# Hadoop Lab report

You will first need to run through your personalised lab script – this is very much a cut/paste exercise so you can see how things look on a ‘real’ system, it should take no more than 30 minutes (depending on how much you review the commands & output)

For the report you will need to take a screen grab of the final part (see the Spark section).

**Note:** Connecting to the server is probably easiest using Putty (this is how I tested the system).

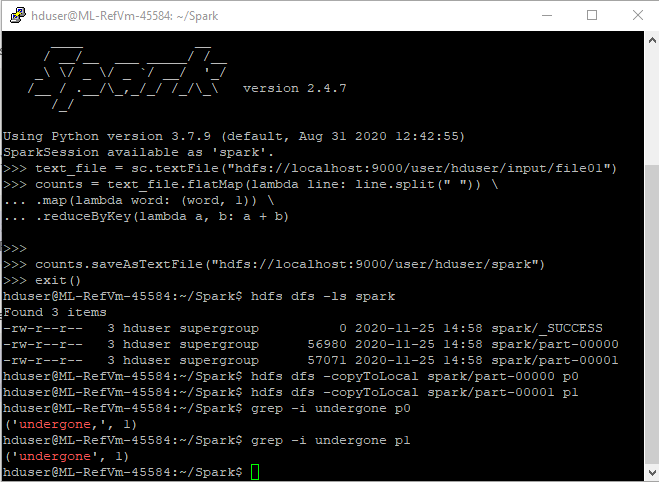
# What is Map Reduce: **Two** sides maximum

Please wite a **MAXIMUM of two sides** outlining what MapReduce is and how it works. You may use up to two figures to support this explanation

# Spark: **One** page maximum

This should consist of two parts

2.1 A screen shot of the output from the last exercise (as per the sample below) – this should take up no more than half of the page



2.2 An explanation of the code as used.

*counts = text\_file.flatMap(lambda line: line.split(" ")) \*

*.map(lambda word: (word, 1)) \*

*.reduceByKey(lambda a, b: a + b)*

When discussing this, you may find it easier to break it into the three steps (line.split, map & reduceByKey)

References: Please provide these on an additional page - they do NOT count towards the three pages