Personal Reflection

Artjola and I formed a highly effective team, demonstrating excellent collaboration throughout our project. Our primary focus was on exploring real-world applications in Big Data, particularly understanding the implementation of Big Data techniques on real-time data. We both shortlisted two intriguing research papers that addressed real-world road traffic problems and the ethics of Smart Cities using Big Data techniques.

We evenly divided the workload, delving into the research papers and reference documents. During our study, we decided to implement a live data stream using Apache Kafka, recognizing it as a fundamental and crucial phase for subsequent predictions and analysis in real-world applications. As data volumes continue to grow, handling and processing data efficiently has become increasingly important, especially considering ethics, data privacy, and security.

In the course of our work, we encountered several challenges related to Intelligent Transportation Systems, which we documented in our final project report. Our discussions on ethics throughout the course were particularly influential, providing motivation and insight as we tackled these challenges.

The visit to the data center was an eye-opening experience. It provided us with invaluable practical insights into the intricate details of scalable computation, data management, and the ethical considerations involved in handling vast amounts of data. Learning from John Licci, the director of the CUNY Data Center, provided us with practical insights into the nuts and bolts of scalable computation, data management, and the ethical considerations of handling vast amounts of data. John's explanations underscored the complexities involved in setting up and maintaining such systems. One concept that particularly caught my attention was Disaster Recovery (DR). This process involves replicating data to an alternate server to ensure continuity in the event of a major production failure. The idea of having a secondary or disaster recovery site, often geographically distant from the primary data center, exemplifies the critical importance of data redundancy and resilience. DR strategies ensure that business operations can quickly resume, minimizing downtime and data loss. This visit helped us grasp the real-world applications of the theoretical concepts we had studied, illustrating how companies safeguard their data and maintain operational continuity amidst potential disruptions.

Another fascinating aspect we learned about was the use of optical cables for data transfer. Optical fibre technology enables the transmission of data as pulses of light, which can travel at extremely high speeds with minimal loss of signal quality. We learned how his technology is critical for supporting the vast amounts of data transfer required in modern data centres, providing the bandwidth necessary for real-time data processing and high-speed internet connections. The efficiency and reliability of optical cables make them an integral part of the infrastructure supporting scalable computation.

Overall, this experience bridged the gap between academic learning and practical application, enriching our knowledge and preparing us to address real-world challenges in data management and scalable computation. I am grateful for this opportunity, which has undoubtedly broadened our perspectives and equipped us with valuable skills for our future careers.

Moreover, the discussions on ethical data management emphasized the responsibility that comes with handling sensitive information. Implementing robust DR systems not only

protects business interests but also ensures compliance with data protection regulations and ethical standards, safeguarding user privacy and maintaining trust.

This experience significantly enriched our project, grounding our theoretical knowledge in real-world practice.

Finally, I would like to express my gratitude to you, Professor, for organizing the data center tour. It was a unique experience that broadened our perspectives and enhanced our understanding of the practical aspects of Big Data and scalable computation.