Welcome to our technical assignment. This assignment has been designed to gauge your programming skills, problem-solving abilities, and thought process. We encourage you to take on these challenges in a manner that truly reflects your expertise and ingenuity.

Upon completion, please submit your code in any format you are comfortable with - this could be a Jupyter notebook, a plain text file, etc. There's no need to rush; focus on demonstrating your skills and abilities. See how far you can go and good luck!

Section 1:

In a language of your choice, implement a simple web crawler that gets a news website as input (e.g. https://www.news24.com/) and crawls the HTML content of up to 100 pages of that site with a breadth-first approach. The downloaded pages should be stored as HTML in a folder in the file system.

Bonus question: The crawler needs to be able to work with up to 50 parallel processes. The number of processes can be passed as a parameter. If no input is given the default value shall be 5 processes.

Section 2:

Establish a local Postgres database and perform the following tasks:

2.1 Identify components from the news website that can be parsed into tabular data (e.g. headings, authors, etc.) and write this data into a maximum of 3 tables of your choosing. This can be incorporated into the codebase from section 1.

2.2 Set up a schema in your database called 'data_science' and restore the provided dump file. Post-restoration, you will find three tables in the 'data_science' schema:

vessel_info

vessel_ports

Vessel_tracking_events

Please ensure to first create the schema before attempting the restoration.

Answer the following questions using PostgreSQL

2.2.1 How many unique ships do we have data on?

2.2.2 At present, we have data on numerous container ships. How many of these ships would be classified as ULCV?

2.2.3 After analyzing the 'ports' table, you noticed that erroneous information has been entered into the 'code' column. Utilizing an alternative column in the table, retrieve a list of rows where most 'code' values are incorrect.

Submit your PostgreSQL queries alongside your code from section 1 and 2.1.