



데이터 경제 활성화의 Bottlenecks 대응전략

© 2015 EN-CORE. All rights reserved.

Data Scientist : (주)엔코어 데이터 서비스 센터장 김옥기
Data Mining Consulting, Data Driven Strategy Consulting
okkim@en-core.com

미국 최대 데이터 브로커


Acxiom Clients

- 12 of the top 15 credit card issuers
- 9 of the top 10 automotive clients
- 9 of the top 10 top telecom / media companies
- 7 of the top 10 retailers
- 7 of the top 10 retail banking companies
- 7 of the top 10 top property / casualty insurance providers
- 6 of the top 8 top brokerage firms
- 3 of the top 5 top pharmaceutical manufacturing companies
- 2 of the top 5 U.S. rankings of life / health insurance providers
- 2 of the top 3 cruise lines, hotels and U.S. casinos

HSBC 



charles SCHWAB

 E·FUNDS™



 macy's way to shop!®



 Nationwide



Sprint®
Together with NEXTEL


A First Data Company

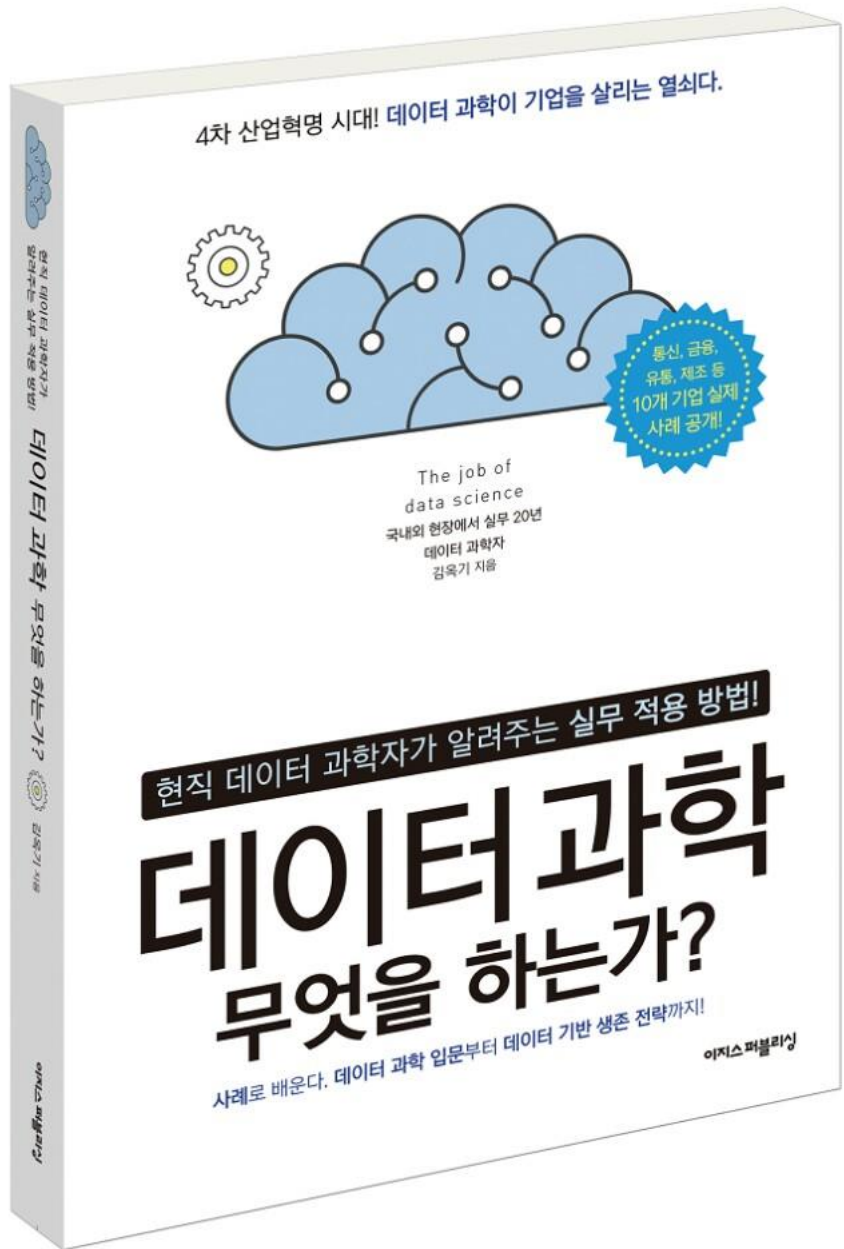

TransUnion™


Unilever




ACXIOM®
WE MAKE INFORMATION INTELLIGENT™

The trademarks and registered trademarks on this page are the property of their respective owners. State updated as of 2/1/07



4차 산업혁명의 핵심 데이터 경제

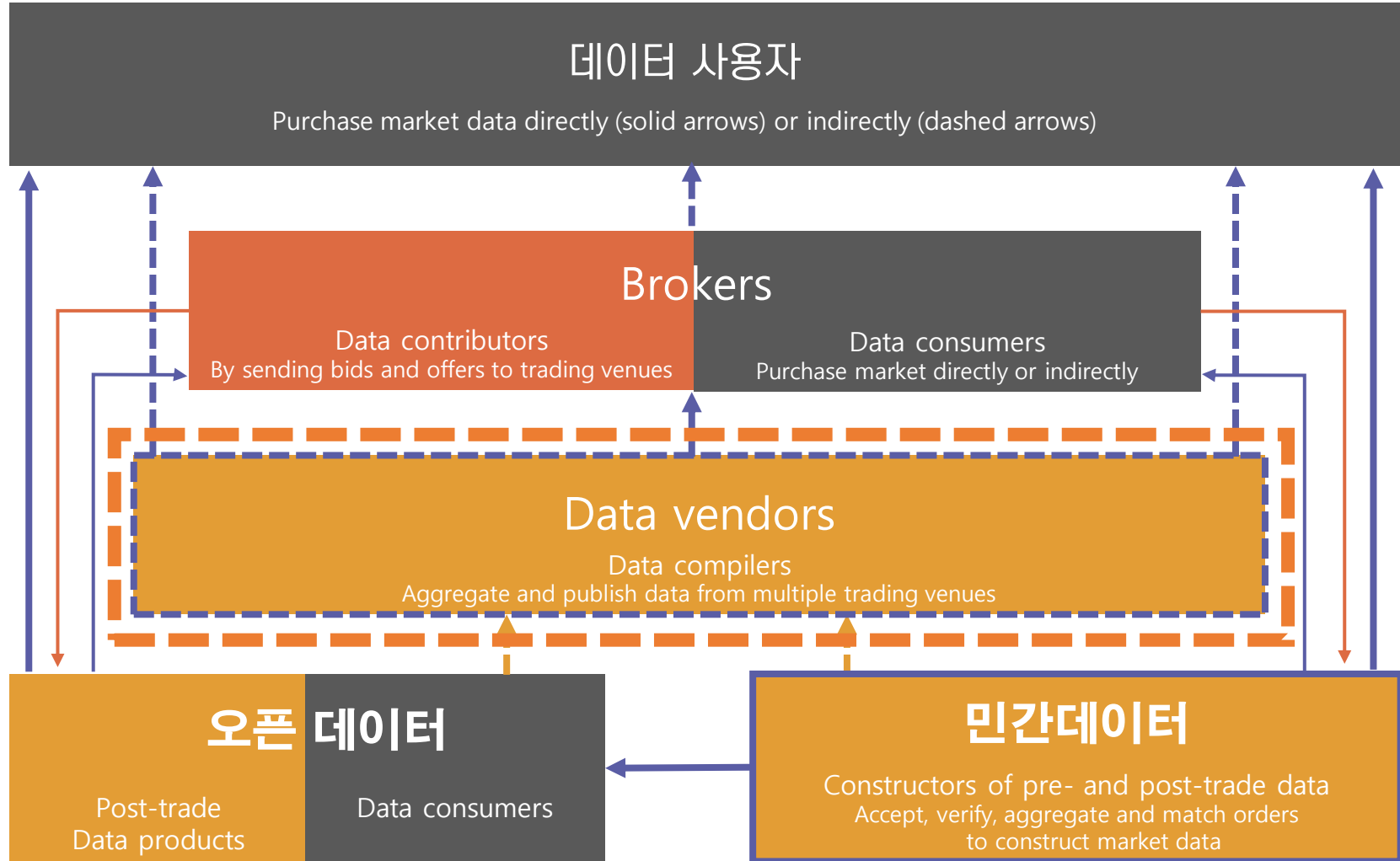
1. 데이터 경제: 기업 VS. 산업
2. 데이터 경제의 Bottlenecks
 - 디지털 전환을 위한 기업의 문화
 - 전사 빅데이터 거버넌스
 - 비즈니스 먼저, 그리고 데이터, SI 구축 나중
 - 기업의 분석 역량의 내재화(Data Literacy)

최근 정권 경제성장률 단위 : %

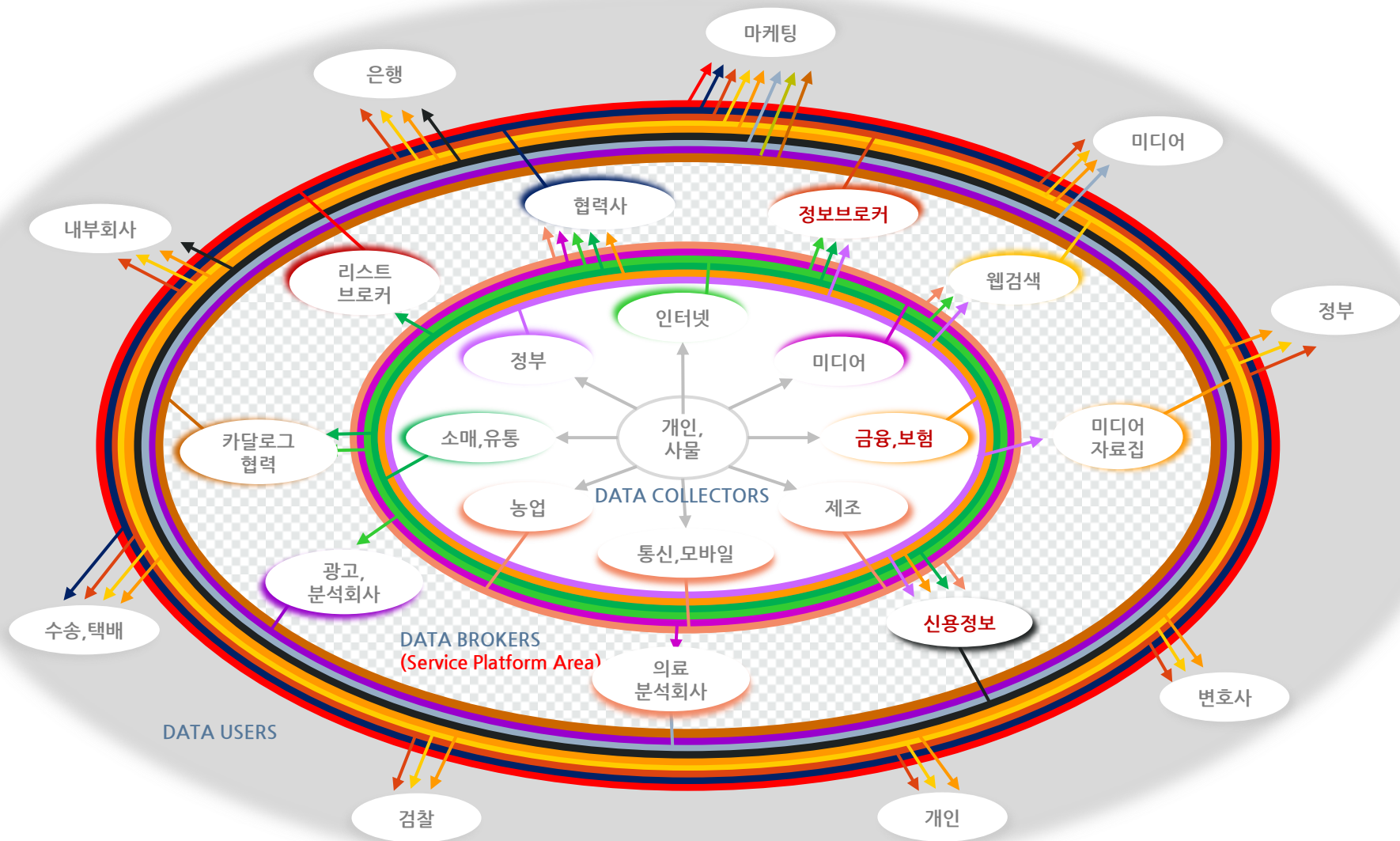


자료: IMF

데이터 경제: 기업 VS. 산업



데이터 산업 측면: 데이터 생태계-개인 데이터 + 사물 데이터



데이터 병목현상(Bottleneck)



기업 문화: 전환, 혁신을 위한 필수조건: 의지와 진정 성 (Digital Transformation)



Change	Transformation
Subscribes to Vision	Prescribes Vision
Fixes the Past and Current	Creates the Future
Driven by Tactics	Driven by Strategy
Focus on Methods and Processes	Focus on Mindsets and Beliefs
External Influence is High	External Influence is Minimal

Made by Hadiqa Aamer



XBR – Extreme Blueprinting & Roadmapping The Blueprint and the Roadmap

XBR (Extreme Blueprinting and Roadmapping) involves a set of methods and activities used to formulate a Future State Blueprint and Roadmap. It is analogous to Extreme Programming, with a focus on iterative delivery, continuous communication and interactive work processes.

This approach provides a clear set of deliverables that aligns strategic to tactical and business to technology. We call this approach using the SAFE Architecture doing "IT Transformation the SAFE Way".

What is the Blueprint?

- A relatively high-level vision of an organisations' envisaged Future-State
- The XBR Blueprint is inclusive of Strategy, Organization, People, Process and Technology and the Solution Definition
- Driven by the established Technology Principles and Business Priorities
- Much of the focus of this phase of work is on the Business Strategy and Technology Architecture
- Once completed, the Blueprint is thought of as a relatively static representation of the Current-State and Future-State with identified intermediate states

What is the Roadmap?

- The Roadmap is derived from the Blueprint
- It is a translation of the Blueprint into a dynamic representation of 'what it takes' to actually do the implementation
- It is a plan that can be directly implemented
- It is complemented in Phase 3 by Foundation Activities – those areas of Data Management and Infrastructure that are required first in the implementation stages of a project

전략과 로드맵

변화, 혁신을 위한 필수조건: 변화의 의지와 진정성

Model for Managing Complex Change



Adapted from Knoster, T. (1991) Presentation in TASH Conference. Washington, D.C. Adapted by Knoster from Enterprise Group, Ltd.

1987, 2003, Mary Lippitt

Why

Who,
What,
When,
Where

기업 목표 및 전략

IT 전략

Data 전략

비즈니스 전략

데이터 거버넌스

- 역할 및 책임
- 정책
- 프로시저 함수
- 비즈니스 규칙
- 데이터 활용 인지도
- 데이터 컨텍스트 인지도
- 업무 흐름
- 데이터 감사
- 측정 및 지표
- 보고

데이터 관리

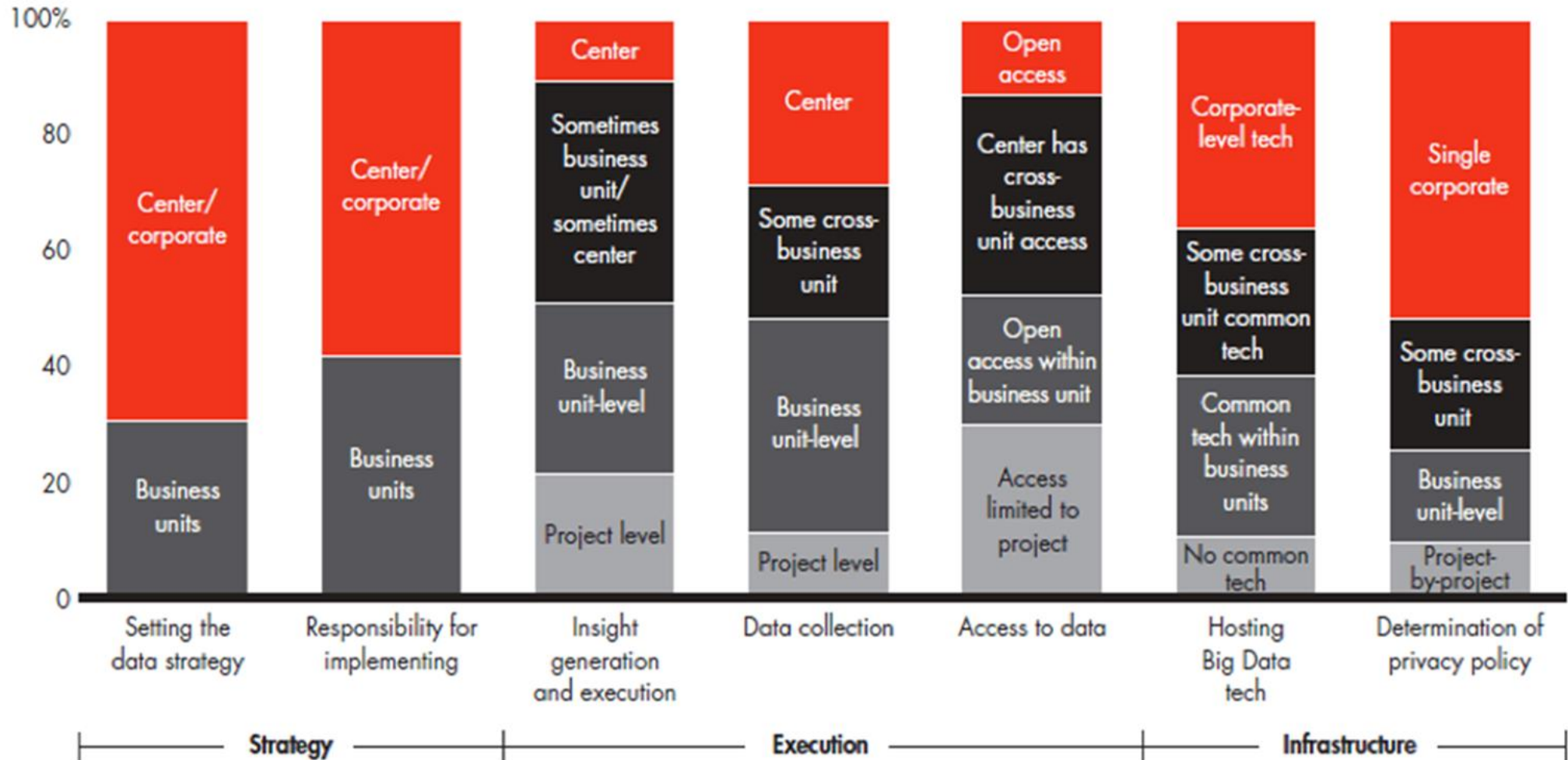
- 데이터 유지관리
(생성, 업데이트, 삭제)
- 데이터 품질
- 데이터 통합
- 데이터 생애주기
(생성 및 아카이브)

How

전통적인
마스터데이터
관리

데이터 위주 프로젝트 구성 요소

Percent of respondents

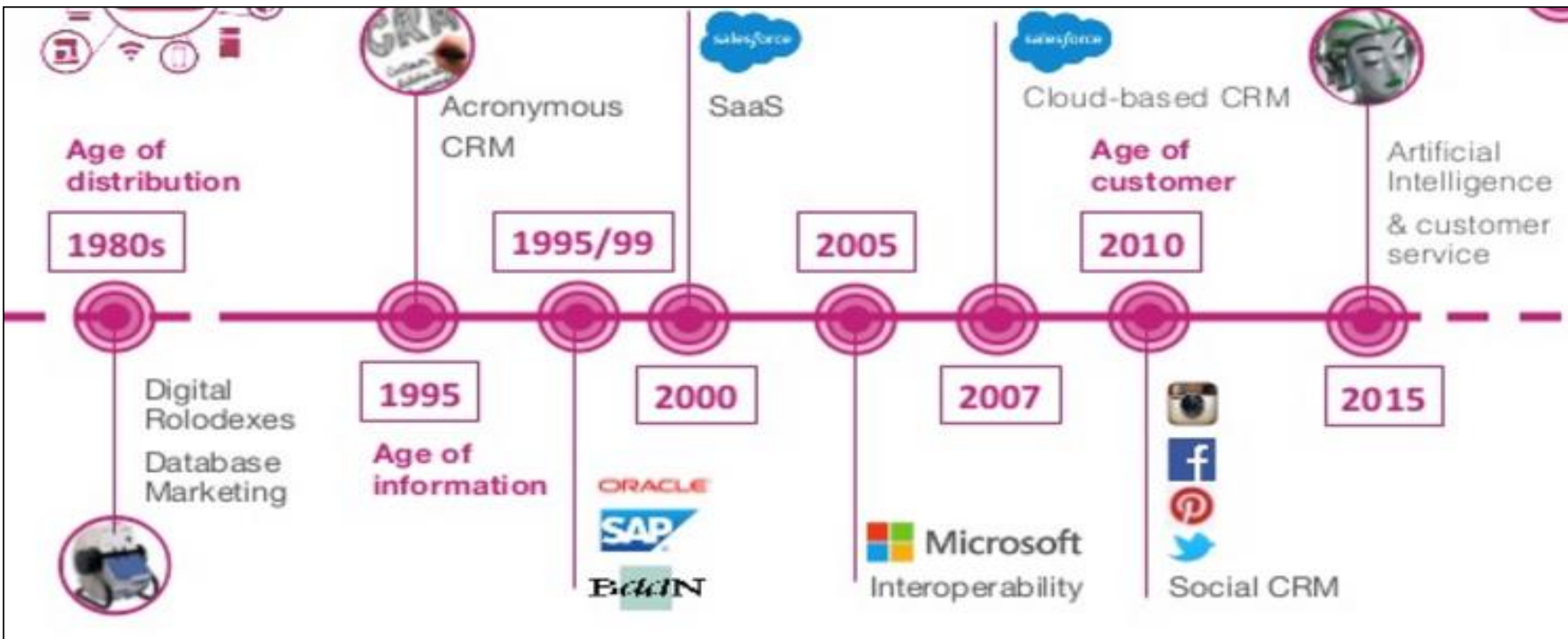




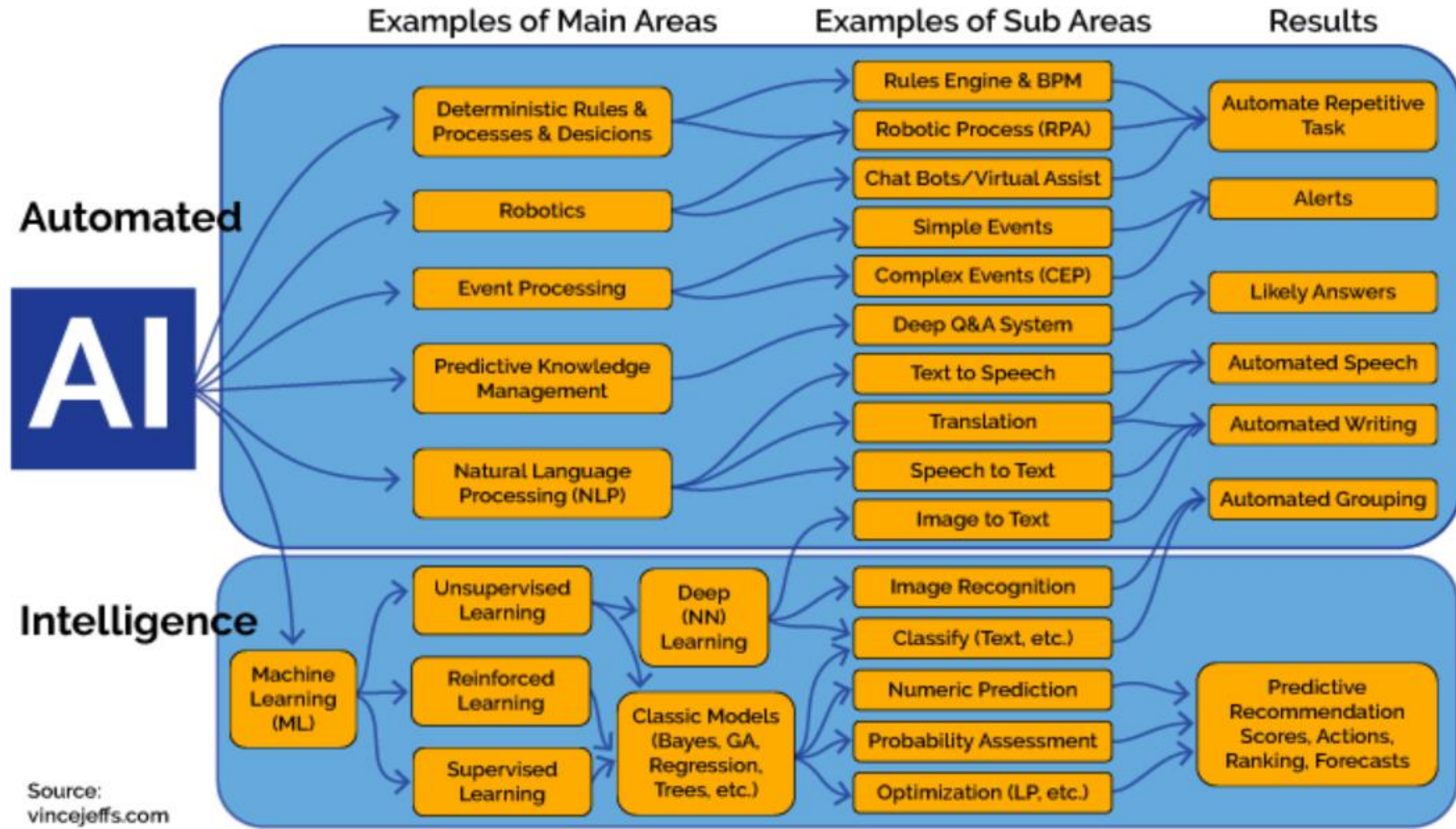
미래 시장 교란 기술들



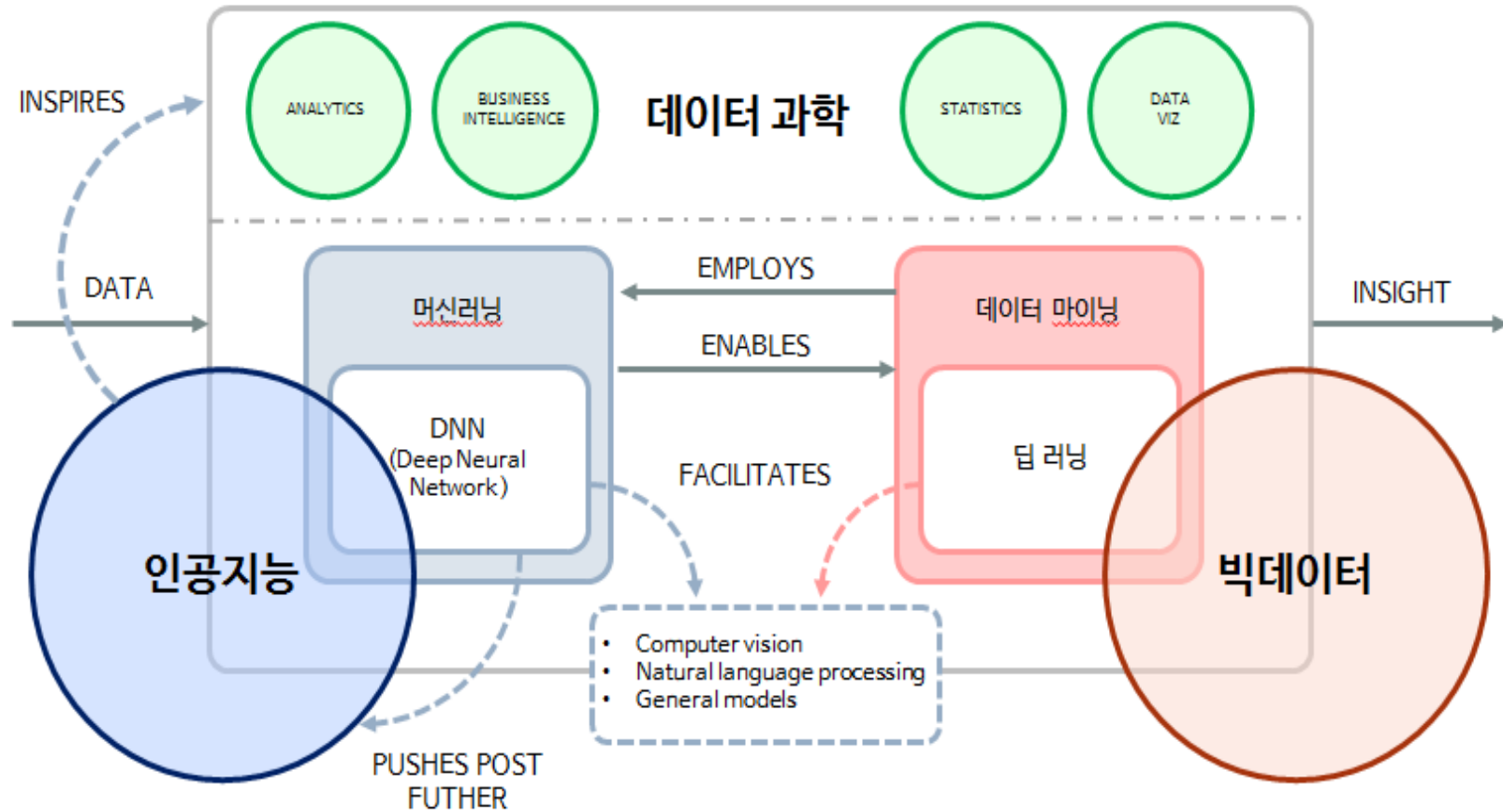
미래 시장 교란 기술들: CRM의 발전



Artificial Intelligence ?



Data Science vs. Artificial Intelligence vs. Big Data (IoT) 관계



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**)



접근

브라우저, 모바일 디바이스, 배치 작업

웹 서비스, FTP, Sockets

분석 결과 데이터베이스

대량 데이터의 최적화 접근

OLAP, RDBMS, Mem/Cached



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



신호 생성

데이터 신호, 패턴 구별

신호 생성 알고리즘들

시계열, 통계, 이벤트, 지리 위치 등



ETL, 가공

데이터 저장 (data store)

데이터 저장 (data store)

데이터 저장 (data store)

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%

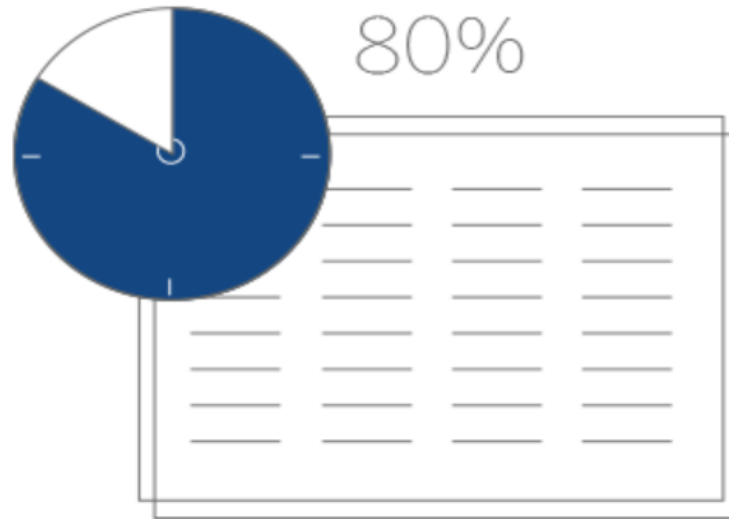
수련된

경험

있는

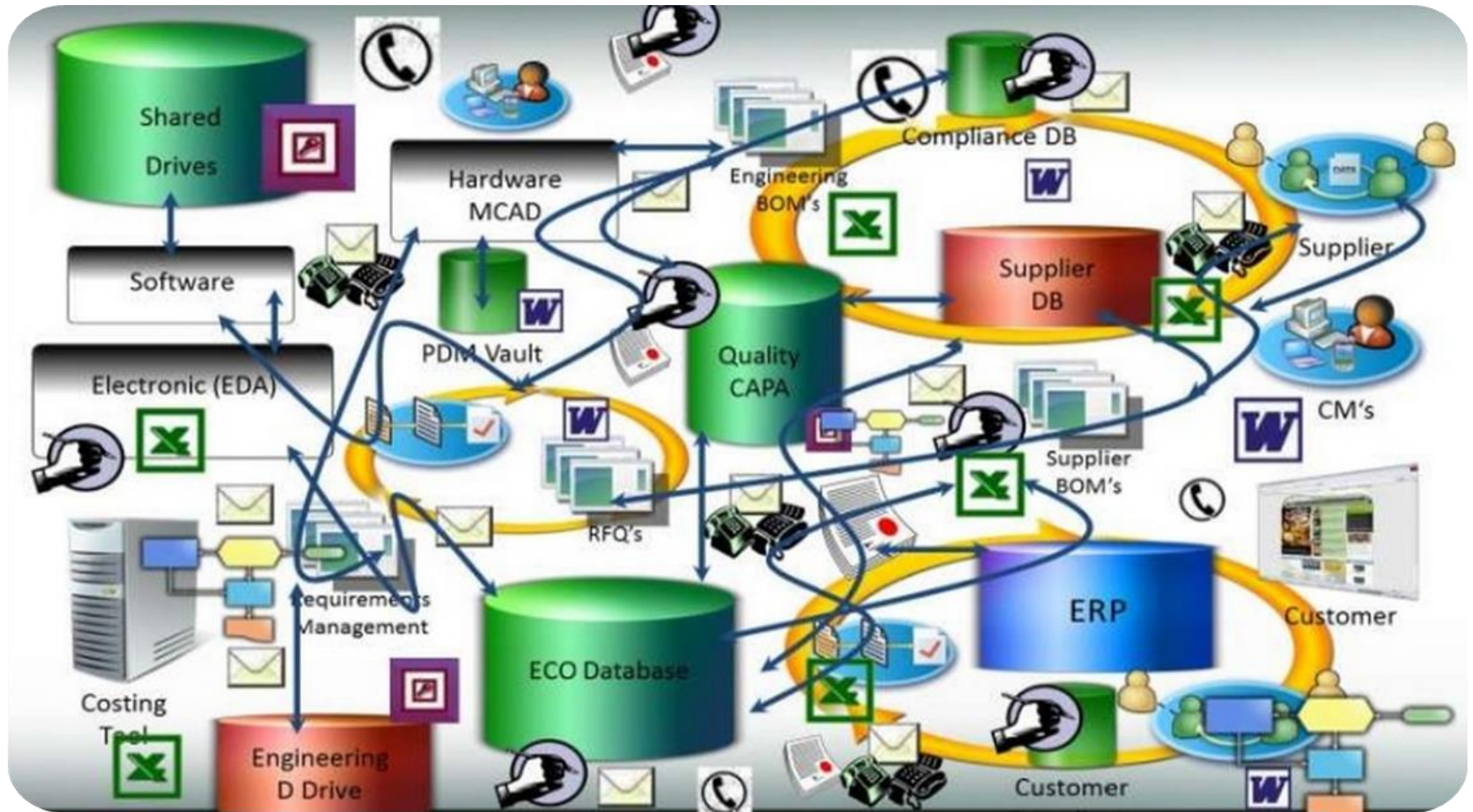
인력이

수행



Data scientists and analysts can spend as much as 80% of their time cleaning and preparing data. (TechRepublic)

데이터 통합의 필요성



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)

The Data Disconnect



데이터 플랫폼 운영상의 문제점: 데이터가 저장만 되어 있고 정제, 가공 융합이 부족



데이터

- 다양한 내·외부 데이터의 안정적 활용
- 경쟁력 있는 데이터 가공, 개발
- 고품질 데이터 사용 가능
- 서비스에 맞는 융합 데이터 구성



지능형 서비스

서비스

- 수요자 맞춤형 융·복합 서비스 가능
- 고객 중심의 지속 가능한 서비스 및 데이터 개발
- 내부 정책 대응 및 의사결정을 위한 데이터 기반 서비스 가능



플랫폼

- 수집-가공-처리-저장-분석-활용에 이르는 전반적인 기능을 수행하는 플랫폼 설계, 확보
- 체계적인 빅데이터 활용 가능 기반 마련



거버넌스

- 조직 및 구성원들에 대한 내재화 및 이해도 향상 가능
- 데이터 활용 관점에 전체 데이터 흐름 프로세스 확립으로 유기적 데이터 활용 가능
- 전사데이터의 통합 및 융합적 분석 가능



데이터 가공 플랫폼 : 데이터 거버넌스의 중요성



빅데이터 활용 기반 데이터 거버넌스를 통한 전략적인 종합 데이터 관리 기반

전략: 기관별 데이터 특성에 맞는 연계, 공유 및 관리 방안 마련

AS IS 데이터 홍수 및 연계 미비로 데이터 활용 저하

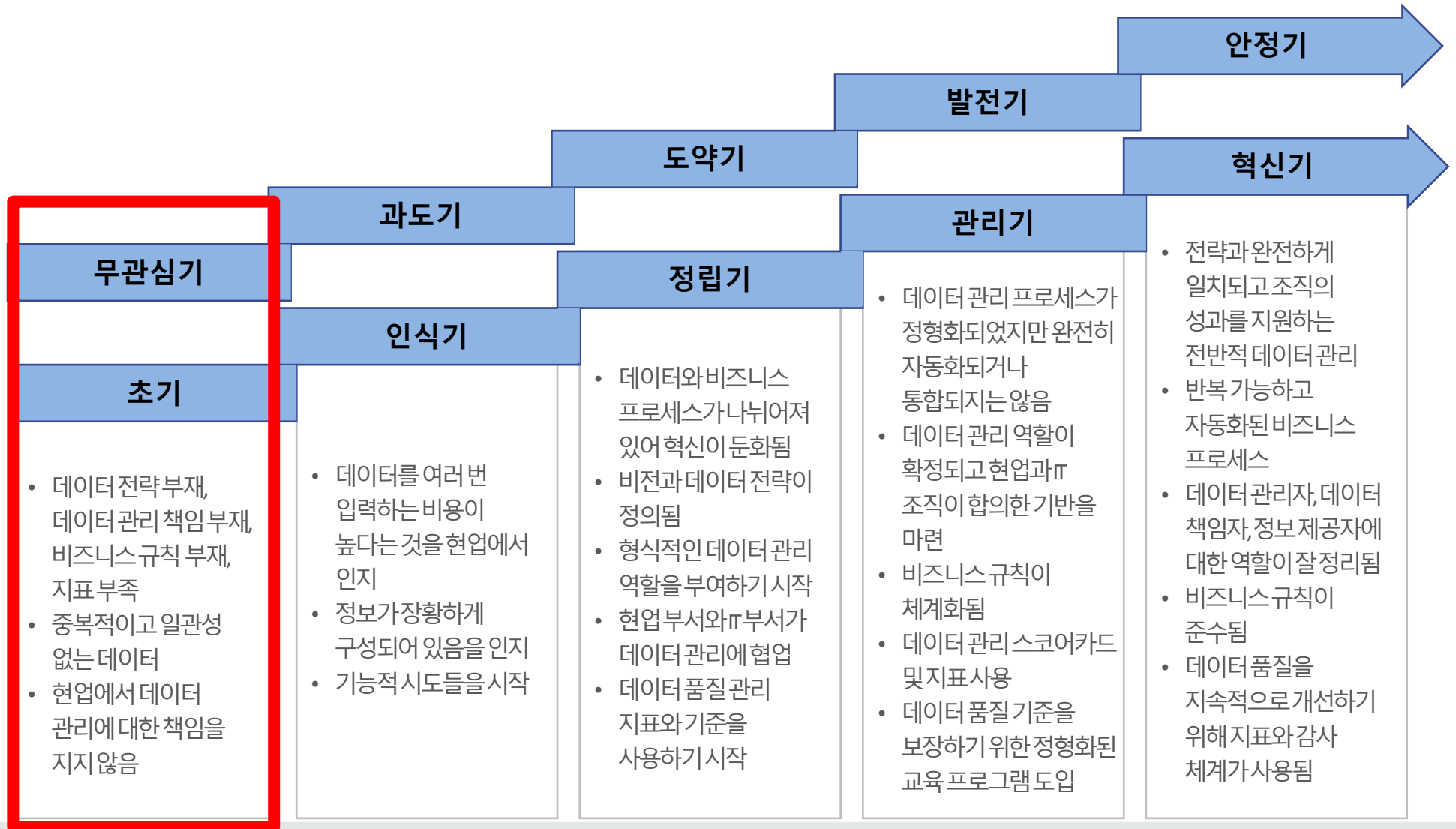
빅데이터 거버넌스 체계 마련

TO BE 데이터 간 융합, 재사용성 증가
내/외부 데이터 간 연결. 가치 발견

제안사 방법론(EDF) 기반

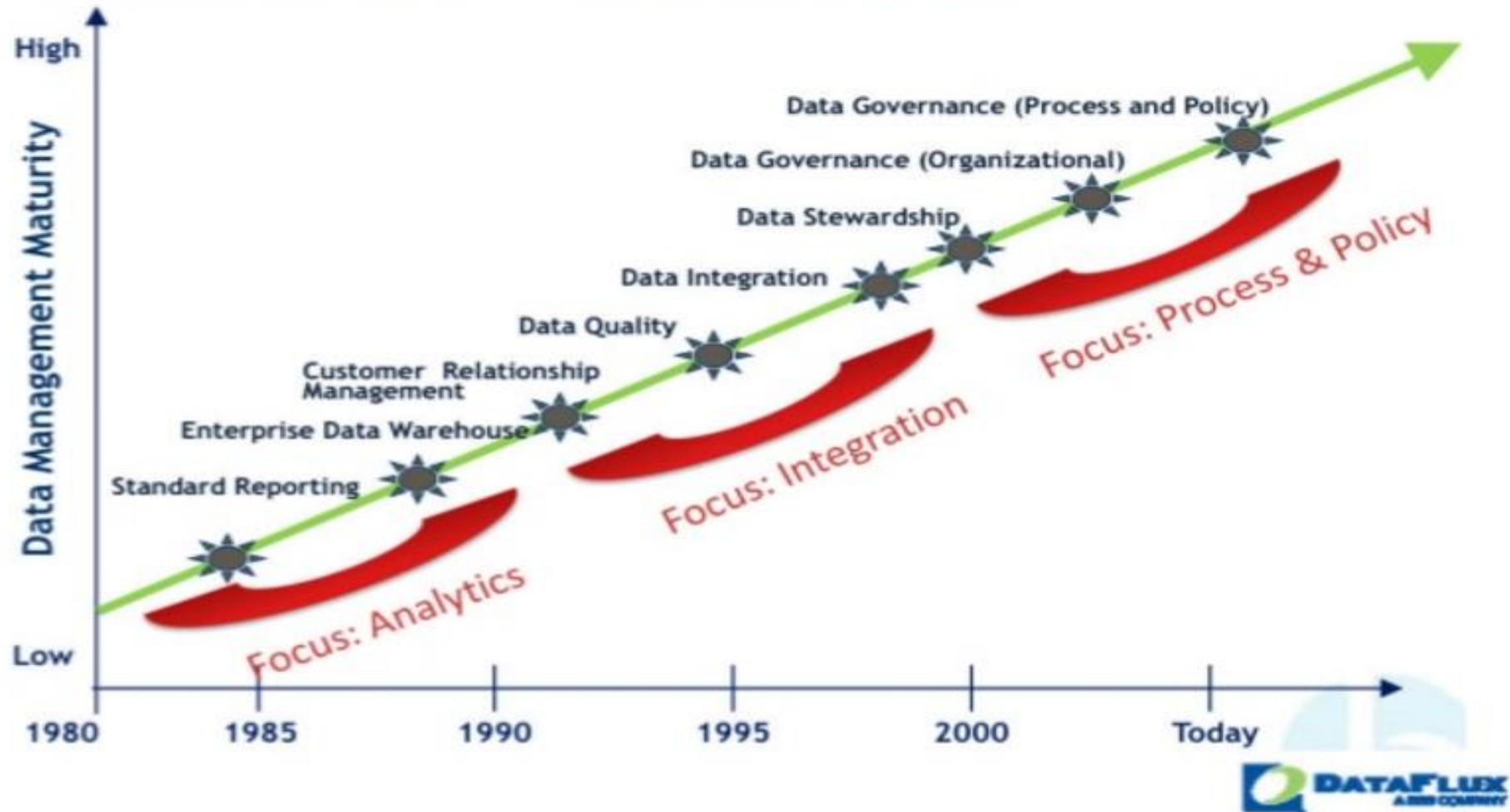
- ①원칙/정책
- ②조직
- ③프로세스
- ④데이터 (콘텐츠)
- ⑤시스템 (인프라)

데이터 거버넌스의 발전 단계와 현황

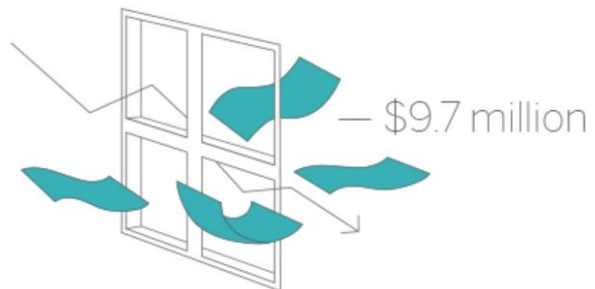


미국의 전사 데이터 거버넌스 발전 역사

데이터 거버넌스 인식의 단계



데이터 가공 플랫폼 : 데이터 거버넌스의 중요성



A 2016 Gartner study found respondent organizations were losing an average of \$9.7 million annually as a result of poor data quality.

DATA

15% have insurance

Only 15% of US companies have an insurance policy specifically for their data. (Ponemon)

Business Drivers for taking a Data-Driven Approach

Address the Data Quality Issues that put you at Risk

Data Quality is and has been a primary problem in project failures – and the issue isn't going away

- According to the Standish Group, in 1998, 74 percent of all data migration projects either overran or failed, resulting in almost **\$100 billion** in unexpected costs
- In a survey of 300 IT executives conducted by Information Week, the majority of the respondents (81 percent) said, "**improving (data) quality was their most important post-year 2000 technology priority**"
- Data Quality issues lead to **87% of projects requiring extra time** to reconcile data – TDWI Data Quality Survey, December, 2001
- Data Quality issues lead to **lost credibility** within a system in 81% of cases – TDWI Data Quality Survey, December, 2001
- A TDWI study shows that **poor data quality** costs U.S. businesses an estimated **\$600 billion a year**. TDWI Data Quality Survey, February, 2002
- According recent studies (2005) to the Gartner and Meta Groups, **55-70% of CRM and 70% of Data Warehouse project failures are due to data quality issues**
- Through 2005, more than **50% of CRM deployments** will suffer limited acceptance, if not **outright failure**, because of the lack of attention to data quality issues., Gartner, 2005
- In a recent report, Gartner predicted 50% of **data warehouse projects** through 2007 will have limited acceptance or be outright failures, as a result of lack of attention to data quality issues

As per the new UK Data Protection Act, "information must be accurate, up-to-date and held for no longer than necessary". Under this new law, companies have to comply with rules for manual and electronic processing of personal data, such as name, addresses and dates of births.

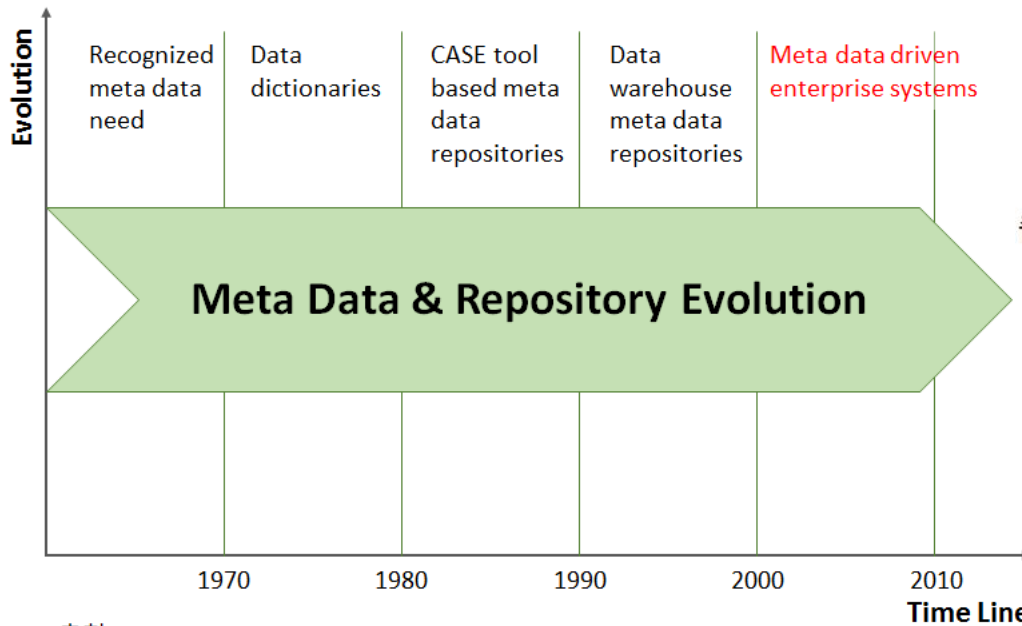
데이터 가공 플랫폼 : 전사 통합 데이터 거버넌스의 중요성



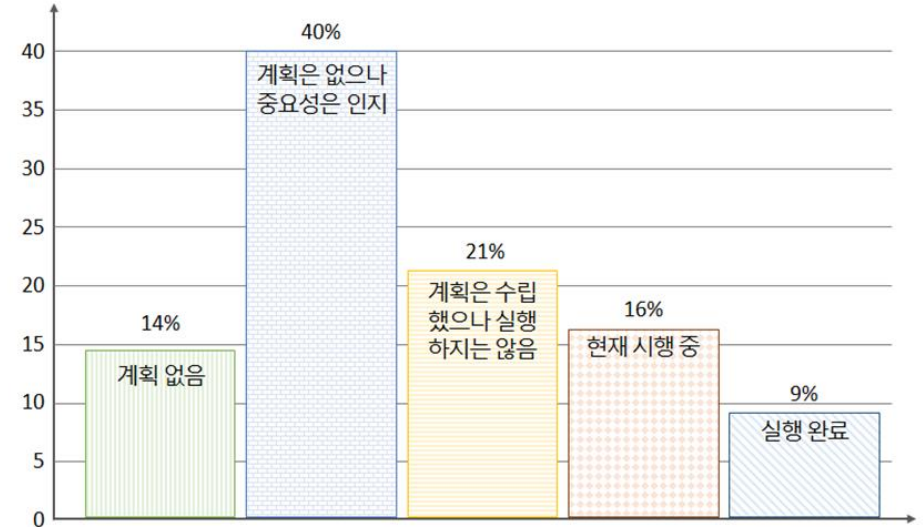
Figure 1 - University of Michigan Medical School Data

메타 데이터의 발전

Evolution of meta data and meta data repositories



Deploying a meta data strategy



출처: The data warehouse Institute(1999) :175명 대상

The Hard Work doesn't go away with Big Data

The 5th “V” - Veracity

- Only through proper Governance, Data Quality Management, Metadata Management, etc., can organizations achieve the 5th “V” – Veracity.
 - **Veracity:** Trust in the accuracy, quality and content of the organizations’ information assets.
- i.e. The hard work doesn’t go away with Big Data

Data Science

Raw data used in Self-Service Analytics and BI environments is often so poor that **many data scientists and BI professionals spend an estimated 50 – 90% of their time cleaning and reformatting data** to make it fit for purpose.¹⁴

Source: DataCenterJournal.com

Data Lakes

The absence of commonly understood and shared metadata and data definitions is cited as one of the main impediments to the success of Data Lakes.

Source: Radiant Advisors

Data Science

Correcting poor data quality is a Data Scientist’s least favorite task, consuming on average 80% of their working day

Source: Forbes 2016

Digitization & Data Quality

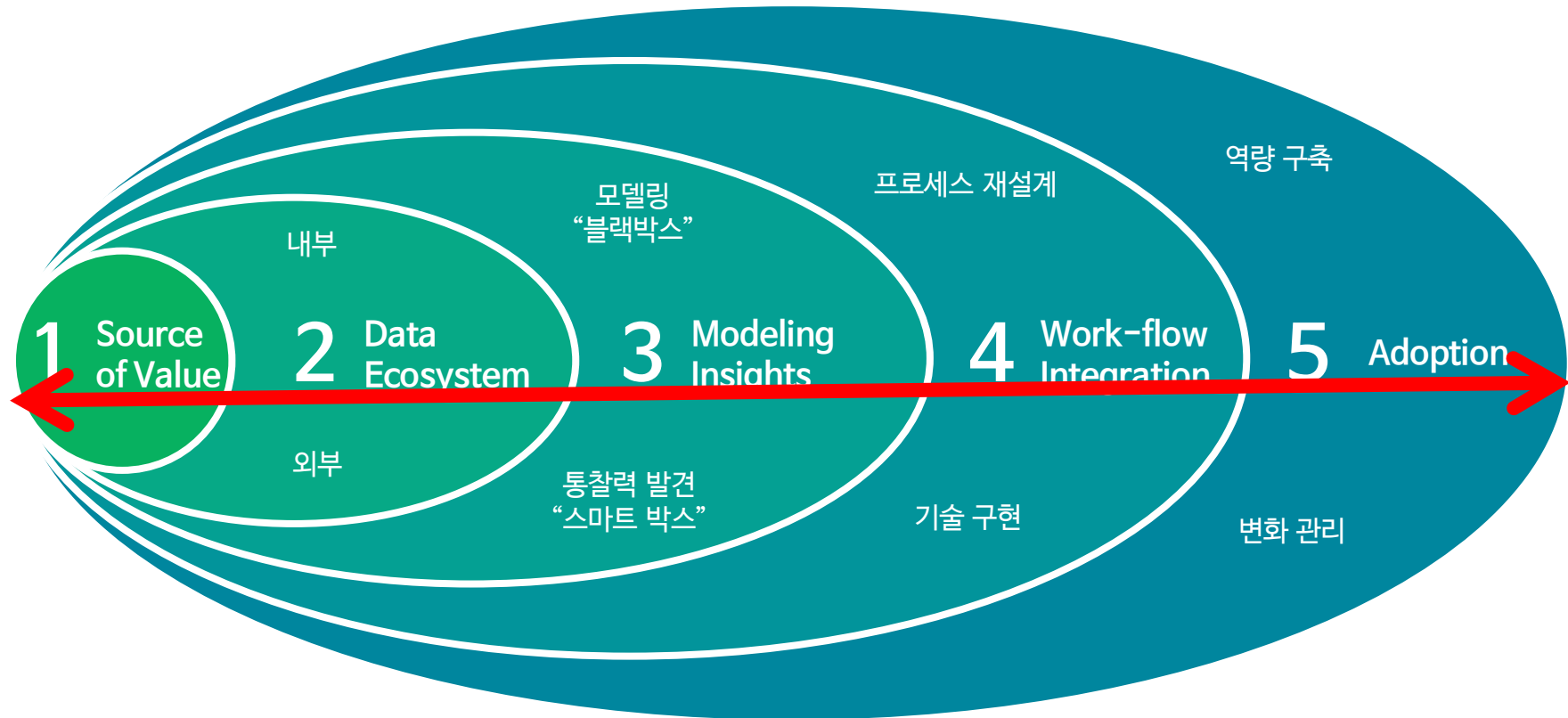
71% of interviewees expect digitization to grow their business. But 70% say the biggest barrier is finding the right data; 62% cite inconsistent data

Source: Stibo Systems



성공적 데이터 활용의 5가지 요소: Closed Loop(데이터는 생물)

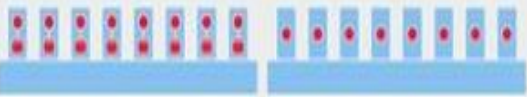


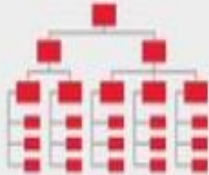





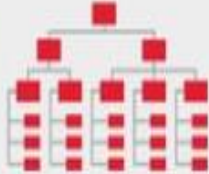





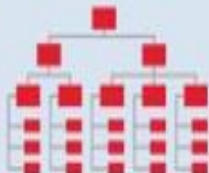


데이터 활용의 성공을 위한 Key Factors



성공적 데이터 활용은 빠르고 정확한 정보 기반의 의사결정으로 인한 기업의 지속 성장(Sustainability)

by Richard Clarke and Ari Libarikian

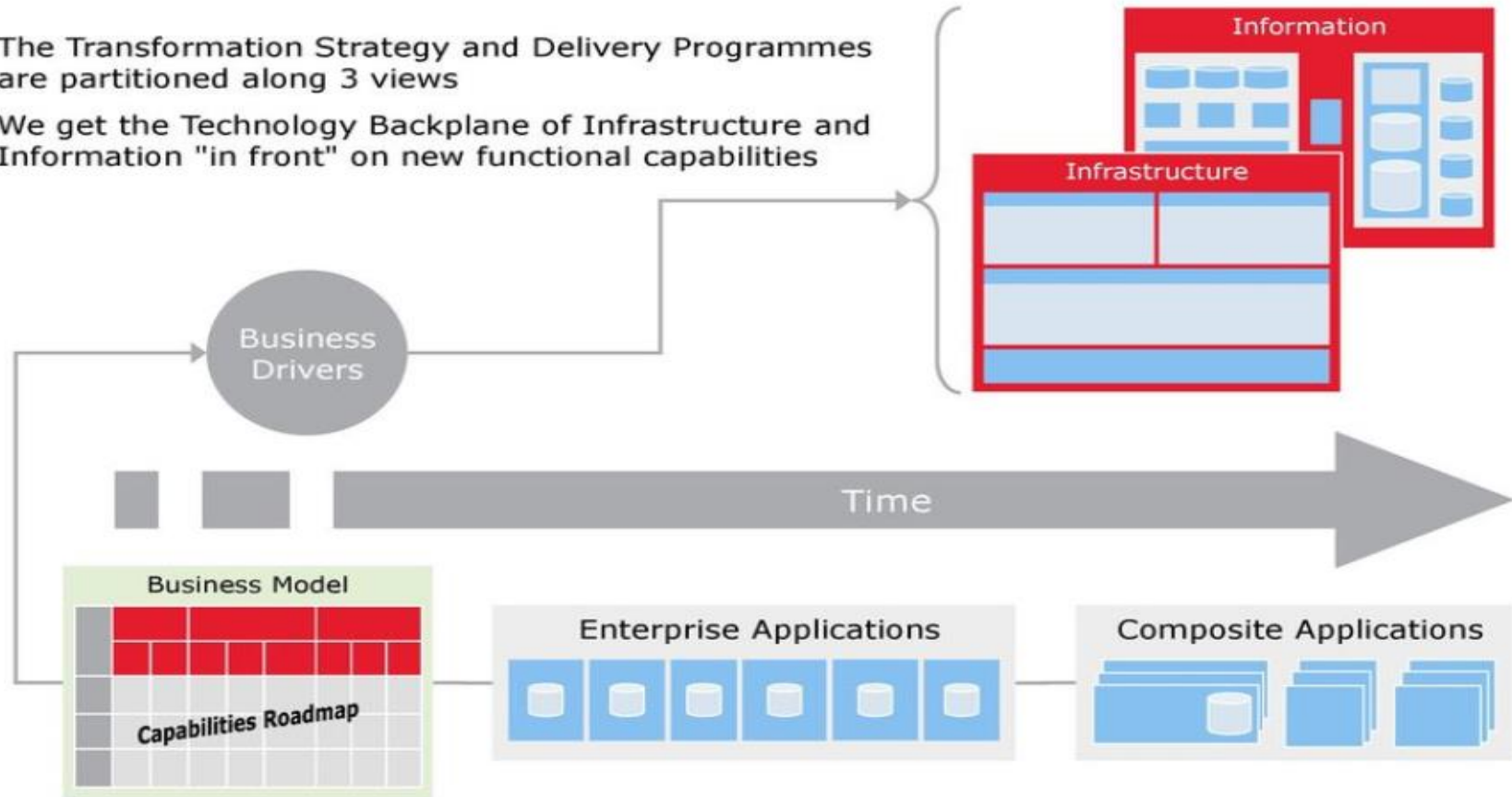
Artificial Intelligence ?

	Strategy	Process	Organisation	Technology	People
 Application Development					
 Infrastructure Development					
 Information Development					

비즈니스 먼저 그리고 데이터, IT SI 구축 나중

The Transformation Strategy and Delivery Programmes are partitioned along 3 views

We get the Technology Backplane of Infrastructures and Information "in front" on new functional capabilities



© 2008 BearingPoint, Inc.

CROSS

데이터 관련 산업의 본질 파악이 최우선

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



빅데이터 활용 혁신 사례: Capital One

start to attract customer accounts worldwide. It is not hard to see why Capital One is investing heavily in digital technologies. It conducts over 80,000 big data experiments a year². Currently, 75% of customer interactions with Capital One are digital, and this number is only expected to grow³. In Q4 2013, Capital One was one of the most visited websites, with 40 million unique online visitors⁴.

“
Capital One conducts over 80,000 big data experiments a year.
”

to make digital part of our business not only with our customers, but also how we operate the company⁷.”

“
Digital is who we are and how we do business.
”

- Richard Fairbank, CEO

This unrelenting focus on digital has underpinned Capital One's strong performance. Despite the global recession, Capital One has maintained sector-leading growth and steady profits. For instance, from 2005 to 2013, Capital One achieved a CAGR in profit-before-tax of 10.78%, significantly

outpace the industry peers. The bank's digital service culture is supported by rapid prototyping capability, which helps deliver new tech-based features faster, as well as real-time analytical tools. These digital capabilities are the deliberate result of a long-term strategy that Capital One has had in play over a number of years, and which we look at in the following section.

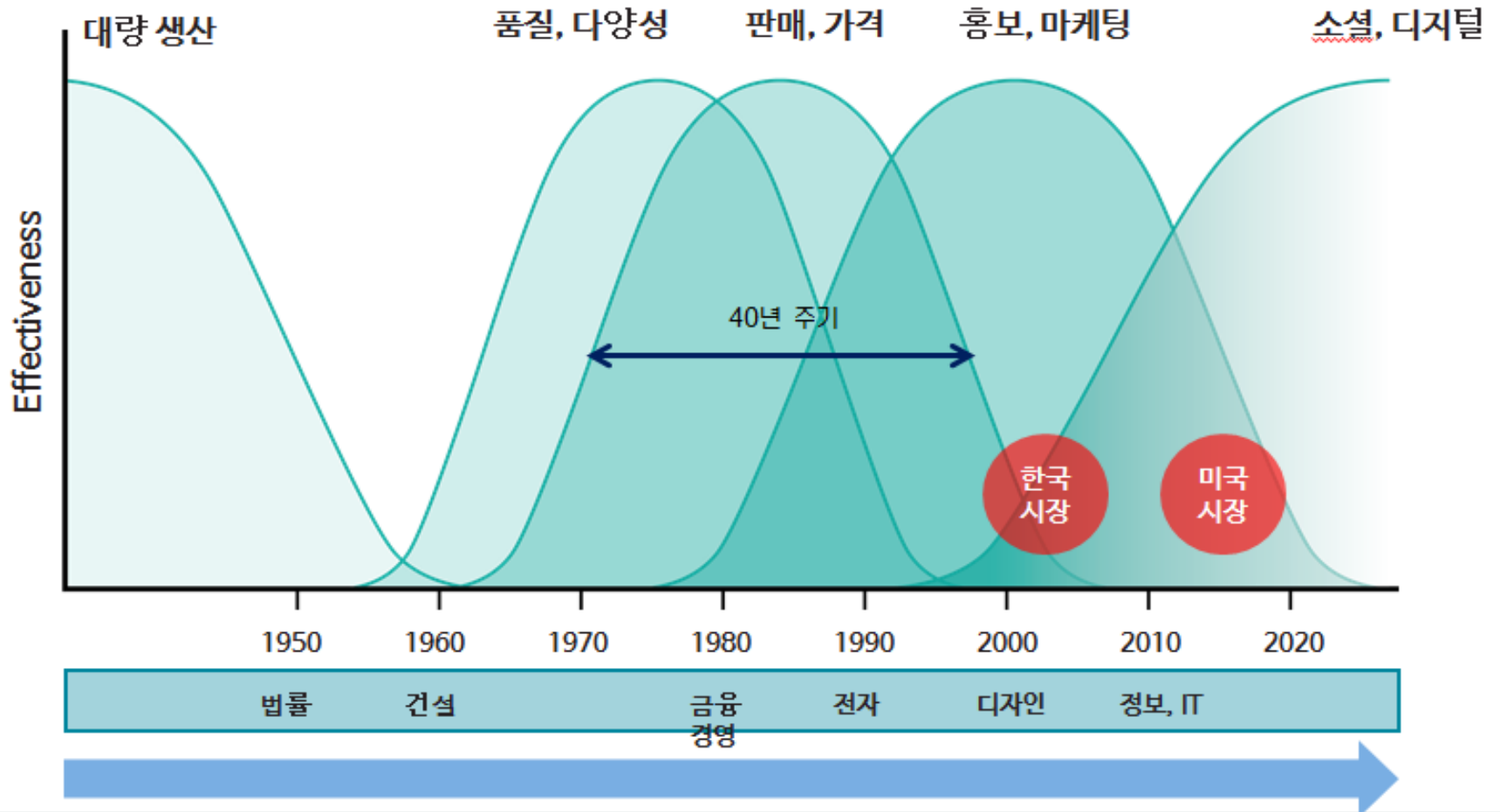
“
I think it's a bit of a fool's errand... to chase digital for the sake of cost reduction.
”

- Richard Fairbank, CEO

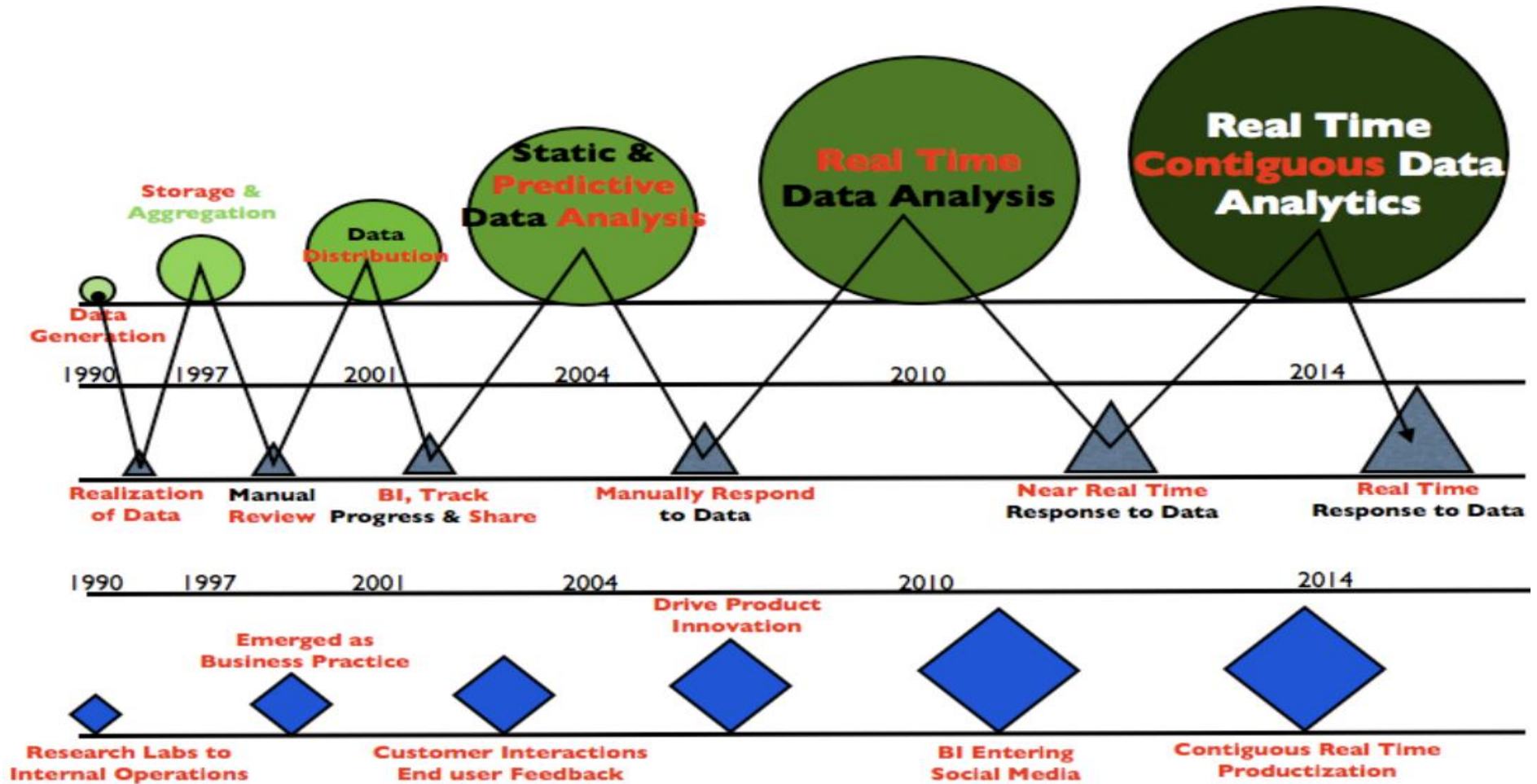
Figure 1: Profit-Before-Tax and Comparative CAGR Analysis of Capital One and Peers

핵심 경쟁 비즈니스 사이클의 따른 분석 트렌드

비즈니스 패러다임의 발전

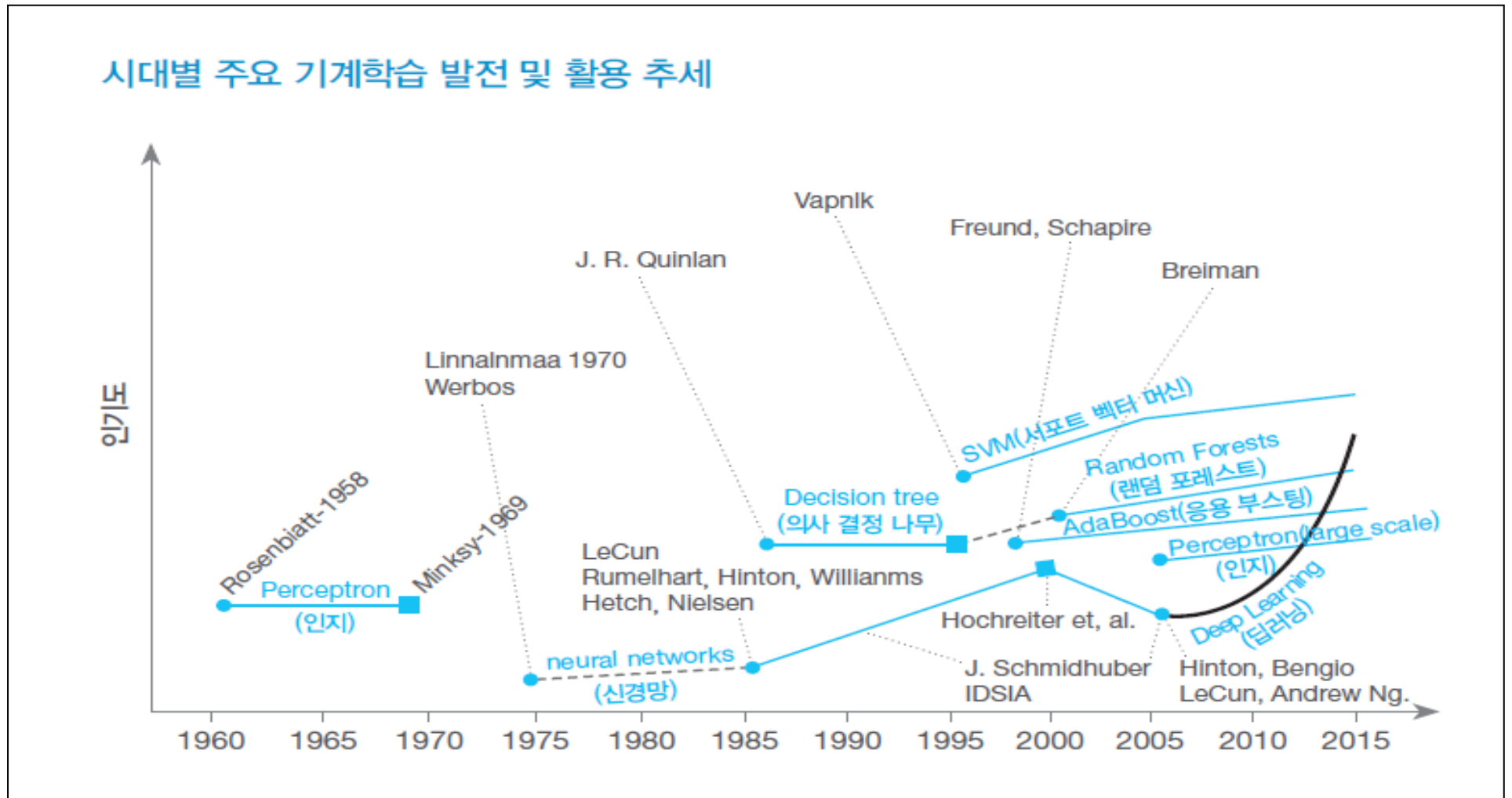


디지털화 되기까지의 분석의 발전



Created by Vamsi Sistla - no formal research was conducted. Timelines are guesstimates. Sept 2009

시대별 데이터 분석(알고리즘) 단계



시대별 데이터 과학의 발전 단계

Computer Science

- Leibniz – Binary Logic.
- Turing machines
- Information Theory
- Weiner & Cybernetics
- Von Neumann Architecture.
- Babbage, Lovelace
- Boolean Algebra
- Punch cards.
- Sort & Search Algorithms – Dijkstra, Kruskal, Shell Sort, ...
- Heuristics – Simulated Annealing, ...
- Graph Algorithms
- Multigrid methods
- Tree based methods.
- Text/ string search
- 1974 Peter Naur "Concise Survey of Computer Methods", **Data Science, Datalogy**
- Knuth – Art of Computer Programming.
- Database Marketing
- Data Mining, Knowledge Discovery
- "Data science, classification, and related methods."

Data Technology

- Cartography
- Astronomical Charts.
- William Playfair
- Charles Minard
- Florence Nightingale.
- First IBM Computers
- DBMS.
- Removable Disk drives
- Relational DBMS.
- Desktop, floppy
- SQL, OOP
- High level languages.
- 1989 First KDD Workshop
- Gregory Piatetsky-Shapiro.
- William Cleveland: Data Science
- Leo Breimann: Statistical Modeling: 2 Cultures.

Visualization

- Optimization Methods
- Fourier and other transforms
- Matrix & Generalizations
- Non-euclidean geometries.
- Applications to Military, manufacturing, Communications.
- Networks
- Assignment Problems
- Automation
- Scheduling
- John Tukey
- Jacques Bertin.
- Edward Tufte.
- Grammar of Graphics
- Word Cloud, Tag Cloud.

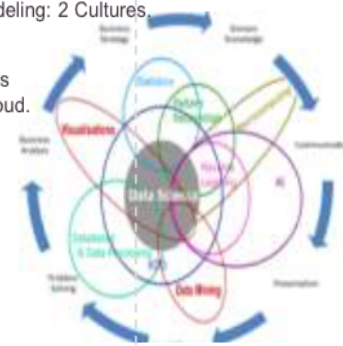
Mathematics/ OR

- Calculus
- Logarithms
- Newton-Raphson.
- Theoretical Foundations of Modern Stats
- Hypothesis, DOE
- Mathematical Statistics.
- 1962 John W. Tukey, Future of Data Analysis
- 1976 – SAS Institute
- 1977 The International Association for Statistical Computing (IASC).
- Decision Science
- Pattern recognition
- Machine learning.

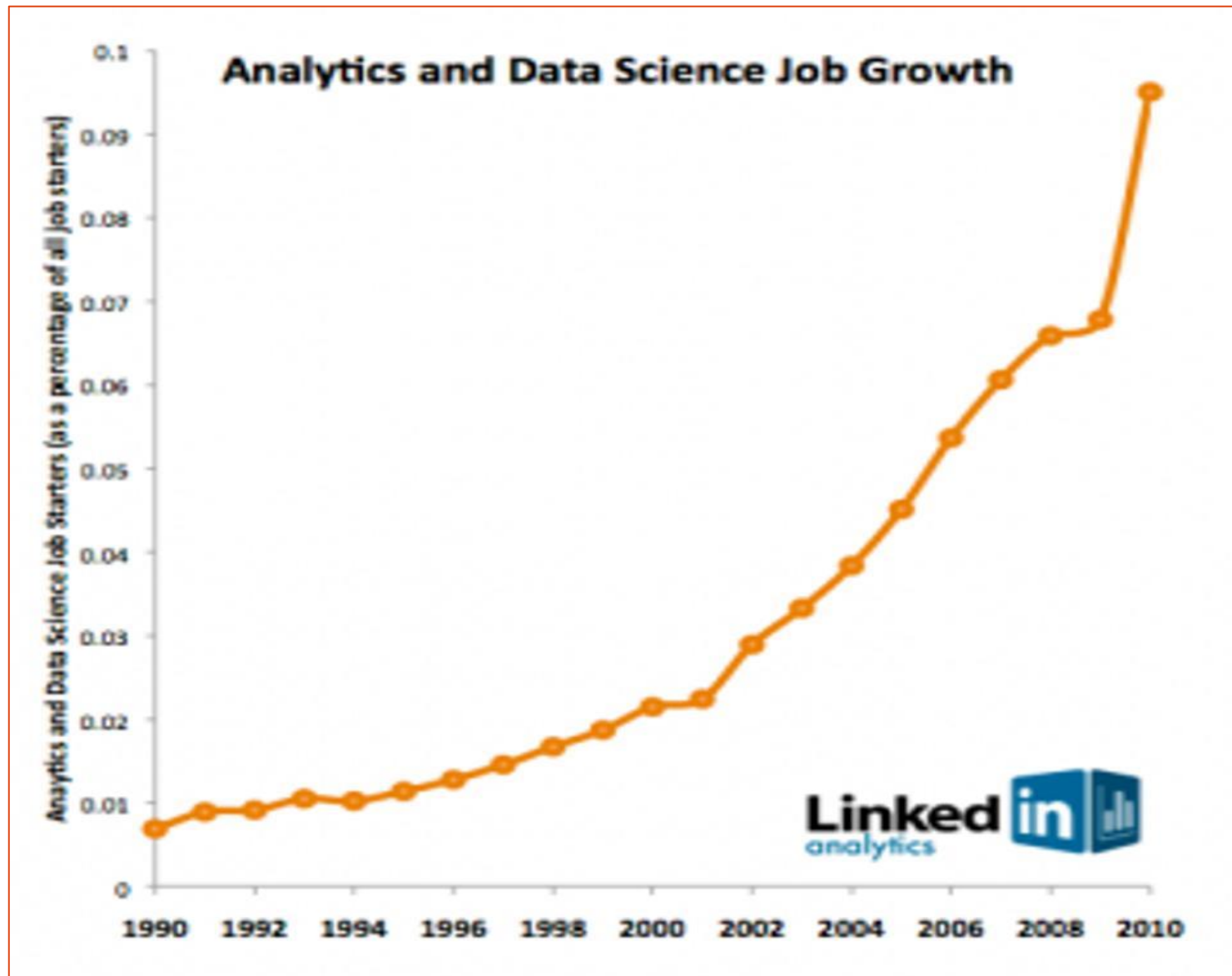
Statistics

- Probability
- Correlation
- Bayes Theorem.
- Regression, Least Squares
- Time Series.
- Bayesian Methods
- Time Series Methods (Box Cox, Survival, etc.)
- Stochastic Methods.
- Simulation, Markov
- Computational Statistics.

Statistics



시대별 데이터 과학 추이



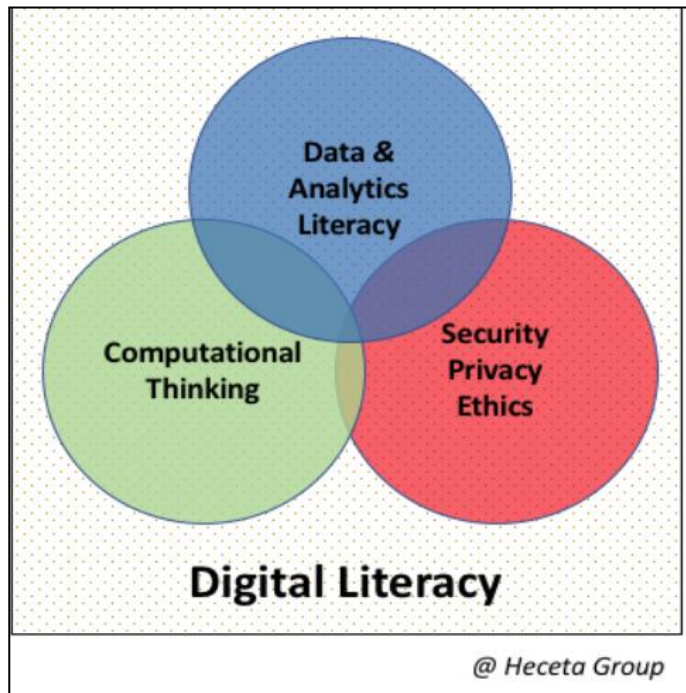
Data Literacy



One Vision. One Durham.

Data-to-Action: Building Middle School Administrators' and Teachers' Data Literacy Capacity in Durham Public Schools

2014 North Carolina Association of Middle Level Educators Conference (NCAMLE)
March 17, 2014



Skills of a Data Literate Person



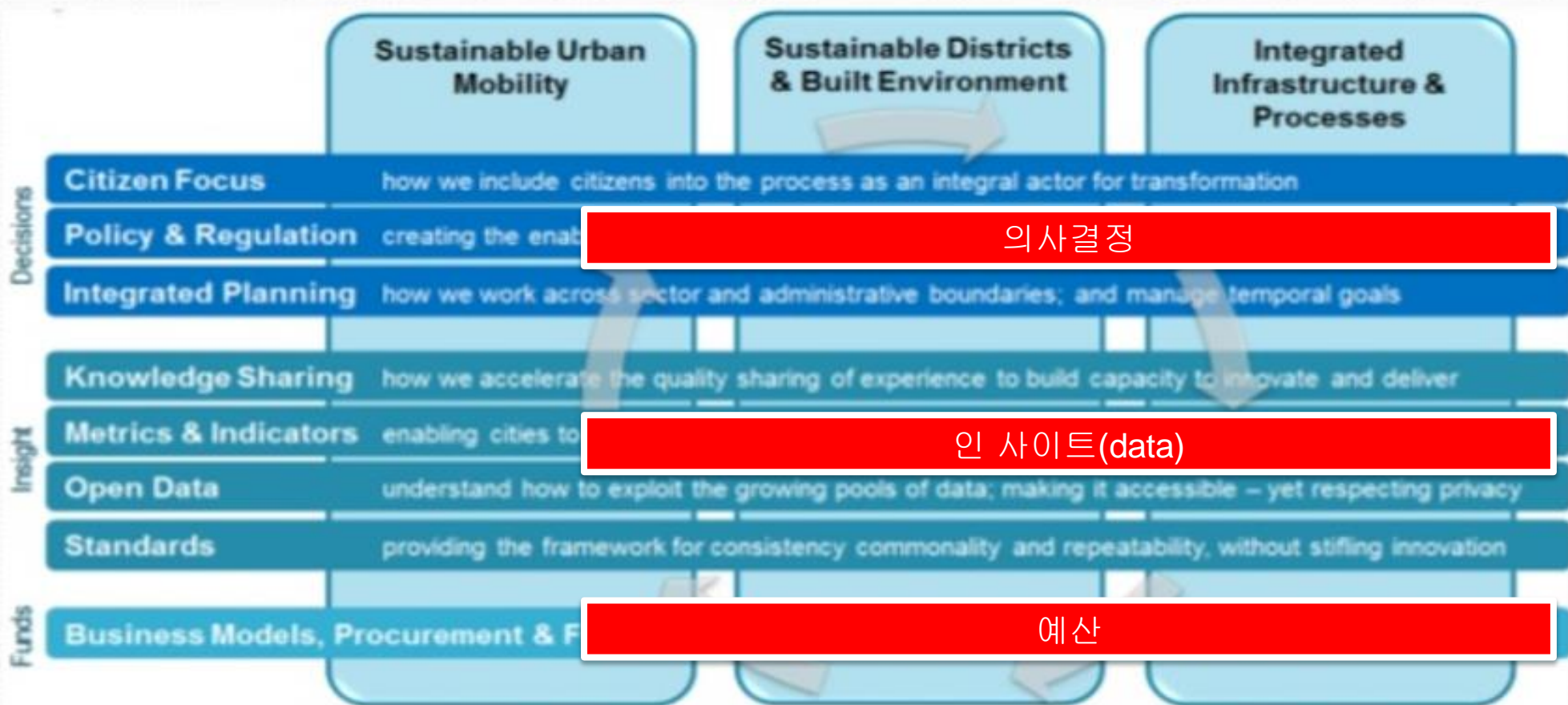
데이터 과학 Maturity Model

Phase	The Old World			The New Era
	Pilot	Departmental Analytics	Enterprise Analytics	Big Data Analytics
Staff Skills (IT)	Little or no expertise in analytics – basic of knowledge BI tools	Data warehouse team focused on performance, availability and security	Advanced data modelers and stewards key part of the IT department	Business Analytics Competency Center (BACC) that includes 'data scientists'
Staff Skills (Business/IT)	Functional knowledge for BI tools	Few business analysts – limited usage of advanced analytics	Savvy analytical modelers and statisticians utilized	Complex problem solving integrated into Business Analytics Competency Center (BACC)
Technology & Tools	Simple historical BI reporting and dashboards	Data warehouse implemented, broad usage of BI tools, limited analytical data marts	In database mining, usage of high performance computing & analytical appliance	Widespread adoption of appliances for multiple workloads. Architecture and governance for emerging technologies
Financial Impact	No substantial financial impact. No ROI Models in place	Certain revenue generating KPIs in place with ROI clearly understood	Significant revenue impact (measured and monitored on a regular basis)	Business strategy & competitive differentiation is based on analytics
Data Governance	Little or none (Skunk works)	Initial data warehouse model and architecture	Data definitions & models standardized	Clear master data management strategy
Line of Business	Frustrated	Visible	Aligned (including LoB executives)	Cross-departmental (with CEO visibility)
CEO Engagement	Hidden	Limited	Involved	Transformative

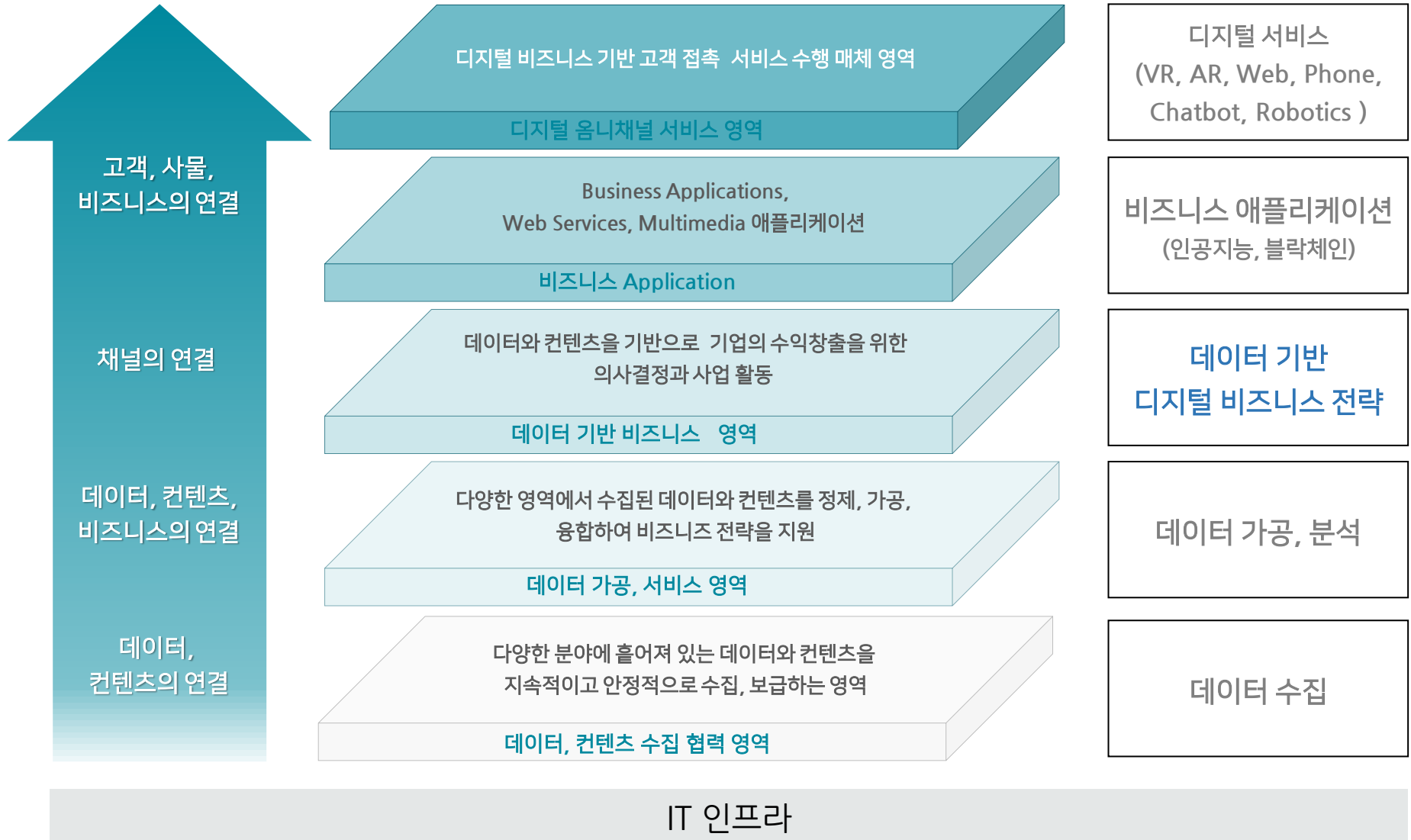
Source: IDC Asia/Pacific Business Analytics Practice (July, 2011)

스마트 시티 성공 요소 사례

- *The High Level Group of the European Innovation Partnership for Smart Cities and Communities* presents a **Strategic Implementation Plan (SIP)**
- **SIP** concentrates on three vertical & eight key horizontal areas [8]



4차 산업 혁명을 위한 **Connectivity** 전략: 데이터 사이언스



Thank you