

# 빅데이터

Big data Framework

# 프레임워크



# 01

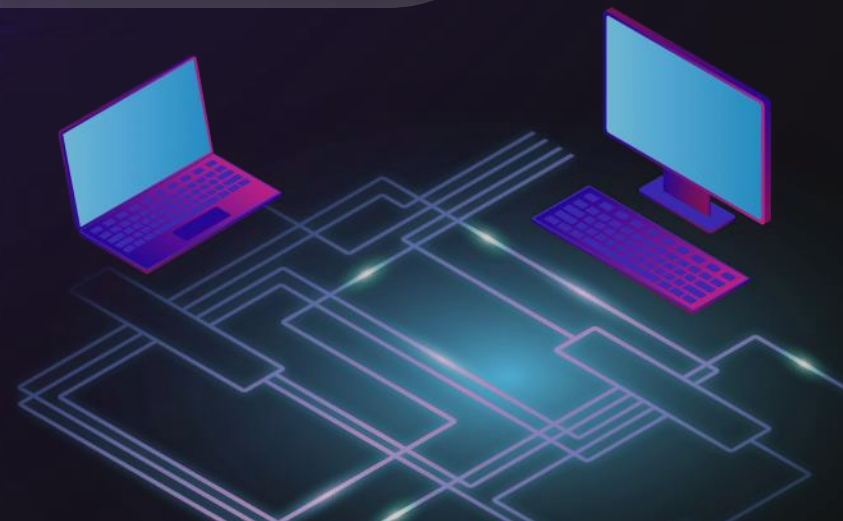
Big data Framework

## 빅데이터 처리 개요



01

# 빅데이터 처리 개요





# 01 | 빅데이터 처리 개요



빅데이터의 속성에 대해 식별할 수 있다.

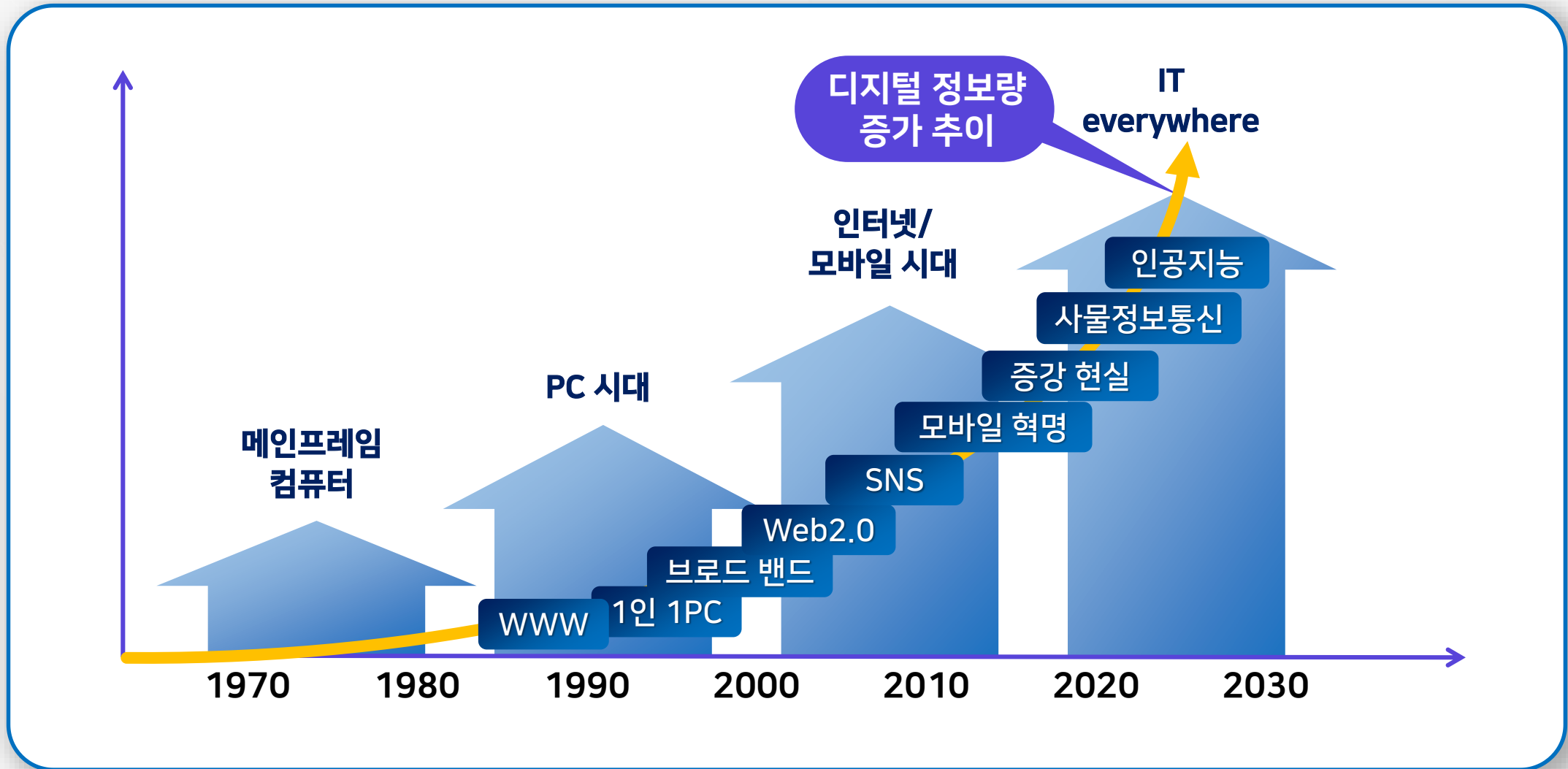


빅데이터의 종류에 대해 설명할 수 있다.

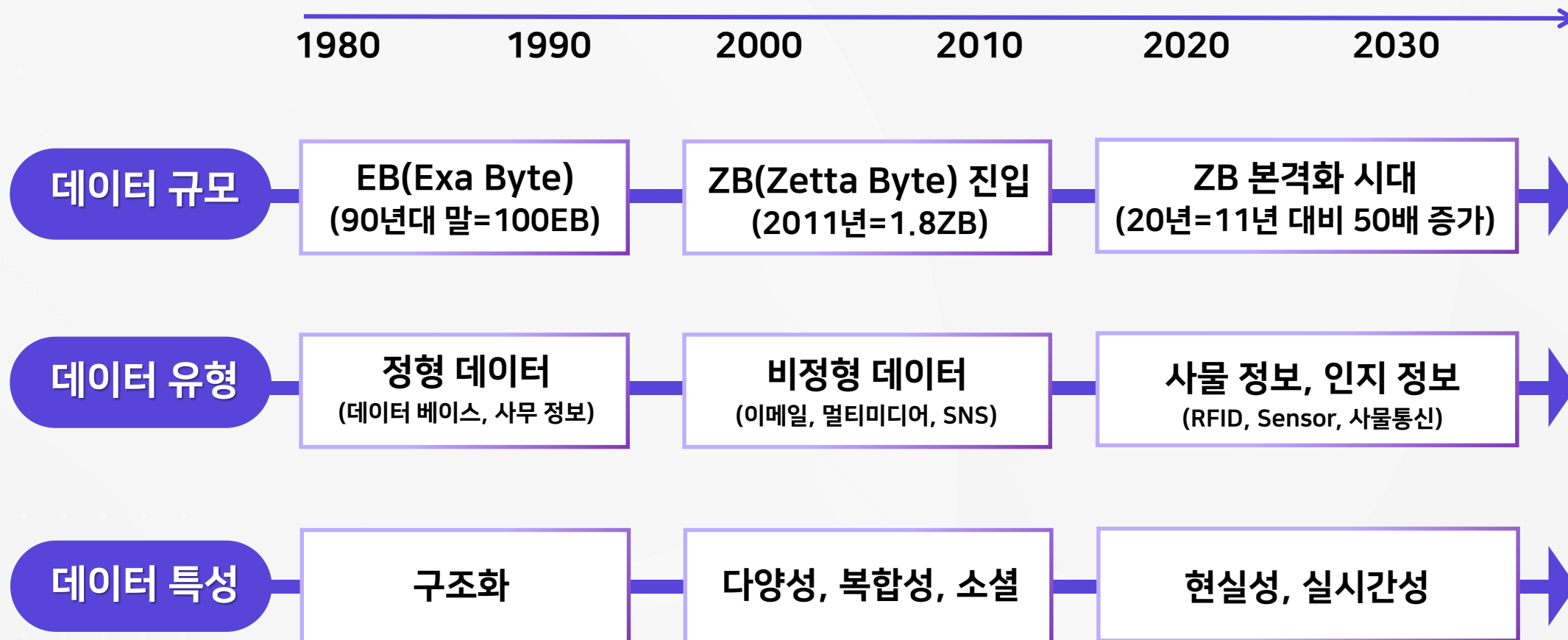


빅데이터 처리의 개념과 절차를 설명할 수 있다.

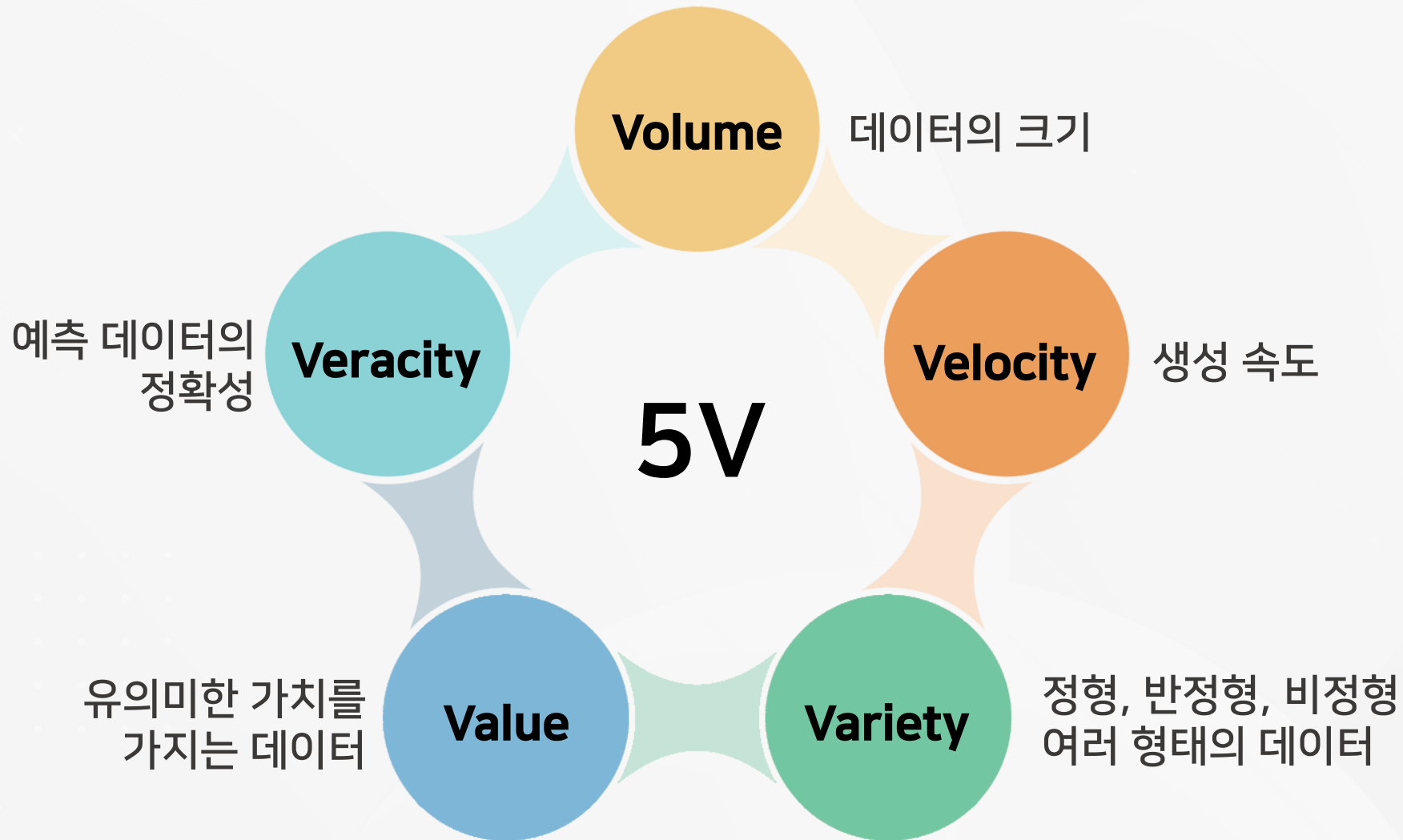
## 1 빅데이터의 등장 배경



## 1 빅데이터의 등장 배경



## 2 빅데이터의 속성



## 3 빅데이터의 속성에 필요한 기술

속성	특성	필요 기술
Volume	저장해야 할 데이터가 계속적으로 늘어남	분산파일 시스템
Variety	다양한 데이터가 생성됨	NoSQL
Velocity	빠른 처리속도를 필요로 함	분산병렬 처리 시스템
Value	시각화를 통해 분석 결과의 새로운 insight를 도출함	시각화 툴
Veracity	분석 정확도를 높여야 함	분석 시스템



### 1 빅데이터의 형태

#### 정형 데이터

- 행과 열을 갖는 표준 데이터 베이스와 같이 관계형 스키마로 구성된 데이터
- 간단한 질의문을 통해 원하는 정보를 획득해 활용

예) 데이터 베이스, 스프레드 시트 등

#### 반정형 데이터

- 고정된 스키마를 갖지 않는 정형 데이터
- 데이터에 레코드 및 필드와 같은 구조를 갖기 위해 태그나 인덱스 등을 포함

예) 시스템 로그, 센서 데이터, HTML 등

#### 비정형 데이터

구조화 되지 않은 임의의 형식

예) 이미지, 동영상, 이메일, 문서 등

### 1 빅데이터 처리의 개념과 절차



## 2 빅데이터의 처리 형태

속성	특성	필요 기술
Volume	저장해야 할 데이터가 계속적으로 늘어남	분산 파일 시스템
Variety	다양한 데이터가 생성됨	NoSQL
Velocity	빠른 처리속도를 필요로 함	분산 병렬 처리 시스템
Value	분석 결과를 시각화를 통한 새로운 insight 도출	시각화 툴
Veracity	분석 정확도를 높여야 함	분석 시스템

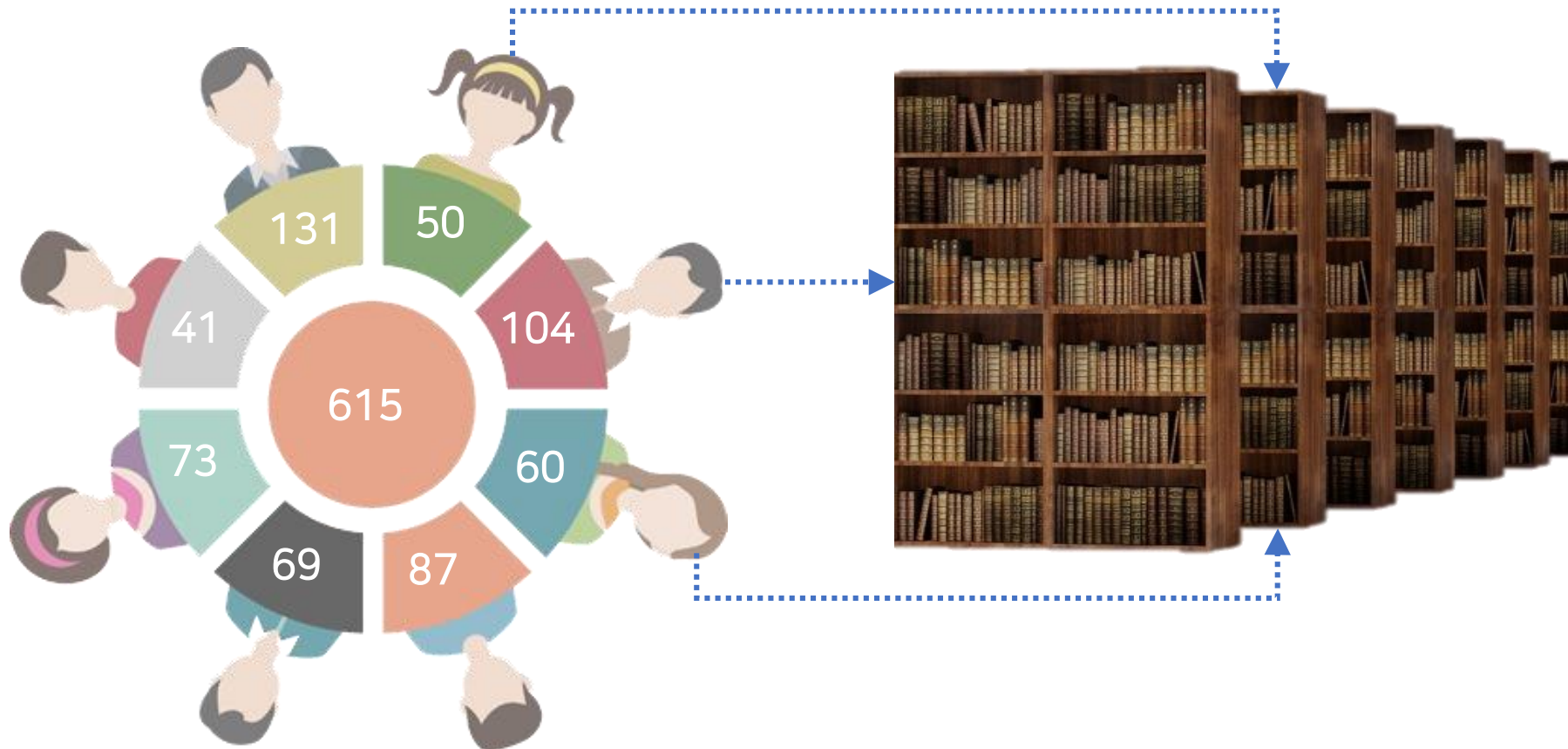
### 3 많은 양의 데이터는 어떻게 처리를 해야 하나?

“ 도서관에 비치된 서재 중, '사람'이라는 단어는 몇 개인지를 세어보자! ”



### 3 많은 양의 데이터는 어떻게 처리를 해야 하나?

🏠 여러 사람과 **협업하기** (여러 태스크를 사람들에게 분할한다면?)





### 3 많은 양의 데이터는 어떻게 처리를 해야 하나?

🏢 컴퓨터도 많은 양의 데이터를 처리하기 위해 협업을 함

- 이를 분산 처리라고 부름



### 3 많은 양의 데이터는 어떻게 처리를 해야 하나?

여러 뇌를 동시에 사용하기



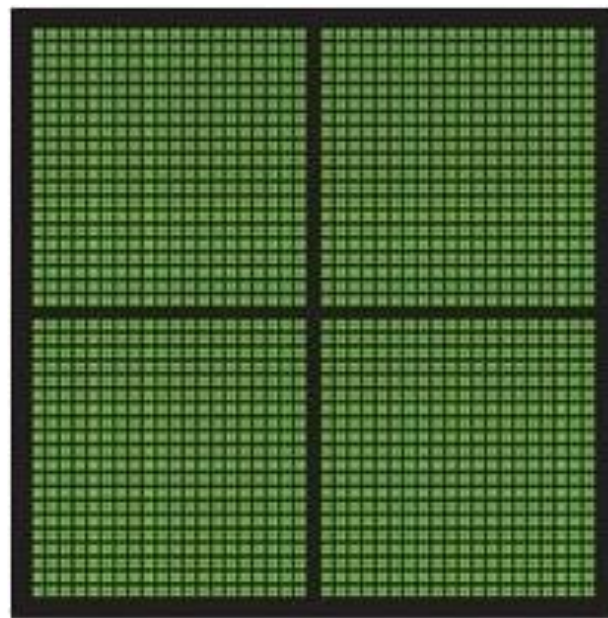
### 3 많은 양의 데이터는 어떻게 처리를 해야 하나?

💡 컴퓨터도 여러 Core를 이용해 여러 태스크를 동시에 처리

- 이를 '병렬 처리'라고 부름



CPU  
MULTIPLE CORES



GPU  
THOUSANDS OF CORES

## 4 분산 컴퓨팅과 병렬 컴퓨팅의 차이

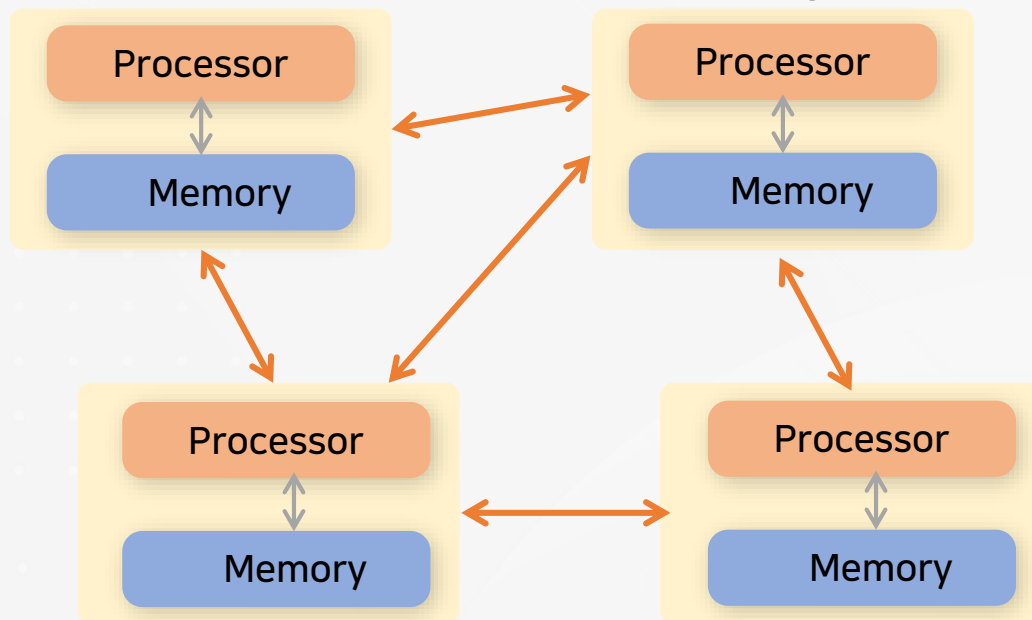
### 분산 컴퓨팅

네트워크로 연결된 여러 대의 컴퓨터가 마치 하나의 컴퓨팅 자원처럼 연산을 수행

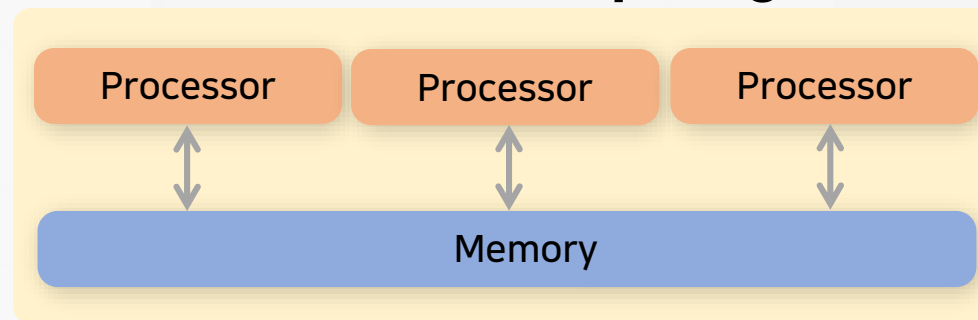
### 병렬 컴퓨팅

메모리를 공유하는 여러 개의 프로세서가 동시에 연산을 수행

### Distributed Computing



### Parallel Computing



## 1 빅데이터의 속성

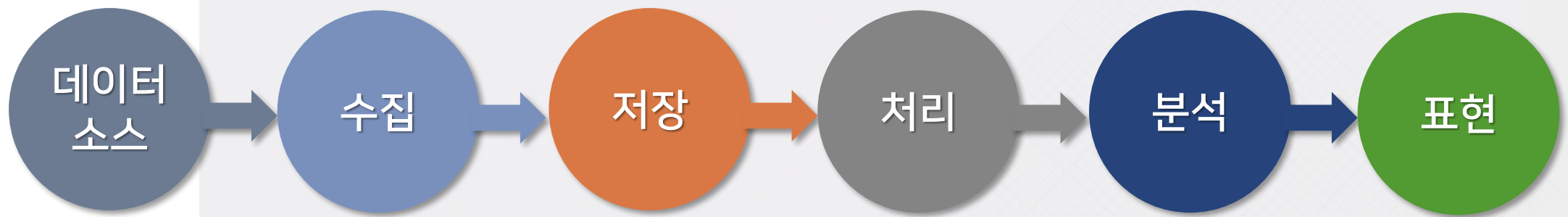
- Volume
- Veracity
- Velocity
- Variety
- Value

## 2 빅데이터의 종류

- 비정형 데이터
- 반정형 데이터
- 정형 데이터



## 1 빅데이터 처리의 절차



- ✎ 빅데이터 컴퓨팅 기술, 2014, 박두순, 한빛미디어
- ✎ 정지선, 가치창출 엔진, 빅 데이터의 새로운 가능성과 대응 전략, 2011, 한국정보화진흥원, IT & Future Strategy