



Data Science and its role in Digital Economy

WE ALL LIVE IN A DIGITAL ECONOMY!

- I. 글로벌 트랜드, 데이터과학, 인공지능
- II. 데이터과학과 디지털경제
- III. 데이터와 인공지능을 활용한 디지털전환

Kyungwon Kim

Assistant Professor
Department of International Trade
College of Global Political Science and Economics
Incheon National University

WE ALL LIVE IN A DIGITAL ECONOMY!

I. 글로벌 트랜드, 데이터과학, 인공지능

- Global “Business” Trends
- Global “Technology” Trends
- Global “Academia” Trends
- “Artificial Intelligence” vs. “Data Science”

II. 데이터과학과 디지털경제

- What is the Communication between Data and the World?
- Why Data Science is in a sudden boom?
- What is a Big Data?
- What is a Big Algorithm?
- Digital Platform with Big Data/Algorithm

III. 데이터와 인공지능을 활용한 디지털전환

- Evolution of Hardware for Digital Platforms
- Evolution of Computing Platforms
- Skills and Jobs in the Data Science
- Driving Digital Transformation with Data Science as a Service

WE ALL LIVE IN A DIGITAL ECONOMY!

I. 글로벌 트랜드, 데이터과학, 인공지능

- Global “Business” Trends
- Global “Technology” Trends
- Global “Academia” Trends
- “Artificial Intelligence” vs. “Data Science”

II. 데이터과학과 디지털경제

- What is the Communication between Data and the World?
- Why Data Science is in a sudden boom?
- What is a Big Data?
- What is a Big Algorithm?
- Digital Platform with Big Data/Algorithm

III. 데이터와 인공지능을 활용한 디지털전환

- Evolution of Hardware for Digital Platforms
- Evolution of Computing Platforms
- Skills and Jobs in the Data Science
- Driving Digital Transformation with Data Science as a Service

Global “Business” Trends

➤ Largest Companies 2008 vs. 2018, a lot has changed

- The US is where we are seeing more rapid changes in its biggest companies than any others.
- The **technology giants** are achieving world domination by investing heavily in developing **new products and services** leading to an explosion in innovation and faster growth.

Largest US companies in 2018 vs 2008

2018				2008			
Rank	Company	Founded	USbn	Rank	Company	Founded	USbn
1.	Apple	1976	890	1.	Exxon	1870	492
2.	Google	1998	768	2.	General Electric	1892	358
3.	Microsoft	1975	680	3.	Microsoft	1975	313
4.	Amazon	1994	592	4.	AT&T	1885	238
5.	Facebook	2004	545	5.	Proctor & Gamble	1837	226
6.	Berkshire	1955	496	6.	Berkshire	1955	206
7.	J&J	1886	380	7.	Google	1998	198
8.	JP Morgan	1871	375	8.	Chevron	1879	192
9.	Exxon	1870	367	9.	J&J	1886	192
10.	Bank of America	1909	316	10.	Walmart	1962	184

Source: Bloomberg, Google

Global “Business” Trends

➤ AI strategy by company type: Service, Product and Chipset

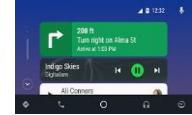
- Service companies increase contact with customers and improve their **convenience and efficiency**.
- Product and chipset companies are aiming to improve the performance of their **product technologies**.

	Company	Main Business Point	AI Application	AI Strategy
Service		검색을 통한 광고 수익	AI 서비스로 모든 제품 내 고객 접점 마련 → 강력한 고객 접점 확보로 광고 수익 ↑	인공지능 기술을 통해 고객 접점 및 생태계를 확대하고 이를 통한 광고, 온라인 커머스, S/W 서비스의 수익 제고
		S/W 솔루션/서비스 판매	AI Agent/App/Service/Infra에 AI 기술 적용 → OS, Cloud 사업의 경쟁력 제고	
		온라인 Retail 마켓 및 Cloud 서비스	Assistant 및 Cloud 서비스에 AI 기술 적용 → 개방형 AI로 더 많은 상품 구매 유도	
		SNS를 통한 광고 수익	자사 SNS 서비스 고도화에 AI 기술 적용 → 서비스 활용도 제고로 광고 수익 ↑	
Product		Mobile/Wearable 단말 판매	개인화, Privacy 강화된 Device 내 AI 기술 적용 → 차별화된 인공지능 가치 제공	차별화된 인공지능 가치 제공을 통한 Device 경쟁력 제고
		가전 기기 및 Mobile 단말 판매	가전 및 로봇 제품에 AI 기술 적용 → 개방형 전략으로 AI 생태계 구축 시도	
		통신 장비 및 Mobile 단말 판매	IoT/Device에 Intelligent 솔루션 제공 → AI + IoT 생태계 확대로 사업 입지 강화	
Chipset		PC/Data센터 Processor 판매	서버/디바이스별 AI 강화된 Processor 개발 → Processor 경쟁력 제고	AI 강화된 Processor를 통해 Chip 경쟁력 제고

Global “Business” Trends

➤ Representative Business Cases – Google (1/2)

- AI that can benefit everyone (Mobile-first □ AI-first □ **AI for Everyone**)
- There are three **pure AI-based services**, but **many of Google's services incorporate AI technology**.
- Strategies for maximizing advertising revenue by **strong customer connections**.

AI Application	Google Assistant	Google Lens	Google Translate	Android Auto
	<ul style="list-style-type: none">• Google이 개발하고 '16년 5월 자체 개발자 컨퍼런스서 발표한 지능형 가상 비서• Duplex, 이중언어 지원 等 차별화 	<ul style="list-style-type: none">• Google I/O 2017 기간 中 발표한 기능의 하나이며 시각 분석을 사용하여 관련 정보 제공 	<ul style="list-style-type: none">• 텍스트 번역을 위해 Google이 무료로 제공하는 다국어 언어 기계 번역 서비스 	<ul style="list-style-type: none">• 운전을 하면서 음성 액션, 네비게이션, 음악 감상을 쉽게 할 수 있는 Android OS 
Product				
AI Tool	TensorFlow		TensorFlow Lite / ML Kit	
Cloud	Google Cloud AI		Cloud IoT Edge	
Processor	Tensor Processing Unit (TPU)		Edge TPU	

Global “Business” Trends

➤ Representative Business Cases – Amazon (2/2)

- The expansion of AI ecosystem through **cloud (AWS) strengths** and opening the **voice AI platform (Alexa)**.
- **Shopping and life convenience** through Alexa Everywhere (Support TV, home appliances, headphones, speakers, door locks, etc.)
- AWS provides various AI services such as **voice recognition, image recognition, and chatbot**.

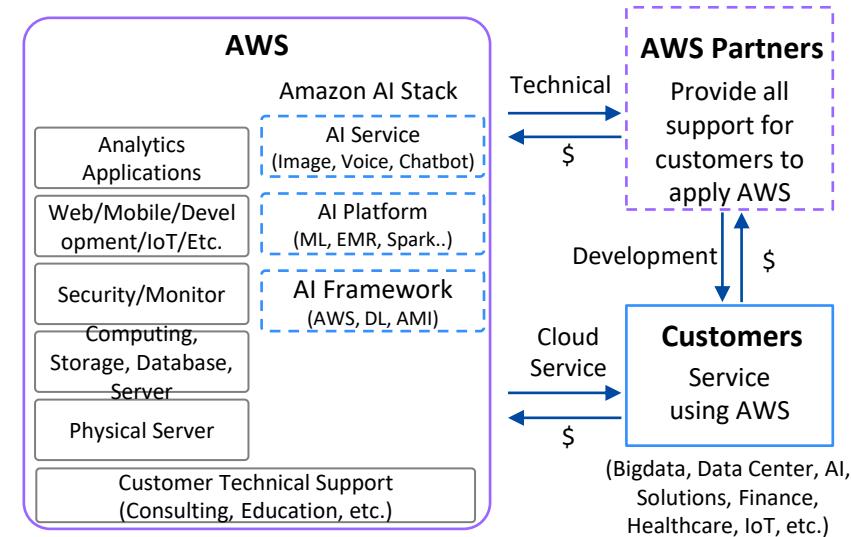
<Alexa Ecosystem>

- 23개 카테고리로 7만개 이상 skill(기능) 제공 ('19.10)
- Alexa 호환 기기 '18년초 4천개 '19년 3만개 이상으로 증가



<Amazon Web Service>

- Cloud 기반 서비스를 위한 Total Solution 제공
- IoT/Edge 기기 대상으로 솔루션 확대



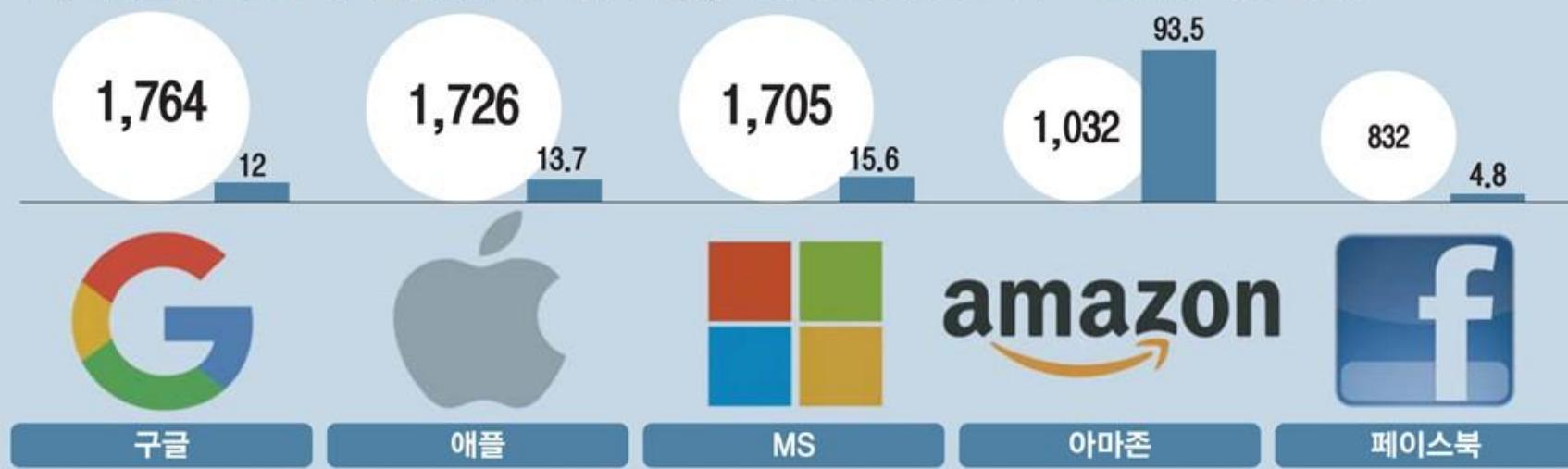
Global “Business” Trends

➤ Future business direction

‘IT 공룡’ 기업들의 주력 사업 확장 흐름

회사	기존 사업	현 성장동력	신규 투자사업
구글	검색	유튜브 · 클라우드	자율주행(Waymo) 등 인공지능 활용 사업
애플	아이폰 등 하드웨어	웨어러블	증강현실 글래스, 5G 폰 등 신규 하드웨어
아마존	전자상거래	클라우드	의약품 배달, 자율주행 전기차, 영화 등 전방위 확대
MS	윈도 등 SW	클라우드 · 게임	기업용 SW 시장
페이스북	소셜미디어	VR 디바이스 등	전자상거래, 결제 시장 등

미국 빅테크의 현금성 자산과 인력 (단위=억달러 · 만명) *올해 1분기 말 기준. 자료=각 사 IR ● 현금성 자산 ■ 인력



Global “Business” Trends

➤ Future business direction

글로벌 시장을 장악한 분야별 플랫폼 기업들

PLATFORM CATEGORIES AND COMPANIES



자료: Peter Fisk, Gamechangers

<https://www.thegeniusworks.com/2018/01/the-platform-revolution-how-network-based-markets-are-transforming-the-rules-of-business/platform-categories-and-companies-900x501/>

Global “Business” Trends

➤ Future business direction

KOREA PROPTech FORUM MEMBERS' MAP

Total 137 | Last updated on Dec, 2019



Copyright 2019, KOREA PROPTech FORUM. All rights reserved.

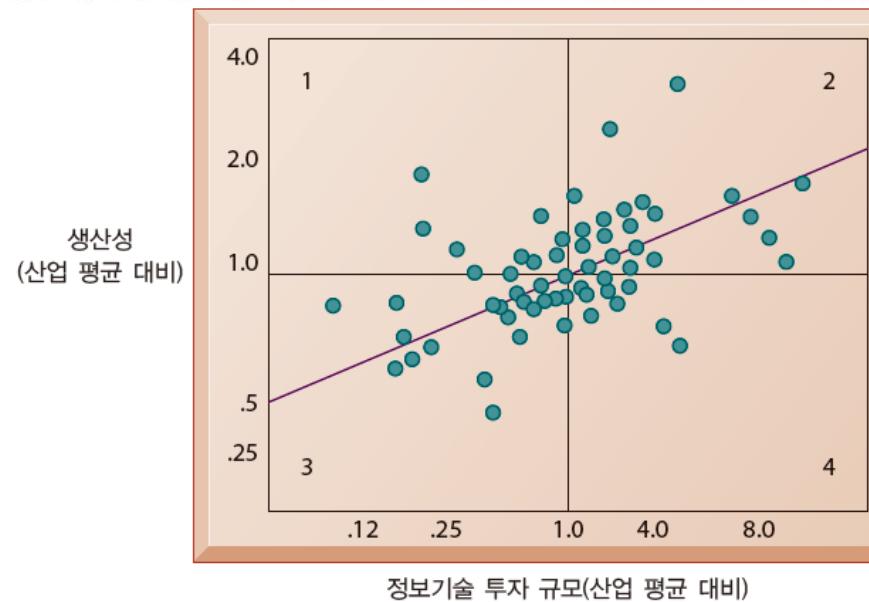
Global “Business” Trends

➤ Future business direction

- An investment becomes an expense if it is not accompanied by an organization's values, structure and behavior, and other complementary changes.
- New business models, processes, management practices, and organizational culture maximize performance. (Brynjolfsson, 2005; Brynjolfsson and Hitt, 2000; Laudon, 1974)

그림 1.8 정보기술 투자 성과에서의 편차

평균적으로 정보기술 투자가 다른 투자보다 높은 성과를 나타내지만 기업에 따라 많은 편차를 보인다.



➤ Current and future business

Is Nike the Future?

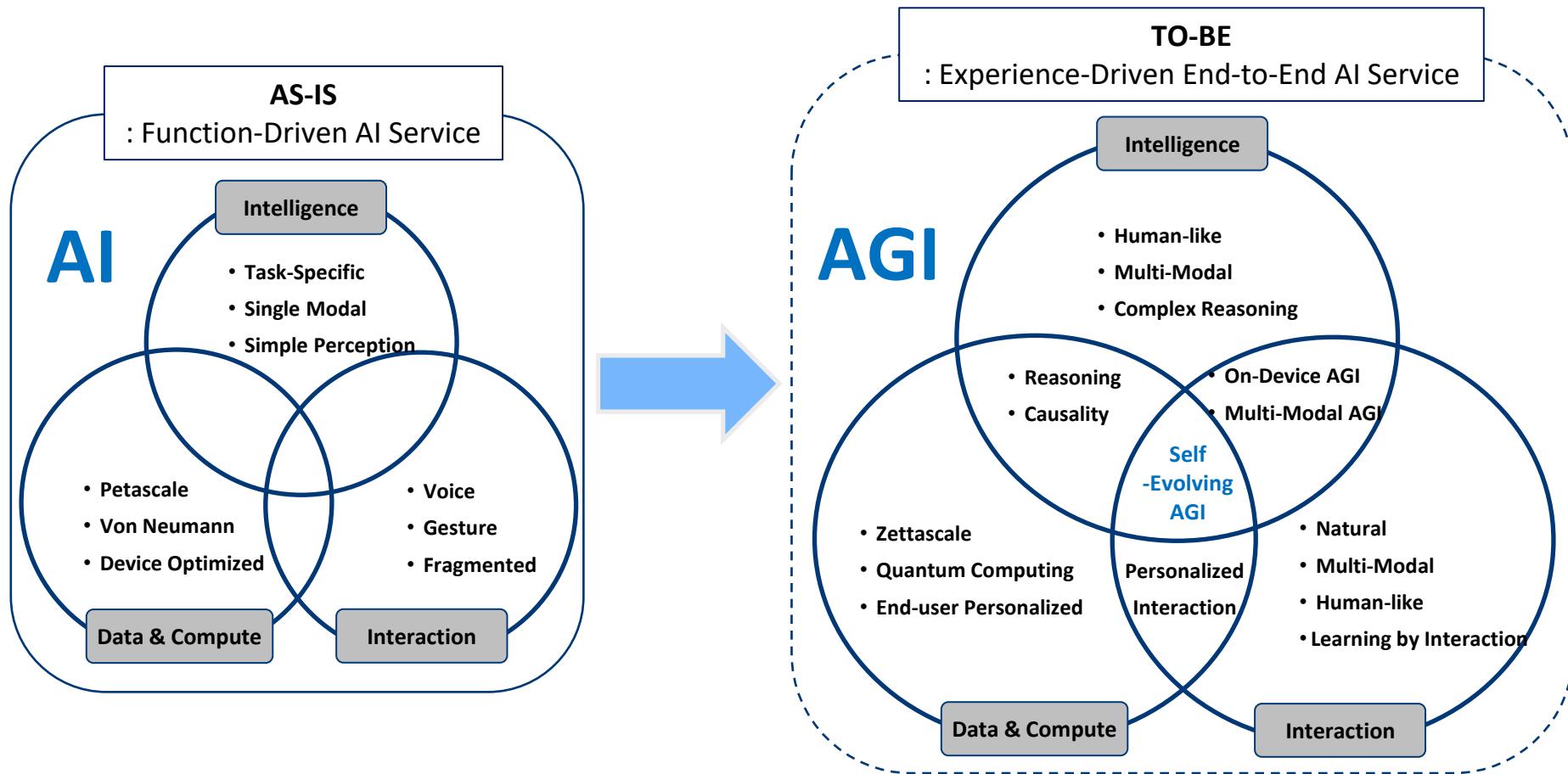


Just Do It.

Global “Technology” Trends

➤ Technology Vision: Beyond AI, Towards Artificial General Intelligence (AGI)

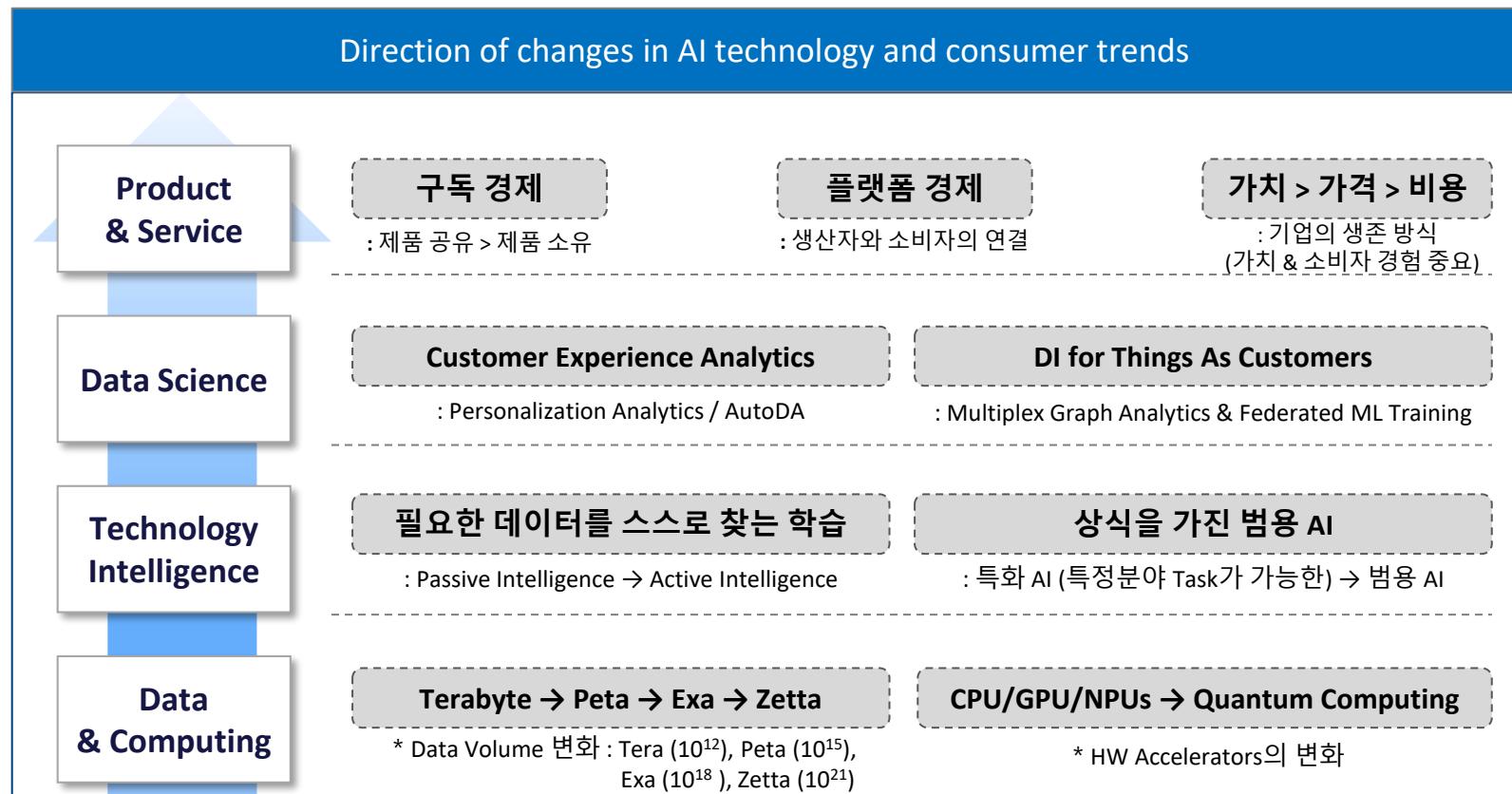
- Enabling new AI technology and **new AI experience** for Data Science Analytics.
- The need for **human-level AI research** in the AL industry including DeepMind, MS and Bengio.



Global “Technology” Trends

➤ Technology Vision: Beyond AI, Towards Artificial General Intelligence (AGI)

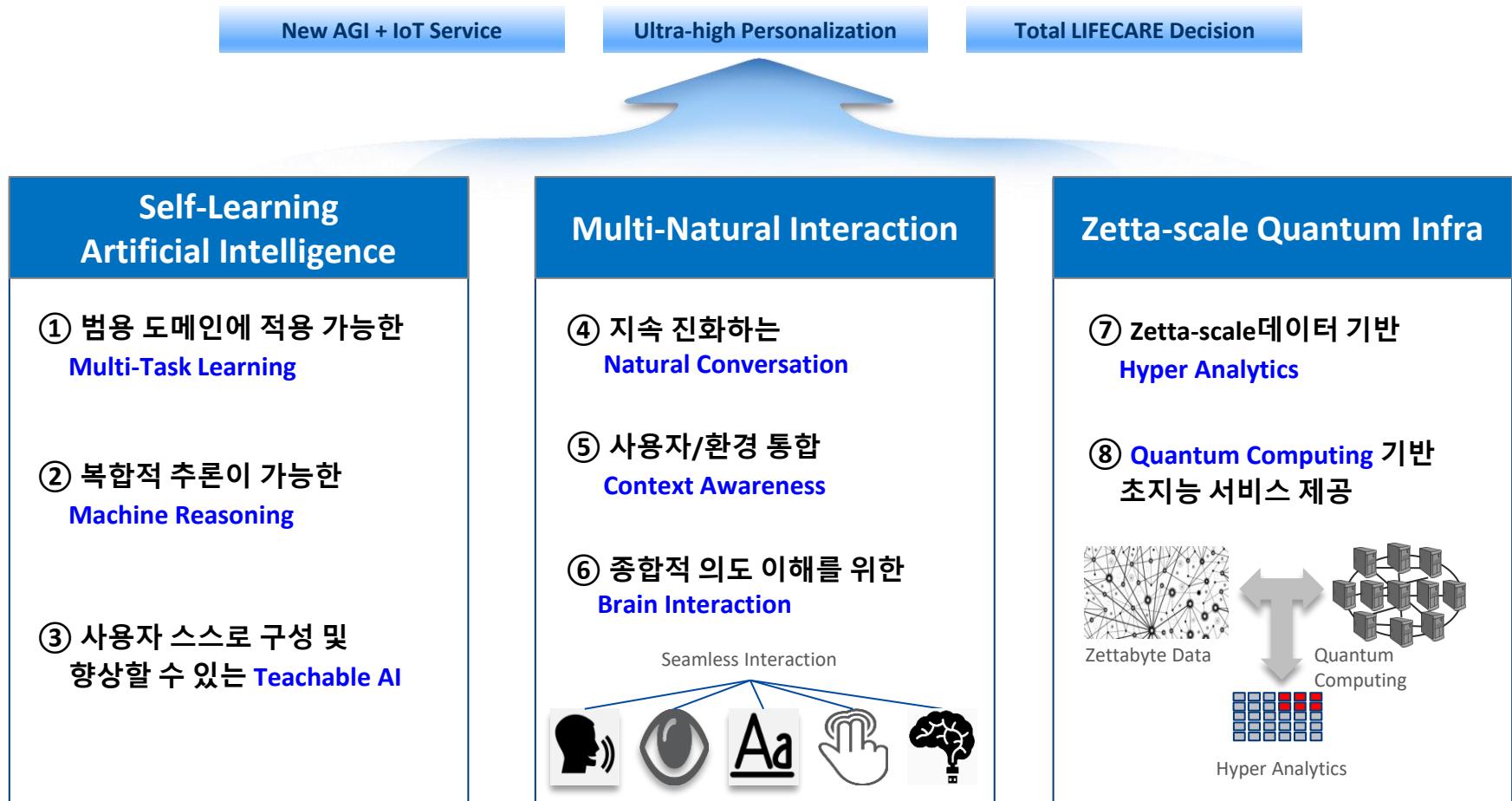
- **AGI:** The hypothetical intelligence of a machine that has the capacity to understand or learn any intellectual task that a **human being can**. (Wikipedia)
 - Current AI solves one specific problem, and **AGI solves multiple problems** in a general way.
 - Current AI responds to problems by pre-learning, and **AGI responds to new, untrained problems**.



Global “Technology” Trends

➤ Technology Vision: Beyond AI, Towards Artificial General Intelligence (AGI)

"AGI Driven Ecosystem"



Global “Technology” Trends

- Technology Vision: Beyond AI, Towards Artificial General Intelligence (AGI)



Global “Technology” Trends

- Technology Vision: Beyond AI, Towards Artificial General Intelligence (AGI)

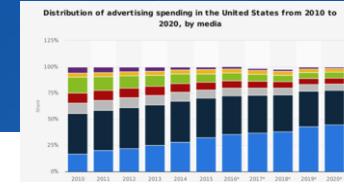


Global “Technology” Trends

- Technology Vision: Beyond AI, Towards Artificial General Intelligence (AGI)



Global “Academia” Trends



➤ Top 연구: Robot > VPA(Virtual Personal Assistant) > 헬스 > 커머스 및 광고 > 운영효율

- **Robot 분야:** 적용 시나리오 및 다른 인공지능 기술과의 접목 가능성이 확대
- **VPA 분야:** 다양한 관련 제품의 시장 진입 후 언어, 영상, 추론 등을 통한 고도화 연구가 추진

Business Analytics

▪ 학계 연구는 Robot, VPA, 헬스, 커머스/광고 순으로 활발한 연구가 진행 중

- Robot, 커머스/광고는 특정 국가 한정이 아닌 글로벌 연구로 확산되는 중
- VPA, 헬스 분야는 아직은 미국 학계를 중심으로 연구가 집중되어있는 추세(데이터 수급 문제)



Technology Research

▪ Robot 분야는 적용처가 다양한 산업군 내 확대되는 추세에 따라 연구 필요성 증가

- 로봇이 수술, Assistant, Social, 자율주행 등으로 활용 분야 및 개념이 점차 확대
- 로봇이 VPA(10개), 자율주행(6개) 등의 분야와 결합한 연구, 또는 일반적인 학습 및 제어를 위한 연구가 꾸준히 진행 중

▪ VPA 활성화로 Interaction 고도화 필요성 확대

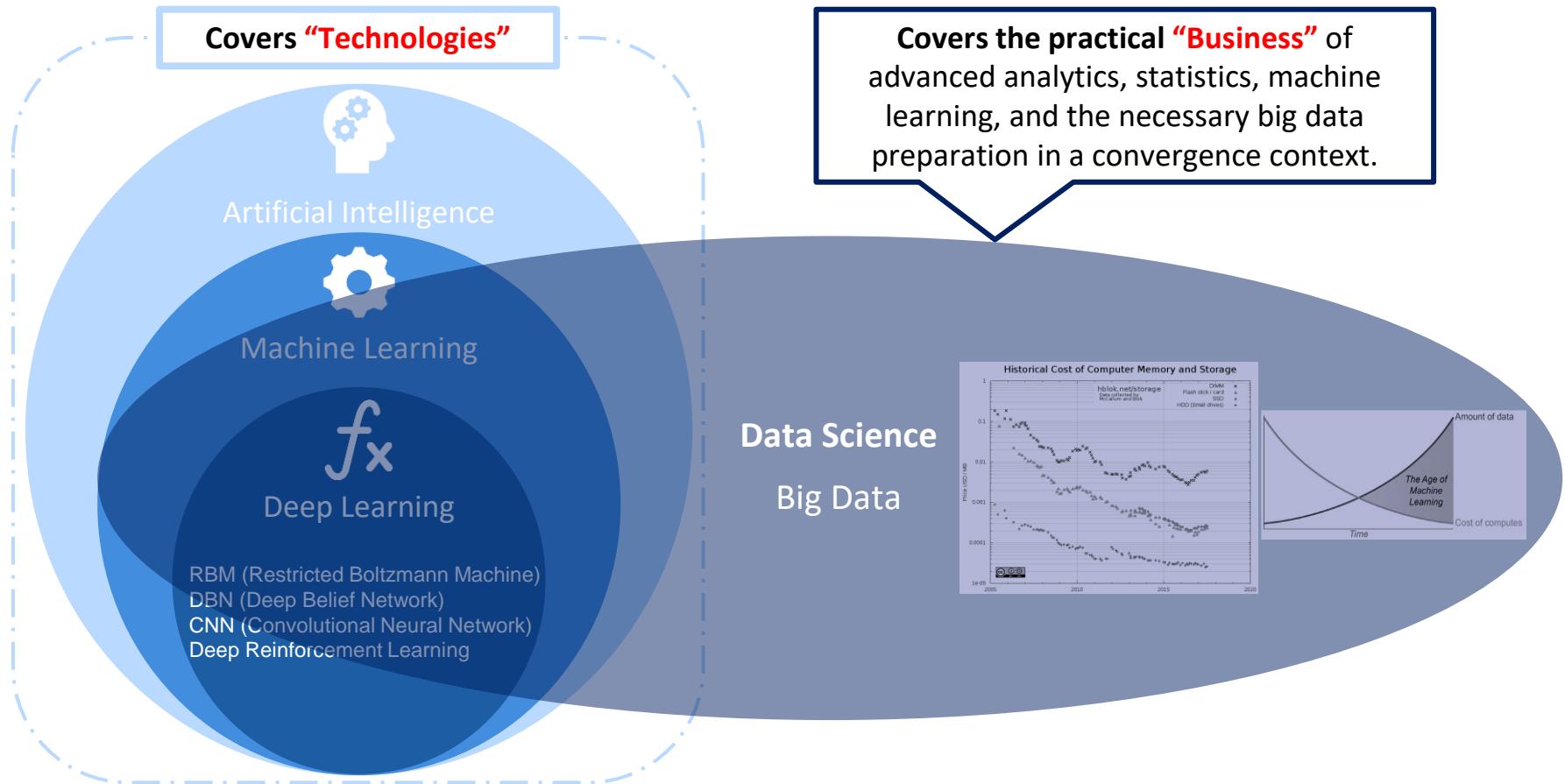
- 언어/영상 처리의 자연스러운 interaction 필요
- Data 처리 및 추론을 통한 VPA 서비스의 질 향상

▪ 헬스 분야는 의료영상 처리, 생체정보 기반 맞춤형 관리, 유전자 분석 연구가 주로 추진

* Data Science and AI Top 500 中 상위 20개 대학 분석

“Artificial Intelligence” vs. “Data Science”

- **Artificial Intelligence:** algorithm or technology which enables computers to behave like a human.
- **Data science:** an convergence application or business that has skills used in various fields (AI, ML, DL).
 - Enables to find meaning and appropriate information from large volumes of data
 - Use data for making key decisions in business, science, technology, and even politics



“Artificial Intelligence” vs. “Data Science”

➤ Global Trend

▪ 용어정리:

- AI(인공지능)는 컴퓨터가 사람처럼 사고하고 행동할 수 있게 해주는 "기술" 도구
- Data Science(데이터과학)은 “빅데이터 재료”에 “AI 기술”을 사용하여 인간과 사회의 문제들을 해결하는 "융합" 도구

▪ 비즈니스와 기술 방향:

- AI기술은 더욱 인간에 가까운 AGI를 향해가고 있으며,
비즈니스는 전통적인 하드웨어와 제품을 넘어 디지털 사회 중심 서비스로 수익구조 다변화,
- 수익구조 내 상위 Top 5는 로봇, 비서, 헬스, 마케팅 및 광고, 운영효율화로,
특정 국가를 막론하고 기업과 학계에서 연구역량을 집중

WE ALL LIVE IN A DIGITAL ECONOMY!

I. 글로벌 트랜드, 데이터과학, 인공지능

- Global “Business” Trends
- Global “Technology” Trends
- Global “Academia” Trends
- “Artificial Intelligence” vs. “Data Science”

II. 데이터과학과 디지털경제

- What is the Communication between Data and the World?
- Why Data Science is in a sudden boom?
- What is a Big Data?
- What is a Big Algorithm?
- Digital Platform with Big Data/Algorithm

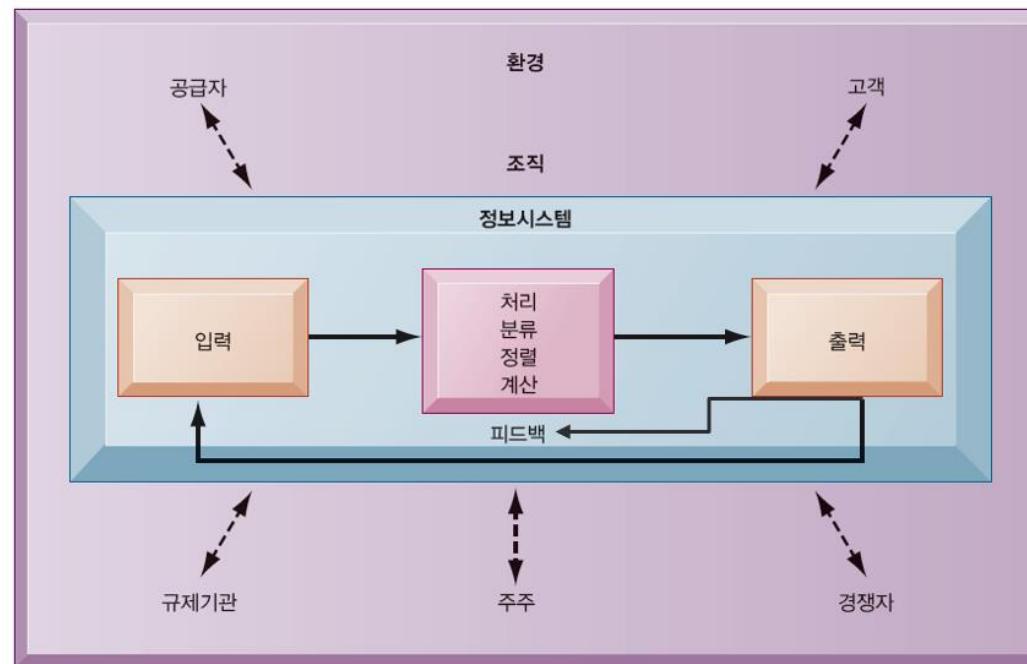
III. 데이터와 인공지능을 활용한 디지털전환

- Evolution of Hardware for Digital Platforms
- Evolution of Computing Platforms
- Skills and Jobs in the Data Science
- Driving Digital Transformation with Data Science as a Service

What is the Communication between Data and the World?

➤ Business Process: Input(X) / Processing(f) / Output(Y) / Evaluation(e)

- All logic in the world is optimized and expressed as **input**, **processing**, **output**, and **verification**.
 - **Input:** collect raw data from the inside or outside the organization
 - **Processing / Function:** changing raw data into meaningful form
 - **Output:** delivering the processed information to the person who will use it or the activity to use it
 - **Evaluation:** Performance indicators by the direction of the company

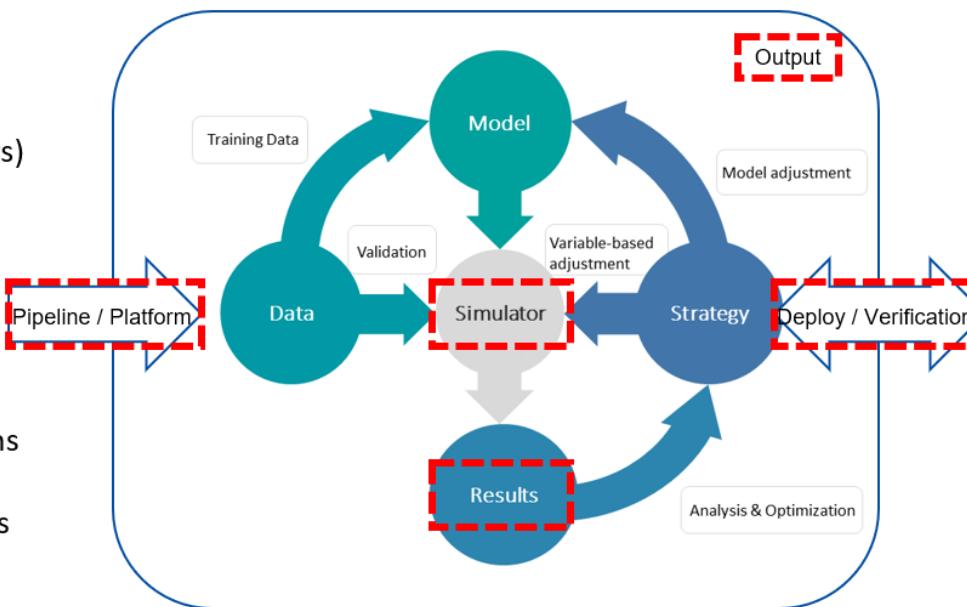


What is the Communication between Data and the World?

➤ Organization Process: Input(X) / Processing(f) / Output(Y) / Evaluation(e)

- All logic in the world is optimized and expressed as **input**, **processing**, **output**, and **verification**.
 - **Input:** collect raw data from the inside or outside the organization
 - **Processing / Function:** changing raw data into meaningful form
 - **Output:** delivering the processed information to the person who will use it or the activity to use it
 - **Evaluation:** Performance indicators by the direction of the company

- ✓ Market conditions (including competitors)
- ✓ Economic indicators (nonfarm payroll, unemployment, CPI, exchange..)
- ✓ Sales and campaigns
- ✓ Retailer and supply
- ✓ Marketing/Promotions
- ✓ Enterprise
- ✓ Customer preferences and demographics

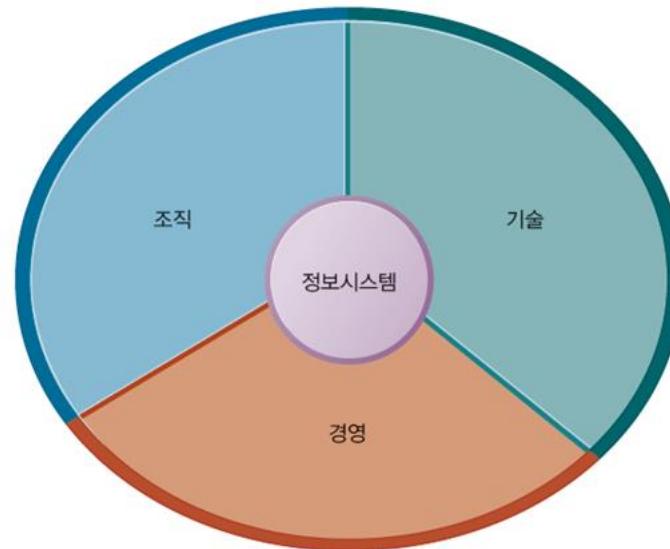


- Financial Management
- Strategy and Product Planning
- Enterprise
- HR and Operation
- Sales and Marketing
- Legal and Regulatory
- Supply Chain
- Quality Management
- R&D and IT
- Manufacturing

What is the Communication between Data and the World?

➤ Organization Process: Input(X) / Processing(f) / Output(Y) / Evaluation(e)

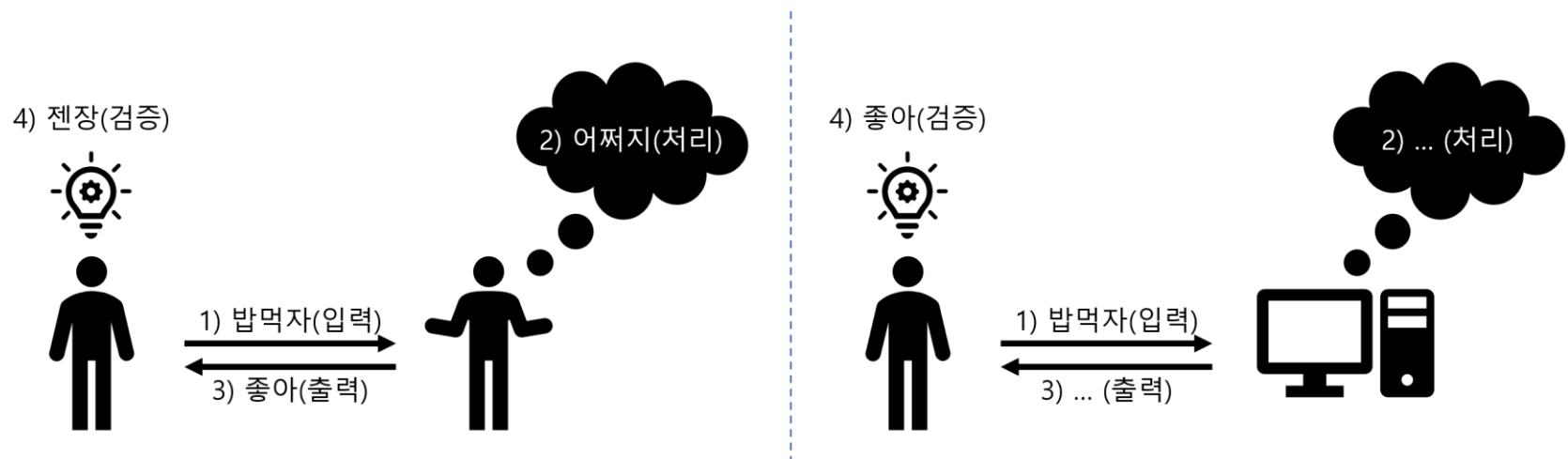
- All logic in the world is optimized and expressed as **input, processing, output, and verification**.
 - **Input:** collect raw data from the inside or outside the organization
 - **Processing / Function:** changing raw data into meaningful form
 - **Output:** delivering the processed information to the person who will use it or the activity to use it
 - **Evaluation:** Performance indicators by the direction of the company



What is the Communication between Data and the World?

➤ Interaction Process: Human vs Computer

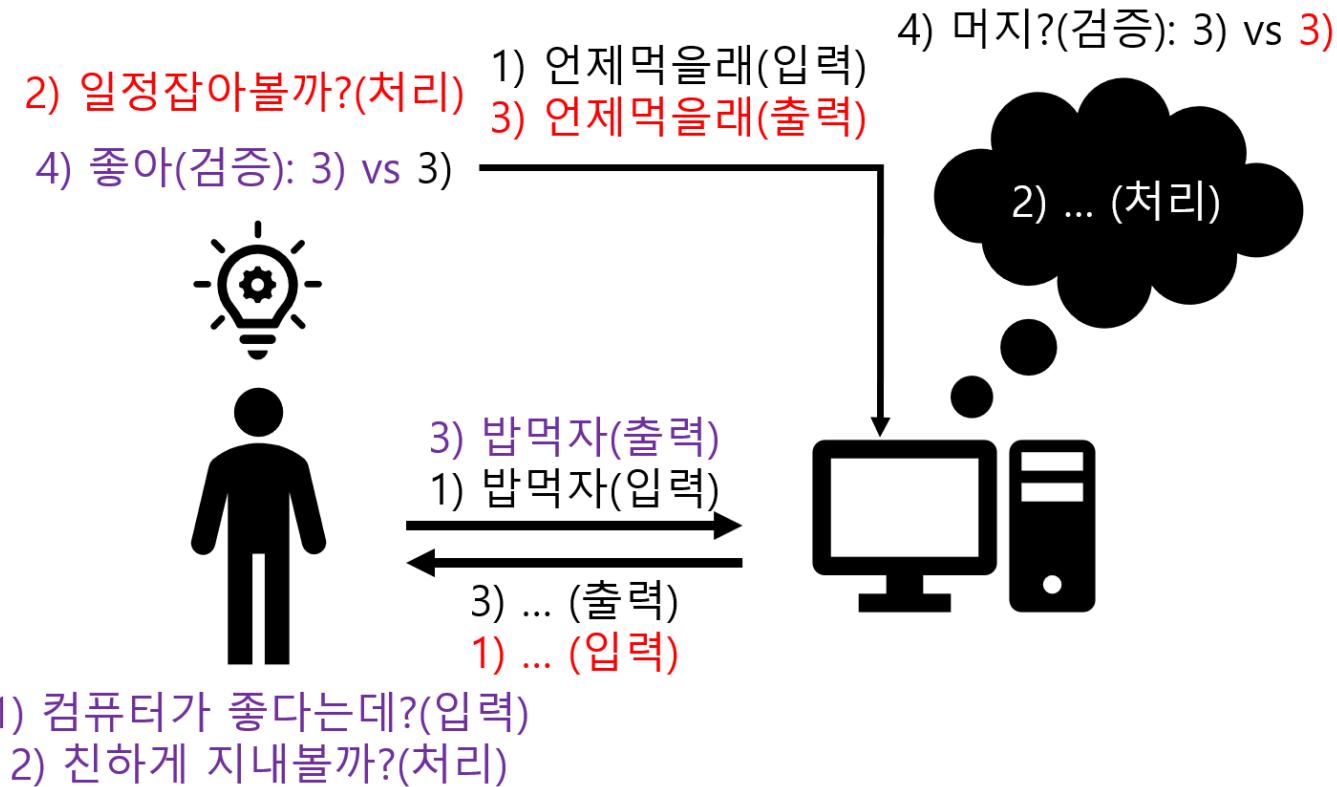
- All logic in the world is optimized and expressed as **input, processing, output, and verification**.
 - **Input:** collect raw data from the inside or outside the organization
 - **Processing / Function:** changing raw data into meaningful form
 - **Output:** delivering the processed information to the person who will use it or the activity to use it
 - **Evaluation:** Performance indicators by the direction of the company



What is the Communication between Data and the World?

➤ Interaction Process: Human vs Computer

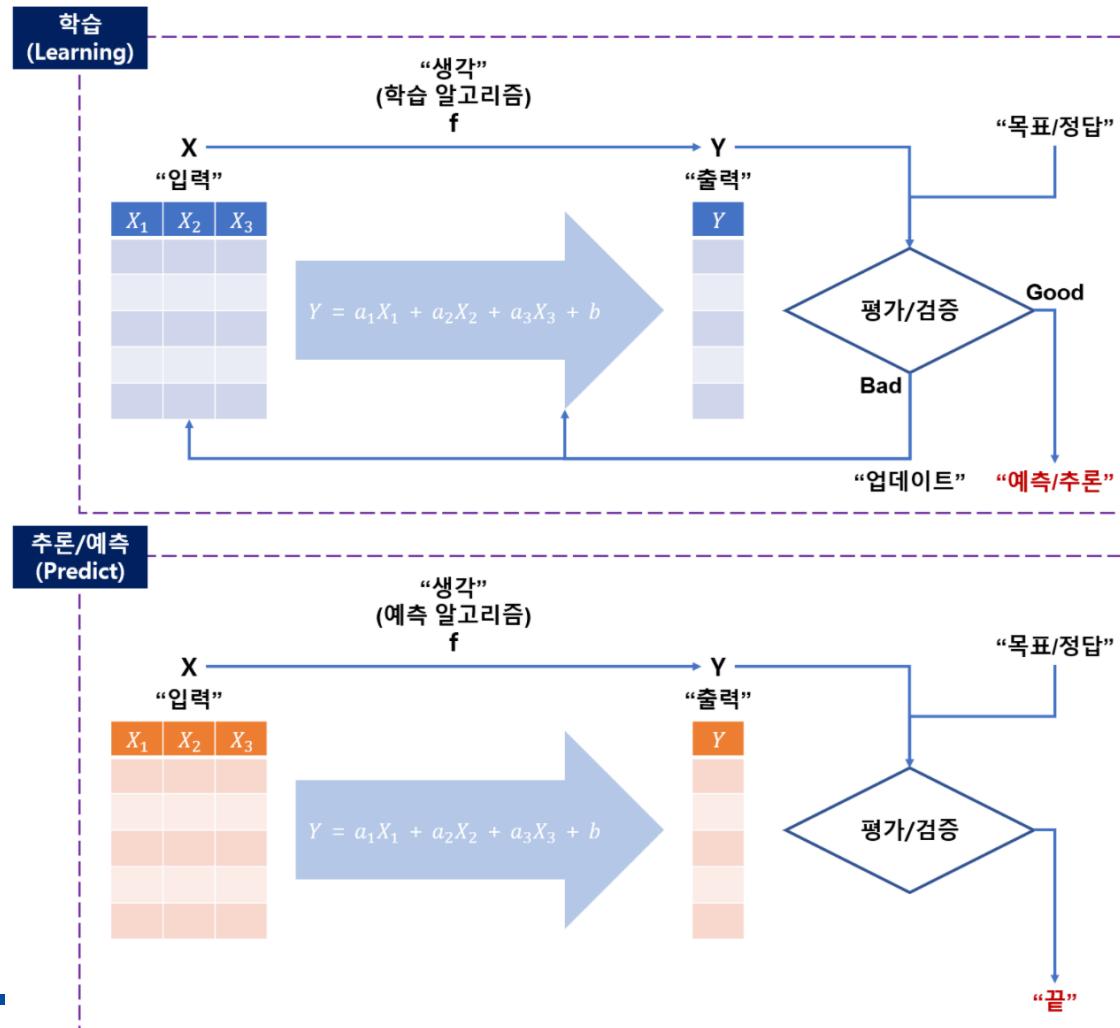
- All logic in the world is optimized and expressed as **input, processing, output, and verification**.



What is the Communication between Data and the World?

➤ Data Science Process

- All logic in the world is optimized and expressed as **input, processing, output, and verification**.



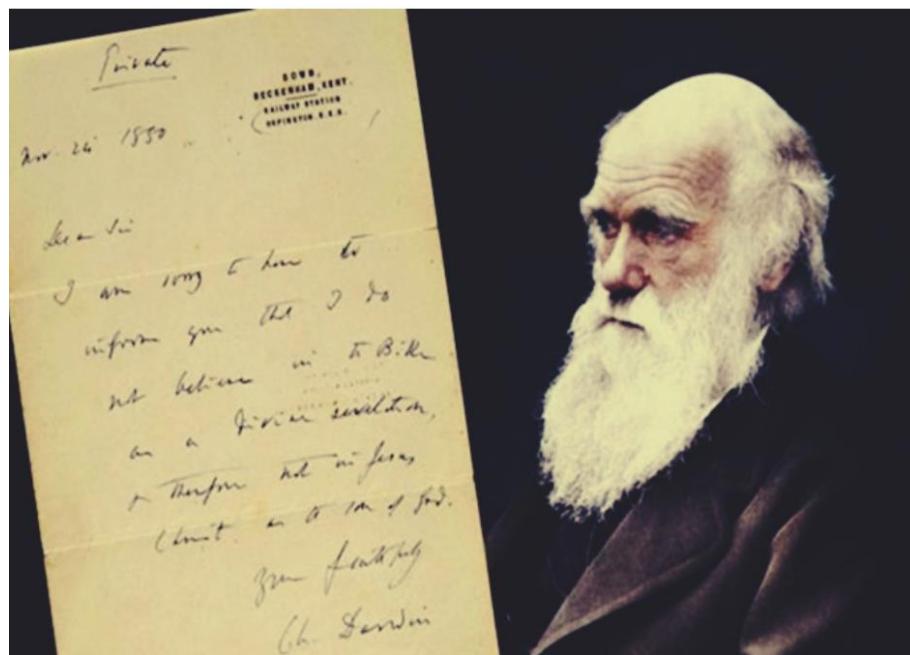
Why Data Science is in a sudden boom?

➤ Data Science is leading these changes

- Regardless of our will, change has already begun, innovation in the technology industry is at the center of change.

가장 강한 종이 살아는 것이 아니다.
가장 두뇌가 뛰어난 종이 살아남는 것도 아니다.
단지 변화에 잘 적응하는 종이 살아남는다.

- 찰스 다윈 -

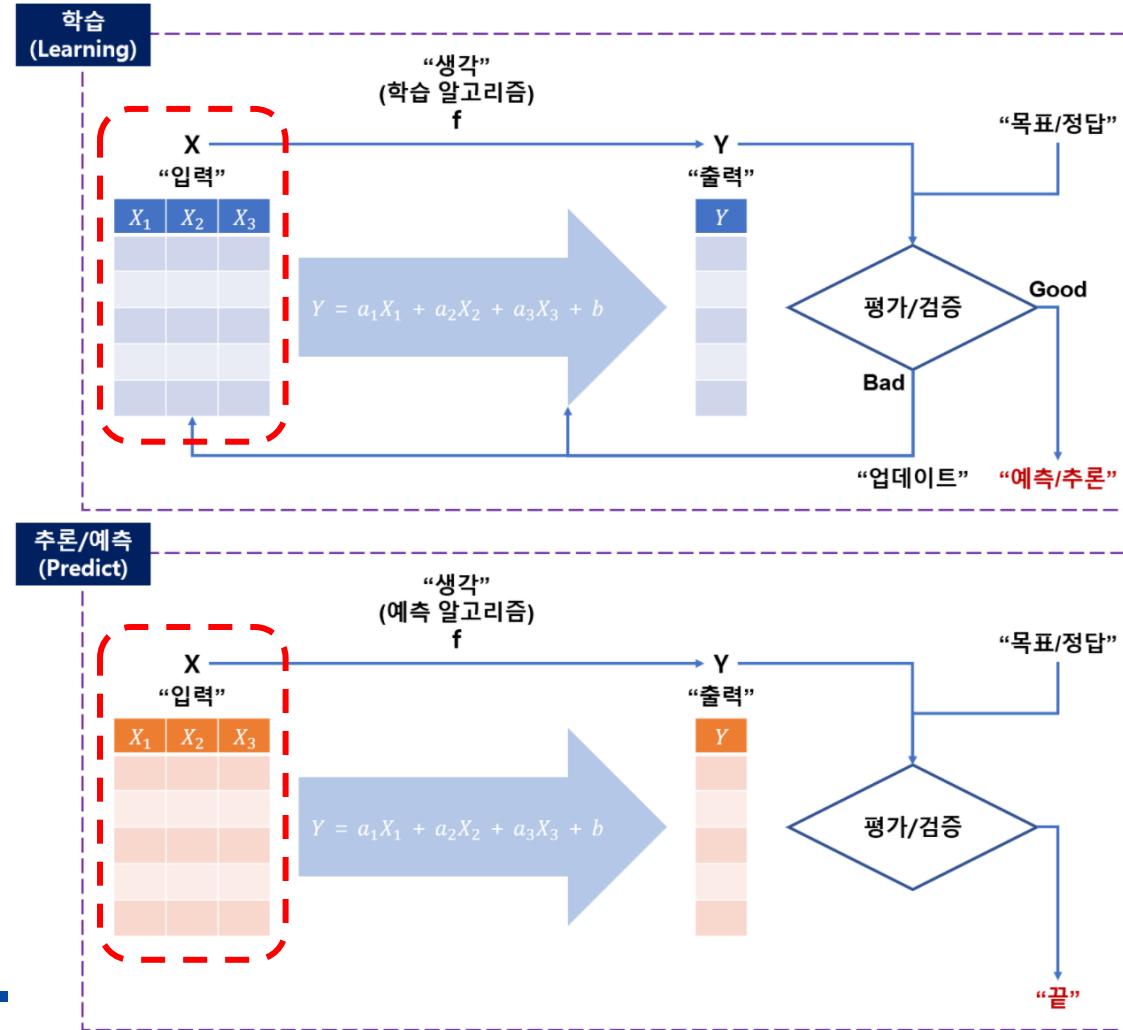


Rank	Brand	Brand Value	1-Yr Value Change	Brand Revenue	Company Advertising	Industry
#1	Apple	\$205.5 B	12%	\$265.8 B	-	Technology
#2	Google	\$167.7 B	27%	\$136.2 B	\$6.4 B	Technology
#3	Microsoft	\$125.3 B	20%	\$110.2 B	\$1.6 B	Technology
#4	Amazon	\$97 B	37%	\$211.4 B	\$8.2 B	Technology
#5	Facebook	\$88.9 B	-6%	\$48.8 B	\$1.1 B	Technology
#6	Coca-Cola	\$59.2 B	3%	\$23.8 B	\$4.1 B	Beverages
#7	Samsung	\$53.1 B	11%	\$221.6 B	\$3.6 B	Technology
#8	The Walt Disney Company	\$52.2 B	10%	\$33.8 B	\$2.8 B	Leisure
#9	Toyota	\$44.6 B	0%	\$190.8 B	\$4.6 B	Automotive
#10	McDonald's	\$43.8 B	6%	\$96.1 B	\$389 M	Restaurants

Why Data Science is in a sudden boom?

➤ Big Data and Big(AI?) Algorithm

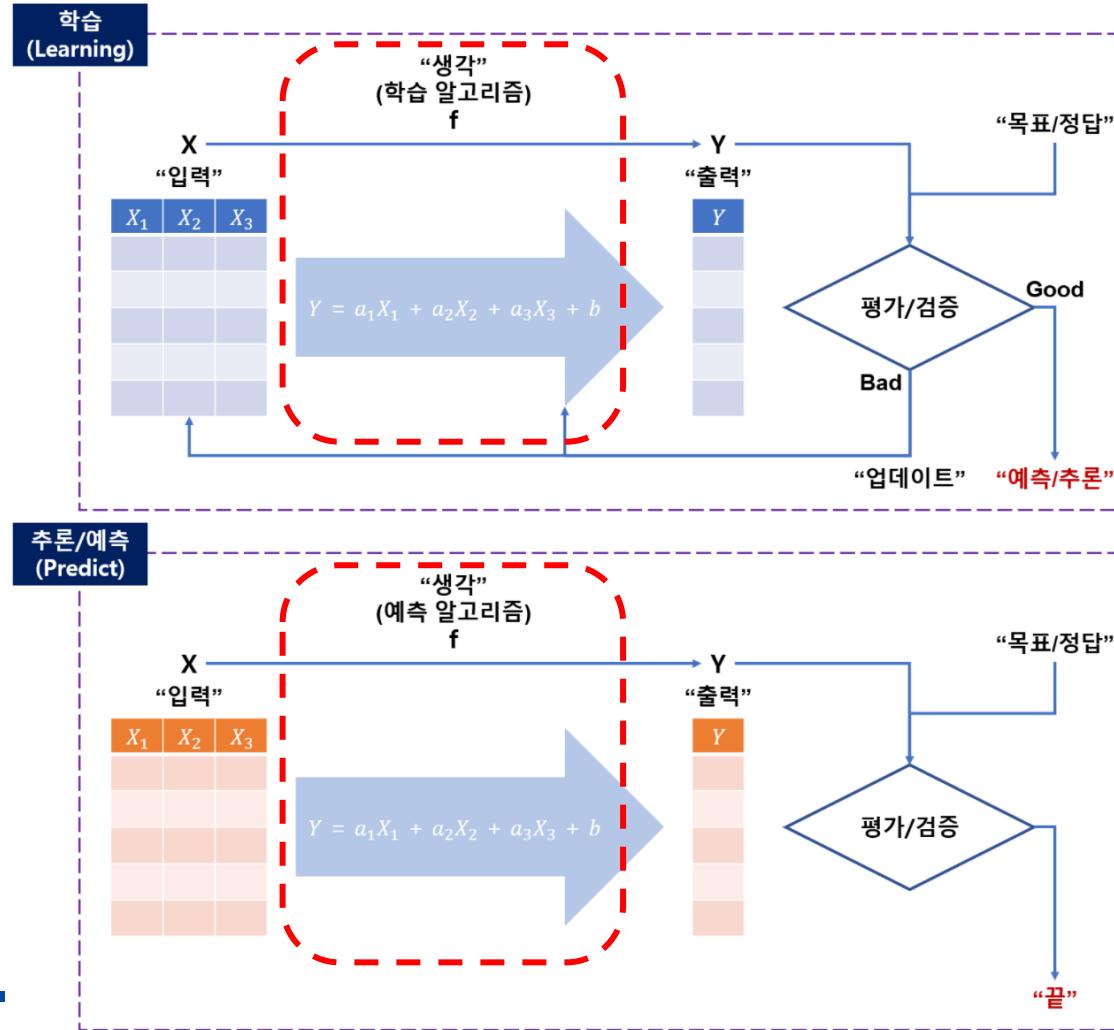
- The evolution of data and algorithms has evolved data science.



Why Data Science is in a sudden boom?

➤ Big Data and Big(AI?) Algorithm

- The evolution of data and algorithms has evolved data science.

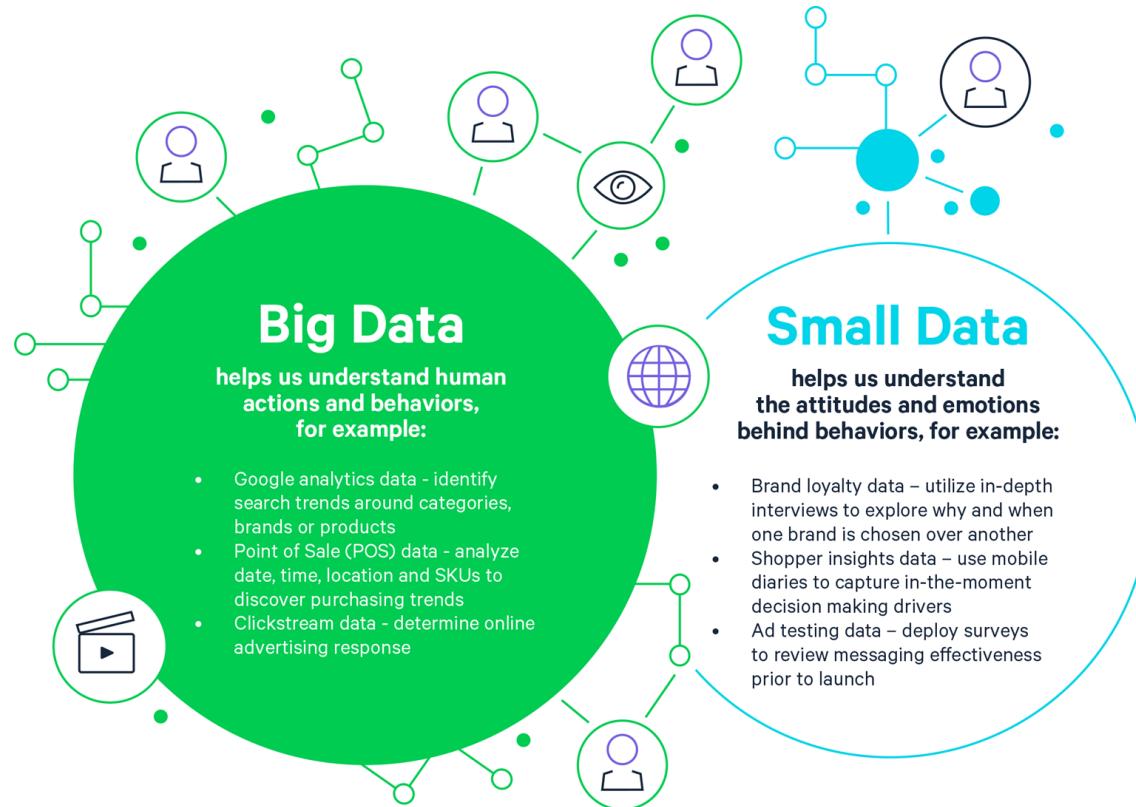


What is a Big Data?

➤ Big Data vs. Small Data

- **Small Data:** Getting machines to do what humans are good at.
- **Big Data:** Feeding an algorithm data to learn and predict something.

=> A Powerful Equation: Real Human Insight = Big Data + Small Data



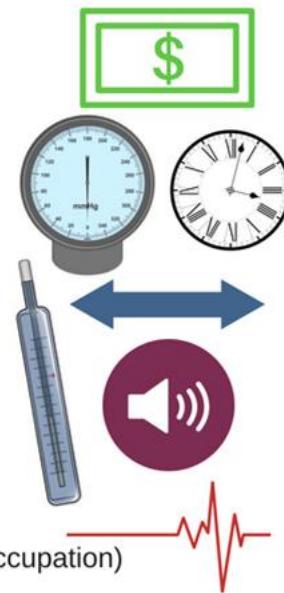
What is a Big Data?

➤ Quantitative vs. Qualitative Data

- **Quantitative Data:** Numerical calculations and measurements.
- **Unstructured Data:** Sensations, feelings, and experiences.

Quantitative Data

- money
- time
- speed
- movement
- height
- length
- area
- volume
- weight
- temperature
- humidity
- pressure
- sound level
- categories
(age, gender, occupation)
- positioning
- status



Qualitative Data

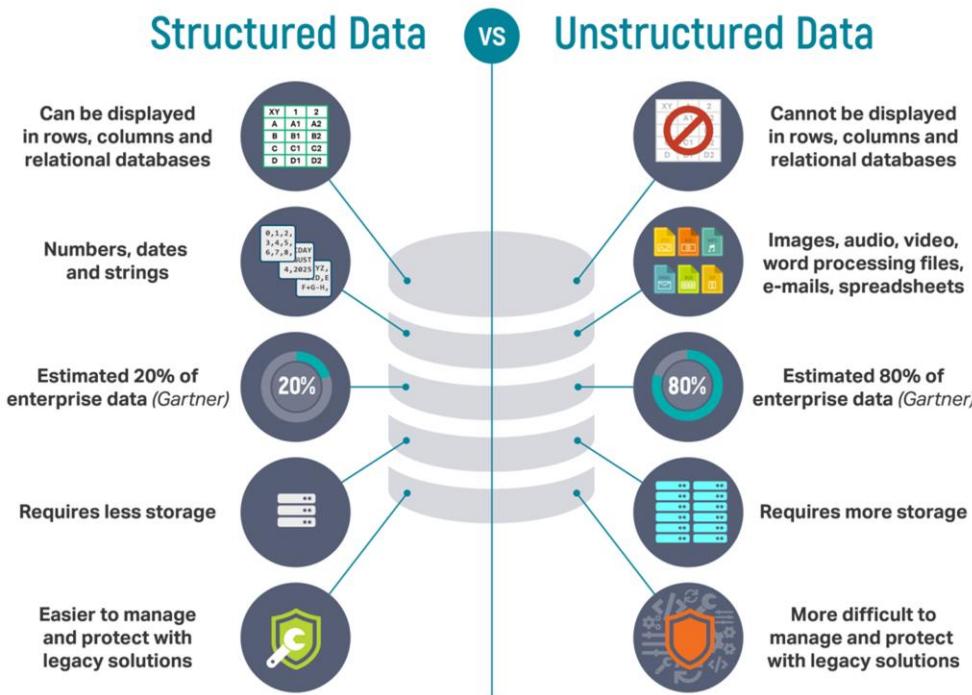
- verbal and written feedback
- narrative story
 - first-hand (direct experience)
 - second-hand (telling someone else)
 - third-hand (outside story-teller)
- visual images, drawings, or models
- experiential sensations
- descriptions of
 - colors
 - textures
 - smells
 - tastes
 - appearance
 - beauty
 - feelings
 - intuition
 - sensations
 - choices
 - values
 - beliefs



What is a Big Data?

➤ Structured vs. Unstructured Data

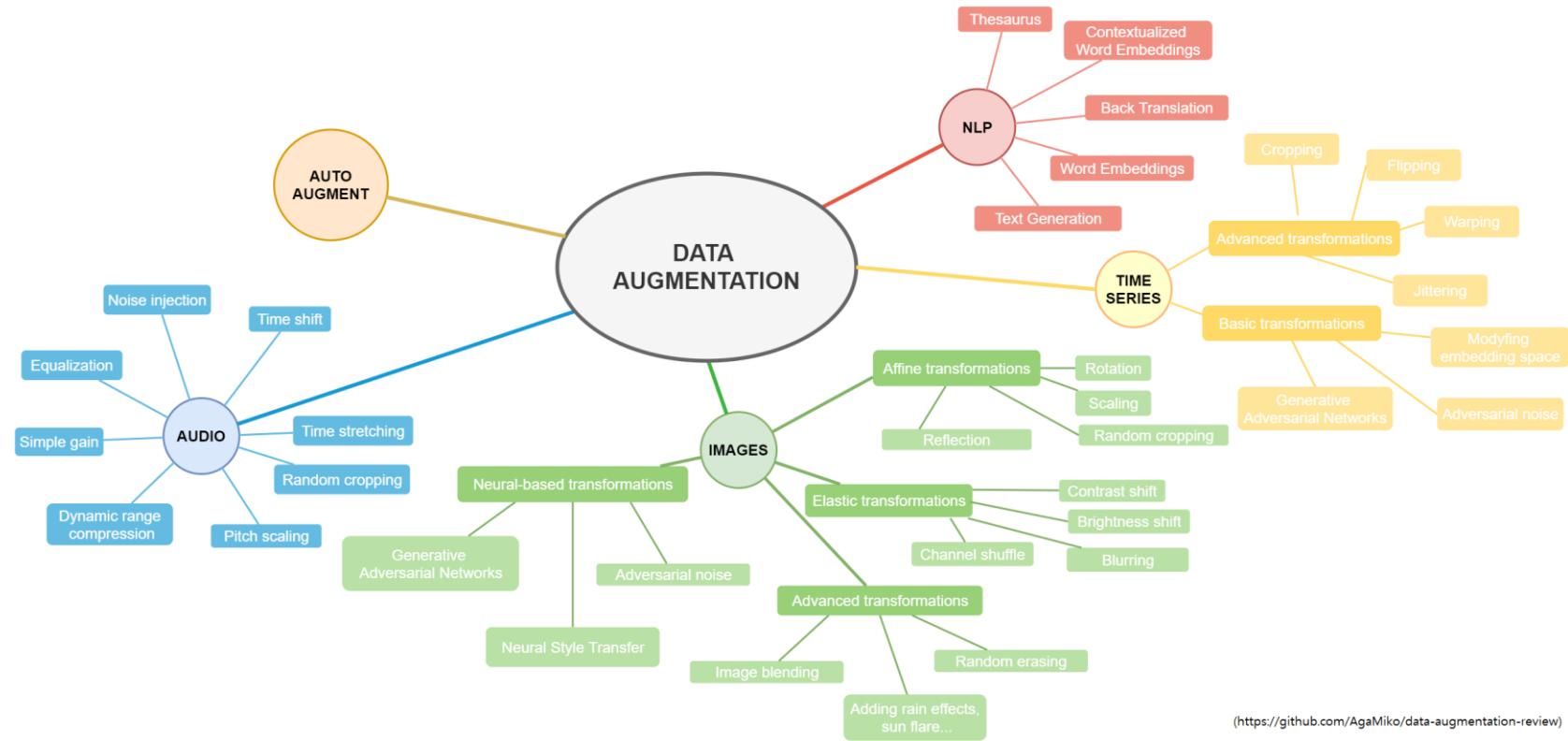
- **Structured Data:** the type of data that fits nicely into a relational database. It's highly organized and easily analyzed. Most IT staff are used to working with structured data.
- **Unstructured Data:** It doesn't fit nicely into a spreadsheet or database. It can be textual or non-textual. It can be human- or machine-generated.



What is a Big Data?

➤ Data Augmentation

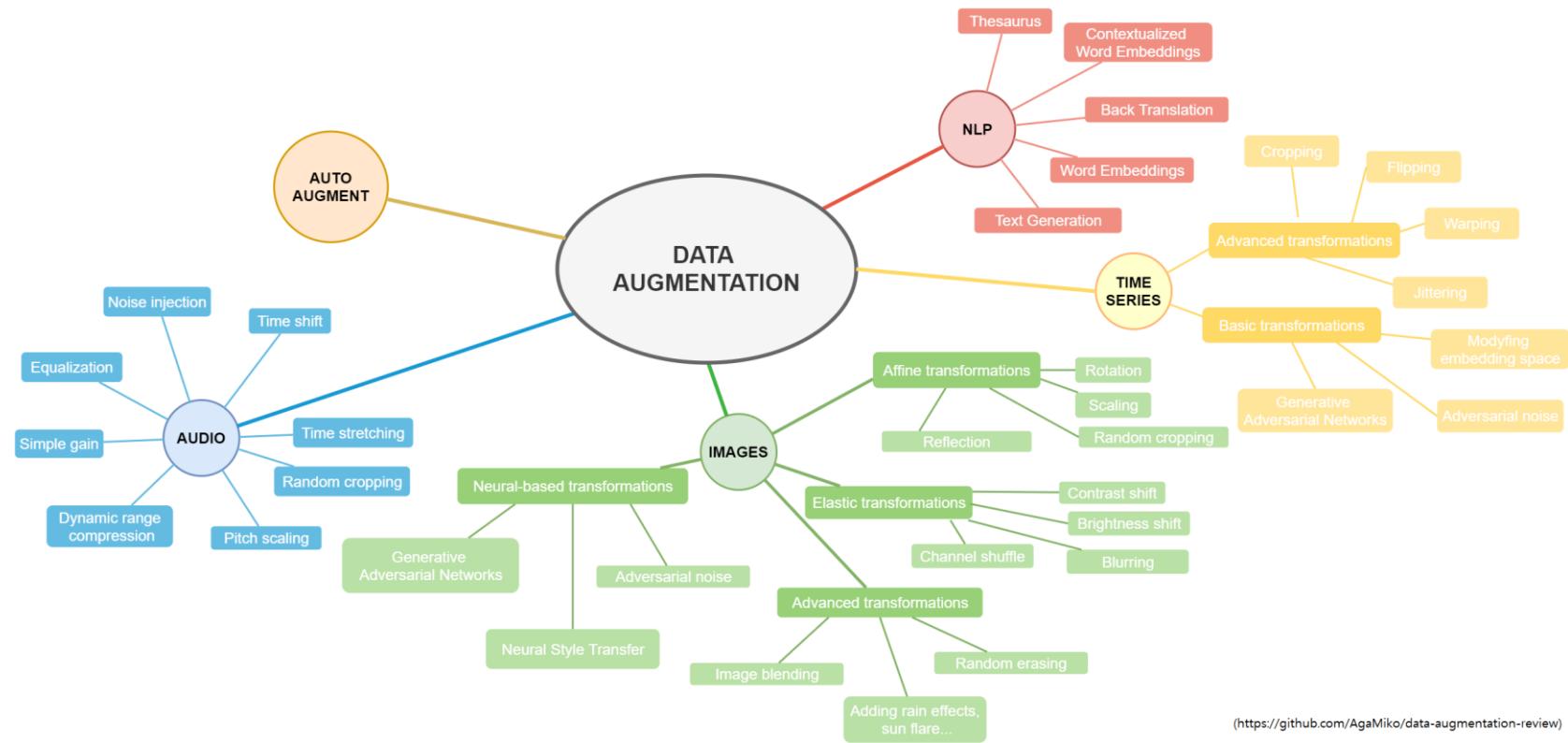
- Techniques used to **increase the amount of data** by adding slightly modified copies of already existing data or newly created synthetic data.
- Simply described as **any method** that makes our dataset larger.



What is a Big Data?

➤ Data Augmentation → Input!

- Techniques used to **increase the amount of data** by adding slightly modified copies of already existing data or newly created synthetic data.
- Simply described as **any method** that makes our dataset larger.



What is a Big Algorithm?

➤ Why use Machine Learning instead of Traditional Statistics?

Traditional Statistics



White-box modelling
simpler computation, emphasis on introspection, form, causal effects and processes, finding a 'correct' model

Machine Learning

Black-box modelling
high computational complexity, emphasis on speed and quality of prediction, finding a 'performant' model

▪ Bayes Theorem:

Thomas Bayes mid 1700's

▪ Regression:

Legendre, Gauss and Galton early 1800's

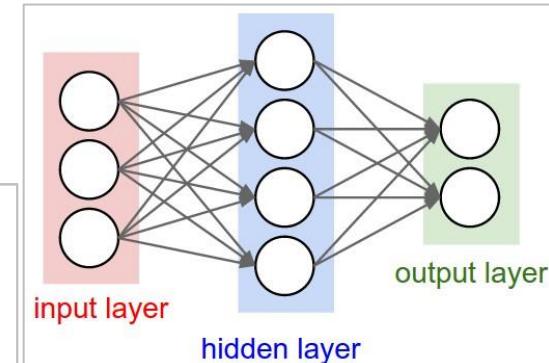
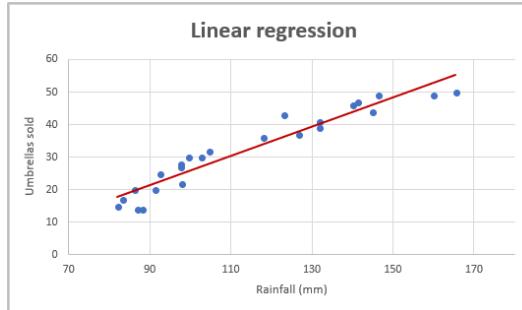
▪ Neural Networks:

McCulloch and Pitts early 1940s

Diagram illustrating Bayes Theorem:

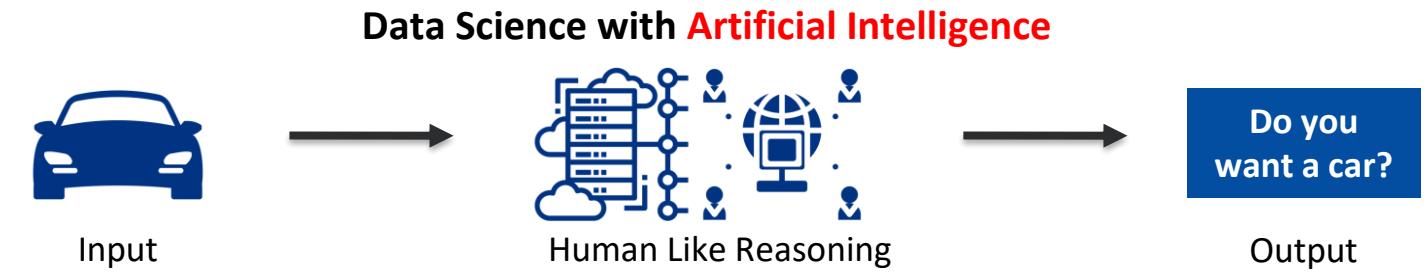
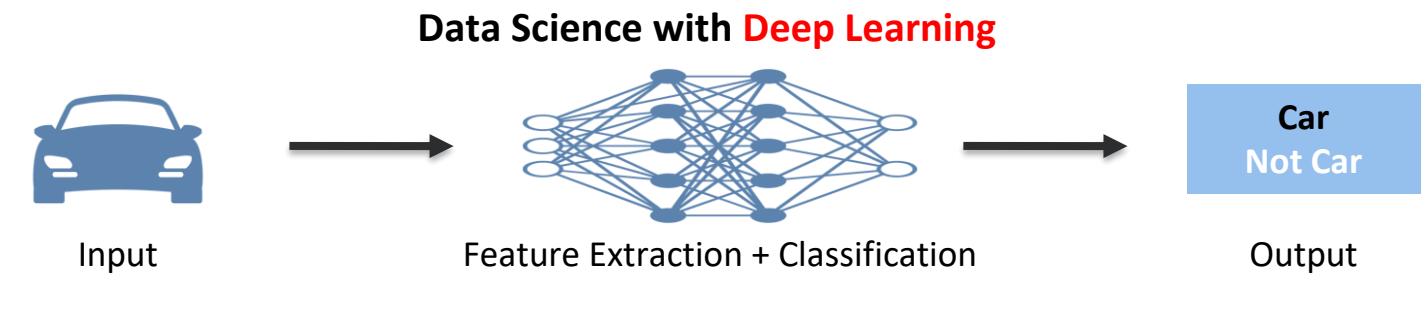
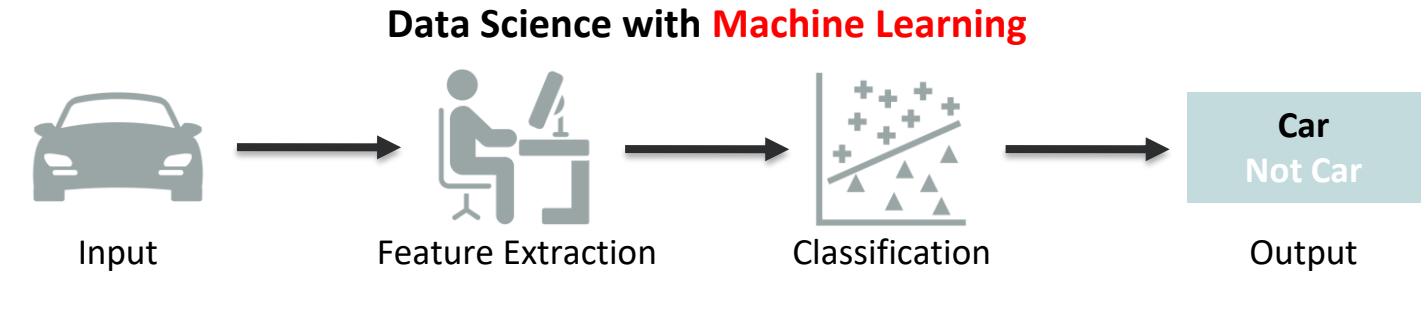
$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

Labels: LIKELIHOOD (The probability of "B" being True, given "A" is True), PRIOR (The probability "A" being True. This is the knowledge.), POSTERIOR (The probability of "A" being True, given "B" is True), MARGINALIZATION (The probability "B" being True).



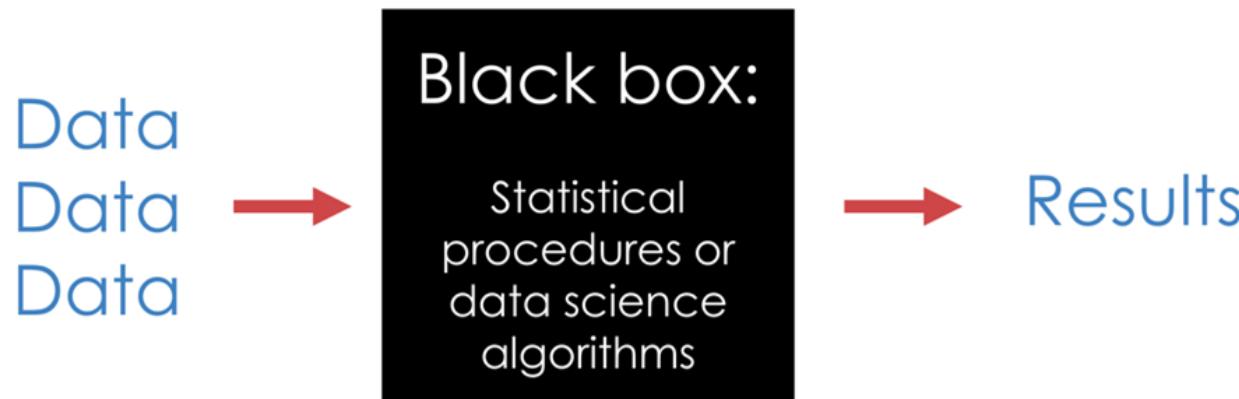
What is a Big Algorithm?

➤ Why use Machine Learning instead of Traditional Statistics?



What is a Big Algorithm?

➤ Why use Machine Learning instead of Traditional Statistics?



- **AI:** Getting machines to do what humans are good at
- **Machine Learning:** Feeding an algorithm data to learn and predict something
- **Deep Learning:** A type of machine learning

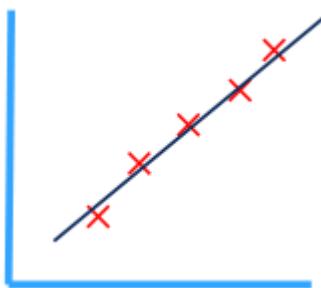


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

What is a Big Algorithm?

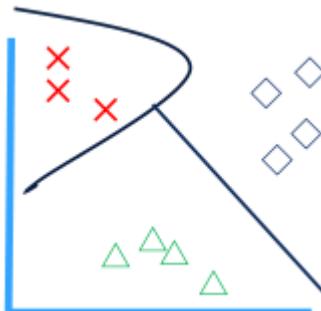
➤ Solution Directions to the Black Box Problem

- How much is the stock of Samsung Electronics tomorrow?



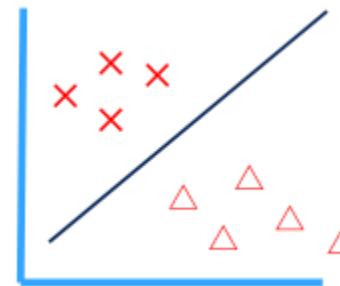
Regression – Looking for a statistical relationship across variables that may give us an estimate of a particular outcome.

- Are Samsung Electronics and Naver similar business companies?



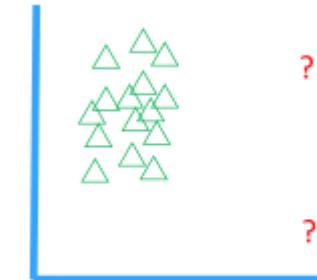
Clustering – Do not have predefined classes but trying to find groups or sets based upon data at hand. (Unsupervised)

- Will Samsung Electronics' stocks rise or fall tomorrow?



Classification – Similar to regression but looking for separations in the data given predefined classes. (Supervised)

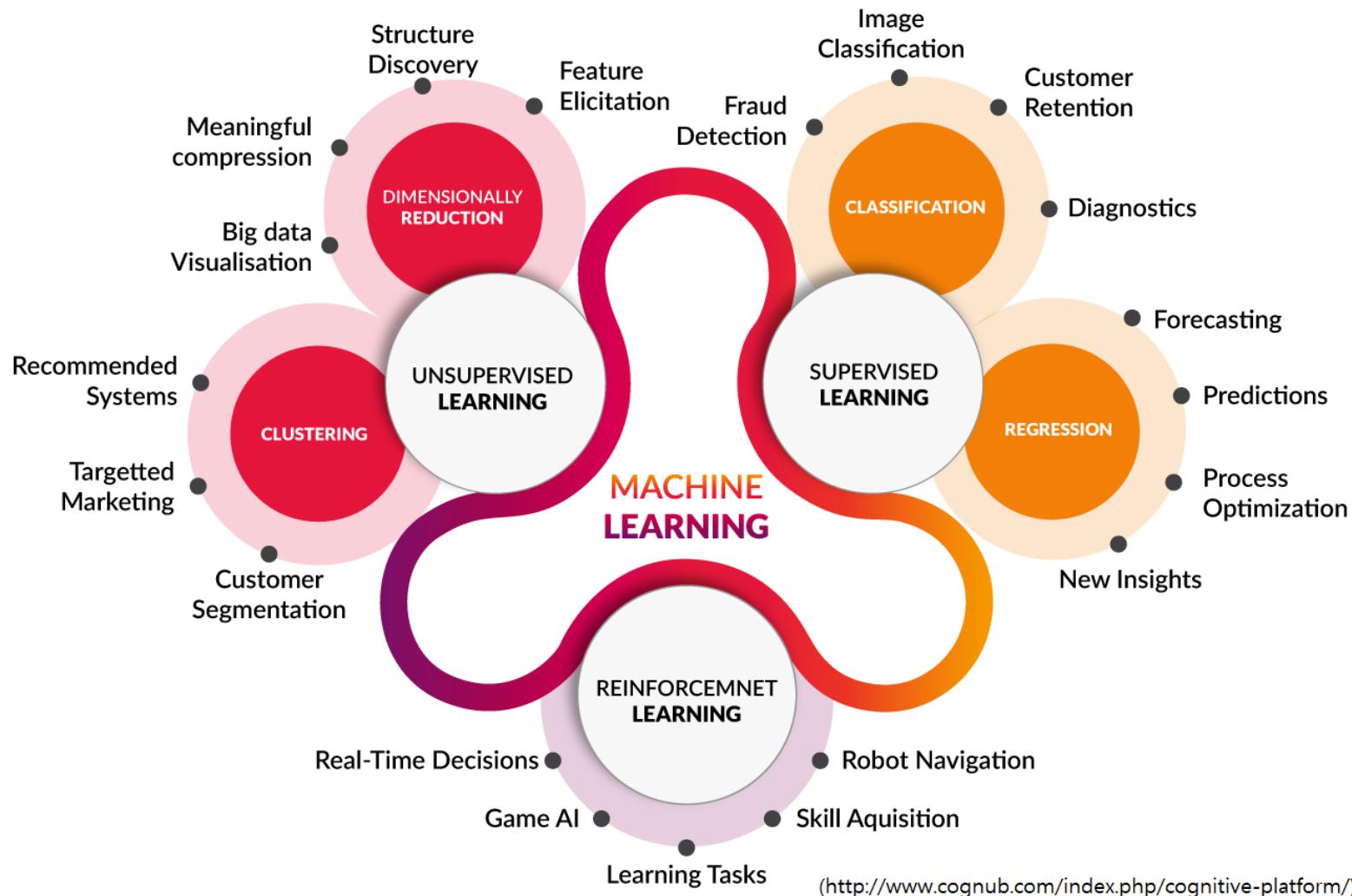
- Is the recent sharp drop in KOSPI outliers or normal?



Anomaly Detection – Identification of outliers based upon expected ranges of data.

What is a Big Algorithm?

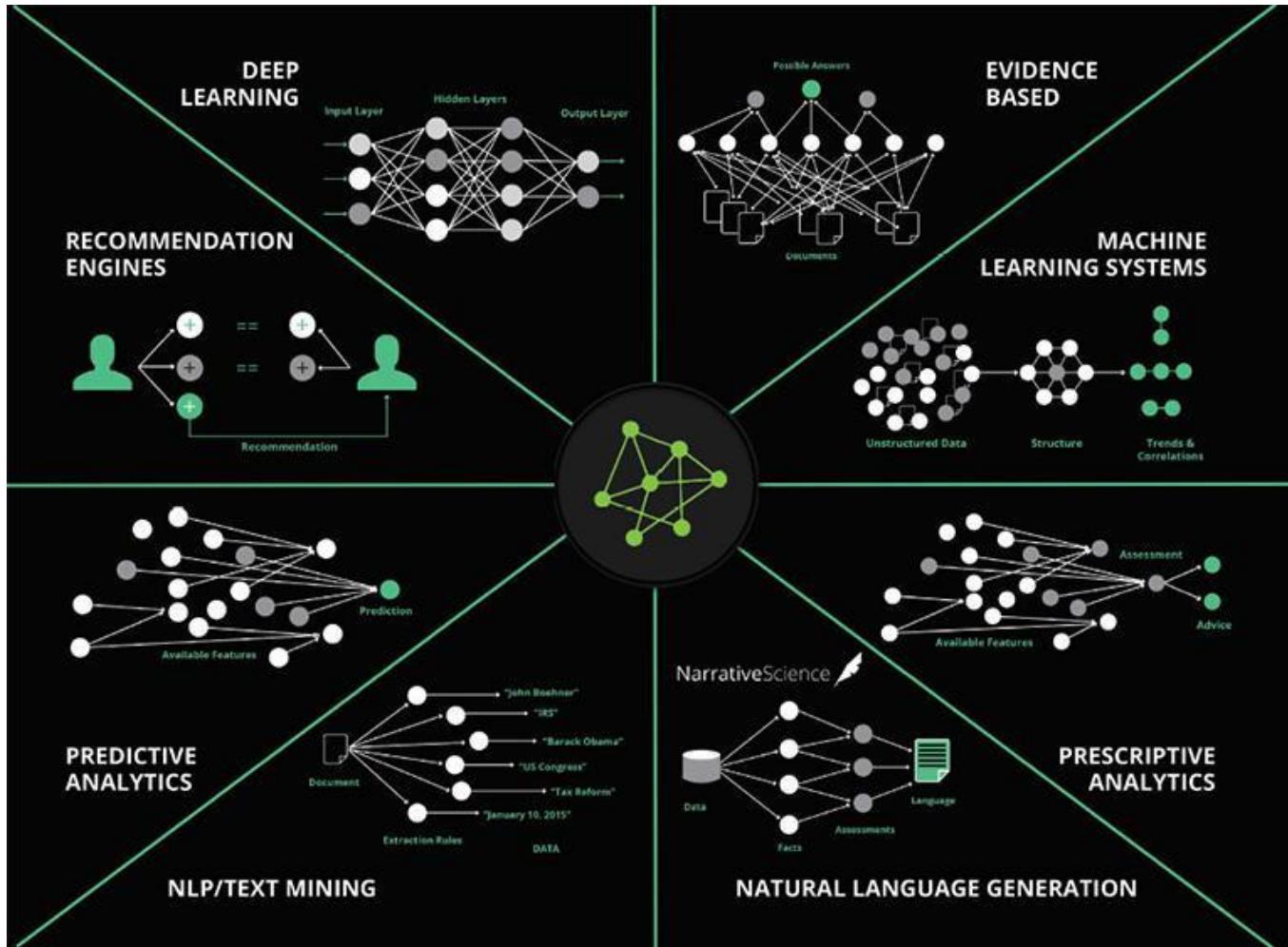
➤ Solution Directions to the Black Box Problem



(<http://www.cognub.com/index.php/cognitive-platform/>)

What is a Big Algorithm?

➤ Different types of Machine Learning Algorithms explained



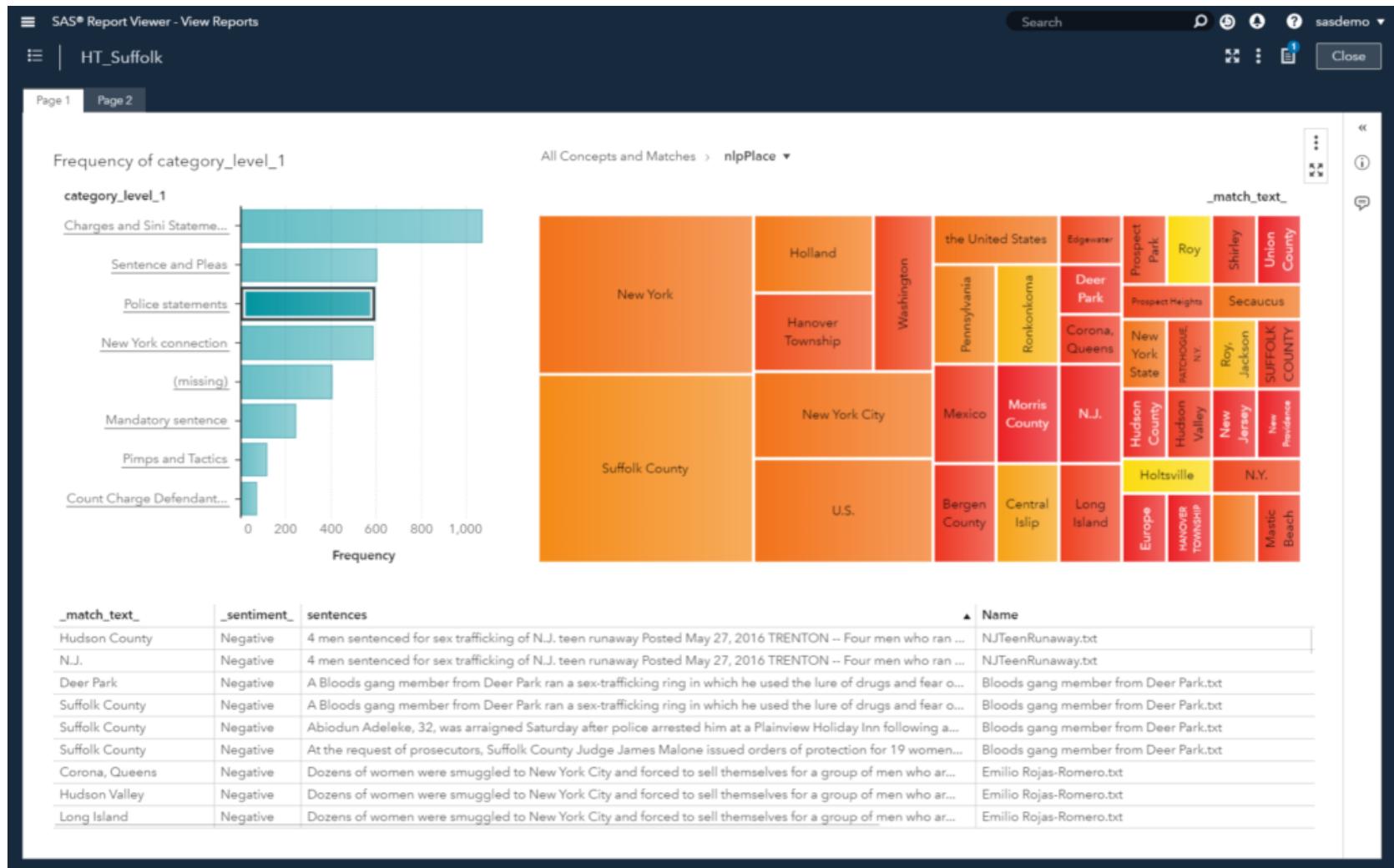
Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era



Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era



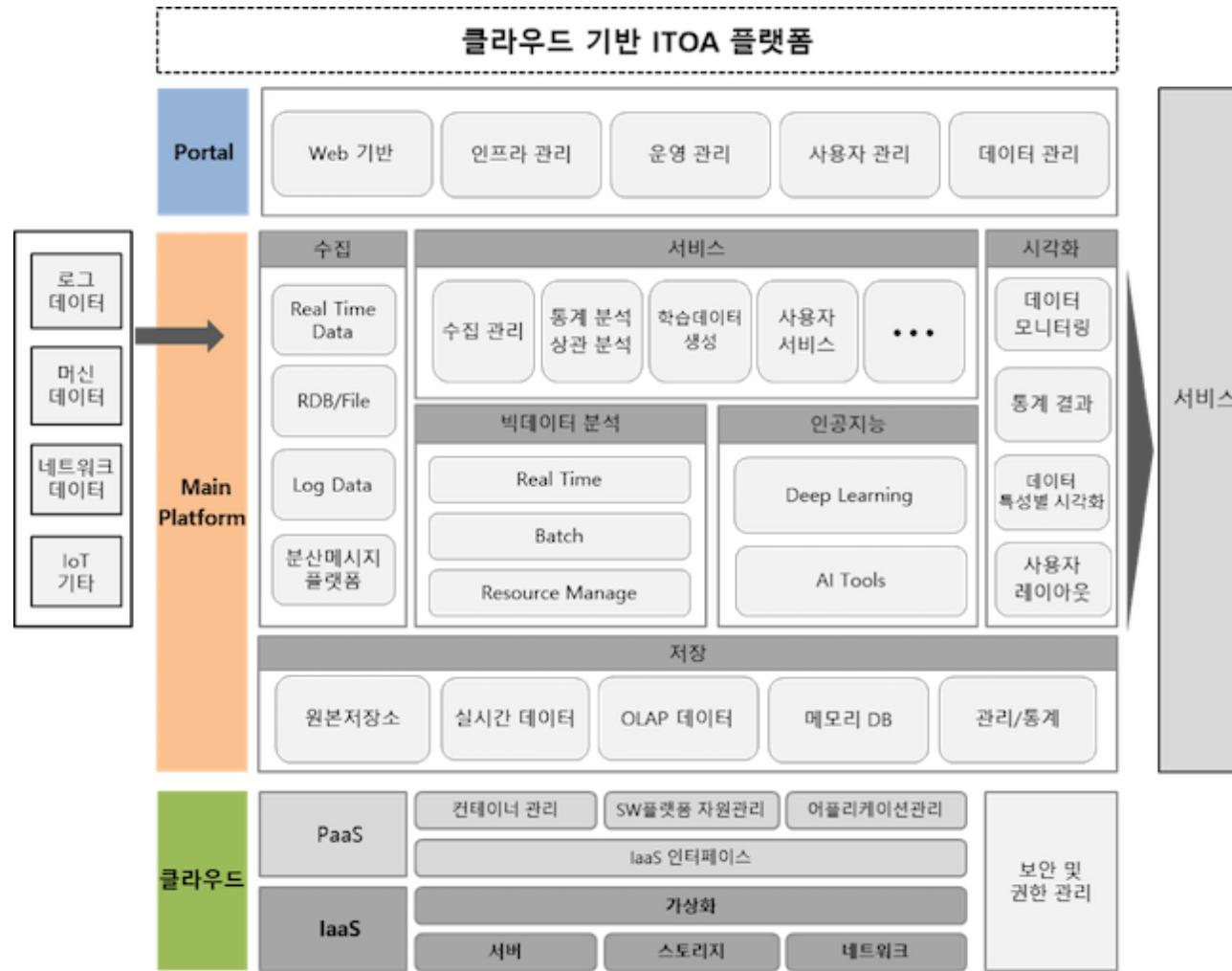
Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era



Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era



Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era

Business
Intelligence
Solutions



Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era



Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era



Autonomous Taxi's
are the future of
city living!

Digital Platform with Big Data/Algorithm

➤ Digital Platforms in the Digital Economy Era



WE ALL LIVE IN A DIGITAL ECONOMY!

I. 글로벌 트랜드, 데이터과학, 인공지능

- Global “Business” Trends
- Global “Technology” Trends
- Global “Academia” Trends
- “Artificial Intelligence” vs. “Data Science”

II. 데이터과학과 디지털경제

- What is the Communication between Data and the World?
- Why Data Science is in a sudden boom?
- What is a Big Data?
- What is a Big Algorithm?
- Digital Platform with Big Data/Algorithm

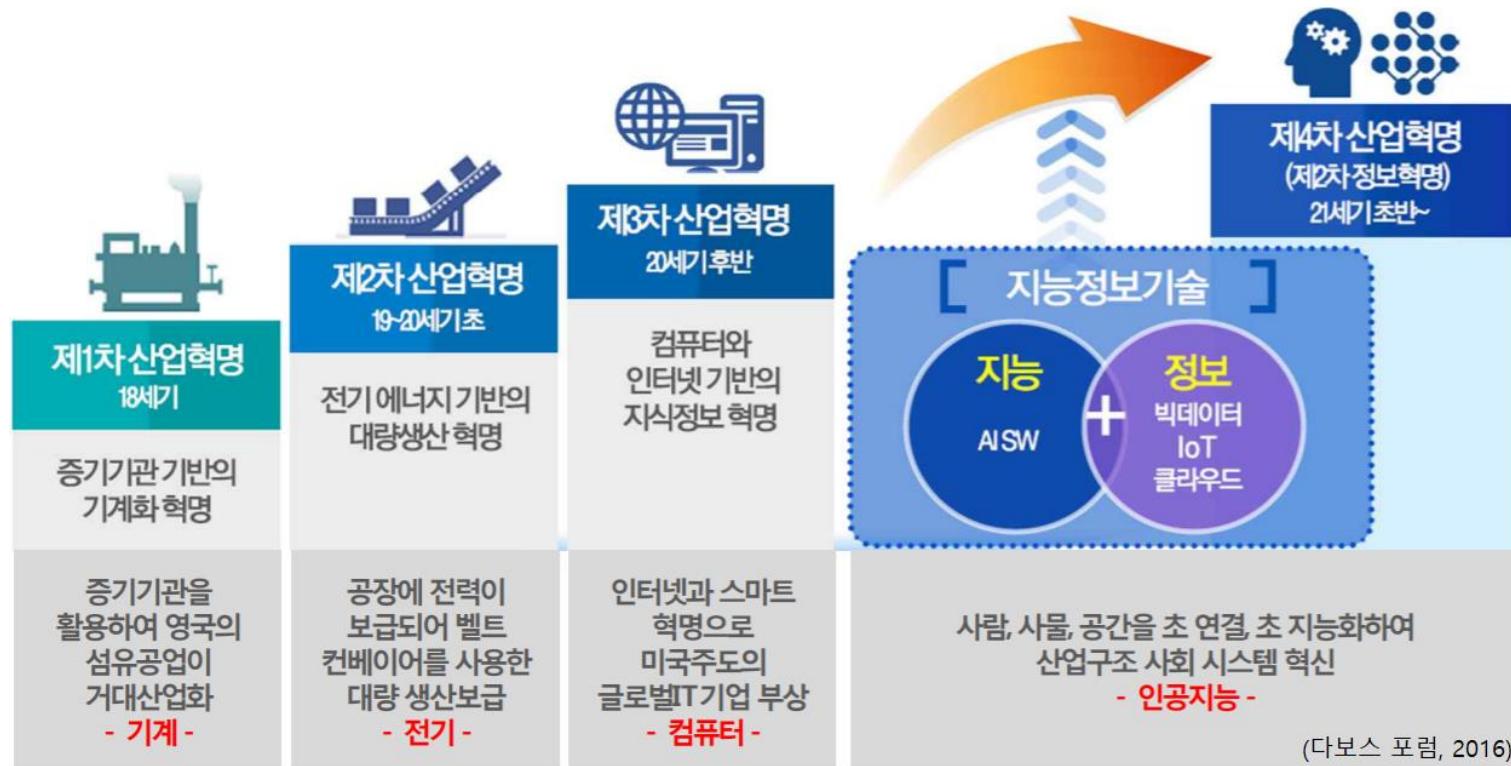
III. 데이터와 인공지능을 활용한 디지털전환

- Evolution of Hardware for Digital Platforms
- Evolution of Computing Platforms
- Skills and Jobs in the Data Science
- Driving Digital Transformation with Data Science as a Service

Evolution of Hardware for Digital Platforms

➤ The Process of the Industrial Revolution

- Certain technologies have led to the development of the industrial revolution.
- The 4th industrial revolution is driven by knowledge information-based big data and artificial intelligence.



Evolution of Hardware for Digital Platforms

➤ The Process of the Industrial Revolution



Evolution of Hardware for Digital Platforms

➤ “산업혁명 요약: 사람은 본능적으로 편안한 것을 좋아한다”

- 1) 노예/수공자에서 기계/관리로 바꾸며 몸이 힘든 일을 덜어주었다
- 2) 몸이 힘든 일보다 머리가 아픈 일이 더 많아졌다 (어떤 때는 몸이 힘든 게 낫다)
- 3) 앞으로는 머리 아픈 일조차 컴퓨터가 대신해준다고 한다 (몸도 머리도 놀 것 같아서 좋을 것 같기도 하다)
- 4) 노동 생산성보다 기계/컴퓨터의 생산성이 비교도 안되게 늘어난다 (몸도 머리도 놀더니 일도 평생 놀란다)
- 5) "변화를 즐기고 새로운 것과 옛날 것 모두를 즐겨야 함" (사람은 익숙함에 능하지 변화는 어려워 한다)

1차 산업혁명	2차 산업혁명	3차 산업혁명	4차 산업혁명
- 사람의 손 ▷ 증기기관 동력 기술혁신	- 증기기관 ▷ 모터, 전기, 석유 등력 기술혁신	- 전자회로, 정밀제어 등 컴퓨터와 인터넷 기술 혁신	- 인공지능, 사물인터넷, 데이터집약 기술혁신
- 면직물 대량 생산 제조업	- 면직물 ▷ 강철, 석유, 자동차, 화학, 전기 대량 생산 제조업	- 모든 산업 대량 생산 ▷ + 문화, 서비스, 네트워크, 커뮤니케이션, 정보 산업 발전	- 모든 산업 데이터화 + 연결 + 인공지능 + 맞춤제품생산
- 농민이나 자영업자 ▷ 육체노동자	- 육체노동자 ▷ 대량 노동자 + 실업	- 육체노동자 + 사무노동자 + 생산소비자(프로슈머)	- 육체노동자 + 사무노동자 + 생산소비자(프로슈머) + 새로운 직업

Evolution of Hardware for Digital Platforms

➤ Background of IT Development

- Computer processing, memory chips, storage, mobile communication, networking hardware and software, and software design have contributed to the advancement of IT, with exponential cost reductions and dramatic improvements in computing performance.

(1) 무어의 법칙과 마이크로프로세싱 파워

(2) 대용량 디지털 저장 법칙(Law of Mass Digital Storage)

(3) 멧칼프의 법칙과 네트워크 경제

(4) 감소하는 통신비용과 인터넷

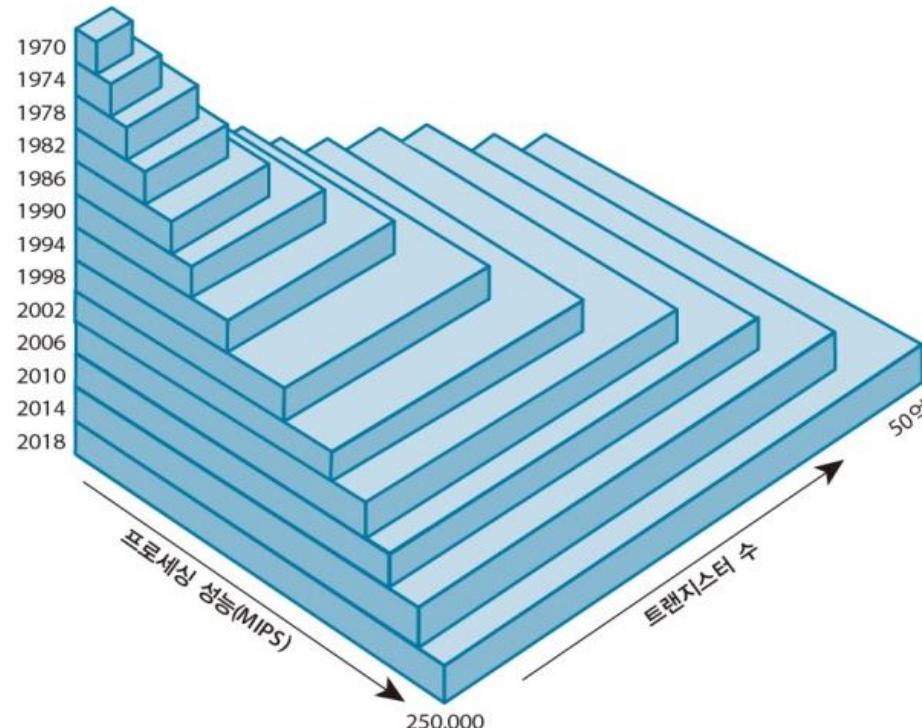
Evolution of Hardware for Digital Platforms

➤ Background of IT Development

(1) 무어의 법칙과 마이크로프로세싱 파워

- 통합 회로 제조업체 페어차일드 반도체(Fairchild Semiconductor)의 연구개발실 책임자 고든 무어(Gordon Moore)는 마이크로프로세서 칩(microprocessor chip)이 나온 1959년 아래 최소의 제조비용으로 제작된 칩의 구성요소 숫자가 매년 2배씩 증가한다 논문을 기고
 - . 마이크로프로세서의 성능은 18개월마다 2배 + 컴퓨팅 성능은 18개월마다 2배 + 컴퓨팅 가격은 18개월마다 절반
- “성장률이 2년마다 2배씩 증가”

무어의 법칙은 더 강력한 성능을 의미한다.



Evolution of Hardware for Digital Platforms

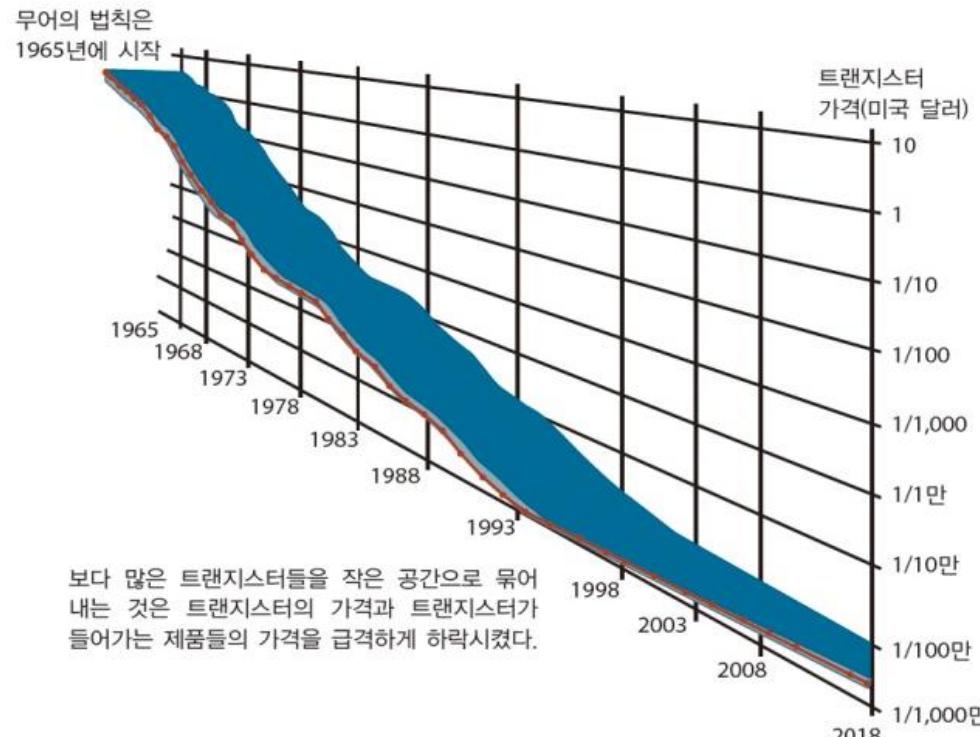
➤ Background of IT Development

(1) 무어의 법칙과 마이크로프로세싱 파워

- 통합 회로 제조업체 페어차일드 반도체(Fairchild Semiconductor)의 연구개발실 책임자 고든 무어(Gordon Moore)는 마이크로프로세서 칩(microprocessor chip)이 나온 1959년 아래 최소의 제조비용으로 제작된 칩의 구성요소 숫자가 매년 2배씩 증가한다 논문을 기고
 - . 마이크로프로세서의 성능은 18개월마다 2배 + 컴퓨팅 성능은 18개월마다 2배 + 컴퓨팅 가격은 18개월마다 절반

→ “성장률이 2년마다 2배씩 증가”

무어의 법칙은 가격의 감소를 의미한다.



Evolution of Hardware for Digital Platforms

➤ Background of IT Development

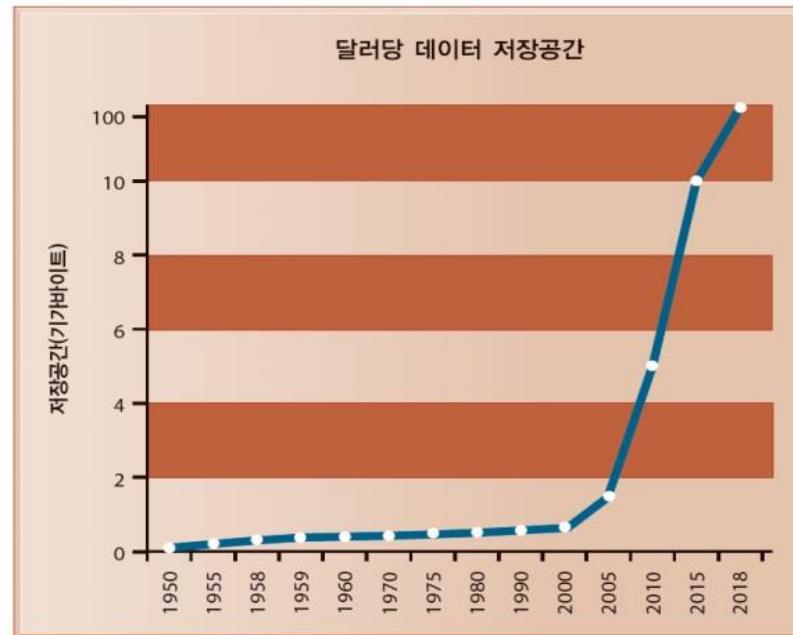
(2) 대용량 디지털 저장 법칙(Law of Mass Digital Storage)

- 디지털 정보의 양은 대충 매년 2배씩 증가한다(Lyman and Varian, 2003)
- 디지털 정보를 저장하는 비용은 1년에 100%씩 기하급수적인 비율로 감소

그림 5.6 급속하게 증가한 달러당 저장공간(1950~2018년)

구글 드라이브와 같은 클라우드 스토리지 서비스는 매달 1.99달러에 100기가바이트의 저장공간을 제공한다.

출처 : 저자 추정

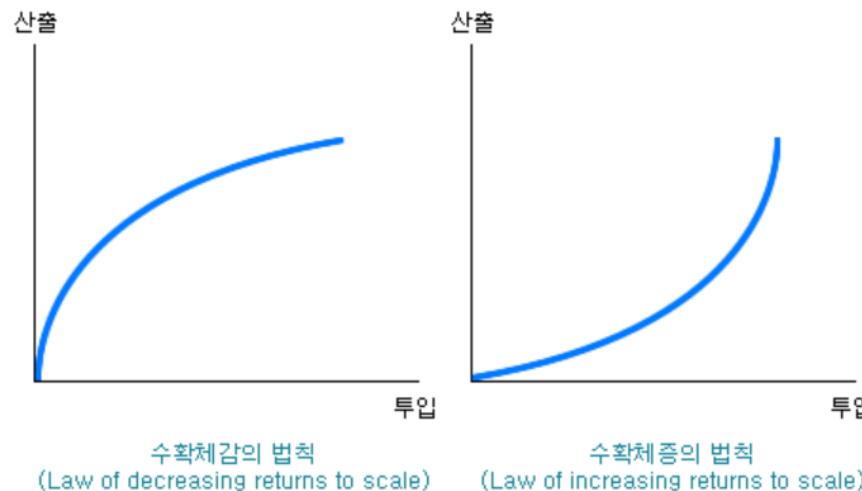


Evolution of Hardware for Digital Platforms

➤ Background of IT Development

(3) 멧칼프의 법칙과 네트워크 경제

- 무어의 법칙과 대용량 저장 법칙은 이제 왜 이렇게 사용 가능한 컴퓨팅 자원들이 많은지를 이해
- 사람들은 왜 더 강력한 컴퓨팅 및 저장 성능을 원하는가?
 - . 이더넷(Ethernet) LAN 기술을 발명한 로버트 멧칼프(Robert Metcalfe)는 1970년에 네트워크 성장의 가치 또는 파워는 네트워크 사용자들의 숫자에 대한 함수로서 급격하게 증가한다고 주장
 - . **수확체증의 법칙(Increasing Returns to Scale):** 네트워크 참여자들이 증가할수록 참여자들은 더 많은 것을 얻을 수 있음
 - . 네트워크의 참여자 수가 선형적 증가함에 따라 전체 시스템의 가치는 급격하게 증가, 이론적으로 영원히 지속



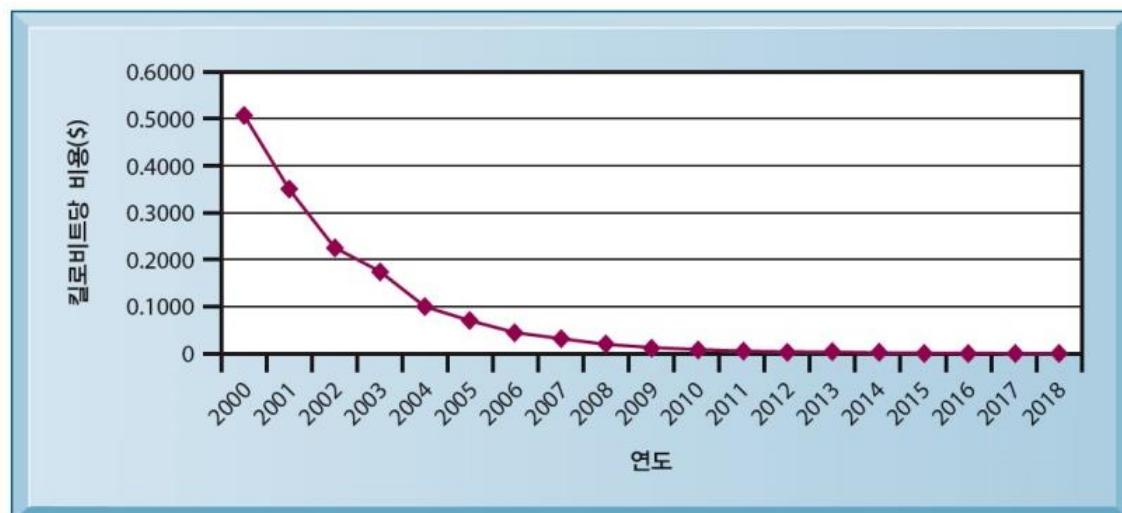
Evolution of Hardware for Digital Platforms

➤ Background of IT Development

(4) 감소하는 통신비용과 인터넷

- 현재 전 세계의 인터넷 접속자는 대략 42억 명 정도다(Internetworldstats, 2018)
- 통신비용 거의 0에 이르게 됨에 따라, 통신 설비와 컴퓨팅 설비의 활용은 폭발적으로 증가
 - . 2018년, 인터넷을 통한 1메가비트 접속비용은 약 2.6달러지만 2000년에 이에 대한 비용은 300달러 이상
 - . 동일한 기간에 가정집 평균 인터넷 속도는 0.2Mbps에서 18Mbps로 증가
- **기업들이 인터넷과 비즈니스 가치를 활용하기 위해 무선 접속 포함 인터넷 접속을 확장시켜야 하며, 클라이언트/서버 네트워크, 데스크톱 클라이언트, 모바일 컴퓨팅 기기의 성능도 크게 증대시켜야**

출처 : 2007–2018: “Average Internet Connection Speed in the United States from 2007 to 2017 (in Mbps), by Quarter” Statista, 2018;
2006 Home Broadband Adoption 2006 BY John B. Horrigan PEW Research 2007; Internet speeds: How Fast Does Internet Speed grow?By Xah Lee, Date: 2006-12-30, Last updated: 2017-01-22, <http://xahlee.info/comp/bandwidth.html>



Evolution of Computing Platforms

➤ Development of IT Infrastructure/Computer

- An organization's IT infrastructure is a **continuous computing evolution** over more than 50 years.

Evolution of Com

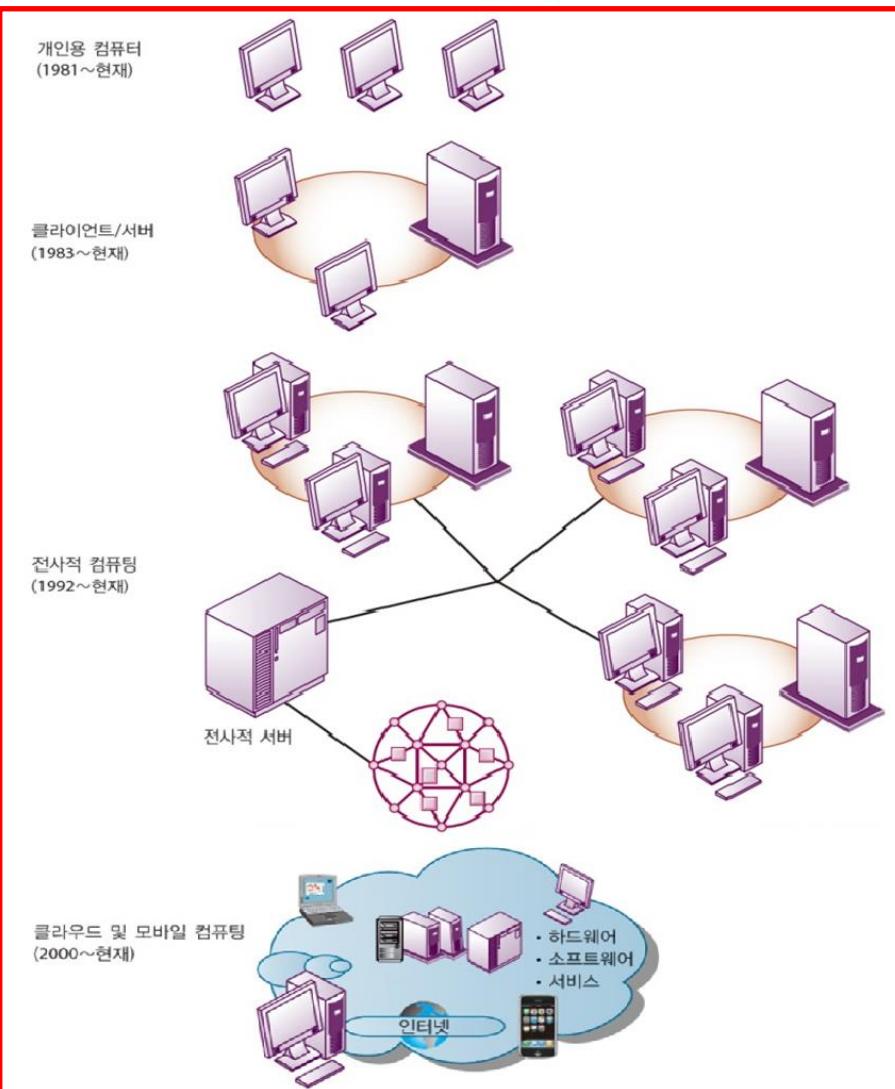
그림 5.2 IT 인프라 발전 과정

여기서는 다섯 가지의 IT 인프라 발전 시대들에 대한 각각의 특징을 보여주는 전형적인 컴퓨팅 구성들을 예시한다.

IT 인프라 발전 단계

➤ Development of IT II

- An organization's IT in



'er more than 50 years.

Evolution of Computing Platforms

➤ Development of IT Infrastructure/Computer

- Cloud and Mobile computing (2000~)
 - 인터넷의 성장은 클라이언트/서버 모델을 “클라우드 컴퓨팅 모델” 방향으로 한 단계 더 업그레이드
 - . 주로 인터넷을 통해 컴퓨팅 자원들(컴퓨터, 저장장치, 애플리케이션, 서비스) 접근을 가능하게 해주는 컴퓨팅 모델
 - . 컴퓨팅 자원의 “클라우드”는 어떤 기기나 위치에서도 필요에 따라 접근 가능
 - . 가장 빠르게 성장하고 있는 컴퓨팅 형태로서 2020년에는 전 세계적으로 공공 클라우드 지출 비용이 4,110억 달러
 - . 2021년경 모든 컴퓨터 작업량의 94%가 클라우드 환경에서 처리될 것으로 전망(Gartner, 2017; Cisco, 2018)
 - . 아마존, 구글, IBM, 마이크로소프트 등은 거대하고 확장 가능한 클라우드 컴퓨팅 센터를 운영함으로써 IT 인프라를 원격관리 하고픈 기업들에게 컴퓨팅 성능, 데이터 저장장치, 그리고 초고속 인터넷 접속을 제공
 - . 구글, 마이크로소프트, SAP, 오라클, 세일즈포스닷컴 등의 기업들은 소프트웨어 애플리케이션을 인터넷을 통해 전달되는 서비스로서 판매

Evolution of Computing Platforms

➤ Development of IT Infrastructure/Computer

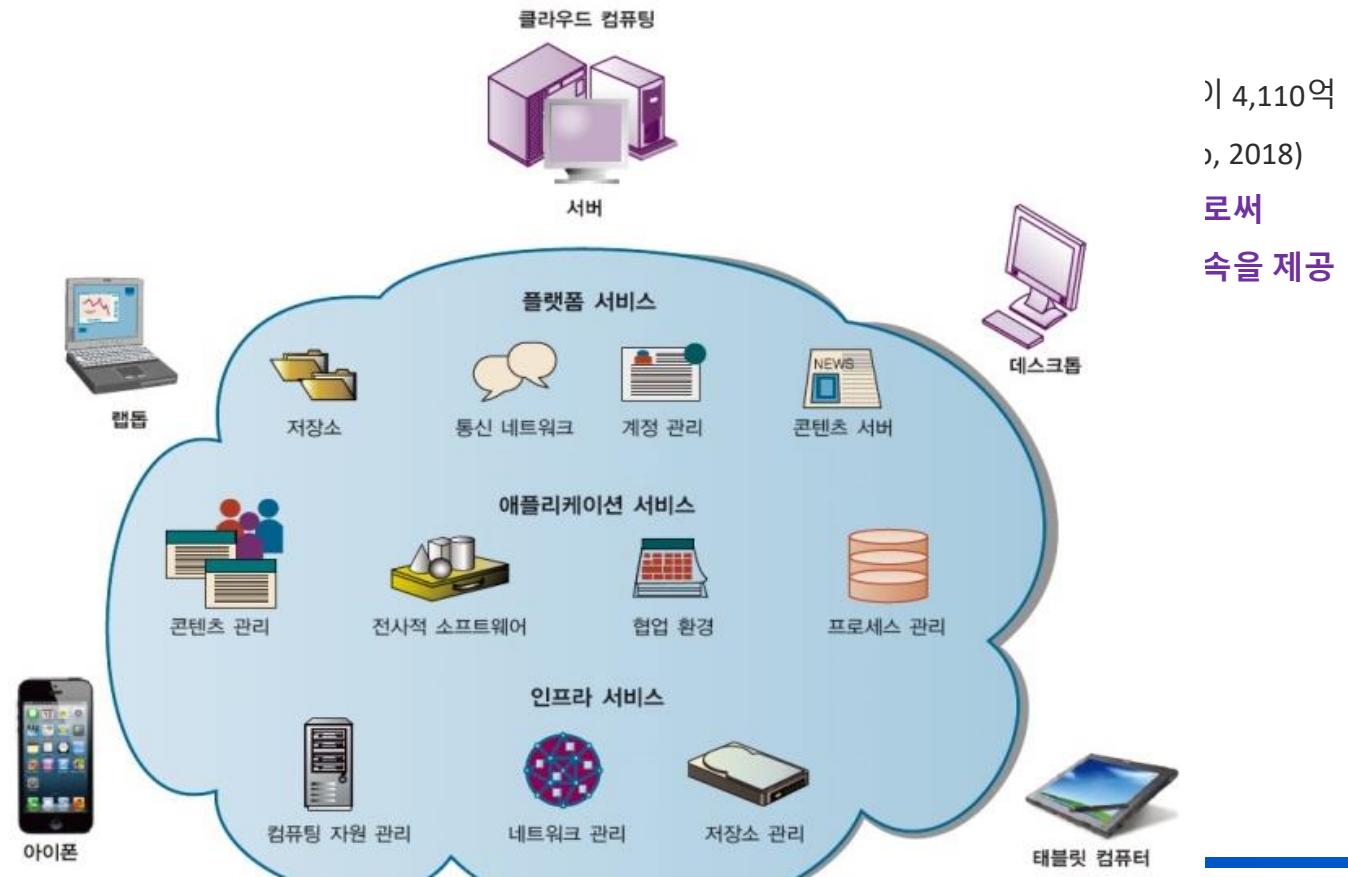
▪ Cloud and Mobile computing (2000~)

- 인터넷의 상

- . 주로 인터넷
- . 컴퓨팅 자원
- . 가장 빠르다
- . 2021년경
- . 아마존, 구글, IT 인프라 구축
- . 구글, 마이크로소프트웨어

그림 5.9 클라우드 컴퓨팅 플랫폼

클라우드 컴퓨팅 환경에서는 하드웨어와 소프트웨어 기능들이 가상화된 자원의 집합으로서 네트워크, 주로 인터넷을 통해 제공된다. 기업들과 직원들은 언제 어디서든 그리고 어떤 기기를 통해서든 애플리케이션과 IT 인프라에 접근할 수 있다.



| 업그레이드
는 컴퓨팅 모델
기 4,110억 달러
, 2018)
로써
속을 제공

Evolution of Computing Platforms

➤ Development of IT Infrastructure/Computer

그림 2. 데이터 시대의 필수 플랫폼: 모든 데이터는 클라우드를 통한다!



자료: 삼성증권

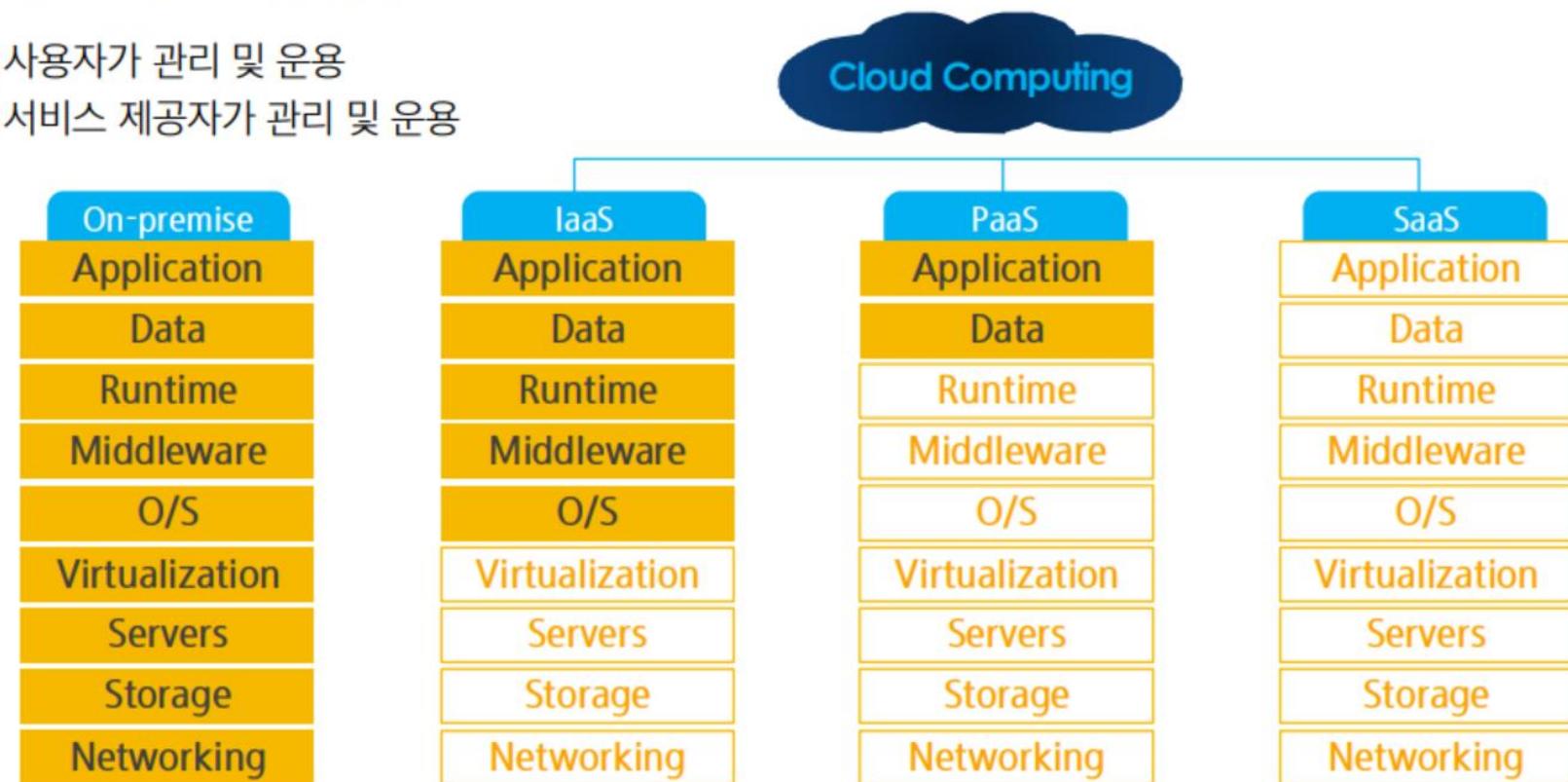
Evolution of Computing Platforms

➤ Development of IT Infrastructure/Computer

그림 15. 서비스 모델에 따른 구분

■ 사용자가 관리 및 운용

□ 서비스 제공자가 관리 및 운용



자료: 삼성증권

Evolution of Computing Platforms

➤ Development of IT Infrastructure/Computer

그림 6. 아마존 클라우드 플랫폼 발전 과정

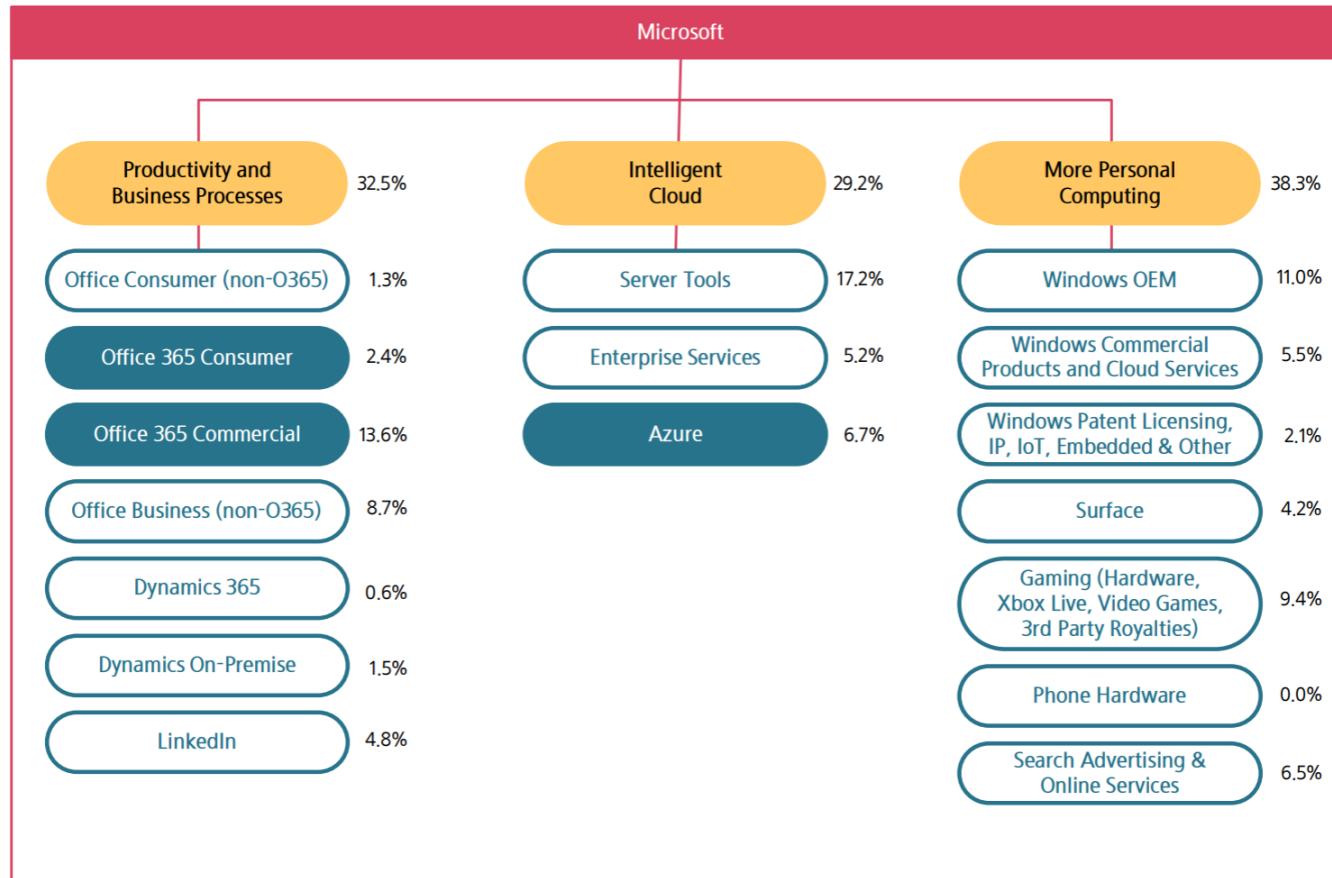
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017										
PaaS	Analytics																						
	App Services																						
	Deployment					Elastic Beanstalk 클라우드 상 앱 관리																	
	Database	SimpleDB				Relational Database																	
IaaS	Storage	Amazon S3																					
	Compute	Amazon EC2																					
	Networking				Route 53 DNS서비스																		
Rekognition 사물인식																							
SageMaker 머신러닝																							

자료: Amazon, 삼성증권

Evolution of Computing Platforms

➤ Development of IT Infrastructure/Computer

그림 76. 사업부문 및 매출비중



참고: FY 2018 매출액 기준

자료: Microsoft, 삼성증권

Evolution of Computing Platforms

➤ Non-relational Database, Cloud Database, Blockchain

기존 중앙 집중형 콘텐츠 유통 플랫폼



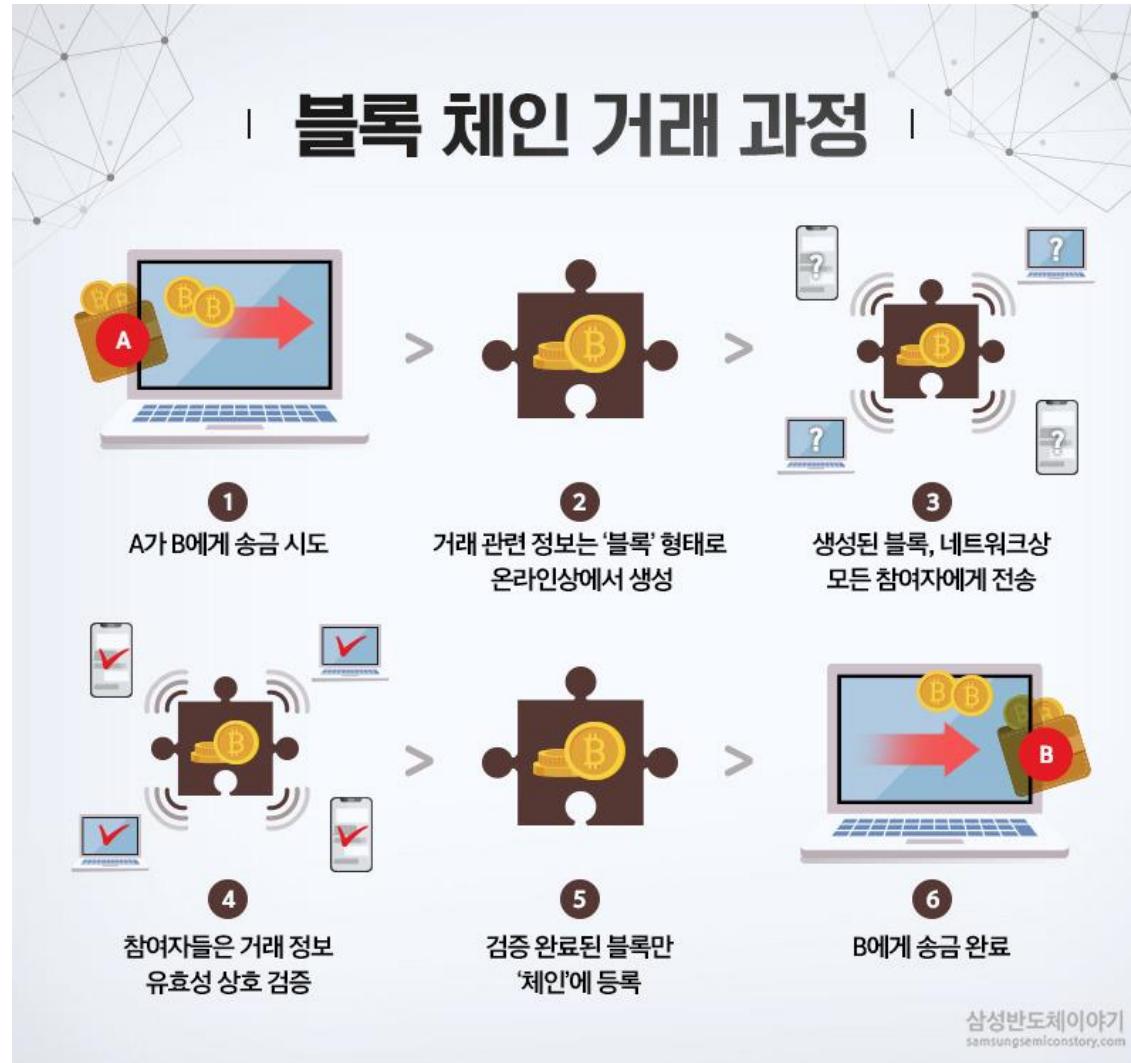
블록체인 기반 분산형 콘텐츠 유통 플랫폼



장점: 1. 공정 거래 2. 다양한 지불 방식 지원 3. 사이버 보안 우수 4. 운영 비용 없음

Evolution of Computing Platforms

➤ Non-relational Database, Cloud Database, Blockchain



Evolution of Computing Platforms

➤ Non-relational Database, Cloud Database, Blockchain

▪ 블록체인(Blockchain):

- 기업과 조직들이 중앙집중적인 권한 없이 거의 즉시 네트워크에서 거래를 생성/확인할 수 있는 분산 데이터베이스 기술
 - . 블록체인은 지속적으로 늘어나는 기록 목록을 보유하는데, 각각의 기록은 블록으로 정의
 - . 거래를 컴퓨터 네트워크의 여러 컴퓨터에 분산 원장으로 저장하고 저장된 정보는 네트워크 컴퓨터들이 지속 조정
 - . 스마트 계약(Smart Contracts) 사용을 통해 기록 거래들을 모두 표준화
 - . 원장에 대한 정당한 변경사항은 몇 초 또는 몇 분 안에 블록체인에 기록되며 암호화를 통해 보호

→ 블록체인은 사용자 검증하고 거래 유효성 인증하는 비용과 많은 기업들에 걸친 거래 정보 저장/처리 관련 위험 대폭 줄여

→ 블록체인 시스템을 기업들에게 활용 가능하고 매력적으로 만드는 요인은 암호화와 행위자 및 참여 기업들에 대한 인증

→ 수많은 기업들은 자체 거래시스템을 구축하고, 그것을 공급업체, 배송업체, 금융기관의 시스템과 통합해 왔지만,

블록체인은 그렇게 하는 대신 참여 기업들을 위한 하나의 간단한 저비용 거래시스템을 제공

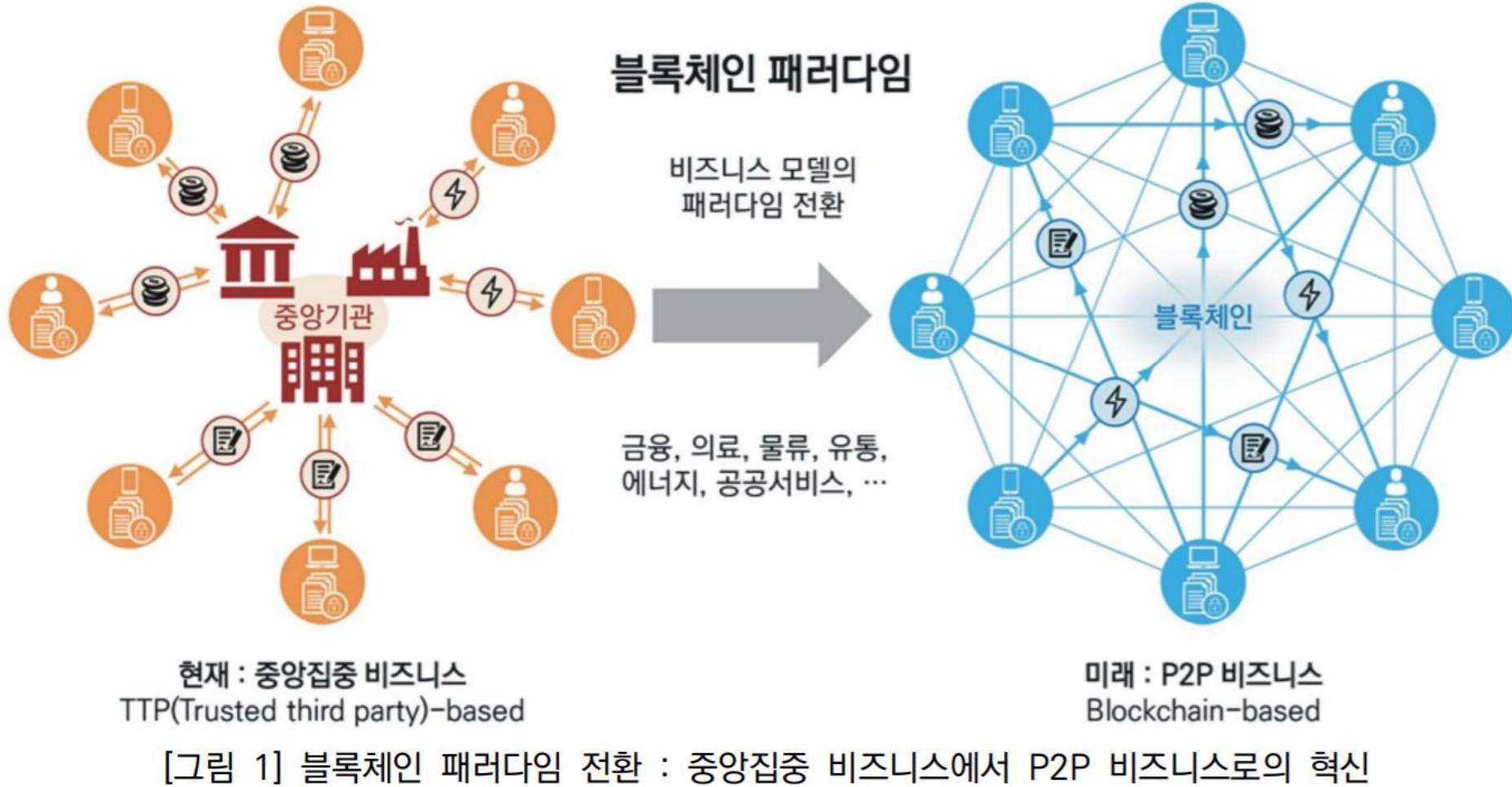
→ 블록체인의 단순성과 보안성은 금융 거래, 공급망 거래, 의료 기록 등 여러 유형 데이터들을 저장/보호하는 데 매력적

→ 블록체인은 비트코인, 이더리움을 비롯한 여타 암호 화폐의 토대가 되는 기술

Evolution of Computing Platforms

➤ Non-relational Database, Cloud Database, Blockchain

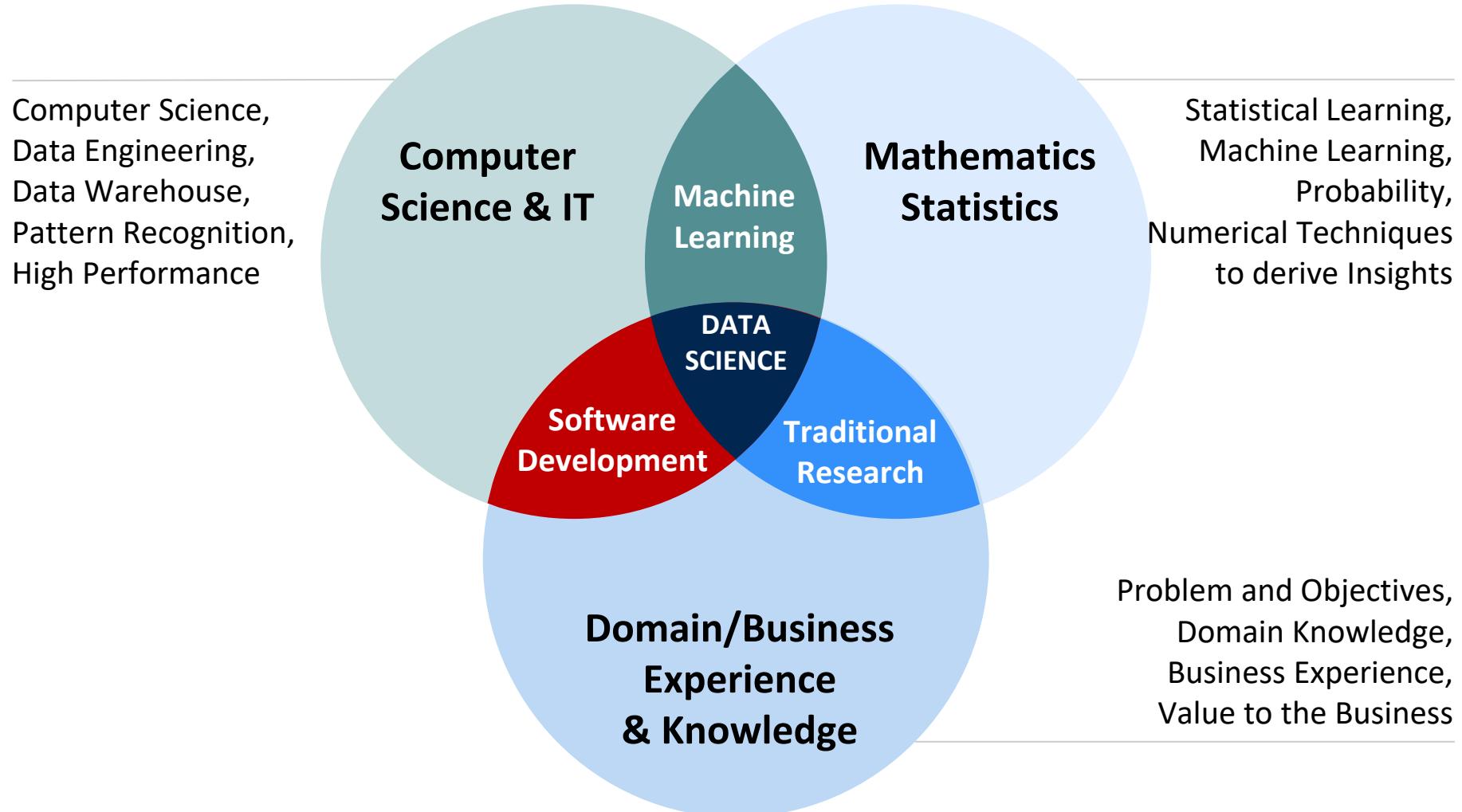
- 블록체인(Blockchain):



[그림 1] 블록체인 패러다임 전환 : 중앙집중 비즈니스에서 P2P 비즈니스로의 혁신

Skills and Jobs in the Data Science

➤ Data Science: Solving Problems with Data (Real Human Insight)



Skills and Jobs in the Data Science

➤ Computer Science, Science, and Data Science

- **Computer Science:** The study of the theory and practice of how computers work.
- **Science:** Focusing on solving problems through the lens of the domain's scientific principles.
- **Data Science:** Interdisciplinary field involving computer science and statistics.

MASTER'S IN COMPUTER SCIENCE VS. DATA SCIENCE

TWO PATHWAYS TO NEXT-GENERATION CAREERS

With today's high demand for software solutions and data analysis, professionals are pursuing graduate education to enhance their knowledge and skills but soon have the dilemma of deciding which master's program is the best fit—computer science or data science. While the programs share some similarities, there are also very differences to consider; explore the benefits of Lewis University's online Master of Science in Computer Science (MSCS) and Master of Science in Data Science (MSDS) programs.

The infographic compares two pathways: Computer Science and Data Science. It features two circular icons: one red for Computer Science showing a monitor and keyboard, and one blue for Data Science showing a monitor and a bar chart. The text highlights that both fields require advanced computing, in-depth experience in developing enterprise-scale applications, database systems, security solutions, and automated systems. It also mentions developing next-generation technology in software, cyber security, and intelligent systems. Computer Science is described as the prime mover of technological innovations. Data Science is described as discovering meaning within big data, using data mining, data visualization, predictive analysis, and efficient data management. Both fields require honing subject-matter expertise and clarifying large data sets for better organization and decision-making. The infographic concludes by noting that data is important to any company and requires experts who can make sense of it.

THE BIG PICTURE

WHAT'S IT ABOUT

Shaping the science of technology

WHAT YOU'LL LEARN

Advanced computing, including in-depth experience in developing enterprise-scale applications, database systems, security solutions and automated systems.

WHAT'S IT FOR

Developing next-generation technology in software, cyber security and intelligent systems.

WHY IT MATTERS

Computer Science is the prime mover of today's technological innovations.

WHAT'S IT ABOUT

Discovering the meaning within big data

WHAT YOU'LL LEARN

The mathematics and computer science of analyzing large data collections using data mining, data visualization, predictive analysis and efficient data management.

WHAT'S IT FOR

Honing subject-matter expertise and the skills needed to clarify the meaning of large data sets, which can help lead to better organization decision-making.

WHY IT MATTERS

Data is important to any company, and the sheer quantity of it requires experts who can make sense of it.

CAREERS

SELECTED CAREERS

- Application Developer
- Computer Engineer
- Computer Programmer
- Database Architect
- Database Developer
- Data Center Manager
- IT Engineer
- Mobile Specialist
- Network Administrator
- Network Architect
- Network Engineer
- Software Engineer
- Systems Architect
- Systems Programmer
- Web Developer

Business Intelligence Manager

Business Systems Analyst

Clinical Researcher

Computational Biologist

Data Analyst

Database Developer

Data Scientist

Data Strategist

Health Informatics Analyst

Financial Analyst

Marketing Analyst

Predictive Modeler

Researcher

Research Analyst

Risk Analyst

Statistician

Skills and Jobs in the Data Science

➤ What should we do?



Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is:

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing package e.g. R
- ★ Databases SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative



COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

Skills and Jobs in the Data Science

➤ Data Science Team



Data Analyst (DA)

: Assist DS with domain understanding, data preprocessing and problem defining.



Data Scientist (DS)

: Prepares data, engineers features, most valuable skill: training models.



Data Engineer (DE)

: Data acquisition focus. Build data pipelines. Not uncommon to have 5:1 ratio DE:DS



Data Application Architect (DAA)

: Design complete solution; deploy and maintain models in production

Skills and Jobs in the Data Science

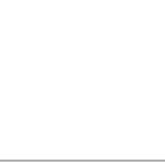
➤ Data Science Team

	Data Analyst	Analyst	(정통) 기획 및 전략	- 기획서, 보고서, 기초분석, 시각화
	Data Analyst	Data Analyst	데이터분석 설계	- 데이터 표준 확립, 구조 및 품질관리, 데이터 분석
	Data Scientist	Statistician	(정통) 통계기반 연구	- 결과가 나온 이유를 연구하는 방향
	Data Scientist	Data Scientist	데이터분석 전반 연구	- Why보다 Result 마련/대응/전략/의사결정 방향
	ML / DL Engineer	ML / DL Engineer	응용 및 구현 연구	- 연구결과를 실질적 서비스 및 비즈니스 집중화
	Data Engineer	Database Administrator	데이터베이스 관리자	- 데이터베이스 운영, 관리, 설계
	Data Engineer	Back-end Engineer	(정통) 백엔드 개발자	- 서버 개발, 데이터베이스 시스템 구현 및 관리 - RDB/NOSQL/Hbase/Spark 등
	Data Engineer	Infra Engineer	인프라 엔지니어	- 데이터 파이프라인 구축 및 운영 - Cloud/Spark/Hadoop 등

Skills and Jobs in the Data Science

➤ Typical Collaboration of Data Science Project

- Makes data science teams more productive
- Broad support for open source libraries in various languages



**Understand
Business
Objectives**

**ID Mapping
Procure
Training
Data**

**Prepare Data
and Build
Features**

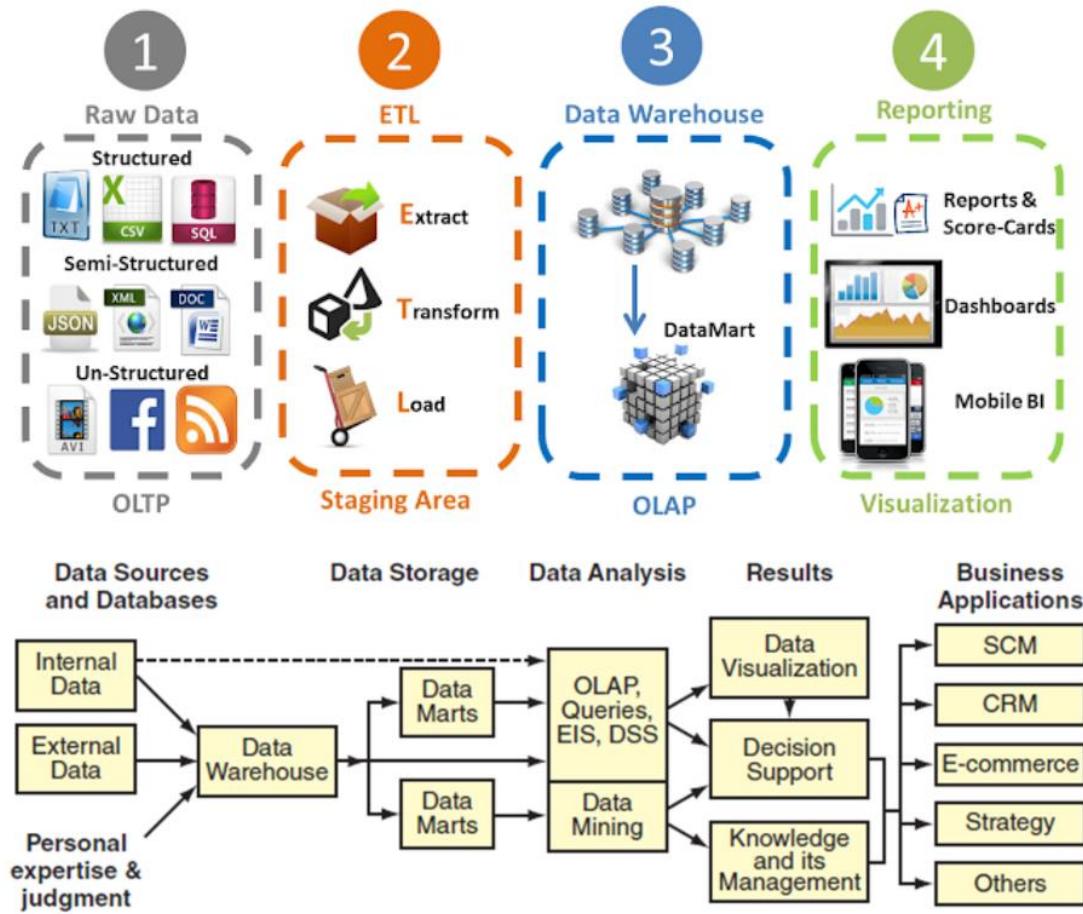
**Train, Tune,
and Test
Models**

**Deploy and
Operationalize
Models**

**Update
Models**

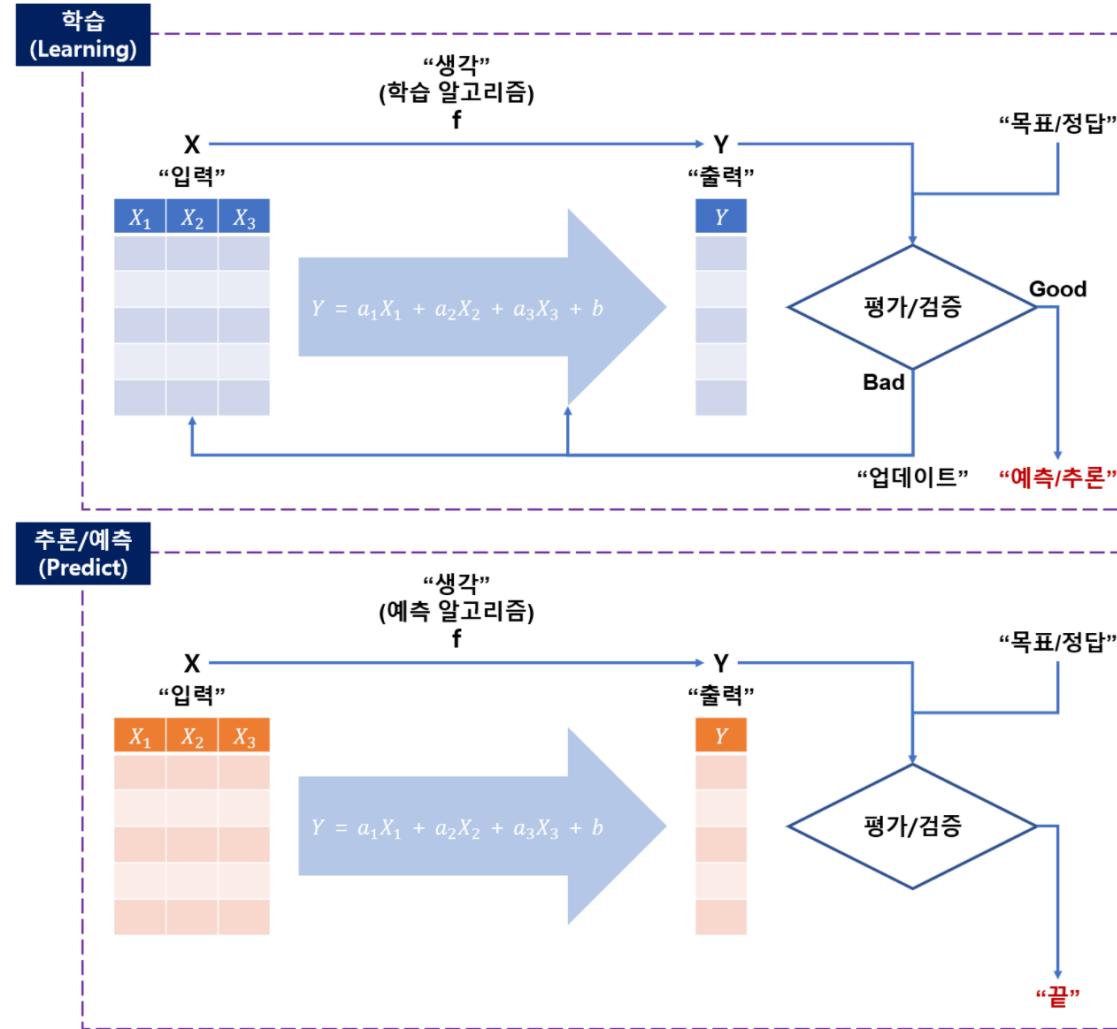
Driving Digital Transformation with Data Science as a Service

➤ Getting from Raw Data to Outcomes



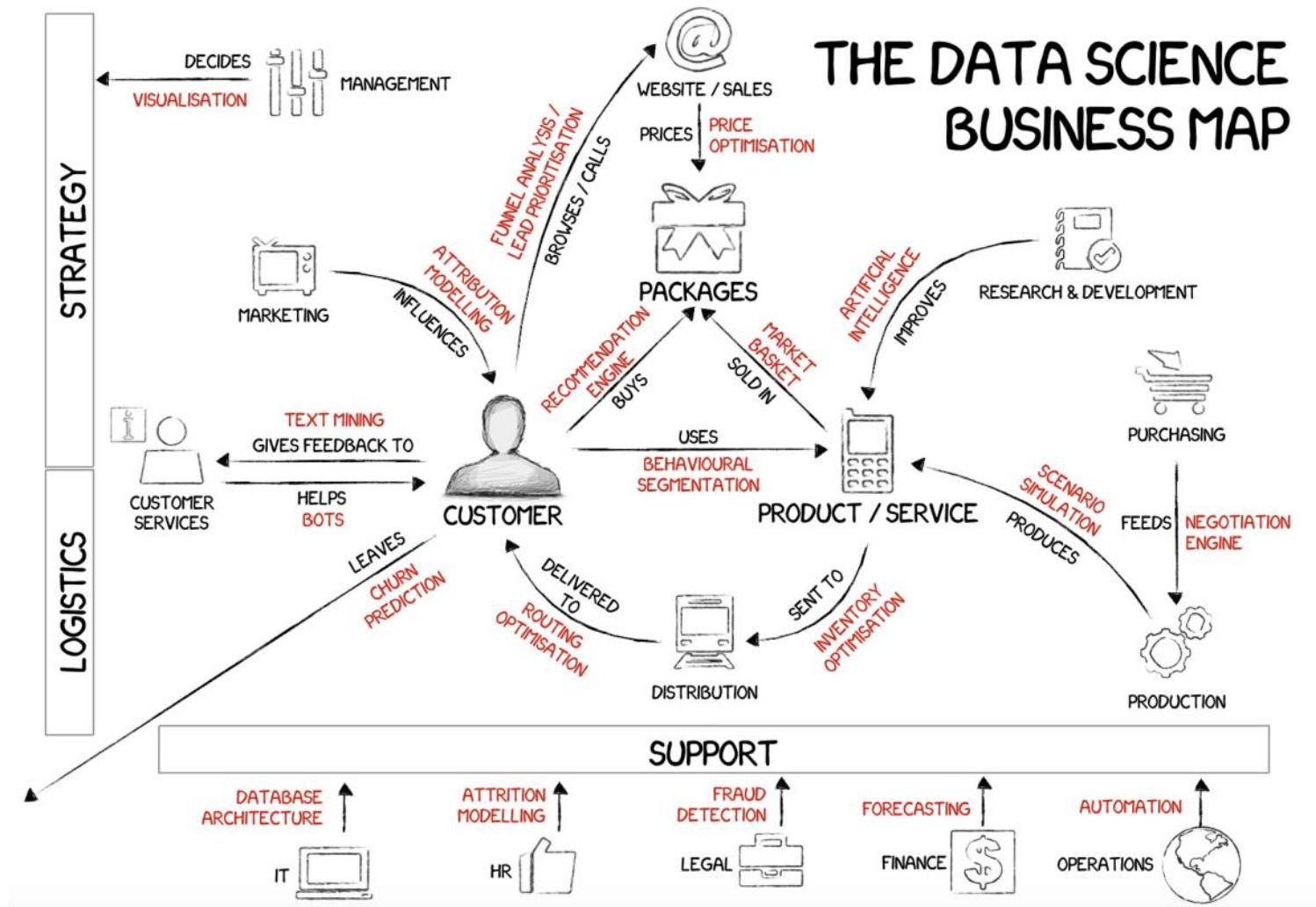
Driving Digital Transformation with Data Science as a Service

➤ Getting from Raw Data to Outcomes



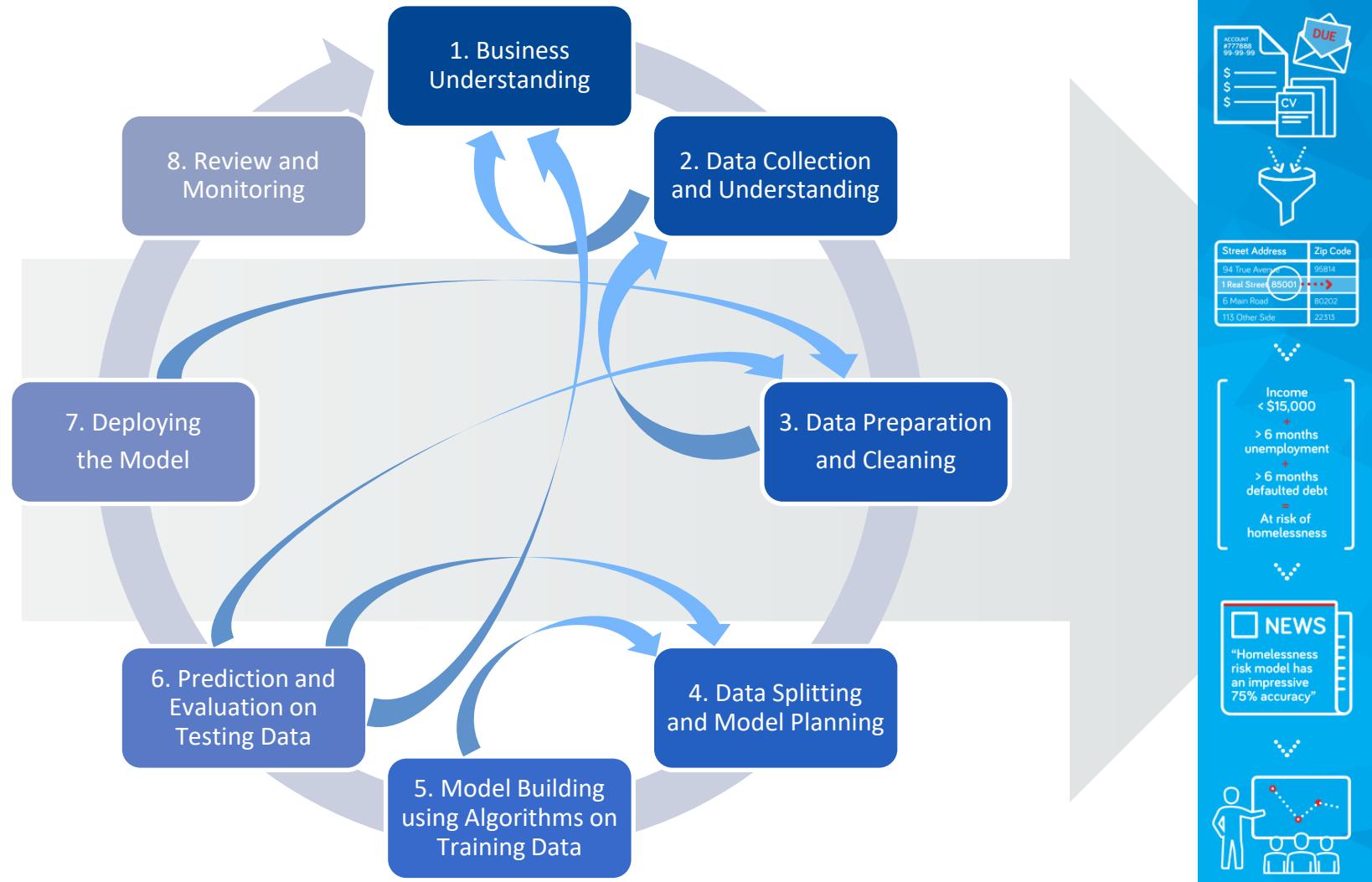
Driving Digital Transformation with Data Science as a Service

➤ The Real World Data Science



Driving Digital Transformation with Data Science as a Service

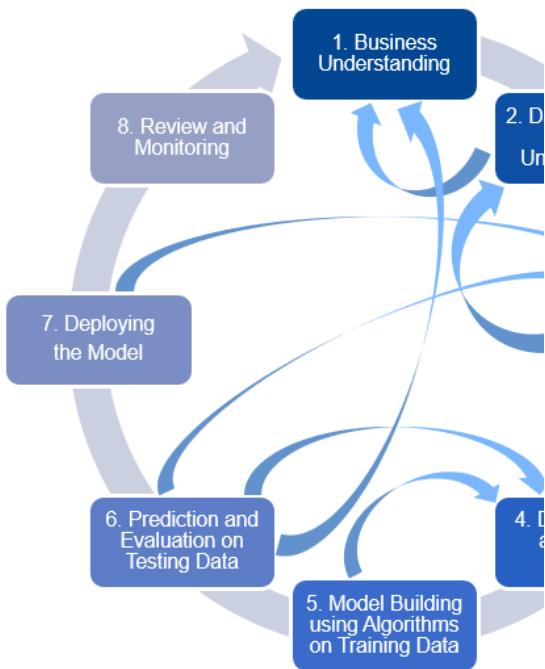
➤ Getting from Raw Data to Outcomes



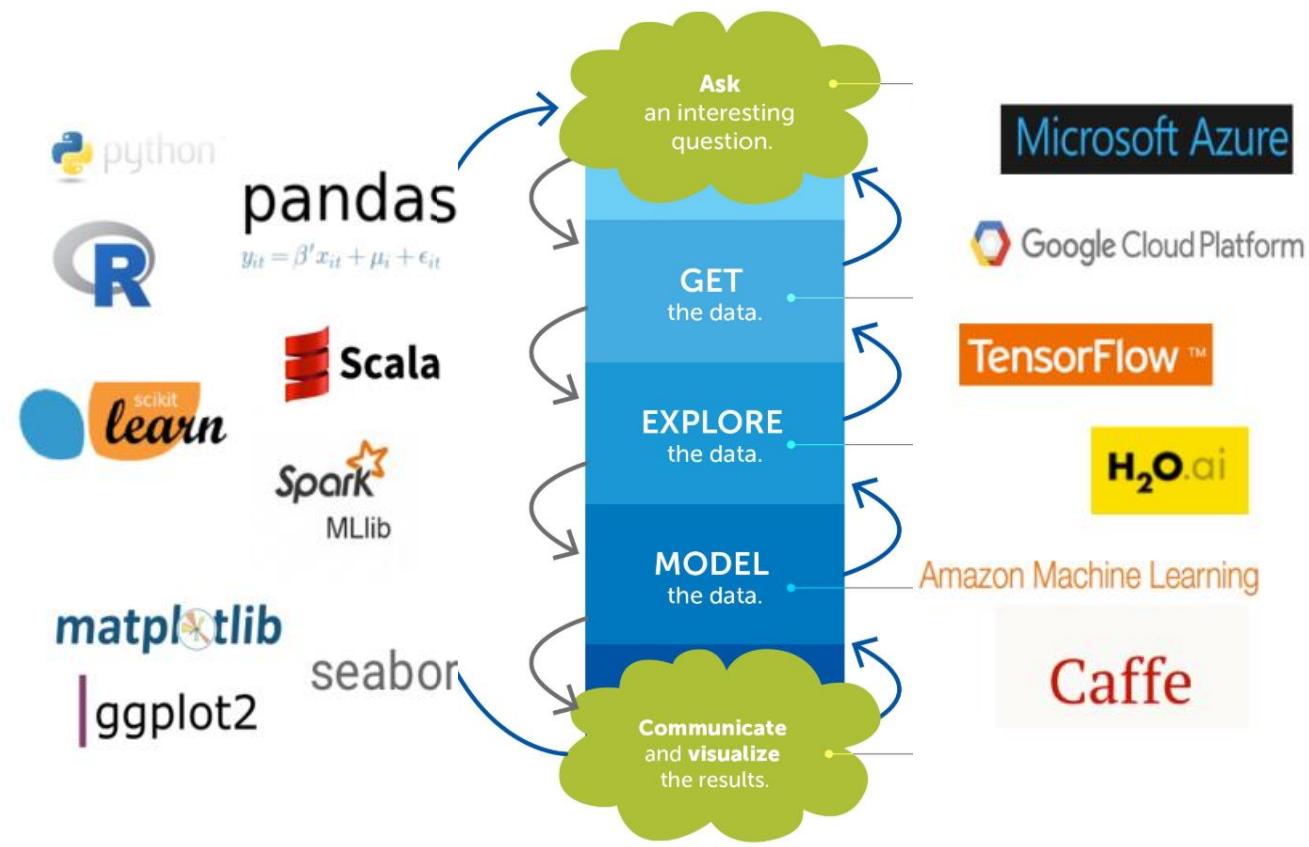
Driving Digital Transformation with Data Science as a Service

➤ Getting from Standard Framework to Data Science

“Standard Framework for Data Mining”



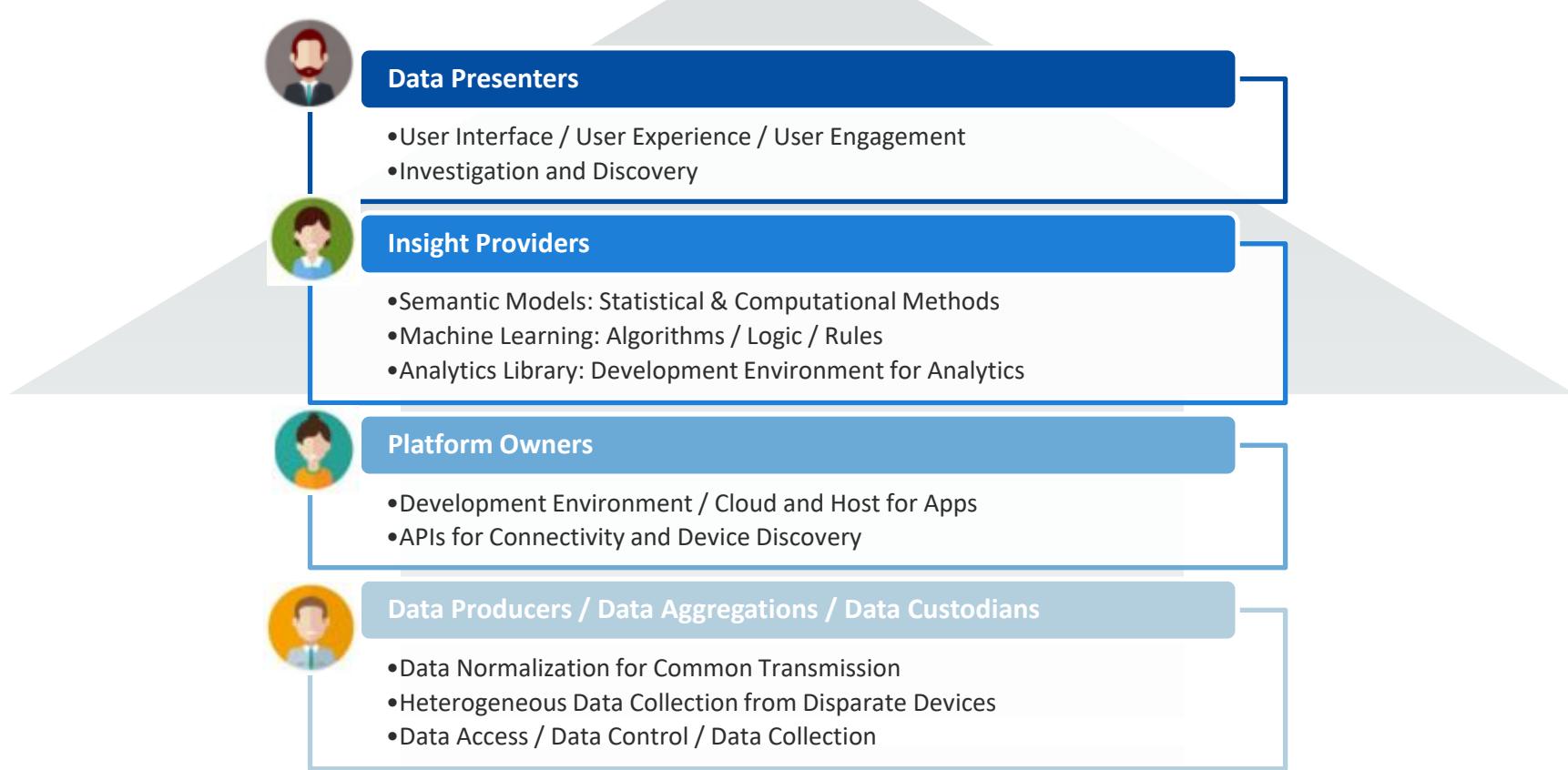
“The Data Science Workflow”



Driving Digital Transformation with Data Science as a Service

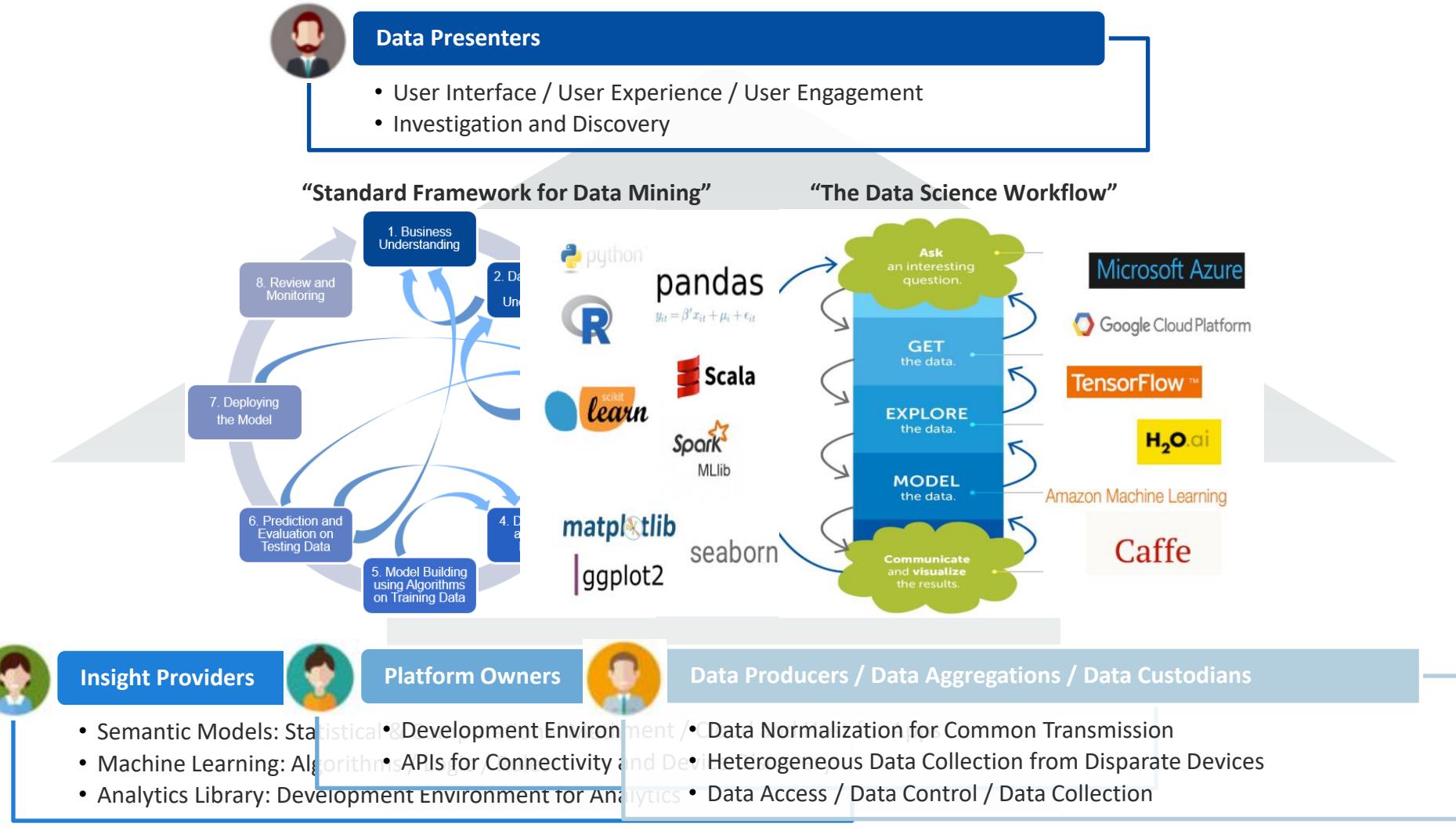
➤ The Insights Revolution?

- A **Data Economy** is a global digital ecosystem in which data is gathered, organized, and exchanged by a network of vendors for the purpose of deriving value from the accumulated information.



Driving Digital Transformation with Data Science as a Service

➤ The Insights Revolution?



Driving Digital Transformation with Data Science as a Service

➤ Data Science as a Service



Driving Digital Transformation with Data Science as a Service

➤ Data Science as a Service



Driving Digital Transformation with Data Science as a Service

➤ Data Science as a Service



Driving Digital Transformation with Data Science as a Service

➤ Data Science as a Service



Driving Digital Transformation with Data Science as a Service

➤ Data Science as a Service

NEW
RETAIL



THANK YOU

Q&A

Properties of Data: What is a Big Data?

➤ Output (Y): Labelled vs Unlabelled

Lets say we want to **Classify Houses by Size**

Given Features or Feature Set

A diagram illustrating supervised learning. At the top, the text "Given Features or Feature Set" is followed by a downward-pointing blue arrow. Below the arrow is a table with five columns: "FullBath", "HalfBath", "Bedrooms", "Home Age", and "Size". The "Size" column is highlighted with a red arrow pointing to it from the right, labeled "Label". The table contains four data rows:

FullBath	HalfBath	Bedrooms	Home Age	Size
1	0	2	56	M
1	1	3	59	L
2	1	3	20	M
2	1	3	19	S

Supervised Learning

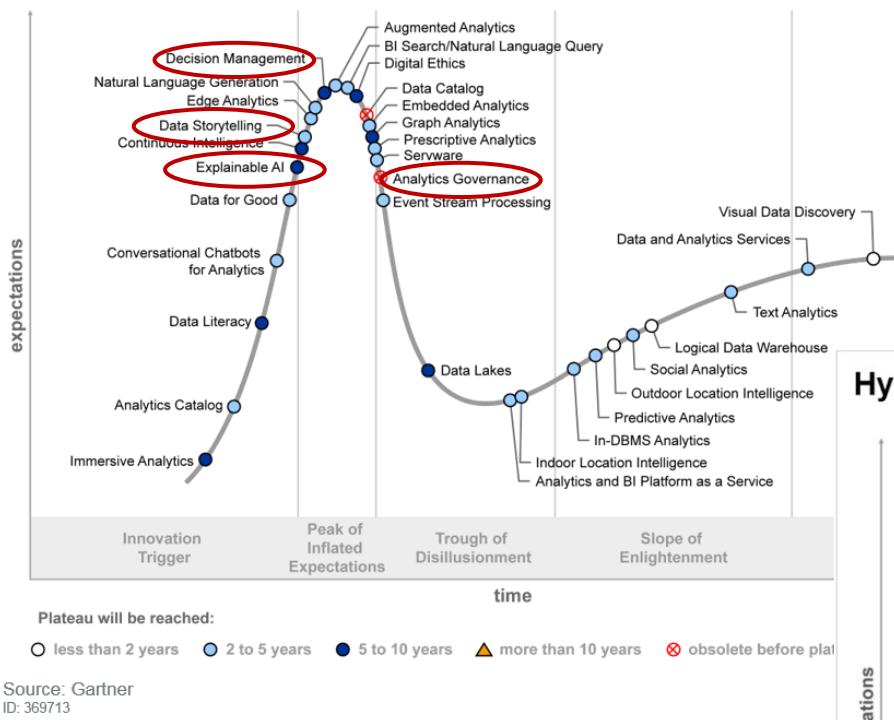
Use the labels to build a model. Model used to classify new house size based ONLY on the known feature set.

Unsupervised

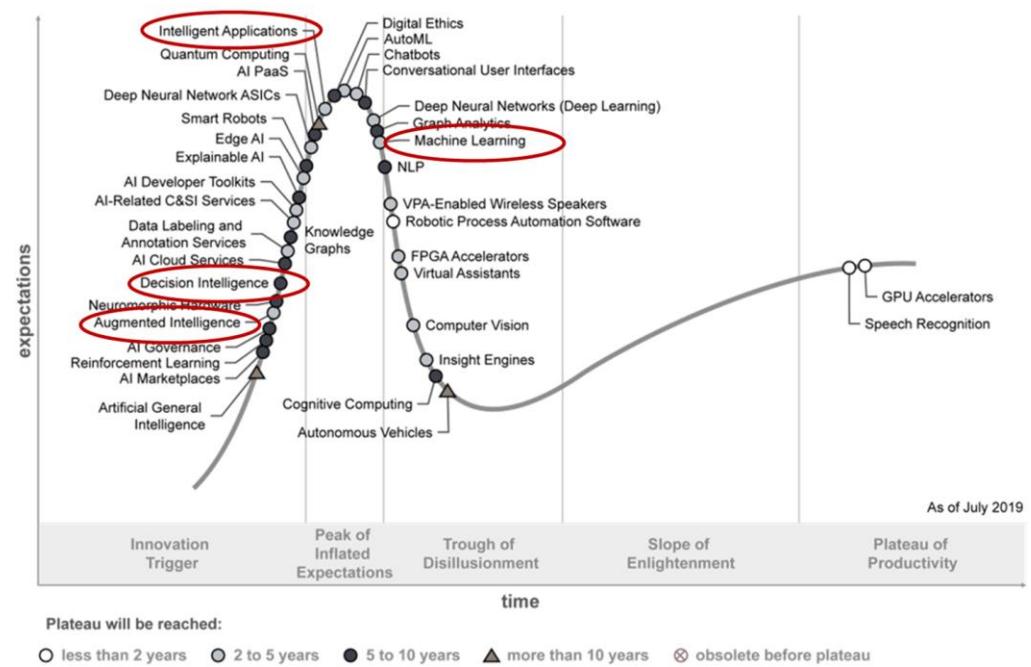
SIZE is missing! We need to look for similarities in the data and group them into clusters.

Global “Technology” Trends

Hype Cycle for Analytics and Business Intelligence, 2019

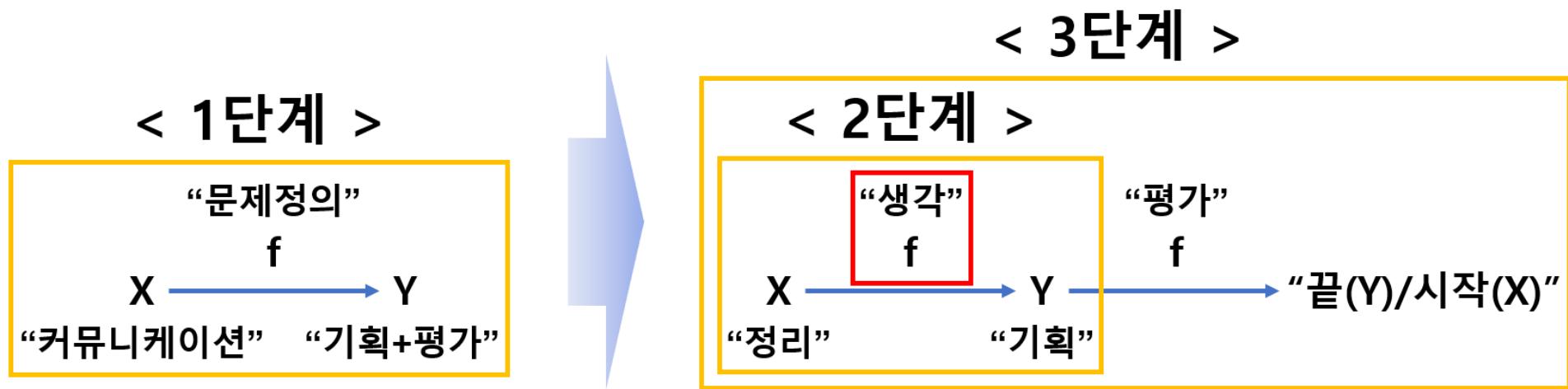


Hype Cycle for Artificial Intelligence, 2019



Data Science Process

➤ Getting from Raw Data to Outcomes



What is Data Science?

➤ The Real World Data Science is not a Kaggle Competition

- It can be worthwhile to step back a little and realize what exactly your **ultimate goal** is.
- The **best performance** might not be equivalent to a model yielding the **best score** in real.

The screenshot shows the Kaggle Competitions website. At the top, it says "Welcome to Kaggle Competitions" and "Challenge yourself with real-world machine learning problems". Below this, there are three main sections: "New to Data Science?", "Build a Model", and "InClass Prediction Competition". The "InClass Prediction Competition" section is highlighted and shows the title "Housing Prices Competition for Kaggle Learn Users", a description "Apply what you learned in the Machine Learning course on Kaggle Learn alongside others in the course.", the number of teams "13,432 teams · 4 months to go", and navigation links for "Overview", "Data", "Notebooks", "Discussion", "Leaderboard", and "Rules". A "Submit Predictions" button is also visible. At the bottom, there are sections for "Description", "Evaluation", and "Frequently Asked Questions", along with a question "What is a Getting Started competition?" and its answer: "Getting Started competitions were created by Kaggle data scientists for people who have little to no machine learning background. They are a great place to begin if you are new to data science or just finished a MOOC and want to get involved in Kaggle."

Welcome to Kaggle Competitions

Challenge yourself with real-world machine learning problems

New to Data Science?

Get started with a tutorial on our most popular competition for beginners, [Titanic: Machine Learning from Disaster](#).

Build a Model

Get the data & use whatever tools or methods you prefer to make predictions.

InClass Prediction Competition

Housing Prices Competition for Kaggle Learn Users

Apply what you learned in the Machine Learning course on Kaggle Learn alongside others in the course.

13,432 teams · 4 months to go

Overview Data Notebooks Discussion Leaderboard Rules Submit Predictions

Overview

Description

Evaluation

Frequently Asked Questions

What is a Getting Started competition?

Getting Started competitions were created by Kaggle data scientists for people who have little to no machine learning background. They are a great place to begin if you are new to data science or just finished a MOOC and want to get involved in Kaggle.

Who is a Data Scientist?

➤ What is a Project of Data Science?



▪ TASKS

: In addition to advanced analytic skills, this individuals are also proficient at integrating and preparing large, varied datasets, architecting specialized database and computing environments, and communicating results.

▪ MISSION

: A data scientist may or may not have specialized industry knowledge to aid in modeling business problems and with understanding and preparing data.

▪ TALENT

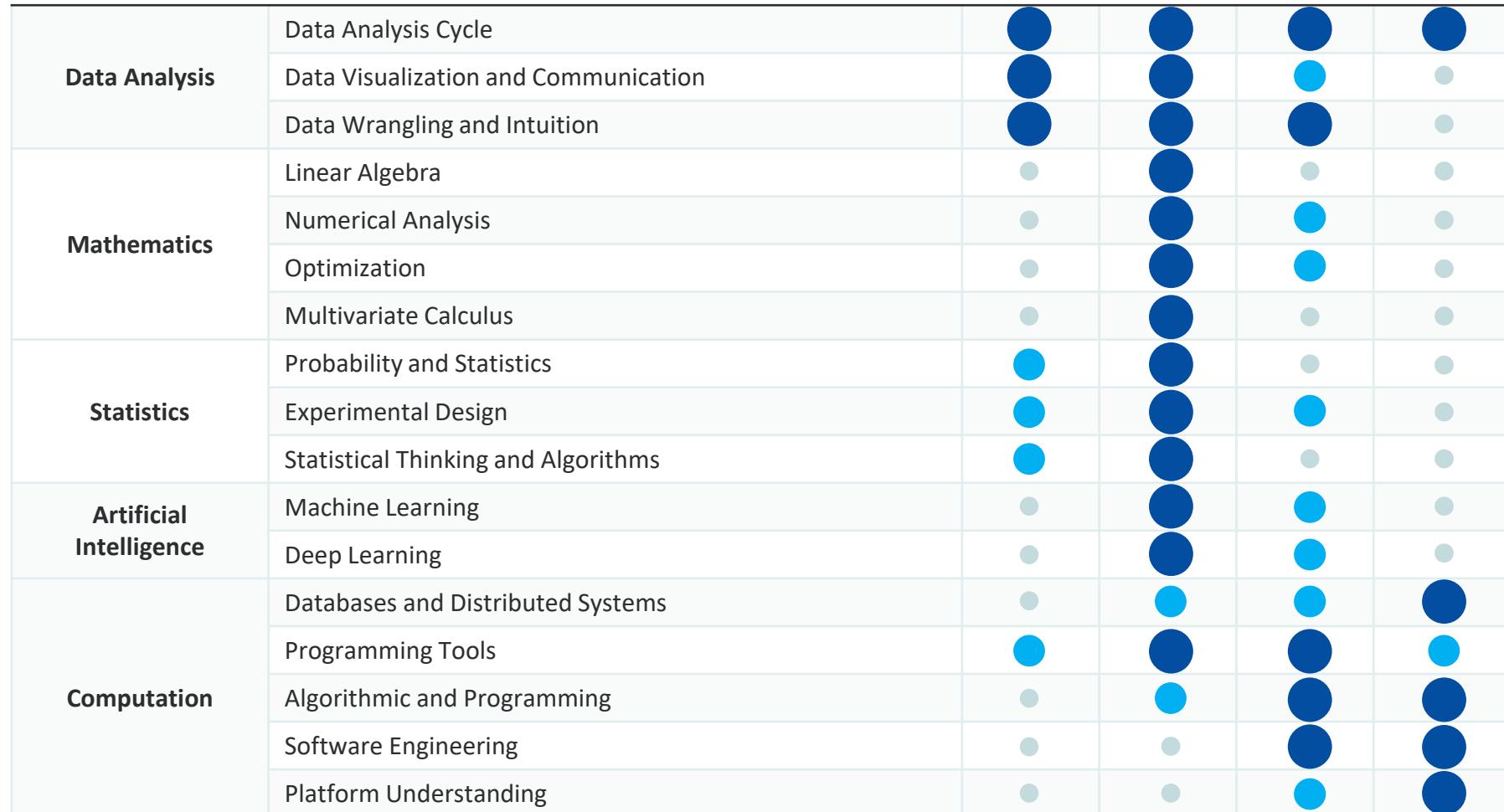
: Creating value from data requires a range of talents from data integration and preparation, to architecting specialized computing/database environments, to data mining and intelligent algorithms.

▪ RESPONSIBILITY

: An individual responsible for modeling complex business problems, discovering business insights and identifying opportunities through the use of statistical, algorithmic, mining and visualization techniques.

Data Science Roadmap

➤ We can be the best Data Science team



● Not that important

● Somewhat important

● Very important

Data Science Curriculum

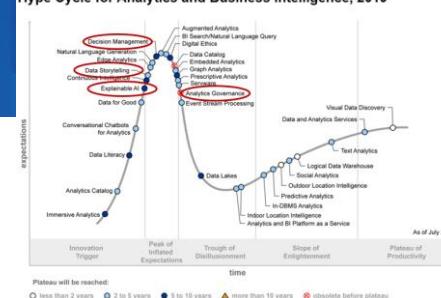
➤ Related 44 Lectures (132 Credits)

[초급] [중급] [고급]

Data Analysis (6)	Mathematics (6)	Statistics (6)	Artificial Intelligence (6)	Computation (10)	Professionalism (10)
<ul style="list-style-type: none">데이터사이언스 개론데이터분석 언어데이터모델링데이터시각화데이터분석 응용데이터사이언스 응용 및 활용	<ul style="list-style-type: none">미적분학 기초선형대수 기초수치해석 기초다면량 미적분학인공지능 수치해석최적화	<ul style="list-style-type: none">확률 및 통계 기초실험계획법통계적 사고 및 알고리즘다면량 자료분석시계열분석시뮬레이션	<ul style="list-style-type: none">데이터마이닝 개론기계학습딥러닝인공지능자연어처리시각처리	<ul style="list-style-type: none">데이터베이스 시스템자료구조 및 알고리즘서버시스템 및 클라우드빅데이터 분산처리소프트웨어 공학웹마이닝미디어콘텐츠 처리컴퓨터비전정보보안 특론빅데이터 플랫폼 특론	<ul style="list-style-type: none">CommunicationTeamworkEthical IssuesPrivacy and ConfidentialityLegal ConsiderationsEconomic ConsiderationsIntellectual PropertyChange ManagementContinuing Professional DevelopmentOn Automation

Who is a Data Scientist?

➤ The Insights Revolution?



Data Presenters

- User Interface / User Experience / User Engagement
- Investigation and Discovery



Insight Providers

- Semantic Models: Statistical
- Machine Learning: Algorithms
- Analytics Library: Development Environment for Analytics
- Development Environment / APIs for Connectivity and Dev
- Data Normalization for Common Transmission
- Heterogeneous Data Collection from Disparate Devices
- Data Access / Data Control / Data Collection

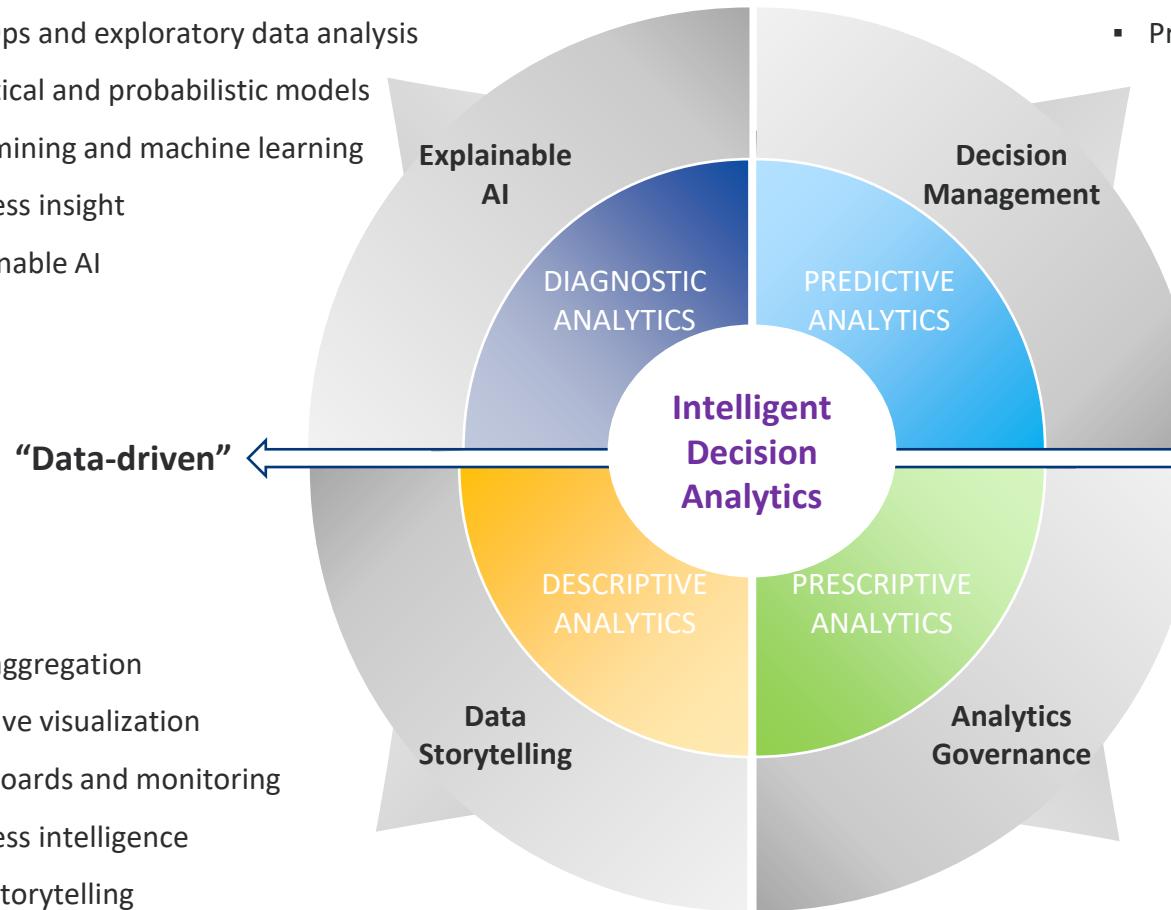
Platform Owners

Data Producers / Data Aggregations / Data Custodians

Who is a Data Scientist?

➤ Why DO it happen?

- DataOps and exploratory data analysis
- Statistical and probabilistic models
- Data mining and machine learning
- Business insight
- Explainable AI



➤ What COULD happen?

- Predictive modeling and Forecasting
 - Automation and warning
 - Business scoring
- Simulation and Expert systems
 - Decision management

➤ What SHOULD happened?

- Recommended actions
- Actionable Intelligence
 - Strategy optimization
 - Business rules
- Analytics governance

➤ What HAS happened?

- Data aggregation
- Effective visualization
- Dashboards and monitoring
- Business intelligence
- Data storytelling