

# MSc Data Analytics

## Database and Analytics Programming (H9DAP)

### Continuous Assessment 1: In-class Test

This test is worth 30% of the module.

#### Instructions:

- You may consult any of the notes and code from the classes or the labs.
- All code must be fully commented
- Upload your code to Moodle
- Any cheating will be reported to the school ethics committee for disciplinary action

#### Question 1:

Theophylline is a drug used for the treatment of diseases such as chronic obstructive pulmonary disease and asthma. This medicine was administered to 20 subjects over a number of days. Each time the drug was administered to the subjects, the following measurements were recorded.

id	a number (1, ..., 20) identifying the subject on whom the observation was made
weight	weight of the subject (kg)
dose	dose of theophylline administered orally to the subject (mg/kg)
time	time since drug administration when the sample was drawn (hr)
conc	theophylline concentration in the sample (mg/L)

There are 1,000 records in total. You have been provided with an XML file containing this data.

- Create a function to import this XML file. Your function should include exception handling clauses. **[15 marks]**
- Use the print function to display the 'id', 'weight' and 'conc' of the first, third, fifth, seventh, ninth and eleventh records in the XML dataset. (Hint: you may use the `range()` function). **[10 marks]**
- Extract the XML data and write it to a CSV file. Your file should also contain the column names. **[30 marks]**

## Question 2:

- Create a NumPy array filled with 800 numbers. Ensure that your array has 200 rows and 4 columns. **[5 marks]**
- Using slicing, split this array into 5 separate arrays. The number of rows in each array should be equal, and there should still be 4 columns. **[10 marks]**
- Reshape these 3 of these arrays into any dimensions of your choice. They should all have different dimensions. **[5 marks]**
- Split 2 of these reshaped arrays horizontally **[5 marks]**

## Question 3:

Given the following string:

Betty Botter bought some butter

But she said the butter's bitter

If I put it in my batter, it will make my batter

bitter. But a bit of better butter will make my batter better

So 'twas better Betty Botter bought a bit of better butter

- Write a single function to highlight the words *"butter"*, *"bitter"*, *"batter"* or *"better"* if they appear at the end of a line. **[15 marks]**  
Your answer should look like this:

```
Betty Botter bought some {butter}
But she said the butter's {bitter}
If I put it in my batter, it will make my {batter}
bitter. But a bit of better butter will make my batter {better}
So 'twas better Betty Botter bought a bit of better {butter}
```

- How would you highlight the word *"Betty"* regardless of wherever it is found in passage? The words in (a) above should also be highlighted. **[5 marks]**  
Your answer should look like this:

```
{Betty} Botter bought some {butter}
But she said the butter's {bitter}
If I put it in my batter, it will make my {batter}
bitter. But a bit of better butter will make my batter {better}
So 'twas better {Betty} Botter bought a bit of better {butter}
```