



## 데이터 특징

<http://www.dbguide.net>

- 데이터가 가진 특징을 명확히 인지할 때,  
데이터를 처리할 수 있는 기술적 사항을 고려하여 설계 가능

### 일반적인 데이터 특징

- 존재론적 특징
- 목적론적 특징

### 수집 단계에서 데이터 특징

- 가역적 데이터
- 불가역적 데이터

## 데이터 특징

<http://www.dbguide.net>

- 데이터가 가진 특징을 명확히 인지할 때,  
데이터를 처리할 수 있는 기술적 사항을 고려하여 설계 가능

### 일반적인 데이터 특징

- 존재론적 특징
- 목적론적 특징

### 수집 단계에서 데이터 특징

- 가역적 데이터
- 불가역적 데이터

# 데이터의 존재론적 특징

<http://www.dbguide.net>

- 정성적 데이터(qualitative data)와 정량적 데이터(quantitative data)로 구분
- 정성적 데이터는 데이터 자체가 하나의 텍스트(“환율이 내리고 있어 올해 목표한 수출 목표의 조기 달성이 가능해 보인다.”)를 이루고 있기 때문에 데이터 하나 하나가 함축된 정보를 가짐
- 정량적 데이터는 여러 속성(이름, 나이, 성별, 주소 등)이 모여 하나의 객체를 형성하고, 각 속성은 속성 하나 혹은 여러 개의 속성이 결합해 측정이나 설명이 가능

| 구분    | 정성적 데이터(Qualitative Data) | 정량적 데이터(Quantitative Data) |
|-------|---------------------------|----------------------------|
| 형태    | 비정형 데이터                   | 정형, 반정형 데이터                |
| 특징    | 객체 하나에 함의된 정보를 가짐         | 속성이 모여 객체를 이룸              |
| 구성    | 언어, 문자 등으로 이루어짐           | 수치, 도형, 기호 등으로 이루어짐        |
| 저장 형태 | 파일, 웹                     | 데이터베이스, 스프레드시트             |
| 소스 위치 | 외부 시스템(주로 소셜 데이터)         | 내부 시스템(주로 DBMS)            |



# 데이터의 존재론적 특징

<http://www.dbguide.net>

- 정성적 데이터(qualitative data)와 정량적 데이터(quantitative data)로 구분
- 정성적 데이터는 데이터 자체가 하나의 텍스트(“환율이 내리고 있어 올해 목표한 수출 목표의 조기 달성이 가능해 보인다.”)를 이루고 있기 때문에 데이터 하나 하나가 함축된 정보를 가짐
- 정량적 데이터는 여러 속성(이름, 나이, 성별, 주소 등)이 모여 하나의 객체를 형성하고, 각 속성은 속성 하나 혹은 여러 개의 속성이 결합해 측정이나 설명이 가능

| 구분    | 정성적 데이터(Qualitative Data) | 정량적 데이터(Quantitative Data) |
|-------|---------------------------|----------------------------|
| 형태    | 비정형 데이터                   | 정형, 반정형 데이터                |
| 특징    | 객체 하나에 함의된 정보를 가짐         | 속성이 모여 객체를 이룸              |
| 구성    | 언어, 문자 등으로 이루어짐           | 수치, 도형, 기호 등으로 이루어짐        |
| 저장 형태 | 파일, 웹                     | 데이터베이스, 스프레드시트             |
| 소스 위치 | 외부 시스템(주로 소셜 데이터)         | 내부 시스템(주로 DBMS)            |

# 데이터의 존재론적 특징

<http://www.dbguide.net>

- 정성적 데이터(qualitative data)와 정량적 데이터(quantitative data)로 구분
- 정성적 데이터는 데이터 자체가 하나의 텍스트("환율이 내리고 있어 올해 목표한 수출 목표의 조기 달성이 가능해 보인다.")를 이루고 있기 때문에 데이터 하나 하나가 함축된 정보를 가짐
- 정량적 데이터는 여러 속성(이름, 나이, 성별, 주소 등)이 모여 하나의 객체를 형성하고, 각 속성은 속성 하나 혹은 여러 개의 속성이 결합해 측정이나 설명이 가능

| 구분    | 정성적 데이터(Qualitative Data) | 정량적 데이터(Quantitative Data) |
|-------|---------------------------|----------------------------|
| 형태    | 비정형 데이터                   | 정형, 반정형 데이터                |
| 특징    | 객체 하나에 함의된 정보를 가짐         | 속성이 모여 객체를 이룸              |
| 구성    | 언어, 문자 등으로 이루어짐           | 수치, 도형, 기호 등으로 이루어짐        |
| 저장 형태 | 파일, 웹                     | 데이터베이스, 스프레드시트             |
| 소스 위치 | 외부 시스템(주로 소셜 데이터)         | 내부 시스템(주로 DBMS)            |

# 데이터의 존재론적 특징

<http://www.dbguide.net>

- 정성적 데이터(qualitative data)와 정량적 데이터(quantitative data)로 구분
- 정성적 데이터는 데이터 자체가 하나의 텍스트(“환율이 내리고 있어 올해 목표한 수출 목표의 조기 달성이 가능해 보인다.”)를 이루고 있기 때문에 데이터 하나 하나가 함축된 정보를 가짐
- 정량적 데이터는 여러 속성(이름, 나이, 성별, 주소 등)이 모여 하나의 객체를 형성하고, 각 속성은 속성 하나 혹은 여러 개의 속성이 결합해 측정이나 설명이 가능

| 구분    | 정성적 데이터(Qualitative Data) | 정량적 데이터(Quantitative Data) |
|-------|---------------------------|----------------------------|
| 형태    | 비정형 데이터                   | 정형, 반정형 데이터                |
| 특징    | 객체 하나에 함의된 정보를 가짐         | 속성이 모여 객체를 이룸              |
| 구성    | 언어, 문자 등으로 이루어짐           | 수치, 도형, 기호 등으로 이루어짐        |
| 저장 형태 | 파일, 웹                     | 데이터베이스, 스프레드시트             |
| 소스 위치 | 외부 시스템(주로 소셜 데이터)         | 내부 시스템(주로 DBMS)            |

## 데이터의 목적론적 특징

<http://www.dbguide.net>

- '데이터'라고 인식하게 되는 객체가 갖고 있는 존재 목적
- 하나의 서비스 혹은 활용(데이터 분석)을 위해 데이터가 존재하고 있다는 목적 인식
- 데이터에서 목적에 맞는 특징을 찾아내는 것
- 데이터는 관점에 따라 여러 종류의 데이터로 인식 가능
- 원본 데이터의 속성 분리, 다른 데이터와의 병합이 발생해 새로운 데이터 객체 생성



## 수집 데이터의 특징 정의

<http://www.dbguide.net>

- 수집에서 일어나는 모든 과정은 데이터를 활용하기 위한 재생산 과정
- 재생산은 원천 데이터를 탐색 · 수집 · 정제 · 저장하는 과정을 거쳐 새로운 데이터를 생산하는 것을 의미
- 수집된 데이터는 원본 데이터와 재생산 데이터의 관계를 고려할 필요가 있음
- 재생산된 데이터는 가역적 데이터와 불가역적 데이터로 구분

## 수집 데이터의 특징 정의

<http://www.dbguide.net>

- 수집에서 일어나는 모든 과정은 데이터를 활용하기 위한 재생산 과정
- 재생산은 원천 데이터를 탐색 · 수집 · 정제 · 저장하는 과정을 거쳐 새로운 데이터를 생산하는 것을 의미
- 수집된 데이터는 원본 데이터와 재생산 데이터의 관계를 고려할 필요가 있음
- 재생산된 데이터는 가역적 데이터와 불가역적 데이터로 구분

# 가역적 데이터 vs. 불가역적 데이터

<http://www.dbguide.net>

## ■ 가역적 데이터

- 생산된 데이터의 원본으로 일정 수준 환원이 가능한 데이터
- 데이터 원본의 내용이 변경되었을 경우 변경 이력에 대한 추적이 가능
- 원본 데이터의 흔적이 재생산된 데이터에도 남아 있음

## ■ 불가역적 데이터

- 생산된 데이터의 원본으로 환원이 불가능한 데이터
- 원본 데이터와는 전혀 다른 형태로 재생산되기 때문에 재생산된 데이터의 원본 데이터에 대한 추적이 불가능
- 원본 데이터의 내용이 변경되었더라도 재생산된 데이터에는 아무런 영향을 미치지 않음

| 구분          | 가역적 데이터          | 불가역적 데이터       |
|-------------|------------------|----------------|
| 환원 가능성      | 가능               | 불가능            |
| 의존성         | 원본 데이터에 의존       | 원본 데이터에 독립적    |
| 원본 데이터와의 관계 | 1:1              | 1:N 혹은 N:1     |
| 처리 과정       | 탐색               | 병합             |
| 활용 분야       | 데이터 웨어하우징, 로그 수집 | 소셜 분석, 텍스트 마이닝 |

# 가역적 데이터 vs. 불가역적 데이터

<http://www.dbguide.net>

## ■ 가역적 데이터

- 생산된 데이터의 원본으로 일정 수준 환원이 가능한 데이터
- 데이터 원본의 내용이 변경되었을 경우 변경 이력에 대한 추적이 가능
- 원본 데이터의 흔적이 재생산된 데이터에도 남아 있음

## ■ 불가역적 데이터

- 생산된 데이터의 원본으로 환원이 불가능한 데이터
- 원본 데이터와는 전혀 다른 형태로 재생산되기 때문에 재생산된 데이터의 원본 데이터에 대한 추적이 불가능
- 원본 데이터의 내용이 변경되었더라도 재생산된 데이터에는 아무런 영향을 미치지 않음

| 구분          | 가역적 데이터          | 불가역적 데이터       |
|-------------|------------------|----------------|
| 환원 가능성      | 가능               | 불가능            |
| 의존성         | 원본 데이터에 의존       | 원본 데이터에 독립적    |
| 원본 데이터와의 관계 | 1:1              | 1:N 혹은 N:1     |
| 처리 과정       | 탐색               | 병합             |
| 활용 분야       | 데이터 웨어하우징, 로그 수집 | 소셜 분석, 텍스트 마이닝 |



## 가역적 데이터 vs. 불가역적 데이터

<http://www.dbguide.net>

### ■ 가역적 데이터

- 생산된 데이터의 원본으로 일정 수준 환원이 가능한 데이터
- 데이터 원본의 내용이 변경되었을 경우 변경 이력에 대한 추적이 가능
- 원본 데이터의 흔적이 재생산된 데이터에도 남아 있음

### ■ 불가역적 데이터

- 생산된 데이터의 원본으로 환원이 불가능한 데이터
- 원본 데이터와는 전혀 다른 형태로 재생산되기 때문에 재생산된 데이터의 원본 데이터에 대한 추적이 불가능
- 원본 데이터의 내용이 변경되었더라도 재생산된 데이터에는 아무런 영향을 미치지 않음

| 구분          | 가역적 데이터          | 불가역적 데이터       |
|-------------|------------------|----------------|
| 환원 가능성      | 가능               | 불가능            |
| 의존성         | 원본 데이터에 의존       | 원본 데이터에 독립적    |
| 원본 데이터와의 관계 | 1:1              | 1:N 혹은 N:1     |
| 처리 과정       | 탐색               | 병합             |
| 활용 분야       | 데이터 웨어하우징, 로그 수집 | 소셜 분석, 텍스트 마이닝 |

# 데이터 수집 정의

<http://www.dbguide.net>

- 서비스 활용에 필요한 데이터를 시스템의 내부 혹은 외부에서 주기성을 갖고 필요한 형태로 수집하는 활동



# 데이터 수집 정의

<http://www.dbguide.net>

- 서비스 활용에 필요한 데이터를 시스템의 내부 혹은 외부에서 주기성을 갖고 필요한 형태로 수집하는 활동



## 서비스 활용

<http://www.dbguide.net>

- 서비스 활용에서 데이터 수집의 역할은 서비스의 품질을 결정할 뿐 아니라 서비스의 생명주기에도 영향을 미침
- 어떤 서비스를 할 것인지 결정했으면 먼저 수집할 원천 데이터 탐색 필요
- 데이터 탐색 과정에서 고려해야 할 사항
  - 수집의 난이도적 측면과 비용적 측면
  - 데이터 수집의 안정성



# 데이터 위치

<http://www.dbguide.net>

- 수집 데이터의 위치에 따라 구축하려는 시스템 내부 혹은 외부의 데이터로 구분

## 내부 데이터

- 내부 시스템에서 데이터를 가져오는 경우
- 대부분 정형 데이터인 경우가 많음
- 외부 데이터 수집과 비교하면 수집에 드는 비용이나 난이도 측면에서 훨씬 유리
- 서비스의 생명주기 관리에 용이

## 외부 데이터

- 협약된 곳에서 수집하는 데이터를 제외하고는 대부분이 비정형 데이터
- 수집에 드는 비용이나 난이도가 내부 데이터 수집과 비교하면 상대적으로 어려움
- 외부 환경은 내부 시스템과는 다르게 통제가 불가능하기 때문에 서비스의 수명주기 관리가 원천 데이터에 의존할 경우 수집이 불가능
- 원천 데이터를 수집할 수 없을 경우에 대비한 서비스 관리 정책이 별도로 필요

# 주기성

<http://www.dbguide.net>

- 데이터 수집의 주기성을 나눌 때 일반적으로 배치(일괄) 혹은 실시간이냐에 따라 주기를 결정 (서비스의 활용적인 측면이지 원본 데이터의 생명주기에 관한 것이 아님)
- 데이터의 수집은 다른 곳에서 생성한 데이터를 복제하는 행위
- 주기를 갖고 계속 수집하는 행위인지 아니면 한 번의 수집으로 끝나는 과정인지를 먼저 고려해야 함
- 일회성 수집이 아니고 일정한 주기를 기준으로 수집 활동이 진행되어야 한다면, 서비스의 종류 및 환경에 따라 실시간으로 수집할 것인지 아니면 배치 작업을 통해 수집할 것인가를 결정해야 함

## 수집 데이터의 저장 형태

<http://www.dbguide.net>

- 서비스 활용에 대해 수집 대상 데이터의 위치 · 주기 · 수집방법이 결정됐으면, 수집한 데이터를 어떠한 형태로 저장할 것인지를 고려해야 됨
- 빅데이터 처리에서 일반적으로 사용되는 저장소는 분산 파일 시스템이지만, 수집된 데이터를 가공 · 처리하기 위해서 DBMS가 사용될 수도 있고, 서비스를 DBMS를 통해 제공할 수도 있으므로 서비스 환경에 맞는 아키텍처를 설계해야 함
- 원본 데이터의 형태에 따라 사용되는 수집 기술이 다르고 수집되는 형태도 다르기 때문에 데이터 저장소의 아키텍처 설계 시 파일 시스템, 관계형 데이터베이스, 분산처리 DBMS 등을 고려해야 함

