WELCOME

제 1강

빅데이터/클라우드의 개념과 현황

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

기업의 정보시스템, 아키텍처를 통한 핵심 강의!

- 다양한 소프트웨어 기업들의 포지셔닝에 대한 이해
- 실리콘밸리 발 뉴스에 대한 이해
- 빅데이터 · 클라우드 강의나 책을 이해하는 길잡이

강의 전 워밍업

1. 이 주제 관련 강의나 책에서 언급하는 데이터 증폭과 그 이용에 관한 장황한 얘기들은 없습니다.
2. 기업의 정보시스템, 핵심 아키텍처로 시작한다. 많이 추상화시켰으니 겁먹을 필요는 없습니다.
3. **모든 것의 플랫폼 = 인터넷**, 모든 것이 연결되고 있다는 생각을 계속하면 좋습니다.
4. 몇 가지 암기가 필요합니다! ^.^

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

키워드 설명 – 1

1. **메인프레임 · 유닉스서버 · X86 PC 클러스터**
2. **다운사이징** 메인프레임 중심에서 유닉스 서버 중심으로 정보시스템을 바꾸는 것
3. **OLTP / Data Warehouse** 예) 계정계 / 정보계, 생산–판매–거래–승인 / 분석–보고
4. **RDBMS** 정형 데이터를 논리적으로 적재하는 데이터베이스
5. **미들웨어**

   분산화로 인한 서버 다양성으로 인해 유저 프로그램(클라이언트) 개별 대응을 단순화 하기 위한 공유 서버 프로그램.

   미들웨어 때문에 데이터베이스를 클라이언트 프로그램에 영향을 주지 않고 바꾸는 것이 가능하고 역시 클라이언트 프로그램을

   데이터베이스에 영향을 주지 않고 바꾸는 것 또한 가능하다. (ODBC, 텍시도, 티맥스, WAS)
6. **패키지 라이선스 판매 / 서비스 판매 (서브스크립션(Subscription) 방식)**
7. **VLDB/Bigdata : very large database/very VLDB** 데이터의 양과 처리능력을 함께 봐야 한다.

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

키워드 설명 – 2

8. **Scale Up / Scale Out**

   Scale Up : 서버 하나에 CPU · 메모리 · HDD를 증설하는 방식

   Scale Out : 서버 스펙을 올리는 것이 아닌 서버 대수를 늘리는 방식

9. **클러스터, 병렬처리 / 가상화**

10. **RDBMS & Hadoop**

11. **Digitizing business** 소셜+모바일+사물인터넷+빅데이터+클라우드 이 5가지가 결합된 비즈니스

12. **IaaS 하드웨어 인프라** x86 클러스터, AWS S3 EC2

13. **PaaS 소프트웨어 인프라** 데이터베이스, AWS RDS

14. **SaaS 유저 프로그램** 구글의 이메일, 엑셀 같은 오피스 프로그램, CRM(Salesforce)

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

목차

1 전형적 기업정보시스템을 만들다

2 인터넷이 만든 빅데이터 처리 필요성

3 검색카운팅, 실시간추천이 클라우드를 만들다

4 세 가지 정보 시스템

5 소셜과 모바일이 클라우드 기반기술의 활용을 높이다

6 소셜+모바일+사물인터넷+클라우드+빅데이터 =digitizing

7 클라우드의 수익모델로의 변화 (IT자원의 서비스화)

8 클라우드를 평가하다

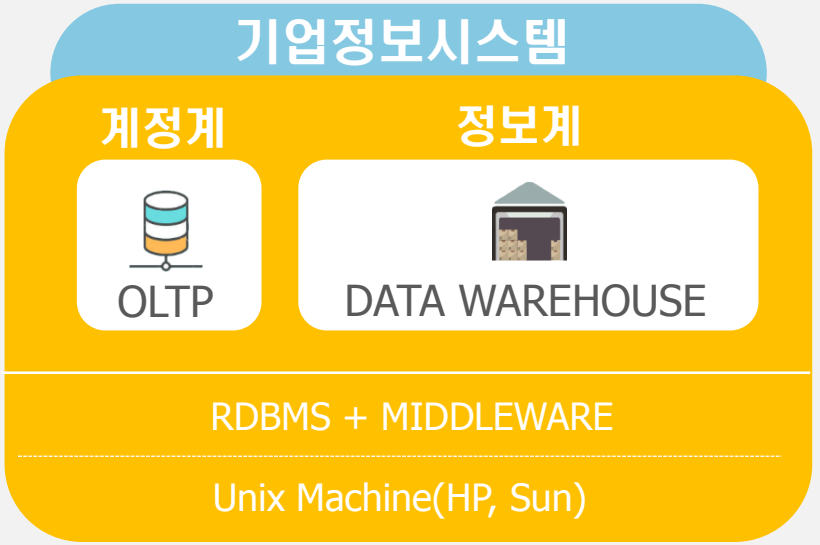
9 엑셈사업모델 1, 2

10 엑셈 빅데이터 사업 추진 방향

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

1990년대
중반 이후

메인프레임
(중앙집중형)



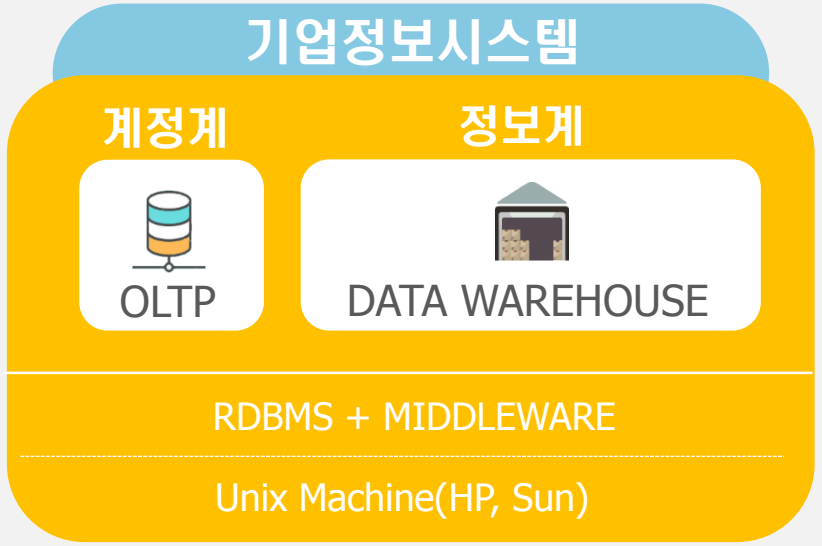
(분산, 다운사이징)

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

1990년대
중반 이후

VLDB

Very Large DataBase



- **VLDB 처리능력 검증**으로 기업정보시스템의 전형으로 자리잡음

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

2000년대
이후

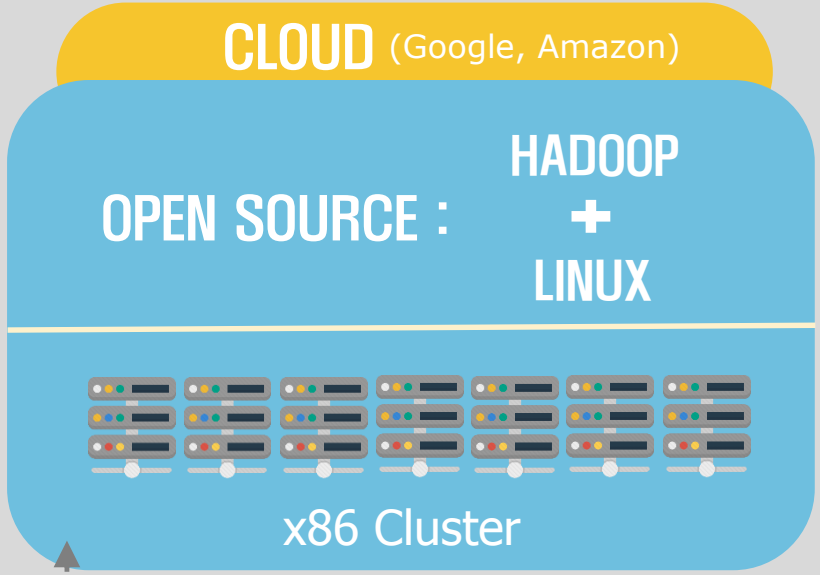
원조 빅데이터의 등장

amazon Google

인터넷이 만든
많은 양의 데이터를 **효과적**으로 다뤄야 한다.

BUT **기업정보시스템의 한계**

클라우드의 탄생!

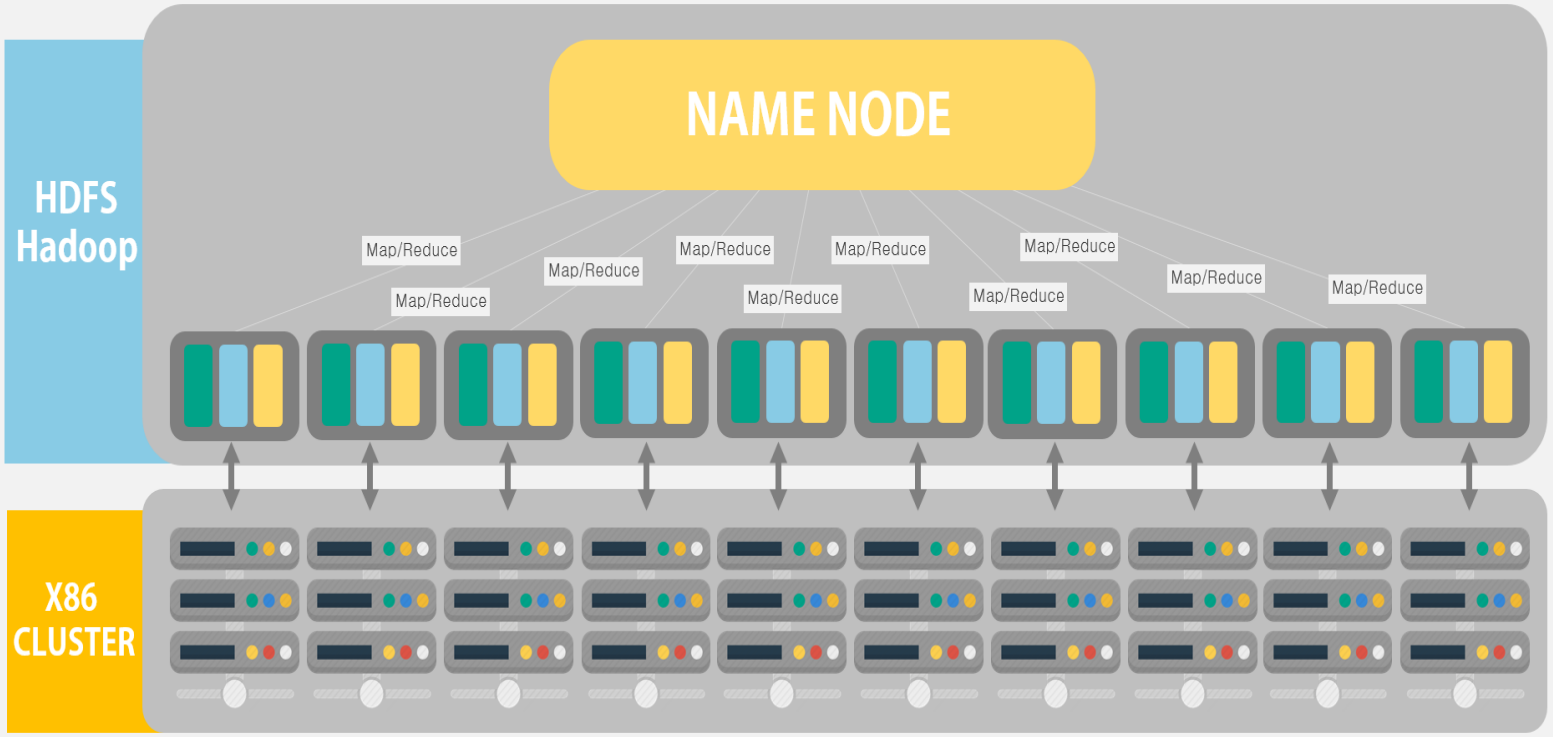


데이터처리능력이 관건!

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

2. 인터넷이 만든 빅데이터 처리 필요성

2000년대
이후



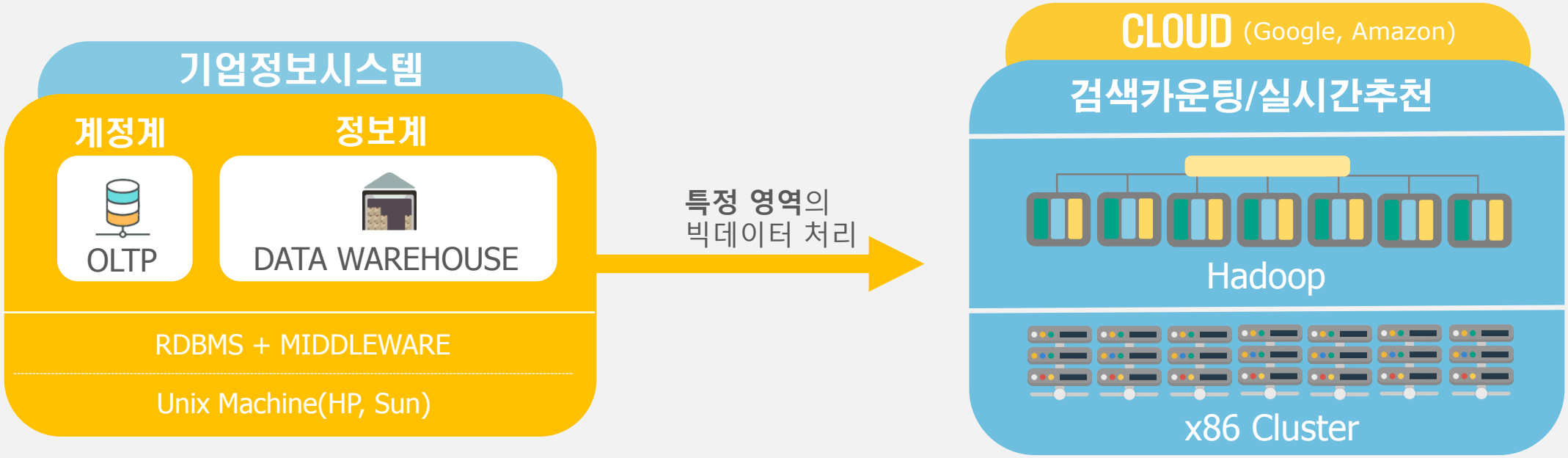
클러스터(병렬처리) 기술

- X86클러스터와 하둡의 분산병렬처리

가상화기술

- 클라우드를 수익모델로 전환!

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.



세계적 규모의 **e커머스** 와 **포탈**의 등장

빅데이터 규모의 데이터 처리 요구

실시간성, 검색, 추천이 새로운 정보시스템을 요구

→ 클라우드는 **아마존**과 **구글**의 정보시스템

는 서비스화를 통해

시작

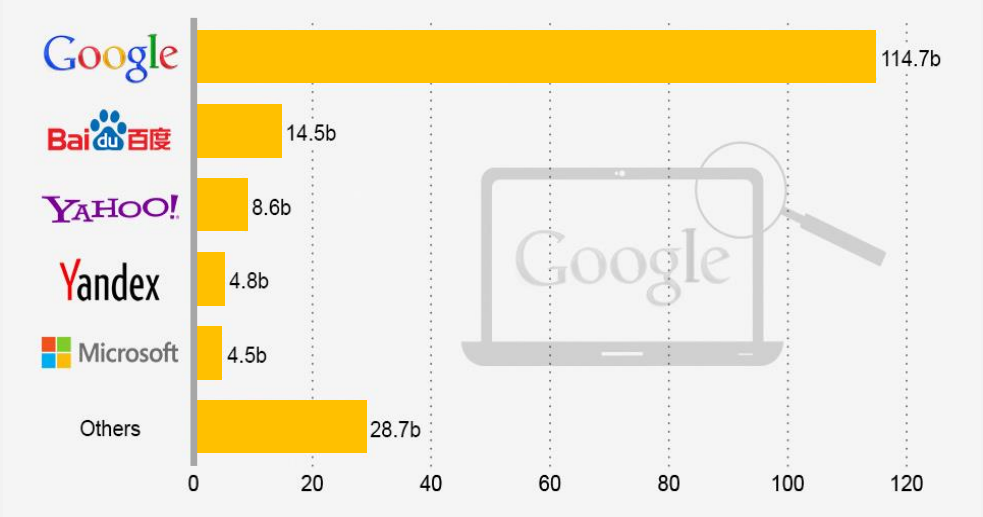
Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

Google Adwords

"Google이 웹페이지 및 블로그에 게재하는 광고"

출처 : comcore

Google 은 한 달에 **1,150억** 검색어 데이터를 관리한다.



- 사용자가 Google에서 서칭해 본 모든 **검색어**를 분석
  - ★ 1년에 나오는 데이터량! : (한달) 1,150억건 * 12 = 13,800억건
- 사용자가 서칭한 **검색어와 관련된 광고** 게재

매출 **$66billion**

user 평균 클릭 단가를 1불로 가정하면

66,000,000,000번 클릭

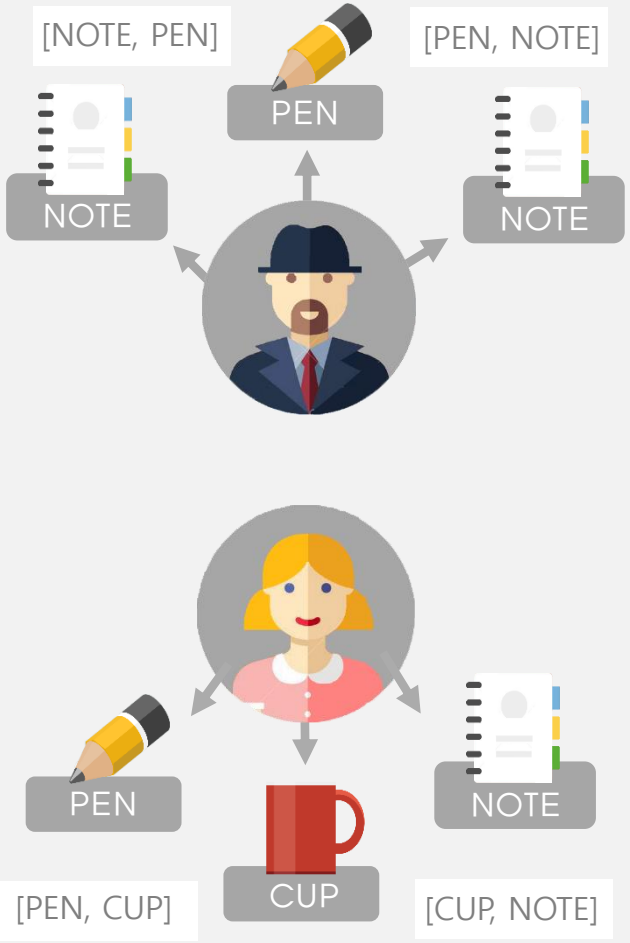
- 사용자가 매번 검색하는 검색어 분석
- 사용자가 매번 클릭하는 광고의 카운트
- 웹페이지와 광고의 검색어를 매핑하는 과정

모든 데이터가 구글에게는 분석해야 하는 도구이므로
대규모 데이터를 처리할 수 있는 병렬 처리 알고리즘이 필요!

★ **MySQL** (Google Adwords의 초기시스템) → **GFS** (Google판 Hadoop)

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

Amazon "Recommendation for You"



- 개인별 성향 비슷한 사용자 그룹을 분석
  - 상품 대 상품 (Item-to-Item) 관계를 이용하여 추천하는 방식
  - 대략 하루에 10만개의 서로 다른 상품이 팔린다 가정한다면..
    Item-to-Item 를 계산하기 위해 **10만 x 10만 =100억** 만큼의 데이터 필요

빅데이터가 될 수밖에 없다!

Item to Item

	CUP	NOTE	PEN
CUP	0	1	1
NOTE	1	0	2
PEN	1	2	0

CUP=[0,1,1] NOTE=[1,0,2] PEN=[1,2,0]

CUP - Amazon → NOTE, PEN

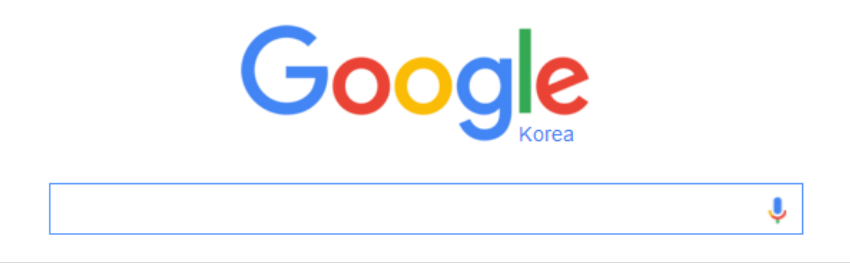
NOTE - Amazon → PEN

PEN - Amazon → NOTE

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

amazon Google

다른 종류의 기업정보 시스템이 필요하다

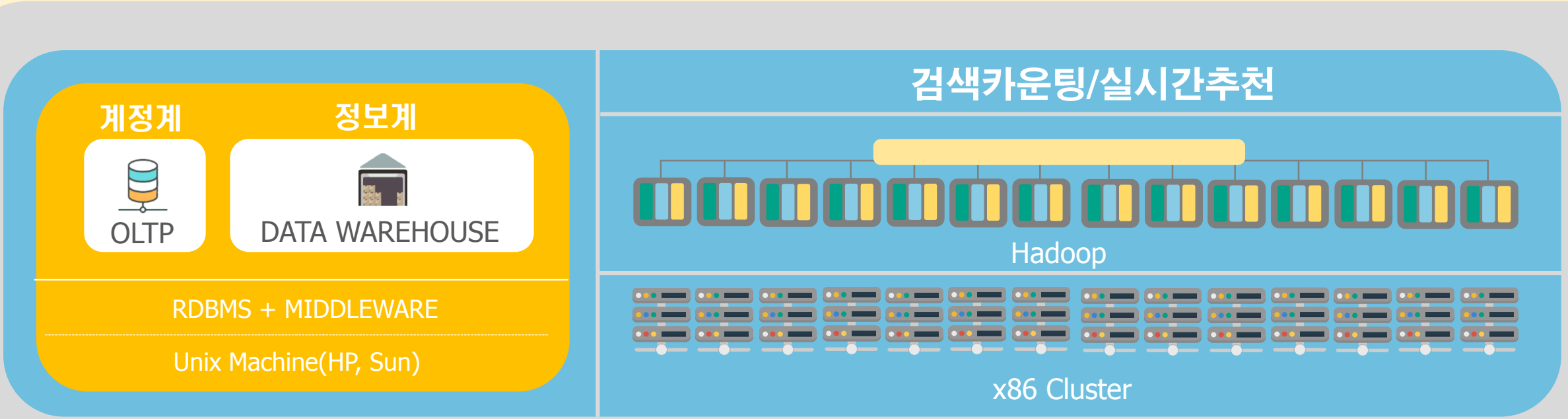


- **검색 카운팅, 검색, 광고와 실시간 추천**
  - **검색**이라는 인터넷의 가장 기본적 행위가 **정보시스템의 진화**
  - **실시간 검색과 추천**
    - 넷플릭스와 아마존의 가장 큰 차별적인 경쟁력
- 기존의 정보시스템이 **처리하기 힘든 수준을 요구**

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

TIP!!

amazon Google



구글이나 아마존도 기업의 전형적인 업무에 대해서는 기존 정보시스템을 사용한다.

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

역사는 반복되는 것인가?

통합 → 분산 → 통합

이 통합의 의미가 다르다!

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

통합

IBM
메인프레임
(중앙집중형)

독재정부

분산

기업정보시스템

계정계

OLTP

정보계

DATA WAREHOUSE

RDBMS + MIDDLEWARE

Unix Machine(HP, Sun)

연립정부

통합

CLOUD (Google, Amazon)

Hadoop

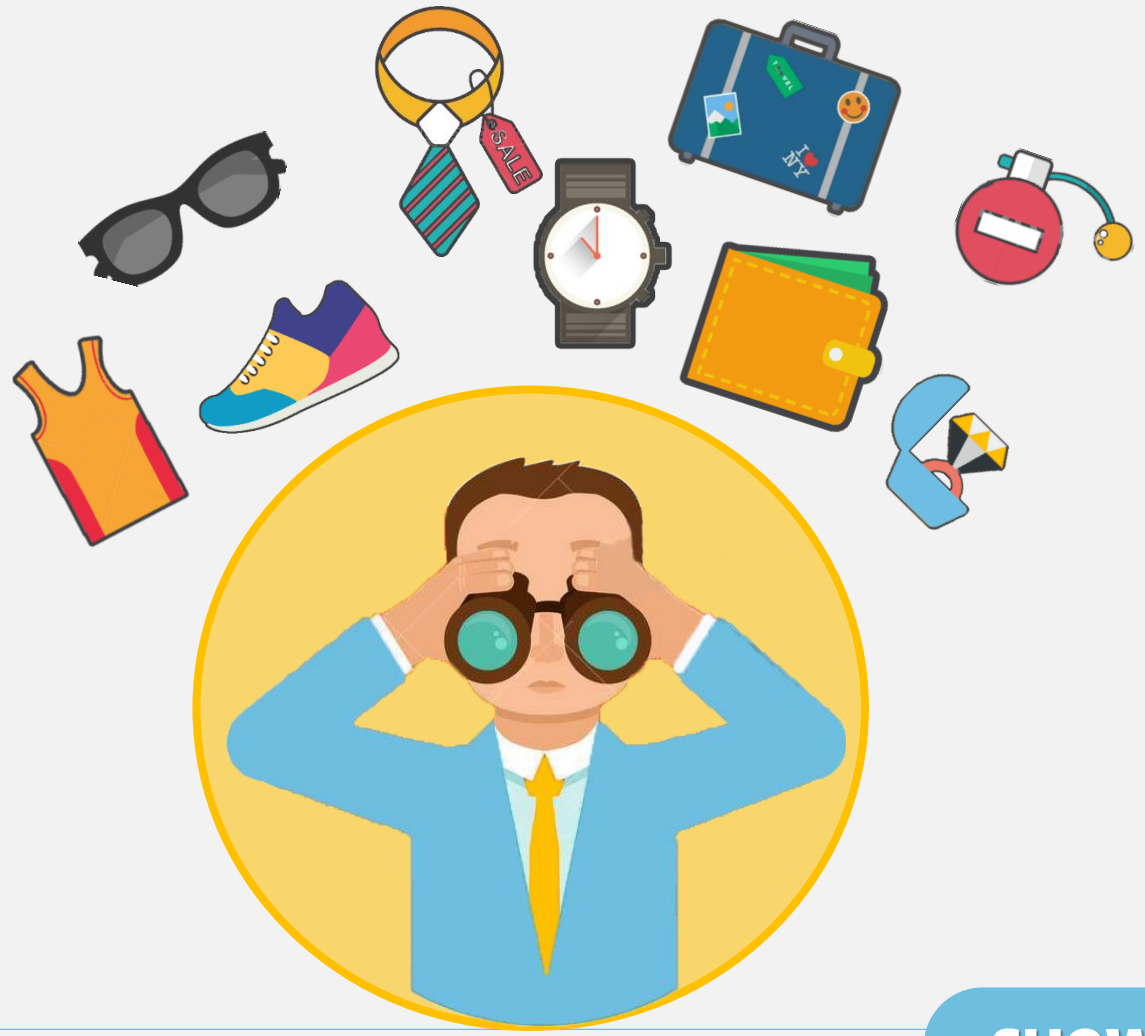
x86 Cluster

민주통합정부

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

2000년대
후반 이후

진짜 빅데이터의 등장



백화점에서 둘러보고
모바일에서 구매하고

SHOW ROOMING

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

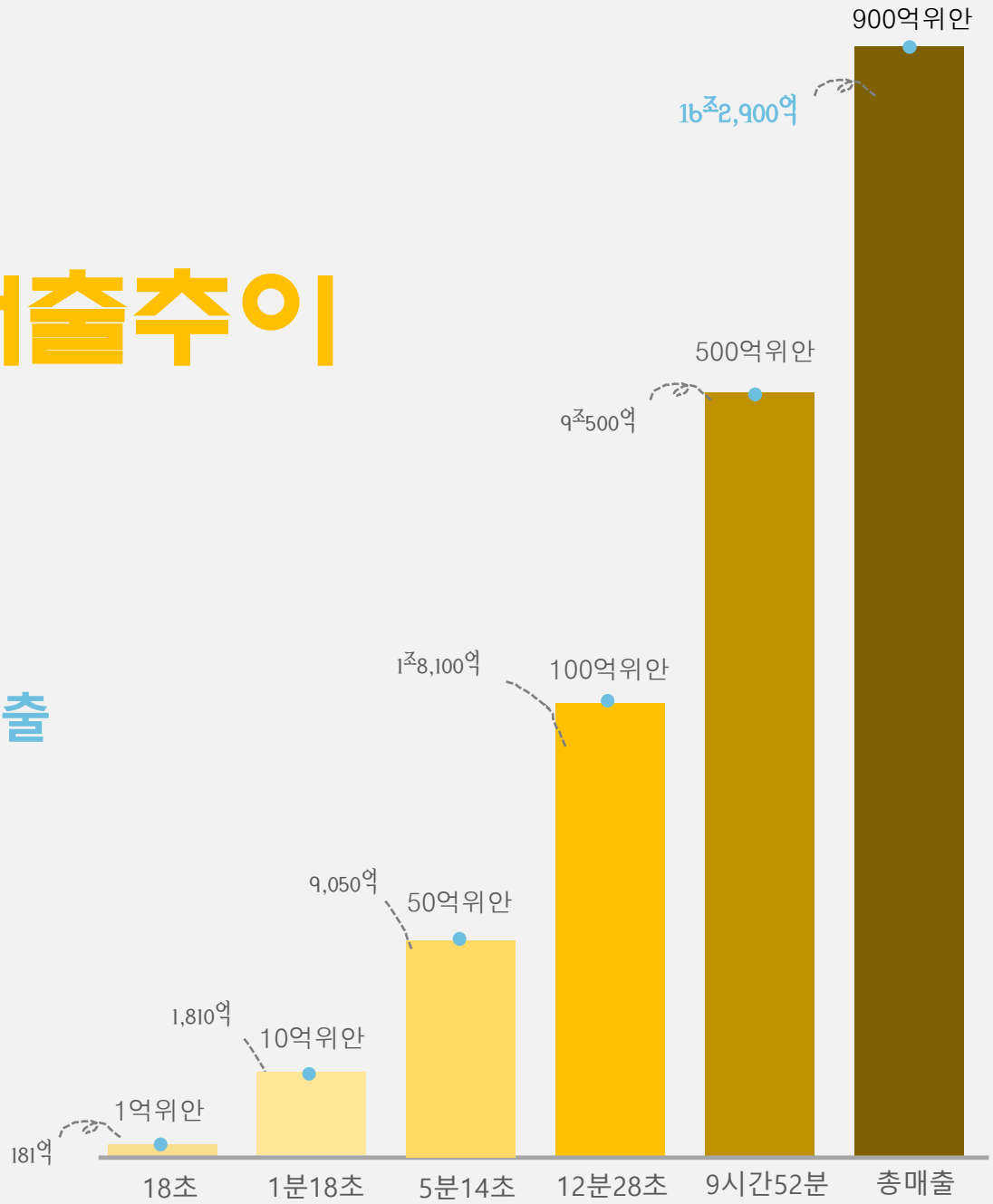
BEACON

백화점에 도착하면
내가 즐겨찾는 곳의 쿠폰이 바로 온다

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

광군절 (11월 11일), Alibaba의 단 하루! 매출추이

- 구매 중 **71%**는 모바일로
- AMAZON **1분기 매출** = Alibaba **하루 매출**
- 서버를 감당하기 위해 **클라우드** 필요



Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

진짜 빅데이터의 등장

2000년대 후반 이후

더 커지는 빅데이터

소셜미디어 + 모바일

=

- 실시간
- 검색, 추천
- 개인별최적화

BUSINESS INSIGHT

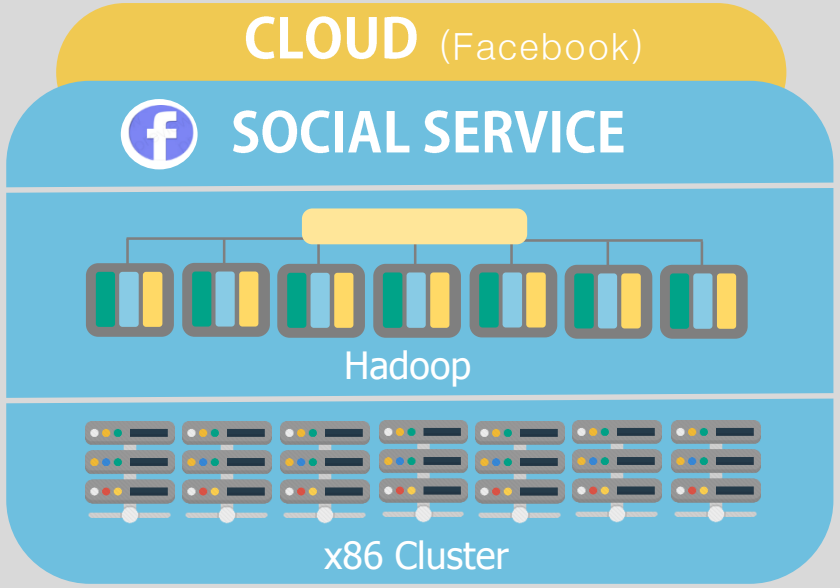
Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

진짜 빅데이터의 등장

2000년대 후반 이후

" 현재 전체 facebook 가입자는 **14억명** ,이고

월간 활동 사용자는 **10억명** 이다."



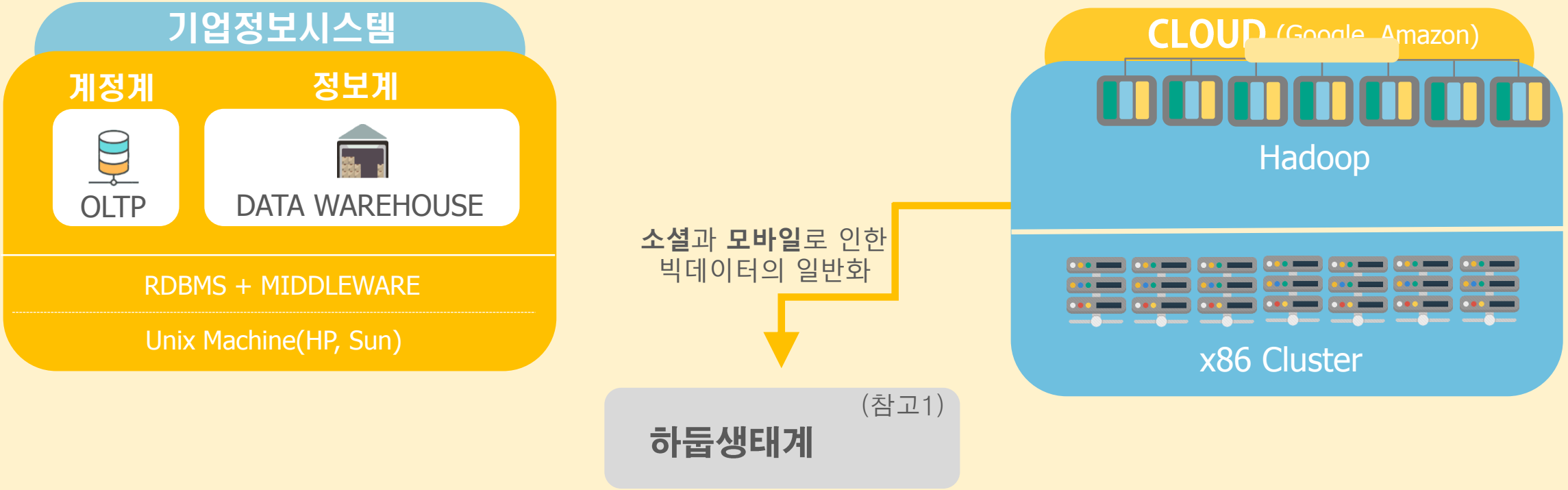
실시간으로 콘텐츠 처리 필요!!

facebook 은 클라우드의 **최상**클래스

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

진짜 빅데이터의 등장

2000년대 후반 이후



구글과 아마존을 넘어 이제 **모든 기업에게 빅데이터처리**를 요구하고 있다

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

더 더 더 ! 커지는 빅데이터

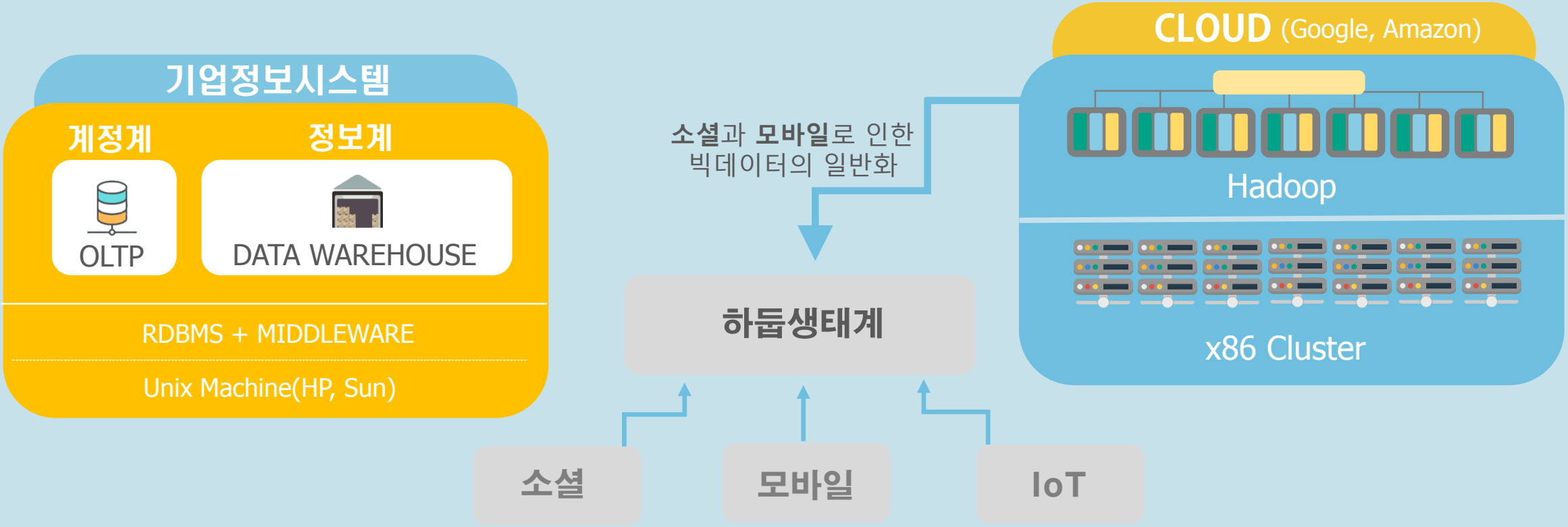
소셜미디어 + 모바일 + 사물인터넷

클라우드 = DIGITIZING BUSINESS 종결자

빅데이터의 가속화

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

6. 소셜+모바일+사물인터넷+클라우드+빅데이터 =digitizing

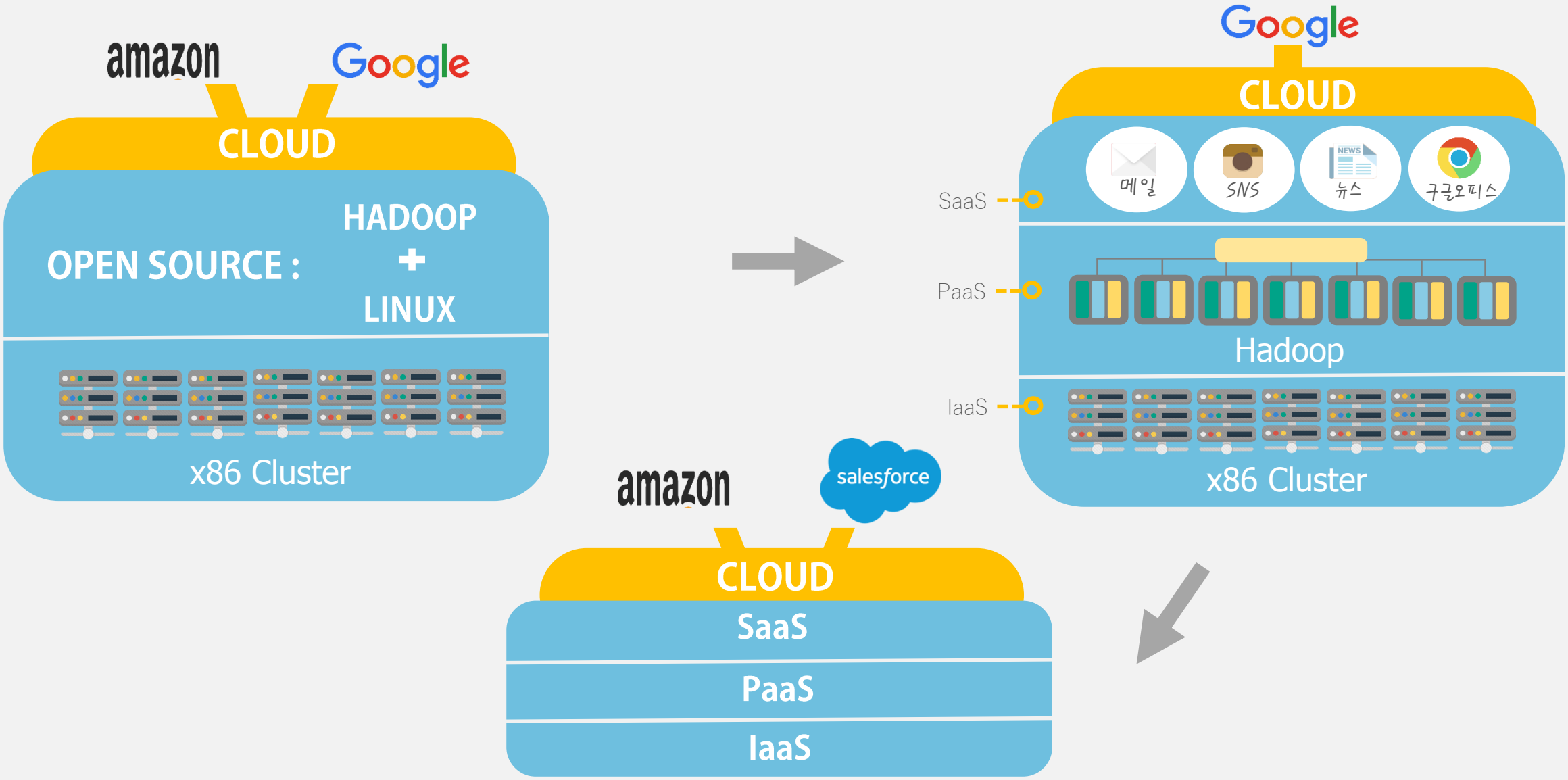


빅데이터의 가속화

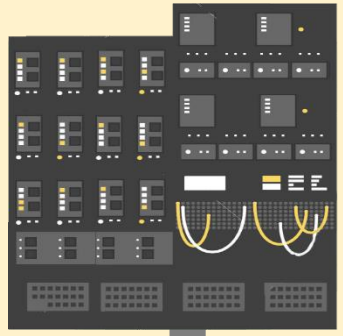
실시간으로 쏟아지는 빅데이터와
그것의 처리가능성은 인공지능의 실현을 이끈다.

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

7. 클라우드의 수익모델로의 변화 (IT자원의 서비스화)



Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.



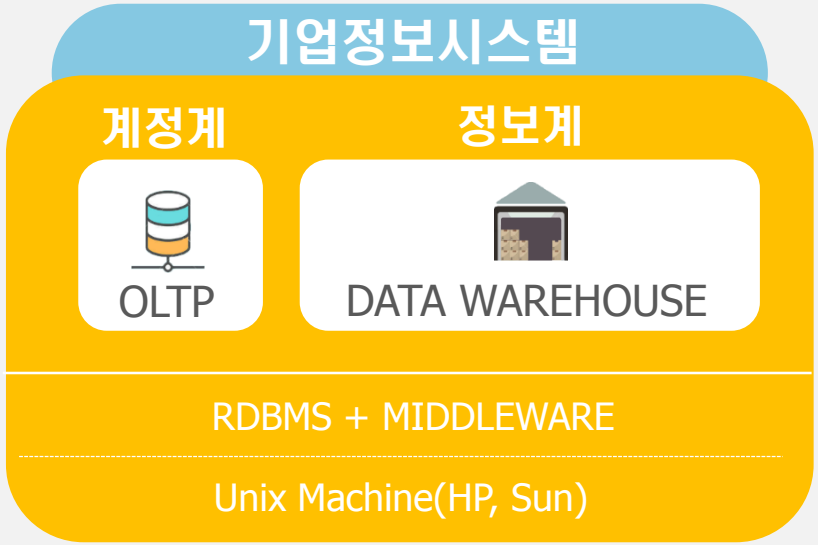
인프라

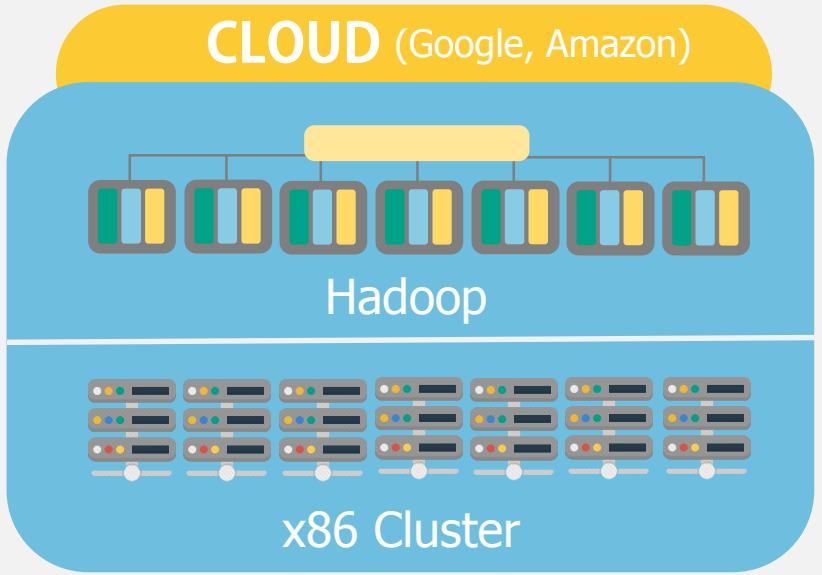
데이터

사업

클라우드를 평가하다

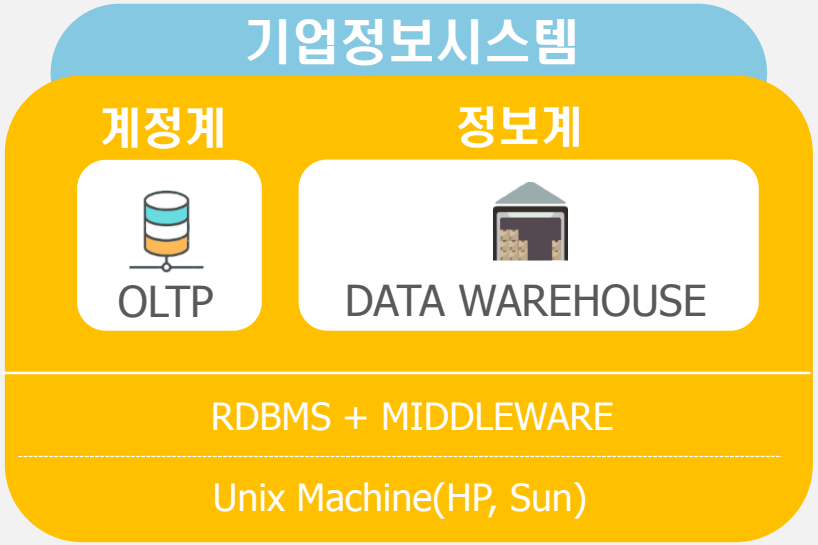
Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

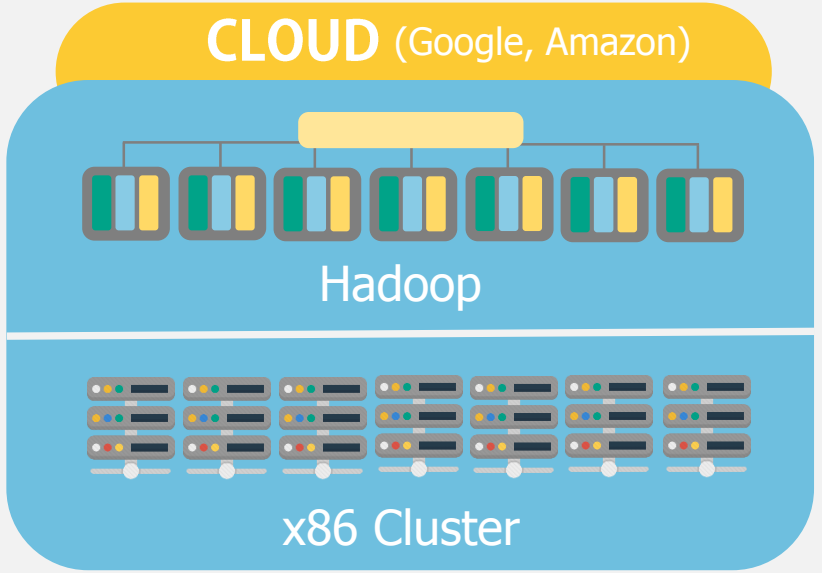




1990년대 중반 이후 **다운사이징**의 표준화된 아키텍처	2000년대 이후 **인터넷 활황시대**의 새로운 아키텍처
유닉스서버와 **상업용 소프트웨어** 패키지 전성시대	**X86**급 PC와 **오픈 소스** 전성시대
HP, 오라클의 전성시대	**구글, 아마존**의 전성시대
유닉스-**오라클/웹로직**	리눅스-**하둡생태계**
Scale up 강점	**Scale out 강점**
병렬처리의 한계	**병렬처리기술, 가상화기술**

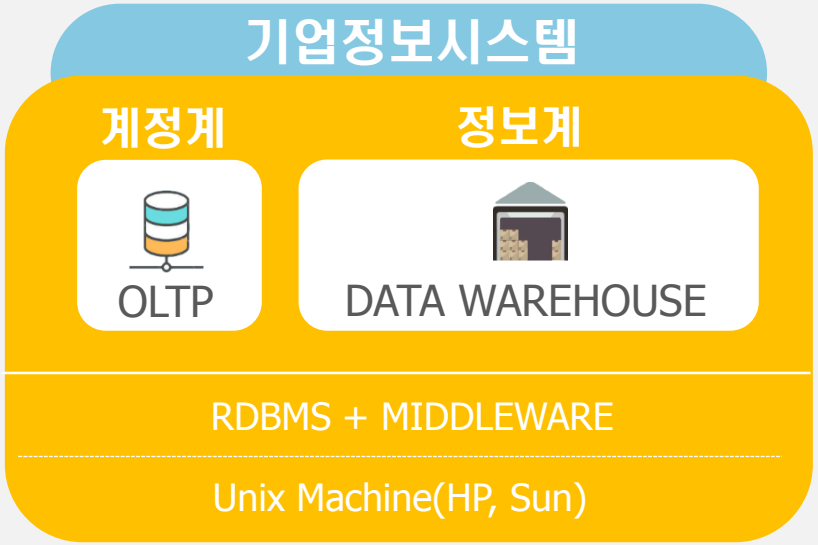
Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

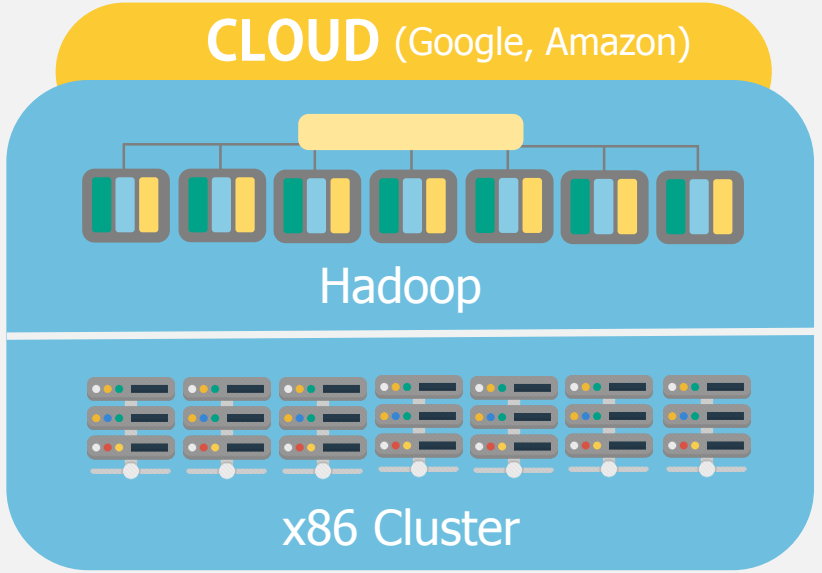




1990년대 중반 이후 **다운사이징**의 표준화된 아키텍처	2000년대 이후 **인터넷 활황시대**의 새로운 아키텍처
정형데이터 처리	**정형/비정형데이터** 처리
정형 데이터의 **논리적 정합성** 처리에 집중	**정형/비정형 데이터**의 **폭발적 증가 처리**가 관건
VLDB	**빅데이터**

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.





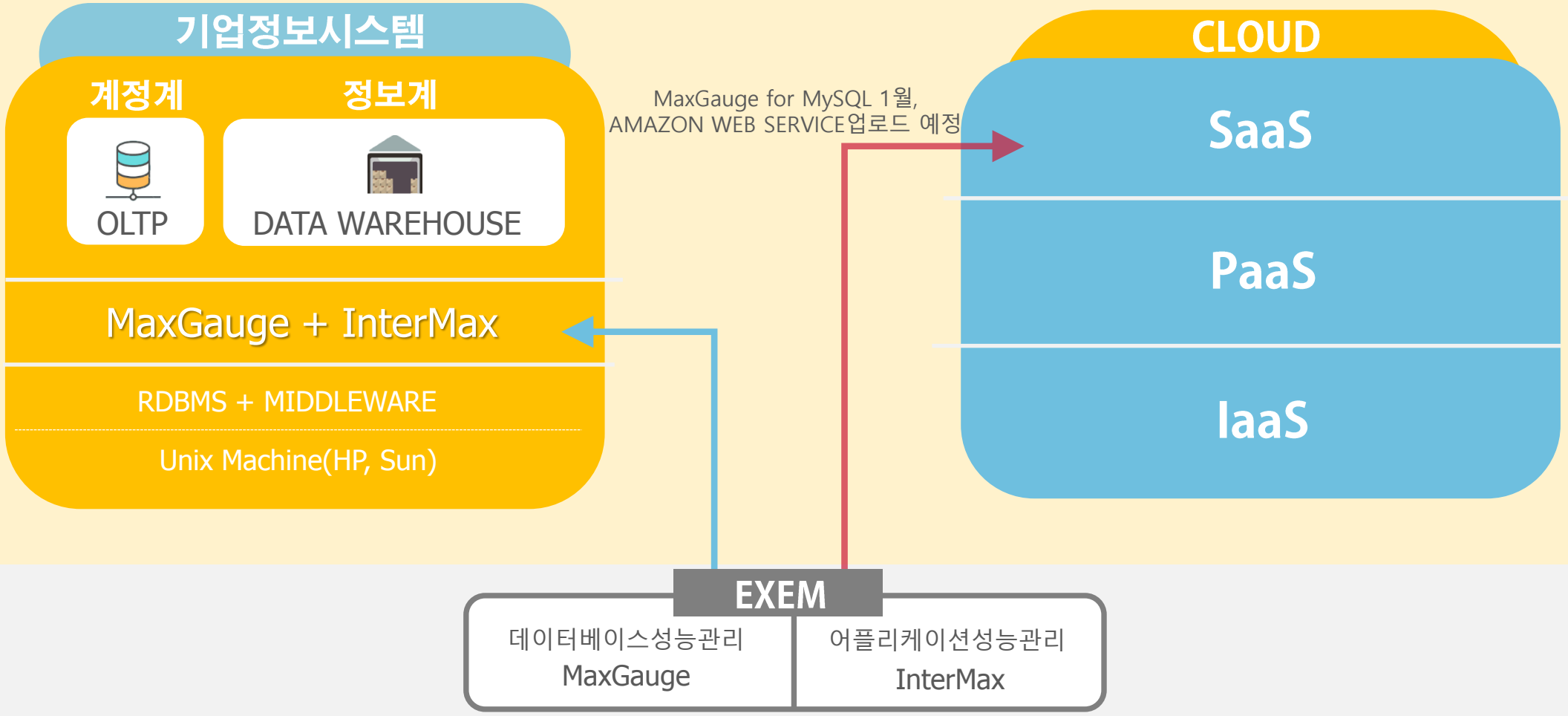
1990년대 중반 이후 **다운사이징**의 표준화된 아키텍처	2000년대 이후 **인터넷 활황시대**의 새로운 아키텍처
소프트웨어 라이선스 모델	**서비스 라이선스** 모델
고정비용모델	**변동비용모델**
정보시스템이 곧 사업프로세스	시뮬레이션의 정교화를 통해 **사업혁신** 프로세스
제조, 금융	**E커머스, 포탈**

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

엑셈 연합군

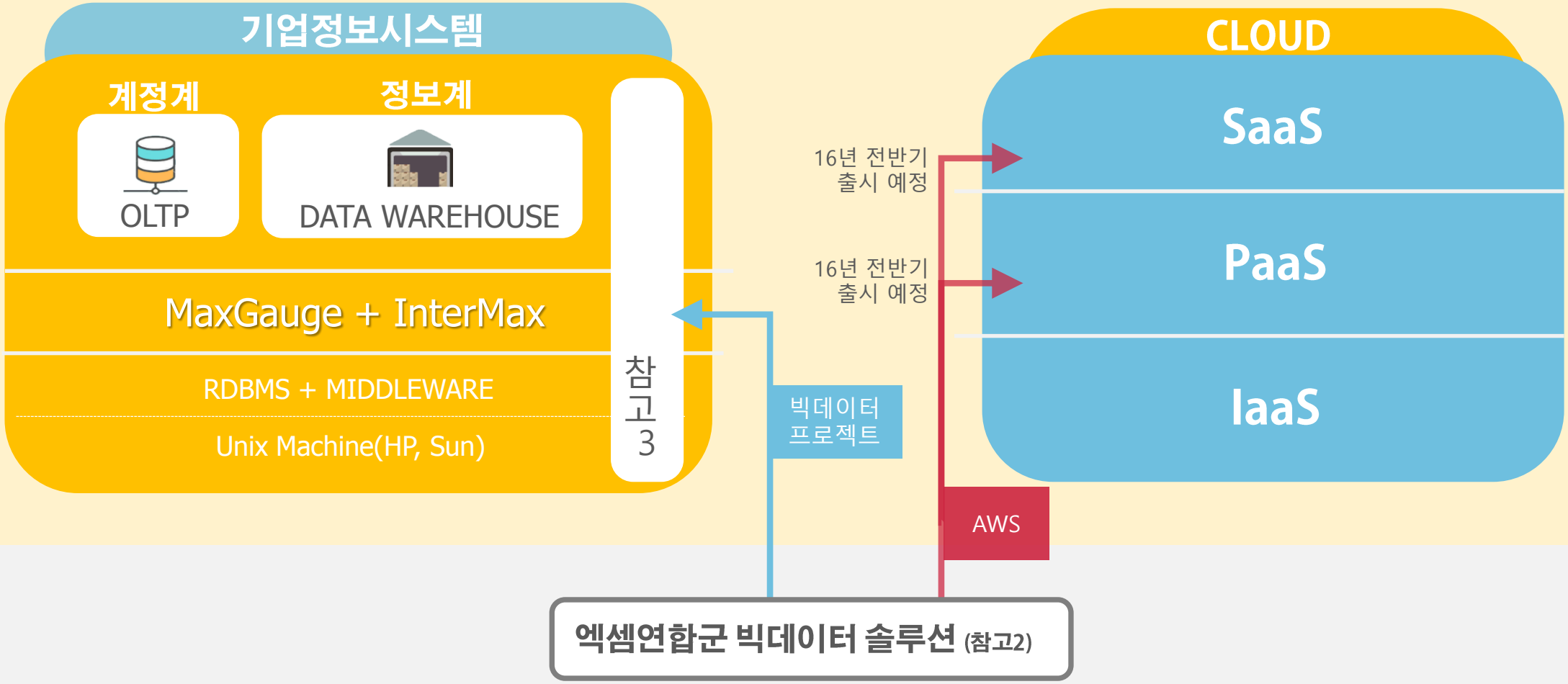
빅데이터와 클라우드 대응

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.



해외 사업을 기존의 라이선스 방식뿐 아니라 **서비스 방식**으로 확장

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.



Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

단기적 경쟁력

- **빅데이터 프로젝트** 참여
  (엑셈연합군빅데이터솔루션)
  엑셈연합군의 빅데이터솔루션의 완성도를 높인다
- **하둡생태계**특정영역최적화 (OPENSIGHT)
  하둡운영/관리 (플라밍고)

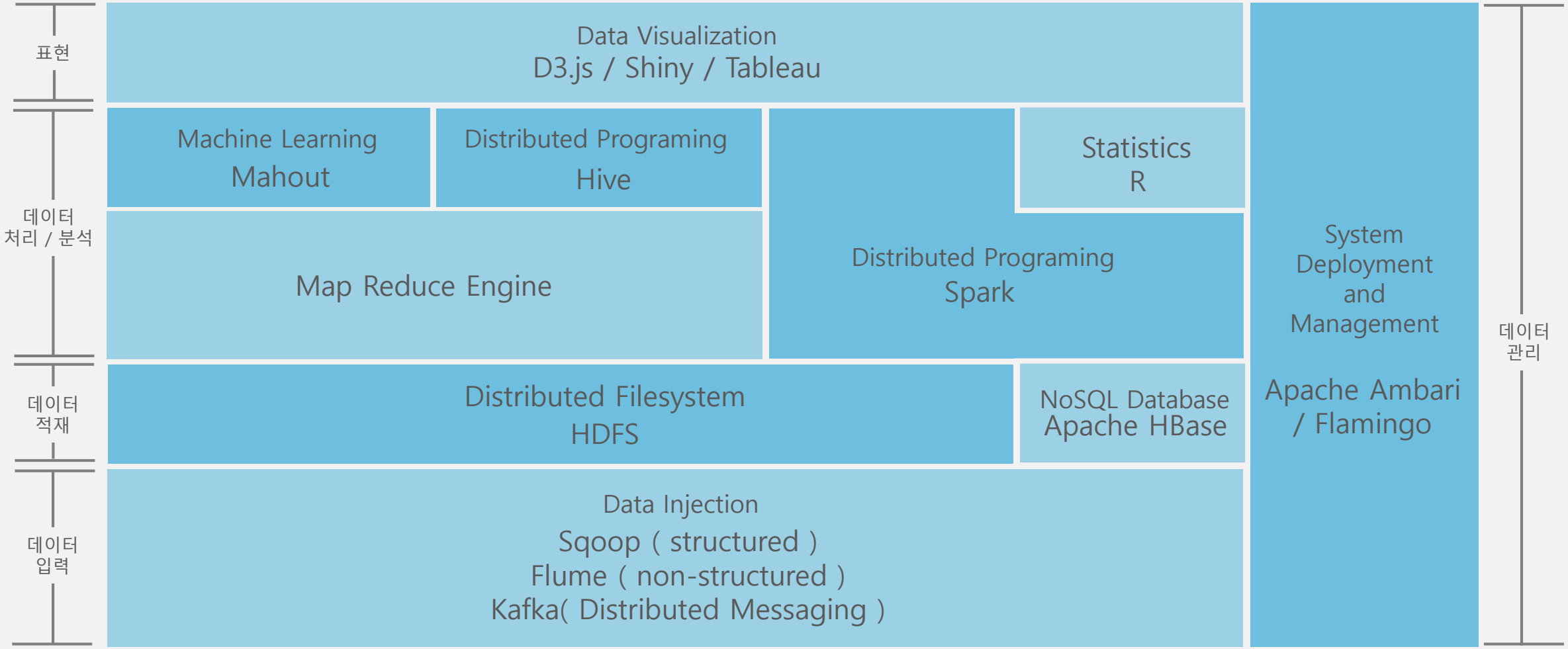
장기적 경쟁력

기술은 범용화의 길을 간다.
하둡생태계의 메인도 결국 범용화된다.
차별적 경쟁력은 무엇인가?

- 결국 **관리**와 **운영**의 효율성이 남는다
- 종합적인 **빅데이터컨설팅** 능력
- 특정 버티컬 영역에 **특화된 분석알고리즘**
  - 자연어처리/이미지처리/관계추출/패턴인식/자동주행……

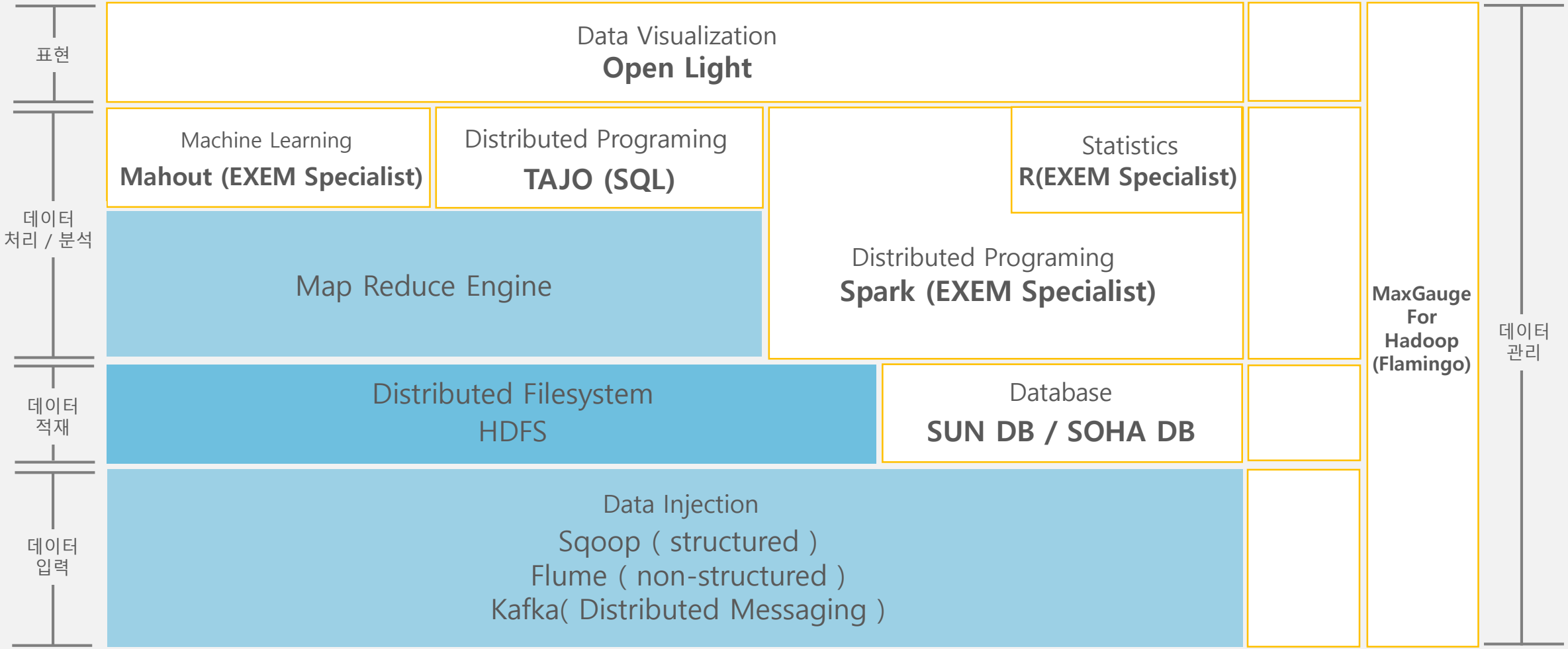
Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

하둡생태계



Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

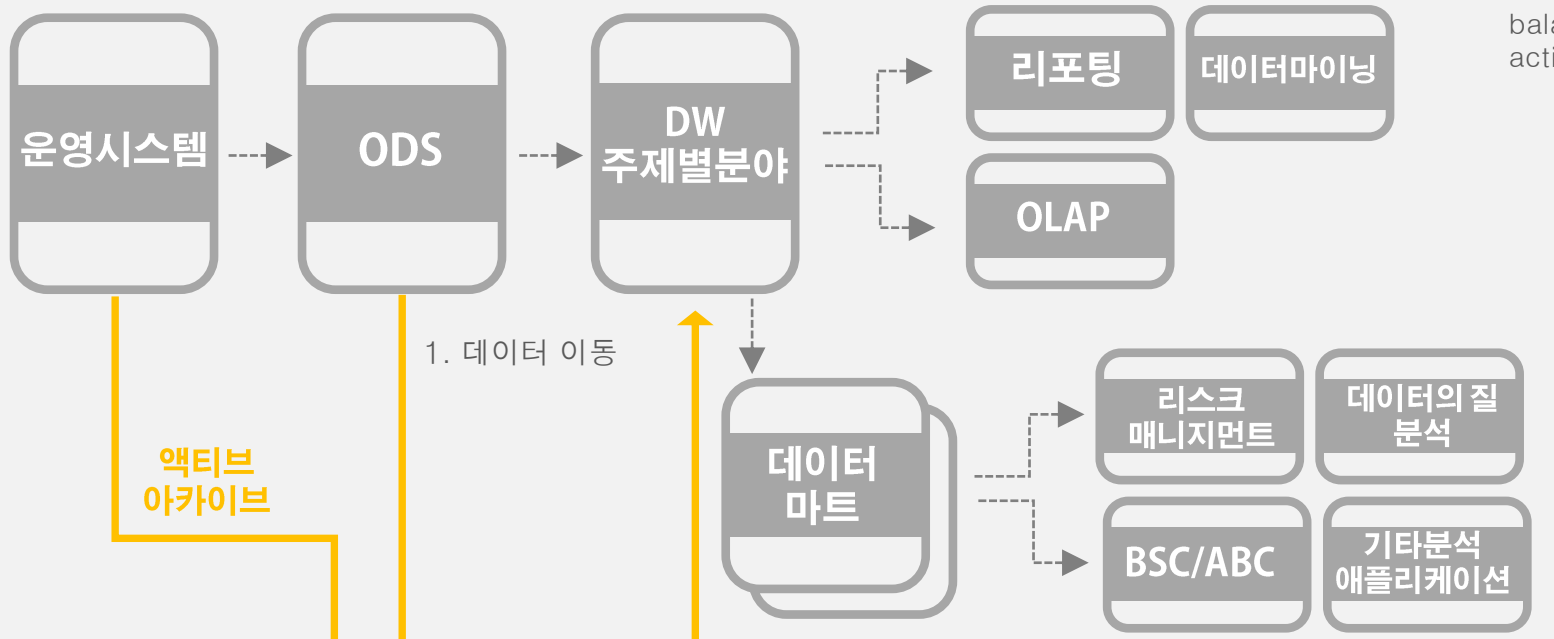
EXEM 연합군 하둡생태계



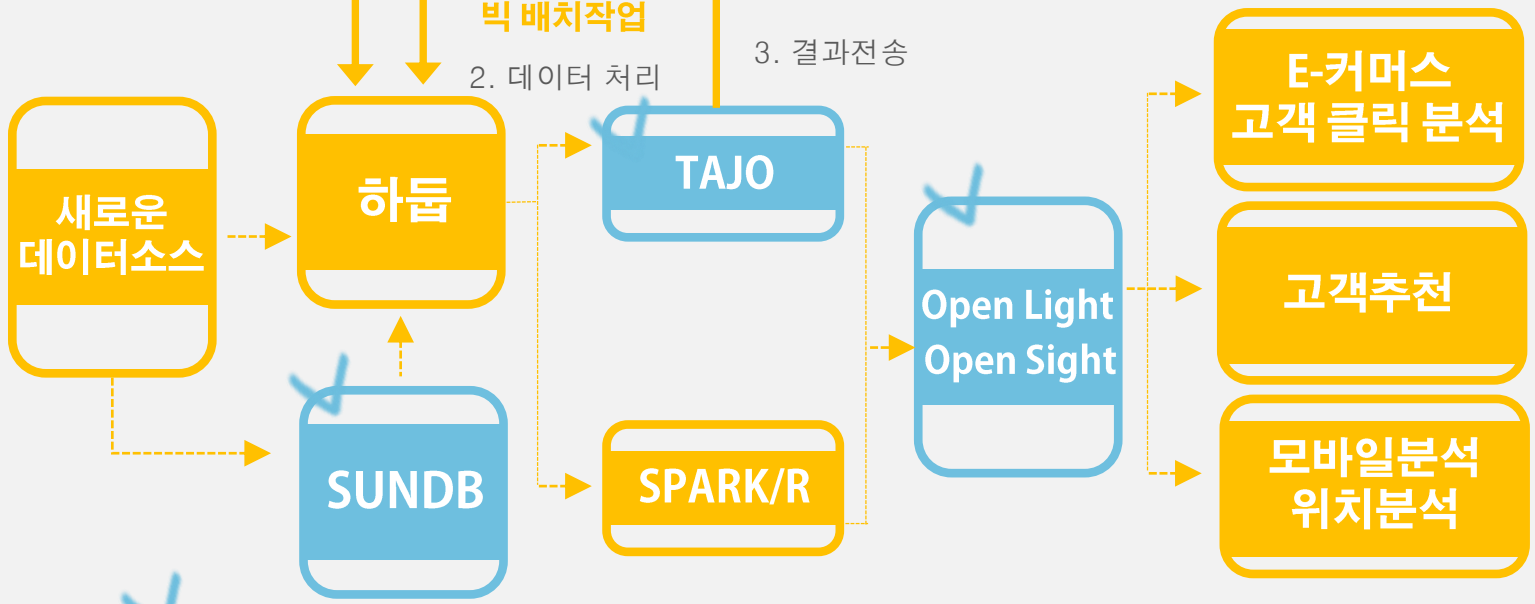
Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

balanced scorecard (BSC)
activity-based costing (ABC)

기존 DW/BI 시스템



신규 빅데이터 시스템



Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

질문 ?

1. 기업 정보시스템의 변화를 가능하게 한 가장 중요한 요소는?
2. 클라우드가 탄생한 이유는? = 빅데이터의 효시는?
3. 3대 정보시스템(데이터센터)의 탄생 이유와 차이를 설명한다면?
4. 최근 빅데이터의 3대 소스는?
5. 빅데이터 처리 시스템으로서 클라우드를 가능하게 한 가장 중요한 기술 혁신 두 가지는?
6. RDBMS 와 Hadoop의 차이는?
7. 클라우드 서비스 모델 세가지는? 그 각각의 의미는?
8. 클라우드는 IT자원을 서비스로 판매하기 위한 데이터센터이다. ○ or X
9. 클라우드는 빅데이터 처리를 위해 탄생한 데이터센터이다. ○ or X
10. 클라우드가 개인/기업에게 가지는 의미는 무엇인가?

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.

THANK YOU!! :)

Copyrights 2001~2015, EXEM CO., LTD. All Rights Reserved.