3 Simple Hadoop Graph Processing

In this section you will run Hadoop MapReduce jobs on a dataset of harbours and routes. You may use the Hadoop environment from previous projects/labs. The data is in the following format: [Harbour, Harbour Number, Route, Route Number]. Harbour names are separated by a "-", route names by a "_".

```
wget --no-check-certificate
↪ 'https://docs.google.com/uc?export=download&id=1SrurEuPrw04S6afIPXl1WKuL73JRuULT'
↪ -O hadoop.csv
```

Using a MapReduce script, complete the following graph exploration tasks. You may "merge" the outputs of multiple MapReduce tasks with further MapReduce tasks. You may also create transformed outputs using MapReduce that generate your final answers (e.g. your final output does not need to be generated from the original bacon.csv file). Remember - your MapReduce tasks can take multiple inputs! You may present multiple mapreduce outputs as your answer, if the answer is split across them (e.g. for Q4 below - you may find A-B and B-C separately, although a better answer might have them in a single output).

1. Create a list of the number of routes that connect to each harbour
2. Route "Wolfsbane_Nine" is associated with only one harbour - what is it?
3. What harbours are connected by route Carnation_Sixty-seven(No.1223).
4. Which harbours fielded emergency routes - these are routes whose route number begins with "911"
5. "Midnightblue-Epsilon" is connected to two other harbours by a route- what are they? As a hint, consider how you could compare multiple MapReduce outputs with different Map conditions to find the link. For this question you do not have to combine your separate operations into one script.