



TERADATA®

차세대정보계 시스템의 비전  
한국테라데이터 장동인 부사장

# 홈플러스 선릉역 가상스토어 Gartner Symposium 2012

“한국의 테스코는 지하철에서 물건을 살 수 있다. QR코드로 스캔해서 주문하면 집으로 배달해 준다. 이것이 digitizing business 예이다”

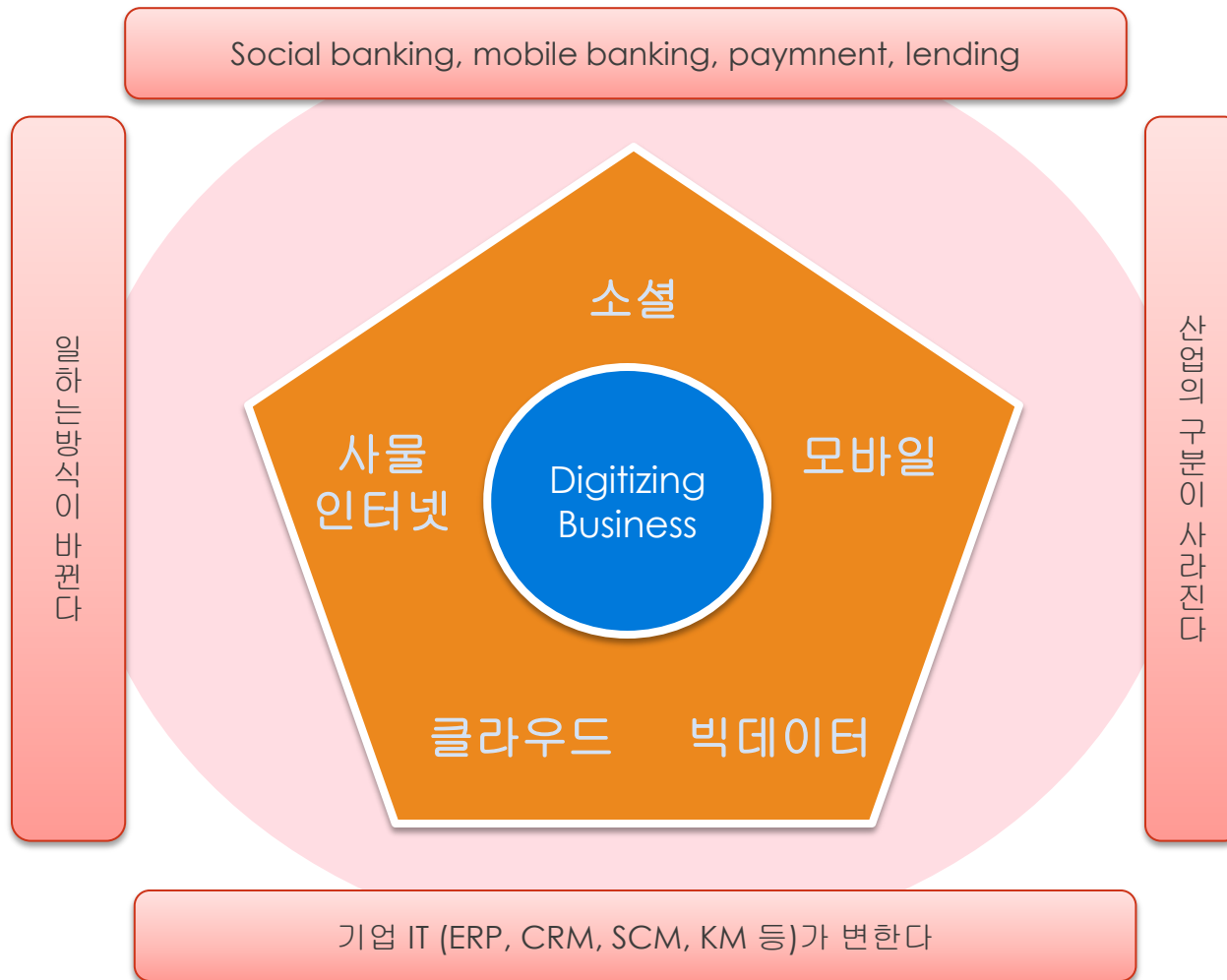


# Business의 변화와 ICT 기술의 명암

시대의 키워드	Technology	Winner	Looser
80' Centralize Business	Mainframe	IBM	Univac, Unisys
90' Down Sizing	분산처리, Relational Database, TCP, Unix	Oracle, MS, Cisco	IBM
00' eBusiness	Internet, eCommerce, Intranet ERP, CRM, SCM	IBM, Oracle, SAP, MS	Old ERP
10' Digitizing Business	Mobile, Social, Big Data, Cloud, IoT	Google, Amazon, Apple, Salesforce	Existing Global IT vendor

# 2015년 ICT 전망 – 은행업의 근본적인 변화

1999년에 인터넷 열풍이 세상의 모든 비즈니스 모델을 바꾸었다.  
15년이 지난 오늘 5개의 통합된 기술이 다시 한번 세상의 비즈니스 모델을 바꿀 것이다.



# Digitizing Business = 파괴적혁신 (제살깎기 혁신)

## 매킨지 Global Institute Interview

[http://www.mckinsey.com/insights/business\\_technology/why\\_every\\_leader\\_should\\_care\\_about\\_digitization\\_and\\_disruptive\\_innovation](http://www.mckinsey.com/insights/business_technology/why_every_leader_should_care_about_digitization_and_disruptive_innovation)

Interview | McKinsey Global Institute

### Why every leader should care about digitization and disruptive innovation

Digitization, automation, and other advances are transforming industries, labor markets, and the global economy. In this interview, MIT's Andrew McAfee and McKinsey's James Manyika discuss how executives and policy makers can respond.

January 2014



PDF Print E-mail Share

#### McKinsey Insights app for iPad



Our latest thinking, delivered seamlessly to your iPad

Disruption everywhere    The employment challenge    Claiming the prize

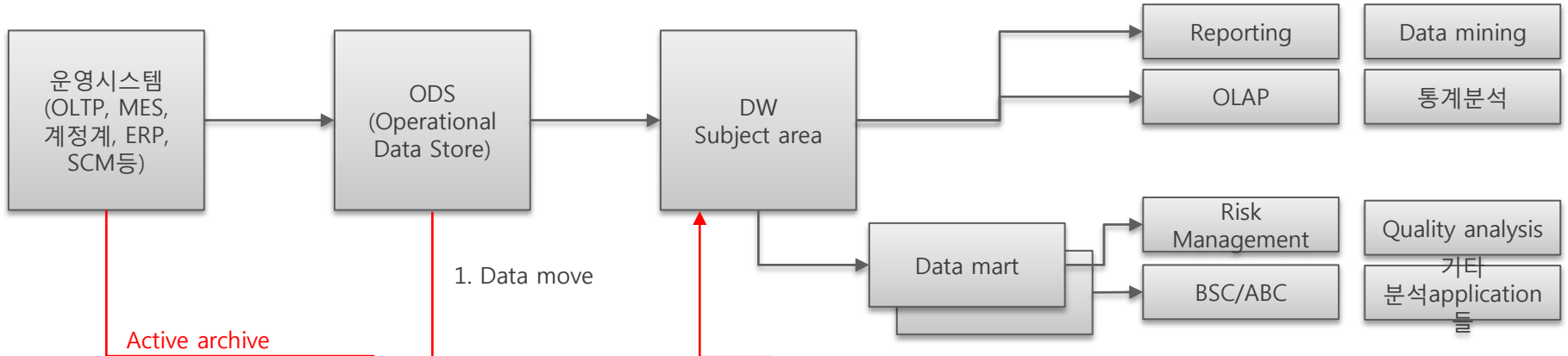
## 기존 정보계 시스템의 이슈

### 정보계는 차세대 시스템 구축에서 제외되었다...

- 현재 EDW의 이슈
  - ✓ Old, out-dated HW, SW (단종된 제품도 존재)
  - ✓ 현상태를 upgrade 할 것이냐 아니면 새로운 기술을 도입할 것인가?
  - ✓ 새롭게 등장하는 빅데이터들을 어떻게 수용할 것인가? 아니면 별도로 가야 하는가?
  - ✓ 현재 빅데이터 기술진보는 어디까지 왔는가? 검토단계
  - ✓ 기존의 시스템을 안전하게 migration할 수 있으며
  - ✓ 기존 시스템과 무리없이 integration 할 수 있는가?
- 이러한 상황에서 빅데이터 기술은 어디까지 왔으며, 앞으로 어떤 방향으로 갈 것인가?

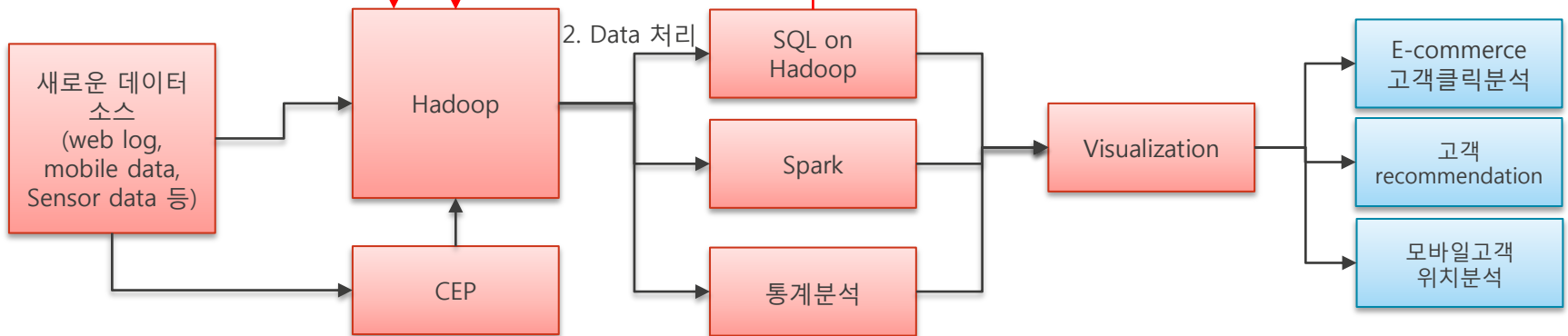
# EDW/BI 와 빅데이터시스템의 병존

## 기존 DW/BI 시스템



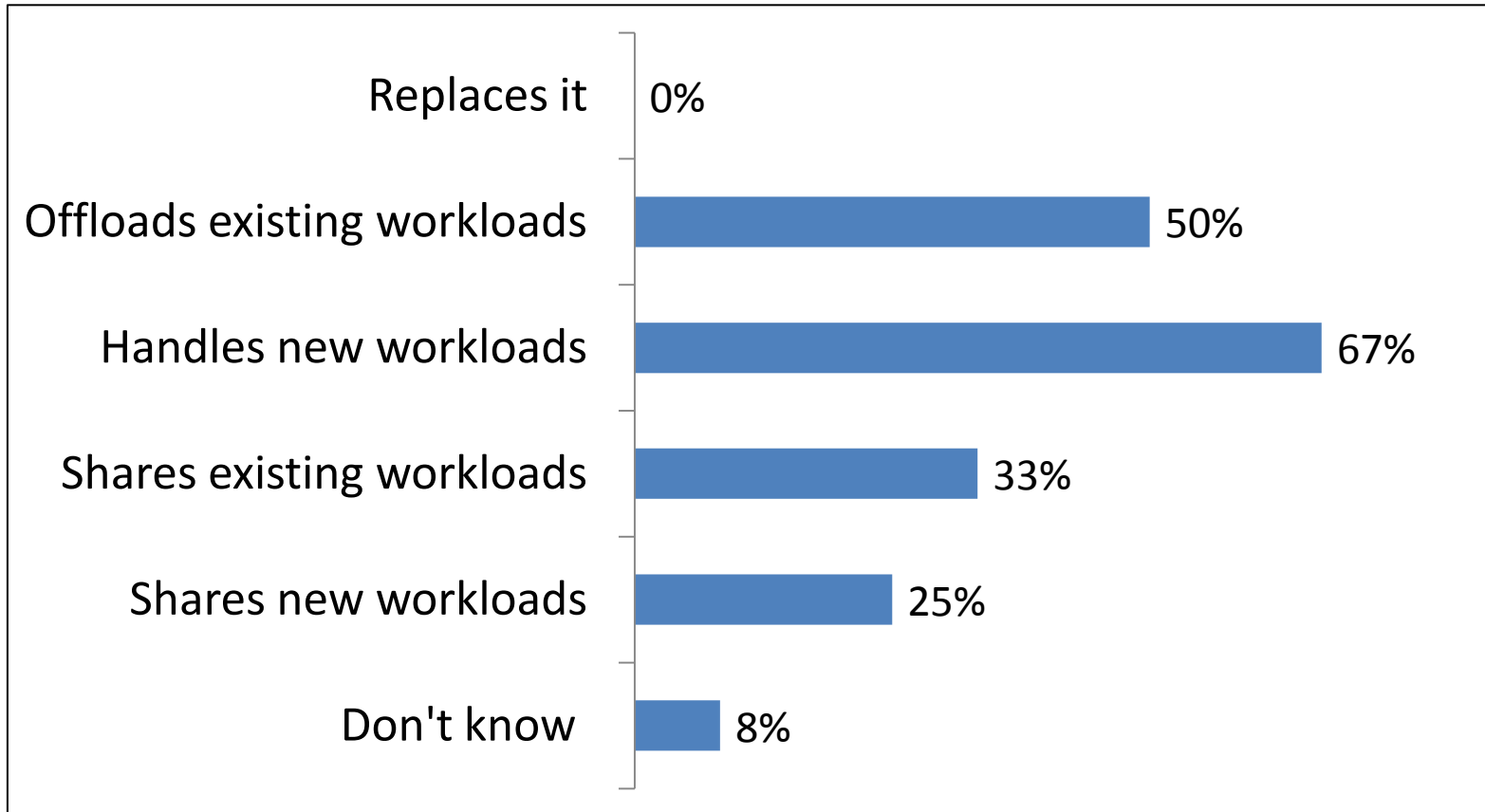
## Big batch job

## 신규 빅데이터 시스템(1)



대부분의 기업 - EDW, Big Data는 별도로 존재

## 하둡을 기존 DW에서 어떻게 활용하는가?



*Based on respondents that have implemented Hadoop. BI Leadership Forum, April, 2012*



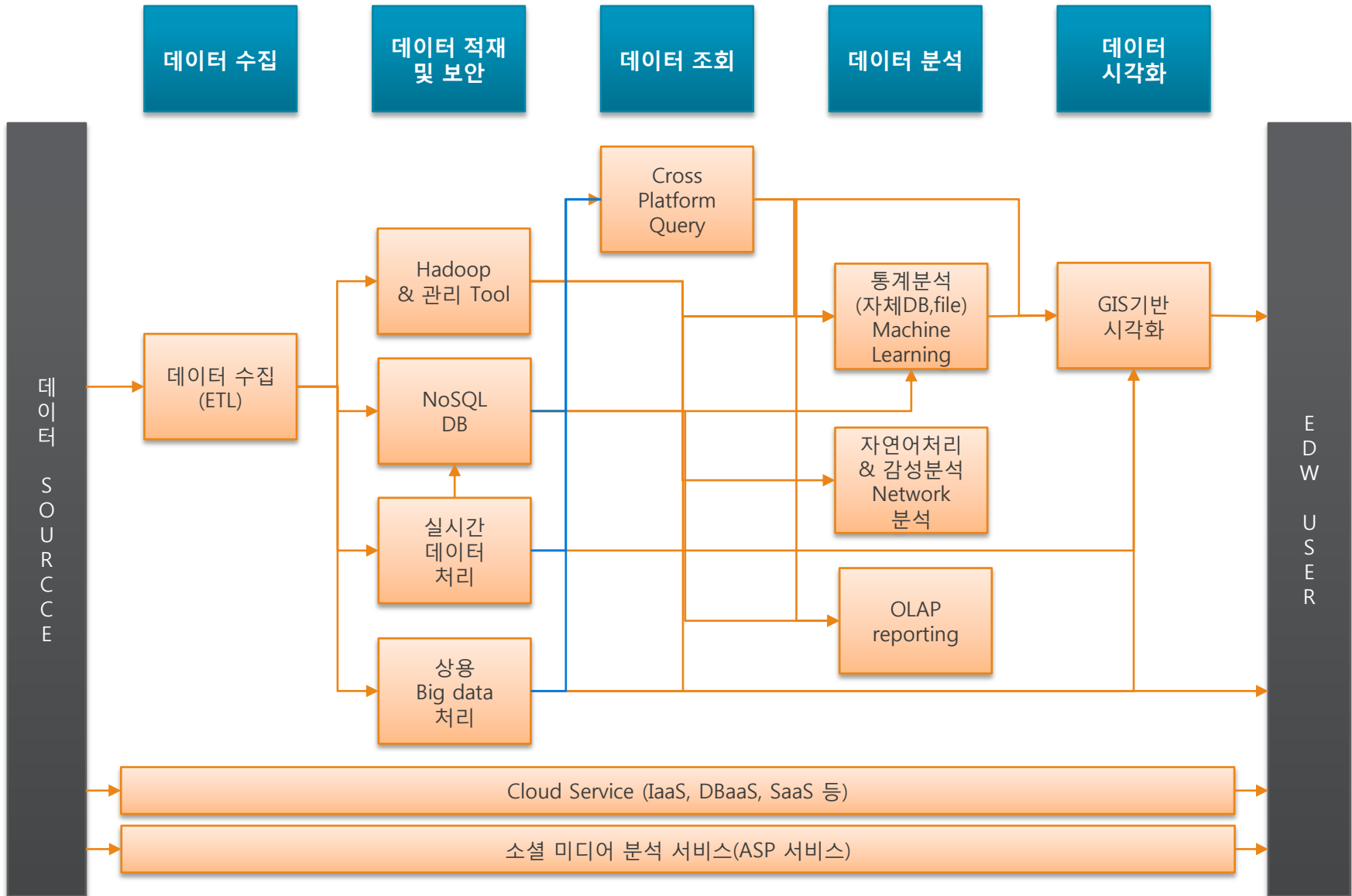
# 도대체 빅데이터 vs. 기존 DW와 어떤 차이가 있는가?

	빅데이터 기술	DW(BI)
데이터	정형, 비정형(SNS, 사진, 동영상, 위치정보, 지리정보, 기타)	정형 데이터
기본기술	Hadoop, NoSQL, 기타	RDB에 근간
특징	수십, 수백 억 건 이상도 처리 가능 HW만 늘러가면 퍼포먼스가 증가함 (Scale-out)	데이터가 수십억 건 이상 되면 HW를 더 많이 사용해도 performance 가 증가하지 않음
서버	싼 일반 x86 서버 사용	주로 UNIX. X86에 비해 고가.
소프트웨어	대체적으로 공개 소프트웨어 (물론 상용 소프트웨어도 있음)	모두 고가 상용 소프트웨어
데이터 처리	3개 노드에 데이터 중복해서 카피함 1~2개의 노드에 문제가 동시에 있어도 자동복구	데이터 중복을 안 함. 대신 시스템을 중복. 고비용
문제점	공개소프트웨어이므로 지원 문제	가격이 비싸고, 데이터처리 양이 늘어나면 한계에 이른다

## 차세대 EDW 시스템에 대한 요구사항

- 한번에 빅데이터에 대한 요구사항이 나오지 않는다.
- 사용자의 요구사항은 활용하면서 나온다.
- 유연한 구조 (Schema free, Data Lake) – 센서, 로그데이터 특징.
- 사용자는 지금까지 쓰던 것, 익숙하던 것을 고집한다. (DW, OLAP)
- DW이든 빅데이터이든, 결국 분석하여 보여지는 것은 유사하다. (insight)
- 변화관리가 생명이다.
- 빅데이터 분야에는 대단히 많은 tool과 solution이 있다.
- 차세대 EDW는 이렇게 많은 tool과 solution을 잘 구사할 수 있는 big data solution architect의 능력에 따라 달라진다.
- Data security, governance는 더욱 중요 같다.
- 기존 DW와 차이점, 유사점을 잘 알아야 한다.
- 결국 기존EDW와 관점을 유지 및 업그레이드해야 한다.

# 차세대 EDW Architecture



\* ETL: Extract, Transport & Load, NoSQL: Not Only SQL, IaaS: Infra as a Service, DBaaS: DB as a Service, DaaS: Data as a Service, SaaS: Software as a Service, OLAP: Online Analytic Processing

# 빅데이터 기술의 시사점

빅데이터의 핵심 기술은... Hadoop만이 전부가 아니다.  
데이터의 성격 따라서 선택되는 기술이 달라져야 한다.

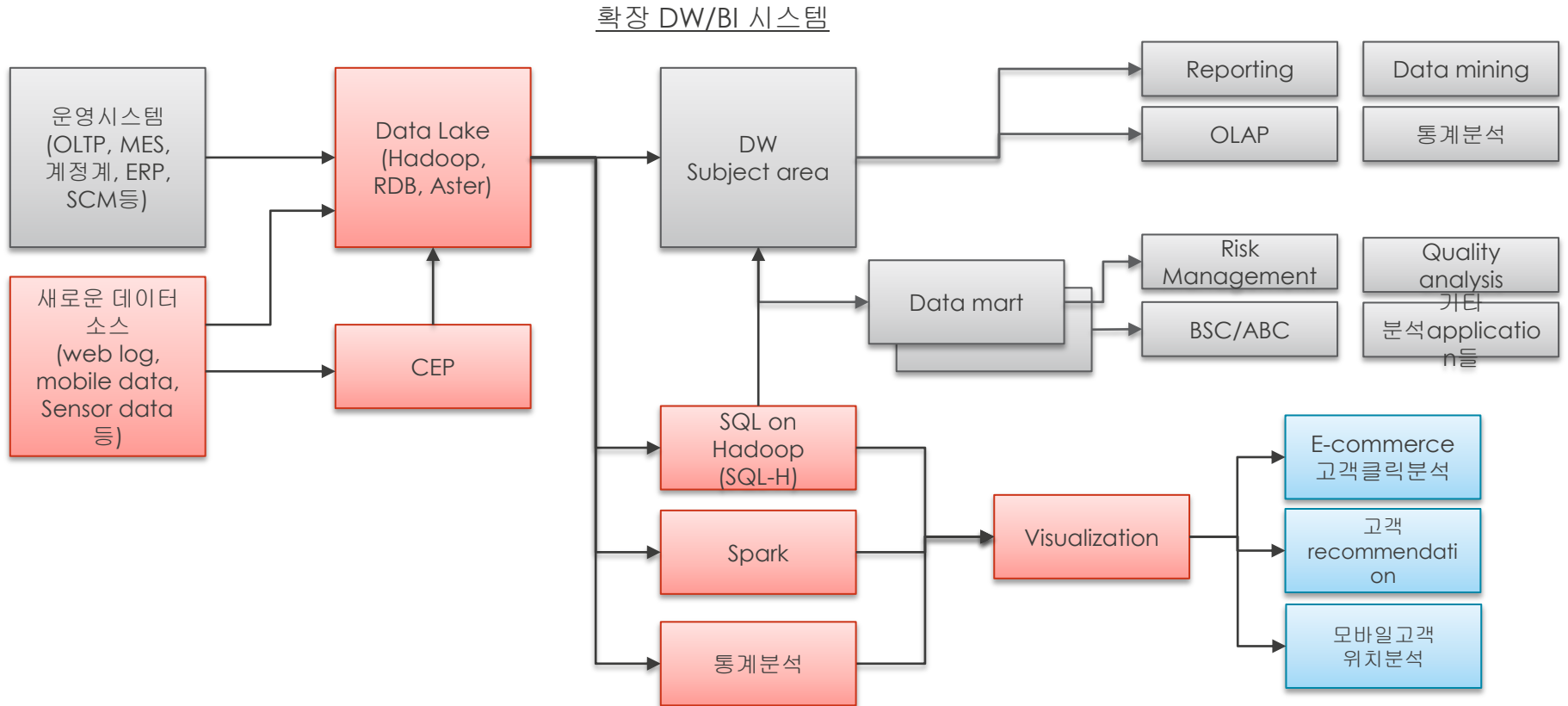
종류	장점	단점	제품
Hadoop	저렴한 비용 언어에 독립성 오픈소스	Map & Reduce가 어렵다 (SQL on Hadoop 필요) 각 회사마다 다른 버전 버전호환성이 낮다 장애시 대처가 어렵다 실시간 분석 안 된다	Apache Hadoop Cloudera Horton Works MapR
SQL on hadoop	어려운 Map & Reduce 대신 SQL을 사용할 수 있다. 데이터는 여전히 Hadoop에 보관한다 RDB기반의 OLAP, 분석툴을 사용할 수 있다 (ODBC connector)	각 제품마다 개발방식이 모 두 다르기 때문에 호환성이 없다.	Hive 그루터 Tajo Cloudera Impala Hortonworks Stinger MapR Drill Facebook Presto
NoSQL	빠르고 유연하다 확장성과 가용성이 좋다	트랜잭션 처리는 어렵다 구조상 Join 이 안 된다 SQL같은 복잡한 Query는 안 된다	CouchDB Mongo DB Casandra DB Hbase
실시간처리	실시간 Rule-base로 처리할 수 있다	많은 데이터를 적재할 수 없다	Esper Storm Apache S4 Spark
상용 빅데이터 처리	특별한 목적으로 개발 데이터 수집에서 시각화까지 suite으로 존재	가격이 비싸다	Aster(Discovery) Splunk (Machine Data) Sumologic (웹로그분석)

## EDW관점에서 빅데이터 기술을 채용한다면...

### EDW+빅데이터는 글로벌 트렌드이다.

- 지금까지는 빅데이터와 EDW 는 서로 별개로 존재
- EDW는 근본적인 upgrade가 필요한 시점
- 빅데이터 기술 역시, 데이터를 핸들링하고 다양한 분석
- 빅데이터 기술은 정형, 비정형 데이터를 통합 분석 가능하다
- 기존 EDW를 그대로 확장하는 것 보다 비용면에서 유리하다
- 기존 빅데이터 TF 진행 상황이 더디다.
- 현재 EDW의 이슈를 빅데이터 기술을 활용해서 업그레이드가 가능하다
- 이러한 상황에서 빅데이터 기술은 어디까지 왔으며, 앞으로 어떤 방향으로 갈 것인가?
- 테라데이타의 방향은 EDW 입장에서 빅데이터를 기술을 포함하여 가고 있는 유일한 업체이다.

# DW/BI 의 입장에서 빅데이터 기술 수용한 Architecture

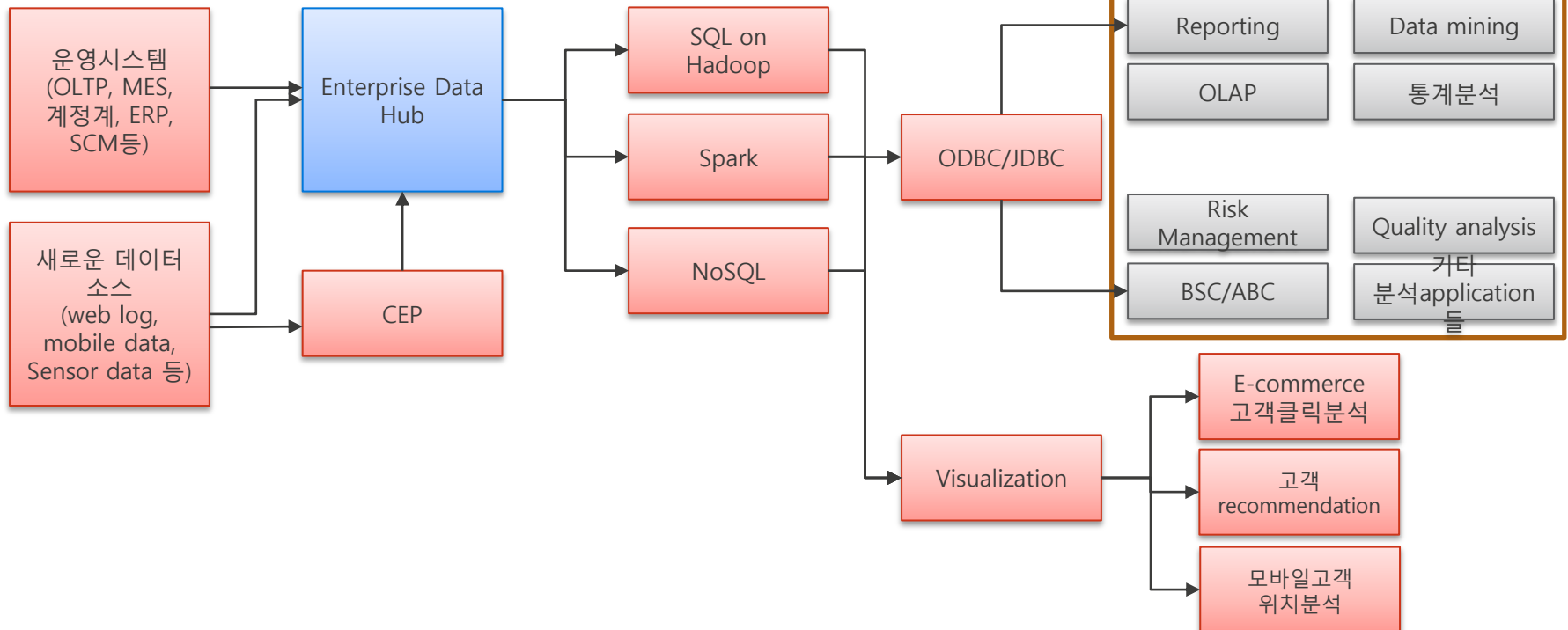


Teradata의 방향: EDW입장에서 big data 기술을 흡수. 기존 시스템과 통합

# Big Data 입장에서 DW/BI 를 흡수 통합한 Architecture

확장 DW/BI 시스템

업그레이드 및 재구축



Cloudera, Pivotal 의 방향; EDW를 Hadoop으로 대체

# 빅데이터 인프라 및 서비스 Map (외산솔루션)

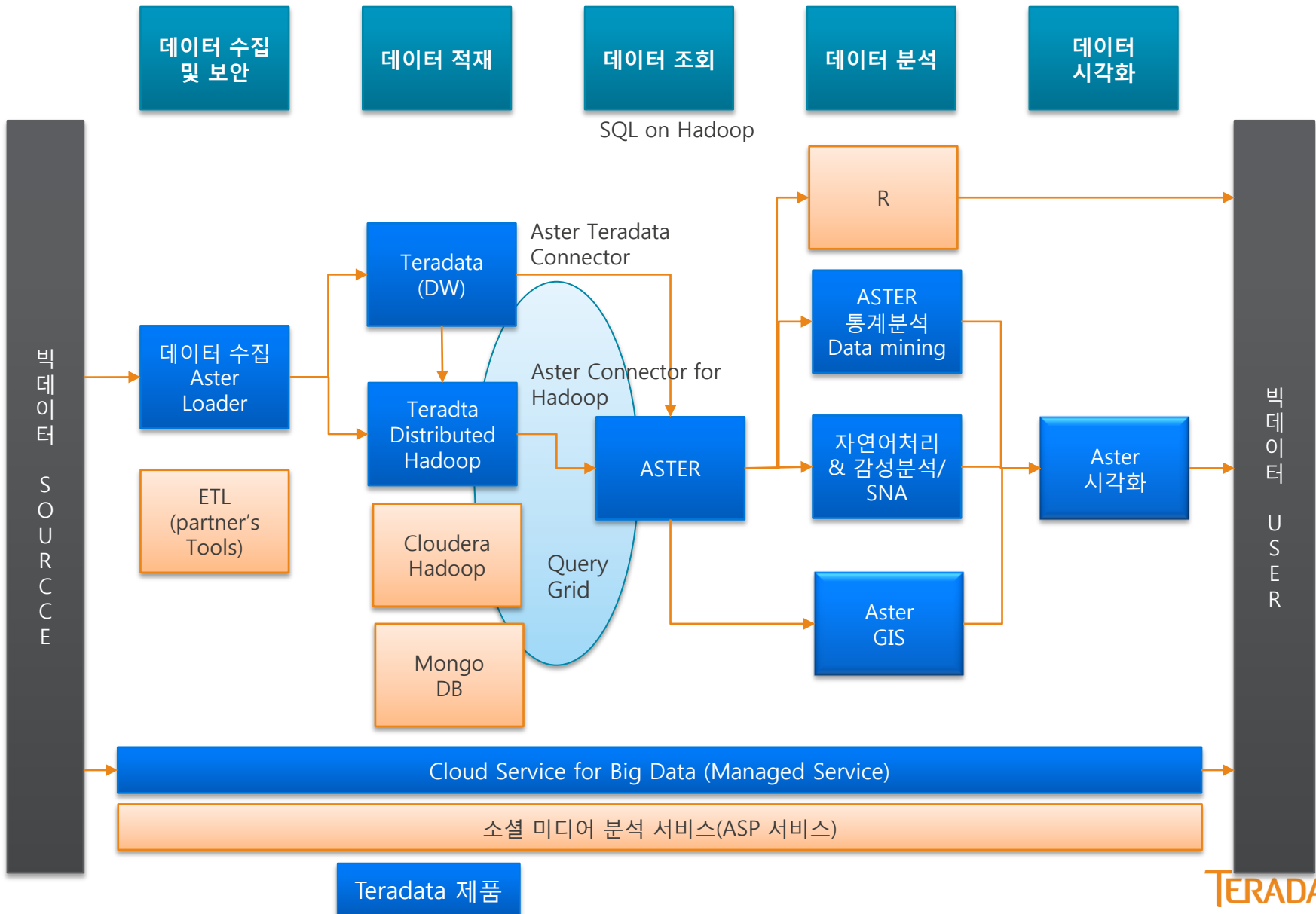
Open Source  
 유료 SW

빅데이터를 처리, 분석하기 위한 각 분야의 오픈 및 테라데이타 Aster위치

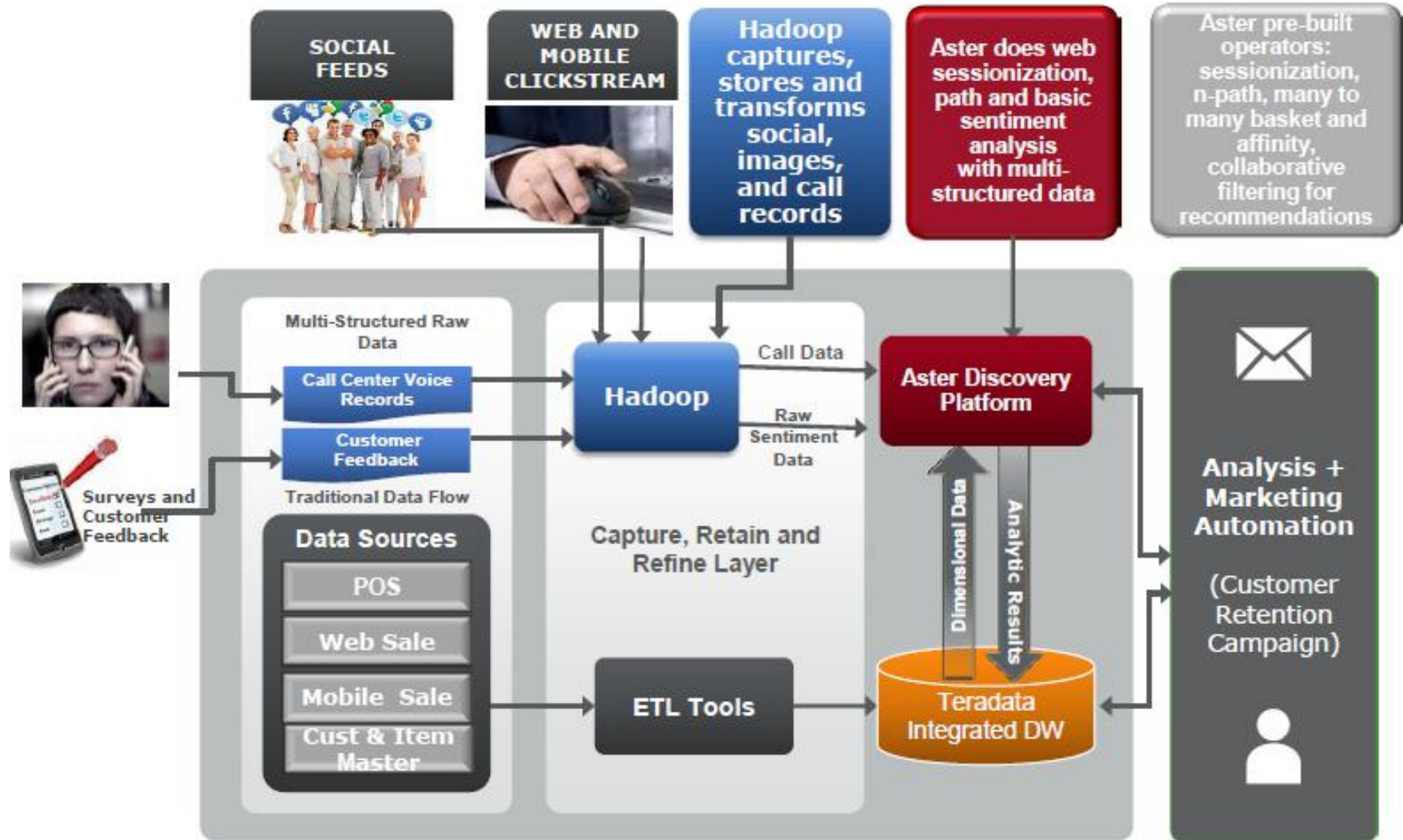
빅데이터 인프라									빅데이터 서비스	
데이터 수집	데이터 적재	데이터 조회 SQL	NoSQL	보안	실시간 데이터분석	통계분석	시각화	Hardware Appliance	Crawling/ 감성분석	Cloud for Big Data
Informatica	Apache Hadoop 2.0		Cassandra	Vormetric	SAP Hana	R	D3/ Visual.ly	Oracle Exadata	Salesforce.com Radian6	Amazon (IaaS+ Hadoop)
Talend (Open Studio)	Cludera (CDH 4.0 Impala 2.0)		Mongo		CEP Esper	SAS	Qliktech	EMC Greenplum	SAS SMA	SoftLayer (IaaS+ Hadoop)
IBM InfoSphere DataStage	HortonWorks (Data Platform 2.0 Stinger)				Oracle	SPSS	Micro Strategy	Teradata Aster		Rackspace (IaaS+ Hadoop)
	MapR (M5 hadoop, M7 hbase)		Riak		Tibco	Tableau		IBM Netizza		Cloudant (DBaaS)
	Teradata Aster					Spotfire				Amazon Dynamo (DBaaS)
	Splunk (proprietary DB)					Aster	Aster		Aster	Sumologic



# Teradata Unified Data Architecture



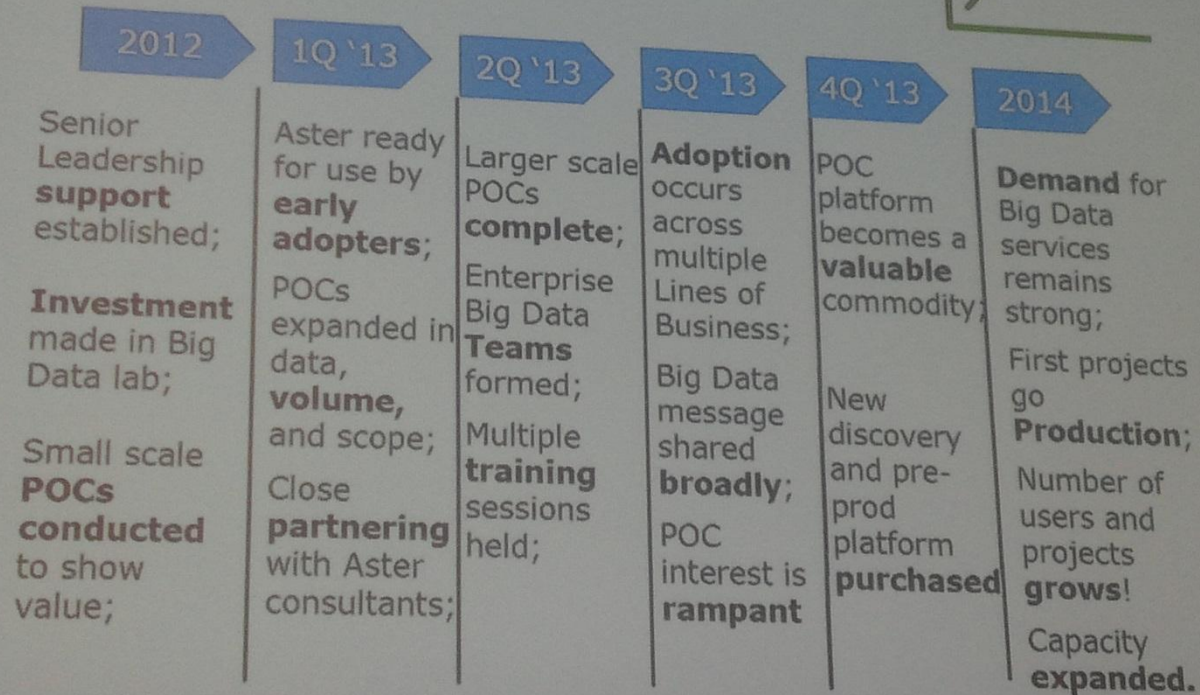
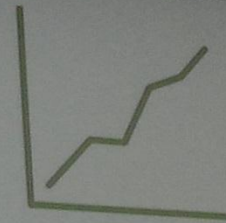
# Wells Fargo Bank



# Wells Fargo Bank

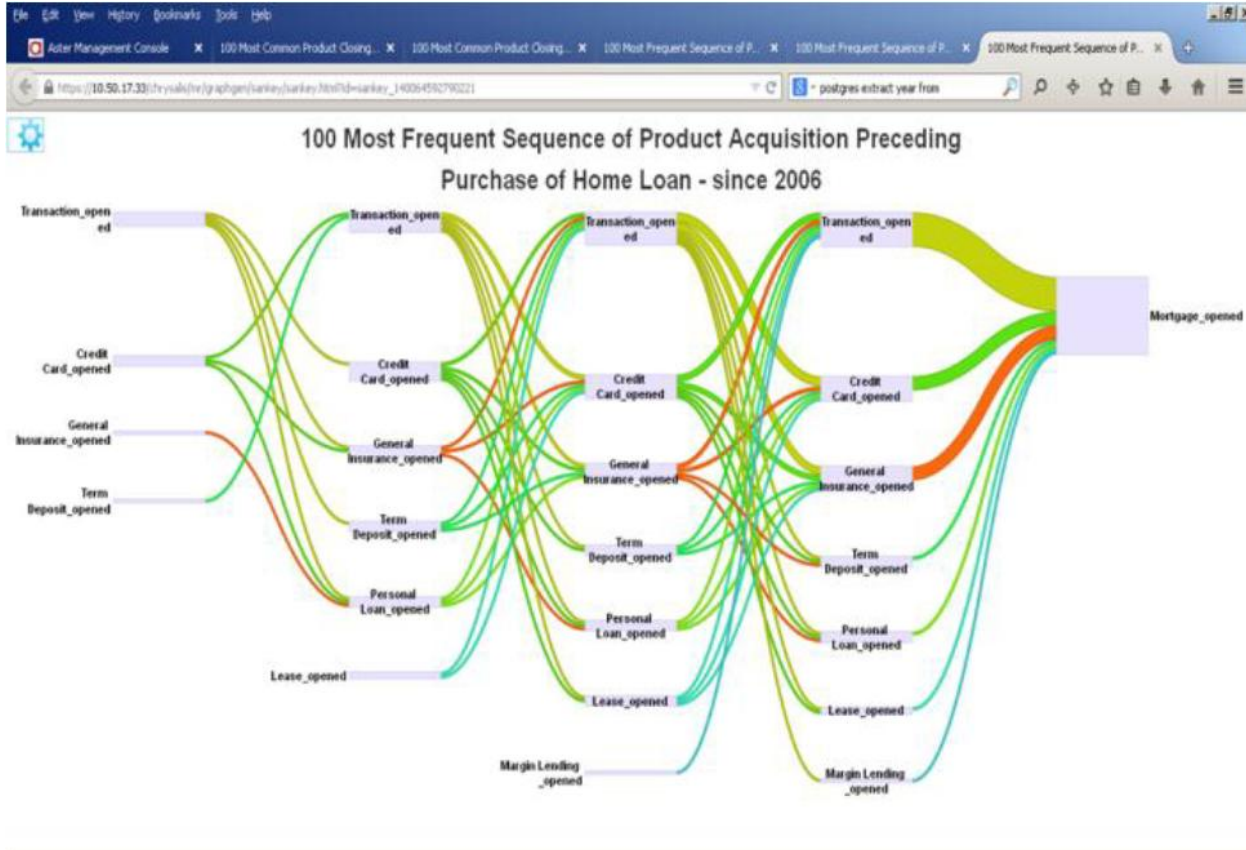
## Big Data Timeline

Fast and Furious



# Westpac Bank

주택담보대출을 받기전 주로 어떤 event가 일어나는가?  
(11년간 데이터 분석)



➔ 주소변경, 반복적인 상당금액의 저금...  
➔ 카드 사용은 어떠한가?

# Westpac Bank

기존 주택담보 상품 이탈 이전에 어떤 일이 발생하는가?



→ 주택가격 평가서 구매, 15%정도가 이탈,

→ 고객만족도가 높을수록 서비스기대치가 높고, 불만제기 횟수가 많을수록 이탈

# Strata Hadoop World 2014, Oct 15-17, New York



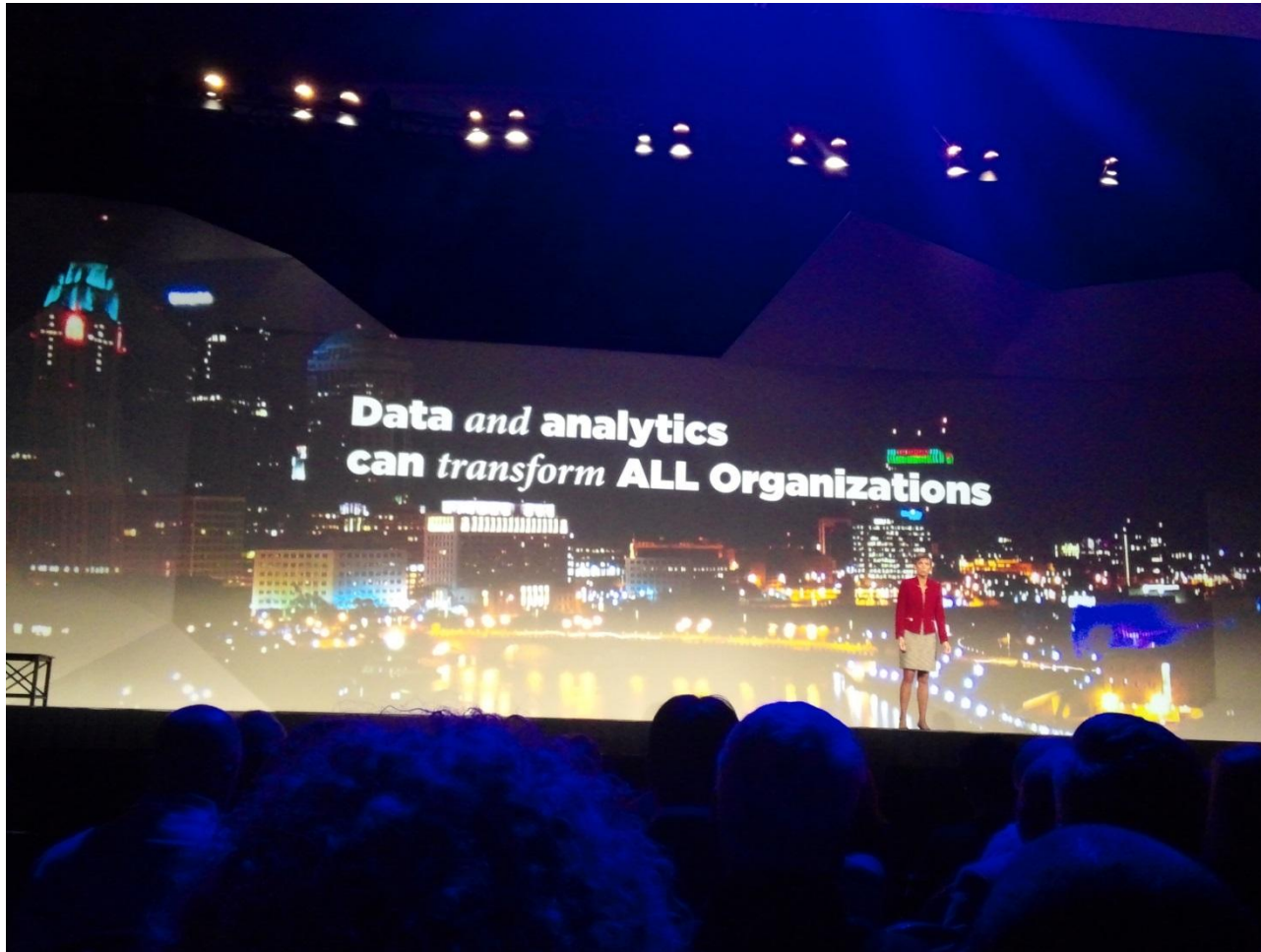
- 중요한 빅데이터 컨퍼런스
- 상상치 않은 조합 발표 (Cloudera + Teradata)
- Hadoop
- Spark



## Strata Hadoop World

- Cloudera 와 O'Reilly 출판사
- 원래 개발자들의 Hadoop Conference
- 2009 500명 → 2014 5000여명 (6회)
- \$2,500 입장권 조기 마감 !!
- 주요 주제; Hadoop Focus, Spark의 탄생
- 참가자: 개발자 위주에서 기업 IT 전략 분야, 30~40 대
- 솔루션 위주 발표
- 기업 IT의 핵심으로 가기 위해서는 앞으로 많이 노력 필요

# Teradata PARTNERS 2014, Oct 19~23, Nashville



Nationwide Insurance, VP, Kathy Koontz, 기조연설



## Teradata PARTNERS 2014

- Teradata 고객들의 conference
- 최초 Keynote – 고객 (Nationwide Insurance VP, Kathy Koonz)
- 개발자, 제품 보다는 고객의 Testimonials (100여 세션)
- 10년 이상, 2014 -- 5000여명
- 주요 주제; EDW, UDA, Aster, Hadoop
- 참가자: 기업 CIO, IT 전략 분야, 40~50 대
- 고객 사례 위주 발표
- 내용에 비해서는 잘 알려지지 않은 컨퍼런스 - 한국 20여명...

## EDW 입장에서 본 2015 Big data 기술의 전망

- 하둡을 각 기업이 customizing → 기업에 맞는 범용 솔루션화
- 하둡의 상용화는 결국 Unix, Linux 와 비슷한 방향으로 갈 것
  - ✓ Open source software의 한계
  - ✓ 그러나, OSS 목적으로 지속적인 개발은 있을 것 (예: MySQL → Maria DB)
- Spark + 하둡의 결합 → 하둡의 상용화/솔루션화
  - ✓ 예: Platfora
- SQL on Hadoop은 계속 개발될 것
  - ✓ DW 분야에 활용되기 위한 기본 기능 (Full SQL support + UDF 등)
  - ✓ Integration : 기존 EDW를 이루고 있는 RDB. (Sqoop을 넘어야)
- RDB/Hadoop/NoSQL cross platform Query
  - ✓ Teradata Query Grid
  - ✓ Oracle Query Franchising
- EDW → Logical Data Warehouse, Data Virtualization, Data Lake
  - ✓ Cross platform metadata, master data management, quality, data governance ...
- Big data platform with IoT



The original big data company

# TERADATA.

장동인

010-5259-9509

[Don.chang@teradata.com](mailto:Don.chang@teradata.com)  
[donchang@hanmail.net](mailto:donchang@hanmail.net)

Facebook:

<http://www.facebook.com/jang.cloud>