

01

빅데이터의 시대

BIG

DATA

1. 빅데이터의 시대



가. 빅데이터 등장배경

- ICT (Information & Communication Technology) 인프라의 고도화 및 성숙에 따른 새로운 투자처에 대한 니즈 증가로 새로운 공급 체계와 시장을 창출 기대
- 기업은 축적된 자산 데이터 (Asset Data)와 소셜 네트워크에 기반한 대규모 외부 데이터의 활용으로 혁신과 기회 포착

1. 빅데이터의 시대



가. 빅데이터 등장배경

21세기 원유로서 중요한 ICT 트렌드 중 하나인 **빅데이터**



IDC(International
Data Corporation)

향후 빅데이터 시장과 빅데이터에 대한
수요는 급증할 것이라고 예측

전세계 총 데이터량은
2013년에는 4.4조 GB이 생산되었고
2020년에 44조 GB에 달할 것이라고 전망



EMC & IDC(2014)

1. 빅데이터의 시대



가. 빅데이터 등장배경

- 최근 10년 사이 인류는 디지털 데이터가 폭증하는 **데이터 홍수(Data Deluge) 현상에 직면**
 - » 2007년 전세계적으로 생성된 디지털 정보량이 사용가능한 저장 공간을 초과하기 시작하여, 2020년에는 관리해야 할 정보량이 현재보다 50배 급증하고 10배 많은 서버가 필요
 - » 기술발전에 따른 데이터 저장 및 처리 비용의 하락, 소셜 네트워크 서비스의 확대 등으로 막대한 데이터 폭발이 진행 중
- 앞으로 도로, 건축물 등에 내장된 **임베디드 시스템(Embedded System)**, **IoT(Internet of Things)**에서 막대한 데이터가 만들어질 것으로 전망

1. 빅데이터의 시대



가. 빅데이터 등장배경

빅데이터 (Big Data)

디지털정보량의 증가에 따라 대규모 데이터가
중대 이슈로 부각되며 등장한 용어로서,
기존의 관리 및 분석체계로는 감당할 수 없을 정도의
거대한 데이터의 집합을 지칭

1. 빅데이터의 시대



가. 빅데이터 등장배경

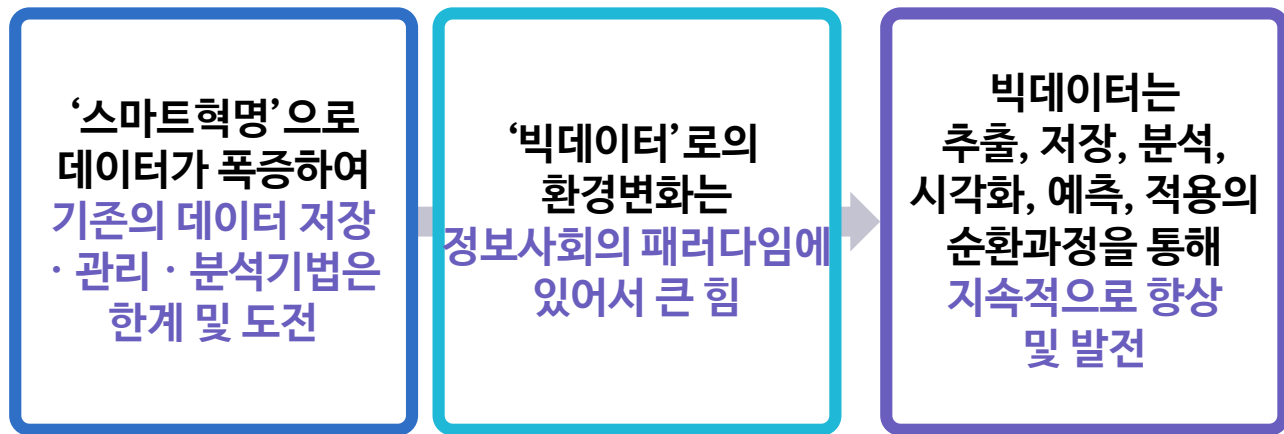
빅데이터 (Big Data)

- 과거 빅데이터는 천문·항공·우주정보·인간게놈 정보 등 특수 분야에 한정됐으나, 정보통신기술의 발달에 따라 전분야로 확산
- 대규모 데이터와 관계된 기술 및 도구(수집·저장·검색·공유·분석·시각화 등)도 포함

1. 빅데이터의 시대



가. 빅데이터 등장배경



1. 빅데이터의 시대



가. 빅데이터 등장배경

④ 정보사회의 패러다임의 변화

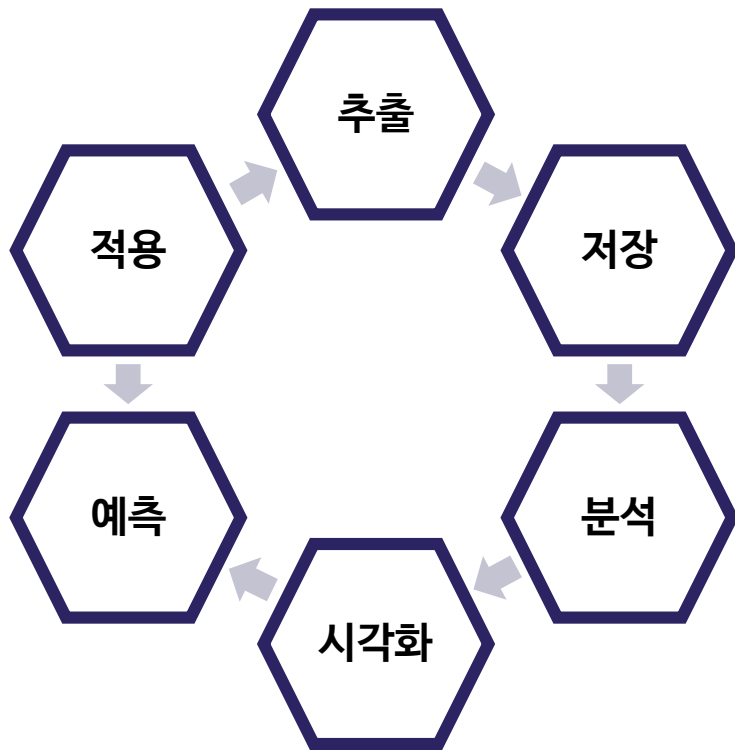
구분	PC 시대	인터넷시대	모바일시대	스마트시대
패러다임 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물정보화
IT이슈	PC, PC통신, 데이터베이스	초고속인터넷, WWW, 웹서버	모바일 인터넷, 스마트폰	빅데이터, 차세대PC, 사물네트워크 (M2M)
핵심분야 (서비스)	PC, OS	포털, 검색엔진, Web2.0	스마트폰, 앱서비스, SNS	미래전망, 상황인식, 개인맞춤형 서비스
대표기업	MS, IBM 등	구글, 네이버, 유튜브 등	애플, 페이스북, 트위터 등	?
IT비전	1인 1PC	클릭 e-Korea	손안의 PC, 소통	IT everywhere, 新가치창출

1. 빅데이터의 시대



가. 빅데이터 등장배경

④ 빅데이터 처리의 순환과정



1. 빅데이터의 시대



가. 빅데이터 등장배경

④ 정보화 시대 VS 스마트 시대 데이터 관련 이슈 변화

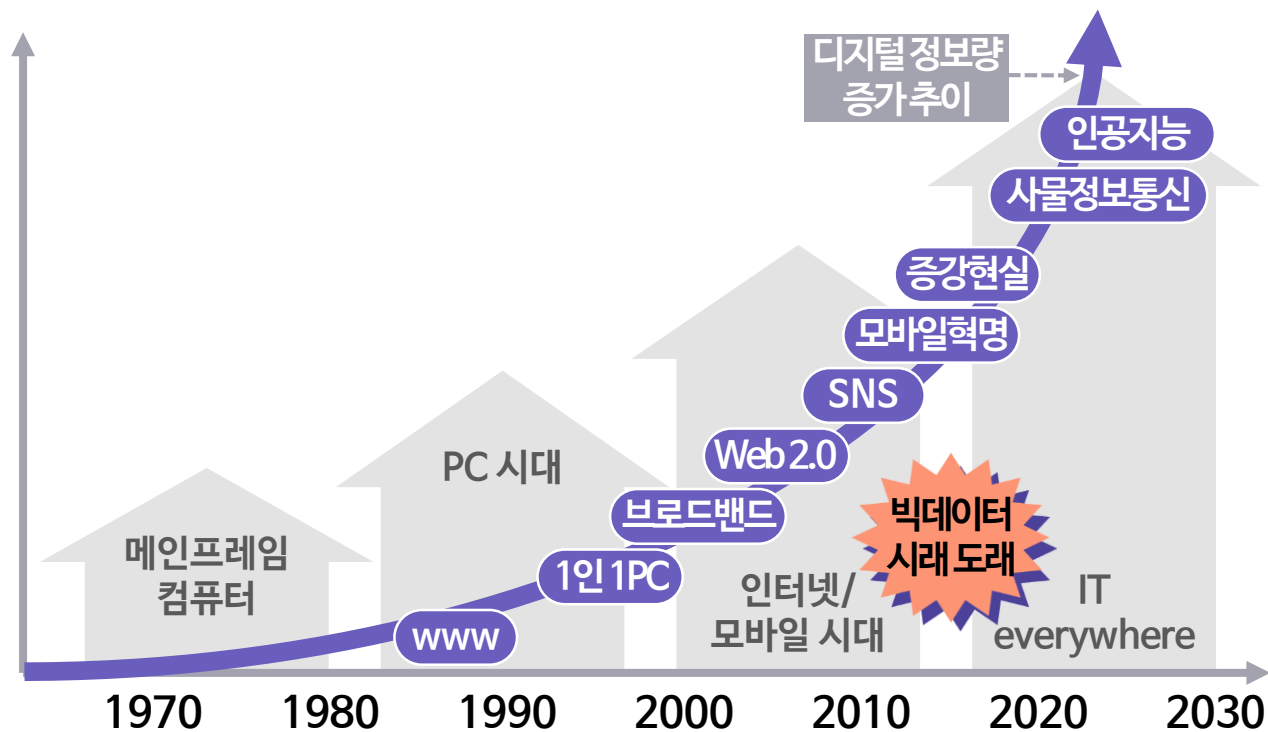
구분	정보화 시대 (1세대)	스마트 시대 (2세대)
저장	관계형/정형 데이터베이스, 데이터웨어하우스	비관계형/비정형 데이터베이스, 가상화, 클라우드 서비스
검색	검색엔진(Text), 포털 서비스	자연어/음성·영상/시멘틱 검색서비스
관리 /공유	KWS, Web2.0	플랫폼, 소셜 네트워크, 집단지성
분석	경영정보/고객정보 /자산정보 분석(ERP, CRM, 데이터마이닝 등)	빅데이터 분석(소셜 분석, 고급 분석, BI, 시각화)
추론	-	상황인식 서비스(미래전망, 사전대응, 자동화 서비스), 개인화 서비스

1. 빅데이터의 시대



가. 빅데이터 등장배경

④ ICT 발전에 따른 데이터의 변화 방향



1. 빅데이터의 시대



가. 빅데이터 등장배경

④ ICT 발전에 따른 데이터의 변화 방향



1. 빅데이터의 시대



가. 빅데이터 등장배경

④ ICT 발전에 따른 데이터의 변화 방향

- 전세계 정보량은 증가하는 추이를 보임

≫ 2013년 전세계 디지털 정보량은 약 **1.82 ZB**

※ 1.8 ZB = 1.8조 GB (2000억 개 이상의 고화질 (HD) 영화를
4700만 년 동안 시청할 수 있는 정도의 정보량)

≫ 2020년에 관리해야 할 데이터의 양은 50배 이상 증가
(IDC&EMC, Digital Universe Study, 2011)

1. 빅데이터의 시대



나. 빅데이터 시장동향

IDC(Internet
Data Center)

빅데이터 기술과 서비스 시장이
2019년까지 연평균 23.1% 성장해
486억 달러에 이를 것으로 전망

2015년 1,220억 달러 규모였던
빅데이터와 애널리틱스 분야의 매출이
2019년 1,870억 달러에
이를 것으로 전망

전세계 반년도 빅데이터 및
애널리틱스 지출 가이드

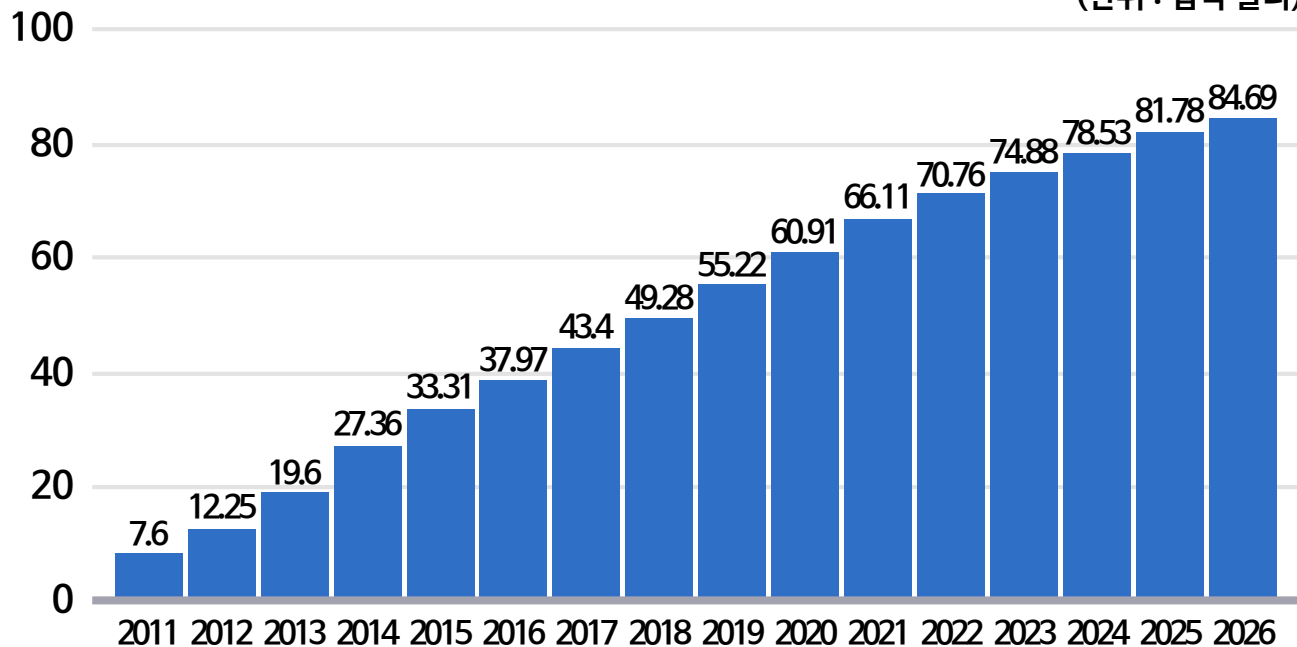
1. 빅데이터의 시대



나. 빅데이터 시장동향

④ 세계 빅데이터 시장 동향 및 전망

(단위 : 십억 달러)



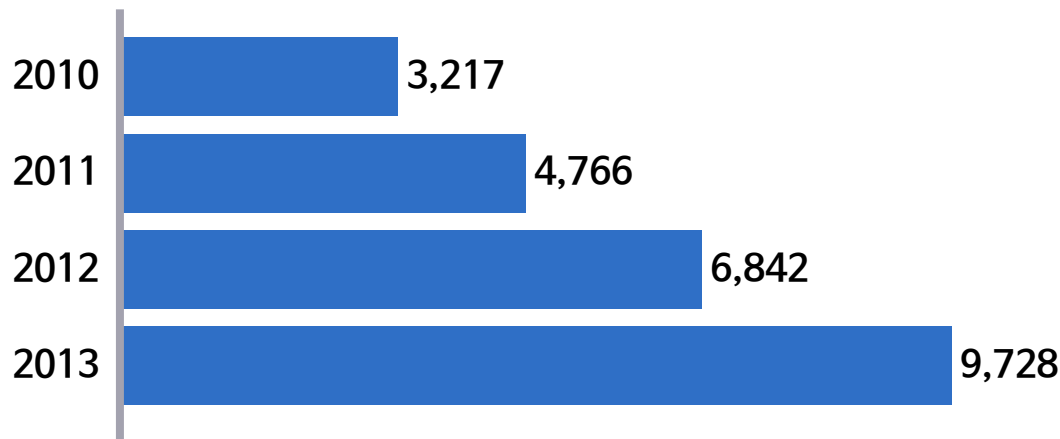
〈출처 : Statista, Forecast of Big Data market size, based on revenue, from 2011 to 2026, 2016 Wikibon, Big Data Market Forecast, 2011-2026, 2016. 재편집〉

1. 빅데이터의 시대



나. 빅데이터 시장동향

④ 산업별 빅데이터 관련 투자현황 및 계획



현재 교육, 운송, 에너지/유틸리티 등에 빅데이터 관련 투자가 많지만, 향후에는 소매, 보험, 통신 및 미디어 등 산업에 빅데이터 관련 투자계획이 많을 것으로 예상

1. 빅데이터의 시대



나. 빅데이터 시장동향

- 2011년 맥킨지의 보고서에 따르면, 산업별로 빅데이터를 활용한 효과가 극대화될 것으로 예상

미국 의료 산업에서
빅데이터를 활용할 경우



매년 3,300억 달러의 가치 창출
(연간 스페인 전체 의료 지출액의 2배)

유럽 공공 분야에서
빅데이터를 활용할 경우



2,500억 유로의 절감 효과
(그리스 GDP와 같은 규모)

1. 빅데이터의 시대



다. 국가별 빅데이터 동향

- 2009년 UN은 빅데이터를 이용하여 보다 나은 세계로 발전하길 기대하며 **글로벌 펄스(Global Pulse)**를 출범
- 2012년 다보스 포럼에서는 새로운 가능성을 내포한 주제 중의 하나로 **빅데이터를 선정해 토의**
- **국가별로 빅데이터의 역량을 강화하기 위해 노력**
 - ≫ 미국은 2012년 3월 ‘빅데이터 연구개발 이니셔티브’를 통해 6개 정부 부처를 중심으로 총 2억 달러를 투자
 - ≫ 우리나라에서는 **정부3.0 정책**을 발표

공공정보를 적극 개방·공유하고, 부처간 칸막이를 없애고 소통 및 협력하여 국정과제에 대한 추진동력을 확보하고 국민맞춤형 서비스를 제공함과 동시에 일자리 창출과 창조경제를 지원

02

빅데이터의 개념



2. 빅데이터의 개념



가. 빅데이터에 대한 다양한 정의



WIKIPEDIA
The Free Encyclopedia

Wikipedia

빅데이터란 ‘기존 데이터베이스
관리도구의 데이터 수집 · 저장 · 관리
· 분석의 역량을 넘어서는
대량의 정형 또는 비정형 데이터세트
및 이러한 데이터로부터
가치를 추출하고
결과를 분석하는 기술’

2. 빅데이터의 개념



가. 빅데이터에 대한 다양한 정의

국가전략위원회

빅데이터란 ‘대용량 데이터를 활용
· 분석하여 가치있는 정보를 추출하고
생성된 지식을 바탕으로 능동적인 대응
또는 변화 예측을 위한 정보화 기술’

빅데이터란 ‘기존 관리 및 분석체계로는
감당 불가능한 거대한 데이터 집합으로
대규모 데이터와 관계된 기술 및 도구를
모두 포함하는 개념’

SERI.org
Knowledge Hub for Creative Solutions

삼성경제연구소

2. 빅데이터의 개념



가. 빅데이터에 대한 다양한 정의

- 빅데이터란 엄청나게 많은 데이터이지만, 양적인 의미를 벗어나 데이터 분석과 활용을 포괄하는 개념으로 사용
- 빅데이터의 정의는 데이터 규모와 기술 측면에서 출발했으나 빅데이터의 가치와 활용효과 측면으로 의미가 확대되는 추세

2. 빅데이터의 개념



나. 빅데이터의 구분

- 빅데이터는 정형화(Structured) 데이터, 반정형화(Semi-structured) 데이터, 비정형(Unstructured) 데이터로 구분

정형화 데이터

- » 일정한 규칙을 갖고 체계적으로 정리된 데이터
- » 매년 통계청에서 발표하는 통계자료, 방송통신 실태조사, 각종 과학적 데이터 등이 해당
- » 그 자체로 의미 해석 가능하며, 바로 활용할 수 있는 정보를 내포

반정형화 데이터

- » 한글글, 마이크로소프트 워드 등의 데이터
- » 인쇄매체의 텍스트 등이 해당
- » 표나 그림이 될 수도 있지만 일반적으로 문자로 서술된 정보

2. 빅데이터의 개념



나. 빅데이터의 구분

- 빅데이터는 정형화(Structured) 데이터, 반정형화(Semi-structured) 데이터, 비정형(Unstructured) 데이터로 구분

비정형화 데이터

- » 스마트기기 등을 통해 형성되는 데이터
- » 페이스북, 트위터, 카카오톡, 라인 등으로 상호 교류되는 정보가 해당
- » 개인, 집단, 사회, 국가 등과 관련된 주제를 스마트 미디어 이용자들이 상호 의견을 교류함으로써 생산
- » 오늘날 빅데이터 중 관심 받는 데이터 형태

2. 빅데이터의 개념



다. 빅데이터의 플랫폼

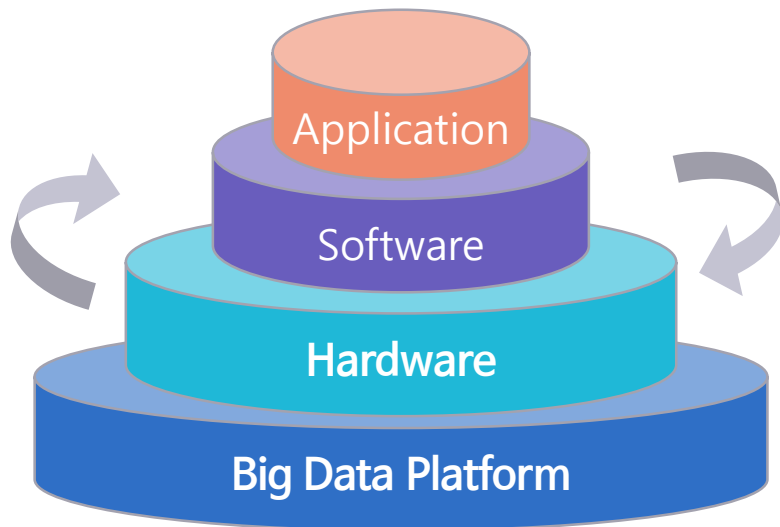
- 빅데이터는 해당 데이터를 분석하고 처리함으로써 기존의 데이터에서 볼 수 없었던 새로운 의미를 산출
 - » 새로운 가치와 의미를 산출하기 위해 축적된 데이터를 갖고 무엇을 분석할 것인가에 대한 **문제제기**가 필요

문제제기는 마이닝과 연결되는데, 정형데이터마이닝(Data Mining), 텍스트마이닝(Text Mining)과 웹마이닝(Web Mining), 소셜마이닝(Social Mining)을 통해 현실마이닝(Reality Mining)에 도달

- » 빅데이터의 현실마이닝은 영화에 나올 법한 미래를 예측할 수 있는 데이터들을 산출한 사후대책이 아니라 사전 방지 시스템을 작동하는 것

2. 빅데이터의 개념

다. 빅데이터의 플랫폼



빅데이터는 빅데이터를 구성하고 있는
하드웨어, 소프트웨어, 어플리케이션 간의
유기적 순환에 의해 가치 창출이 가능

03

빅데이터의 의미

BIG

DATA

3. 빅데이터의 의의



가. 빅데이터의 이슈화 요인

- 1 모바일 스마트 기기 보급이 활성화 되면서 센서, 원격감지기술, 소프트웨어 등을 통해 비정형화 데이터 수집이 용이
- 2 클라우드 서비스를 통해 개인과 조직의 데이터가 한 곳으로 축적되고, 저장된 데이터를 분석하여 활용하고자 하는 요구 증가
- 3 소셜 미디어 활용이 일상화되면서 쌍방향 커뮤니케이션이 활발해지면서 정보유통구조가 새롭게 재편

3. 빅데이터의 의의

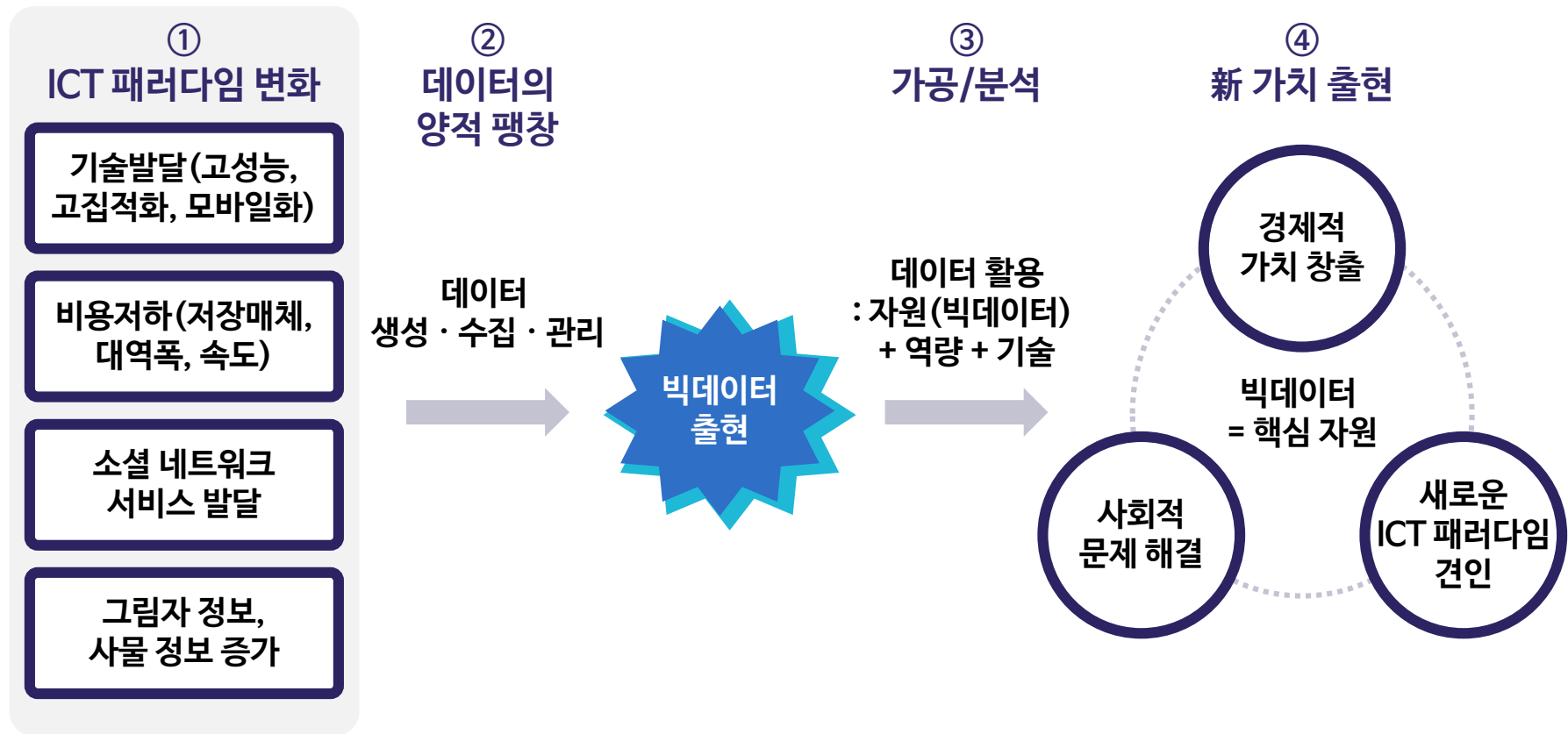


나. 빅데이터의 출현과 新 가치 창출

- ICT 패러다임의 변화는 데이터의 양적 팽창을 가져왔으며
데이터 범람이 새로운 기회와 편익을 창출하기 시작
- 빅데이터는 잠재적 가치와 잠재적 위험이 공존하며
사회·경제적으로 성패를 좌우하는 핵심 원천이 될 것으로 평가

3. 빅데이터의 의의

나. 빅데이터의 출현과 新 가치 창출



04

빅데이터의 특징

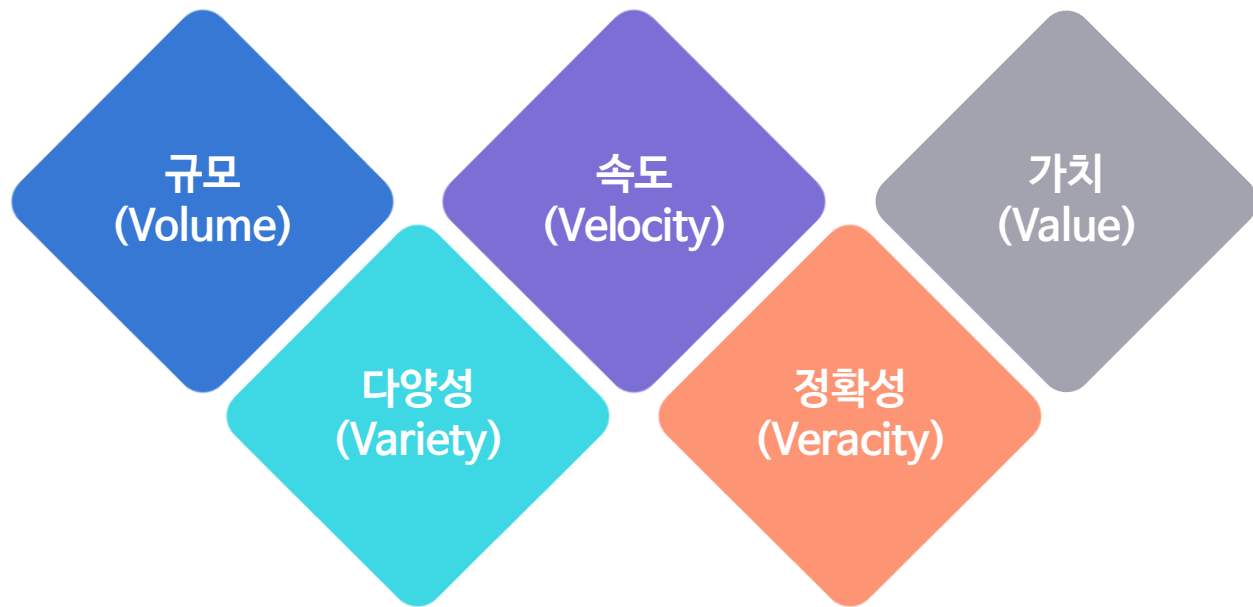
BIG

DATA

4. 빅데이터의 특징



가. 빅데이터의 구성요소



특정 규모(Big Volume) 이상을 넘어서,
빅데이터에서 원하는 가치(Big Value)를
얻을 수 있는 정도의 상대적인 해석이 필요

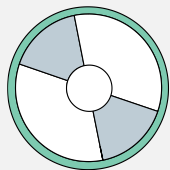
4. 빅데이터의 특징



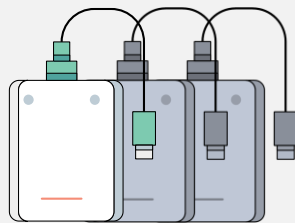
가. 빅데이터의 구성요소

01 규모(Volume)

- 기술적인 발전과 ICT의 일상화가 진행되면서
해마다 디지털 정보량이 기하급수적으로 폭증
≫ ZB(제타바이트) 시대로 진입



MB, GB 단위



TB, PB, EB, ZB 단위

4. 빅데이터의 특징

가. 빅데이터의 구성요소

02 다양성 (Variety)

- 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터 종류의 증가
- 멀티미디어 등 비정형화 데이터 유형의 다양화



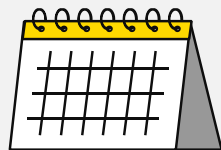
4. 빅데이터의 특징



가. 빅데이터의 구성요소

03 속도(Velocity)

- 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간성 정보 증가
- 실시간성으로 인한 데이터 생성, 이동(유통) 속도의 증가
- 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터를 처리 및 분석 속도가 중요



유통활용주기
수 시간~주



분, 초 단위
유통 활용

4. 빅데이터의 특징



가. 빅데이터의 구성요소

04 정확성 (Veracity)

- 빅데이터의 특성상 방대한 데이터들을 기반으로 분석 수행
- 데이터 분석에서 질 높은 데이터를 활용하는 것이 분석의 정확도에 영향

4. 빅데이터의 특징



가. 빅데이터의 구성요소

05 가치(Value)

- 빅데이터가 추구하는 것은 가치 창출
- 빅데이터 분석을 통해 도출된 최종 결과물은 기업이 현재 당면하고 있는 문제를 해결하는 데에 통찰력 있는 유용한 정보를 제공

4. 빅데이터의 특징



나. 데이터 처리 방식의 차이점

	기존 데이터 처리방식	빅데이터 처리방식
데이터 트래픽	» TB(테라바이트) 수준	» PB(페타바이트) 수준(최소 100 TB 이상) » 장기간의 정보수집 및 분석 » 방대한 데이터 처리량
데이터 유형	» 정형데이터 중심	» 비정형 데이터의 높은 비중(SNS 데이터 로그파일, 클릭스트림 데이터 등) » 처리의 복잡성 증대
프로세스 및 기술	» 단순한 프로세스 및 기술 » 정형화된 처리/ 분석결과 » 원인/결과규모 중심	» 다양한 데이터소스 및 복잡한 로직처리 » 데이터처리 복잡도가 높아 분산처리 기술 필요 » 새롭고 다양한 처리방법 개발 필요(정의된 데이터 모델/상관관계/절차 등이 없음) » 상관관계 규명 중심 » Hadoop, R, NoSQL 등 개방형 소프트웨어