

# Introduction to Data Science

Bộ môn Khoa học dữ liệu  
Khoa Công nghệ thông tin  
Trường Đại học Công nghiệp thành phố Hồ Chí Minh-IUH

## Bài 1:

Với mỗi loại file sau hãy thực hiện các yêu cầu:

- Tạo một file theo đúng định dạng
- Đọc file và hiển thị một phần nội dung của file
- Thêm dữ liệu và ghi vào file

Các loại file:

- 1.1 txt
- 1.2 log
- 1.3 config
- 1.4 ini
- 1.5 JSON
- 1.6 YAML
- 1.7 XML
- 1.8 CSV

## Bài 2:

Viết chương trình sử dụng cấu trúc dữ liệu Danh sách liên kết đơn – SLL quản lý Tàu (Trainees) theo thông tin như

**Train:**

1. tcode (string): mã của tàu (mỗi tàu có **một mã duy nhất**).
2. train\_name (string): tên của tàu.
3. seat (integer): số ghế trên tàu (**seat > 0**).
4. booked (integer): số ghế đã đặt (**booked >= 0 and booked ≤ seat**).
5. depart\_time (double): Thời gian khởi hành (**depart\_time >= 0**).
6. depart\_place (string): Nơi khởi hành.

*Thực hiện các yêu cầu như dưới đây, chú ý phải xử lý Exception cho toàn chương trình:*

2.1. Đọc dữ liệu từ file input.txt có nội dung như sau:

```
B03 | Sug | 12 | 3 | 11 | PA
B01 | Mil | 10 | 5 | 5.7 | PC
B02 | App | 5 | 2 | 4 | PB
B05 | Roo | 7 | 6 | 15 | PE
B07 | Bee | 11 | 3 | 12 | PF
```

Dòng đầu tiên có ý nghĩa: tcode = B03, train\_name = Sug, seat = 12, booked = 3, depart\_time = 11 và depart\_place = PA

- 2.2. Nhập và thêm dữ liệu vào cuối danh sách
- 2.3. Hiển thị dữ liệu
- 2.4. Lưu dữ liệu vào 1 file
- 2.5. Tìm kiếm dữ liệu theo tcode
- 2.6. Xóa dữ liệu theo tcode
- 2.7. Sắp xếp dữ liệu theo tcode
- 2.8. Thêm 1 dữ liệu sau vị trí k
- 2.9. Sửa dữ liệu theo tcode
- 2.10. Xóa 1 node trước node có tcode = xCode

### Bài 3:

Dữ liệu được cung cấp trong file *car.csv* được thu thập từ năm 1990 đến năm 2020 tại Việt Nam.

- 3.1 Đọc dữ liệu và hiển thị 20 dòng đầu tiên của dữ liệu, cho biết kiểu dữ liệu của mỗi cột
- 3.2 Cho biết các dữ liệu còn trống (null) trong mỗi cột
- 3.3 Tìm hiểu thông tin về dữ liệu Số lượng xe của mỗi loại

| Loại xe               | Số lượng |
|-----------------------|----------|
| Sedan                 |          |
| SUV                   |          |
| Hatchback             |          |
| Crossover             |          |
| Van/Minivan           |          |
| Bán tải / Pickup      |          |
| Truck                 |          |
| Coupe                 |          |
| Convertible/Cabriolet |          |
| Wagon                 |          |

- 3.4 Tìm hiểu số lượng xe của mỗi loại xe mới/cũ theo từng loại xe (“new\_old”)

| <b>Loại xe</b>        | <b>Xe cũ</b> | <b>Xe mới</b> |
|-----------------------|--------------|---------------|
| Sedan                 |              |               |
| SUV                   |              |               |
| Hatchback             |              |               |
| Crossover             |              |               |
| Van/Minivan           |              |               |
| Bán tải / Pickup      |              |               |
| Truck                 |              |               |
| Coupe                 |              |               |
| Convertible/Cabriolet |              |               |
| Wagon                 |              |               |

- 3.5 Vẽ biểu đồ thể hiện số lượng xe của mỗi loại xe mới/cũ qua từng năm và đưa ra nhận xét.
- 3.6 Tạo bảng dữ liệu mới chỉ có 3 vùng (TP HCM, Hà Nội và Đà Nẵng) và lưu kết quả vào 1 file mới
- 3.7 Từ bảng dữ liệu đã thu được ở câu 3.6, hãy so sánh các đại lượng thống kê giá trị trung bình, trung vị, độ lệch chuẩn của giá xe, vẽ biểu đồ cột và đưa ra nhận xét
- 3.8 Vẽ đồ thị phù hợp thể hiện phân phối của các loại màu sắc bên trong (in\_color), qua đó bạn hãy đưa ra bình luận của mình.
- 3.9 Vẽ mối tương quan giữa giá xe và số km đã đi đối với xe cũ. Theo bạn hai đại lượng này có quan hệ với nhau hay không? Giải thích.
- 3.10 Chọn loại đồ thị phù hợp để khảo sát mối quan hệ giữa giá của xe SUV và Sedan theo từng năm, đưa ra nhận xét của bạn.