

# Curs 3: Pachetul Pandas, reprezentari grafice

## 3.1. Pandas

Deși NumPy are facilitati pentru incarcarea de date in format CSV, se prefera in practica utilizarea pachetului Pandas:

```
In [2]: import pandas as pd
        pd.__version__

import numpy as np
```

O comparatie a eficientei Pandas DataFrame si PostgreSQL este data [aici](https://blog.thedataincubator.com/2018/01/pandas-vs-postgresql/) (<https://blog.thedataincubator.com/2018/01/pandas-vs-postgresql/>). Concluzia articolului este ca la ora aceea, DataFrame era in multe incercari mai rapid decat PostgreSQL, ruland bine pe seturi mari de date.

## Pandas Series

O serie Pandas este un vector unidimensional de date indexate.

```
In [3]: data = pd.Series([0.25, 0.5, 0.75, 1.0])
        data

Out[3]: 0    0.25
        1    0.50
        2    0.75
        3    1.00
        dtype: float64
```

Valorile se obtin folosind atributul values, returnand un NumPy array:

```
In [4]: data.values

Out[4]: array([0.25, 0.5 , 0.75, 1.  ])
```

Indexul se obtine prin atributul index. In cadrul unui obiect Series sau al unui DataFrame este util pentru adresarea datelor.

```
In [5]: data.index

Out[5]: RangeIndex(start=0, stop=4, step=1)
```