



Nội dung

- ❑ Giới thiệu
- ❑ Series
- ❑ DataFrame

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
TRUNG TÂM TIN HỌC



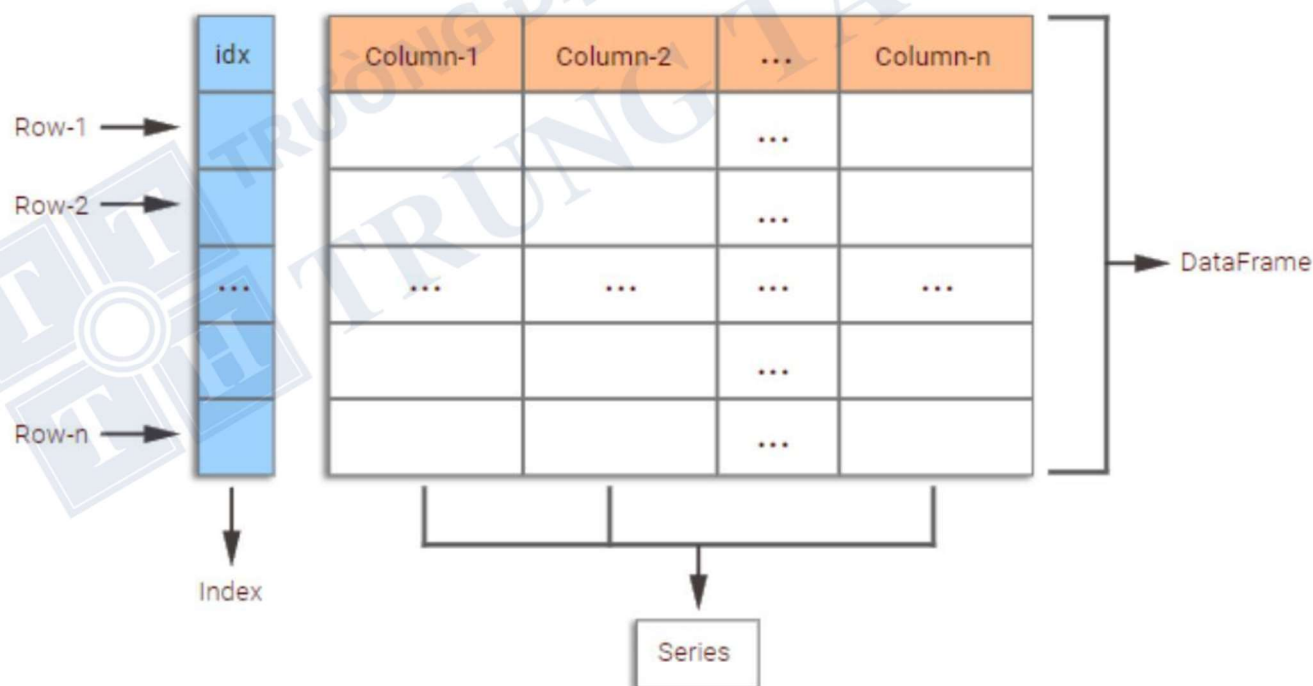


DataFrame

❑ **DataFrame** là một cấu trúc dữ liệu 2D, được tổ chức theo dòng và cột.

❑ **Đặc điểm**

- Các cột có nhiều kiểu dữ liệu khác nhau
- Trục được gán nhãn (dòng và cột)
- Có thể thực hiện các phép tính số học theo dòng và cột

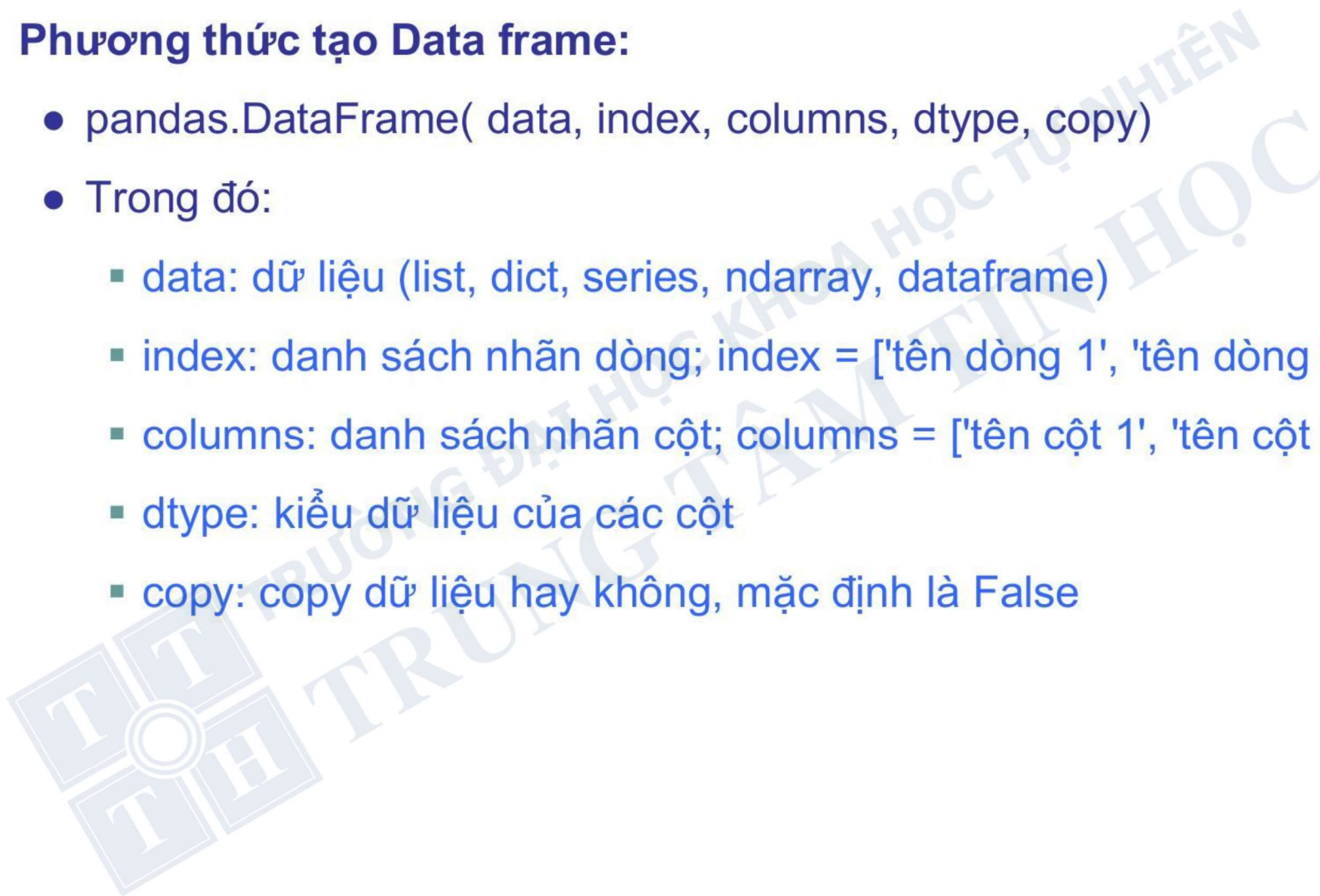




DataFrame

❑ Phương thức tạo Data frame:

- `pandas.DataFrame(data, index, columns, dtype, copy)`
- Trong đó:
 - `data`: dữ liệu (list, dict, series, ndarray, dataframe)
 - `index`: danh sách nhãn dòng; `index = ['tên dòng 1', 'tên dòng 2', ...]`
 - `columns`: danh sách nhãn cột; `columns = ['tên cột 1', 'tên cột 2', ...]`
 - `dtype`: kiểu dữ liệu của các cột
 - `copy`: copy dữ liệu hay không, mặc định là `False`





DataFrame

❑ Tạo DataFrame

- Tạo dataframe từ list

```
lst = [['Tom',25], ['Mike',27], ['Rose',32], ['Alice',26], ['Bill',25]]  
df = pd.DataFrame(lst, columns=['name','age'])  
df
```

	name	age
0	Tom	25
1	Mike	27
2	Rose	32
3	Alice	26
4	Bill	25



DataFrame

❑ Tạo DataFrame

- Tạo dataframe từ list

```
lst = [['Tom', 25, 1200], ['Mike', 27, 1500], ['Rose', 32, 3000], ['Alice', 26, 1200], ['Bill', 25, 1500]]  
  
df = pd.DataFrame(lst,  
                  columns=['name', 'age', 'salary'],  
                  index=['T01', 'M02', 'R03', 'A04', 'B05']) # cấp các index dòng  
df
```

	name	age	salary
T01	Tom	25	1200
M02	Mike	27	1500
R03	Rose	32	3000
A04	Alice	26	1200
B05	Bill	25	1500



DataFrame

❑ Tạo DataFrame

- Tạo dataframe từ dictionary

```
dict = {  
    'name': ['Thanh Thảo', 'Huyền Trân', 'Tuấn Tú', 'Phương Quang', 'Phương Nhi', 'Nhu Quỳnh'],  
    'age': [18, 20, 18, 21, 20, 19],  
    'mail': ['thao@gmail.com', 'htran@edu.vn', 'tu@yahoo.com', 'phuongquang@gmail.com', 'nhi@google.com', 'quynh@edu.vn']  
}
```

```
df = pd.DataFrame(dict)  
df
```

	name	age	mail
0	Thanh Thảo	18	thao@gmail.com
1	Huyền Trân	20	htran@edu.vn
2	Tuấn Tú	18	tu@yahoo.com
3	Phương Quang	21	phuongquang@gmail.com
4	Phương Nhi	20	nhi@google.com
5	Nhu Quỳnh	19	quynh@edu.vn



DataFrame

❑ Tạo DataFrame

- Tạo dataframe từ dictionary

```
dict = {  
    'name': ['Thanh Thảo', 'Huyền Trân', 'Tuấn Tú', 'Phương Quang', 'Phương Nhi', 'Nhu Quỳnh'],  
    'age': [18, 20, 18, 21, 20, 19],  
    'mail': ['thao@gmail.com', 'htran@edu.vn', 'tu@yahoo.com', 'phuongquang@gmail.com', 'nhi@google.com', 'quynh@edu.vn']  
}  
  
df = pd.DataFrame(dict, index=['0967123657', '0986111222', '0698453214', '1112223339', '9998887770', '0765341111'])  
df
```

	name	age	mail
0967123657	Thanh Thảo	18	thao@gmail.com
0986111222	Huyền Trân	20	htran@edu.vn
0698453214	Tuấn Tú	18	tu@yahoo.com
1112223339	Phương Quang	21	phuongquang@gmail.com
9998887770	Phương Nhi	20	nhi@google.com
0765341111	Nhu Quỳnh	19	quynh@edu.vn



DataFrame

❑ Tạo DataFrame

- Tạo dataframe từ series

```
name_ser = pd.Series(['Bảo Ngọc', 'Thanh Bình', 'Ngọc Thủy', 'Đức Duy', 'Toàn Trung'])
salary_ser = pd.Series([2000, 1200, 1500, 1700, 1500])
department_ser = pd.Series(['KT', 'KT', 'IT', 'IT', 'KT'])
```

```
employees_df = pd.DataFrame({'ten':name_ser, 'luong':salary_ser, 'phong':department_ser})
employees_df
```

	ten	luong	phong
0	Bảo Ngọc	2000	KT
1	Thanh Bình	1200	KT
2	Ngọc Thủy	1500	IT
3	Đức Duy	1700	IT
4	Toàn Trung	1500	KT