


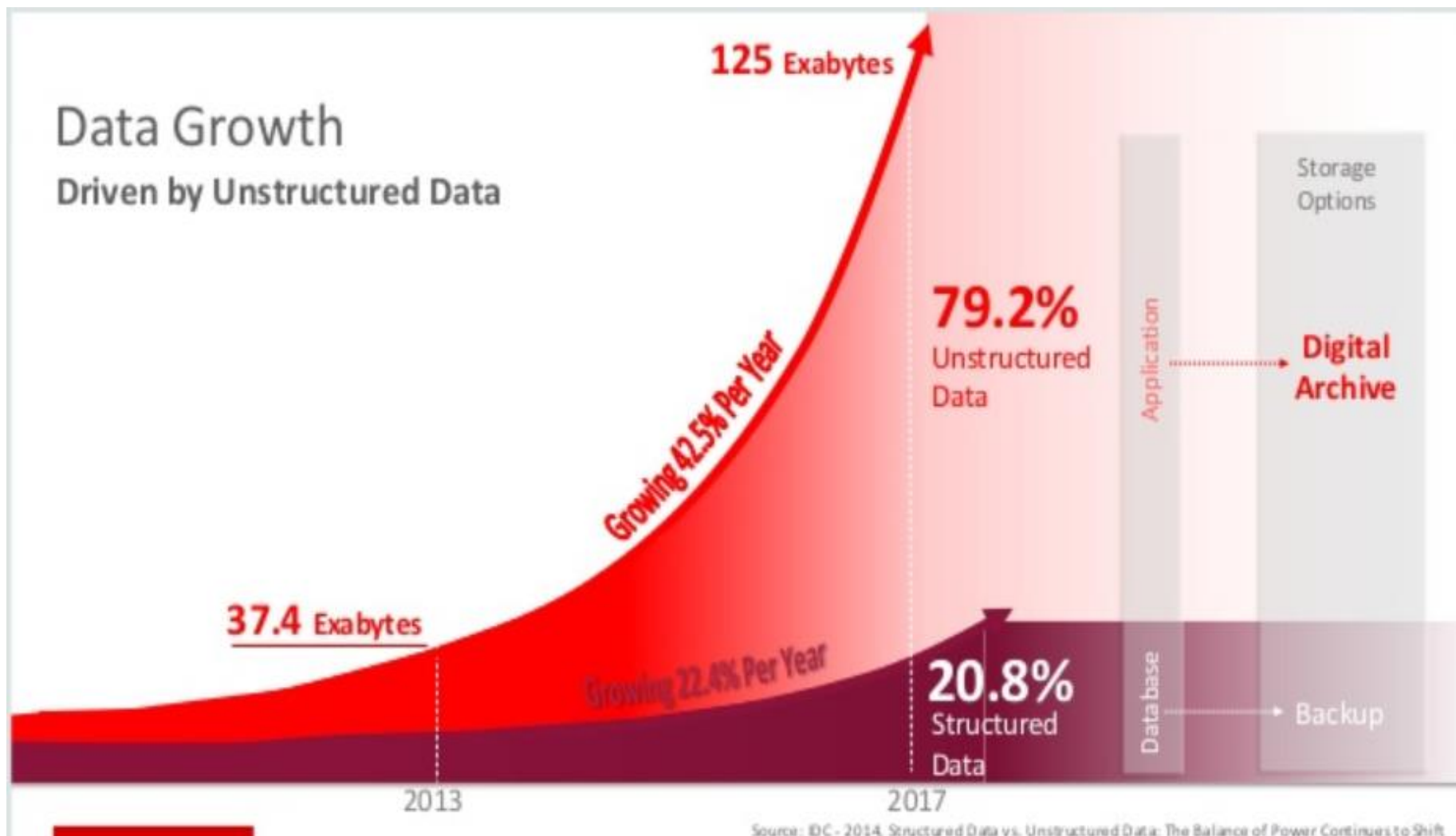


# 데이터, 어떻게 활용 할 것인가?

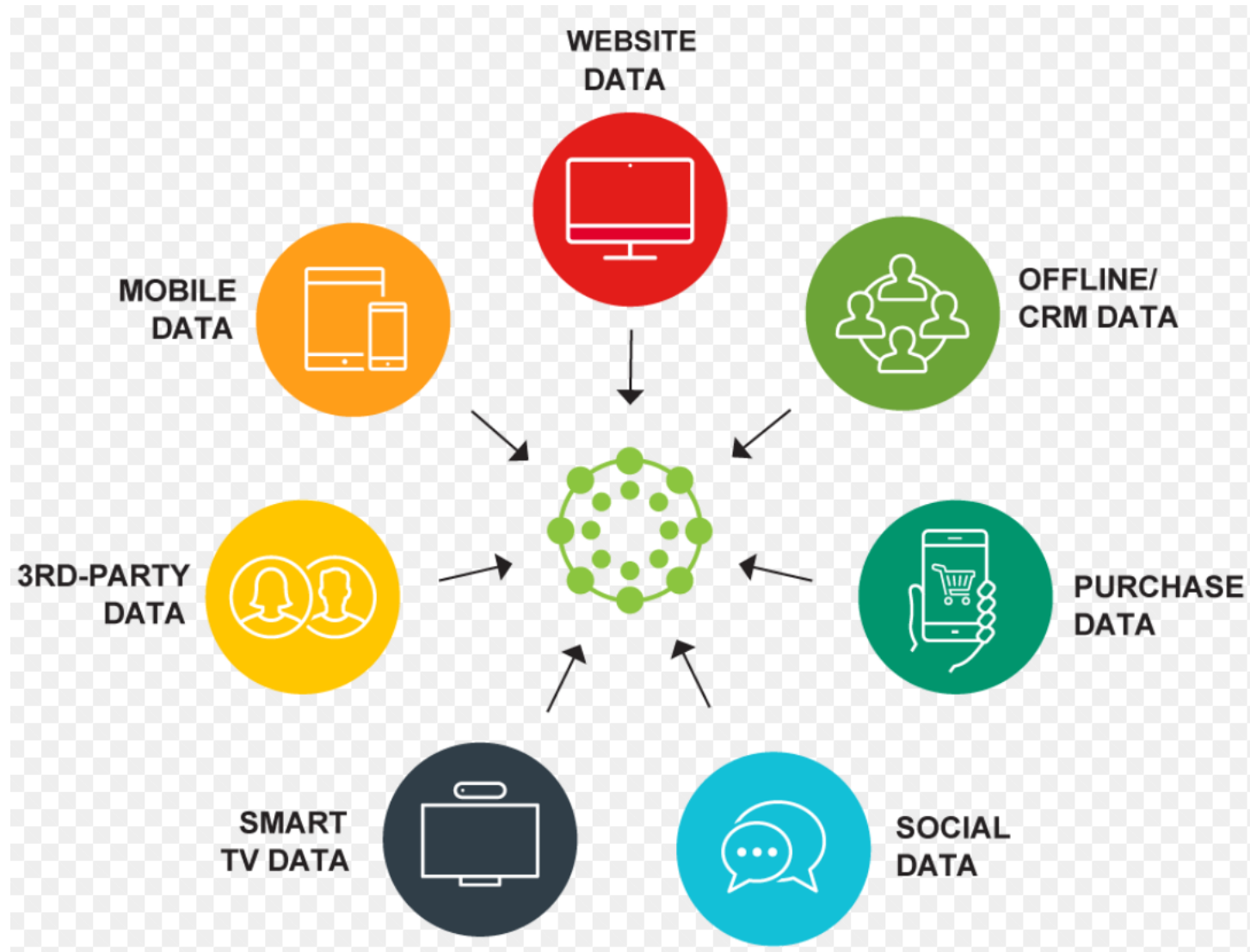
EN-CORE Inc. is "Total Data Service Provider." We, as one of the leading data scientists groups in Korea, are offering a wide-range of services such as consulting, education, solutions, and technology information based on data ranging from planning the corporation IT systems to analyzing, designing, developing, transitioning, managing, performing governance, and data servicing.

- 
1. Data란
  2. Data 와 IT
  3. 데이터 가공 플랫폼
  4. 데이터 활용 전략

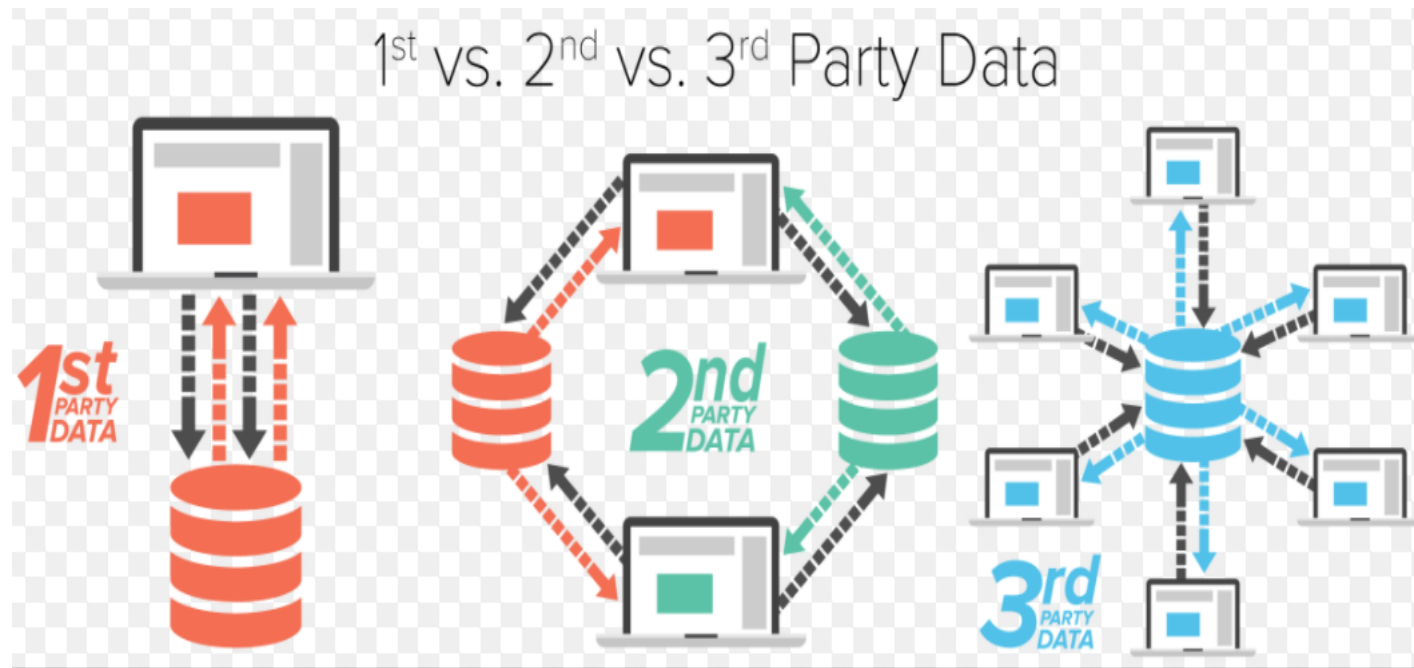
# Data의 성장



# Source of Data



# Data 종류





## 데이터 활용 사례



# 데이터 활용 사례

## How two names and a sheaf of newspaper cuttings revealed the 9/11 team

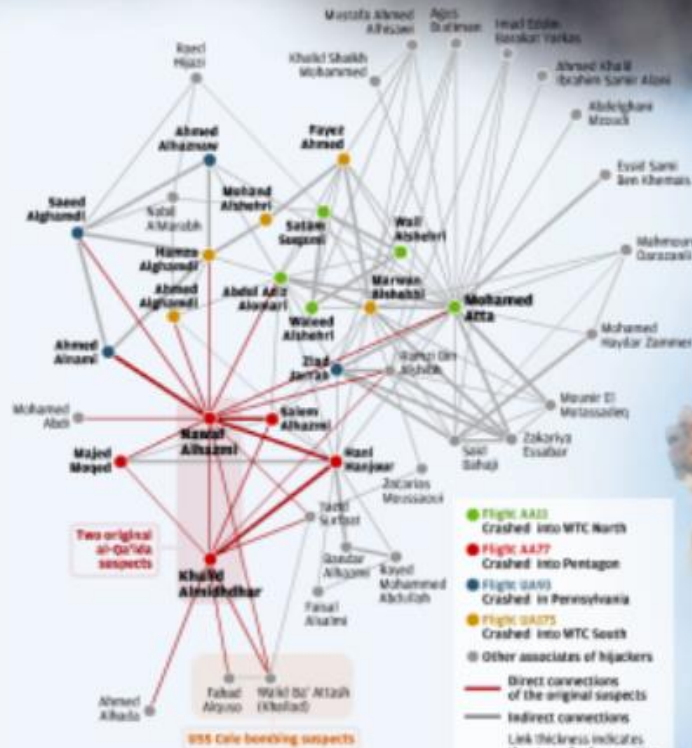
This social network of the 19 hijackers behind the 9/11 attacks in the United States, and their associates, was drawn up at the end of 2001. Valdis Krebs, a commercial consultant in network analysis, started with newspaper reports of the two original terrorist suspects, Nawaf Alhazmi and Khalid Almihdhar. He then plotted the position of the other hijackers and associates. His analysis highlighted the central role played by Mohamed Atta. It also shows the close associations between the "Hamburg cell" that Atta set up, as well as the close links with the two original suspects - critical information that may have helped to avert an attack had it been known.



Mohamed Atta, Nawaf Alhazmi, Khalid Almihdhar



Emergency services attend the scene after Flight AA77 crashes into the Pentagon



The second plane, Flight UA175, crashes into the South Tower of the World Trade Centre



## Suspected Hijackers

### American Airlines Flight 77

8:10 a.m. departed Washington Dulles for Los Angeles  
9:39 a.m. crashed into the Pentagon



**Khalid Almihdhar**  
Possible Saudi national

Possible resident of San Diego, California, and New York

Alias: Sannan Al-Makki; Khalid Bin Muhammad; 'Addallah Al-Mihdhar; Khalid Mohammad Al-Saqaf



**Majed Moqed**  
Possible Saudi national

Alias: Majed M.GH Moqed; Majed Mashaan Moqed



**Nawaf Alhazmi**  
Possible Saudi national

Possible resident of Fort Lee, New Jersey; Wayne, New Jersey; San Diego, California

Alias: Nawaf Al-Hazmi; Nawaf M.S. Al Hazmi



**Salem Alhazmi**  
Possible Saudi national

Possible resident of Fort Lee, New Jersey; Wayne, New Jersey

Alias: Hani Saleh Hanjour; Hani Saleh; Hani Hanjour, Hani Saleh H. Hanjour



**Hani Hanjour**  
Possible resident of Phoenix, Arizona, and San Diego, California

Alias: Hani Saleh Hanjour; Hani Saleh; Hani Hanjour, Hani Saleh H. Hanjour

Alias: Hani Saleh Hanjour; Hani Saleh; Hani Hanjour, Hani Saleh H. Hanjour

### United Airlines Flight 93

8:01 a.m. departed Newark, New Jersey for San Francisco  
10:10 a.m. crashed in Stony Creek Township, Pennsylvania



**Saeed Alghamdi**  
Possible residence: Delray Beach, Florida

Alias: Abdul Rahman Saeed Alghamdi; Ali S Alghamdi; Al-Gamdi; Saad M.S. Al Ghamdi; Saeed Al Ghamdi; Saheed Al-Ghamdi; Seed Al Ghamdi



**Ahmed Al Haznawi**  
Possible Saudi national

Date of birth used: October 11, 1980

Possible residence: Delray Beach, Florida

Alias: Ahmed Alhaznawi



**Ahmed Alnami**  
Possible residence: Delray Beach, Florida

Alias: Ali Ahmed Alnami; Ahmed A. Al-Nami; Ahmed Al-Nawi



**Ziad Samir Jarrah**  
Believed to be a pilot

Alias: Zaid Jarrah; Zaid Samir Jarrah; Ziad S. Jarrah; Ziad Jarrah; Jarrah, Ziad Samir Jarrah

### American Airlines Flight 11

7:45 a.m. departed Boston for Los Angeles  
8:45 a.m. crashed into North Tower of the World Trade Center



**Satam Al Suqami**  
Possible Saudi national

Dates of birth used: June 28, 1976; Last known address: United Arab Emirates



**Walid M. Alshehri**  
Possible Saudi national

Dates of birth used: September 13, 1974; January 1, 1976; March 3, 1976; July 8, 1977; December 20, 1978; May 11, 1979; November 5, 1979

Possible residence(s): Hollywood, Florida; Orlando, Florida; Daytona Beach, Florida

Believed to be a pilot



**Wail M. Alshehri**  
Date of birth used: September 1, 1968

Possible residence(s): Hollywood, Florida; Newton, Massachusetts

Believed to be a pilot



**Abdulaziz Alomari**  
Possible Saudi national

Dates of birth used: December 24, 1972 and May 28, 1979

Possible residence(s): Hollywood, Florida

Believed to be a pilot



**Mohamed Atta**  
Possible Egyptian national

Date of birth used: September 1, 1968

Possible residence(s): Hollywood, Florida; Coral Springs, Florida; Hamburg, Germany

Believed to be a pilot

Alias: Mehan Atta; Mohammad El Amir; Muhammad Atta; Mohamed El Sayed; Mohamed Elsayed

### United Airlines Flight 175

7:58 a.m. departed Boston for Los Angeles  
9:05 a.m. crashed into South Tower of the World Trade Center



**Marwan Al-Shehhi**  
Date of birth used: May 9, 1978

Possible residence(s): Hollywood, Florida

Believed to be a pilot

Alias: Marwan Yusuf Muhammad Rashid Al-Shehhi; Marwan Yusuf Muhammad Rashid Lakrab Al-Shehhi; Abu Abdullah



**Fayed Banihammad**  
Possible residence(s): Delray Beach, Florida

Alias: Fayed Ahmad; Banihammad Fayed Abu Dhabi

Fayed Rashid Ahmad; Banihammad Fayed; Rasid Ahmed Hassen Alqadi; Abu Dhabi Banihammad Ahmed Fayed; Faez Ahmed



**Ahmed Alghamdi**  
Alias: Ahmed Salah Alghamdi



**Hamza Alghamdi**  
Possible residence(s): Delray Beach, Florida

Alias: Hamza Al-Ghamdi; Hamza Ghamdi; Hamzah Alghamdi; Hamza Alghamdi Saleh



**Mohand Alshehri**  
Possible residence(s): Delray Beach, Florida

Alias: Mohammed Alshehri; Mohand Alshehri; Mohaid Alshehri



Acxiom은 세계 최대 개인 정보 (개인 1인당 1,500가공정보와 1,000개의 추정정보:  
정형 빅 데이터) 보유하고 다국적 글로벌 회사에 데이터 판매와 컨설팅

## History of Acxiom Offerings (the portfolio grows)



Business Information Grid™
Data Factory
Integrated Marketing Services
Personix® International
Fraud Prevention & Risk Mitigation
Customer Information Infrastructure (CII)
Sentrix®
Personix®
Optix®
Acxiom Interactive Solutions
AbiliTec® & Interactive Delivery
Marketing Services & Consulting
Internet & Online
Data Warehousing
Data Access
Data Content
Data Center Outsourcing
List Enhancement
Marketing Databases
Mailing List Processing



현재 Acxiom 최고상품 개발 담당  
(2012년 4월 입사)  
전 구글 최고 상품 개발 담당  
전 마이로 소프트 상품 개발 담당

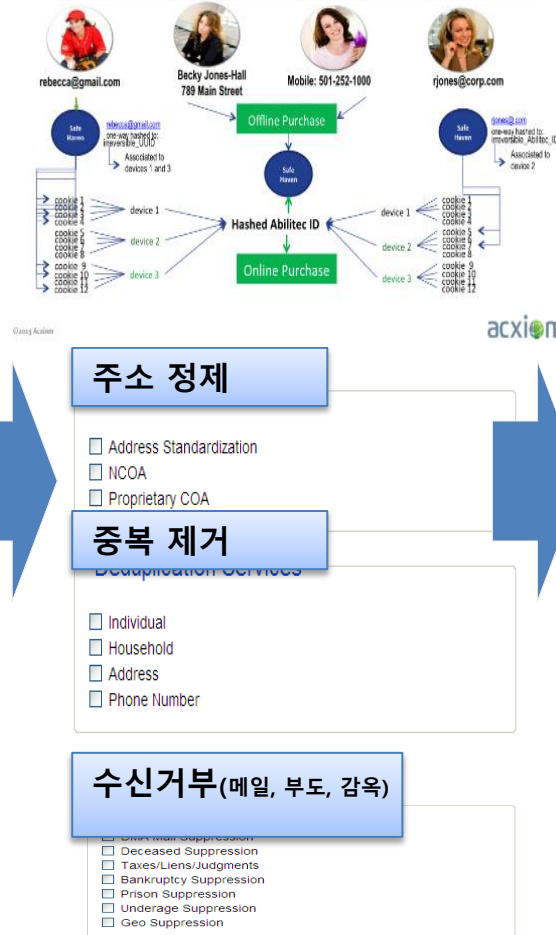
# 미국의 데이터 가공 및 브로커 산업 역할

데이터 수집, 가공, **데이터 인사이트** 필요(속도와 효율성 핵심 경쟁력)

## 1. 고객 소비 데이터 수집



## 2. 데이터 정제



## 3. 가공 데이터

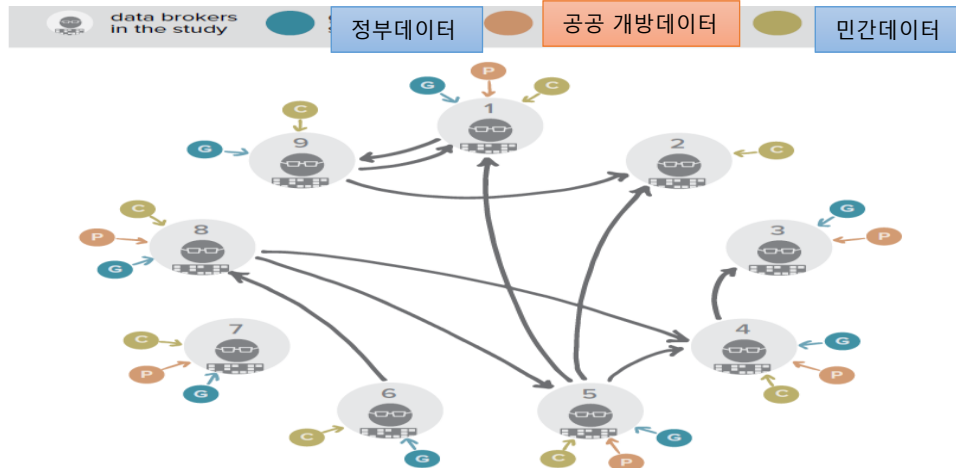
### 고객 세분화



### 고객 지수화

### 마케팅 관련 지수





## 원천 데이터 예시

check all    uncheck all

<input type="checkbox"/> Number of Adults	<input type="checkbox"/> Political Party - 2nd Individual	<input type="checkbox"/> Education - 1st Individual - 100 percent
<input type="checkbox"/> Household Size	<input type="checkbox"/> Political Party - 3rd Individual	<input type="checkbox"/> Presence of Children
<input type="checkbox"/> Marital Status in the Household	<input type="checkbox"/> Political Party - 4th Individual	<input type="checkbox"/> LifeStages Code
<input type="checkbox"/> Name / Gender - 1st Individual	<input type="checkbox"/> Political Party - 5th Individual	<input type="checkbox"/> Occupation - 3rd Individual
<input type="checkbox"/> Generations in Household	<input type="checkbox"/> Congressional District	<input type="checkbox"/> Occupation - 4th Individual
<input type="checkbox"/> Working Woman	<input type="checkbox"/> Smoker	<input type="checkbox"/> Occupation - 5th Individual

## 수집 데이터 카테고리

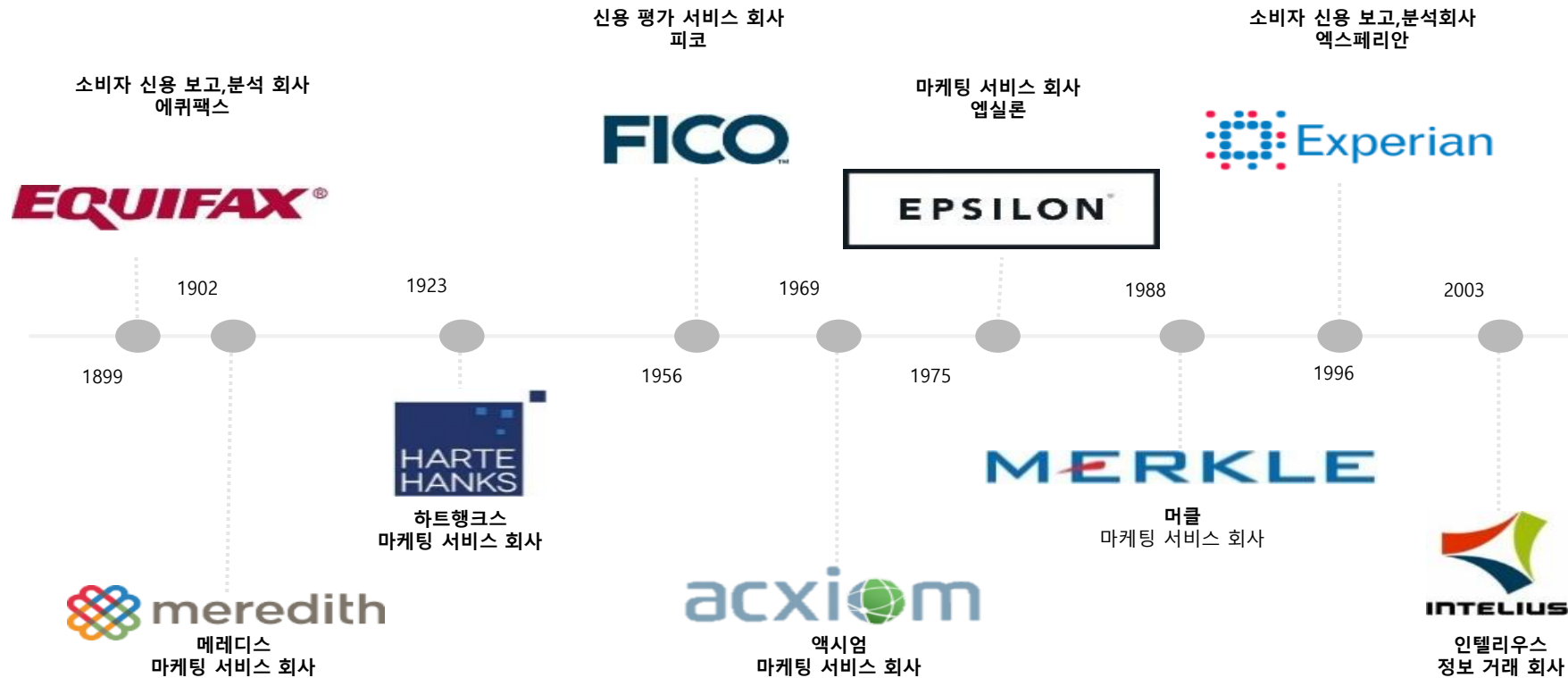
사용자 식별 정보	미디어나 기술관련 데이터(소유 전자제품)
인구 통계	부동산, 자동차, 취미, 여행
법, 공공 관련 기록	금융관련 , 구매행태, 건강(흡연자)

## 가공 데이터

1270	Personicx Refresh - Cluster Code & Indicator(그룹 세분 코드)	\$30
1271	Personicx Refresh - LifeStage Group Code & Indicator	\$30
1274	Personicx Refresh - Insurance Groups & Indicator(그룹 보험관련)	\$30
1275	Personicx Refresh - Financial Groups & Indicator(그룹 금융관련)	\$30
2705	Social Network - Number of Sites - Input Individual(SN 사용 사이트 수)	\$25
2706	Commercial Network Membership - Input Individual(쇼핑 사이트 멤버 수)	\$25
2829	Income Deciles(소득)	\$7
8851	Home Market Value Quality Indicator (replaces 8850)(부동산 가치)	\$10

Element	Description	\$/M Matched	Available
2701	Social Network - First Reported Activity	\$15	Now
2702	Social Network - Last Reported Activity	\$15	Now
2707	Social Network Interest	\$15	August
2705	Social Network - Number of Sites	\$35	Now
2704	Social Network - Number of Friends	\$35	Now
2703	Social Network - Membership	\$35	Now
2706	Commercial Network - Membership	\$35	Now
2708	Social Network - Facebook	\$35	July
2709	Social Network - LinkedIn	\$35	July
2710	Social Network - MySpace	\$35	July
2711	Social Network - Twitter Followers	\$35	August
2712	Social Network - Twitter Following	\$35	August
2714	Social Activity Segments	\$35	August

# 외부, 구매 데이터의 기업 설립 역사







데이터: 순간의 경험과 지식, IT: 전달 하는 수단

## Bimodal IT = Marathon Runners + Sprinters

Think  
Marathon Runner



### Mode 1

Reliability

Goal

Price for performance

Value

Waterfall, V-Model,  
high-ceremony IID

Approach

Plan-driven,  
approval-based

Governance

Enterprise suppliers,  
long-term deals

Sourcing

Good at conventional  
process, projects

Talent

IT-centric, removed  
from customer

Culture

Long (months)

Cycle times

### Mode 2

Agility

Revenue, brand,  
customer experience

Agile, Kanban,  
low-ceremony IID

Empirical, continuous,  
process-based

Small, new vendors,  
short-term deals

Good at new and  
uncertain projects

Business-centric,  
close to customer

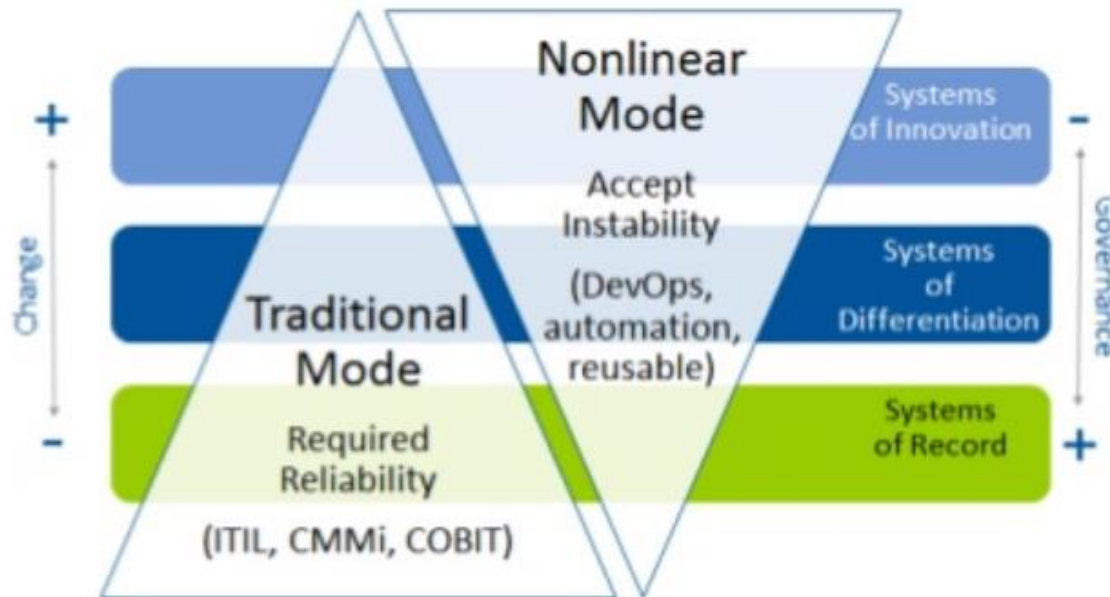
Short (days, weeks)

Think  
Sprinter



Gartner.

## The IT Response: No More One Size Fits All



**45%**

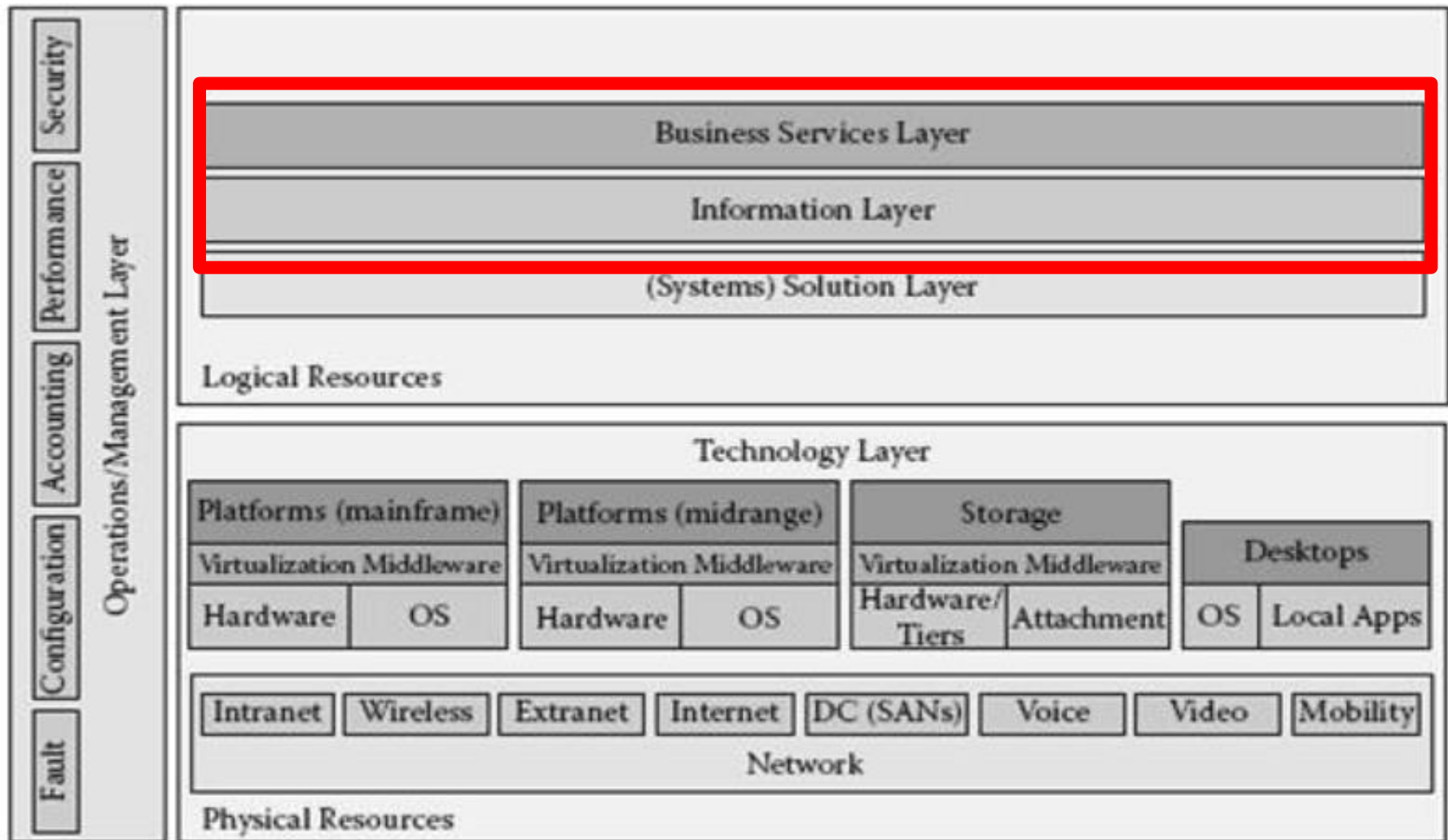
of CIOs currently  
have a second  
fast/agile mode  
of operation

2017년 75%

Source: Gartner annual CIO Survey, 2014  
@doudpundit  
© 2014 Gartner, Inc. and/or its affiliates. All rights reserved.

**Gartner**

# 기업의 IT 아키텍처





# Solution Architecture for bigdata

Business Architecture

Information Architecture

Infrastructure Architecture

Data Architecture

Integration Architecture

Service Architecture

# Information, Business Layers

	기술 및 인력	활용	
Service Management Platform	AI, Web service, Smart factory, Mkt service, BI...	공장, 도시, 홈, 쇼핑, 건강, 병원, 엔터테인먼트, 운송 등 일상의 변화 주도	데이터 서비스
DaaS: 데이터 공개, 공유 서비스			
Analytics Management Platform	Business analyst, Data Scientist, Analytic tool...	데이터 활용 서비스를 위한 데이터 분석 플랫폼 서비스	데이터 과학, 분석
Data Management Platform	Data Scientist, Data engineer, Data Architecture...	데이터 분석을 위한 데이터 가공 서비스 플랫폼	데이터 가공, 정제
Data Collection	IoT, Cloud, Big Data, Mobility...	만물인터넷을 통한 광범위 네트워킹	물리적 it 설비/SW

# 데이터 활용 프로세스

기업, 기관의 운영 속도와 효율성 높여 수익 극대화 및 최적화 제공이 핵심

정제 → 가공 → 신호, 패턴(인사이트) → 분석 모형 개발 → 모형 적용 → 수행 프로세스정립 → 평가

## 데이터 플랫폼 통합수행 내역(용역 측면)

인공지능

조치(Action)

시각화, 대시보드, 시뮬레이션 (**Human Action**)

Decision, Alerts, Curricula (**Machine Action**)

브라우저, 모바일 디바이스, 배치 작업

웹 서비스, FTP, Sockets

접근

분석 결과 데이터베이스

대량 데이터의 최적화 접근

OLAP, RDBMS, Mem/Cashed

분석

모형(Model) 적용, 수행

의사결정을 위한 비즈니스 룰

분석 모형 결합(Ensembles)

분석모형, 기계학습 개발

SVM, Neural Nets, Regression, K-NN, SVD, Matrix Factorization, GEO-Distance 등등

데이터 신호, 패턴 선택

PCA, Decision tree, Chart, Clustering, 감성 분석, Regression, Outlier 등등

정제된 데이터베이스

데이터 접근의 최적화

NOSQL, 인 메모리, RDBMS

신호 생성

데이터 신호, 패턴 구별

신호 생성 알고리즘 들

시계열, 통계, 이벤트, 지리 위치 등

데이터 저장(data store)

데이터 저장, 데이터 전송

데이터 저장, 데이터 전송

ETL, 가공

데이터 가공(Transform)

데이터에 맞는 툴 사용

SAS, Hadoop, 구매 ETL 툴, 자체 개발 툴

데이터 수집(Extract)

융통성 있는 데이터 수집 인터페이스

웹 서비스, 소켓(Sockets), FTP, SQL., SAP

데이터 소스

전사운영데이터:  
고객,오라클,SAP 등

구매데이터:  
실시간 구매, 배치 구매

웹 데이터:  
정형, 비정형 데이터

기계, 위치 데이터:  
정형, 비정형 데이터

전체 작업의 60~70%

수행된  
정립  
것  
인력이  
수행

# 데이터 활용의 트렌드

빅데이터  
검색어

...

사물 인터넷  
검색어

...

인공지능  
검색어

...

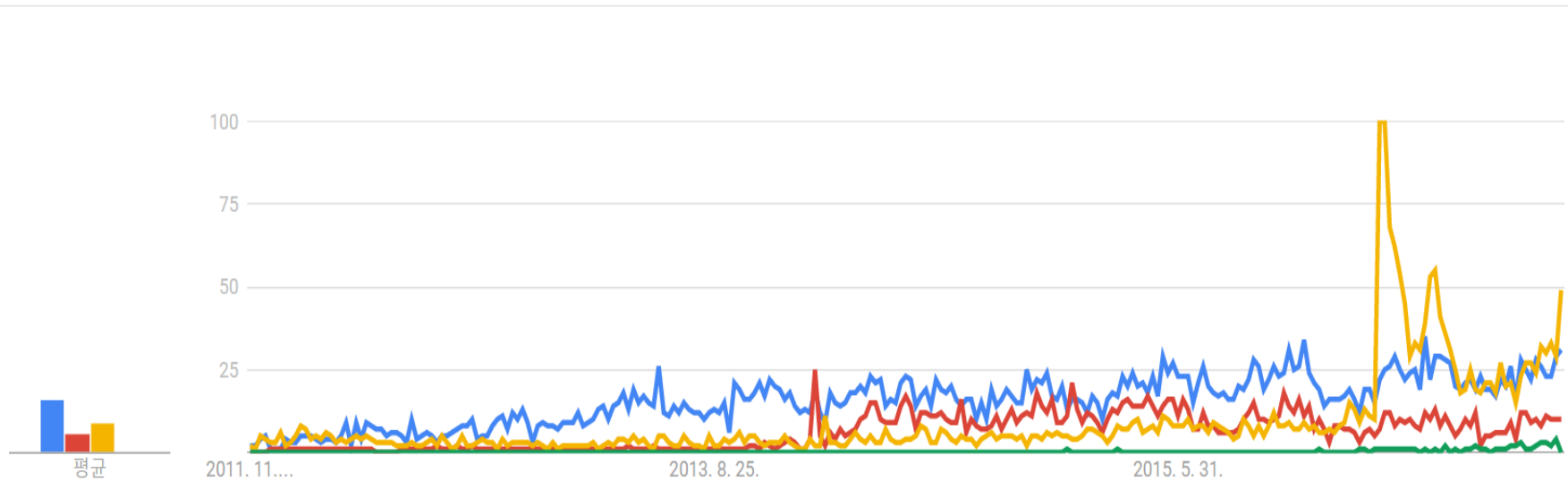
4차산업혁명  
검색어

...

+

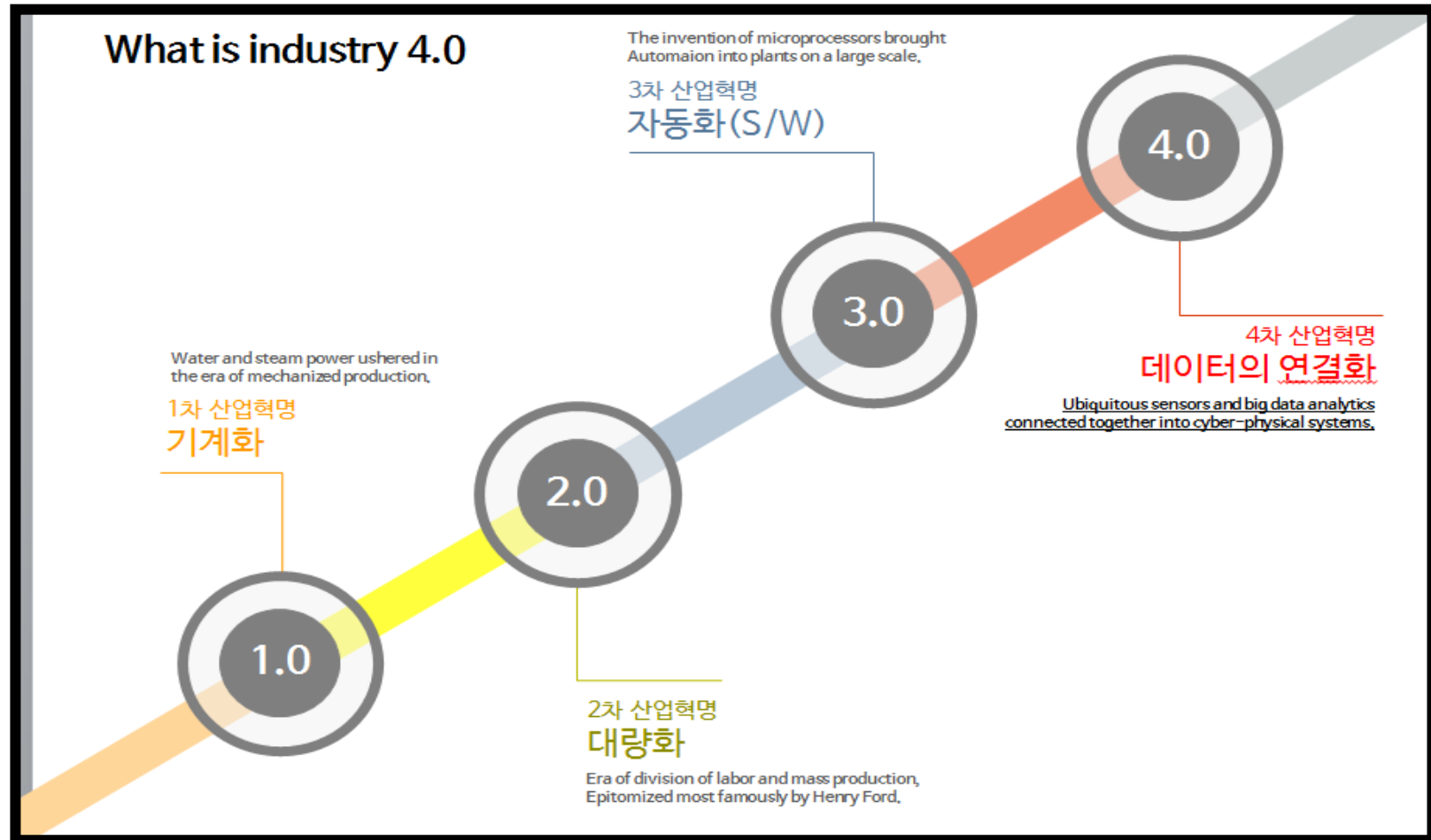
전 세계 ▼    지난 5년 ▼    모든 카테고리 ▼    웹 검색 ▼

시간 흐름에 따른 관심도 변화 ?















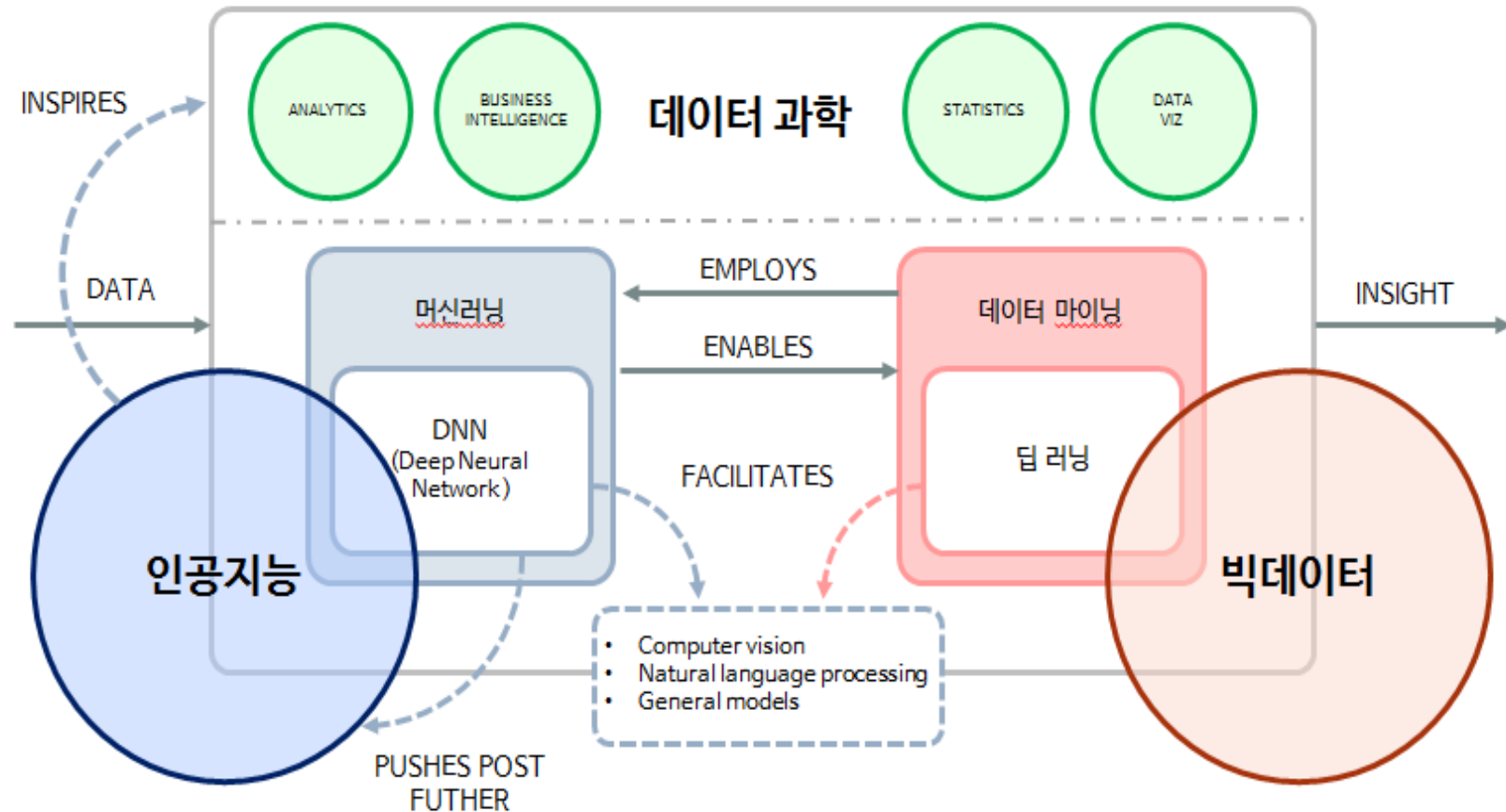
# 산업혁명 4.0: 데이터와 연결화



# 데이터 혁명

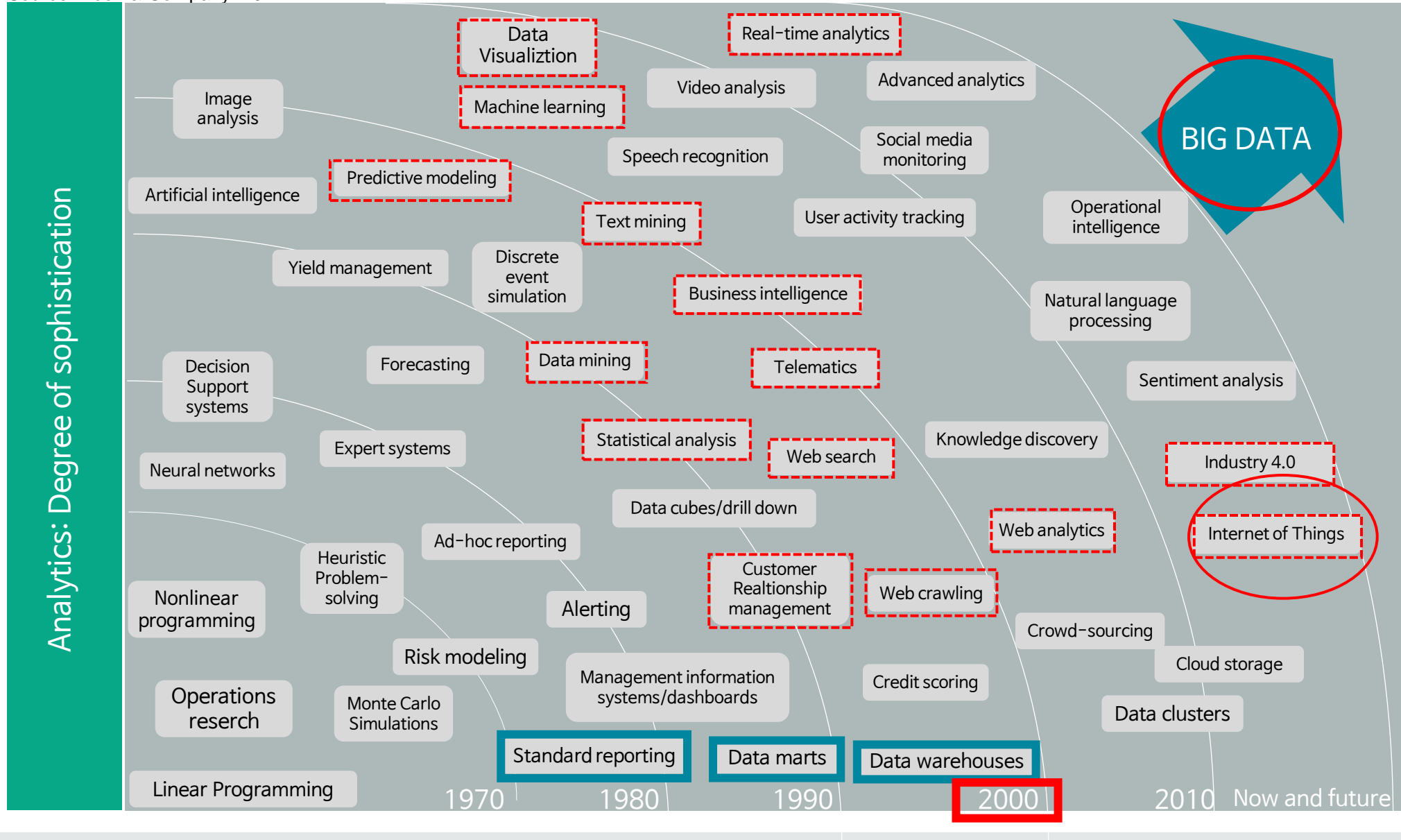
전통 기업의 데이터혁명	농업혁명	제조혁명	의료혁명
	MONSANTO 	 ZARA 	 Google
	물류혁명	유통혁명	금융혁명
	 FedEx	coupang 	 TransferWise SoFi 
스타트업 기업의 데이터혁명	DNA 데이터분석	이상감지, 데이터분석	데이터가공/유통
	 23andMe	 Palantir	acxiom
	O2O	개인자산관리	트위터뉴스
	 UBER	 PERSONAL CAPITAL	 Dataminr

# Data Science vs. Artificial Intelligence vs. Big Data(IOT) 관계



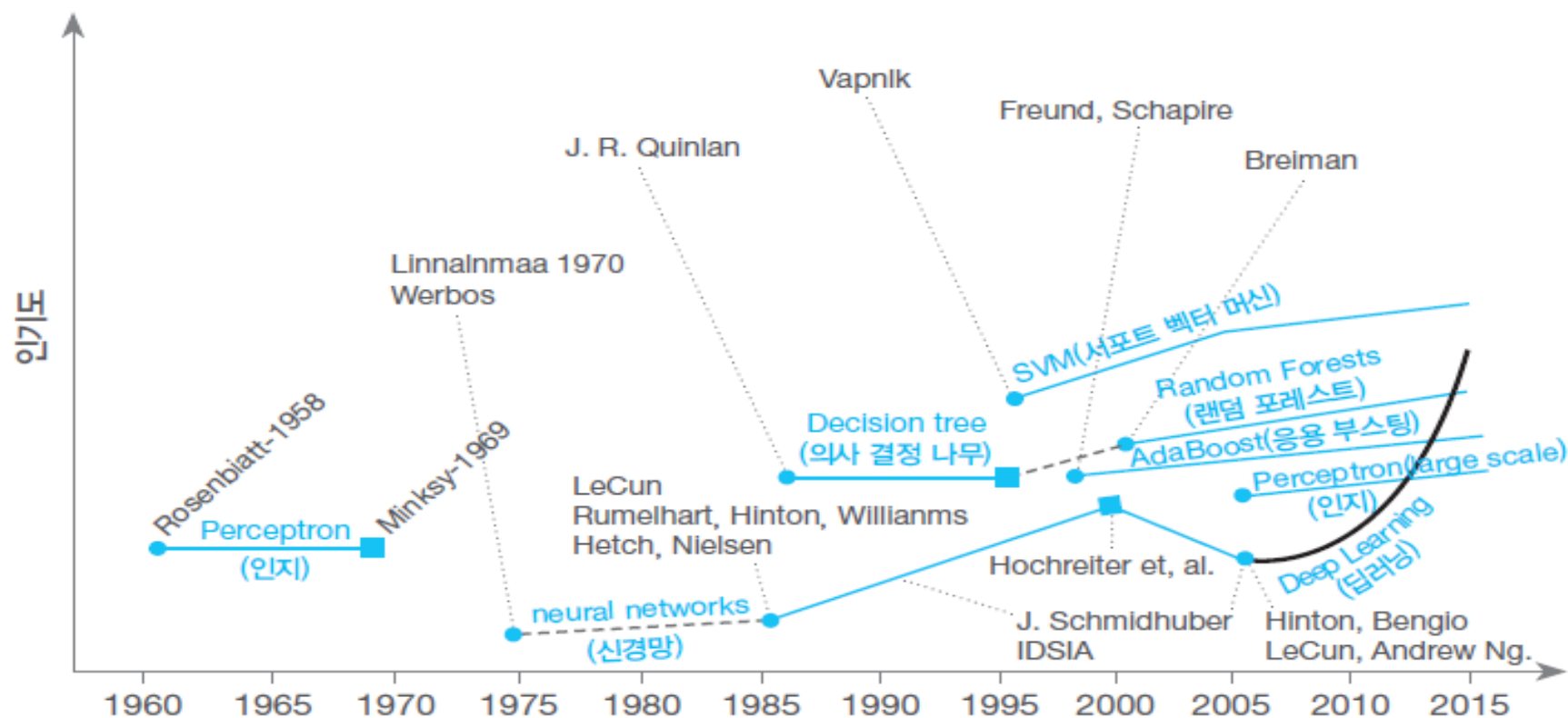
# 시대별 데이터 분석(알고리즘) 단계

Source: Booz & Company. 2014



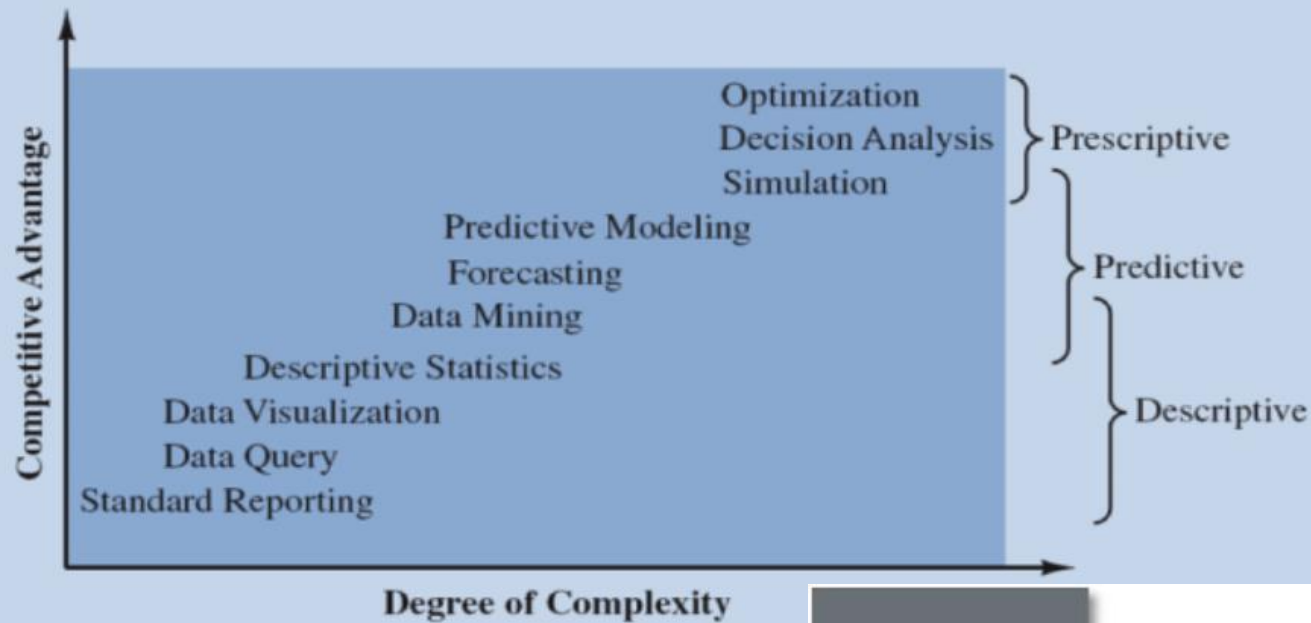
# 시대별 데이터 분석(알고리즘) 단계

시대별 주요 기계학습 발전 및 활용 추세

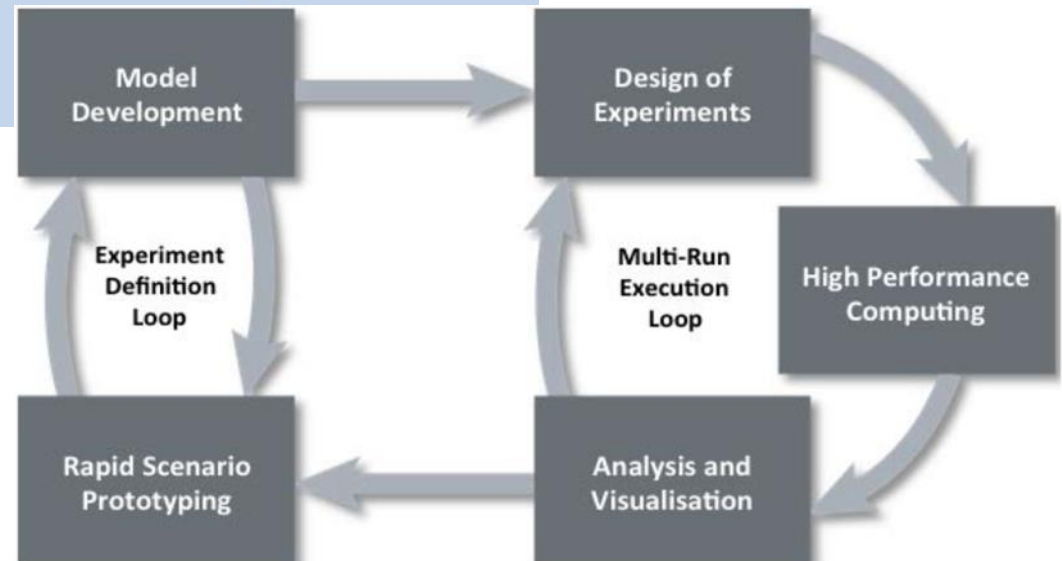




# Data Mining vs. Data Farming



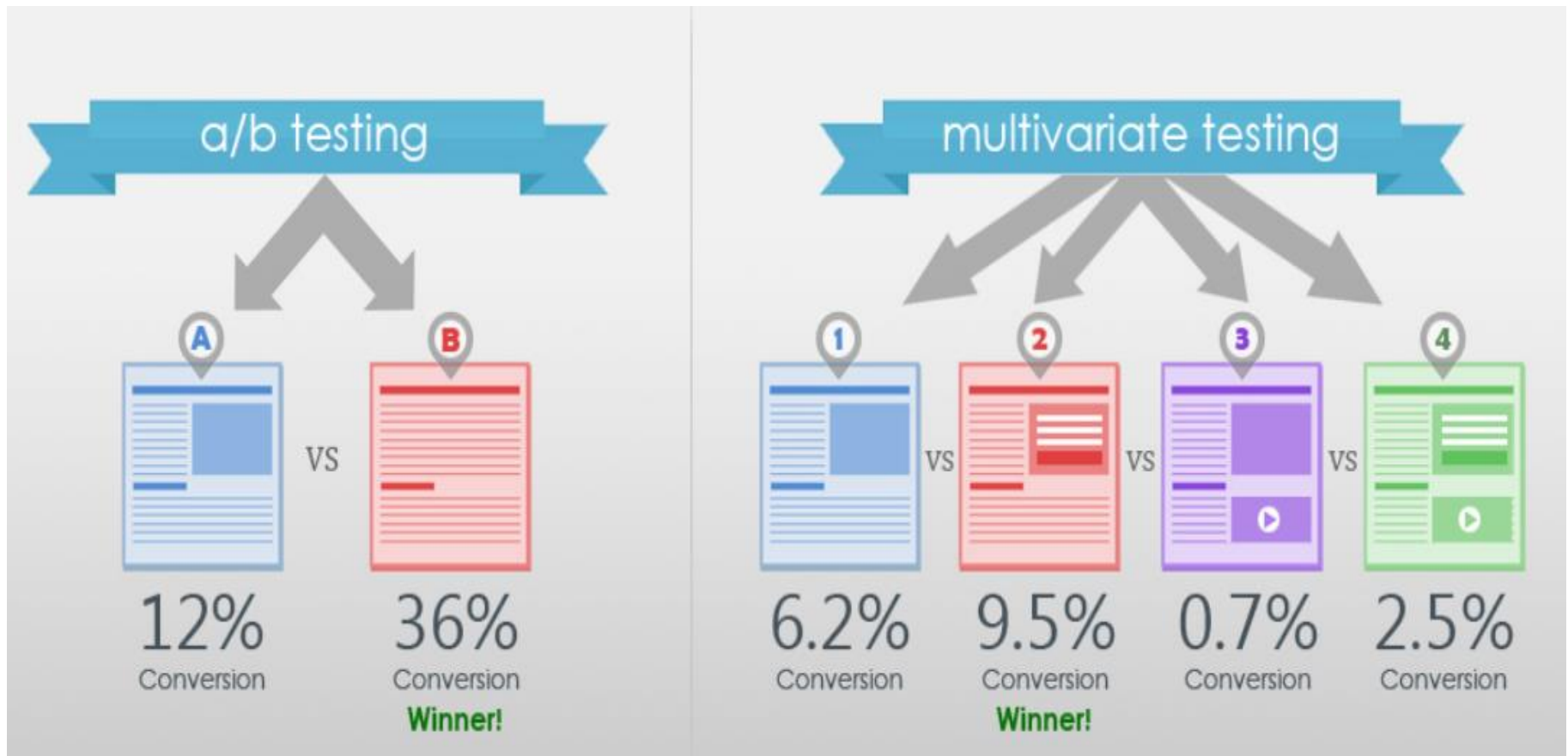
Source: Adapted from SAS.



Marine Corp's Project Albert

SEED Center for Data Farming/  
The Naval Postgraduate School

## 효과 테스트(A/B Test)



# 데이터 활용 프로세스

기업, 기관의 운영 속도와 효율성 높여 수익 극대화 및 최적화 제공이 핵심

정제 → 가공 → 신호, 패턴(인사이트) → 분석 모형 개발 → 모형 적용 → 수행 프로세스정립 → 평가

## 데이터 플랫폼 통합수행 내역(용역 측면)

조치(Action)



접근



분석



신호 생성



ETL, 가공



데이터 소스

시각화, 대시보드, 시뮬레이션 (Human Action)

Score, Decision, Alerts, Curricula (Machine Action)

브라우저, 모바일 디바이스, 배치 작업

웹 서비스, FTP, Sockets

분석 결과 데이터베이스

대량데이터의 최적화 접근

OLAP, RDBMS, Mem/Cached

모형 (Model) 적용, 수행

의사결정을 위한 비즈니스 룰

분석 모형 결합 (Ensembles)

분석모형, 기계학습 개발

SVM, Neural Nets, Regression, K-NN, SVD, Matrix Factorization, GEO-Distance 등등

데이터 신호, 패턴 선택

PCA, Decision tree, Chart, Clustering, 감성 분석, Regression, Outlier 등등

정제된 데이터베이스

데이터 접근의 최적화

NOSQL, 인 메모리, RDBMS

데이터 신호, 패턴 구별

신호 생성 알고리즘들

시계열, 통계, 이벤트, 지리 위치 등

데이터 저장 (data store)

현업에 맞는 환경 구축

Hadoop, 인 메모리, RDBMS

데이터 가공 (Transform)

데이터에 맞는 룰 사용

SAS, Hadoop, 구매 ETL 룰, 자체 개발 룰

데이터 수집 (Extract)

융통성 있는 데이터 수집 인터페이스

웹 서비스, 소켓 (Sockets), FTP, SQL., SAP

전사운영데이터:  
고객, 오라클, SAP 등

구매데이터:  
실시간 구매, 배치 구매

웹 데이터:  
정형, 비정형 데이터

기계, 위치 데이터:  
정형, 비정형 데이터

숙련된

경험

이

있는

인력만이

수행

가능

# 데이터 관련 산업의 본질 파악이 최우선

1. 우리가 보는 성공 사례들은 B2C 영역
2. 빅데이터 관련 산업은 B2B 영역(플랫폼, 데이터 유통)
3. 빅데이터 수요자(중간자 B)가 아직 준 내부활용 미진한 상태(기업이나 기관)



# 기업의 데이터 활용 5단계

## 기업 데이터 활용 프레임 워크와 발전 단계

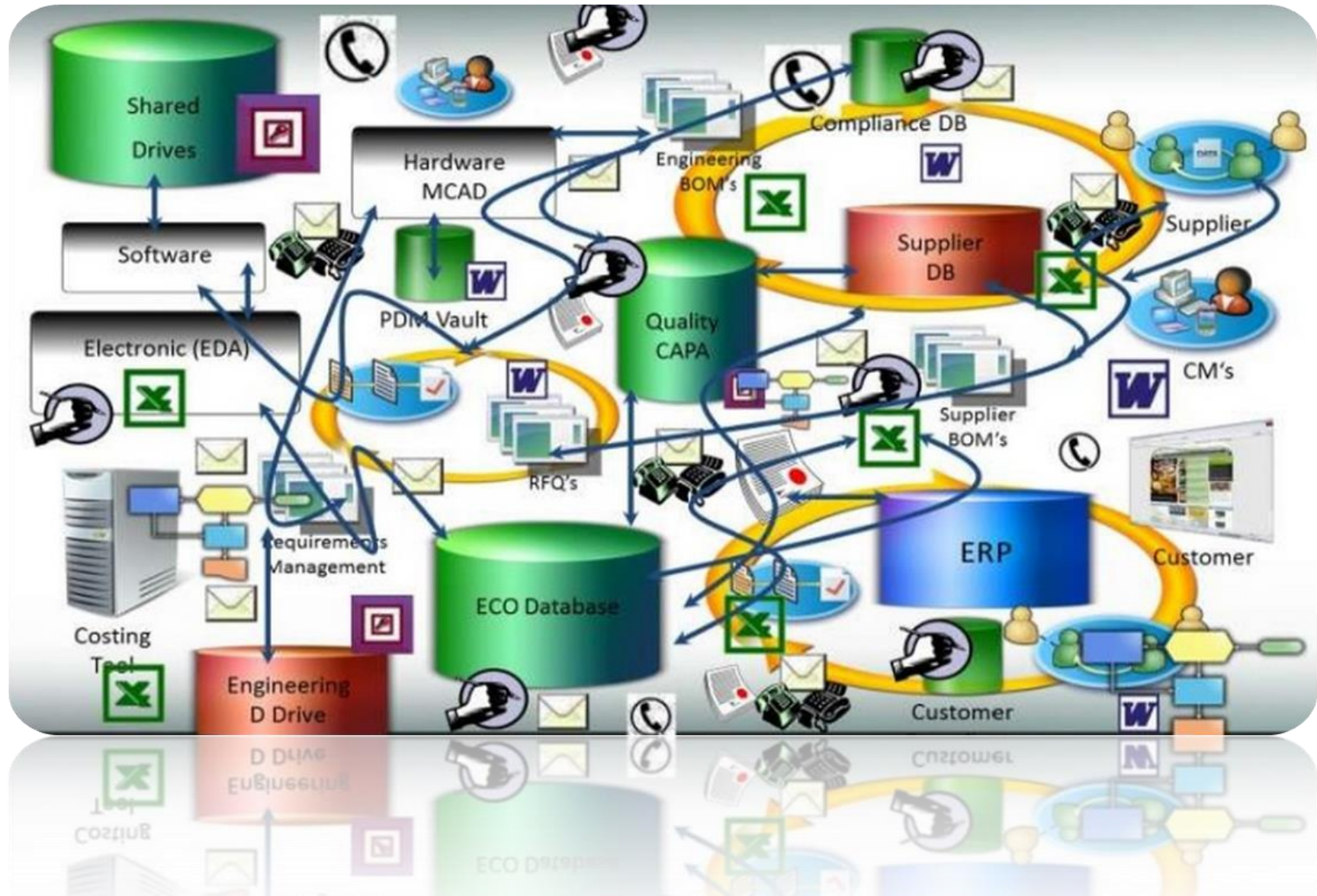
	1단계	2단계	3단계	4단계	5단계
비전 (Vision)	없음	생산 성과 비전 시도	각 영역별 채널의 효율성	전사 내부 통합	연결된 가치 부여 및 인식
전략 (Strategy)	없음	고립적 프로젝트, 아래로부터 시도	좀더 협력적 생각, 여전히 사일로 존재	전사적 통합 CRM 프로그램	서로의 이익을 위해 협력하는 가치 인식
고객 경험 (Customer Experience)	개념 없음	개념 없음	이해하는 사일로 수준에서 집중	각 영역별 연결된 비즈니스로 이해하고 집중	좀더 넓은 영역 이해, 협력
협력 (Collaboration)	내부 영역별로 집중, 사일로 (Silo) 구조	초기 편협적 고객 위주, 사일로적	문화나 동기의 변화, 여전히 사일로적	한국 고객 중심, 분야별, 영역별, 재 구조 조정	고객 중심 공유, 목표 연합적 협력
프로세스 (Process)	내부 영역별로 집중, 사일로적	초기 자동화 시기, 사일로적	사일로 수준에서 비용과 가치의 최적화	전사 수준에서 비용과 가치의 최적화	초기부터 끝까지 실제적 최적화 프로세스
정보 공유 (Information)	기본적 정보의 산재	팀 기준, 산재, 최소의 인사이트	사일로 수준 정보 공유, 인사이트 발전 과정	전사 관여 정보 공유 및 인사이트	미국 기업을 넘어선 인사이트와 정보 공유, 외부 데이터 활용
기술 (Technology)	몹시 산재되고 미미한 기술	산재 존재, 한정된 역할과 집중	사일로 내에서 높은 수준의 역할	전사 통합 수준의 높은 수준의 역할	기업을 넘어선 높은 수준의 역할
집중 영역 (Metrics)	적은 내부적 집중 영역	산재되고 한정된 집중 영역, 운영 내부적 집중	사일로 내에서 효율적, 고객 집중 부족	전사적 고객 집중/균형 있는 구조	목표 공유, 균형 있는 구조, 잘 연결, 정리된

미국 2000년 초

빅데이터 (IoT, 외부 데이터, 소셜)



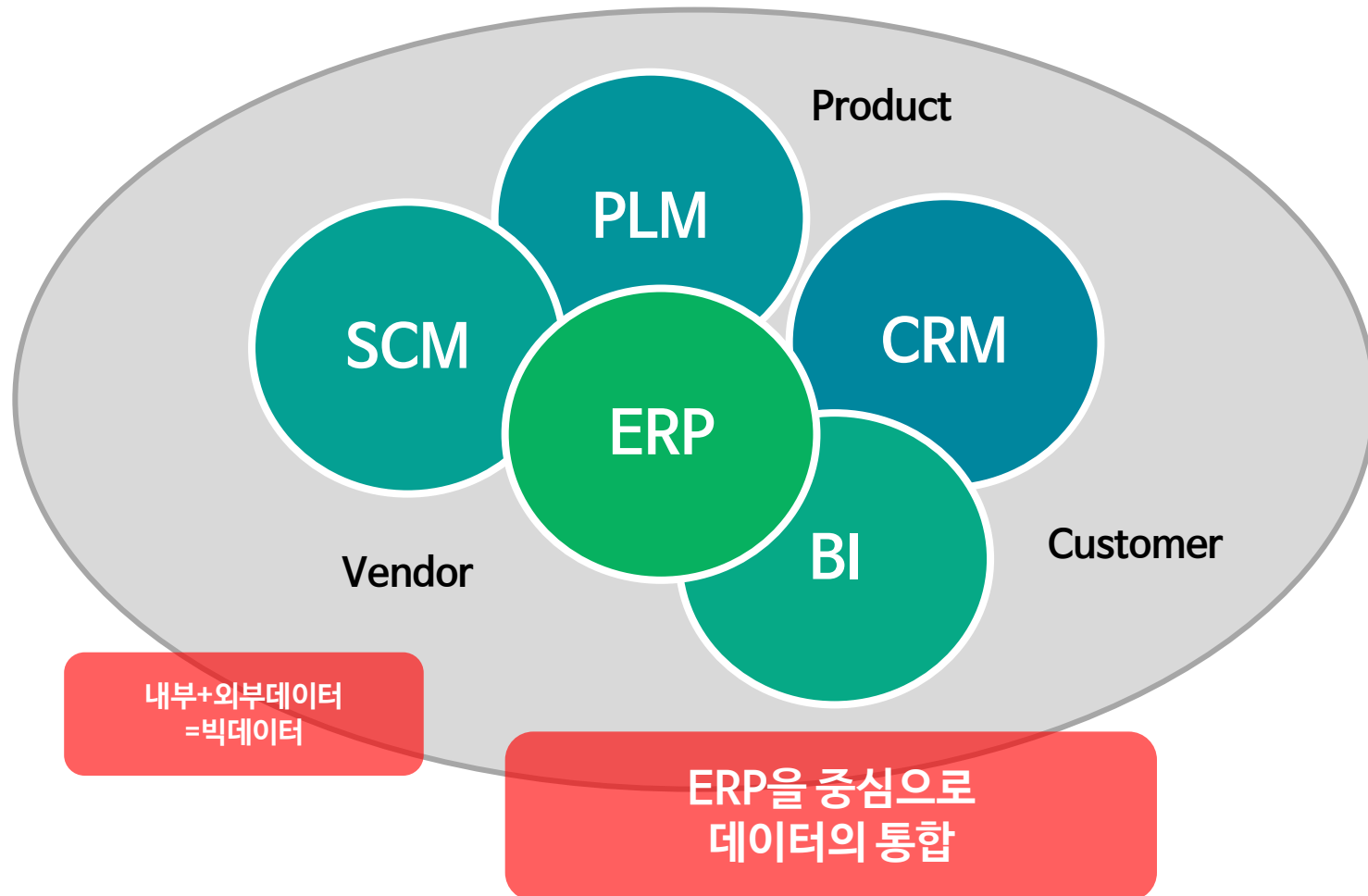
# 데이터 통합의 필요성



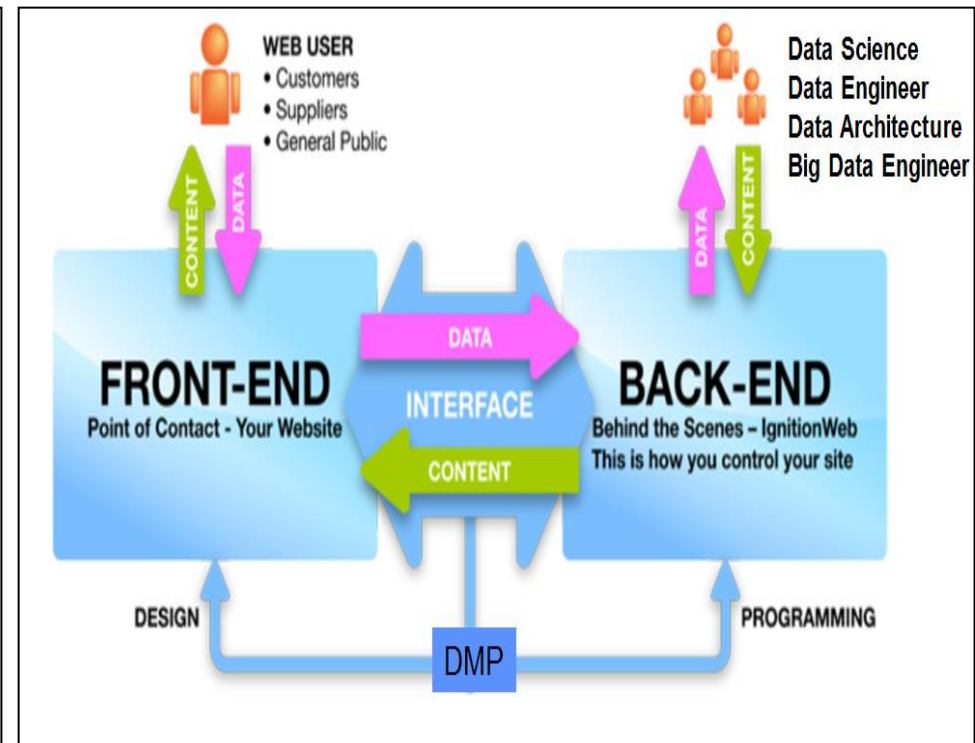
## 데이터 가공 플랫폼(DMP): data siloes effects



## 데이터 통합 → 빅데이터



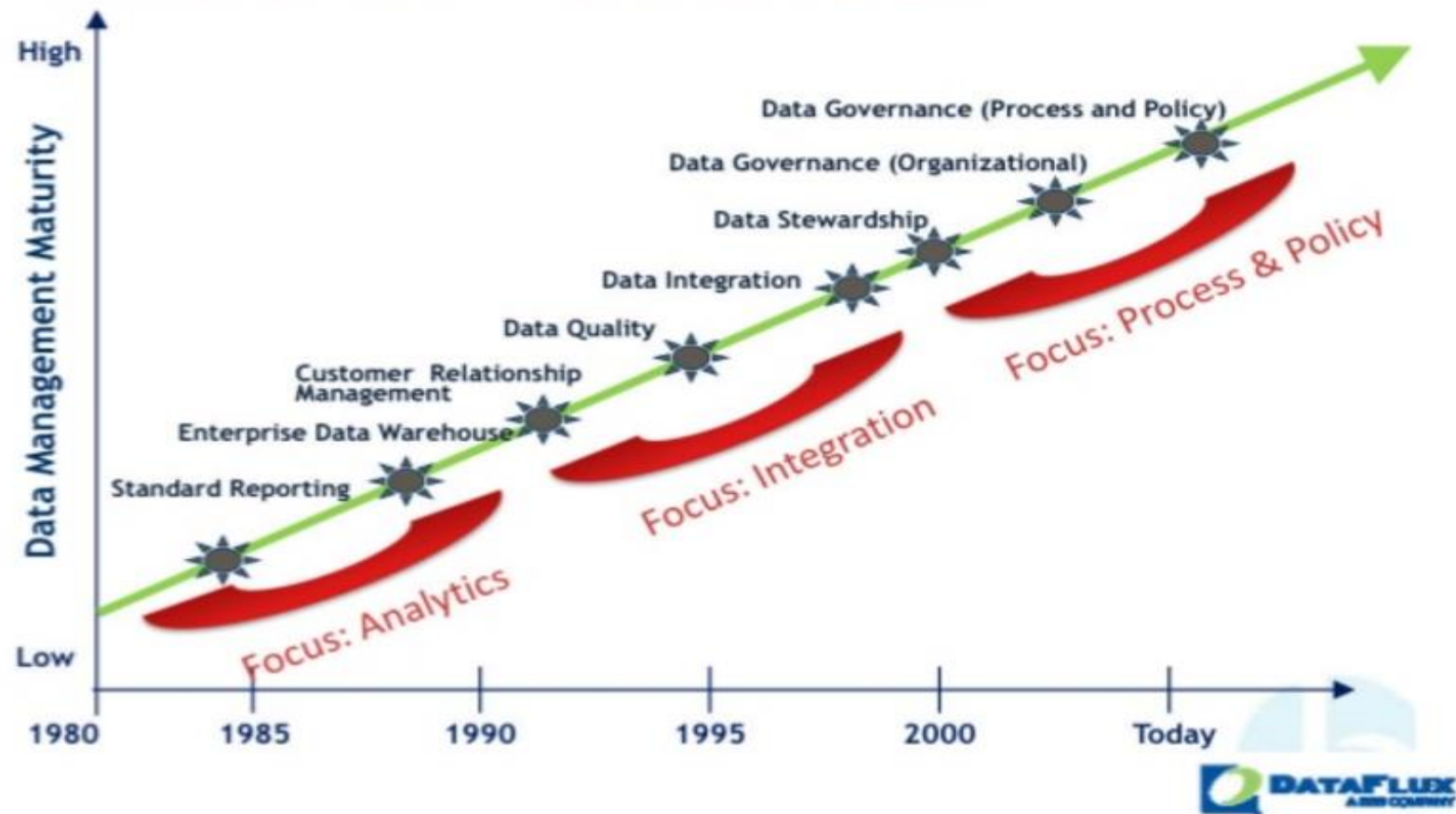
# 데이터 가공 플랫폼(Data Management Platform)



1. 복잡성 -> 단순화
2. 일관성과 점진적, 지속적 발전 유지
3. 빠른 적용, 적응력
4. 유연성 향상
5. 반복적 데이터 접근 패턴
6. 반복적 통합 패턴
7. 반복적 시스템 적용 패턴
8. 반복적 데이터 디자인 패턴

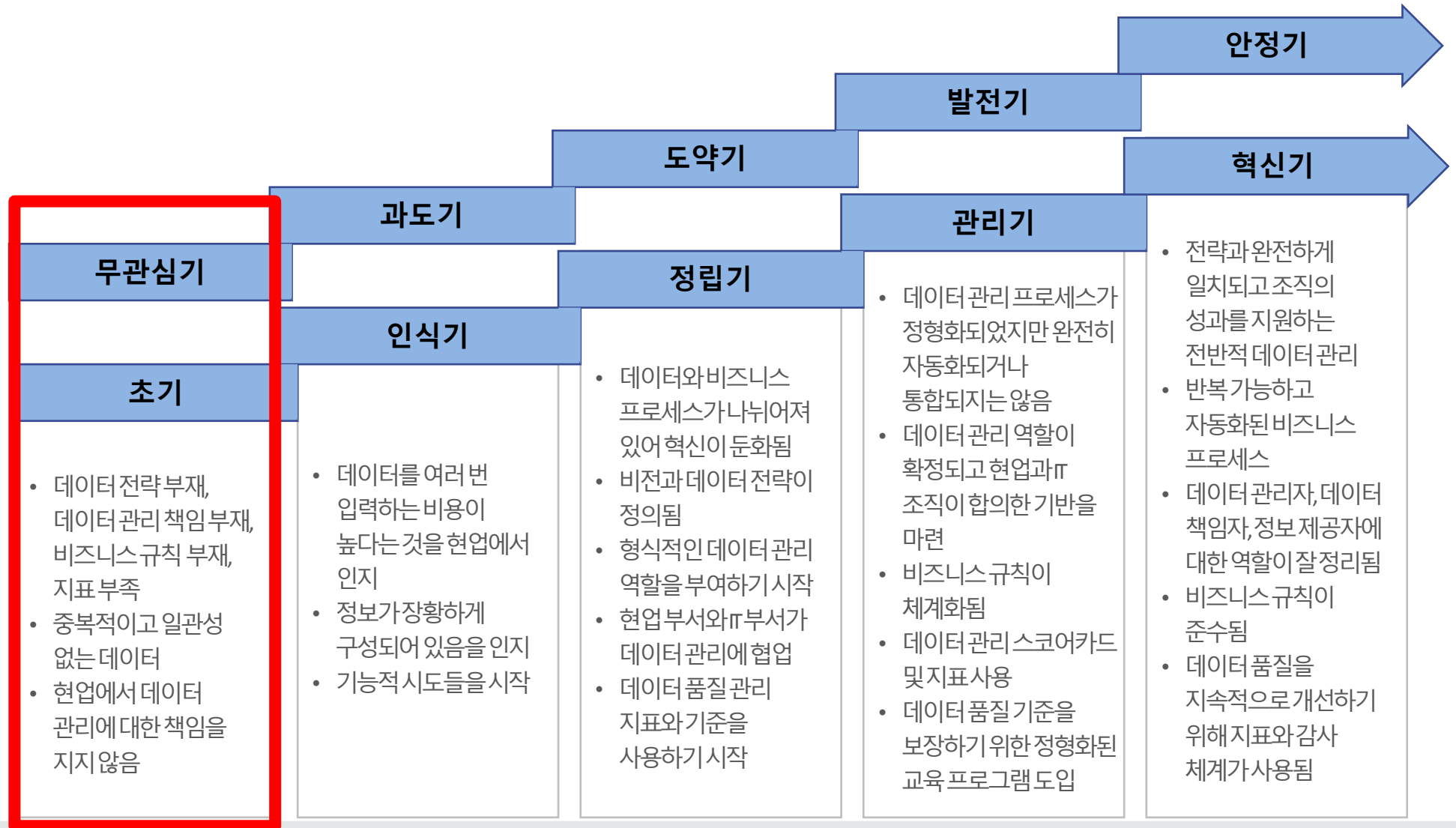


## 데이터 자산화 에 대한 인식의 단계

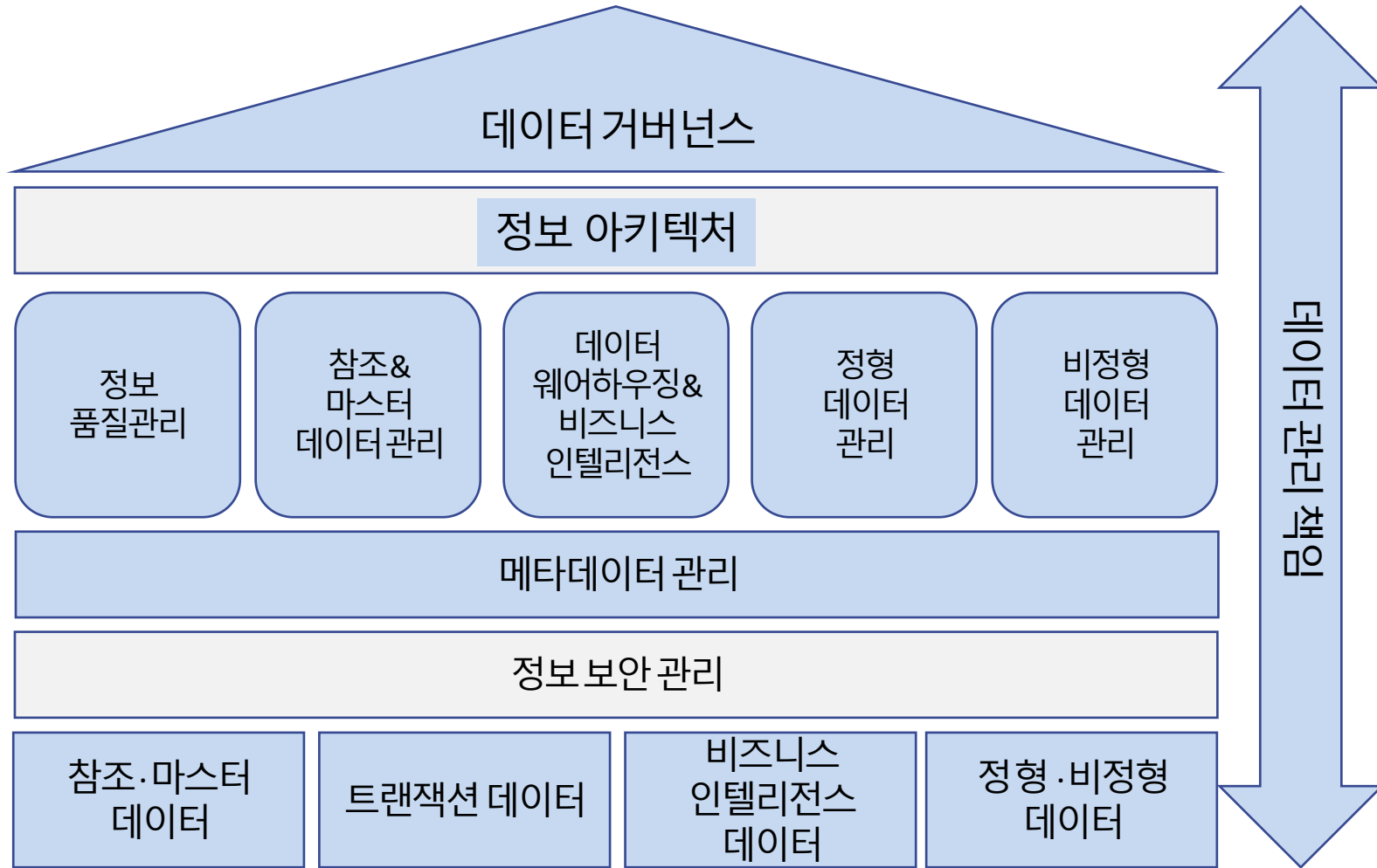




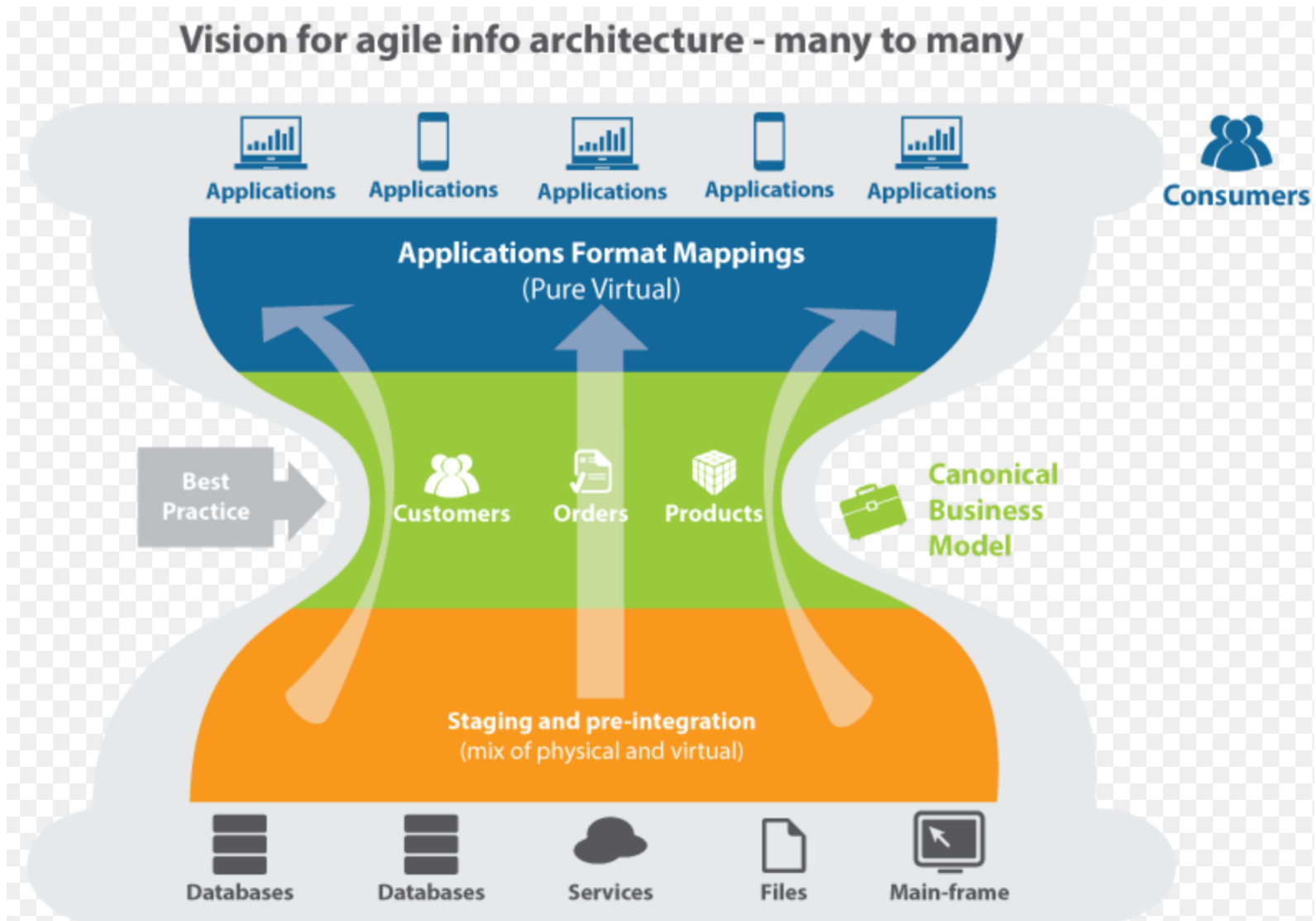
# 데이터 거버넌스의 발전 단계



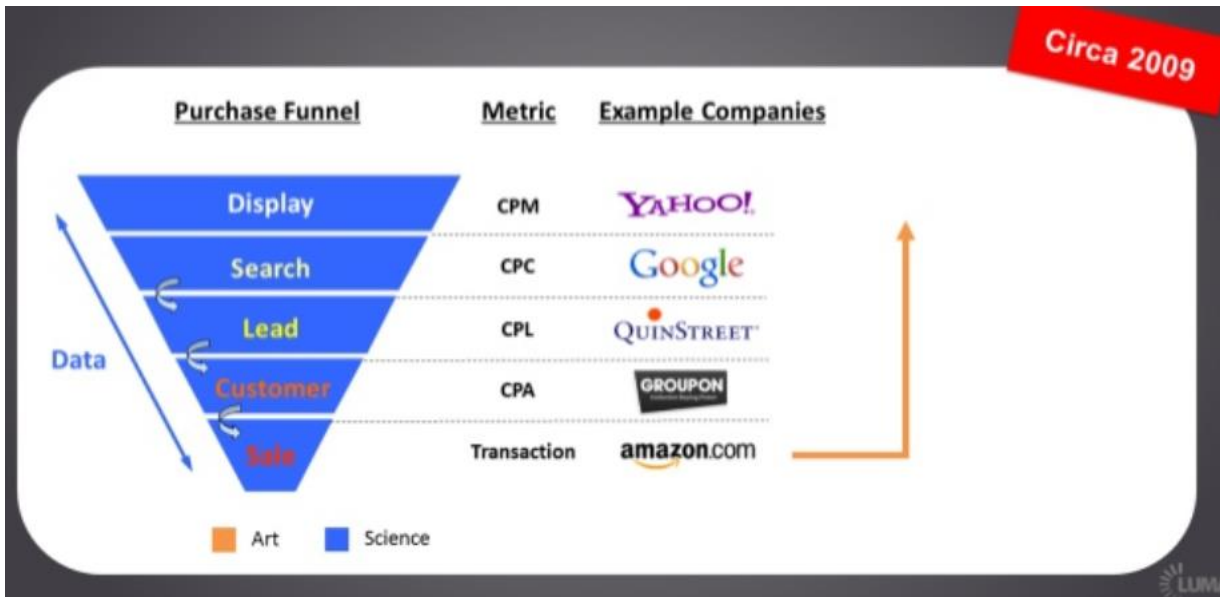
# 데이터 거버넌스의 틀



# 논리적 데이터 아키텍처 vs. 물리적 데이터 아키텍처



# 모바일, 미디어, 마케팅, 테크놀로지 발전



## Amazon Is Playing a Different Game...



I know who you are...



I know what you might be looking for...



I know what you bought and what you will buy next.

## Amazon Is Focused On Its Top-of-Funnel Business

amazon

Expected Advertising Revenue in 2018  
\$5 Billion

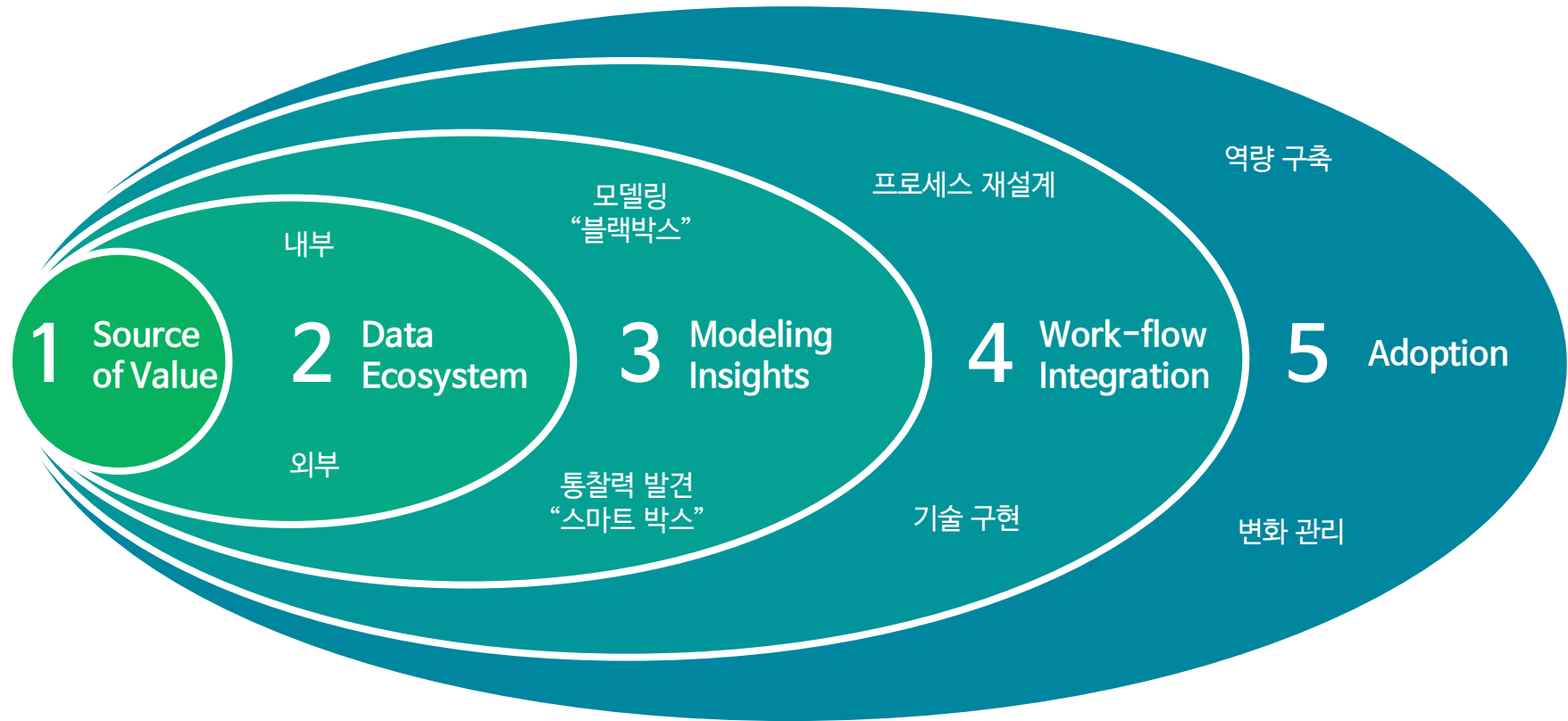
amazon advertising

DSP & Header Bidding Solution

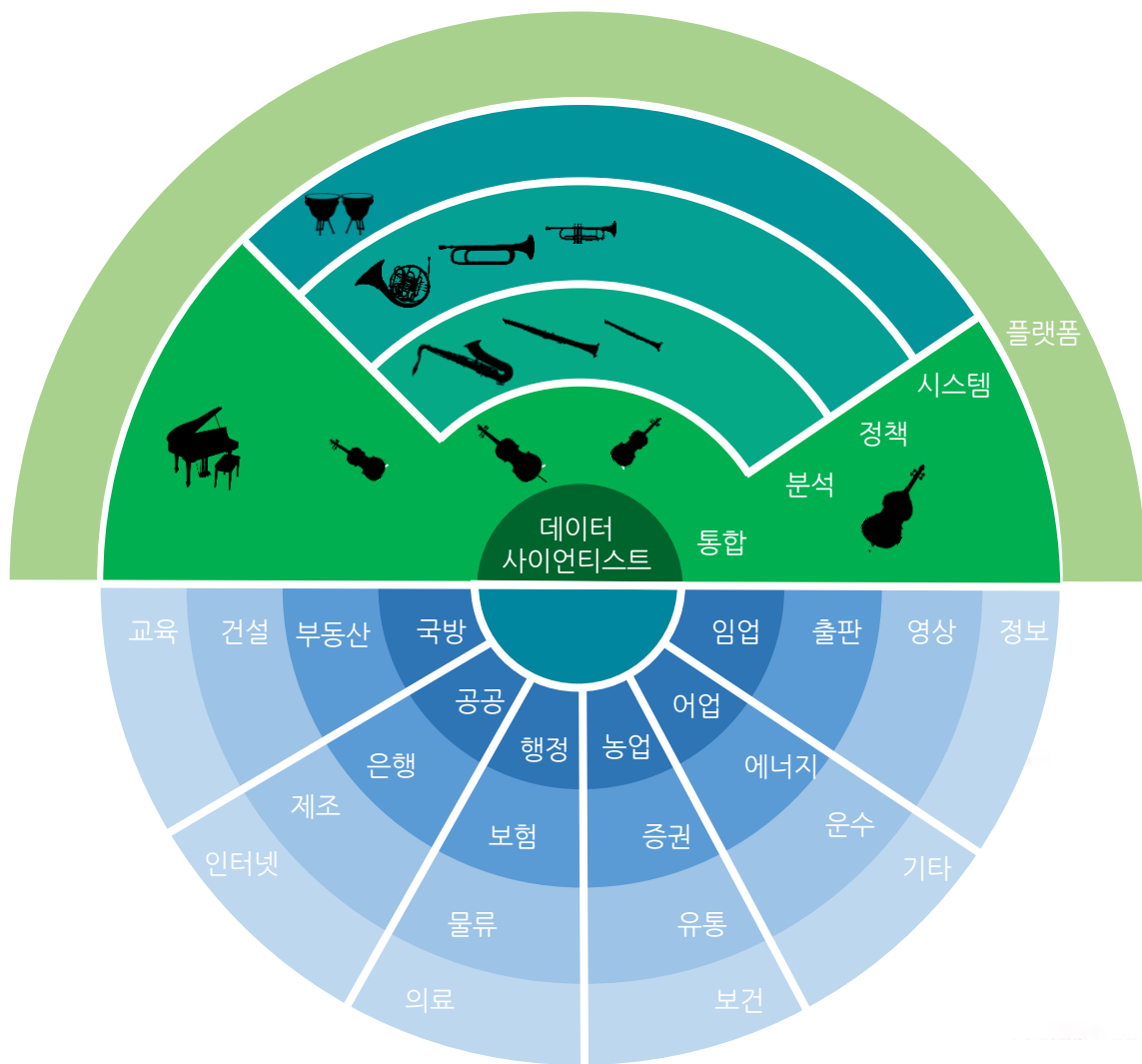
amazon web services

amazon echo

## 데이터 활용의 성공을 위한 Key Factors

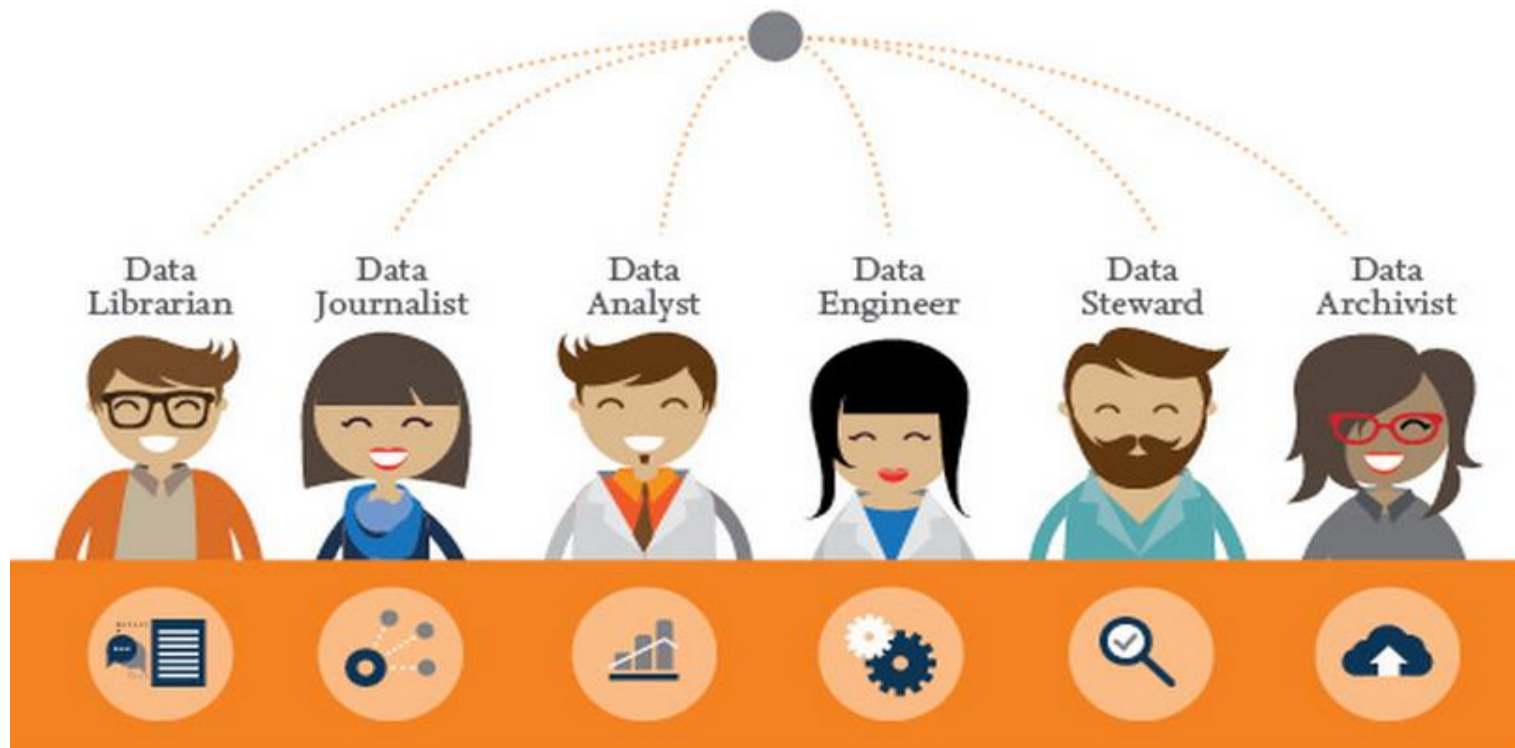


# 데이터 과학이란?

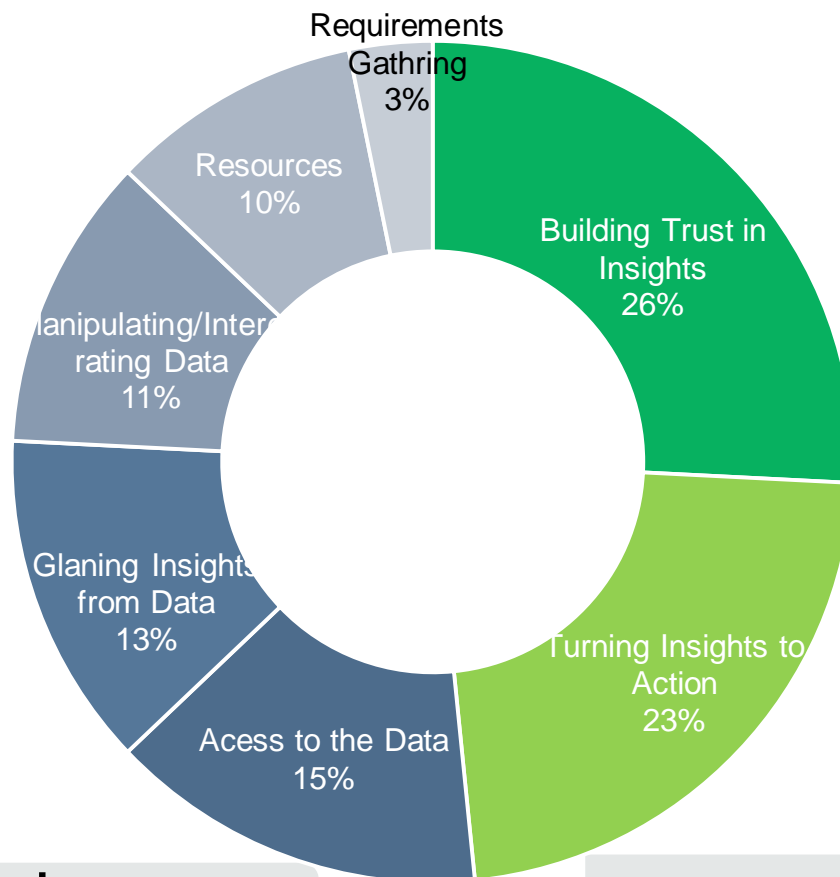




# DATA SCIENCE ROLES

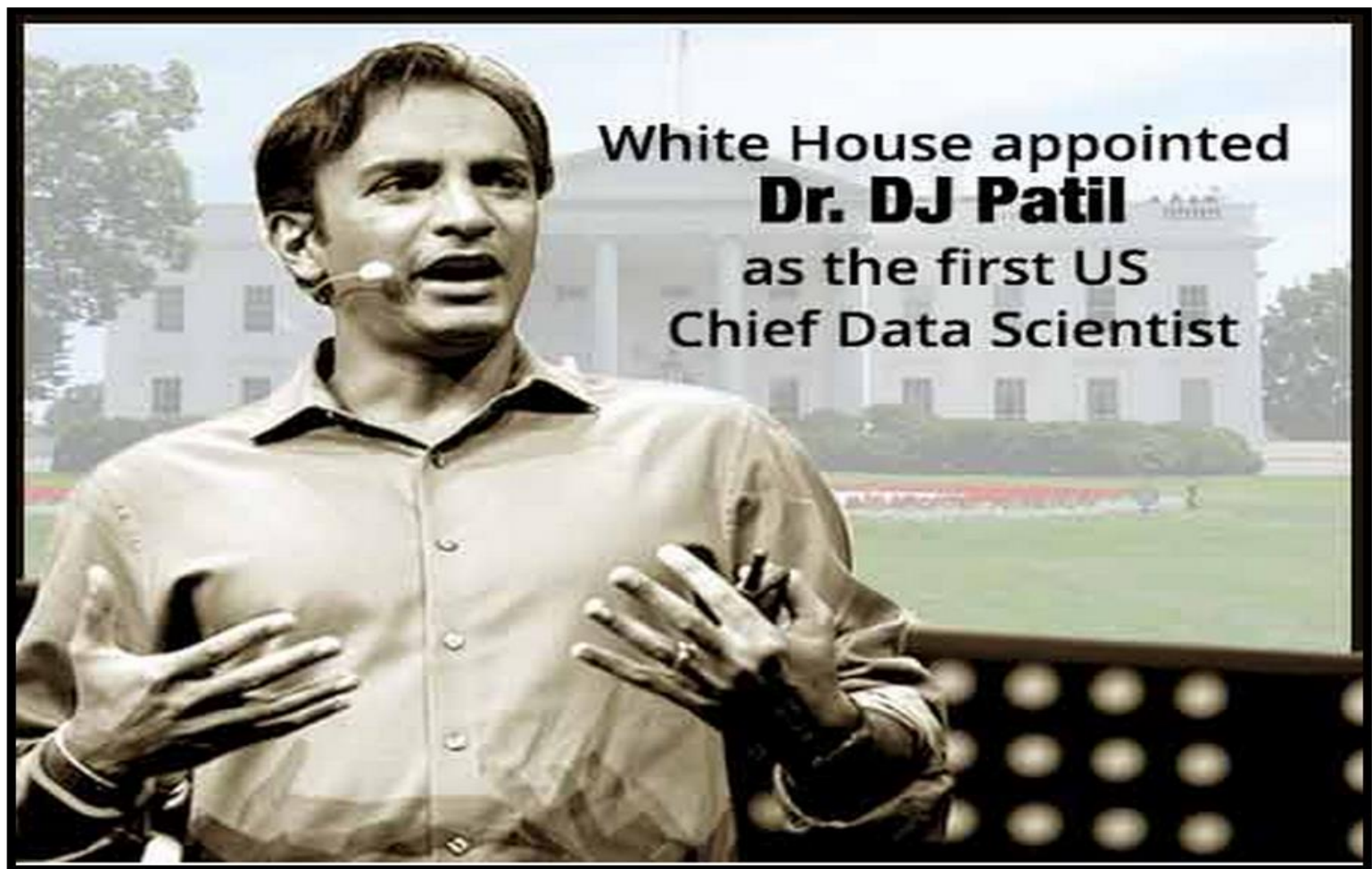


## 데이터 과학자로서의 가장 큰 도전은 무엇입니까?

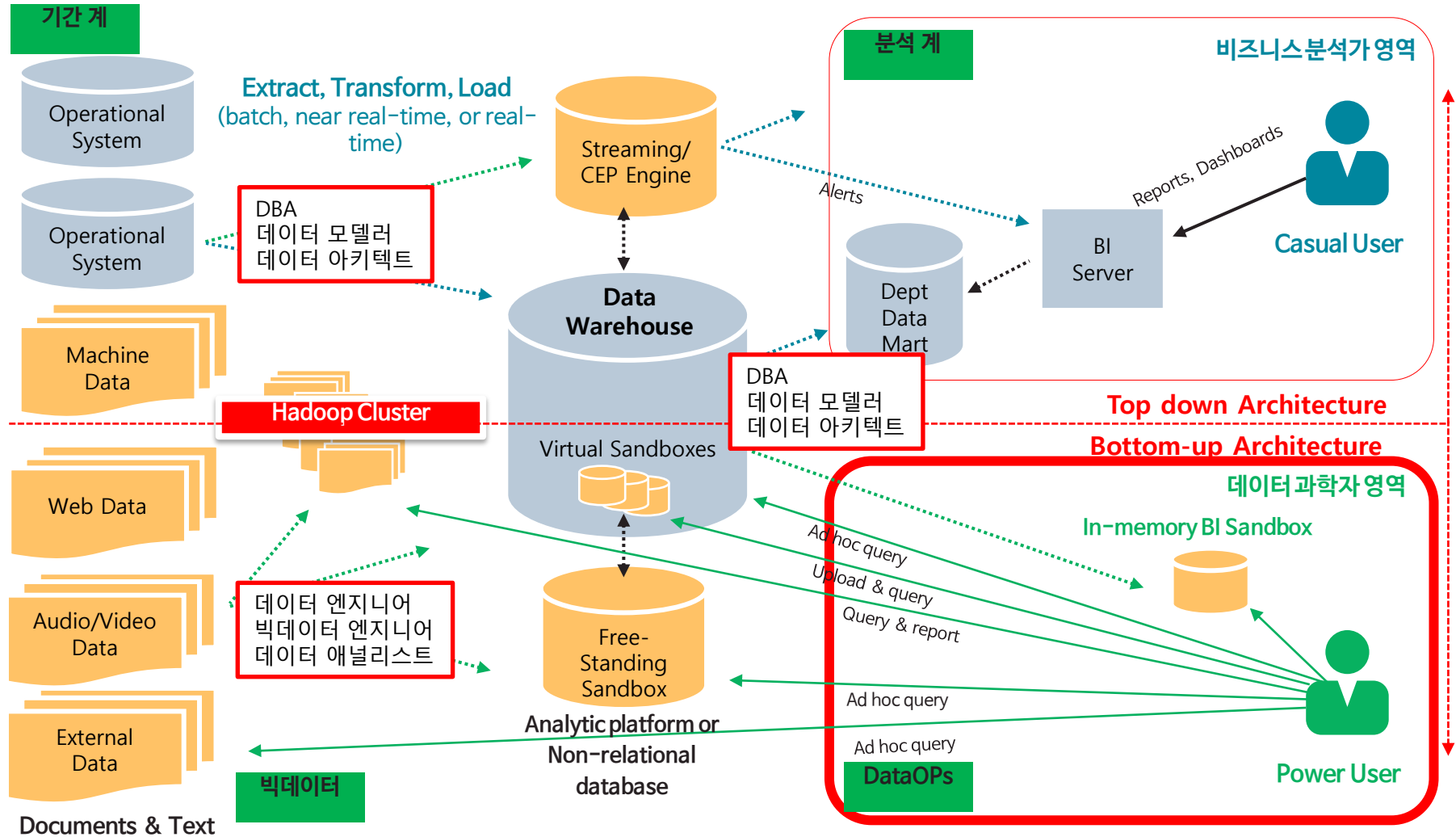


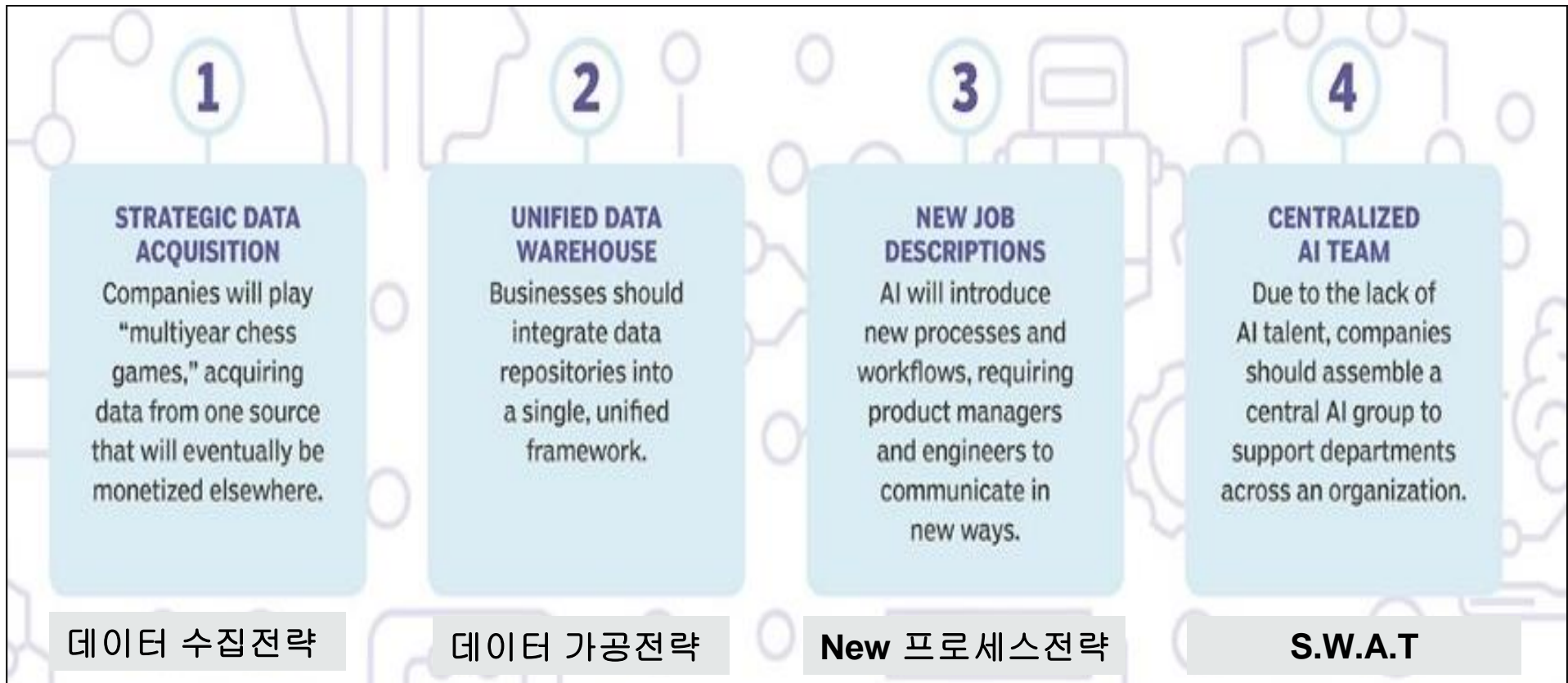
데이터

비즈니스



# 빅데이터 생태계에서의 데이터과학 역할





## IT 주도 모델-> 연대 협력 모델

