Reflective Piece: Deciphering Big Data

Over the past three months, I have participated in the "Deciphering Big Data" course, which has been an enlightening journey into the intricate techniques of data extraction, exploration, and cleaning. I have also gained many skills from doing the group project with my class mates this semester. This reflection essay will delve into my experiences, contributions, and the impact this course has had on both my professional and my personal development.

The course covered a wide array of topics, starting from the basics of data extraction such as using API for parsing data between two systems , Webscaraping and Data sources i.e. Application data stores, such as relational databases, Static files produced by applications, such as web server log files, Real-time data sources, such as IoT devices(Huxley et al , 2020).The module on data cleaning was particularly enlightening as it emphasized the importance of data quality in the analytics process.

The topics I covered in this module helped me greatly during the development team project .Our project was titled: PHI Inventory Management Database and as the name suggests, it was a logical design of an aircraft parts inventory management database. Our task was to act as consultants and design an Inventory Database Management System for the organization. The main correspondence for the group was done via WhatsApp; however we had several meetings over zoom to discuss the truly important aspect of our project. For example in our initial Zoom meeting we discussed and divided the task to be carried out by each member in order to put together a comprehensive report on the proposed database report. According to Knight (2016), making a list of all the work that needs to get done and then assigning tasks according to each worker's specific function, position, and strengths is key to the success of a project.

Courtney Somerville had the task of covering the first section of the report which was about introducing the organization, its activities/operations, the needs of the stakeholders, and how the database would meet them. My task was to do the conceptual database design and explain the type of data to be captured, our data source, our choice of DBMS, etc. Aneil Maharaj was to provide the public datasets from his workplace since he works for PHI Americas, and to also do the User Interface Design and discuss the application programs that use and process the database. However I took on the role of team leader in terms of making suggestions about

the schema of the database since I have experience with Databases form my certification as an IBM Data Analyst.

As part of the development team, I appreciated the opportunity to collaborate with peers from diverse backgrounds and skill sets. Our team dynamic was characterized by open communication, and a shared commitment to producing our deliverable. By leveraging each other's strengths and collective expertise, we were able to tackle the project and deliver our report 5 hours prior to the deadline.

Emotionally, working on the project was both exhilarating and challenging. There were moments of frustration and uncertainty, particularly when trying to come up with a comprehensive schema for the database and deciding which fields to keep and which ones to discard. We finally used the principles of Normalising such that we had related and normalised tables which suited every member for our database.

However, these challenges served as valuable learning opportunities, pushing me out of my comfort zone and teaching me how to work in harmony as a team by considering everyone’s suggestions. We ended up using Google docs as a platform for writing the report so that every member was able to make updates to the document. A good example is when Aneil suggested that we used MySQL Workbench for creating the database and populating the table. Initially I was against the idea, I suggested opening a connection to the database using Jupyter notebook .But when I was implementing the database I saw that Workbench has quite a number of features and tools for database creation, data importing etc .That is why I ended up using the software as my main tool for the database build. MySQL Workbench simplifies database design and maintenance, automates time-consuming and error-prone tasks, and improves communication among DBA and developer teams (MySQL, 2024).

I realized the importance of embracing a growth mindset and adopting a problem-solving approach. I actively sought feedback from my peers when I was finally doing my individual project of designing and building the logical database, refining my work based on their constructive criticism. One of the most important learning outcomes of this project was the development of practical skills and knowledge that are directly applicable in my professional life. I honed my skills in the use of APIs, Normalisation and data cleaning, and working with MySQL.

In conclusion, the "Deciphering Big Data" course has been an interesting journey that has significantly enhanced my skills, knowledge, and mindset in the field of big data analytics. From mastering data extraction and cleaning techniques to collaborating effectively in a team setting, the course has equipped me with the tools and insights needed to tackle the various tasks in the data-driven world. Moving forward, I am excited to apply these concepts and tools in my professional career in the field of big data analytics.

REFERENCES:

- Huxley et al. (2020) Big Data Architectures: https://learn.microsoft.com/en-us/azure/architecture/databases/guide/big-data-architectures [Accessed 30 th November 2023].
- Knight, R. (2016) Delegating: https://hbr.org/2016/11/make-sure-your-teams-workload-is-divided-fairly [Accessed 12 January 2024].
- MySQL. (2024) MySQL Workbench: Visual Database Design: https://www.mysql.com/products/workbench/design/#:~:text=MySQL%20Workbench%20simplifies%20database%20design,among%20DBA%20and%20developer%20teams. [Accessed 3 February 2024].