1st June

Today is the first day of working on my dissertation project. I’ve been looking forward to this for a while, largely because I’ve been wanting to do something more practical than the endless reading I’ve been doing while revising for my exams.

My job is to build an emulator that emulates the properties of BASE transactions over an ACID database. Just before the end of last term I submitted an annotated bibliography on the project. My first plan of action for at least today and tomorrow is to begin reading through the resources that I’ve gathered, reminding myself of the different approaches that I considered, and choosing where to go next. While working on the bibliography I decided that the first BASE property that I should try to emulate would be that of eventual consistency. I still plan to stay with this original objective. I also picked out two potential real cloud databases on which to base this model, and I now need to look at the eventual consistency algorithms for both of these and decide which one I will actually try to implement. My supervisor suggested the possibility of maybe using ideas from both of the approaches, and I would certainly like to do this if it were possible, because then it might make the second half of the project, where I have to compare my solution against real cloud systems, more interesting.

I’d also like to take the opportunity to point out how relevant this project is to the rest of the modules I studied as part of my degree. The language I plan to use for my implementation is the Scala programming language, which is based upon Java and runs on the Java virtual machine, and so I will be putting the skills acquired in CO871 to good use. Because BASE transactions are heavily used in cloud-based databases, my project has obvious links to CO846. I also plan to make use of a large number of the tools introduced in CO894, as I believe that this will ensure a product of better quality.

Scala is a functional programming language, so I will be able to put the recursion skills learnt in CO884 to good use. This project lends itself well to concurrency, and therefore my skills learnt in CO890 will also be highly valuable.

2nd June

I spent most of yesterday going through the two main approaches that I found during my initial project research, which were the two eventual consistency models supported by CouchDB and DynamoDB. I have set myself a small target that by the end of today, I would like to have written up both a small description of the project and a technology review for the eventual consistency section of the project.

So far, both approaches seem to be very interesting and I can see how each one of them is effective in what it is doing. After reading through the two approaches more carefully, however, I think that both of them have various tradeoffs and I’d need to consider which one I’d want to implement very carefully.

With the approach offered by DynamoDB, all pieces of data are stored under version control. In implementation terms, this would be relatively straight forward to implement, as every time a piece of data is changed, one can just record the change and the time it happened in the database. While this would be straightforward to implement algorithmically, it wouldn’t scale well. This is because if we had to track every time something was changed in the database, this would take a lot of space relatively quickly.

The other approach is that of using vector clocks, which would be more complex to implement algorithmically, as it involves comparing these lists to determine which changes come before another. On the other hand though, it would be easy to store a vector clock in the database for each object, as this is just a list of timestamps and other metadata, so maybe I can compress that into a more serialisable form. This would, however, make more of a challenge in implementing the actual algorithms.

4th June

Yesterday I had a meeting with my supervisor, We talked through the feedback of my annotated bibliography and I was able to get a clearer idea of what needed to do and what then might be needed to get there.

I am now working on my project plan, I have decided that I will use an agile approach to my software development. This is partly because this kind of project is unfamiliar to me, and so by making it incremental, I will be able t assess my progress based upon my targets each week and check to see whether I am on target or not. I am currently in the middle of writing some broad goals, Once this has been done, I will write more specific user stories and try to establish a plan from week to week. This will also be accompanied by the creation of UML use case Diagrams to illustrate the requirements in a more visual form.