



Nội dung

- ☐ Giới thiệu
- ☐ Series
- ☐ DataFrame

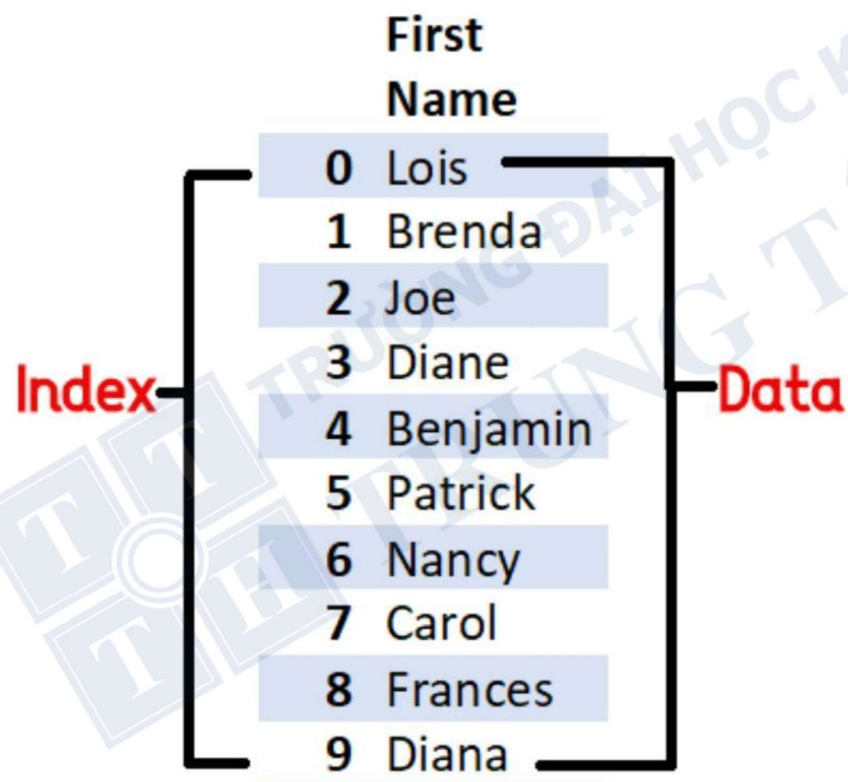
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
TRUNG TÂM TIN HỌC





Series

- ❑ **Series** là mảng một chiều có gán nhãn, có khả năng chứa các phần tử là một trong các kiểu dữ liệu object, int64, float64, bool, datetime64...
- ❑ Các axis label của Series được gọi là index.





Series

❑ Phương thức tạo series

- `pandas.Series(data, index, dtype, copy)`
- Trong đó:
 - `data`: dữ liệu từ nhiều dạng khác nhau như `ndarray`, `list`, `constants`
 - `index`:
 - Index tường minh: người dùng xác định giá trị index (label)
 - Index ngầm định: sẽ có giá trị trong `np.arange(<số phần tử>)`
 - `dtype`: kiểu dữ liệu (tùy chọn)
 - `copy`: tạo bản copy (tùy chọn)



Series

❑ Tạo series

Tạo series từ list

```
# tạo series từ list --> series
age_lst = [25, 27, 24, 28, 30]
age_ser = pd.Series(age_lst)
```

```
print(age_ser)
```

```
0    25
1    27
2    24
3    28
4    30
dtype: int64
```

Tạo series từ array

```
# tạo series từ array --> series
age_arr = np.array([25, 27, 24, 28, 30])
age_ser = pd.Series(age_arr)
```



Series

❑ Tạo series với index mặc định

```
# tạo series age_ser là tuổi của 5 nhân viên  
age_ser = pd.Series([25,27,24,28,30])
```

```
# in ra age_ser  
print(age_ser)
```

```
0    25  
1    27  
2    24  
3    28  
4    30  
dtype: int64
```

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
TRUNG TÂM TIN HỌC



Series

- ❑ Tạo series với index được gán nhãn (label index)

```
# tạo series age_ser là tuổi của 5 nhân viên  
age_ser = pd.Series([25,27,24,28,30], index=[101,102,103,104,105]) # gán nhãn cho index
```

```
print(age_ser)
```

```
101    25  
102    27  
103    24  
104    28  
105    30  
dtype: int64
```



Series

❑ Tạo series với index được gán nhãn

```
# tạo series score_ser là điểm của 5 sinh viên  
score_ser = pd.Series([7.5, 8.0, 6.0, 7.0, 3.5], index=['std1', 'std2', 'std3', 'std4', 'std5'])
```

```
print(score_ser)
```

```
std1    7.5  
std2    8.0  
std3    6.0  
std4    7.0  
std5    3.5  
dtype: float64
```

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
TRUNG TÂM TIN HỌC





Series

❑ Tạo series từ một giá trị

```
# tạo series từ một giá trị  
ser = pd.Series(10, index=[0,1,2,3,4])  
print(ser)
```

```
0    10  
1    10  
2    10  
3    10  
4    10  
dtype: int64
```

❑ Tạo series từ tập tin txt

```
# Tạo series từ tập tin txt là chiều cao các vận động viên bóng chày, tính theo inch  
height_arr = np.loadtxt('heights_1.txt', delimiter=',')
```

```
height_ser = pd.Series(height_arr)
```




Series

Thuộc tính	Ý nghĩa
dtype	Kiểu dữ liệu của series
size	Số phần tử của series
shape	Giá trị kiểu bộ (tuple) cho biết số chiều và số phần tử của series
index	Danh sách chỉ mục của series
values	Mảng giá trị của series, kiểu ndarray
loc, iloc	Truy xuất phần tử trong series theo nhãn index (loc) hay vị trí (iloc)
...	



Series

❑ Xem thông tin series

series age_ser

101	25
102	27
103	24
104	28
105	30
dtype: int64	

```
print(age_ser.size) # in ra số phần tử
```

5

```
print(age_ser.shape) # in ra số chiều và số phần tử
```

(5,)

```
print(age_ser.dtype) # in ra kiểu dữ liệu
```

int64

```
print(age_ser.ndim) # in ra số chiều
```

1

```
# in ra các index  
print(age_ser.index)
```

Int64Index([101, 102, 103, 104, 105], dtype='int64')