



데이터 수집



데이터 가공



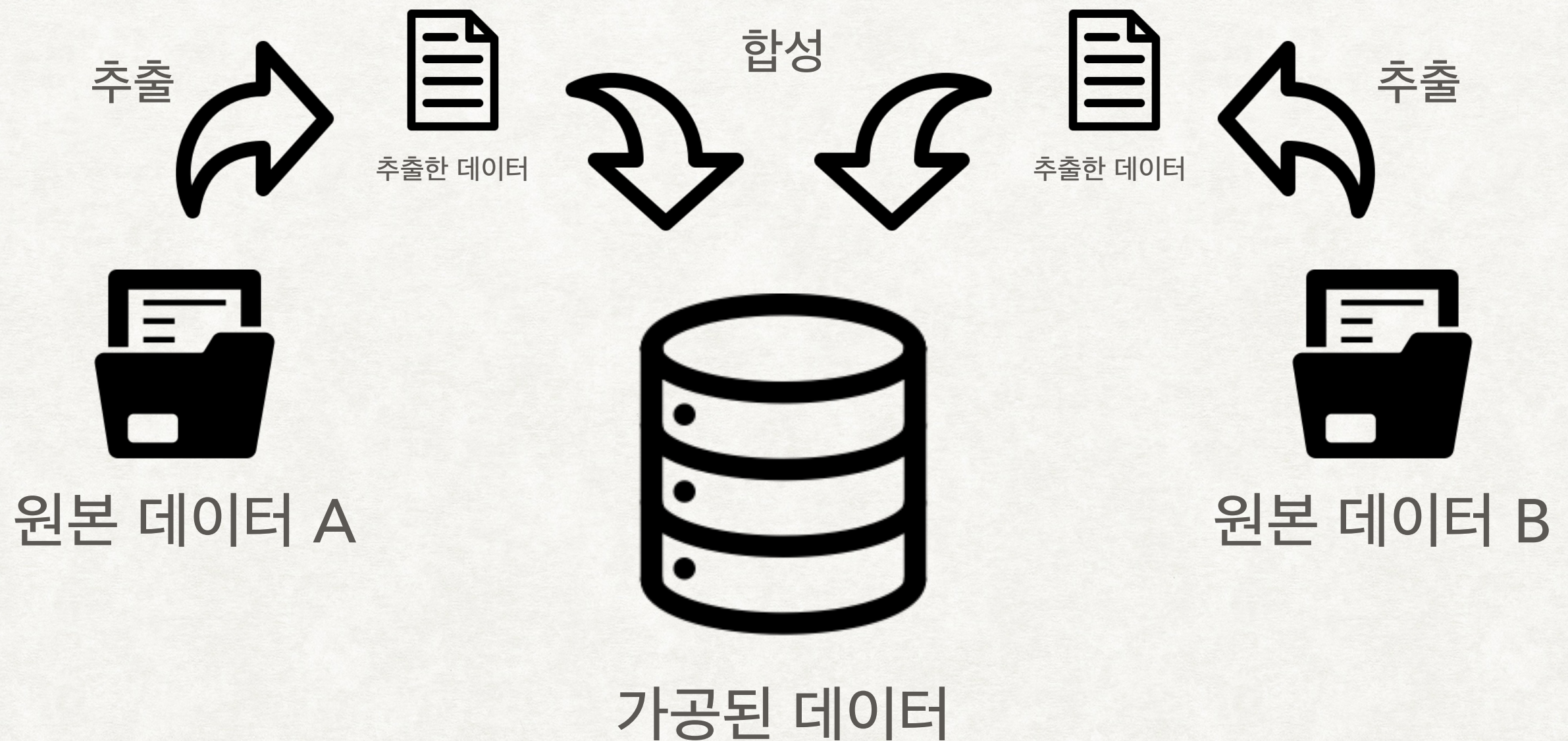
데이터 분석



결론 도출

데이터 가공

데이터 가공



- 원본 데이터들을 분석하기 좋게 처리하는 과정(전처리)

파일에서 데이터 불러오기

- `import numpy as np`
 - 파이썬에서 수치연산을 쉽고 편하게 구현하기 위한 라이브러리
- `import pandas as pd`
- `data = pd.read_csv(파일경로).to_numpy()`
 - csv파일에서 데이터 읽어오기
- `print(data)`
 - 데이터 출력

행, 열 기준으로 데이터 추출하기

`data[행범위, 열범위]`

`data[1:7, 2:7]`

2:7

	00	01	02	03	04	05	06	07	08	09	
	10	11	12	13	14	15	16	17	18	19	
	20	21	22	23	24	25	26	27	28	29	
	30	31	32	33	34	35	36	37	38	39	
	40	41	42	43	44	45	46	47	48	49	
	50	51	52	53	54	55	56	57	58	59	
	60	61	62	63	64	65	66	67	68	69	
	70	71	72	73	74	75	76	77	78	79	
	80	81	82	83	84	85	86	87	88	89	
	90	91	92	93	94	95	96	97	98	99	

1:7

data =

- **주의할 점!** 행, 열 기준으로 자르는 연산은 numpy 배열만 가능(리스트 불가)

데이터 모양 바꾸기

`reshape(차원 모양)`

`data[:, 0]`

`data = [00 10 20 30 40 50 60 70 80 90]`

`data.shape`

`(10,)`

`data.reshape(5, 2)`

`data =`
$$\begin{bmatrix} 00 & 10 \\ 20 & 30 \\ 40 & 50 \\ 60 & 70 \\ 80 & 90 \end{bmatrix}$$

- 데이터는 유지하면서 데이터의 형상(모양)만 바꾸는 함수

행, 열 합치기

`hstack, vstack`

$$a = \begin{bmatrix} 12 & 13 & 14 \\ 22 & 23 & 24 \end{bmatrix} \quad b = \begin{bmatrix} 57 & 58 & 59 \\ 67 & 68 & 69 \end{bmatrix}$$

$$hstack(a, b) \Rightarrow \begin{bmatrix} 12 & 13 & 14 & 57 & 58 & 59 \\ 22 & 23 & 24 & 67 & 68 & 69 \end{bmatrix}$$

$$vstack(a, b) \Rightarrow \begin{bmatrix} 12 & 13 & 14 \\ 22 & 23 & 24 \\ 57 & 58 & 59 \\ 67 & 68 & 69 \end{bmatrix}$$