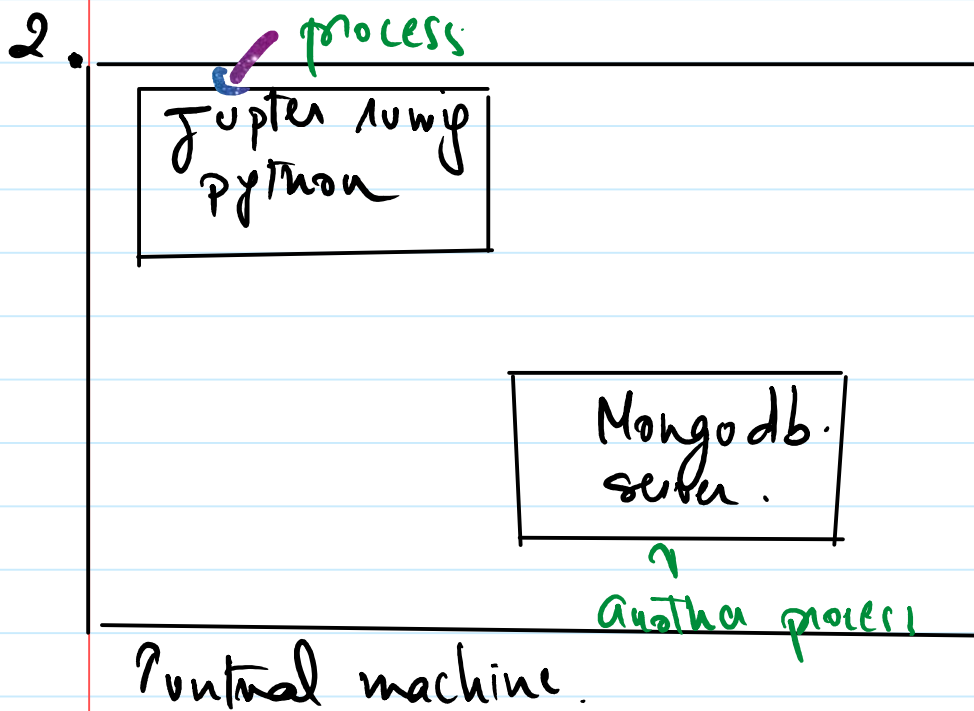


1. when we get result from mongodb database we get them in form of dict and pretty printer helps us to print them more organized.

`pp = PrettyPrinter(indent=2)`



- ① Jupyter python and mongodb server are independent process running on the same virtual machine.
- ② Jupyter python is a client to connect to mongodb server need:
  - 1. on which machine the mongodb is running: local cal host:

...ing. what are n...

2. which port the mongoDb is using: in our case (27017 default)

```
client = MongoClient(host="localhost", port=27017)
```

3) so the iterator in python is basically a container that hold object but iterate only one object at time for one time

=> how it works: basically it allocate a place in memory for one item then when you want the next item it will reallocate the same space for the next item.

=> we use it when we have a large amount of data instead of just load it in list **take much memore** we use iterator.

memory we use iterator.

Q: To go through iterator:

- we make it a list  $\Rightarrow$  `list(it)`.
- $\forall n \in \text{iter}$ :

④ to get all all the database:

`client.list_databases()`

↑  
iterator that we have to iterate through it

met1:

```
pp.pprint(list(client.list_databases()))
```

```
[{'empty': False, 'name': 'admin', 'sizeOnDisk': 40960},  
{'empty': False, 'name': 'air-quality', 'sizeOnDisk': 4190208},  
{'empty': False, 'name': 'config', 'sizeOnDisk': 12200},  
{'empty': False, 'name': 'local', 'sizeOnDisk': 73728},  
{'empty': False, 'name': 'wqu-abtest', 'sizeOnDisk': 585728}]
```

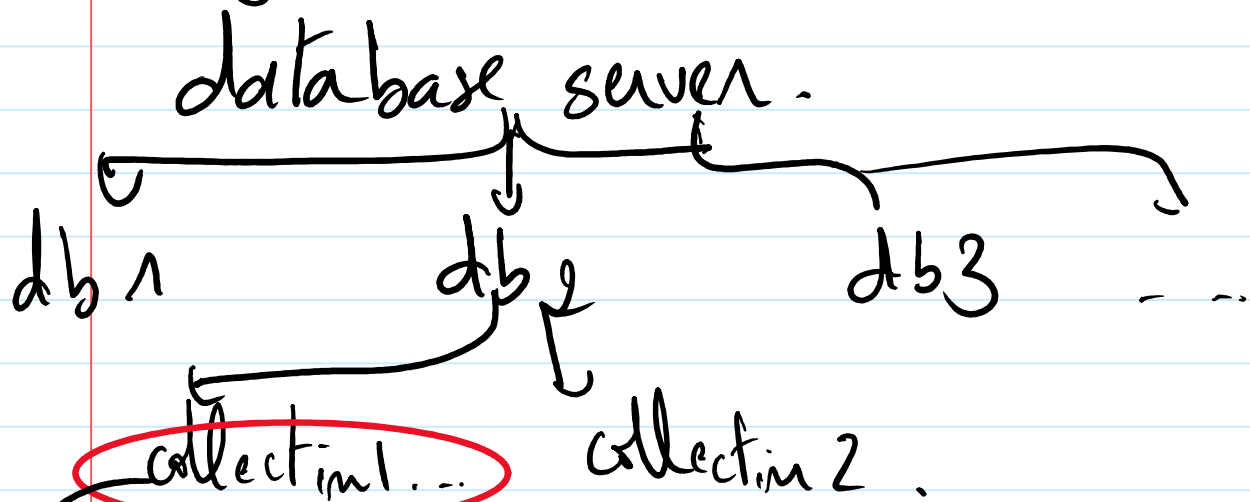
big  
size

the one  
that we need it

met2:

```
for d in client.list_databases():  
    pp.pprint(d)
```

⑤ The structure of database in MongoDB:



↳ we can think of collection as dataframe

↳ file that have many observation have common attribute

↳ documents ⇒ is a row in dataframe

⑥ To access to specific database in some case "nairobi-air"

=>

```
db = client["air-quality"]
```

⑥ we use  
`db.list_collections()`  
iterate so we have  
to iterate through it.

```
for c in db.list_collections():  
    pp.pprint(c["name"])
```

⑦ To retrieve a document from  
collection: `collection.find_one()`.  
the structure of document.

```
pp.pprint(result)
```

```
{ '_id': ObjectId('6525d772f44bfedd842a6fcc'),  
  'metadata': { 'lat': -1.3,  
                'lon': 36.785,  
                'measurement': 'temperature',  
                'sensor_id': 58,  
                'sensor_type': 'DHT22',  
                'site': 29},  
  'temperature': 16.5,  
  'timestamp': datetime.datetime(2018, 9, 1, 0, 0, 4, 301000)}
```

primary key (identifier)

```
'timestamp': datetime.datetime(2018, 9, 1, 0, 0, 4, 301000)}
```

data.

basically data about data.

analogy with emails:

the content of email: data

who send  
when was sent  
time was send } metadata