

빅데이터의 이해와 활용

Understanding and Using Big Data

01

빅데이터 개요

학습 내용

- 01 데이터와 정보의 이해
- 02 빅데이터 정의 및 활용
- 03 빅데이터 처리
- 04 빅데이터 미래

학습 목표

- 데이터와 정보의 개념과 특성을 설명할 수 있다.
- 빅데이터 정의 및 활용 방안에 대해 설명할 수 있다.
- 데이터 처리 기술에 대해 설명할 수 있다.
- 빅데이터 발전 미래에 대해 설명할 수 있다.

생각
해보기

빅데이터의 최신 트렌드는?

**BIG
DATA**





01

데이터와 정보의 이해

- 1) 데이터
- 2) DBMS

1) 데이터

● 데이터란?

데이터

1646년 영국 문헌에 처음 등장하였으며, 라틴어인 Dare(주다)의 과거 분사형으로 '주어진 것'이란 의미

01 과거의 관념적이고 추상적인 개념에서 기술적이고 사실적인 의미로 변화

02 추론과 추정의 근거를 이루는 사실

03 다른 객체와의 상호관계 속에서 가치를 갖는 것

1) 데이터

● 데이터의 유형

정성적 데이터

- 언어, 문자 등
- 저장, 검색, 분석에 많은 비용이 소요됨

예) 닳은 얼굴 구분하기

↳ 주관적이며,
통계분석이 어려움

정량적 데이터

- 수치, 도형, 기호 등
- 정량화가 된 데이터로 비용 소모가 적음

예) 키, 나이, 인구 수,
생산성 지표 등

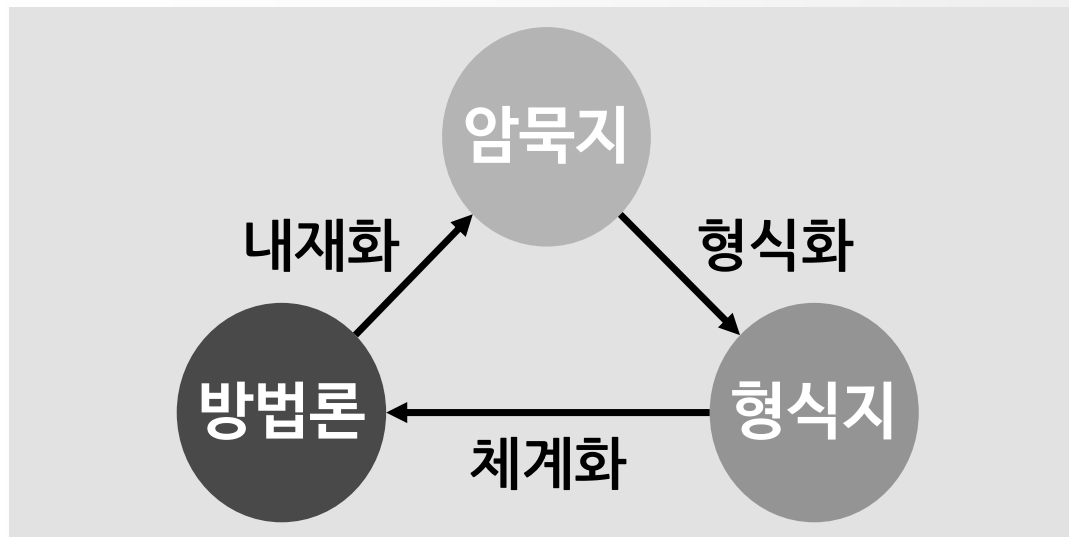
↳ 객관적이며,
통계분석이 용이함

1) 데이터

● 지식경영과 암묵지/형식지

데이터

- 지식경영의 핵심 이슈인 암묵지(Tacit Knowledge)와 형식지(Explicit Knowledge)의 상호작용에 있어 중요한 역할을 함



1) 데이터

● 지식경영과 암묵지/형식지

암묵지

- 학습과 경험(체험)을 통해 개인에게 체계화되어 있지만 겉으로 드러나지 않는 지식
- 암묵지는 현장작업과 같은 작업을 통해 획득할 수 있는 지식(Kim, 1997; Nonaka, 1995)

예

과학적 발견

1) 데이터

● 지식경영과 암묵지/형식지

암묵지

- 상호 작용
 - 공통화, 내면화

예 스킬, 노하우,
판단/행동, 기준,
가치체계 등

형식지

- 상호 작용
 - 표출화, 연결화

예 규정, 업무 처리 절차,
매뉴얼, 소프트웨어,
데이터베이스 등

2) 데이터

- 지식경영과 암묵지/형식지

데이터 (Data)	개별 데이터 자체로는 의미가 중요하지 않은 객관적인 사실
정보 (Information)	데이터의 가공, 처리와 데이터간 연관관계 속에서 의미가 도출된 것
지식 (Knowledge)	데이터를 통해 도출된 다양한 정보를 구조화하여 유의미한 정보를 분류하고 개인적인 경험을 결합시켜 고유의 지식으로 내재화된 것
지혜 (Wisdom)	지식의 축적과 아이디어가 결합된 창의적인 산물

2) DBMS

- DBMS의 정의

DBMS

문자, 기호, 음성, 화상, 영상 등 상호 관련된 다수의 콘텐츠를
정보 처리 및 정보통신 기기에 의하여 체계적으로 수집,
축적하여 다양한 용도와 방법으로 이용할 수 있도록 정리한
정보의 집합체

2) DBMS

● DBMS의 특징



실시간 접근성



다수의 사용자 또는 프로그램의 요구에
즉시 응답하거나 처리할 수 있음



지속적인 변화



데이터의 삽입(Insert), 수정(Update),
삭제>Delete) 등을 통해 데이터의 내용을
최신의 내용으로 유지할 수 있음

2) DBMS

● DBMS의 특징



동시 공유



다수의 사용자가 서로 다른 목적으로 동시에 동일한 데이터를 이용할 수 있음(Transaction, Lock Mechanism)



내용 참조



주소나 특정위치가 아닌 데이터의 내용을 참조할 수 있음

2) DBMS

- DBMS의 특징

통합된 데이터

저장된 데이터

공용 데이터

지속적인
변경 데이터

2) DBMS

- DBMS 자료구조

원자성
(Atomicity)

일관성
(Consistency)

격리성
(Isolation)

내구성
(Durability)

2) DBMS

● DBMS 비교(트랜잭션 관점)

구분	OLTP (On-line Transaction Processing)	OLAP (On-line Analytic Processing)
데이터 구조	복잡	단순
데이터 갱신	동적으로 순간적	정적으로 주기적
응답 시간	수 초 이내	수 초~수 분
데이터 범위	수 십일 전후	오랜 기간 저장
데이터 성격	정규적인 핵심 데이터	비정규적이며, 읽기 전용

2) DBMS

● DBMS 비교(트랜잭션 관점)

구분	OLTP (On-line Transaction Processing)	OLAP (On-line Analytic Processing)
데이터 크기	수 기가 바이트	수 테라 바이트
데이터 내용	현재 데이터	요약된 데이터
데이터 특성	트랜잭션 중심	주제 중심
데이터 액세스 빈도	높음	보통
질의 결과 예측	주기적이며 예측 가능	예측하기 어려움



02

빅데이터 정의와 활용

- 1) 빅데이터 정의
- 2) 빅데이터 특징
- 3) 빅데이터 활용

1) 빅데이터 정의

- 빅데이터 등장 배경

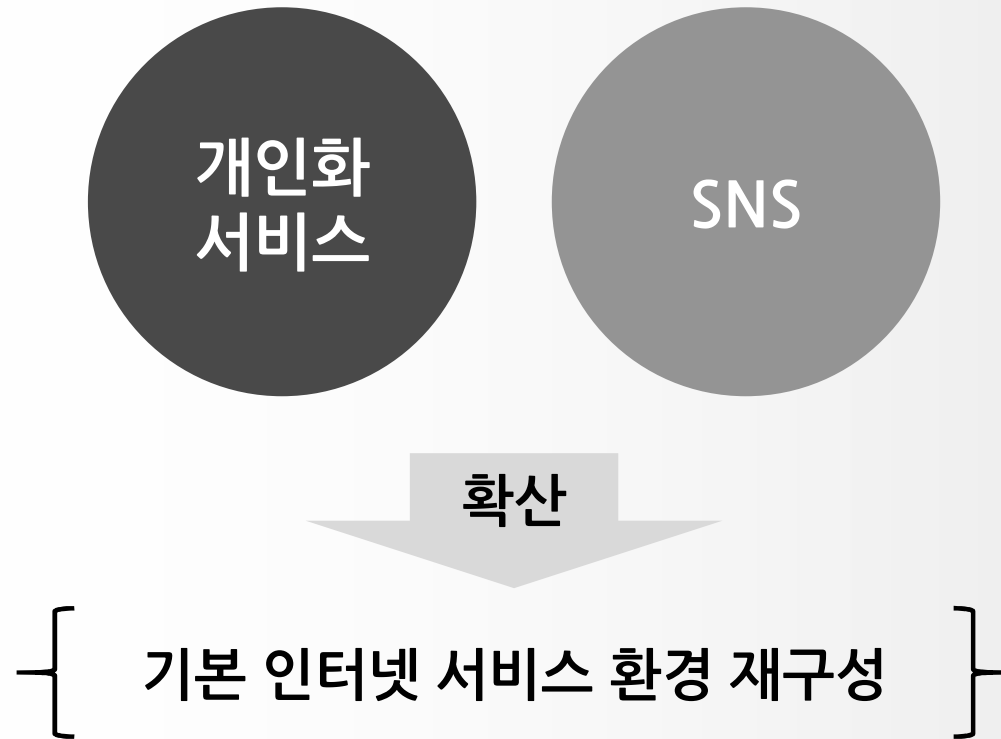
BIG DATA

1990년 이후 인터넷이 확산되면서
정형화된 데이터와 비정형화된 데이터가 무수히 발생

정보 홍수 개념이 등장하고
오늘날 빅데이터 개념으로 이어짐

1) 빅데이터 정의

- 빅데이터 등장 배경

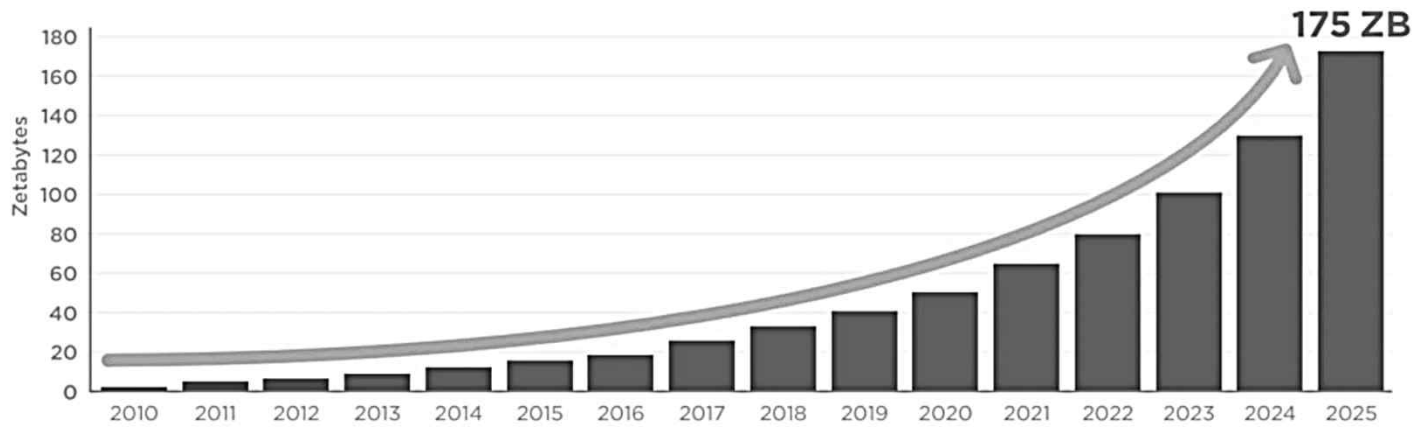


1) 빅데이터 정의

● 빅데이터 등장 배경

전 세계 디지털 데이터 양이 증가하여 2025년이 되면
한해 동안 생성되는 데이터 총량이 현재의 10배인
175ZB가 될 것이라고 함

연도별 전 세계 데이터 생산량



〈출처 : Data Age 2025, IDC〉

1) 빅데이터 정의

● 빅데이터 등장 배경

데이터 양 증가



기존의 데이터 저장, 관리,
분석기법으로 데이터 처리에
한계



정보 기술의 패러다임 변화



빅데이터 등장

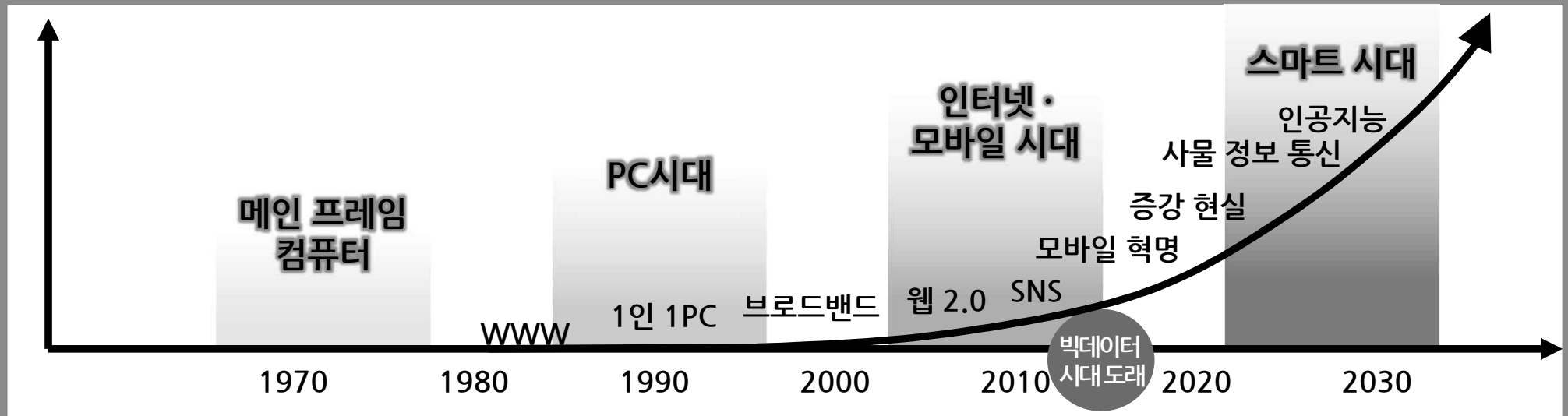


1) 빅데이터 정의

● 빅데이터 등장 배경

	PC 시대	인터넷 시대	모바일 시대	스마트 시대
패러다임의 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물정보화
정보기술 이슈	PC, PC 통신, 데이터베이스	초고속 인터넷, www, 웹 서버	모바일 인터넷, 스마트폰	빅데이터, 차세대 PC, M2M
핵심분야 (서비스)	PC, OS	포털, 검색엔진, Web 2.0	스마트폰, 웹 서비스, SNS	미래 전망, 상황인식, 개인화 서비스
대표 기업	MS, IBM	구글, 네이버, 유튜브	애플, 페이스북, 트위터	구글, 삼성, 애플, 페이스북,
정보 기술비전	1인 1PC	클릭 e-korea	손 안의 PC, 소통	IT everywhere

▶ 정보 통신 기술 발전에 따른 빅데이터 변화 방향



데이터 규모	EB(엑사 바이트) (1990년대 말=100EB)	ZB(제타바이트) 진입 (2011년대 말 =1.8ZB)	ZB 본격화 시대 (2020년=2011년 대비 50배 증가)
데이터 유형	정형 데이터 (데이터베이스, 사무정보)	비정형 데이터 (이미지, 멀티미디어)	사물 정보, 인지 정보 (RFID, 센서, 사물정보)
데이터 특성	구조화	다양성, 복합성, 소셜	현실성, 실시간성

1) 빅데이터 정의

- 빅데이터 등장 배경



빅데이터가 차세대 이슈로 떠오르는 이유는?

정보 통신 기술의 주도권이 데이터로 이동

구분	정보화 시대(1세대)	스마트 시대(2세대)
저장	관계형(정형) DB, 데이터 웨어하우스	비관계형(비정형) DB, 가상화, 클라우드 서비스
관리	지식관리시스템, 웹 2.0	플랫폼, SNS, 집단 지성
분석	경영 정보, 고객정보, ERP	빅데이터 분석

1) 빅데이터 정의

- 빅데이터 등장 배경



빅데이터가 차세대 이슈로 떠오르는 이유는?

빅데이터



공간, 시간, 관계, 세상 등을 담음



미래 경쟁력과 가치 창출의 원천

1) 빅데이터 정의

- 관점에 따른 정의

“빅데이터를 보는 관점에 따라
3가지로 정의

01

3V로 요약되는 데이터 자체의 특성 변화에 초점을
맞춘 좁은 의미

02

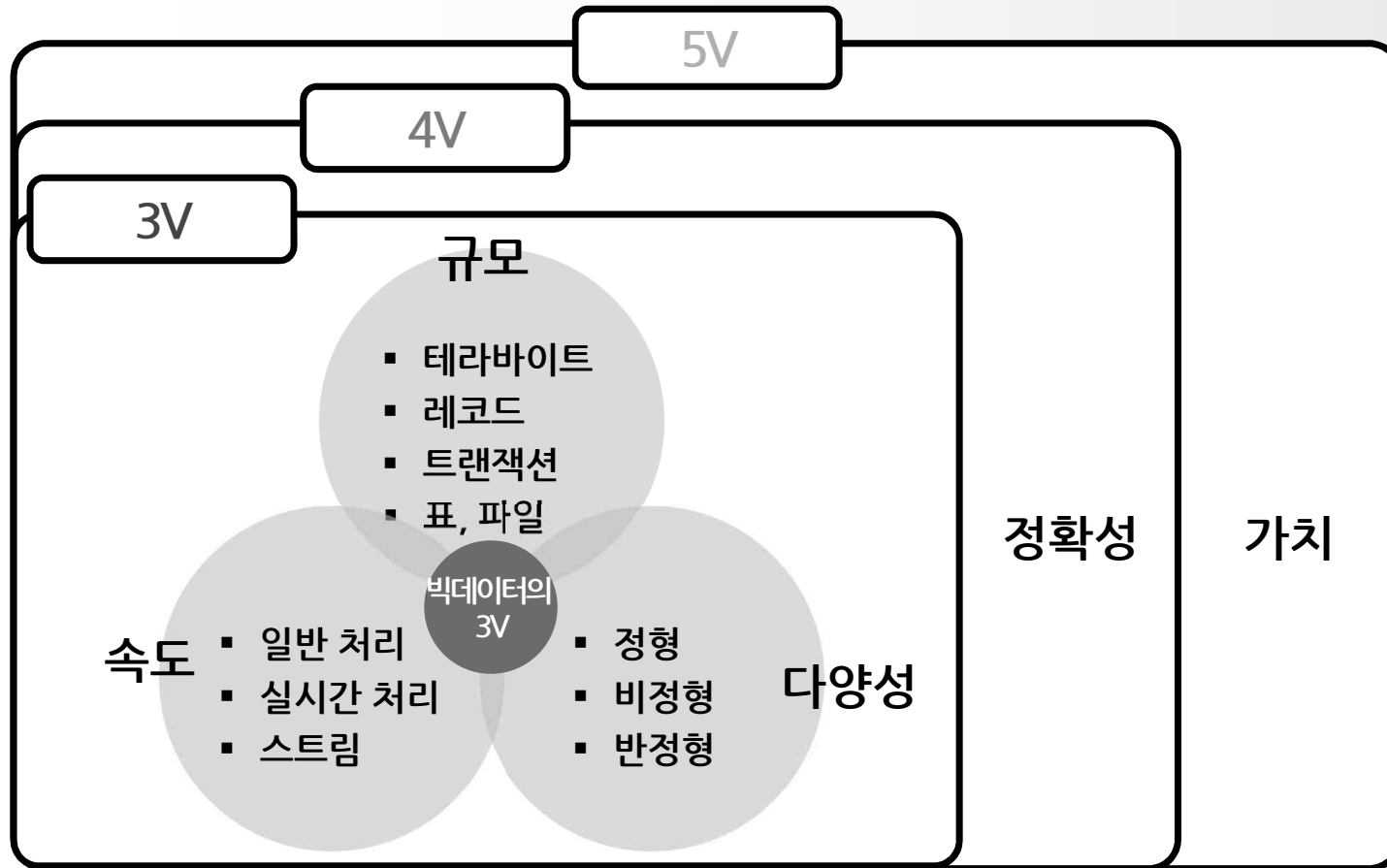
데이터 자체 뿐만 아니라 처리, 분석 기술적 변화까지
포함된 중간 범위

03

다른 객체와의 상호관계 속에서 가치를 갖는 것

1) 빅데이터 정의

● 관점에 따른 정의



〈출처 : 빅데이터 컴퓨팅 기술〉

1) 빅데이터 정의



데이터 규모에 중점을 둔 정의(맥킨지, 2011)



빅데이터는 일반적인 데이터베이스 소프트웨어로 저장, 관리 분석할 수 있는 범위를 초과하는 규모의 데이터



분석 비용 및 기술에 초점을 맞춘 정의(IDC, 2011)



빅데이터는 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처

1) 빅데이터 정의



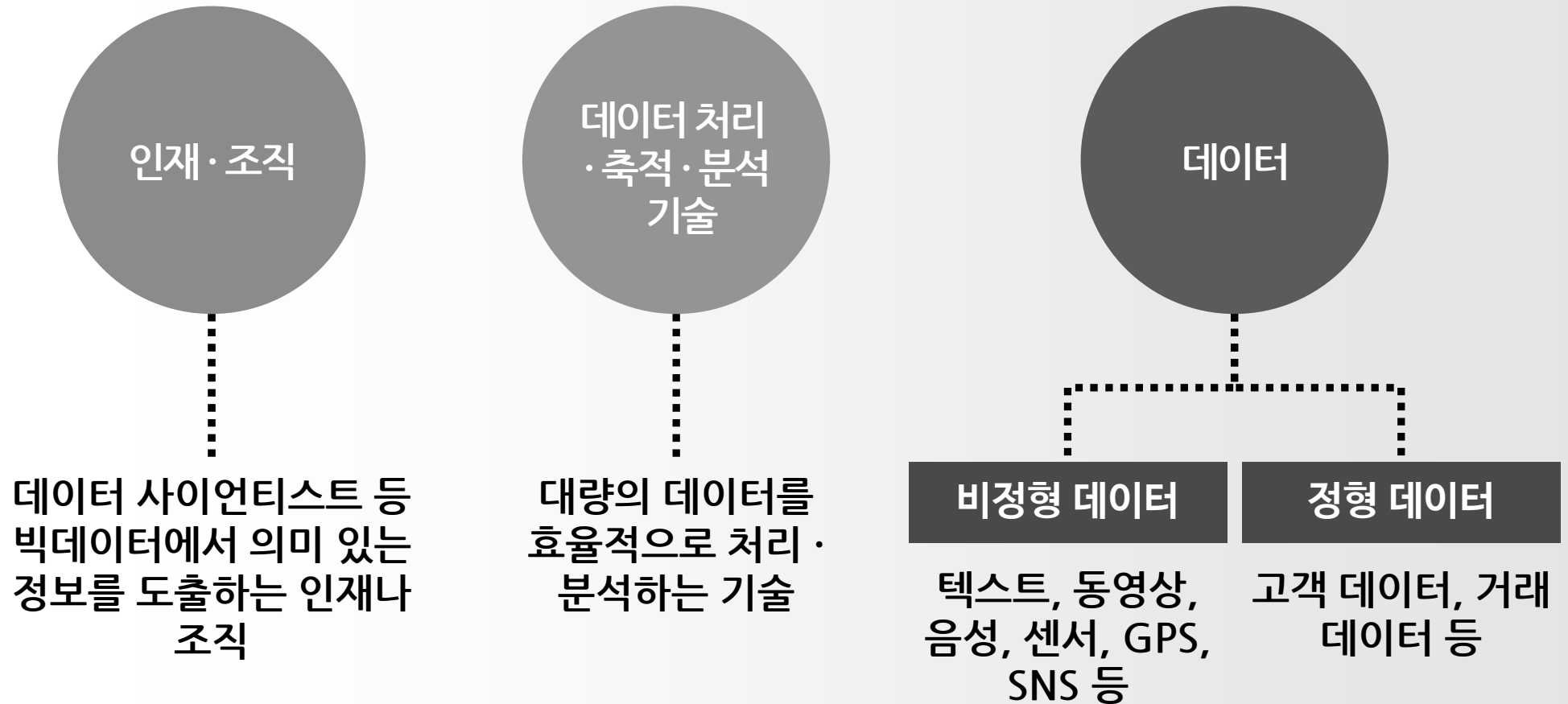
통찰과 가치 추출에 초점을 맞춘 정의
(메이어-웬베르거와쿠키어, 2013)



빅데이터란 대용량 데이터를 활용해
작은 용량에서는 얻을 수 없었던 새로운 통찰이나
가치를 추출해 내는 일

1) 빅데이터 정의

● 광의의 빅데이터 정의



1) 빅데이터 정의

- 빅데이터의 위치

- ▶ 규모와 다양성에 따른 빅데이터의 위치



1) 빅데이터 정의

- 빅데이터의 종류와 유형 변화

정형	고정된 필드에 저장된 데이터, 관계형 DBMS ↳ 사무정보
비정형	데이터, 텍스트 분석이 가능한 문서, 이미지, 동영상 등 ↳ 이메일, 콘텐츠, SNS
반정형	고정된 필드에 저장되어 있지는 않지만, 메타데이터나 스키마 등을 포함하는 데이터 ↳ XML, HTML

1) 빅데이터 정의

● 빅데이터의 종류와 유형 변화

구분	전통적 데이터	빅데이터
데이터 원천	<ul style="list-style-type: none">▪ 전통적 정보 서비스	<ul style="list-style-type: none">▪ 일상화된 정보 서비스
목적	<ul style="list-style-type: none">▪ 업무와 효율성	<ul style="list-style-type: none">▪ 사회적 소통, 자기표현, 사회기반 서비스
생성주체	<ul style="list-style-type: none">▪ 정부 및 기업 등 조직	<ul style="list-style-type: none">▪ 개인 및 시스템
데이터 유형	<ul style="list-style-type: none">▪ 정형데이터▪ 조직 내부 데이터▪ 주로 비공개 데이터	<ul style="list-style-type: none">▪ 비정형 데이터 (비디오, 오디오 등)▪ 조직 외부 데이터▪ 일부 공개 데이터

1) 빅데이터 정의

● 빅데이터의 종류와 유형 변화

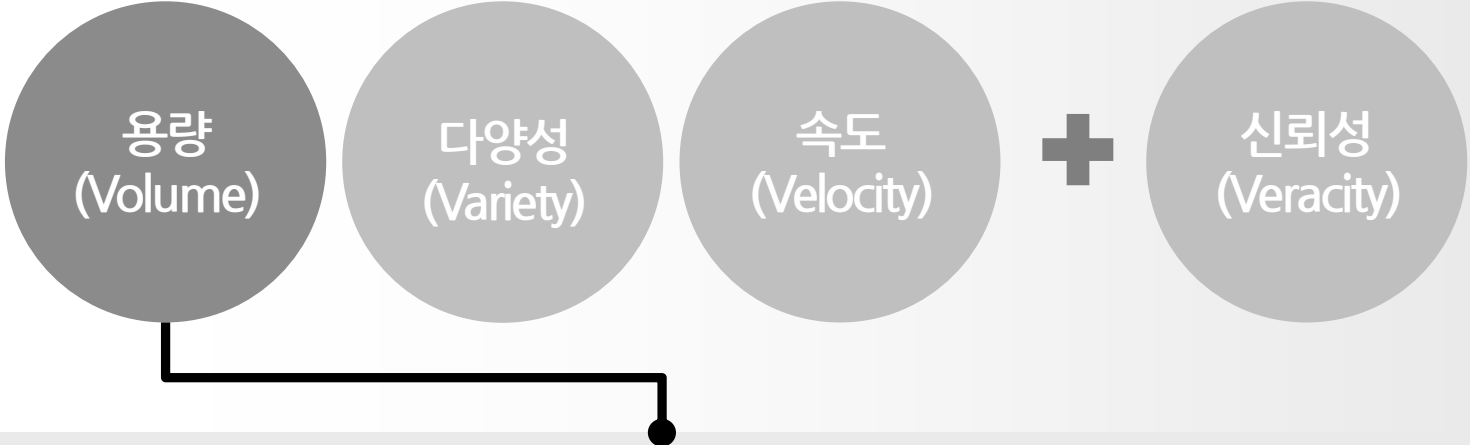
구분	전통적 데이터	빅데이터
데이터 특징	<ul style="list-style-type: none">▪ 데이터 증가량 관리 가능▪ 신뢰성 높은 핵심 데이터	<ul style="list-style-type: none">▪ 기하급수로 양적 증강▪ 쓰레기 데이터 비중 높음▪ 문맥 정보 등 다양한 데이터
데이터 보유	<ul style="list-style-type: none">▪ 정보, 기업 등 대부분 조직	<ul style="list-style-type: none">▪ 인터넷 서비스 기업▪ 포털▪ 이동통신회사

2) 빅데이터 특징



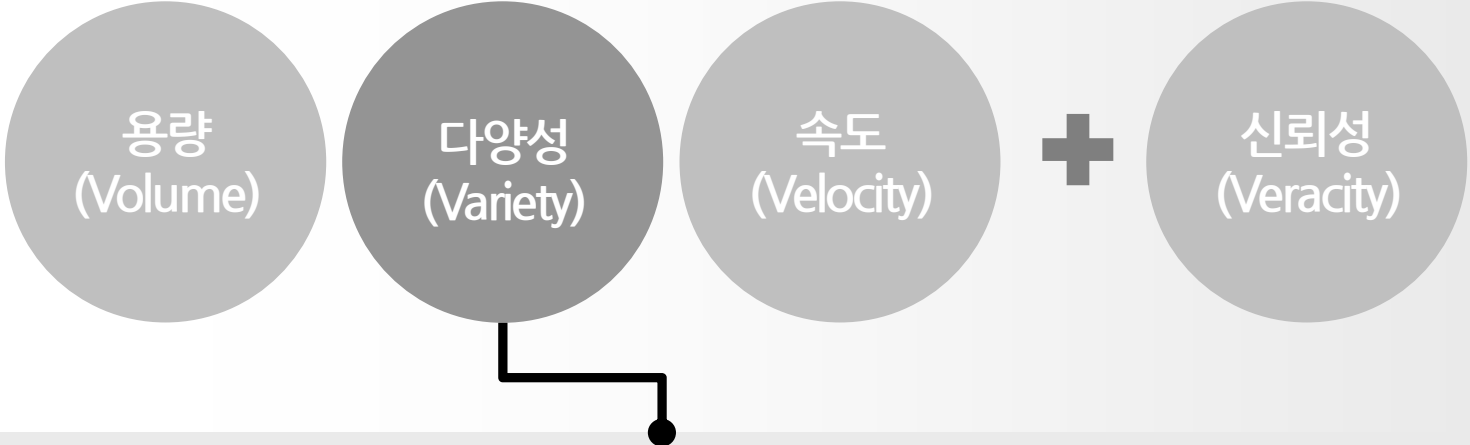
의미 없는 데이터에 대해
문제 핵심과 가치를 정확히 분석하고,
새로운 관계를 발견하도록 제공

2) 빅데이터 특징



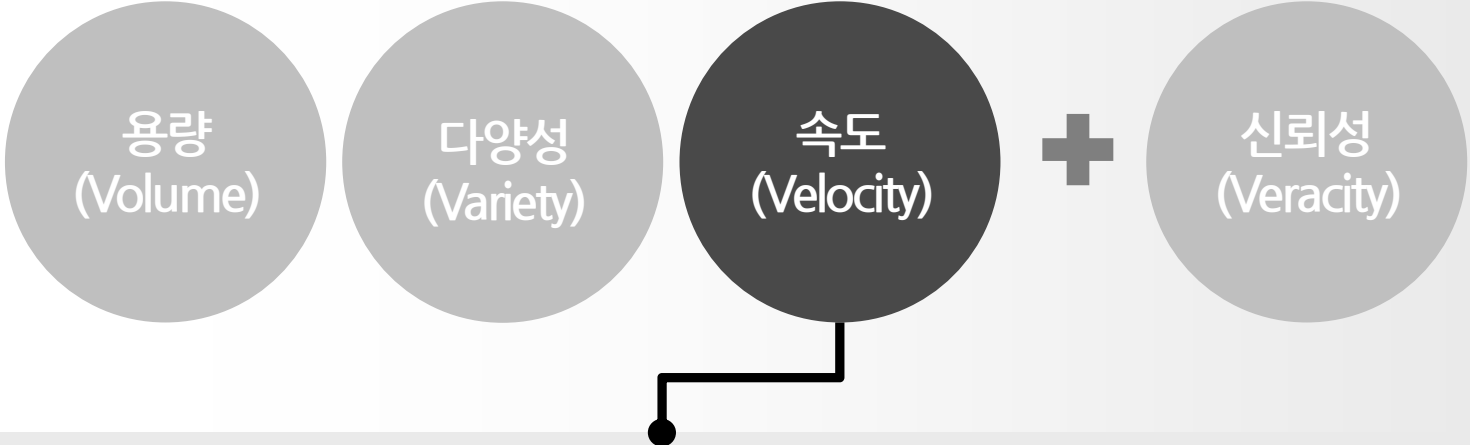
특성	대량의 산재되어 있는 데이터
제약	데이터 관점에서 관리 부재
방향	분산되어있는 대량 데이터 분석 통한 가치 제공

2) 빅데이터 특징



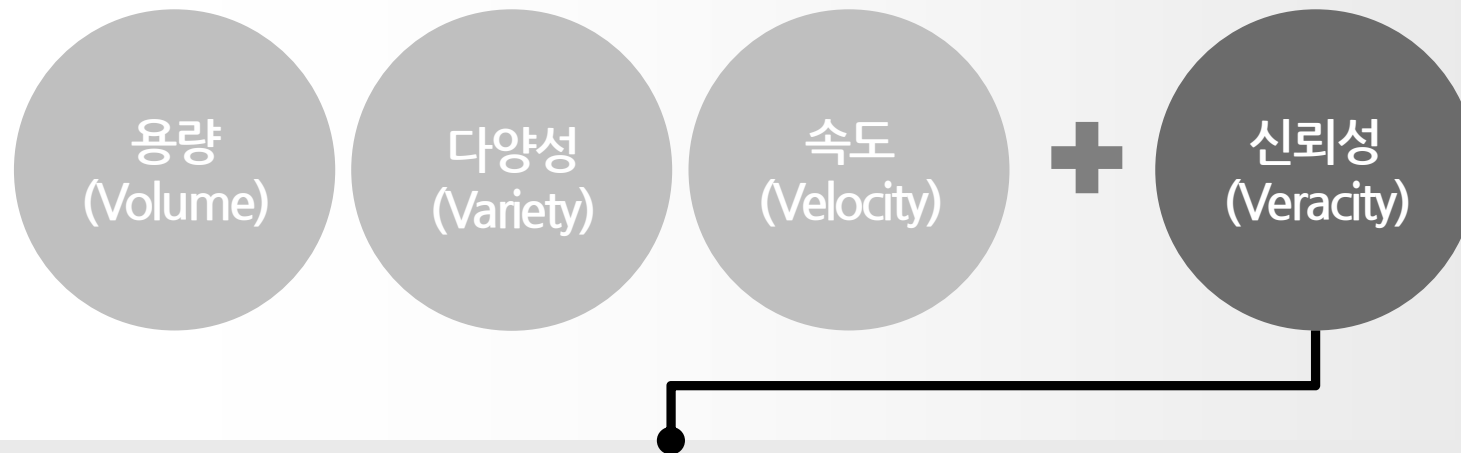
특성	분산된 데이터 중 비정형 80% 이상 차지
제약	특정 분산 병렬 처리 방법 제공
방향	다양한 비정형 데이터 처리 방법 제공

2) 빅데이터 특징



특성	데이터 생성 후 유통에서 활용까지 수분 소요
제약	일괄처리(실시간 처리 미지원)
방향	실시간 데이터 조회와 처리

2) 빅데이터 특징



특성	흩어져 있는 방대한 데이터를 분석
제약	분석의 정확도와 신뢰도 향상에 노력
방향	다양한 분석기법 및 시뮬레이션 처리

3) 빅데이터 활용

- 빅데이터 활용 분야



미래 예측

- 실시간 대용량 데이터 분석으로 정확도 높은 실시간 예측

예

미국 포드사

커넥티드 자동차 주행시스템에 운전자의 주행이력과 패턴을 분석하여 최적, 최단 주행경로 제안



숨은 필요 발견

- 대규모 데이터로부터 패턴을 발견해 숨은 필요사항 발견

3) 빅데이터 활용

- 빅데이터 활용 분야



리스크 경감

- 정량적 정보를 대폭 확보하여 위험의 감소와 정확도 증대

예

VISA

카드 부정검지모델에서 부정패턴의 갱신을 매일
수회 실시하여, 정밀도를 향상



맞춤형 서비스

- 고객 개인별로 차별화해 유용한 정보 제공

3) 빅데이터 활용

- 빅데이터 활용 분야



실시간 대응

- 실시간 모니터링과 즉각적인 대응이 가능

예

일본 Nishitetsu Store

매일 단일상품별 원가율과 원가변동 추리를 분석하여,
이익률이 높은 상품에 대한 일자별 주력 마케팅 정책
수립

3) 빅데이터 활용

산업	산업별 분석 애플리케이션(일차원분석)
금융서비스	신용점수 산정, 사기탐지, 가격 책정, 클레임 분석, 고객 수익분석
소매업	판촉, 매대 관리, 수요 예측, 재고 보충, 가격 및 제조 최적화
제조업	공급사슬 최적화, 수요 예측, 보증서 분석, 맞춤형 상품 개발, 신상품 개발
운송업	일정 관리, 노선 배정, 수익 관리
헬스케어	약품 거래, 예비 진단, 질병 관리

3) 빅데이터 활용

산업	산업별 분석 애플리케이션(일차원분석)
병원	가격 책정, 고객 로열티, 수익 관리
에너지	트레이딩, 공급 및 수요 예측
커뮤니케이션	가격계획 최적화, 고객 보유, 수요 예측, 생산능력 계획, 네트워크 최적화, 고객 수익성관리
서비스	콜센터 직원관리, 서비스 vs. 수익 사슬관리
정부	사기탐지, 사례 관리, 범죄 방지, 수익 최적화
온라인	웹 매트릭스, 사이트 설계, 고객 추천
모든 사업	성과관리



03

빅데이터 처리

1) 빅데이터 처리 과정

1) 빅데이터 처리 과정

● 빅데이터 처리 특징

구분	처리특징
의사 결정 속도	빠른 의사 결정이 상대적으로 덜 요구되어 장기적, 전략적 접근 필요
처리 복잡도	다양한 데이터 소스, 복잡한 로직처리, 대용량 데이터 처리로 처리 복잡도가 높아 분산 처리 기술 필요
데이터 규모	처리할 데이터 규모가 방대, 즉, 고객 정보 수집 및 분석을 장기간에 걸쳐 수행해야 함으로 처리해야할 데이터양이 방대
데이터 구조	비정형 데이터의 비중이 높음
분석 유연성	처리 · 분석 유연성이 높음
처리량	동시 처리량이 낮음

▶ 빅데이터 속성과 처리 특징

페이스북, 구글 플러스

다른 데이터 형식

- 비정형 데이터
- 약한 관계형 스키마
- 텍스트, 이미지, 비디오, 로그

대용량 데이터 규모

- 수백 테라바이트에서 수십 페타바이트

새로운 형태의 질문과 통찰

- 이 제품은 얼마나 인기 있나?
- 가장 뛰어난 광고는?
- 이 거래는 사기가 아닐까?

실시간 처리

- 데이터 발생 속도 가속화
- 스트림 데이터 폭증
- 신속한 데이터 처리

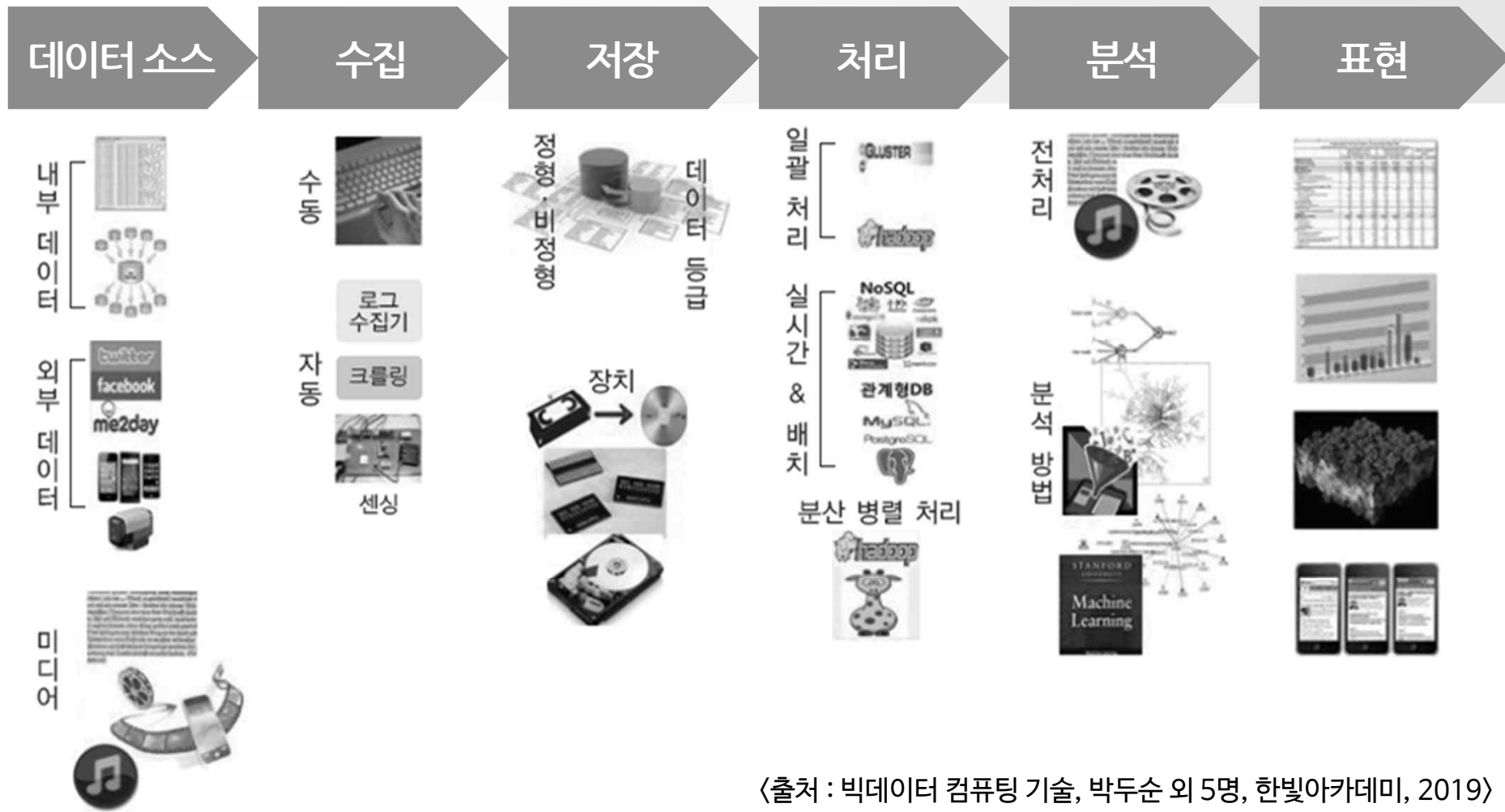
신경제학

- 초저가 대규모 프로세싱
- 분석 클라우드 컴퓨팅 환경

새로운 데이터 소스

- 센서
- 장비
- 기존 응용
- 웹 서버
- 공공 데이터

1) 빅데이터 처리 과정



〈출처 : 빅데이터 컴퓨팅 기술, 박두순 외 5명, 한빛아카데미, 2019〉

1) 빅데이터 처리 과정

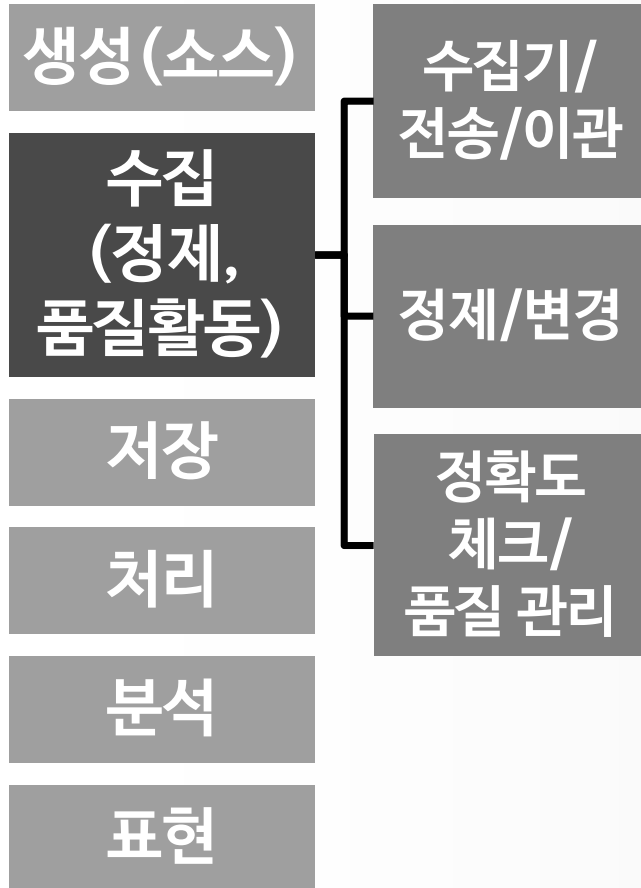
과정	영역	개요
생성	내부데이터	DB, 파일 관리 시스템
	외부데이터	인터넷으로 연결된 파일, 멀티미디어, 스트림
수집	크롤링	검색 엔진의 로봇을 사용한 데이터 수집
	ETL(extraction, transformation, loading)	소스 데이터의 추출, 전송, 변환, 적재
저장	NoSQL DB	비정형 데이터 관리
	스토리지	빅데이터 저장
	서버	초경량 서버

1) 빅데이터 처리 과정

과정	영역	개요
처리	맵리듀스	데이터 추출
	프로세싱	다중 업무 처리
분석	NLP	자연어 처리
표현	가시화	데이터를 도표나 그래픽적으로 표현
	획득	데이터의 획득 및 재해석

1) 빅데이터 처리 과정

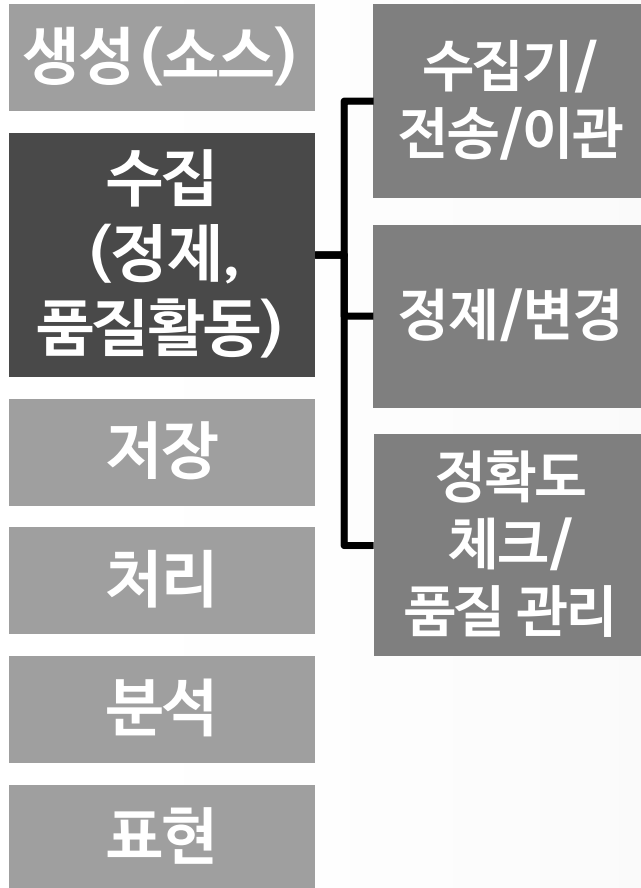
● 빅데이터 수집 기술



구분	설명
로그 수집기	<ul style="list-style-type: none">내부에 있는 웹 서버의 로그를 수집즉, 웹로그, 트랜잭션 로그, 클릭 로그, DB의 로그 데이터 등 수집
크롤링	<ul style="list-style-type: none">주로 웹 로봇으로 거미줄처럼 얽혀 있는 인터넷 링크를 따라다니며 방문한 웹 사이트의 웹 페이지라든가 소셜 데이터 등 인터넷에 공개되어 있는 데이터 수집
센싱	<ul style="list-style-type: none">각종 센서로 데이터 수집

1) 빅데이터 처리 과정

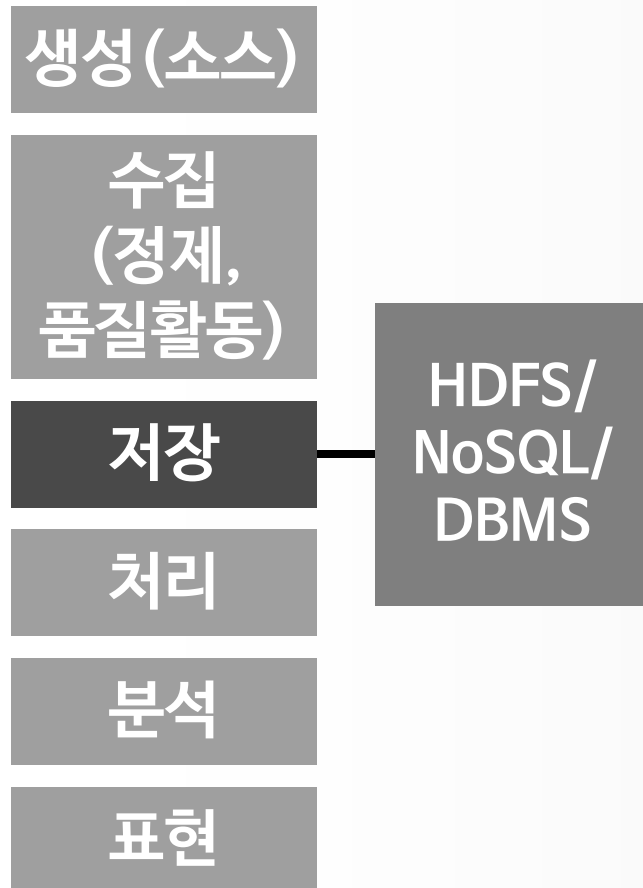
● 빅데이터 수집 기술



구분	설명
RSS 리더 /오픈 API	▪ 데이터의 생산 · 공유 · 참여 환경인 웹 2.0을 구현하는 기술로 필요한 데이터를 프로그래밍으로 수집
ETL	▪ 데이터의 추출 · 변환 · 적재의 약자로, 다양한 소스 데이터를 취합해 데이터를 추출 ▪ 하나의 공통된 형식으로 변환하여 데이터웨어하우스에 적재하는 과정 지원

1) 빅데이터 처리 과정

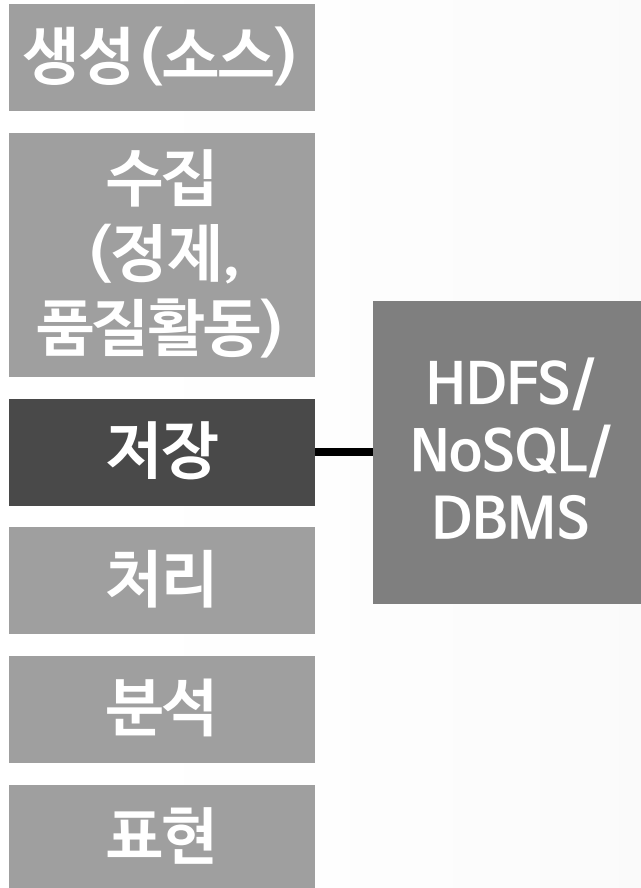
● 빅데이터 수집 기술



접근방식	설명	제품
분산파일 시스템	컴퓨터 네트워크로 공유하는 여러 호스트 컴퓨터 파일에 접근할 수 있는 파일 시스템	<ul style="list-style-type: none">▪ GFS(Google File System)▪ HDFS(Hadoop Distributed File System)▪ 아마존 S3 파일 시스템
NoSQL	데이터 모델을 단순화해서 관계형 데이터모델과 SQL을 사용하지 않는 모든 DBMS	<ul style="list-style-type: none">▪ Cloudata, Hbase, Cassandra

1) 빅데이터 처리 과정

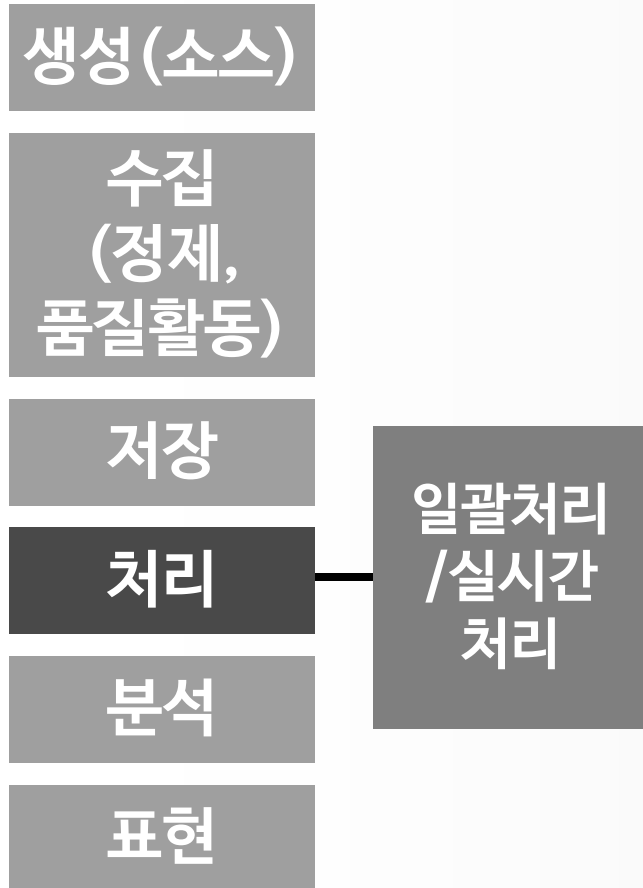
● 빅데이터 수집 기술



접근방식	설명	제품
병렬 DBMS	다수의 마이크로프로세서를 사용하여 여러 디스크의 질의, 갱신, 입출력 등 데이터베이스 처리를 동시에 수행하는 DBMS	▪ VoltDB, SAP HANA, Vertica, Greenplum, Netezza
네트워크 구성 저장 시스템	서로 다른 종류의 데이터 저장 장치를 하나의 데이터 서버에 연결하여 총괄적으로 데이터를 저장 및 관리	▪ SAN(Storage Area Network) ▪ NAS(Network Attached Storage)

1) 빅데이터 처리 과정

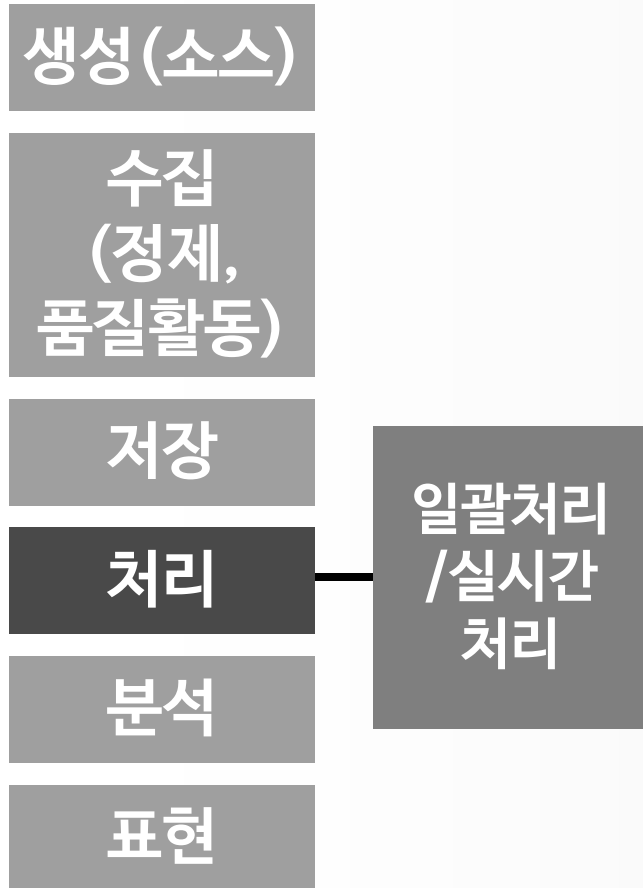
● 빅데이터 수집 기술



구분	설명
빅데이터 일괄 처리 기술	<ul style="list-style-type: none">일괄 처리 기술은 거대한 양의 데이터를 빨리 처리하기 위해 사용되며 단일 시스템에서는 처리용량보다 만들어지는 데이터가 많음적시 처리를 위해 그때그때 쌓인 데이터를 여러 개의 서버에 분산하여 나누어서 처리함이것을 다시 모아서 결과를 내는 분산·병렬 기술 방식을 사용 <ul style="list-style-type: none">구글의 맵리듀스(MapReduce), 하둡의 맵리듀스, MS의 드라이애드(Dryad)

1) 빅데이터 처리 과정

● 빅데이터 수집 기술



구분	설명
빅데이터 실시간 처리 기술	▪ 실시간 처리 기술은 데이터가 생기는 즉시 처리하여 실시간으로 생성되는 데이터를 적정 구간으로 나누어 스트림 데이터가 들어오는 대로 그 결과를 실시간으로 제공
	▪ IBM의 인포스피어 스트림스, 트위터의 스톰, 야후의 S4

1) 빅데이터 처리 과정

● 빅데이터 수집 기술

생성(소스)	용어	설명
수집 (정제, 품질활동)	텍스트 마이닝	▪ 자연어 처리 기술을 사용해 인간의 언어로 쓰인 비정형 텍스트에서 유용한 정보를 추출하거나 다른 데이터와의 연계성을 파악하며, 분류나 군집화 등 빅데이터에 숨겨진 의미 있는 정보를 발견하는 것
저장	웹 마이닝	▪ 인터넷에서 수집한 정보를 데이터 마이닝 기법으로 분석하는 것
처리	오피니언 마이닝	▪ 다양한 온라인 뉴스와 소셜 미디어 코멘트 사용자가 만든 콘텐츠에서 표현된 의견을 추출, 분류, 이해하고 자산화 하는 컴퓨팅 기술
분석		
표현		

1) 빅데이터 처리 과정

● 빅데이터 수집 기술

생성(소스)	용어	설명
수집 (정제, 품질활동)	리얼리티 마이닝	<ul style="list-style-type: none">▪ 휴대폰 등 기기를 사용하여 인간관계와 행동 양태 등을 추론하는 것▪ 통화량, 위치, 상태, 내용 등을 분석하여 사용자의 인간관계, 행동 특성 등 정보를 찾아냄
저장	소셜 네트워크 분석	<ul style="list-style-type: none">▪ 수학의 그래프 이론을 바탕으로 소셜 네트워크 서비스에서 소셜 네트워크 연결▪ 구조와 연결 강도를 분석하여 사용자의 명성 및 영향력을 측정하는 것
처리		
분석		
표현		

1) 빅데이터 처리 과정

● 빅데이터 수집 기술

생성(소스)	용어	설명
수집 (정제, 품질활동)	분류	<ul style="list-style-type: none">▪ 미리 알려진 클래스들로 구분되는 훈련 데이터군을 학습시켜 새로 추가되는 데이터가 속할 만한 데이터군을 찾는 지도 학습 방법▪ 가장 대표적인 방법으로 KNN이 있음
저장	군집화	<ul style="list-style-type: none">▪ 특성이 비슷한 데이터를 합쳐 군으로 분류하는 학습 방법▪ 분류와 달리 훈련 데이터군을 이용하지 않기 때문에 비지도 학습 방법▪ 트위터에서 주로 사진/카메라를 논의하는 사용자군과 게임에 관심 있는 사용자군 등 관심사나 취미에 따라 분류
처리		
분석		
표현		

1) 빅데이터 처리 과정

● 빅데이터 수집 기술

생성(소스)	용어	설명
수집 (정제, 품질활동)	기계 학습	<ul style="list-style-type: none">인공지능 분야에서 인간의 학습을 모델링한 것컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하여 수신한 이메일의 스팸 여부를 판단할 수 있도록 훈련
저장	감성 분석	<ul style="list-style-type: none">문장의 의미를 파악하여 글을 내용에 긍정/부정, 좋음/나쁨을 분류하거나 만족/불만족 강도를 지수화, 그런 다음 이 지수를 이용하여 고객의 감성 트렌드를 시계열적으로 분석하고 고객 감성 변화에 기업의 신속한 대응 및 부정적인 의견의 확산을 방지하는 데 사용
처리		
분석		
표현		

1) 빅데이터 처리 과정

● 빅데이터 수집 기술

생성(소스)

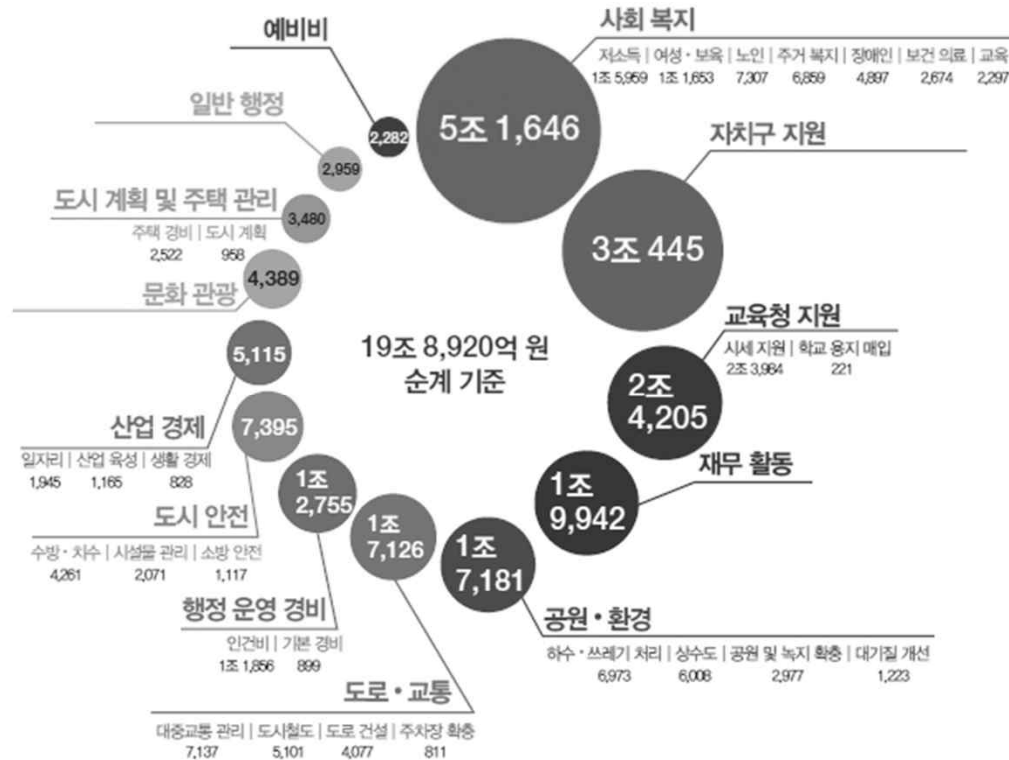
수집
(정제,
품질활동)

저장

처리

분석

표현



단위 %	점유비	증감률	단위 %
26.0	사회 복지		13.3
15.3	자치구 자원		4.8
12.2	교육청 자원		1.6
10.0	재무 활동		-1.3
8.6	공원·환경		-3.3
8.6	도로·교통		-2.9
6.4	행정 운영 경비		7.2
3.7	도시 안전		44.3
2.6	산업 경제		0.5
2.2	문화 관광		6.7
1.7	도시 계획 및 주택 관리		-15.0
1.5	일반 행정		13.4
1.2	예비비		-22.5

〈출처 : 빅데이터 컴퓨팅 기술, 박두순 외 5명, 한빛아카데미, 2019〉

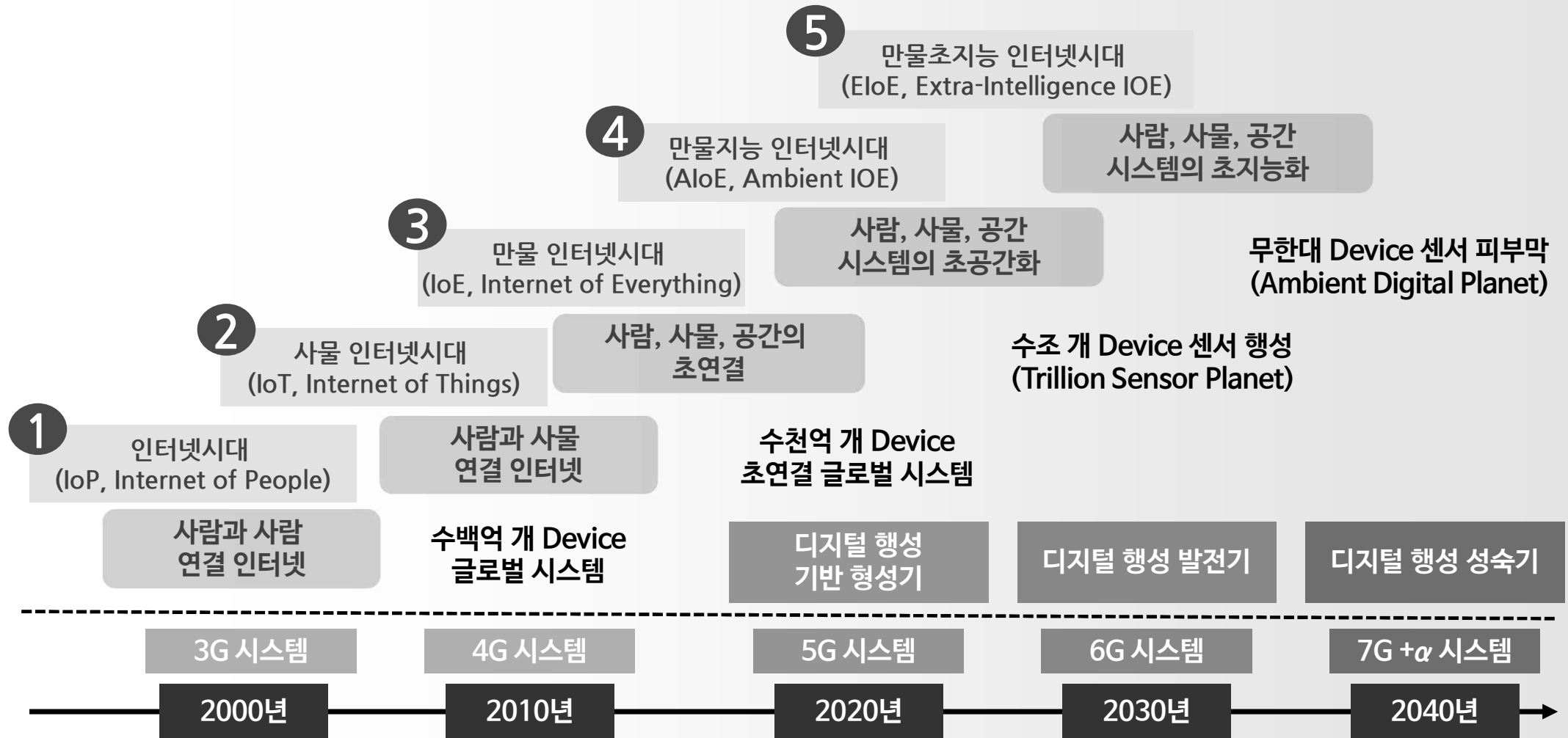
The background features a dark, abstract design with glowing white lines and geometric shapes. A prominent white hexagon is located in the upper center, with lines radiating from it. To the left, there are several overlapping, glowing white diamond shapes. The overall aesthetic is futuristic and technological.

04

빅데이터 미래

- 1) 빅데이터 발전
- 2) 빅데이터 트렌드

1) 데이터 발전



〈출처 : 하원규, '디지털 행성 시대의 메가 트렌드와 창조국가 전략', 〈국정관리연구〉 제 10권 제2호 (2015년 8월), p.46〉

2) 빅데이터 트렌드

● 데이터 폭발



4차 산업혁명의 디지털 주요기술



빅데이터, 인공지능, IOT



폭발적인 데이터 성장



- 개인적이며 어디서나 점점 더 접근 가능한 데이터 생성됨
- 물건이나 장치에서도 데이터를 생성



2) 빅데이터 트렌드

- 데이터 폭발

“

Big Data로부터 가치를
창출할 수 있는 가능성은 높아짐

”

BUT,

지속 가능한 Big Data 혁신 안을 내놓기
쉽지 않음



빠르게 행동 하지 않으면 결과는 실패 뿐임

2) 빅데이터 트렌드

● 빅데이터 시대 준비



데이터 경제 시대를 대비하는 '연결과 협력'



- 데이터는 무한한 자원이나 활용 가능한 자원의 영역은 상호 연결과 협력으로 더욱 확장 가능



빅데이터 핵심 역량인 '창의적 인력'의 양성



- 빅데이터는 많은 데이터를 수집하는 것보다 무엇을 분석할 것인지 분명한 목적의식과 통합적 사고력, 해석력이 중요함
- 향후 미래에 데이터 관리자 등 분석 인력이 부족해짐에 따라 전문가 양성이 시급함

2) 빅데이터 트렌드

● 빅데이터 시대 준비



데이터 신뢰 환경의 구축



- 데이터에 민감한 개인 사용자 정보의 노출 없이도 타당한 수준의 분석을 도출할 수 있는 방안을 고려해야 함

2) 빅데이터 트렌드

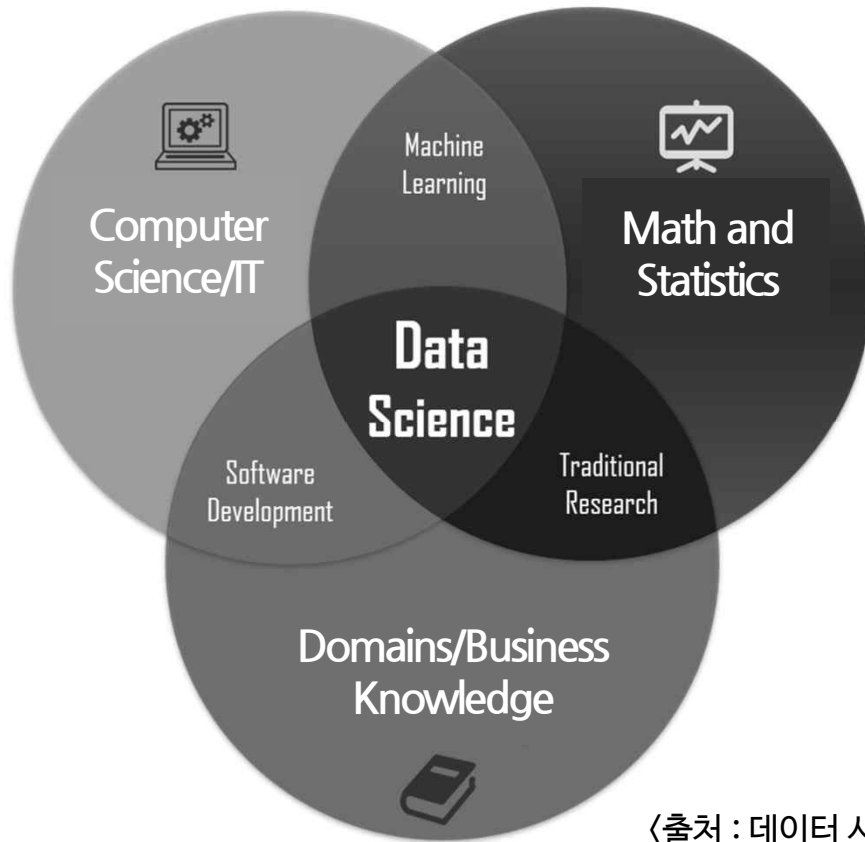
● 빅데이터 시대 준비

쇼퍼블 콘텐츠	음성 검색 광고	프로그래미틱 광고	브랜딩 퍼포먼스 마케팅	O4O 서비스
인플루언스를 통한 경험 마케팅의 증가, 이전 콘텐츠에서 바로 쇼핑할 수 있다.	날씨를 물었더니 영어 표현을 알려준다? 사용자 발화에 따라 다양한 제품/서비스 추천	비싼 TV 프로그램에 붙인 광고도 원하는 타겟에게만 보여줄 수 있는 프로그래매틱 TV광고	브랜딩과 퍼포먼스는 한 몸이 될 수 없다? No! 브랜딩 퍼포먼스라면 가능하다	언택트 서비스의 증가, 이전 오프라인을 위한 온라인 시대가 열린다. Online For Offline

〈출처 : 미디어랩 인크로스〉

2) 빅데이터 트렌드

● 빅데이터 분석 트렌드



〈출처 : 데이터 사이언스 컨셉, Micheal Barber〉

학습 평가

Q1

데이터를 통해 도출된 다양한 정보를
구조화하여 유의미한 정보를 분류하고
개인적인 경험을 결합시켜 고유의
으로 내재화된 것을 이라고 한다.

Q1

Q2

학습 평가

Q1

Q2

Q1

데이터를 통해 도출된 다양한 정보를
구조화하여 유의미한 정보를 분류하고
개인적인 경험을 결합시켜 고유의 지식
으로 내재화된 것을 지식 이라고 한다.

정답

지식

해설

지식은 유의미한 정보를 분류하고 개인적인
경험을 결합시켜 내재화 된 것입니다.

학습 평가

Q2

정성적 데이터는 비정형 데이터로,
주관적 내용이며, 통계분석이 용이하다.

Q1

Q2

O

X

학습 평가

Q1

Q2

Q2

정성적 데이터는 비정형 데이터로,
주관적 내용이며, 통계분석이 용이하다.

O

X

정답

X

해설

정성적 데이터는 통계분석에 어려움을 겪습니다.

정리 하기

데이터와 정보의 이해

✓ 데이터의 정의

- 데이터는 단순히 객체로써의 가치뿐만 아니라 다른 객체와의 상호관계 속에서 가치를 갖는 것으로 설명되고 있음
- 정성적 데이터는 주관적이며, 통계분석이 어려움
- 정량적 데이터는 객관적이며, 통계분석이 용이함

✓ 정보의 정의

- 데이터의 가공, 처리와 데이터간 연관관계 속에서 의미가 도출된 것

정리 하기

데이터와 정보의 이해

✓ DBMS의 정의

- 문자, 기호, 음성, 화상, 영상 등 상화 관련된 다수의 콘텐츠를 정보 처리 및 정보통신 기기에 의하여 체계적으로 수집, 축적하여 다양한 용도와 방법으로 이용할 수 있도록 정리한 정보의 집합체

정리 하기

빅데이터 정의와 활용

- ✓ 빅데이터의 정의
 - 기존의 DBMS로는 처리하기 어렵고
굉장히 큰 규모의 데이터
- ✓ 빅데이터 3요소
 - 용량(크기) = Volume
 - 다양성 = Variety
 - 속도 = Velocity



정리 하기

빅데이터 정의와 활용

- ✓ 빅데이터 특징
 - 비정형성
 - 중요 패턴에 대한 분석 불가
 - 불필요한 정보가 급증

정리
하기

빅데이터 처리

✓ 빅데이터 처리 과정



정리 하기

빅데이터 미래

- ✓ 빅데이터 시대 준비
 - 데이터 경제 시대를 대비하는 '연결과 협력'
 - 빅데이터 핵심 역량인 '창의적 인력'의 양성
 - 데이터 신뢰 환경의 구축



- 다음 시간에 살펴 볼 내용 -

02강 빅데이터 활용

수고하셨습니다.