

최근 DW/BI 프로젝트 Trend 및 대응방안

2011.10.06

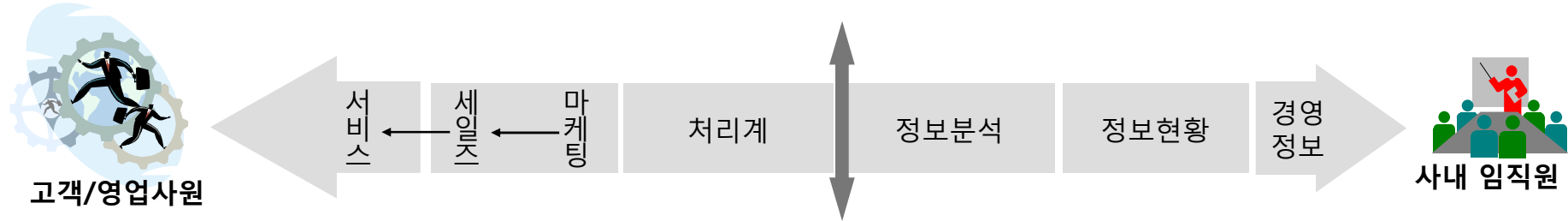
(주)데이터스트림즈

발표 : 박시영 이사



DW/BI 시스템 등장 배경

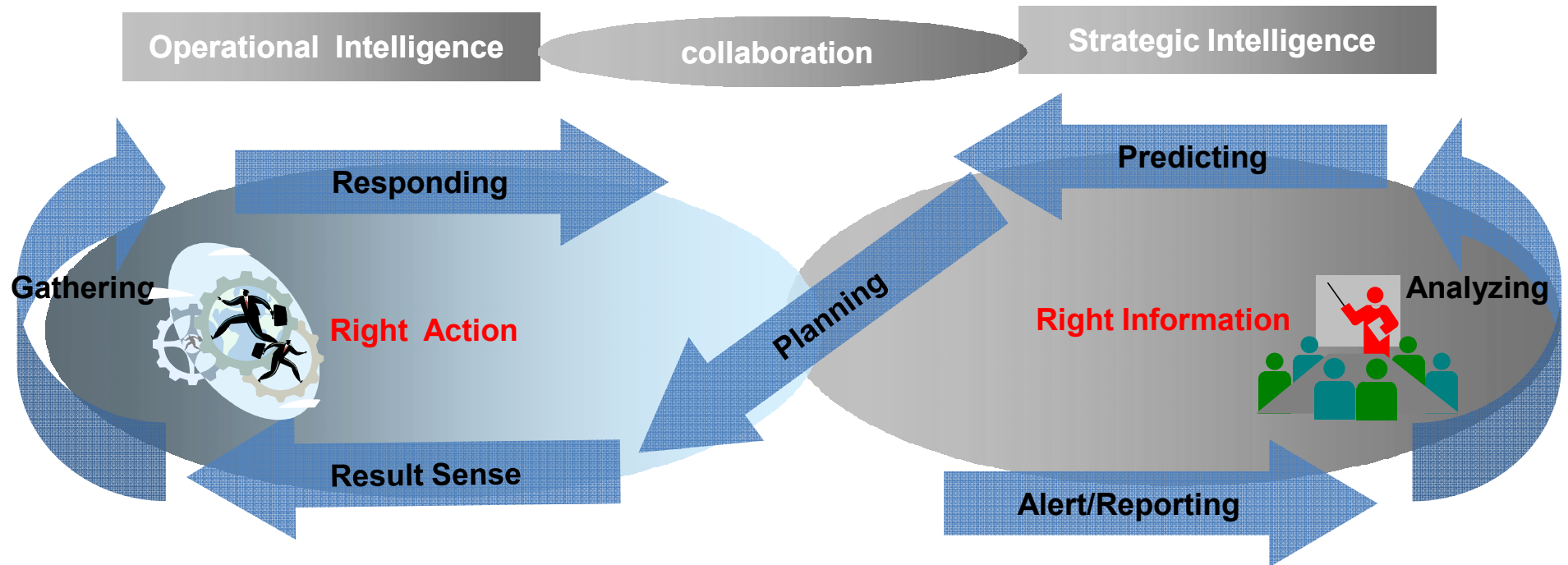
❖ 과거 **Computer**의 역할은 업무 처리를 지원하는 프로세스개선과 시간절감의 가치를 지녔으나 이후 데이터의 축적으로 정보활용 측면의 가치가 새롭게 부각되면서 정보분석 관점의 새로운 아키텍처가 개발됨



구분	사무 처리계(OLTP)	정보 분석계(OLAP)
목적	영업활동 Process 지원 중심 - 마케팅, 세일즈, 서비스 제공 - 비즈니스 운영("What"에 초점)	의사결정지원 data 관계 중심 - 데이터 분석 및 통찰력 - 비즈니스 방향설정("Why"에 초점)
활용영역 (예)	전략결과 실행 - 캠페인 실행 - 세일즈, 서비스 실행	전략수립 결정 - 캠페인 기획 - 고객행동 예측(Data Mining)
주 사용자	실무자, 영업사원	임직원, 분석가
정보처리	영업활동 관점: 어플리케이션(업무영역)별 데이터 집합 - 데이터 생성 원천 - 자금, 마감 처리	전사적 관점: 주제영역별 데이터 집합 - 데이터 결과 분석 - 데이터 가공 처리
정보축적	거래처리중심의 데이터 처리, 관리 - 현재 까지(최신 또는 현행 데이터) - 실시간 데이터 처리	주제영역중심의 데이터 통합 - 시계열 : 미래 데이터 생성(이력 데이터) - D-1일 기준(필요시 실시간 분석가능)

BI (Business Intelligence) 개념

❖ BI(Business Intelligence)는 데이터를 기반으로 다양한 정보 분석을 통해 의사결정과 관련된 **Business process**를 접목하여 필요로 하는 사용자(**right people**)에게 필요한 정보(**right Information**)를 적시에 제공하여(**right time**) 보다 고도화된 서비스 (**right action**)를 지원해야 함

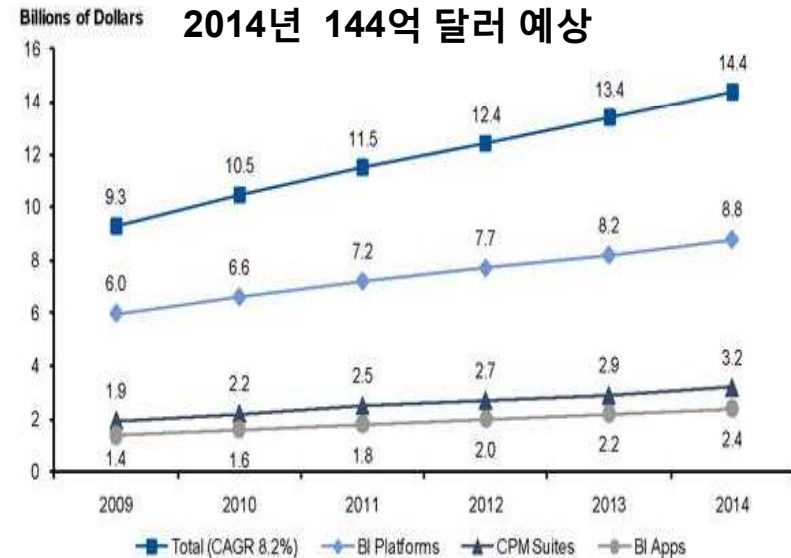


❖ 최근 BI시장의 통계를 보면 점점더 고객은 자신의 업무 프로세스와 BI기술을 접목하기를 기대하고 있으며 6~50명 정도의 BI사용자 WorkGroup 규모의 요구가 증대되고 BI부문에 대한 예산을 계속 확대할 전망으로 나타남

◆ BI Market

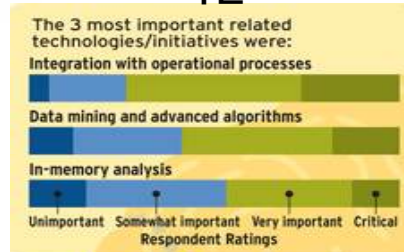


2014 BI, Analytics and Performance Management Revenue, Worldwide, 2009-



Source: Gartner (May 2011)

기술



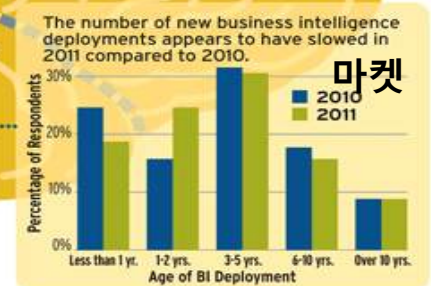
People



예산



Workgroups

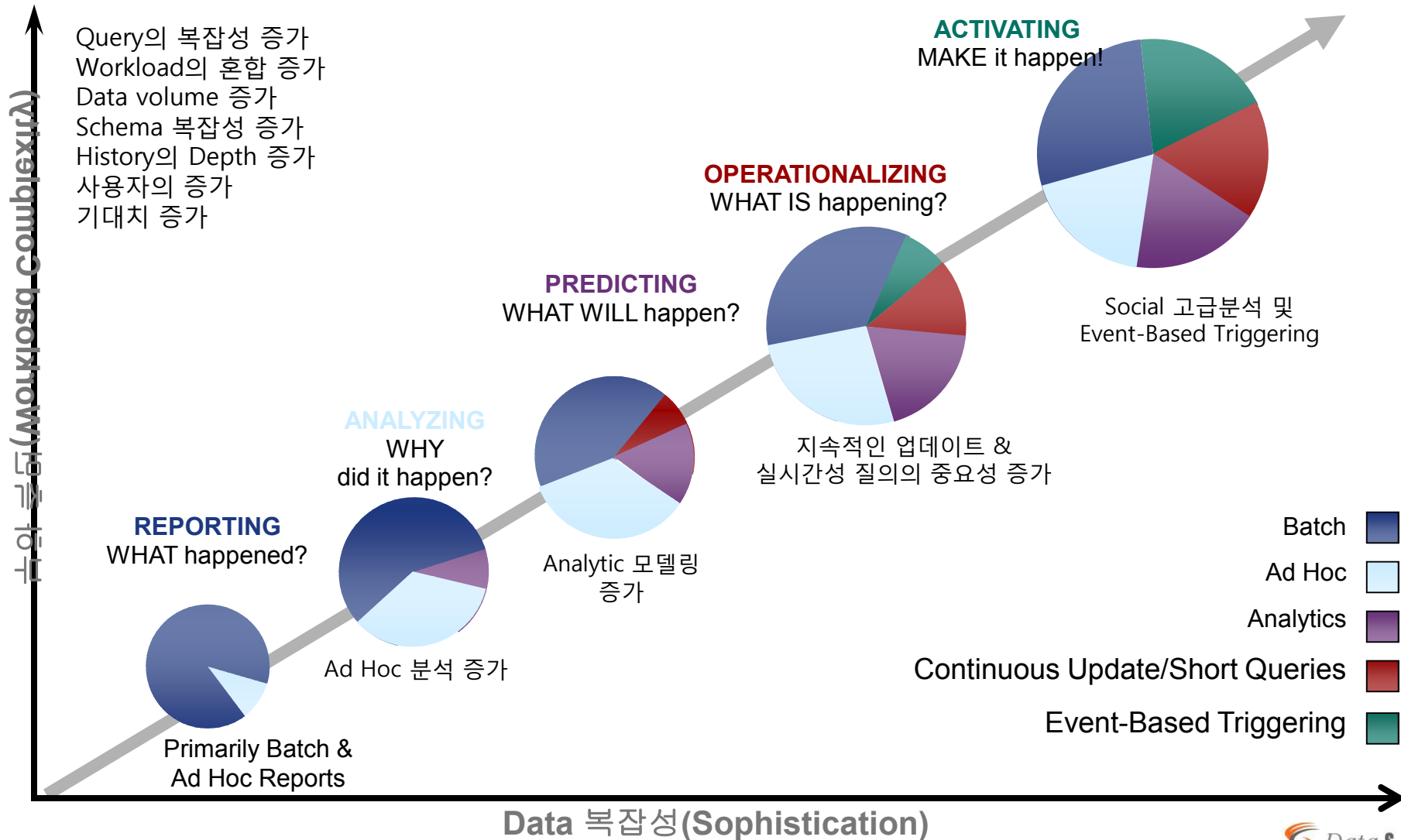


마켓

Source: Wisdom of Crowds Business Intelligence Market Study, Dresner Advisory Services, LLC 2011

DW/BI 진화 방향

❖ DW/BI는 점점더 대용량화되고 분석의 깊이, 데이터구조의 복잡성이 증대되는 반면 점점 더 배치에서 실시간에 가까운 비즈니스 반영을 요구하고 있는 추세로 발전하고 있음



❖ 국내 DW/BI시장은 이미 기구축된 대형 사이트를 중심으로 재구축 고도화 하는 추세이며 모바일, 규제대응, 분석요구수준 증대, 데이터량의 증가, BI 솔루션 기술의 진보를 고도화 요건으로 반영하고 있다.

시장 환경의 변화

- 계정계 중심의 차세대 완료에 따라 정보계 시스템 부분에 비중
- 금융지주회사 체제 전환에 따른 통합 정보 전략
- 자본시장통합법 시행 및 규제(Compliance) 강화
- 모바일 환경의 급격한 확대

데이터의 변화

- 데이터의 폭증 (S카드/H카드 차세대 BI 규모 170TB (압축) 예상)
- SNS 가 기업의 마케팅 전략에 흡수
- 과거 의미없던 데이터의 저장 및 분석 요구 증가
- Big Data (SNS, 이미지, 동영상, Log분석, RFID, GIS)

EDW 고도화

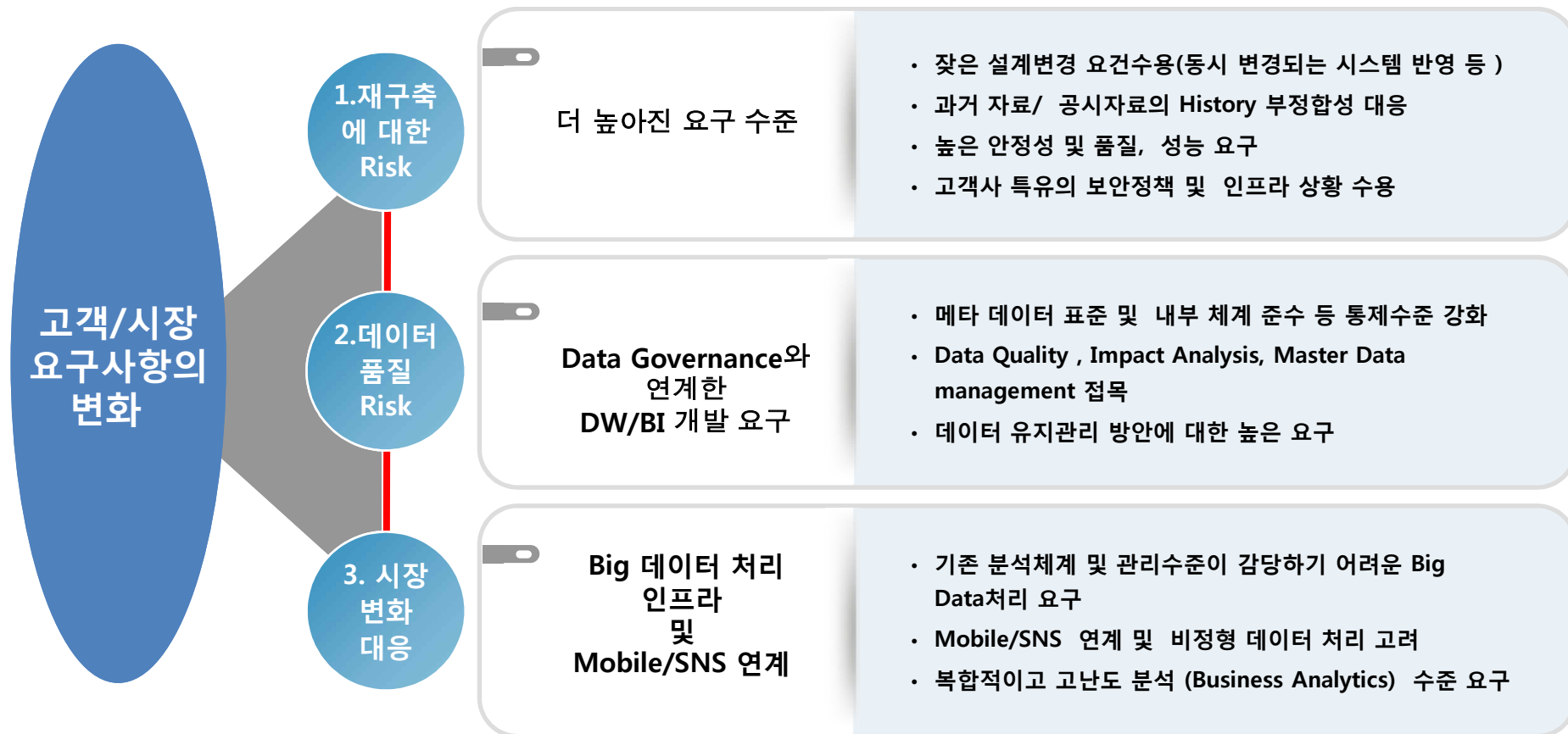
분석 환경의 변화

- 고급 분석 (Advanced Analytics) 필요성 증가
- 효율적 리포팅 도구로서의 사용자 편의성 및 시각화
중점에서 데이터 분석에 의한 과거와 현재
비즈니스를 심도있게 비교하고 미래 상황을
예측하여 진화된 의사 결정 지원

DW 솔루션의 변화

- DW 어플라이언스 등장
Vertica, Teradata, GreenPlum, Exadata, Netezza
- 서버, 스토리지 및 DBMS의 통합을 최적화하여 성능
극대화
- Collabolation 기능 확대로 메타데이터 증가

❖ 최근 과거의 DW/BI를 재구축하는 프로젝트가 많아지고 있으며 고객은 한번 DW/BI의 실패를 경험하거나 운영 경험을 토대로 높은 기대수준을 안고 재구축 사업계획을 추진하고 있으며 더불어 IT환경의 급격한 변화는 수행입장의 고민을 더욱 가중시키고 있다.

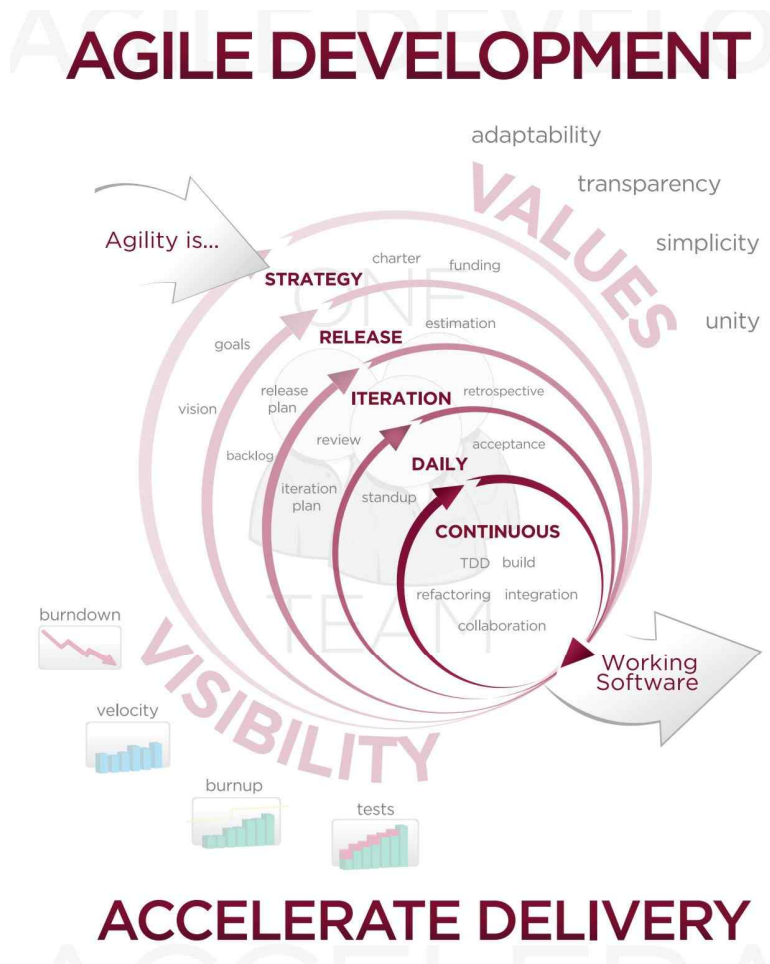


Issue-1. 재구축 Risk 대응

❖ 재구축 리스크는 이미 BI프로젝트와 사용문제점을 경험한 고객의 요구수준이 점점 높아진 상태에서 단순한 BI개발방식으로는 문제를 해결하기가 어려우며 좀더 Customized된 수행기법과 향상된 UI통합기술 등 프로젝트 특성을 감안하여야 함

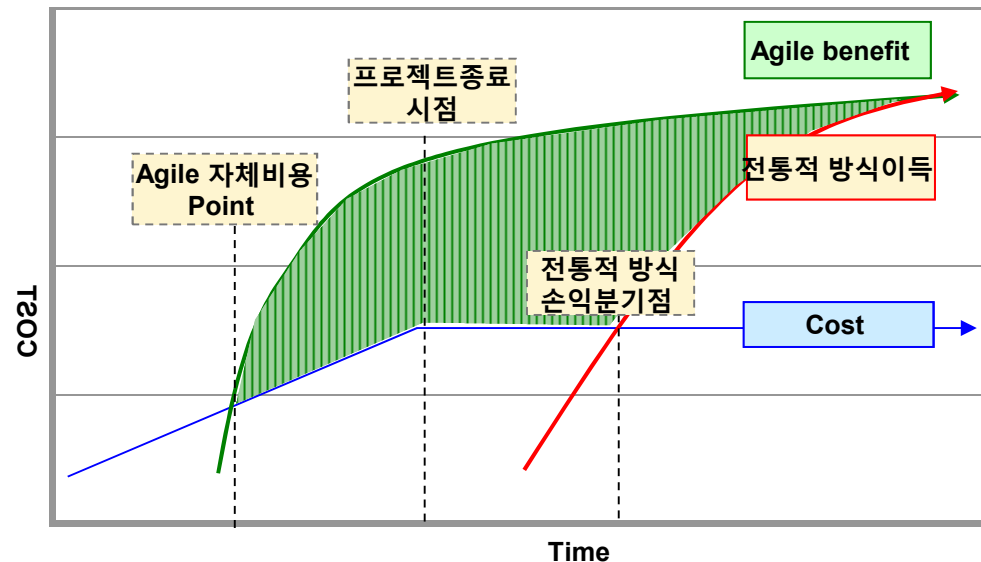
	Risk	대안
재구축 Risk 대응	동시 변경되는 소스시스템을 고려한 개발 형태	<ul style="list-style-type: none">• 개발 중인 소스(원천)시스템의 정보획득 경로를 프로젝트 초기 혹은 사전 확보• 오픈 일정에 대한 단계별 오픈 방식 고려
	과거 자료/ 공시자료의 History 부정합성 대응	<ul style="list-style-type: none">• 분석관점을 현재시점과 과거시점 모두를 충족하는 단일 History 데이터 모델 생성은 데이터량 및 수행 로드를 감안 특정 중요 테이블에 한정
	높은 안정성 및 ETL 성능 요구	<ul style="list-style-type: none">• 아키텍처에 이중화를 HW레벨뿐 아니라 실 작업 상황 고려한 Software적 설계 고려• Tuning 전문 엔지니어를 반드시 참여
	고객사 특유의 보안정책 및 인프라 상황 고려	<ul style="list-style-type: none">• 고객사 특유의 보안 정책상 구현 문제점을 사전 확인• 운영인프라 개발인프라, 이행 테스트 인프라에 대한 고려

❖ Agile DW/BI 개발 방법론은 최근 대두되고 있는 방법이기도 하나 근본은 과거 RAD/JAD 방법론에 근간한 프로토타입 및 인수 테스트를 결합한 방법론을 띄고 있으며 IT 개발자 기반 Cloud Computing 서비스와 맞물려 최근의 BI 개발 해법으로 등장하고있다.



주요 특징

- Processes and tools → **Individuals and interactions**
- Comprehensive documentation → **Working software**
- Contract negotiation → **Customer collaboration**
- Following a plan → **Responding to change**



Traditional Waterfall vs Agile method

❖ Agile DW/BI 개발 방법론은 기존의 전통적 분석-설계-개발-테스트-이행 등의 순서로 진행되는 폭포수 Waterfall 방식에 비해 다음과 같은 12가지 특징과 차별점을 갖고 있다.

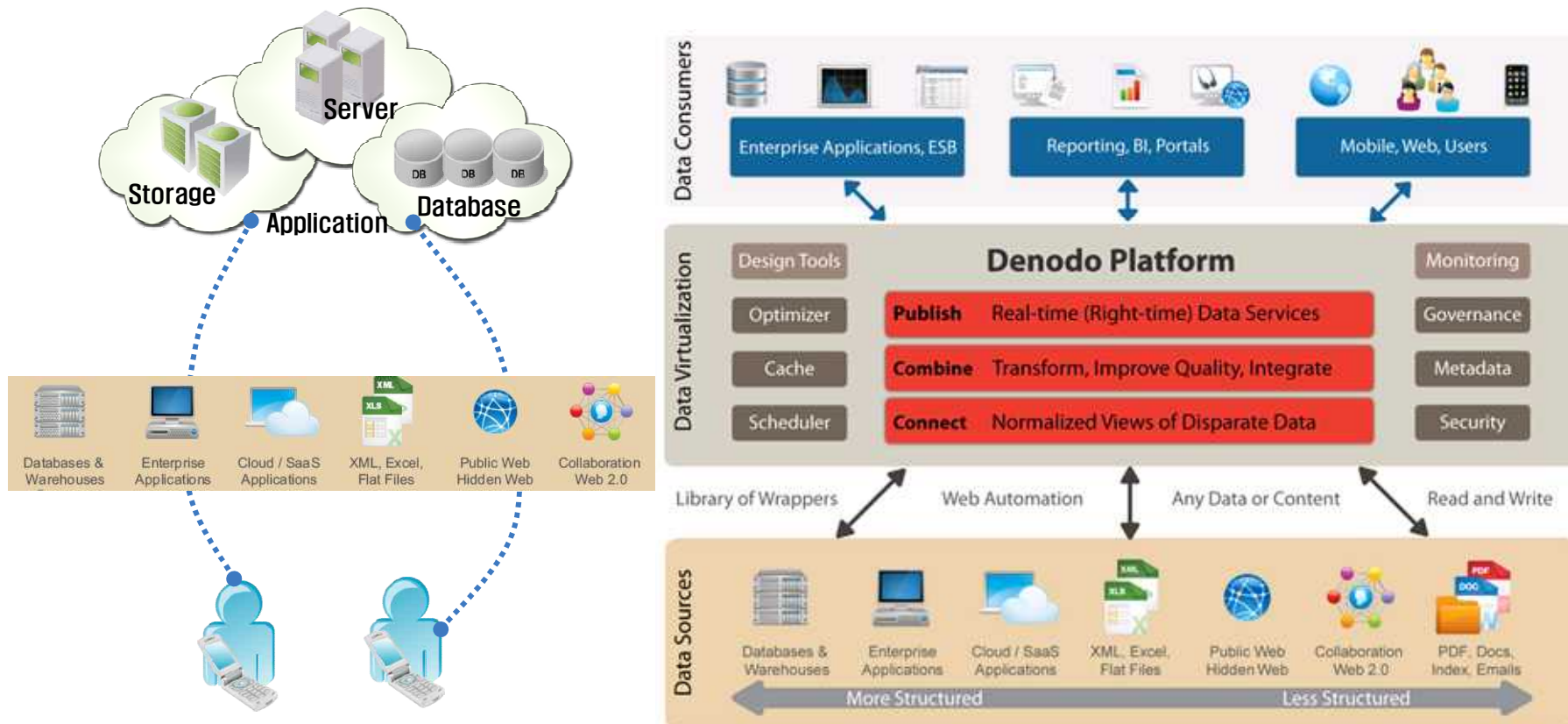
Twelve principles underlie the Agile Manifesto,

1. Customer satisfaction by rapid delivery of useful software : 신속한 Output
2. Welcome changing requirements, even late in development : 즉각적 변경 요구 수용
3. Working software is delivered frequently (weeks rather than months) : 짧은 주기의 feedback
4. Working software is the principal measure of progress : 진척도의 관리기준은 Software (not 문서)
5. Sustainable development, able to maintain a constant pace : 일정한 개발 속도 유지 및 지속적 개발
6. Close, daily co-operation between business people and developers : 사용자와 개발자의 일일 협업
7. Face-to-face conversation is the best form of communication (co-location) : 같은 공간에서 얼굴 대화
8. Projects are built around motivated individuals, who should be trusted : 신뢰와 동기부여된 팀원
9. Continuous attention to technical excellence and good design : 기술적 우수성과 좋은 설계에 지속적인 관심
10. Simplicity : 단순함
11. Self-organizing teams : 자발적 팀구성
12. Regular adaptation to changing circumstances : 변화상황에 상시 적응

Agile 과 Platform-as-a-Service

❖ Salesforce.com denodo 등 클라우드 컴퓨팅기반의 SaaS, PaaS 서비스 업체의 등장과 Agile DW/BI 방법론이 부각되고 있으며 이는 많은 제도적인 변화와 개발 형태의 변경을 요구하고 있으며 우리나라 IT업체도 준비가 필요한 부문임

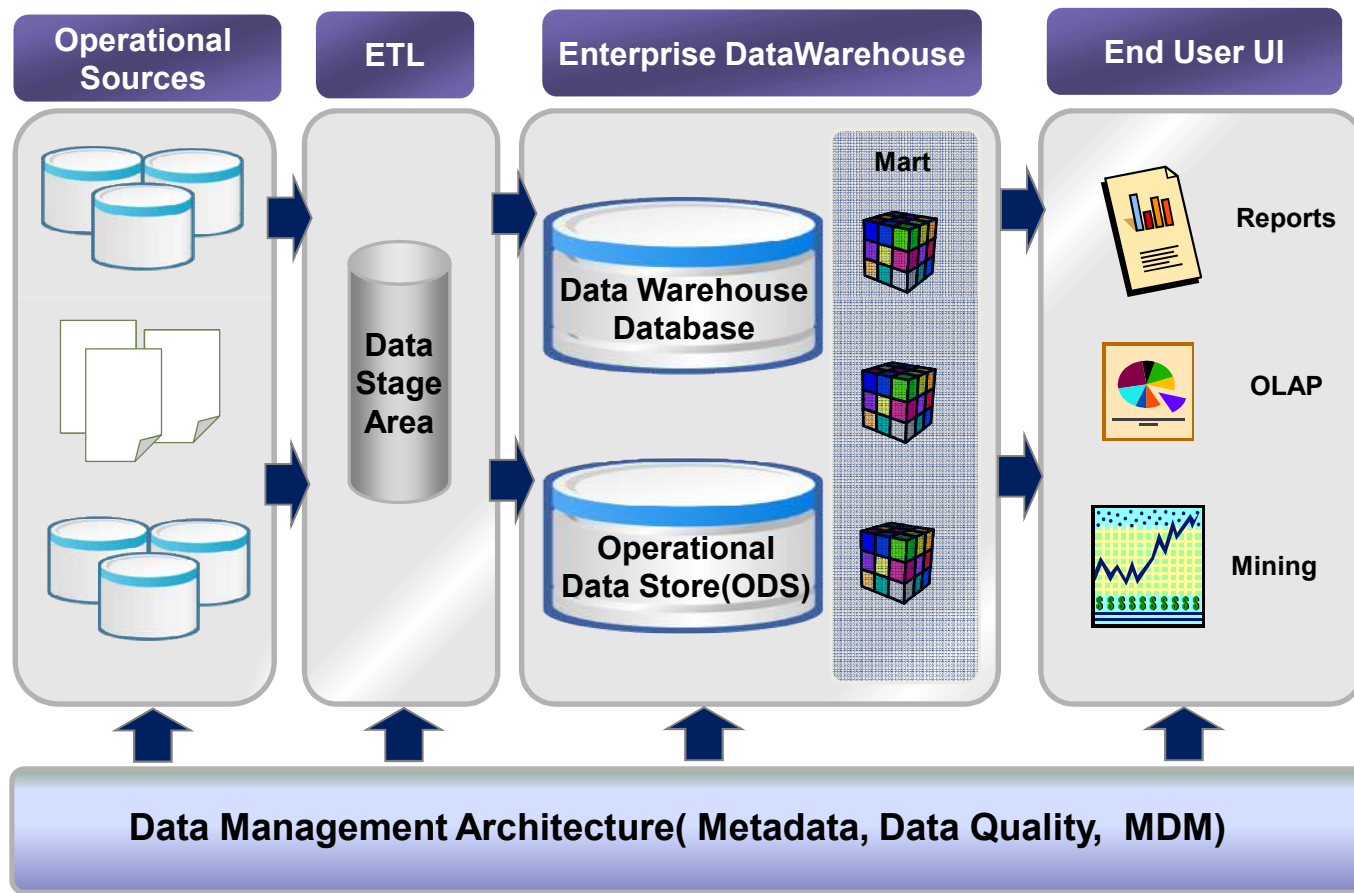
“Data Virtualization has doubled the number of BI projects we completed on time; as a DataWarehouse extension Denodo reduces cost and enriches BI with new data across internal and external systems.”



Issue-2. 데이터거버넌스와 BI접목

❖ 데이터 표준, 메타, 품질 기반의 데이터 거버넌스 아키텍처를 BI와 접목하여 통합 리포지토리 환경으로 구축하는 것이 BI 신뢰도와 효과를 배가시키는 방안임

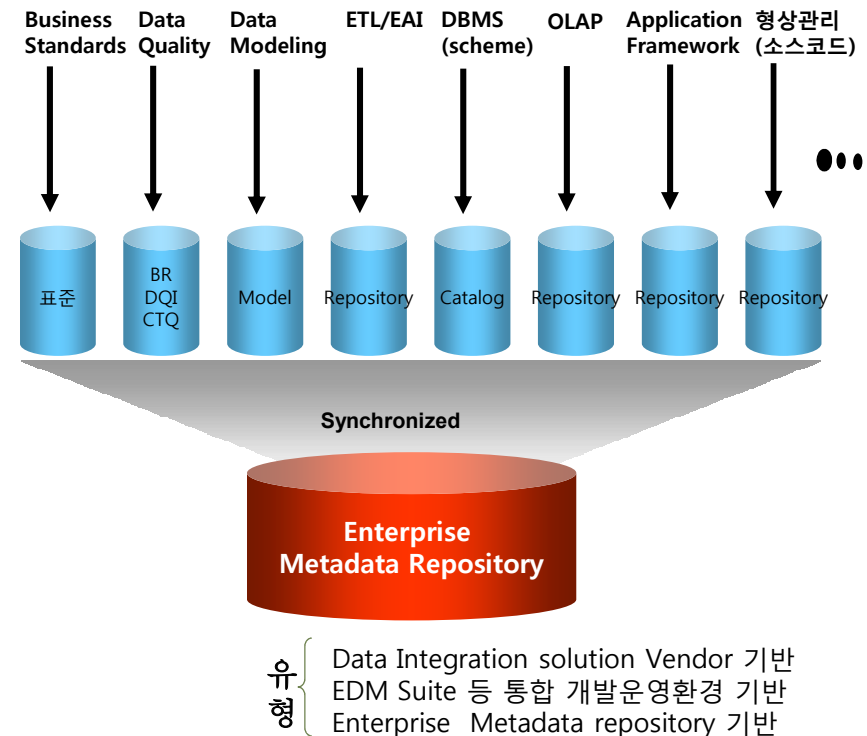
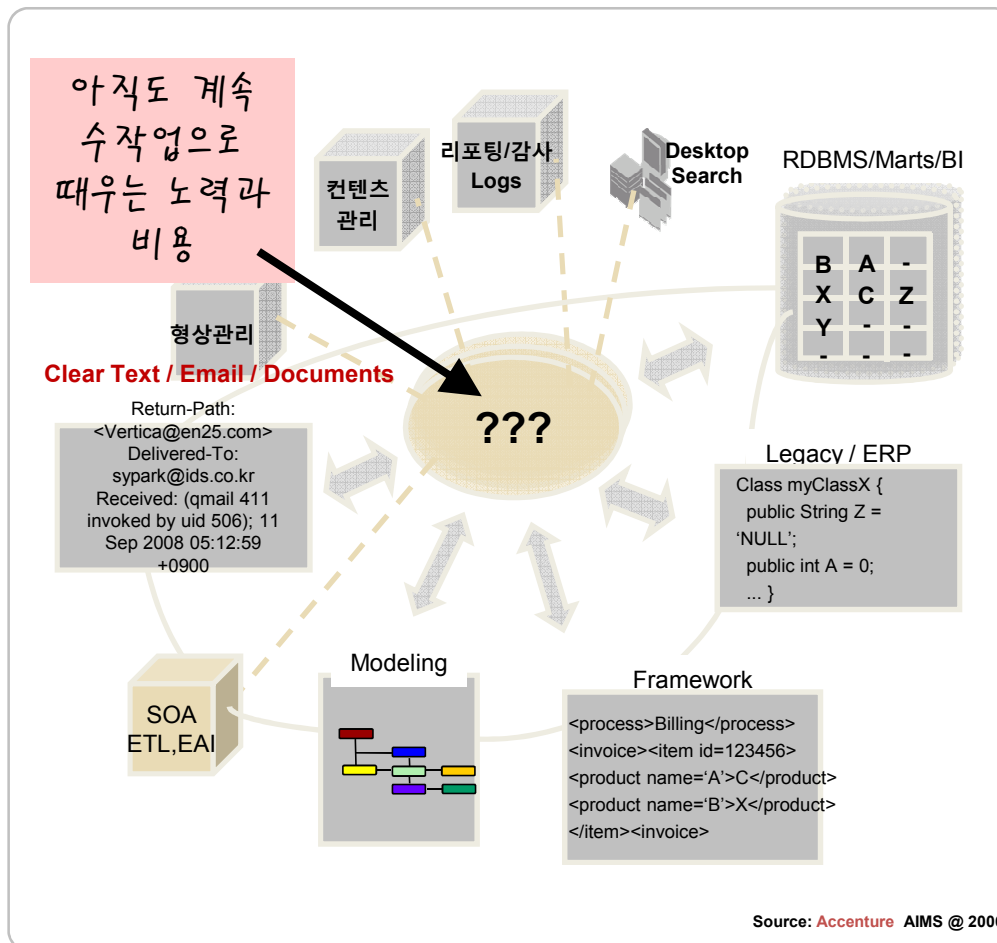
데이터
품질 Risk
대응



- DMA 와 BI의 결합 효과**
- 사용자 신뢰도/만족도 향상
 - 데이터 정확도 및 품질향상
 - Application 형상관리
 - 유연한 데이터 구조 변화관리
 - 시스템간 통합 용이성
 - 적정 성능 유지관리 용이
 - 스케줄링 모니터링 용이
 - 보안성 강화

Meta 관점의 BI 통합고려

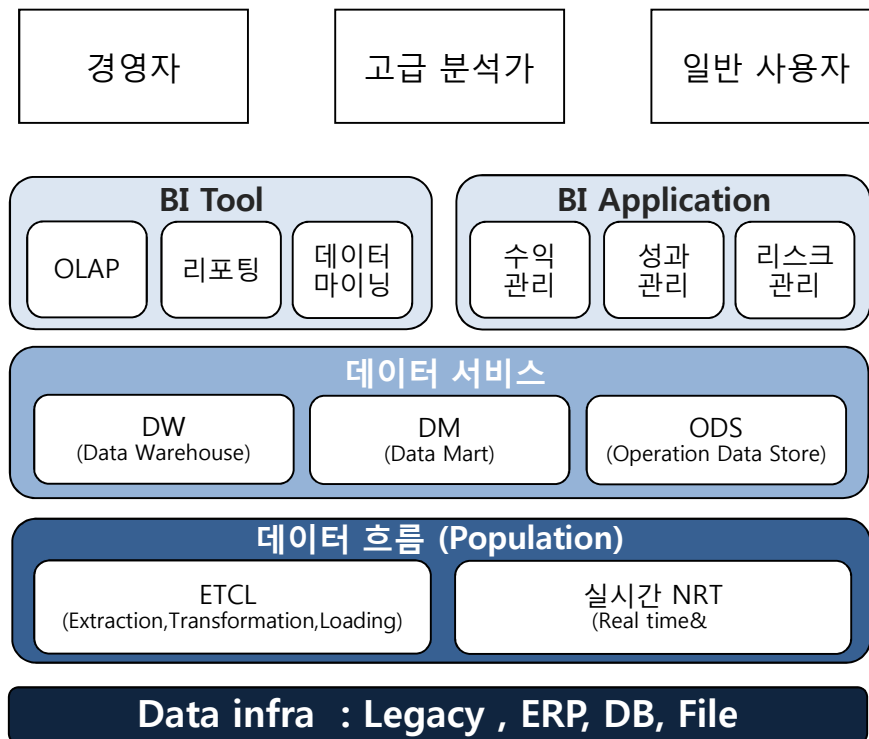
❖ 데이터 품질 리스크를 해소 하기위한 방안으로 Meta관점의 DW/BI 구축 접근이 필요하며 이는 데이터 표준,품질,흐름 등 모든 데이터의 구조와 과정, value에 대한 Single Repository 화를 목표로 한다.



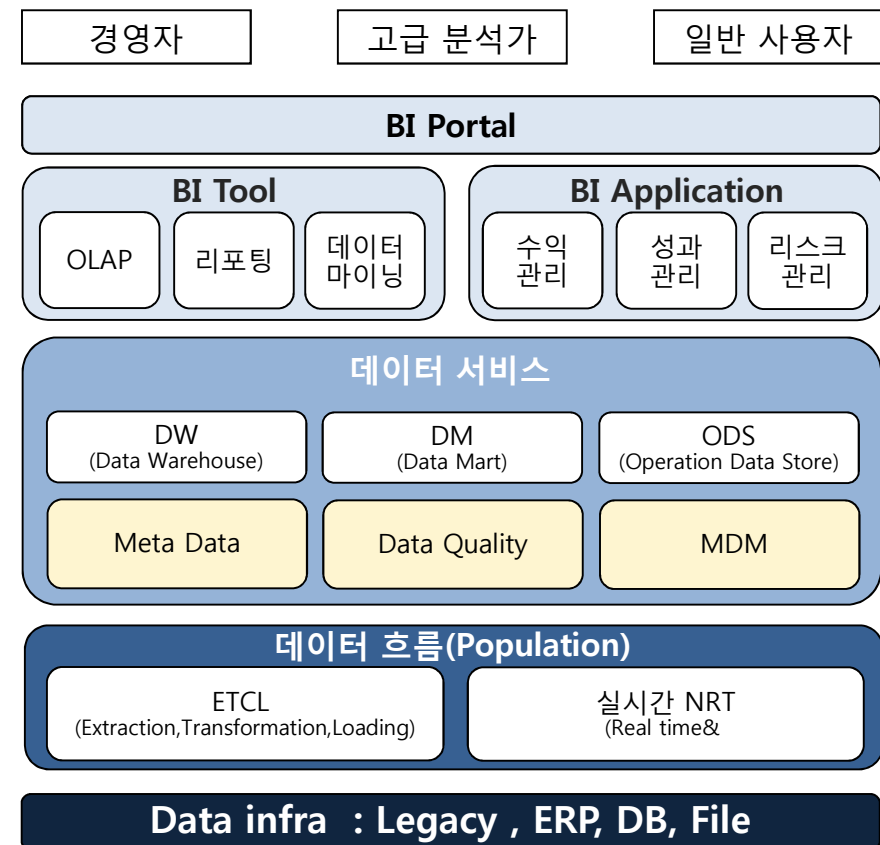
BI framework 변화

❖ BI아키텍처는 UI부문에 대한 End user 접점의 분석 Application 영역과 Data Service 를 위한 데이터 통합(Data integration) 영역으로 구성되어 최근 BI portal 및 데이터 표준/품질 등의 관리부문이 추가 통합되는 추세

일반적인 BI 프레임워크



21c 형 DW/BI 관점



※ BI : Business Intelligence

Issue-3. 미래변화의 대응

❖ Big 데이터 이슈로 Appliance기반의 DW/BI전용 DW DBMS와 Hadoop / mapreduce 분산 비정형 Query 기술의 지원을 통한 제품이 속속 등장하고 있어 수행입장에서 이 부문에 대한 기술 준비가 필요함

1. 분산처리 조회 분석기술



시장변화 대응

OLTP-based DBMS

Columnar DBMS



MPP-based DW Appliance

MPP-based Columnar DBMS

2. Appliance형태의 DW전용 DBMS

“Next Generation BI는 MPP 기술과 Column 기반 기술을 융합한 형태로 진화”



Greenplum



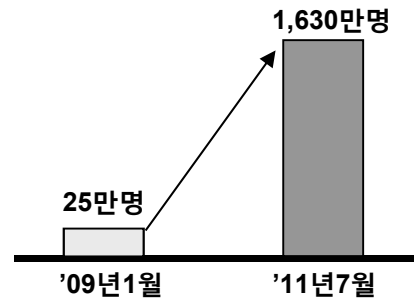
Issue-3. 미래변화의 대응

❖ 다양한 채널의 등장에 대한 유입속도 증가로 이에 대한 BI 차원의 준비가 필요함

스마트폰



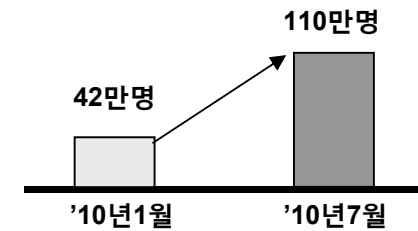
국내 스마트폰(3G)가입자수 증가



SNS



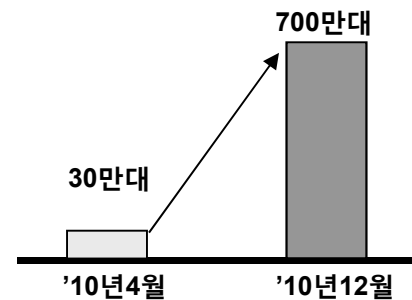
국내 SNS이용자수 증가



Tablet PC



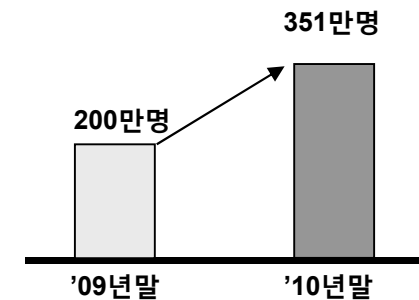
국내 Tablet pc 판매수 증가



Smart TV

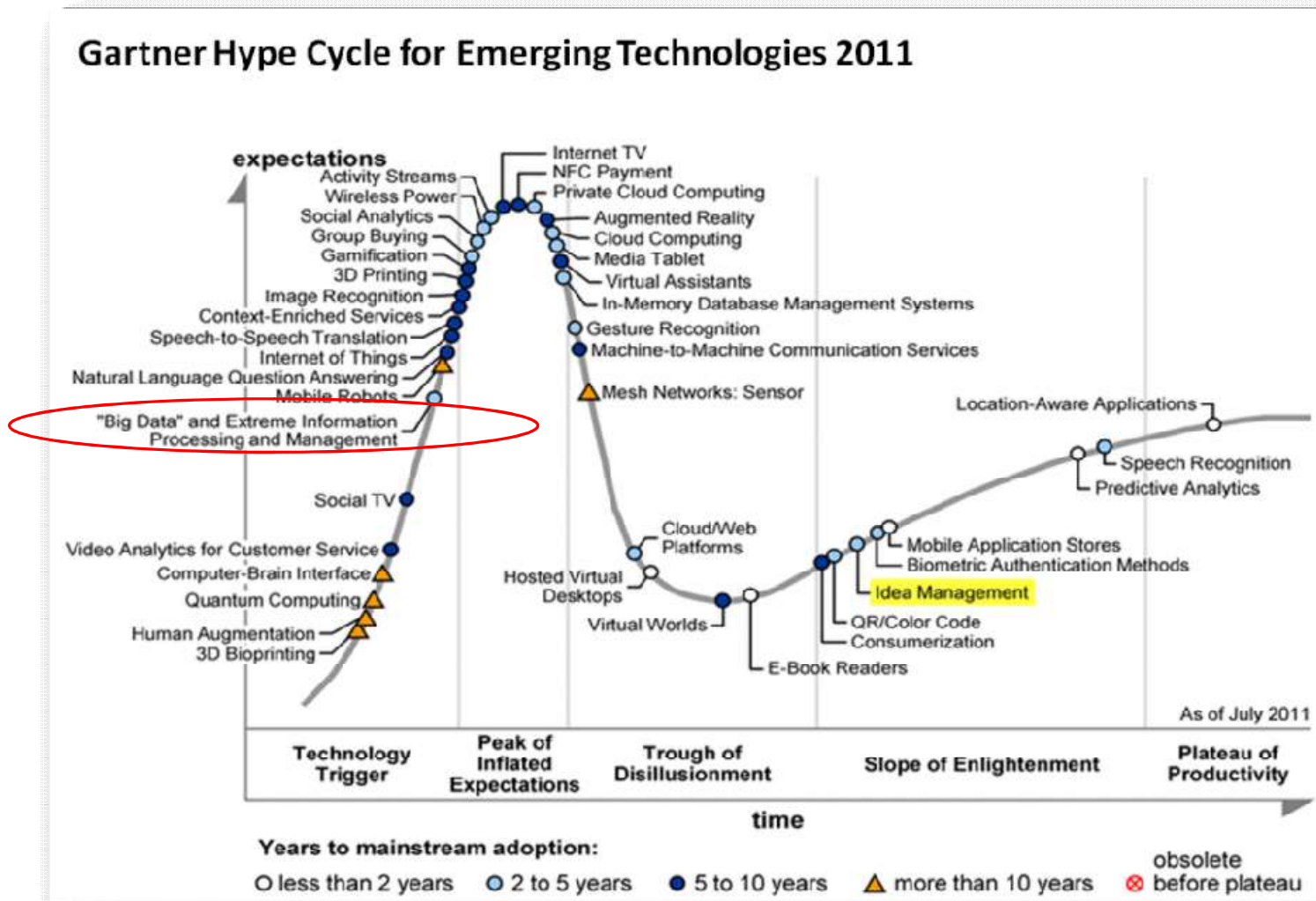


국내 Smart TV 가입 수 증가



Big data와 BI Market

- ❖ BI Market은 BI Platform 기술의 발전과 더불어 꾸준한 성장을 통해 다시 떠오르는 기술로 재 등장하고 있으며 이는 Big data 환경의 대응 기술과 고급 분석기술이 앞으로의 시장을 견인하리라 추측하고 있음



(Source: Gartner, 2011)

❖ Big 데이터란 기존의 관리 및 분석 체계로는 감당하기 어려운, 가용 자원의 한계를 벗어난 자료량을 의미하며 이미 수년전부터 스마트 모바일의 등장, 저장장치 가격의 하락, 소셜 네트워크 서비스 등장, Log데이터 및 ClickStream 분석증가, 각종 Compliance에 대응을 위한 Business Transaction data 분석 증가 등으로 DW시작 시점부터 이미 예견된 일이었다.

“Big data is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single data set”
2011 - Wikipedia -

오늘날 Gartner는 전신인 2000년 초 MetaGroup시절 애널리스트 Doug Laney 가 한 말을 인용 Big data의 모델로 사용함

“Data growth challenges (and opportunities) as being three-dimensional, ie increasing volume (amount of data), velocity (speed of data in/out), and variety (range of data types, sources). Gartner continues to use this model for describing "big data.”

1990' 중반 DW의 원조 Bill. Inmon은 이미 big datawarehouse등장에 대해 암시함

“A big data warehouse is a data warehouse that is an order of magnitude greater than the one you have today...grasshopper”

❖ 소비자 시장에서 개인화된 콘텐츠의 증가 및 유비쿼터스의 등장으로 정보 Access 범위가 급격히 확대됨에 따라 디지털 콘텐츠의 활용은 폭발적인 추세로 가속화되어 Big data 시대는 Memory 기반 DBMS, Cloud 컴퓨팅이라는 서비스 형태의 솔루션과 Appliance 기반의 데이터 처리 기술 시장의 성장을 예견하고 있음

정보 활용 측면의 요구에 부응하기 위한 자체 전문 인력이 부족한 기업을 상대로 한 중소기업 조직을 위한 SaaS/PaaS 서비스 등장

SaaS: Software as a Service
PaaS: Platform as a Service

Cloud Service Market



Appliance Market

• Netezza, GreenPlum, Vertica, Exadata,

Big data Query (비정형기반)

• Hadoop, MAPReduce

시장 혁신을 주도하고 있는 비전통적인 Big Data 영역은 콘텐츠 / 텍스트 분석, In-Memory DBMS, columnar DBMS 등이 지원함
- Gartner, 2011-

❖ Big 데이터 부문(Area)의 분석 활용분야는 산업별로 다양한 분야에서 기존방식과는 다른 고급 분석이 가능할 것으로 여겨짐

Big data
Analysis

금융

- 신규고객 창출 및 세분화
- 고객 이익률/신용/위험관리
- 개인화 Banking 서비스
- 각종 규제준수
- 부정방지 (Fraud Detection)

통신

- 매출관리 및 분석
- 고객성향 분석 및 이탈방지
- 서비스 품질 및 상품라이프 사이클
- 각종 규제준수
- 네트워크 트래픽 분석 예측

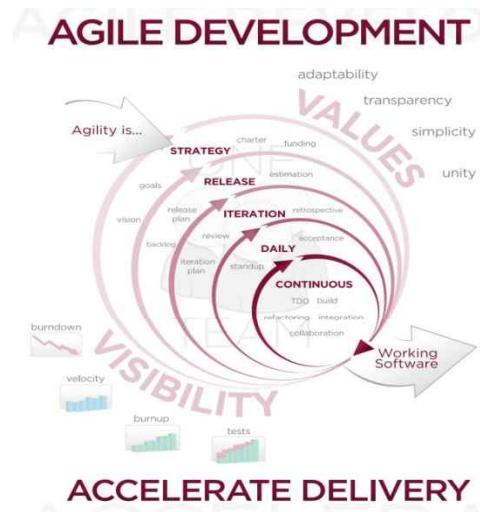
의료(Healthcare)

- 임상/진료데이터분석(Clinical 분석)
- 질병 리스크 관리
- Healthcare 진단지원
- 자연어 처리 / 음성인식 진료

제조 유통

- GIS기술 활용 위치 기반 마케팅
- SNS 마케팅
- RFID활용 추적 기능 물류관리
- 심화된 공정/수율 관리

“Big Data시대의 변화는 피할 수 없으며 먼저 준비하든가 이끌려가든가 선택해야 할 것”
 - 우리 소프트웨어 개발자의 미래는? -



데이터 서비스		
DW (Data Warehouse)	DM (Data Mart)	ODS (Operation Data Store)
Meta Data	Data Quality	MDM

Thank You!
www.datastreams.co.kr