Winston Shine

ILC Database Theory/Security

Week 6 Progress Report

2/19/2023

My goal is to study Database Theories and Security through the book Translucent Databases. Some additional smaller goals I am trying to accomplish are doing my work this quarter with docker containers and If time allows study a bit about document-oriented databases.

This week I continued studying quantization. I didn’t make as much progress as I hoped to this week so I will be doing more work on understanding quantization next week, specifically how to use it in databases since so far, I have just been learning about how quantization works and not how to apply it.

One of the topics I learned about this week was adaptive quantization. This is a method commonly seen in compressions algorithms the example used the most in the book was the compression of digitized images. The idea of adaptive quantization is to continuously separate the data in smaller and smaller groups and then find one value to represent each group, and replace the original data with the value that represents the group it’s a part of. The example the book gave that helped me understand this is the gif file format. The size of the file can be reduced by reducing the total number of colors one pixel can represent, in the case of gif each pixel is 8 bits and can be one of 256 colors. To achieve this reduction in size, all the original colors in the image would be continuously separated into smaller groups of similar colors until there were 256 groups. Then one color would be chosen to represent all the datapoints in that group. Two methods that could be used to determine what color represents the group would be to find the datapoints closest to the ‘center’ of the group (easier to understand when thinking of it as a graph), or to take the average of all the members of the group and use that instead. Reading about this reminded me a lot of talking about classification in Machine Learning. I think I have a solid understanding of this idea, but I need to study a bit more about the math behind it next week.

The other topic I read about was Projection. Projection is taking data and representing it in a smaller dimension. The algorithm used to accomplish this is singular value decomposition, an algorithm used in matrix algebra. The largest singular values are the used for the best way to view the data, so the algorithm finds the largest singular values (number of singular values corresponds to number of dimensions in the result) and discards the rest. I do not quite understand what is meant by singular value. Next week I’ll have to read more about this since I did not follow the math very well.

M = UDVt

M is a matrix representing the data

D is a matrix with the largest singular values set along its diagonal and every other value being 0

The goal of the singular value decomposition would be to find U and V.

Next week I will continue learning about quantization with a focus on how it applied in databases, and then I will move on to read chapter 7 of Translucent databases titled ‘Coordinating Users’.