**1. After watching online lectures 1-3, discuss in approximately 500 words the following question.**

***“In the context of computer science, why is BIG DATA an issue now, and what factors in your opinion have lead to this so-call Tsunami of data”* (70%)**

"Big Data" when it comes to computer science means that there is a lot of data which is being recorded using many types of devices and utilities.

The factors that have led to this are the devices that are recording data, are increasing. The people that have these devices are increasing and therefore the data is increasing. Data comes in many forms particularly images, videos, pictures etc.

All this data is captured and stored. But we need to be able to write it faster than capture it. After this we need to be able to process it, if not, it sits waiting to do work.

We then analyse that data and we need to move that data so usually wherever we record it, we need to move that with the data.

The challenge is to store this ever growing data store.

Databases wouldn't initially be set up to store that kind of data in an indexed way and hence we need new technology to address this.

This ever growing humungous amounts of data also cannot be processed all the time since we cannot reserve enough computing power for this and we also can't have data staying idle due to the costs associated with it.

There is also part of keeping data secure but it should be accessible to restricted number of people.

There are people that farm this data and bring it to one location and then decide what can be done with this tsunami of data.

Most of this data must be in exabytes. Another challenge that exists in the computer science community is that exabytes is a territory which hasn't been addressed yet and the highest data that a single device can hold would be 8 terabytes at this moment.

We need a new kind of computer science that can process more data that can reside in a single device. This should lead to more data that can reside in a grouped set of service. Since the data needs to be processed to be useful, we apply some kind of processing to transform the data where the search for key marks within that which would be the apps we write residing in a cloud.

The fact that this tsunami of data needs to be capable of being:

1. Moved

2. Processed

3. Stored

4. Secured

5. Searchable

requires a tremendous amount of effort, power and cost. Since in order to generate any sort of theory, we need data, and data is currently going through many factors and challenges which would most likely delay a lot of time and research associated with this.

Data also needs to be sorted and placed in such a way that it should be easily found. Many people have different requirements on how data should be sorted and time and money is put into storing data in a certain way. Due to this tsunami of data, this would take longer or even become quite complex.

A lot of challenges and factors arise from this tsunami of data and the real solution would be to first find the means of actually being able to transfer and store exabytes.

**2. Explain any areas of difficulty in this week’s lab (100 words or less) (10%)**

The virtual machine image from last week had disappeared from my U drive and my supervisor told me that some changes had taken place so I had to re-download everything all over and install the packages which took time.

There were problems with OKEANOS virtual machines since they couldn't detect IP addresses and hence I could not use them after having them built and running.

Instead of cloning my repo using the OKEANOS virtual machines, I decided to use the one from last week instead.

There wasn't enough time to complete the Euler python problems even though I created a python file to start my work.

**3. Explain what went well in this week’s lab (100 words or less) (10%)**

I already have a Github account and I made the repo and shared with Paul.

Instead of using OKEANOS virtual machines, I used my Ubuntu from last week and was successful in cloning and making changes to my files. I also made a change to a file on my github repo and then made a change on my Ubuntu using nano. This created a conflict with the files giving me error in regards to my "head" being ahead of my other commit etc. I resolved it by merging the files. I was also informed that I could have resolved this conflict by deleting some edit too.

I also created a Nitrous account very easily and was capable of cloning my git repo into the Python IDE. I made changes inside my repo from Nitrous using git. I then created an 'Euler.py' file to start my Python problem and pushed the commit into git.

**4. Provide a link to your GITHUB Repo (10%)**

<https://github.com/sjbarlas/CloudComputingDT228-3>

**5. Any other comments?**

The lab was confusing and hectic because my Ubuntu virtual image disappeared and then OKEANOS wasn't working. This took most of my time when it came to downloading the Ubuntu image and downloading the packages. Hence it did not let me complete my Euler problems.

I also thought that looking at the videos also took most of my time and hence the lab wasn't completed. Instructions involving steps on a pdf would have been better but since the video showed each process in real time, it would have been effective if there was more time.

Overall I enjoyed the lab since I learned something completely new and different!!!