**Cloud Computing**

**Lab 2**

1. **Overview:**

Big data is the processing of large amounts of data. Because computers are becoming more advanced and are processing larger and larger projects or bigger networks, it makes data storage and retrieval a problem.

The reason why there is always such a demand for more data is because there is constantly more data being generated by taking lots of measurements and recording of events.

In more recent years Big Data has become more of an issue, causing huge tsunami of data. I believe this is due to a number of reasons which include the increase of use and recent innovations in Social Media, Search Engine Technology, Cloud Computing Services, and Mobile Smart Phones etc. most of which are now processing billions of photos and images each day, including HD videos which use far more data than conventional digital videos which were more used only a few years ago, all of which have caused a massive influx of new users and people to consume data at an increased rate.

Looking into big data is isn’t just huge social media networks and cloud computing services that are using vast amounts of data but rather scientific technology such as observatories, CERN’s LHC and Climate Computing, as vast amounts of data is gathered to make these calculations on weather, space and atomic observations.

To get a scope of how incredibly huge the amount of data being processed every day we looked at the amount of data Google, Netflix, Facebook, eBay, CERN’s LHC and German Climate Computing Stores process every day and then taking these huge figures and compare them against figures we’re more familiar with e.g. Megabytes, Gigabytes, Terabytes etc.

Exabyte is seen as the new challenge to Computer Science. The big problem that has arisen in the realisation of the necessity to start processing data at a Exabyte level Is the capture versus analysis rate of the data, which involves the ability to store the data, search through the data, have an elastic capacity and keeping the data secure. Keeping this in mind the big data is often broken down into organised groups or network. When it comes to actually doing some science on the data all the sources must be brought together or farmed. Which more problems can arise from even trying to farm massive amounts of data due to big data being so large now it’s not possible to store it on a single device even though 8TB drives exist.

To allow all this transferring of data better computer science methods must be created to ensure the data will all arrive in one piece on a secure line as fast as possible which possess a big problem, which must also be processed for the data to become useful.

1. **Areas of Difficulty:**

Was not able to set up a Okeanos server. This was due to a problem with distributing further IPs. To compromise I had to create a clone of my Ubuntu Virtual Box. This provided a similar situation that the server would have been able to provide and I was able to work on Github on both Virtual Boxes.

Had problems pushing and pulling with git on the virtual machines, this was due to a syntax error when specifying the remote origin master, this took me a while to solve as the error that was provided said that my access was denied which led me to believe it was successfully connecting to my repo but something was preventing the ability to allow me to download or upload to it.

1. **What went well:**

I was very happy with how the Project Euler tasks went as it was my first time ever using Python, managing to pick up on it much quicker than I had anticipated. I first solved the problem in Java using Eclipse and then converted the Java code into Python using Nitrous as my IDE.

Using Nitrous.io was straightforward for me as I’ve used it in the past for managing work through sharing of a workspace online.

Learning about Big Data and how it worked was fairly straight forward to get my head around, with use of the lectures notes and the videos.

1. **My Github Account:** https://github.com/chdonncha