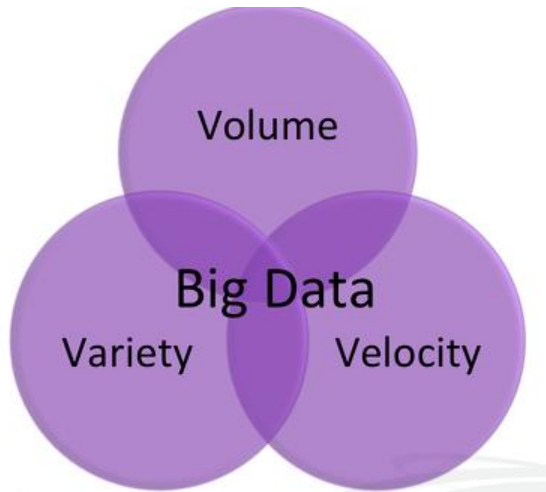


# 빅데이터 전체 목차

---

- **빅데이터개념소개**
  - 빅데이터개요
  - 빅데이터속성
  - 빅데이터처리기술소개
- **Hadoop**
  - HDFS
  - MapReduce
- **BigDatawith Python**
  - 파이썬기초문법
  - 데이터수집
  - 데이터저장(csv, mysql, MongoDB)
  - 데이터분석및시각화
  - 머신러닝
- **MongoDB**
  - NoSQL
  - 데이터분산저장
- **빅데이터응용Project**

# 1. 빅데이터 소개



1. 빅데이터 개념
2. 빅데이터 처리 과정
3. 빅데이터 처리 기술
4. 빅데이터 활용 사례

# 1. 빅데이터 개념

---

## • 등장배경

- 빅데이터를 소셜 미디어 데이터로 오인하기도 함
- 1990년 이후 인터넷이 확산되면서 정형화된 데이터와 비정형화된 데이터가 무수히 발생하면서 정보 홍수 개념이 등장, 이것이 오늘날 빅데이터 개념으로 이어짐
- 개인화 서비스와 SNS의 확산으로 기본 인터넷 서비스 환경 재구성
- 전 세계 디지털 데이터 양이 제타 바이트 단위로 2년마다 2배씩 증가하여 2020년에는 약 40제타 바이트가 될 것이라고 함
- 특히 스마트폰의 보급으로 데이터가 매우 빠르게 축적되어 제타바이트 시대를 스마트 시대라고도 함

### NOTE\_ 디지털 데이터 단위

- |                                |                               |
|--------------------------------|-------------------------------|
| • 1테라바이트(TeraByte; TB)=1024GB  | • 1페타바이트(PetaByte; PB)=1024TB |
| • 1엑사바이트(ExaByte; EB)=1024PB   | • 1제타바이트(ZetaByte; ZB)=1024EB |
| • 1요타바이트(YottaByte; YB)=1024ZB |                               |

# 1. 빅데이터 개념

## • 등장배경 : 정보기술의 패러다임 변화

- 데이터 양의 급속한 증가로 기존의 데이터 저장, 관리, 분석기법으로는 데이터 처리에 한계가 있어 정보 기술의 패러다임도 [표1-1]과 같이 변화

표 1-1 정보 기술의 패러다임 변화 [01]

	PC 시대	인터넷 시대	모바일 시대	스마트 시대
패러다임 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물 정보화
정보 기술 이슈	PC, PC통신, 데이터베이스	초고속 인터넷, www, 웹 서버	모바일 인터넷, 스마트폰	빅데이터, 차세대 PC, 사물 네트워크 Machine to Machine; M2M
핵심 분야(서비스)	PC, OS	포털, 검색 엔진, Web 2.0	스마트폰, 웹 서비스, SNS	미래 전망, 상황 인식, 개인화 서비스
대표 기업	MS, IBM	구글, 네이버, 유튜브	애플, 페이스북, 트위터	구글, 삼성, 애플, 페이스북, 트위터
정보 기술 비전	1인 1PC	클릭 e-Korea	손 안의 PC, 소통	IT everywhere, 신가치창출

# 1. 빅데이터의 개념

## • 정보 통신 기술 발전에 따른 데이터의 변화 방향



그림 1-1 정보 통신 기술 발전에 따른 데이터의 변화 방향 [01]

# 1. 빅데이터 개념

- 빅데이터가 차세대 이슈로 떠오르는 이유
  - 정보통신 기술의 주도권이 데이터로 이동

표 1-2 정보화 시대와 스마트 시대의 데이터 처리 변화 [01]

구분	정보화 시대(1세대)	스마트 시대(2세대)
저장	관계형(정형) 데이터베이스, 데이터웨어하우스	비관계형(비정형) 데이터베이스, 가상화, 클라우드 서비스
관리	지식 관리 시스템 Knowledge Management System; KMS, 웹 2.0 Web 2.0	플랫폼, 소셜 네트워크, 집단지성
분석	경영 정보, 고객 정보, 자산 정보 분석(ERP, CRM, 데이터 마이닝 등)	빅데이터 분석 (소셜 분석, 고급 분석, 시각화)

- 공간, 시간, 관계, 세상 등을 담은 빅데이터
- 빅데이터는 미래 경쟁력과 가치 창출의 원천

# 1. 빅데이터의 개념

---

## • 빅데이터의 개념과 속성

### • 속성

- 규모(Volume) - 데이터의 크기  
페이스북의 하루 사진 데이터는 30페타 바이트  
뉴욕 증권거래소의 하루 데이터는 1테라 바이트
- 다양성(Variety) - 다양한 종류의 데이터  
수치화(정형화)하기 어려운 데이터 포함
- 속도(Velocity) - 빠르게 데이터가 생성되고 사라짐, 짧은 시간 유의미
- 정확성(Veracity) - 표본이 크므로 신뢰성도 높음
- 가치(Value) - 빅데이터 분석 결과가 유의미한 정보를 제공해야 함

# 1. 빅데이터의 개념

## • 빅데이터의 속성

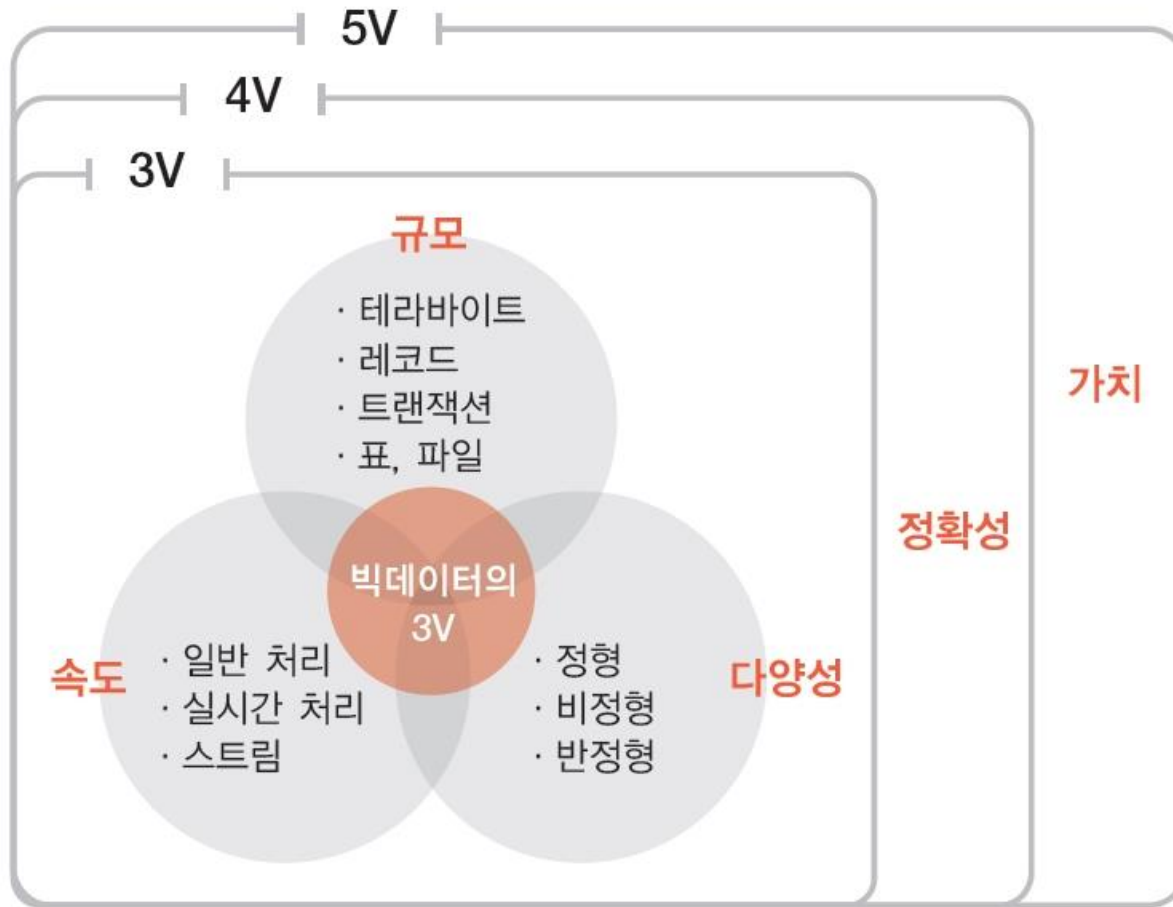


그림 1-2 빅데이터의 속성 [02]



# 1. 빅데이터의 개념

## • 빅데이터의 정의

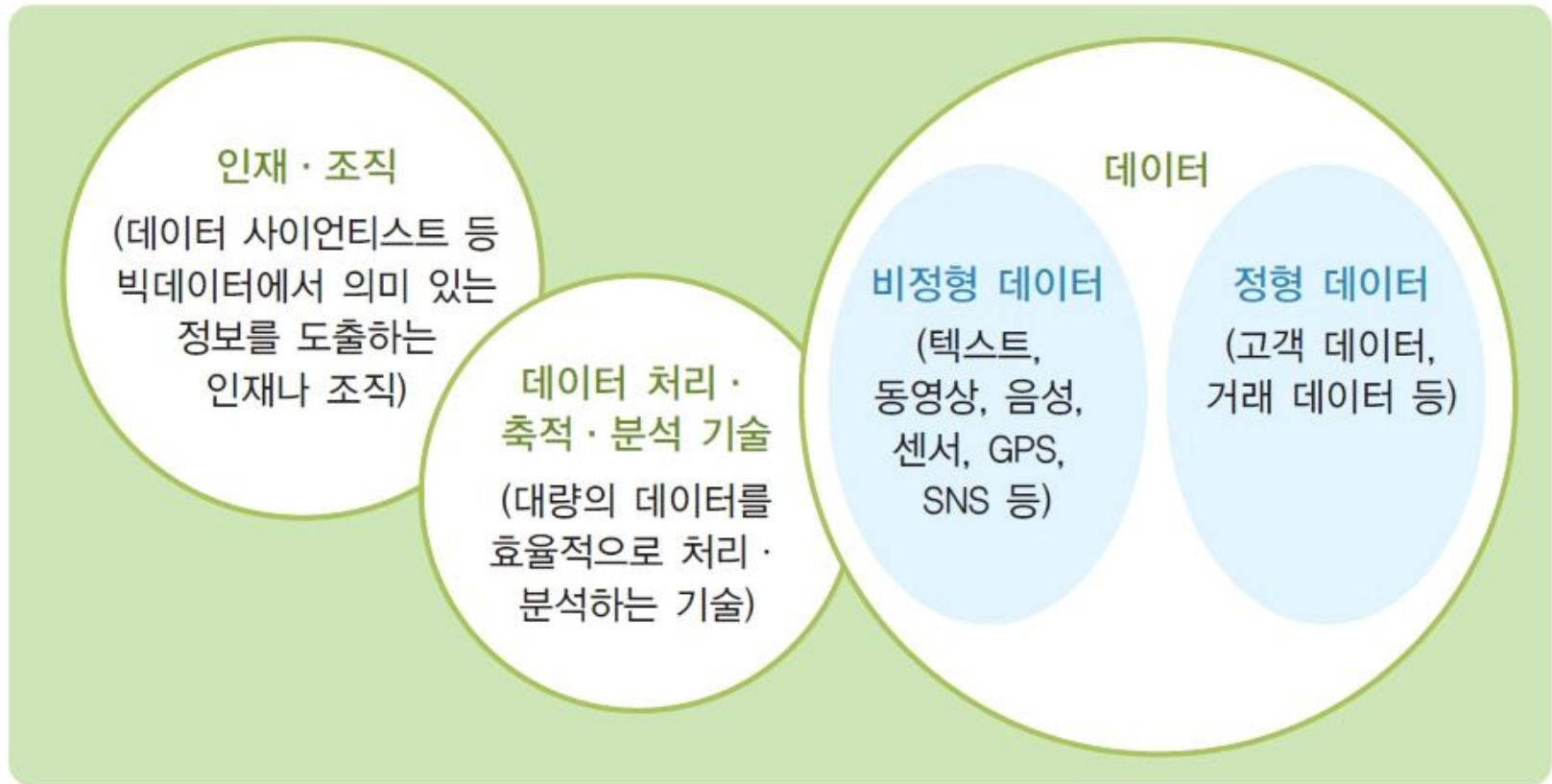


그림 1-3 광의의 빅데이터 정의 [03]

# 1. 빅데이터의 개념

- 빅데이터 위치

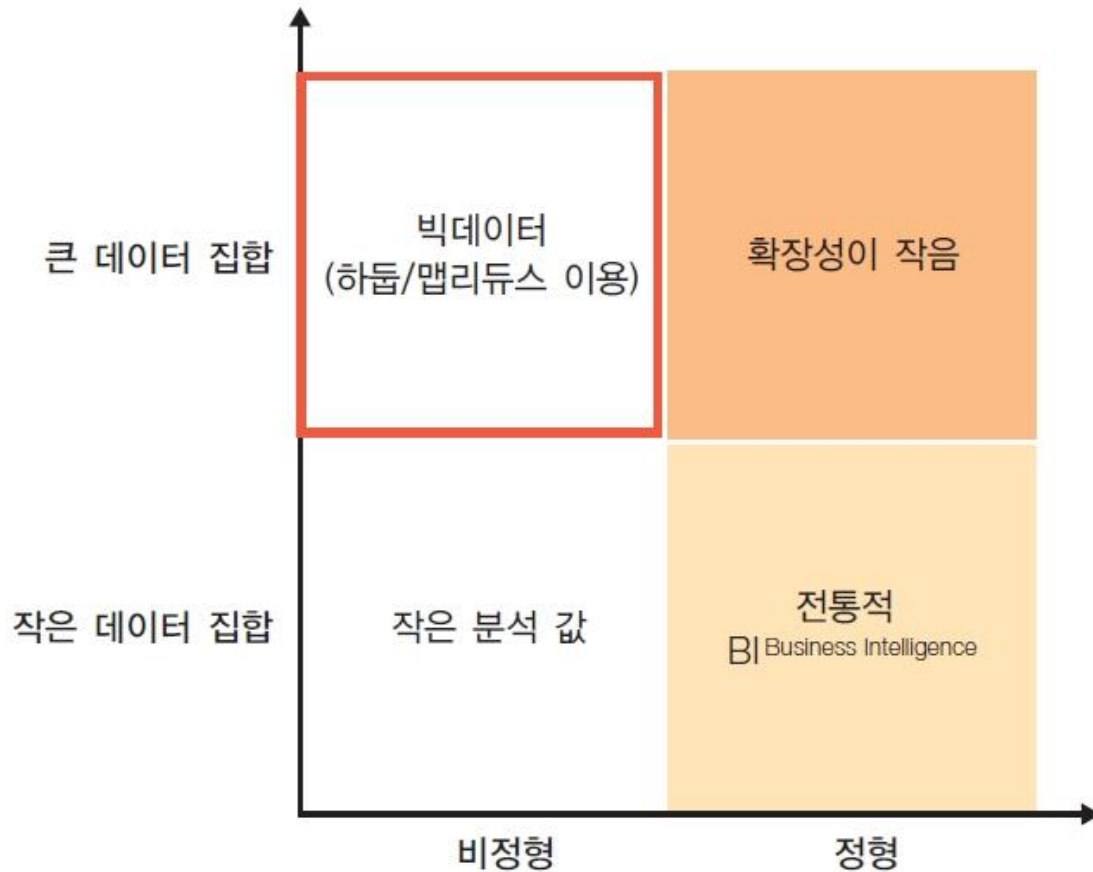


그림 1-4 규모와 다양성에 따른 빅데이터의 위치 [04]

# 1. 빅데이터의 개념

---

- 빅데이터 종류와 유형 변화

표 1-3 빅데이터 종류 [05]

종류	설명
정형	고정된 필드에 저장된 데이터 예 관계형 데이터베이스, 스프레드시트
반정형	고정된 필드에 저장되어 있지는 않지만, 메타데이터나 스키마 등을 포함하는 데이터 예 XML, HTML 텍스트
비정형	고정된 필드에 저장되어 있지 않은 데이터 예 텍스트 분석이 가능한 텍스트 문서, 이미지 · 동영상 · 음성 데이터

# 1. 빅데이터 개념

## • 정형과 비정형 데이터 유형의 변화

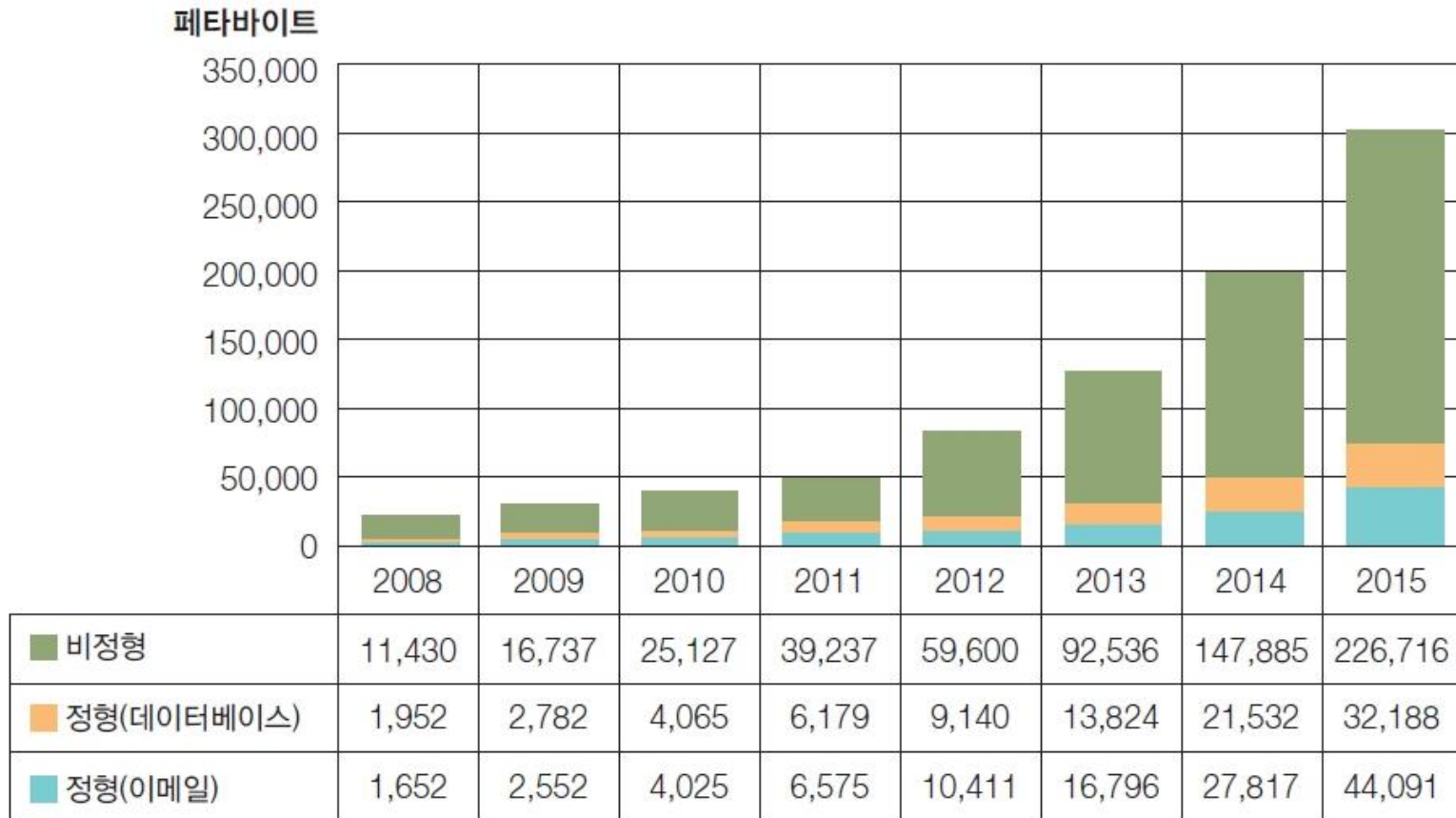


그림 1-5 정형과 비정형 데이터 유형의 변화 [06]

# 1. 빅데이터 개념

## • 전통적 데이터와 빅데이터 특징 비교

구분	전통적 데이터	빅데이터
데이터 원천	전통적 정보 서비스	일상화된 정보 서비스
목적	업무와 효율성	사회적 소통, 자기표현, 사회 기반 서비스
생성 주체	정부 및 기업 등 조직	개인 및 시스템
데이터 유형	<ul style="list-style-type: none"><li>■ 정형 데이터</li><li>■ 조직 내부 데이터(고객 정보, 거래 정보 등)</li><li>■ 주로 비공개 데이터</li></ul>	<ul style="list-style-type: none"><li>■ 비정형 데이터(비디오 스트림, 이미지, 오디오, 소셜 네트워크 등 사용자 데이터, 센서 데이터, 응용 프로그램 데이터 등)</li><li>■ 조직 외부 데이터</li><li>■ 일부 공개 데이터</li></ul>
데이터 특징	<ul style="list-style-type: none"><li>■ 데이터 증가량 관리 가능</li><li>■ 신뢰성 높은 핵심 데이터</li></ul>	<ul style="list-style-type: none"><li>■ 기하급수로 양적 증가</li><li>■ 쓰레기(Garbage) 데이터 비중 높음</li><li>■ 문맥 정보 등 다양한 데이터</li></ul>
데이터 보유	정부, 기업 등 대부분 조직	<ul style="list-style-type: none"><li>■ 인터넷 서비스 기업(구글, 아마존 등)</li><li>■ 포털(네이버, 다음 등)</li><li>■ 이동 통신 회사(SKT, KTF 등)</li><li>■ 디바이스 생산 회사(애플, 삼성전자 등)</li></ul>
데이터 플랫폼	정형 데이터를 생산·저장·분석·처리할 수 있는 전통적 플랫폼 예) 분산 DBMS, 다중처리기, 중앙 집중 처리	비정형 대량 데이터를 생산·저장·분석·처리할 수 있는 새로운 플랫폼 예) 대용량 비정형 데이터 분산 병렬 처리

## 2. 빅데이터 처리 과정

### • 빅데이터 처리 특징

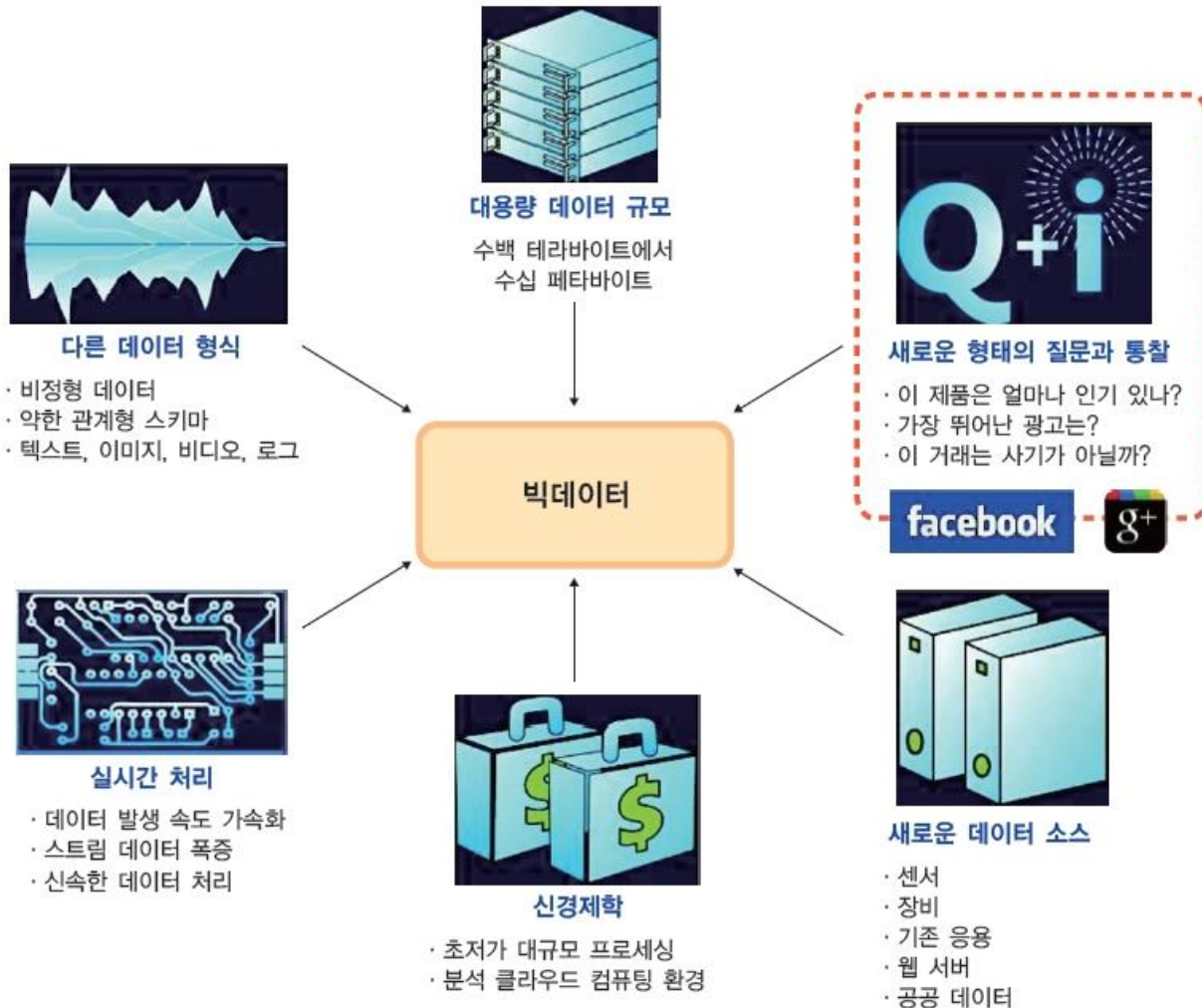
표 1-5 빅데이터의 처리 특징 [02]

구분	처리 특징
의사 결정 속도	빠른 의사 결정이 상대적으로 덜 요구되어 장기적 · 전략적 접근 필요
처리 복잡도 <small>Processing Complexity</small>	다양한 데이터 소스, 복잡한 로직 처리, 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 하므로 처리해야 할 데이터양이 방대
데이터 구조	비정형 데이터의 비중이 높음. 즉, 소셜 미디어 데이터, 로그 파일, 스트림 데이터, 콜센터 로그 등 비정형 데이터 파일의 비중이 높음
분석 유연성 <small>Analysis Flexibility</small>	처리 · 분석 유연성이 높음. 즉, 잘 정의된 데이터 모델, 상관관계, 절차 등이 없어 기존 데이터 처리 방법에 비해 처리 및 분석 유연성이 높음
처리량 <small>Throughput</small>	동시 처리량이 낮음. 즉, 대용량 및 복잡한 처리가 가능하여 동시에 처리할 수 있는 데이터양이 적어 실시간 처리가 보장되어야 하는 데이터 분석에는 부적합



## 2. 빅데이터 처리과정

### • 빅데이터 속성과 처리 특징



## 2. 빅데이터 처리 과정

### • 빅데이터 처리 과정과 기술

데이터 통합



그림 1-7 빅데이터 처리 과정 [09]



## 2. 빅데이터 처리 과정

### • 빅데이터 처리 과정별 기술 영역

표 1-6 빅데이터 처리 과정별 기술 영역 [10]

과정	영역	개요
생성	내부 데이터	데이터베이스 Database, 파일 관리 시스템 File Management System
	외부 데이터	인터넷으로 연결된 파일, 멀티미디어, 스트림
수집	크롤링 Crawling	검색 엔진의 로봇을 사용한 데이터 수집
	ETL Extraction, Transformation, Loading	소스 데이터의 추출 · 전송 · 변환 · 적재
저장	NoSQL 데이터베이스	비정형 데이터 관리
	스토리지 Storage	빅데이터 저장
	서버 Server	초경량 서버
처리	맵리듀스 MapReduce	데이터 추출
	프로세싱 Processing	다중 업무 처리
분석	NLP Neuro Linguistic Programming	자연어 처리
	기계 학습 Machine Learning	기계 학습으로 데이터의 패턴 발견
	직렬화 Serialization	데이터 간의 순서화
표현	가시화 Visualization	데이터를 도표나 그래픽적으로 표현
	획득 Acquisition	데이터의 획득 및 재해석

### 3. 빅데이터 기술

#### • 빅데이터 수집 기술

- 내부 데이터 수집 : 자체적으로 보유한 내부 파일 시스템, 데이터베이스 관리 시스템, 센서 등에서 정형 데이터를 수집
- 외부 데이터 수집 : 인터넷으로 연결된 외부에서 비정형 데이터를 수집

표 1-7 빅데이터 자동 수집 방법 [07]

방법	설명
로그 수집기	내부에 있는 웹 서버의 로그를 수집. 즉, 웹 로그, 트랜잭션 로그, 클릭 로그, DB의 로그 데이터 등 수집
크롤링	주로 웹 로봇으로 거미줄처럼 얽혀 있는 인터넷 링크를 따라다니며 방문한 웹 사이트의 웹 페이지라든가 소셜 데이터 등 인터넷에 공개되어 있는 데이터 수집
센싱	각종 센서로 데이터 수집
RSS 리더/오픈 API	데이터의 생산·공유·참여 환경인 웹 2.0을 구현하는 기술로 필요한 데이터를 프로그래밍으로 수집
ETL Extraction, Transformation, and Loading	데이터의 추출, 변환, 적재의 약자로, 다양한 소스 데이터를 취합해 데이터를 추출하고 하나의 공통된 형식으로 변환하여 데이터웨어하우스에 적재하는 과정 지원

### 3. 빅데이터 기술

#### • 빅데이터 저장 기술

표 1-8 대용량 데이터를 저장하는 다양한 접근 방식 [07]

접근 방식	설명	제품
분산 파일 시스템	컴퓨터 네트워크로 공유하는 여러 호스트 컴퓨터 파일에 접근할 수 있는 파일 시스템	GFS <sup>Google File System</sup> , HDFS <sup>Hadoop Distributed File System</sup> , 아마존 S3 파일 시스템
NoSQL	데이터 모델을 단순화해서 관계형 데이터 모델과 SQL을 사용하지 않는 모든 DBMS 또는 데이터 저장 장치	Cloudata, HBase, Cassandra
병렬 DBMS	다수의 마이크로프로세서를 사용하여 여러 디스크의 질의, 갱신, 입출력 등 데이터베이스 처리를 동시에 수행하는 데이터베이스 시스템	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 <sup>Storage Area Network</sup> , NAS <sup>Network Attached Storage</sup>

### 3. 빅데이터 기술

---

- 빅데이터 처리 기술

- 맵리듀스

- 분산 병렬 데이터 처리 기술의 표준, 일반 범용 서버로 구성된 군집화 시스템을 기반으로 <키,값> 입력 데이터분할 처리 및 처리 결과 통합 기술, job 스케줄링 기술, 작업 분배 기술, 태스크 재수행 기술이 통합된 분산 컴퓨팅 기술

- R

- R 언어와 개발 환경으로 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현 및 개선이 가능

- 하둡

- 정형·비정형 빅데이터 분석에 가장 선호되는 솔루션

- *NoSQL*<sup>No SQL; Not-only SQL</sup>

- 전통적인 관계형 데이터베이스 RDBMS와는 다르게 설계된 비관계형 데이터베이스

### 3. 빅데이터 기술

#### • 빅데이터 처리 기술

표 1-9 빅데이터 처리 기술

용어	설명
빅데이터 일괄 처리 기술	<ul style="list-style-type: none"><li>■ 빅데이터를 여러 서버로 분산하여 각 서버에서 나누어 처리하고, 이를 다시 모아서 결과를 정리하는 분산·병렬 기술 방식</li><li>■ 구글 맵리듀스(구글에서 분산 컴퓨팅을 지원할 목적으로 제작·발표한 소프트웨어 프레임워크, 함수형 프로그래밍에서 일반적으로 사용되는 맵<sup>Map</sup>과 리듀스<sup>Reduce</sup> 함수를 기반으로 주로 구성), 하둡 맵리듀스, 마이크로소프트 드라이애드<sup>Dryad</sup> 등이 있음</li></ul>
빅데이터 실시간 처리 기술	스트림 처리 기술로 강화된 스트림 컴퓨팅을 지원하는 IBM의 InfoSphere Streams <sup>인포스피어 스트림즈</sup> , 분산 환경에서 스트리밍 데이터를 분석할 수 있게 해주는 트위터의 스톰 <sup>Storm</sup>
빅데이터 처리 프로그래밍 지원 기술	분산 데이터를 처리하는 프로그래밍 언어인 구글의 소잘 <sup>Sawzall</sup> 과 병렬 처리를 하는 고성능 데이터-플로우 언어와 실행 프레임워크인 하둡 Pig

### 3. 빅데이터 기술

#### • 인프라 기술을 포함한 빅데이터와 연계된 기술들

표 1-10 인프라 기술을 포함한 빅데이터와 연계된 기술들 [11]

용어	설명
Cassandra <sup>카산드라</sup>	<ul style="list-style-type: none"><li>■ 분산 시스템에서 대용량 데이터를 처리할 수 있도록 설계된 오픈 소스 데이터베이스 관리 시스템</li><li>■ 원래 페이스북에서 개발했으며 지금은 아파치 소프트웨어 재단에서 한 프로젝트로 관리</li></ul>
Hadoop <sup>하둡</sup>	<ul style="list-style-type: none"><li>■ 분산 시스템에서 대용량 데이터 처리 분석을 지원하는 오픈 소스 소프트웨어 프레임워크</li><li>■ 구글이 개발한 맵리듀스를 오픈 소스로 구현한 결과물</li><li>■ 야후에서 최초로 개발했으며, 지금은 아파치 소프트웨어 재단에서 한 프로젝트로 관리</li><li>■ 주요 구성요소로는 하둡 분산 파일 시스템인 HDFS, 분산 컬럼 기반 데이터베이스인 HBase, 분산 컴퓨팅 지원 프레임워크인 맵리듀스 포함</li></ul>
HBase <sup>H베이스</sup>	<ul style="list-style-type: none"><li>■ 구글의 '빅테이블'을 참고로 개발된 오픈 소스 분산 비관계형 데이터베이스</li><li>■ 파워셋에서 개발했으며, 현재는 아파치 소프트웨어 재단에서 한 프로젝트로 관리</li></ul>
MapReduce <sup>맵리듀스</sup>	<ul style="list-style-type: none"><li>■ 분산 시스템에서 대용량 데이터 세트를 처리하려고 구글이 제안한 소프트웨어 프레임워크</li><li>■ 하둡에서도 구현</li></ul>
NoSQL	<ul style="list-style-type: none"><li>■ Not-only SQL 또는 No SQL을 의미</li><li>■ 전통적인 관계형 데이터베이스와 다르게 설계된 비관계형 데이터베이스</li><li>■ 대표적인 NoSQL 솔루션으로는 Cassandra, HBase, MongoDB 등이 있음</li></ul>



### 3. 빅데이터 기술

#### • 빅데이터 분석 기술

표 1-11 빅데이터 분석 기술

용어	설명
텍스트 마이닝 Text Mining	자연어 처리 Natural Language Processing 기술을 사용해 인간의 언어로 쓰인 비정형 텍스트에서 유용한 정보를 추출하거나 다른 데이터와의 연계성을 파악하며, 분류나 군집화 등 빅데이터에 숨겨진 의미 있는 정보를 발견하는 것
웹 마이닝 Web Mining	인터넷에서 수집한 정보를 데이터 마이닝 기법으로 분석하는 것
오피니언 마이닝 Opinion Mining; 평판 분석	<ul style="list-style-type: none"><li>■ 다양한 온라인 뉴스와 소셜 미디어 코멘트, 사용자가 만든 콘텐츠에서 표현된 의견을 추출 · 분류 · 이해하고 자산화하는 컴퓨팅 기술</li><li>■ 텍스트 속의 감성과 감동, 여러 가지 감정 상태를 식별하려고 감성 분석 사용</li><li>■ 마케팅에서는 버즈 Buzz 입소문 분석이라고도 함</li></ul>
리얼리티 마이닝 Reality Mining	<ul style="list-style-type: none"><li>■ 휴대폰 등 기기를 사용하여 인간관계와 행동 양태 등을 추론하는 것</li><li>■ 통화량, 통화 위치, 통화 상태, 대상, 내용 등을 분석하여 사용자의 인간관계, 행동 특성 등 정보를 찾아냄</li></ul>
소셜 네트워크 분석 Social Network Analysis	수학의 그래프 이론 Graph Theory을 바탕으로 소셜 네트워크 서비스에서 소셜 네트워크 연결 구조와 연결 강도를 분석하여 사용자의 명성 및 영향력을 측정하는 것
분류 Classification	<ul style="list-style-type: none"><li>■ 미리 알려진 클래스들로 구분되는 훈련 데이터군 Group을 학습시켜 새로 추가되는 데이터가 속할 만한 데이터군을 찾는 지도 학습 Supervised Learning 방법</li><li>■ 가장 대표적인 방법으로 KNN K-Nearest Neighbor이 있음</li></ul>

### 3. 빅데이터 기술

---

#### • 빅데이터 분석 기술

---

##### 군집화 Clustering

- 특성이 비슷한 데이터를 합쳐 군 Group으로 분류하는 학습 방법
- 분류와 달리 훈련 데이터를 이용하지 않기 때문에 비지도 학습 Unsupervised Learning 방법
- 트위터에서 주로 사진/카메라를 논의하는 사용자군과 게임에 관심 있는 사용자군 등 관심 사나 취미에 따라 분류

##### 기계 학습 Machine Learning

- 인공지능 분야에서 인간의 학습을 모델링한 것
- 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하여 수신한 이메일의 스팸 여부를 판단할 수 있도록 훈련
- 결정 트리 Decision Tree 등 기호적 학습, 신경망이나 유전자 알고리즘 등 비기호적 학습, 베이저안 Bayesian이나 은닉 마코프 Hidden Markov 등 확률적 학습 등 다양한 기법이 있음

##### 감성 분석 Sentiment Analysis

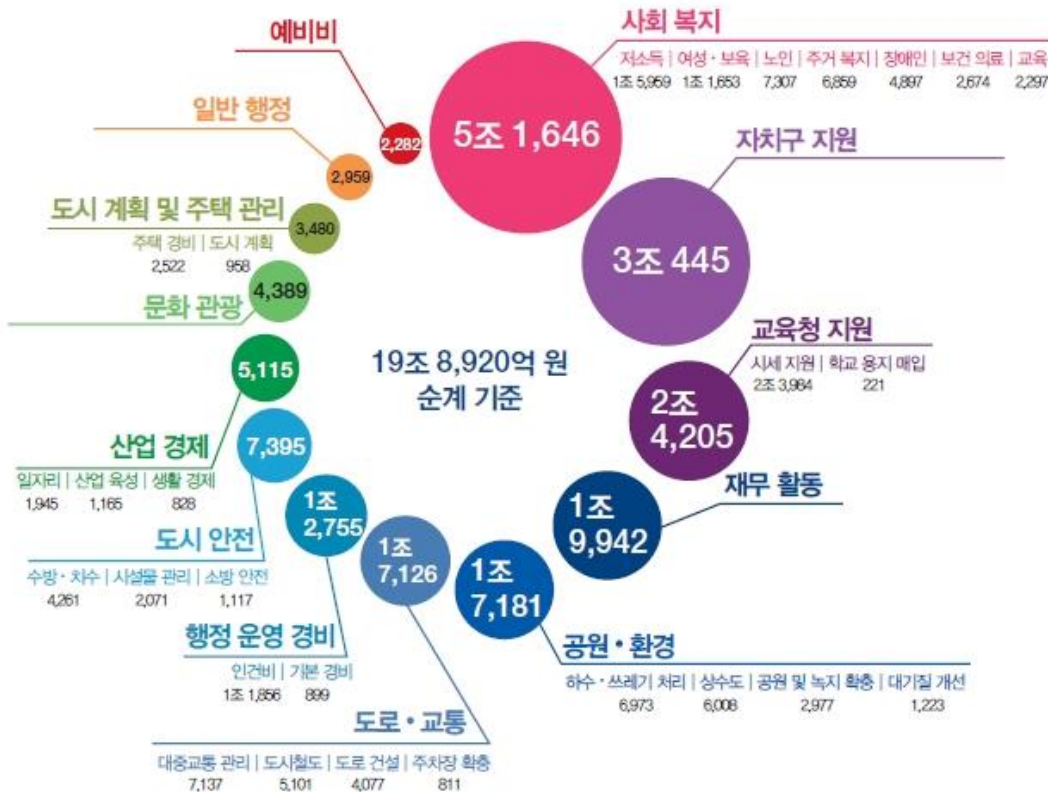
문장의 의미를 파악하여 글의 내용에 긍정/부정, 좋음/나쁨을 분류하거나 만족/불만족 강도를 지수화. 그런 다음 이 지수를 이용하여 고객의 감성 트렌드를 시계열적으로 분석하고 고객 감성 변화에 기업의 신속한 대응 및 부정적인 의견의 확산을 방지하는 데 활용

---



### 3. 빅데이터 기술

#### • 빅데이터 표현 기술



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	

순계 : 총예산에서 회계 간 중복분 제외한 실질적인 예산

그림 1-8 정보 표현의 간단한 예 [12]

## 4. 빅데이터 활용 사례

### • 빅데이터 활용 분야와 기대 효과

표 1-12 맥킨지에서 제시한 빅데이터 활용 분야 [13]

도메인	분석 대상 데이터	예상 효과
미국의 의료 산업	제약사 연구 개발 데이터, 환자 치료 · 임상 데이터, 의료 산업의 비용 데이터	연간 \$3조로, 0.7% 생산성 향상
유럽의 공공 행정	정부의 행정 업무에서 발생하는 데이터	연간 \$4.1조로, 0.5% 생산성 향상
소매업	고객의 거래 데이터, 구매 경향	<ul style="list-style-type: none"><li>▪ \$1조+서비스 업자 수익</li><li>▪ \$7조 소비자 이익</li></ul>
제조업	고객 취향 데이터, 수요 예측 데이터, 제조 과정 데이터, 센서 활용 데이터	<ul style="list-style-type: none"><li>▪ 60% 마진 증가</li><li>▪ 0.5~1.0% 생산성 향상</li></ul>
개인 위치 데이터	개인과 차량의 위치 데이터	<ul style="list-style-type: none"><li>▪ 개발 및 조립 비용 50% 감소</li><li>▪ 운전 자본 7% 감소</li></ul>

## 4. 빅데이터 활용 사례

---

### • 활용 분야

- **공공분야**- 국가적 차원에서 방대한 양의 데이터로 수자원 관리, 스마트 그리드, 재난 방재 영역 등을 포괄적으로 포함
- **과학분야** - 산발적으로 흩어진 과학 데이터를 국가 차원에서 수집, 가공, 유통, 재 활용할 수 있는 기반을 마련
- **의료 분야** - 의료 기록의 전자화, 병원 간 연구 데이터 공유로 빅데이터 도입과 활용이 확대됨
- **도소매 분야** - 이미 데이터를 활용 중이며 빅데이터 분석으로 수요 예측 및 선제적 경영 지원에 초점을 둠
- **제조분야** - 보유 데이터양 이 많고, 불량품 개선 비용 등 적용 효과를 계량화하여 빅데이터의 유용성을 확인할 수 있는 분야
- **정보 통신 분야** - 이동통신의 발전과 개인 단말기의 증가로 생성된 디지털 공간의 개인 데이터로 목표 마케팅, 개인화 서비스 확대

## 4. 빅데이터 활용 사례

---

- 기대 효과

- **이상현상 감지**- 업무에서 발생하는 이벤트를 기록하여 ‘정상 상태,’ ‘비정상 상태’를 표시하는 패턴을 파악하고, 이 패턴을 기초로 새로운 이벤트가 발생할 경우 이상 현상 여부를 판단함, 마케팅 분야에도 활용
- 고객 이탈을 사전에 감지한 T-Mobile
- 위키리크스 데이터 분석으로 효과적인 전술 정보 제공
- 아마존닷컴의 추천 상품 표시, 구글 및 페이스북의 맞춤형 광고

## 4. 빅데이터 활용 사례

---

- 빅데이터 시대 준비

- 데이터 경제 시대를 대비하는 ‘연결과 협력’
  - 데이터는 무한한 자원이나 활용 가능한 자원의 영역은 상호 연결과 협력으로 더욱 확장될 수 있음
- 빅데이터 핵심 역량인 ‘창의적 인력’의 양성
  - 빅데이터는 많은 데이터를 수집하는 것보다 무엇을 분석할 것인지 분명한 목적 의식과 통합적 사고력, 해석력이 중요함
  - 향후 미래에 데이터 관리자 등 분석 인력이 부족해짐에 따라 전문가 양성이 시급함
- 데이터 신뢰 환경의 구축
  - 데이터에 민감한 개인 사용자 정보의 노출 없이도 타당한 수준의 분석을 도출할 수 있는 방안을 고려해야 함