Jason Domzalski

Dr. Shamsuddin

CIS-2532-NET01

14 May 2022

**Assignment 12 Question 1**

Ever since the explosion of computing technology more and more things have become digital. This digital age has in turned created vast amounts of data, or “big data” as it is often referred to today. However, not all data is the same, some is in the form of structured data which are your relational databases which are manipulated with SQL and vast majority is unstructured data in the forms of JSON and XML documents. To cope with unstructured data there has a been a rise in new database types such as NoSQL databases, NewSQL databases and more not mentioned here. The end goal for most of this data is the same, be able to sort through the data and run queries to find information we are specifically searching for. And while running queries on big data is important, nothing is more important than maintaining the integrity of the data itself within our databases. As mentioned in the textbook, ACID, which stands for Atomicity, Consistency, Isolation, Durability are key points of relational databases. Without ACID, there could be the potential for someone to withdraw money they don’t have from an ATM or attempt two transactions at once to get double the money. However, in applications where these requirements are not needed such as some NoSQL databases a BASE method is followed which stands for Basic Availability, Soft-State, Eventual Consistency. An important aspect of Big Data is the three v’s, which are Volume, Velocity, and Variety. Volume is the data itself rather it be transaction data, data from social media posts or even data from flight sensors in an aircraft. Velocity refers to how fast the data is generated and how fast it can be dealt with. Data can be generated from sensors every millisecond and information from online platforms can suddenly spike from something trending or a sale and the big data needs to account for changes in velocity of data. Variety refers to the different types of data rather it be structured data from relational databases or data from social media which is unstructured or from emails which are also unstructured.