**Reflective Piece**

Understanding the fundamentals of big data and its potential impact on an organization is crucial for any business looking to stay competitive in today’s data-driven economy. However, data needs to be processed appropriately to gain beneficial insights from it, and the heterogeneous nature of this data makes this increasingly more difficult and time-consuming. The exponential growth of data generated is far more than human processing capabilities and thus computing methods need to be automated to scale effectively. (Ugur et al, 2020).

This module covers several essential and fundamental facets of data management and manipulation, each of which is crucial to data science, analytics, and a wide range of other disciplines. It revolved around data storage, web scraping, data cleaning, transformation, automation, and utilizing APIs for data parsing, focusing primarily on constructing normalized tables and building databases. Data storage is essential for efficient data work, as it allows for quicker retrieval and processing, saving time and resources. Web scraping is crucial for gathering non-necessarily available data, while data cleaning and transformation are essential for ensuring accuracy and usability. Automation can speed up repetitive tasks and reduce human error in data manipulation. APIs are used for data parsing, allowing for data collection from various sources. Database normalization is crucial for organizing tables and columns, reducing redundancy, and improving data integrity, especially when working with relational databases.

The module was intriguing and rigorous in terms of the complexity of the concepts and the level of critical thinking required; however, it was very challenging and provided many opportunities for learning and growth. It pushed the boundaries of my prior knowledge and experience, requiring me to step out of my comfort zone. There were instances when the tasks seemed daunting, and the learning curve was steep. Yet, these challenges were also what made the module so rewarding. This module provided me with a transformative experience that has enriched my knowledge, honed my skills, and broadened my perspectives in the field.

A standout aspect of this module for me was the creation of a normalized database. This component gave me a deep understanding of the complexities involved in data storage. It brought to light the paramount importance of meticulous planning and careful implementation in averting data redundancy and preserving data integrity (Coronel et al., 2023). Through the construction of normalized tables, my comprehension of data relationships was enhanced, and I realized the critical need for optimising data storage to ensure efficient data retrieval. The process offered valuable insights into the minutiae of database design and underscored its role in effective data management.

One particularly challenging yet enriching segment of this module was the Development Team Project. This project exposed me to a wide array of practical skills related to database design, data management, and implementation, including proposal writing. The collective approach we took to the project made these challenging aspects manageable. We divided the workload, and each team member's unique knowledge and skills became a shared resource, catalysing our collective learning. One such experience was learning to use the 'MySQL-connector-python module' in Python to establish a connection to a MySQL database, a skill imparted by a teammate with expertise in that area. This real-time learning experience further emphasized the power of diverse skills within a team, which not only enhanced my technical proficiency but also underscored the importance of collaboration in problem-solving.

Additionally, I gained invaluable insights into setting clear team goals and the critical role effective communication plays in expediting project completion. This team experience offered a platform for translating theoretical knowledge into practical, real-world application. The project was not merely an academic exercise; it mirrored real-world scenarios that deepened my understanding of the concepts and equipped me for future workplace tasks.

Furthermore, working as part of the team gave me a profound understanding of how to balance different tasks within a project, a critical skill in project management. The responsibility of each team member was clearly defined and evenly distributed, ensuring every aspect of the project was adequately covered. In summary, this team project served as a dynamic learning platform, enhancing my practical skills, fostering collaborative learning, and improving my ability to manage and divide tasks effectively. It has been instrumental in my growth and preparation for similar tasks in a professional setting.

The collaborative discussions were very informative and enriching. It helped me gain a deep and thorough understanding of the opportunities and challenges associated with IoT and the importance of data security in today's digital landscape. It helped me better understand the rationale behind IoT, the risks associated with data collection, and the need for robust data protection regulations like GDPR. Moreover, comparing different data protection laws and reading about the different laws in each country by the different participants offered insights into the variations in standards and regulations across different regions, highlighting the complexities and challenges in global data protection and cybersecurity.

As stated in my e-portfolio, the integration of web scraping, data cleaning, data normalization, and database design plays a significant role in the hospitality industry. It facilitates efficient extraction and analysis of data from various sources, such as booking portals, thereby allowing a comprehensive understanding of industry trends. Automation skills in these areas provide an effective means of managing large datasets, improving data quality, and deriving valuable insights. A well-structured, normalized database enhances data integrity and consistency, which is paramount in customer-focused sectors like hospitality. It further simplifies data modelling in platforms like Power BI, leading to detailed visualizations and quicker data refresh times. In addition, it ensures data security and scalability. Evaluating tools such as the ReviewPro API aligns with the industry's need for analysing customer feedback. Ultimately, these practices foster improved data management, operational efficiency, and business insights within the hospitality industry.

Overall, the module was intense, challenging and required a lot of effort but I feel that by studying this module, I gained a comprehensive set of skills that would enable me to collect, clean, transform, store, and analyse data effectively and efficiently. These are highly sought-after skills in many fields, including data science, data analytics, machine learning, and more and I look forward to applying what I've learned to future projects and continuing to explore these topics further.

**References:**

Uğur, N.G. and Turan, A.H. (2020) ‘Understanding big data’, Big Data Analytics for Sustainable Computing, pp. 1–29. doi:10.4018/978-1-5225-9750-6.ch001.

Oladipupo, M.A. et al. (2023) ‘An automated python script for data cleaning and labeling using machine learning technique’, Informatica, 47(6). doi:10.31449/inf.v47i6.4474.

Coronel, C. and Morris, S. (2023) Database systems: Design, implementation, &amp; management. Boston, MA: Cengage.

Pennay, A.E. and Measham, F.C. (2016) ‘The normalisation thesis – 20 years later’, Drugs: Education, Prevention and Policy, 23(3), pp. 187–189. doi:10.3109/09687637.2016.1173649.

What is an application programming interface (API)? (no date) IBM. Available at: https://www.ibm.com/topics/api (Accessed: 24 July 2023).

Basim Alwan, H. and Ku-Mahamud, K.R. (2020) ‘Big data: Definition, characteristics, life cycle, applications, and challenges’, IOP Conference Series: Materials Science and Engineering, 769(1), p. 012007. doi:10.1088/1757-899x/769/1/012007.