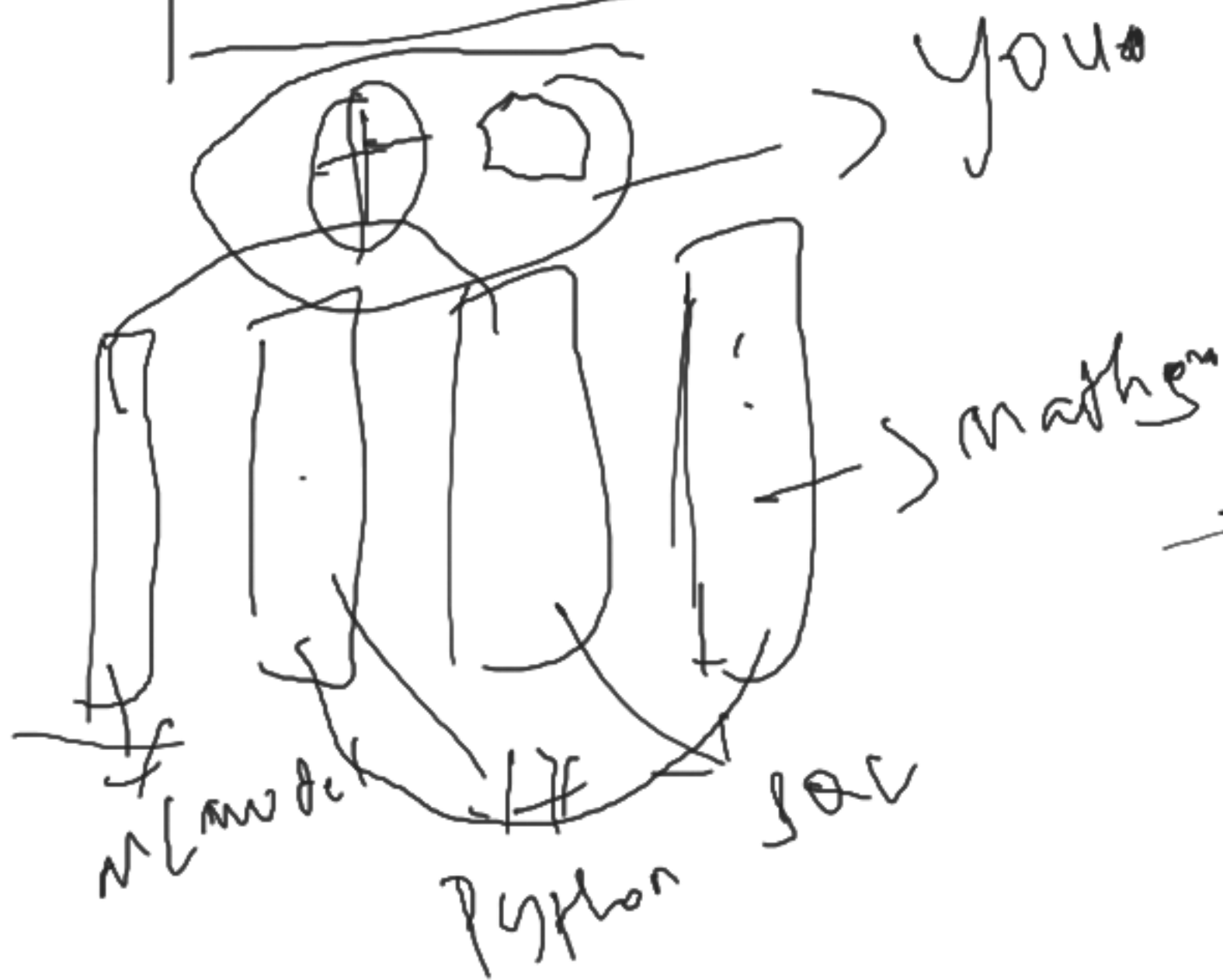




VSc-case

Output

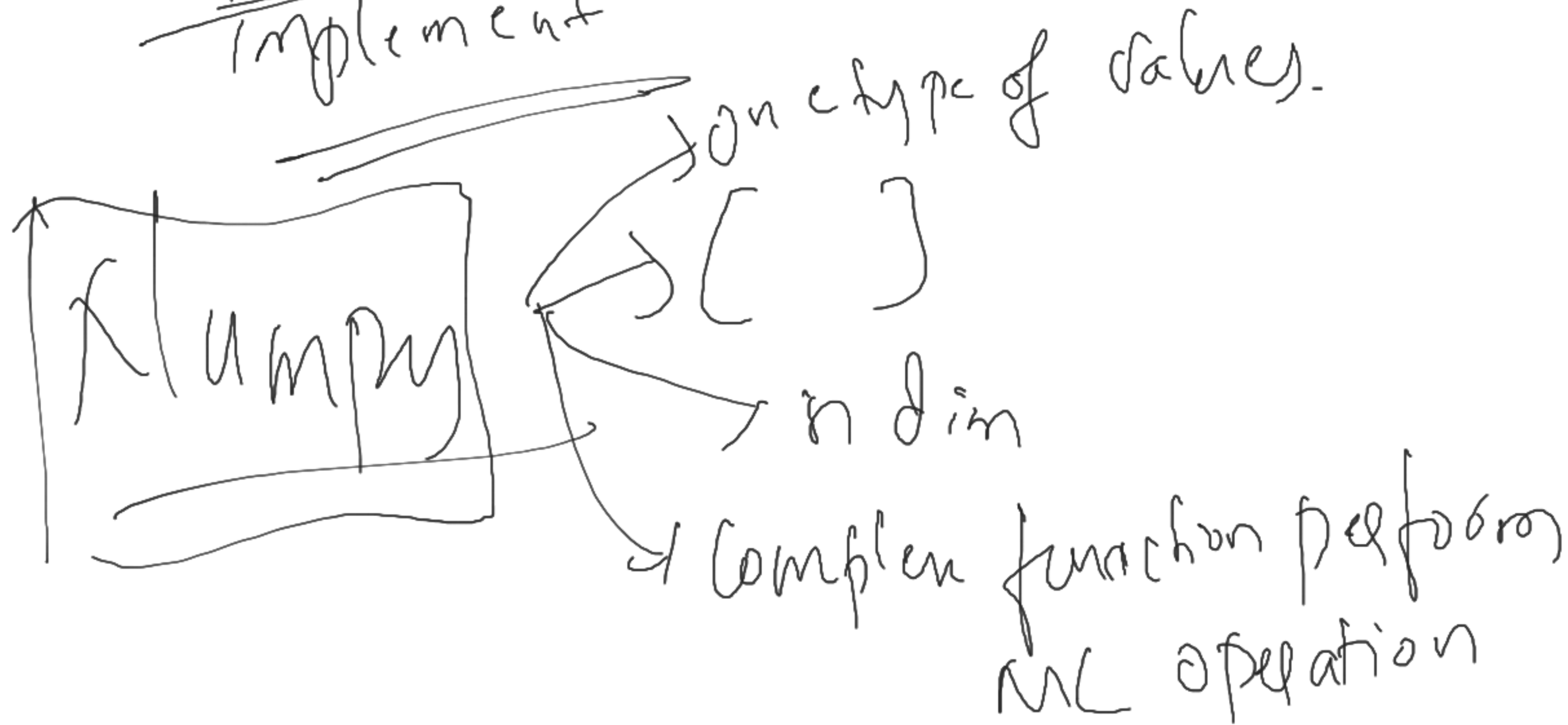


tools

Basics

foundation

Pandas is very easy understand &  
implement



Student ID	Gender	Marks Scored
121010	M	80
121011	F	70
121012	F	60
121013	M	60
121014	M	70
121015	F	90

String

numbers.

Republic =

[M, 80], -> String  
 [F, 70], -> String  
 [F, 60], Casting  
 [F, 90]

defaut 10

id	Gender	Studento.
0	M	80
1	F	70
2	M	60
3	R	50

Gender  
Studento.

row index & column index

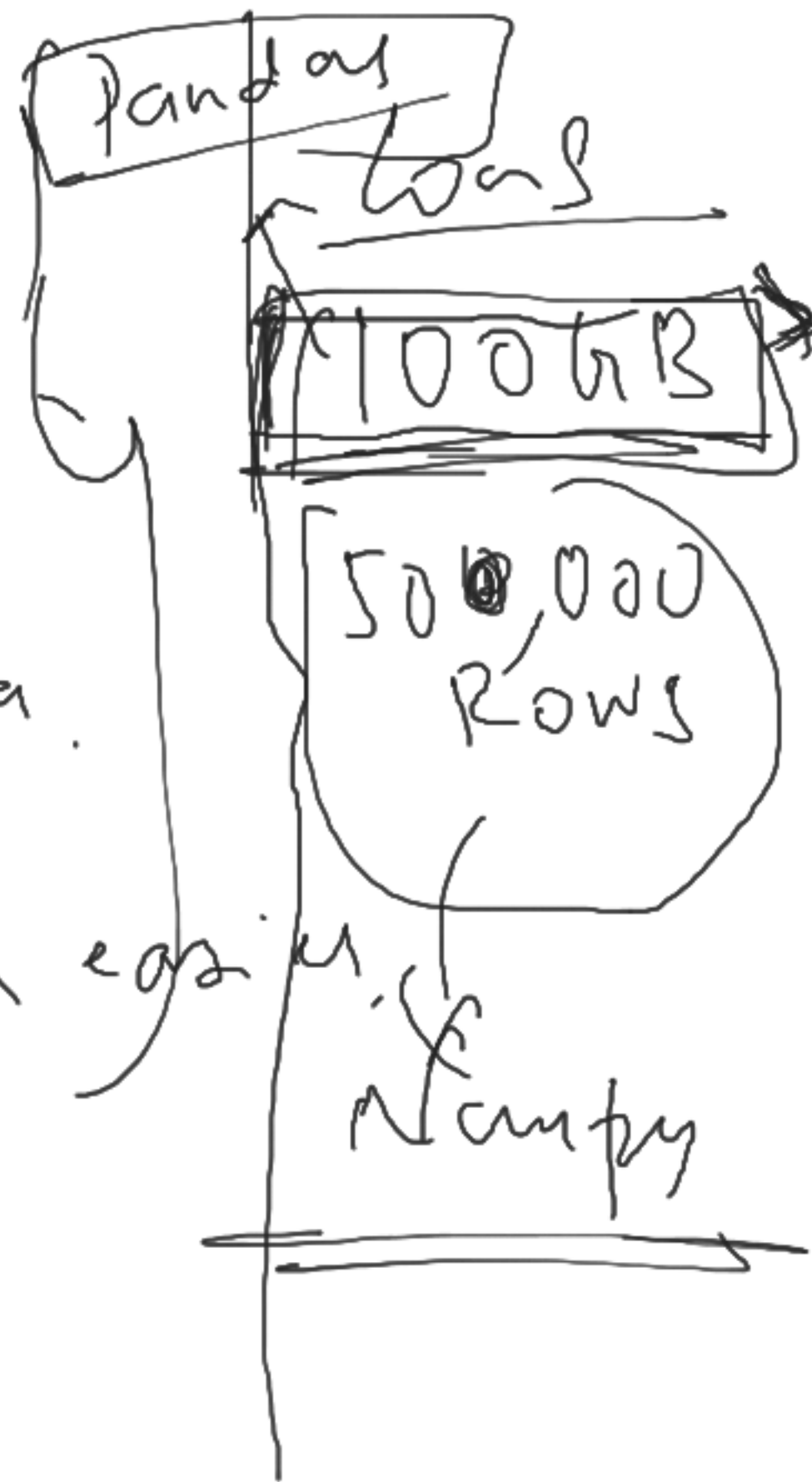
= Republic will not have names

Row Column.

Republic

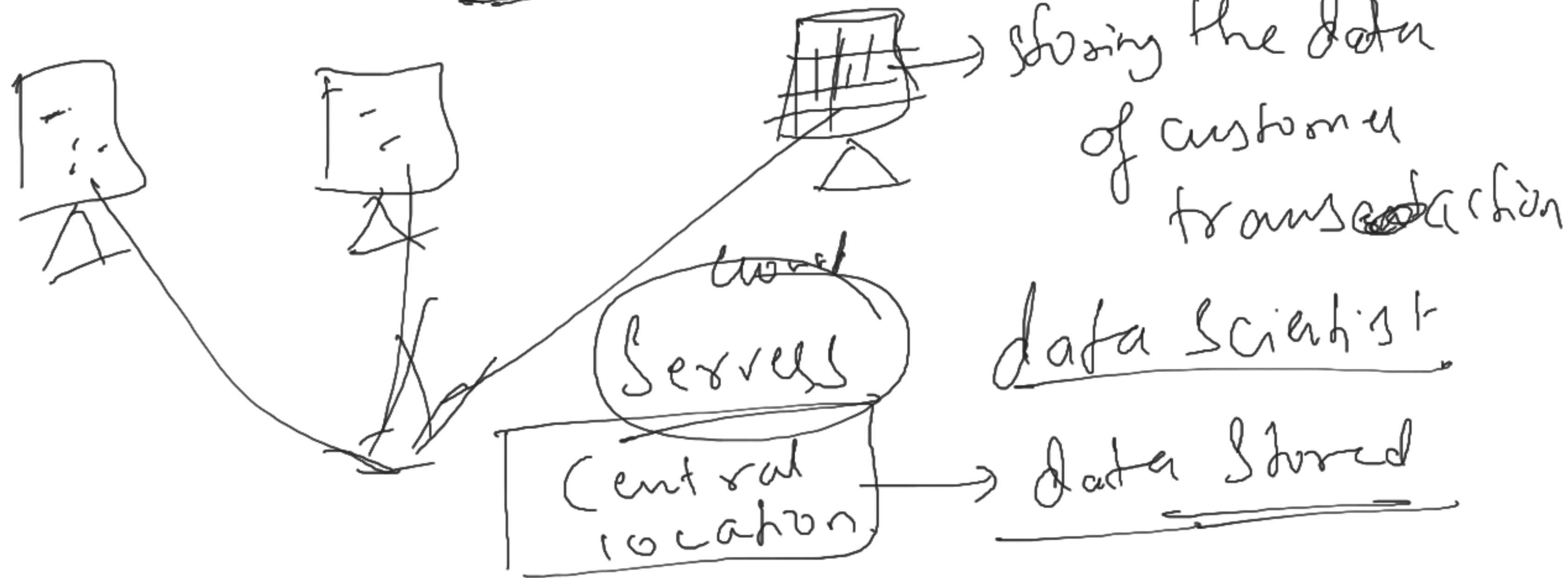


- i ~~different~~ Different datatypes.
- ii Naming the index
- iii Neat & Clean display of data.
- iv Cleaning up data is much easier.
- ~~v~~



Storing the data.

→ More / DMART Beng



Cleaning → 12m das stored

3 item

customer 2	milic	40
customer 1	Curd	50
customer 1	bread	30
customer 2	Biscuit	20
" 2	carpet	100

Cleaning

Sum / Avg

min / max

Aggregated

cur  
cur  
cur

3

2

Total  
120  
120

10GB < <sup>works</sup> faster

10GB > slows down

Pandas

sk

10GB

60GB

95GB

Pandas

TB

Databricks

Ray

TensorFlow

Administration



# Type of data

100%

op ① - Structured

unstructure. <sup>img</sup> <sub>man</sub>

Text  
Image  
Audio  
Video

9%

Semi-Structured

text message  
email Boxes

Text

2%

Structured

Strnd 1

M1	M2	M1	M4

90%

index  
[3, 2, 0, 1] → 1D → numpy array  
column with no name

Apples

3
2
0
1

column name  
= Pandas Series

$\begin{bmatrix} 1 \\ 3, 0 \end{bmatrix}$ ,  
 $\rightarrow [2, 3]$ ,  
 $\rightarrow [0, 7]$ ,  
 $\rightarrow [1, 2]$

$\rightarrow$  numpy array with no. column many

$\boxed{2D}$

$\downarrow \downarrow$  (column)

	Apple	Orange	
rows $\rightarrow 0$	3	0	
$\rightarrow 1$	2	3	
$\rightarrow 2$	0	7	
$\rightarrow 3$	1	2	

Pandas DataFrame (2D)

$\boxed{(4, 2)}$

$\rightarrow 3$

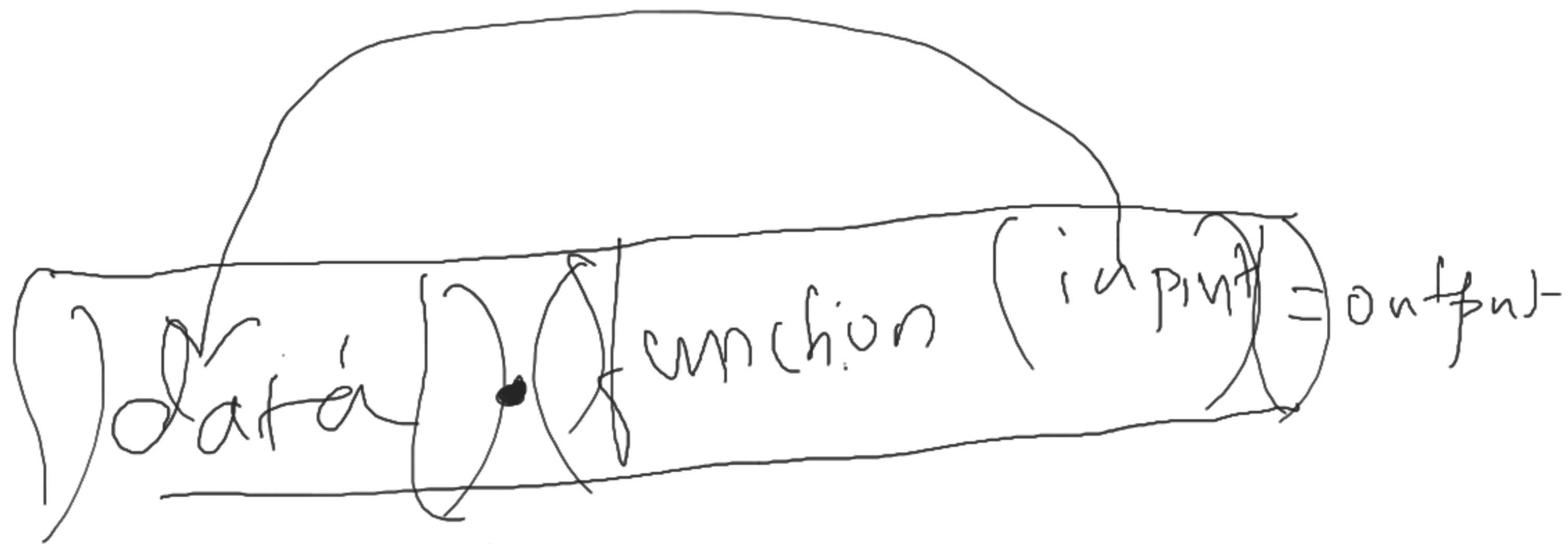
Numpy arrays  
list of list } → pd.DataFrame(arrayd  
list, column)

Dictionary → pd.DataFrame(dictionary)

keys → columns

value → column values.

---



Pipeline.

df.rename(columns = { Old Name : New name  
↓  
with error

inplace = True).

index

• head

• tail

• info

• column

• rename

→ Pandas attributes

---

list, Dict, NumPy array



Pd.DF



Data will be Stored in .csv

In Real-world

Pd.DF

infor  
& name

Server

