DATA8005 – Assignment 1 Part 2 – Neo4j

**Due:** End of week 7 (November 4th)

**Worth:** 20%

# Overview

This is the 2nd and final part of your first assignment, to be handed up with part 1.

Before attempting this assignment, you should have completed the tutorials in the Neo4j web console, which features a variety of node and edge creation and query commands.

You have been given the Tate dataset containing thousands of artists and artworks. You must create a graph containing artists and artworks and other potentially useful nodes.

The movie database example created a graph database containing movies and people with the people having various roles, e.g. actor or director. Now you must create a tate database. You can reference the file movie.cypher and how it creates nodes and edges (relationships).

Remember, Neo4j is a *property* graph database and so relationships can hold data too.

Choose 5 artists and about 15 to 20 artworks. Try to get some artists with a range of artworks created from 1 to 5. Create the nodes the “CREATED\_BY” relationship between them.

Now you need to examine the following typical questions often asked of the Tate data:

1. Which artworks were painted by Edinburgh-born artists?
2. Which artworks feature royalty, or subjects with their arms raised, or sea monsters?
3. I like a particular “Watercolour on paper” artwork featuring a boat. Can you recommend other watercolours featuring boats?

For this assignment, you need to submit the following:

* A file called tate.cypher containing the commands to create the artist, artwork and other node types.
* A document with:
  + a sketch of some of the nodes (you don’t need to sketch all concrete nodes, just a selection, e.g. 2 or 3 artists, 5 or 6 artworks and the other nodes you chose to add to help answer questions 1 to 3. Include relationship names and any properties of interest on nodes and edges (you don’t need to include every single property, just the ones you might use to answer questions 1 to 3;
  + A brief discussion of why you added the extra node types you did;
  + 3 sample queries to answer questions 1 to 3, which can be run on your created graph database.

# Marking scheme

* The Cypher commands to create the database – 8 marks
* The sketch of the graph database – 3 marks
* Discussion about reasons for extra nodes – 3 marks
* The 3 Cypher queries – 6 marks