A picture containing logo

Description automatically generated

***Data Storage Solutions for Data Analytics: CA1***

Submitted By : Vaibhav Tiwari  
Student Number : 10621051  
Lecturer: Dr Shazia A Afzal

**In this report, I have outlined my contributions and learning experiences regarding the assigned Continuous assessment. Our team evaluated multiple datasets and ultimately chose the IPL dataset, which contains all the match details played from 2008 to 2019, including match winner, player of the match, toss decision, win by runs, and other important match information. By analysing this dataset, we were able to provide valuable insights to our stakeholders on player and team performance. Fantasy cricket platforms could also utilize this data to estimate team capabilities based on statistical performance. Investors could use this data analysis to avoid overspending and choose key players wisely.**

**We sourced the data from a CSV file and created an ERD, we decided to adopt star schema when building the data warehouse since it lowers data redundancy, makes it easier to retrieve data, and enhances query efficiency. To build the star schema we used Draw.io to create the Relational Diagram. We identified the parent and child parameters by considering the categories of records and entities. Through the various stages of the ETL process, I learned about concepts such as lookup, derived column, connection manager, and various functions available in the SSIS toolbox. We then used SSRS for reporting and divided the reports among team members, which helped me improve my understanding of report development.**

**As a team member, I utilized Tableau to generate visual representations of our database needs. I designed a few reports using drag-and-drop functionality, which helped me become familiar with the user-friendly interface and various visualizations based on distinct data combinations. I integrated all the reports into a single dashboard for a comprehensive summary of our visuals.**

**This assignment allowed me to gain confidence in SQL, ETL, and related technologies such as SSIS, SSRS, and Tableau. I also learned how to approach data from a business perspective and manipulate it to meet business requirements. Additionally, I had the opportunity to learn about a new type of database called Neo4j from scratch. I found help regarding the same from the lecture recordings and on-demand lessons available on moodle. Still learning and exploring the same. My teammates guided me through each step on this process and brought me on par with them. I used Neo4J to learn how to do CQL queries. I discovered how to build the nodes and relationships among the tables. I also gained knowledge about how to retrieve and display data. The queries in CQL utilize less lines of code than the ones in SQL. Additionally, CQL allows us to visualize node relationships in a way that SQL does not.**

**We compared the performance of Relational and Graph databases once we created the reports for both databases. We found that Relational databases are better for structured data and managing transactions, while Graph databases are more proficient in handling intricate, interrelated data with connections. Choosing between the two depends on the unique requirements of the application and the type of data being managed.**

**After completing the assessment, I can conclusively understand a data in terms of the different stages of CRISP-DM and can implement various stages of the same, I am keen on understanding how SSIS, SSRS and Neo4j work in daily enterprise solutions.**