelerator)

BigInsights Engine

- Map Reduce + 인덱싱 (Map Reduce + Indexing)
- 워크로드 관리 (Workload Management)
- 보안 (Security)
- Apache Hadoop

Integration

- Databases
- Content Management
- Information Governance

✓ **성능 & 워크로드 최적화**
 - Adaptive MapReduce, Compression, Indexing, Flexible Scheduler

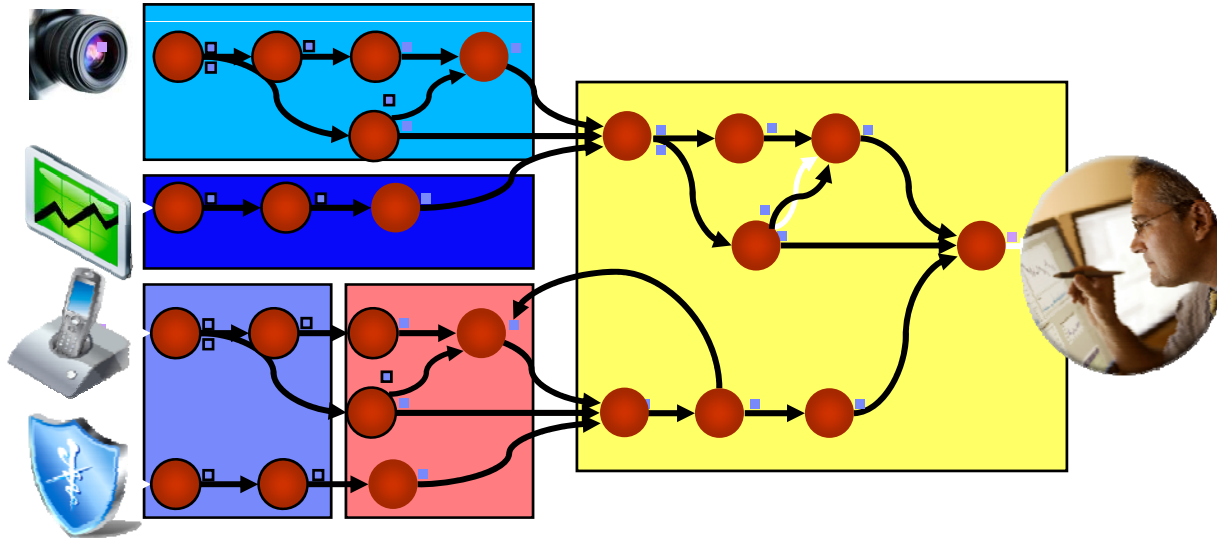
✓ **분석 역량 강화**
 - 데이터 검색 및 탐구를 위한 Spreadsheet
 형식의 시각화 툴 Big Sheets, 텍스트 분석 엔진, 분석 Accelerator

✓ **Enterprise 환경을 위한 기능 개선**
 - Role 기반의 보안, 타 시스템과의 연계를 위한 Connector

✓ **쉬운 관리 및 개발 환경**
 - 웹 관리 콘솔, 클러스터/시스템/Job 모니터링, Eclipse 개발 환경, 쉬운 개발을 위한 Jaql 언어



스트림 컴퓨팅을 통한 실시간 분석 > InfoSphere Streams



✓ 높은 데이터 전송 성능

- 매우 낮은 지연시간, 대용량 데이터 처리

✓ 다양한 데이터 유형 지원

- 어떠한 형태의 데이터라도 처리가능, 전통적으로 처리하기 어렵거나 빠른 응답을 요구하는 데이터도 처리

✓ Operator 분산처리

- 효율적인 CPU Core 사용, 초고속 데이터 교환

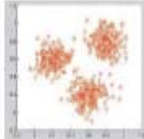
✓ 손쉬운 관리 및 개발

- 자동 Operator 수행 배치, 다운타임 없이

Application 추가 가능, 빌트인 어댑터, 운영 중 분석 로직 추가 및 변경

데이터 마이닝

(In Streams)

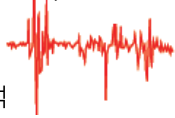


텍스트 분석
(In Streams)

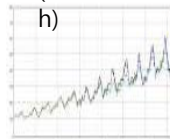
Text
(listen, verb),
(radio, noun)

음향 분석

(Research)



예측 분석
(Research)



통계 분석

(In Streams)

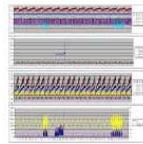
$$\sum_{population} R(s_i, a_i)$$

지리 데이터
(Research)



고급 산술 모델

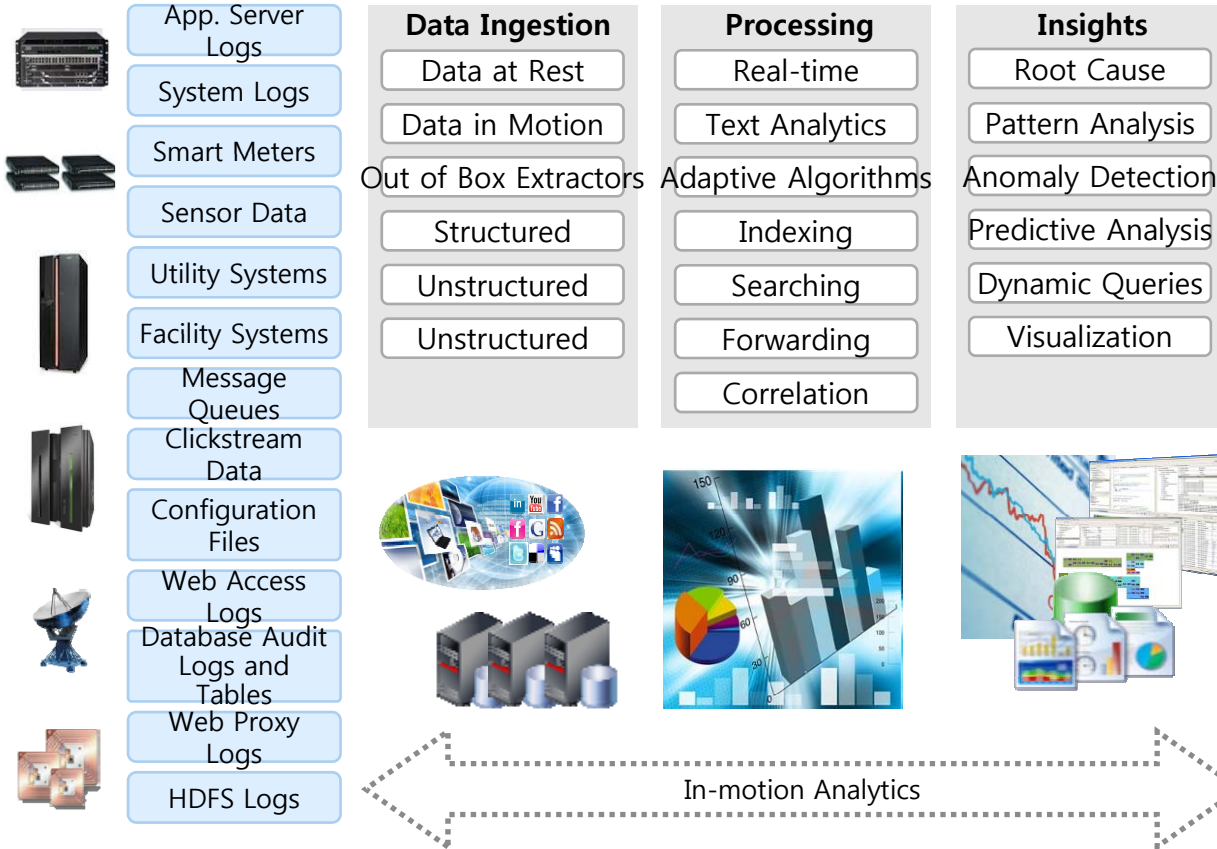
(Research)



이미지, 비디오
(Oper 연계)



스트림 컴퓨팅을 통한 실시간 분석 > InfoSphere Streams



✓ 다양한 데이터 소스

- 비정형 포맷
- 대용량 machine data의 분석 저장

✓ 실시간 수집/처리

- micro-second 수준의 데이터 수집 및 처리 결과 응답
- Terabyte의 초당 처리 volume

✓ 분석

- 대용량 인덱싱
- 다른 데이터 유형 간 연관/고급 분석
- Root cause 분석

✓ Visualization

- 실시간 대시보드



정형 분석 최적화 > PureData System for Analytics

PureData
System for Analytics

분석을 위한 데이터 서비스 제공

✓속도

- 기존 맞춤형 시스템 대비 10-100배 빠른 속도
- 전술적 쿼리의 경우 20배 향상된 동시성과 처리량
- 특허받은 하드웨어 가속화 기술

✓간소화

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

✓확장성

- 페타급 데이터 사용자 용량

✓지능성

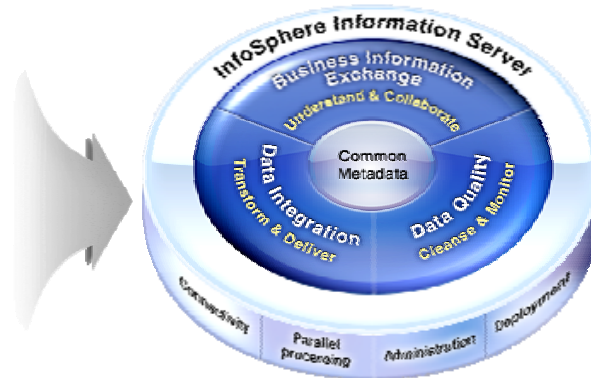
- 몇 분 이내에 복잡한 분석을 실행할 수 있도록 설계
- 가장 강력한 In-Database 분석



신뢰성 있는 정보생성을 위한 데이터 통합 > InfoSphere Information Server

정보 통합 플랫폼을 통하여 단일 환경하에서 정보 통합 프로세스 자체 뿐만 아니라, 데이터 품질 측정 및 표준화를 통한 정보의 신뢰성 확보

시스템 구성 요소별 과도한 중복
 - Data
 - Applications
 - Infrastructure (서버 및 디스크)
 단일한 전사 데이터의 부재
 비체계적인 데이터 통합 작업
 비효율적 데이터 통합에 의한 비용 상승



- ✓ 정보의 적시 제공
- ✓ 생산성 증대
- ✓ TCO 절감

비즈니스 정보 교환

정보에 대한 이해 및 협업

- Information blueprints
- 데이터 소스간 연관관계 분석
- IT와 Business간 맵핑

데이터 품질

정제 및 모니터링

- 분석 및 검증
- 데이터 정제
- 데이터 품질 규칙 생성 및 관리

데이터 통합

변환 및 전달

- 확장성
- 복잡한 로직 구현
- 추적 관리
- 실시간 데이터 캡처
- 데이터 전송의 다양화
- 빅데이터 처리



빅데이터 플랫폼 Use Case

통합된 고객 정보 관리

- 소셜 미디어 고객 감성 분석
- 프로모션 최적화
- 고객 세분화
- 고객 수익성 분석
- Click-stream 분석
- CDR 프로세싱
- 다중 채널 상호작용 분석
- 로열티 프로그램 분석
- 고객 이탈 방지



신속하고 폭넓은 신상품 도입 연구

- 소셜 미디어 - 상품/브랜드 감성 분석
- 브랜드 전략
- 마켓 분석
- RFID 트래킹 및 분석
- 상품 및 서비스 오퍼링 트랜잭션 분석



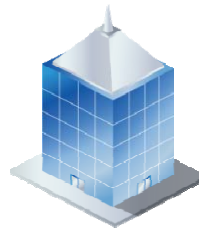
실시간 운영

- 스마트 그리드/미터 관리
- 분산 부하 예측
- 영업 리포팅
- 재고 및 판매 최적화
- 선물 옵션 트레이딩
- ICU 환자 모니터링
- 재난 감시
- 교통망 최적화
- 스토어 성능
- 환경 분석
- 실험용 연구



실시간 사기 및 리스크 감지

- 감시 다중모델 적용
- 사이버 보안
- 사기 모델링 및 감지
- 리스크 모델링 및 감지
- 취약성 리포팅



기계적 자산 활용

- 네트워크 분석
- 기계적 자산 관리와 이슈 예측 해결
- 웹사이트 분석
- IT 로그 분석



빅데이터 플랫폼 적용 사례

	필수 요건	주요 적용 제품	비즈니스 가치
	Run Zero Latency Operations	InfoSphere Data Explorer	30개 이상의 데이터 저장소의 통합 시각화를 통하여 비즈니스의 능률적 운영 및 지원
	Run Zero Latency Operations	InfoSphere BigInsights	효율적 turbine 설치를 통하여 운영 비용 감소 및 차별성 확보
	Know Everything about your Customers	InfoSphere Streams	실시간 프로모션 제공을 위한 실시간 통화 분석으로 이윤 증가 및 고객 이탈 감소
	Exploit Instrumented Assets	IBM Netezza	네트워크 장비의 문제점 확인 및 수정을 통하여 네트워크 가용성 증가
	Instant Awareness of Risk and Fraud	IBM Netezza	2PB 의 분석 시간이 26시간에서 2분으로 감소

IBM 빅데이터 플랫폼 Advantage



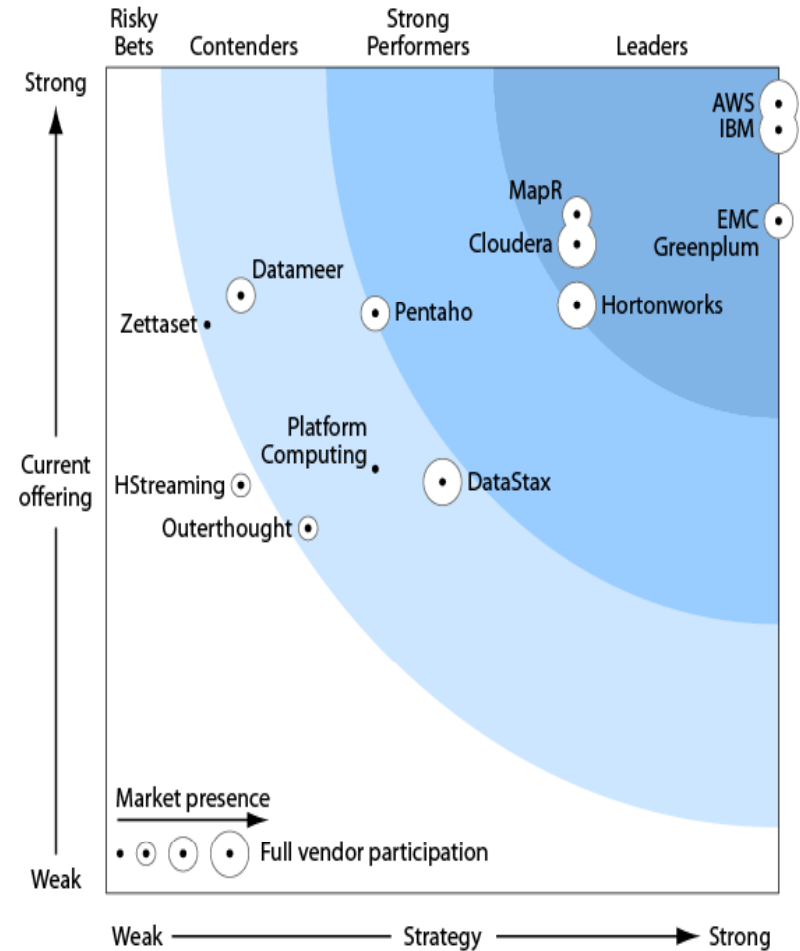
- ✓ 초기 빅데이터 프로젝트로 후 Next 프로젝트로의 확장 시 이점 확보
- ✓ 컴포넌트를 공유하고 각 시스템간의 통합을 통하여 도입 및 개발 비용 감소
- ✓ 또 다른 기술적 이점
 - 스트리밍과 하둡 간의 텍스트 분석의 재사용
 - 빅데이터 플랫폼 내의 정보 통합
 - 메타 데이터와 통합된 거버넌스
 - 각 엔진을 아우르는 Accelerator 사용 – common analytics, models, and visualization



IBM 빅데이터 플랫폼 외부기관 평가

		IBM	ORCL	SAP	EMC	TD
The Forrester Wave™	Enterprise Data Warehousing Platforms	●	●	●	●	●
	Enterprise Hadoop Solutions	●	○	○	●	○
	Database Auditing And Real-Time Protection	●	◐	○	○	○
	Enterprise ETL	●	●	●	○	○
Gartner Magic Quadrant	Data Integration Tools	●	●	●	○	○
	Data Quality Tools	●	◐	●	○	○
	Data Warehouse Database Management Systems	●	●	●	●	●
	Master Data Management of Customer Data Solutions	●	●	◐	○	○
	Master Data Management of Product Data Solutions	●	●	◐	○	○
	Enterprise Information Archiving	●	○	○	◐	○

Source : Forrester Waves, Gartner Magic Quadrants



Source: February 2012 "The Forrester Wave™: Enterprise Hadoop Solutions, Q1 2012"



THINK

BIG