**Part 1: Design Document**

**A.  Select one of the provided scenarios to complete the following:** Scenario 2

**1.  Describe a business problem that can be solved with a database solution and is in alignment with the chosen scenario.**

EcoMart is an emerging company that is in the marketplace industry. Their goal is to foster sustainability and environmental consciousness. EcoMart has a platform where customers can find ethically sourced, sustainable, and eco-friendly products, such as groceries, apparel, home goods, and personal care items. The business problem that can be solved is to allow flexibility in scaling their platform and also allow for faster lookups of products by the customers, all in an effort to reach their goal of being sustainable. The solution will be to implement a relational database in third normal form. In addition, the database solution will provide robust security measures (encryption, access controls, and audit logging), ensure data consistency, and ensure long-term stability and reliability of the database implementation.

**2.  Propose a data structure to solve the identified business problem.**

A data structure to solve the business problem is to format the data into a relational database. Currently, the data is in a raw csv file, so it will need to be ingested into the database and be normalized into third normal form to eliminate transitive dependencies. Transitive dependencies can cause database issues, data redundancy, anomalies, and inconsistencies (Bhalla, 2023).

In order to solve the business issue with a relational database, the raw data will be ingested, then it will be ensured that all the values are atomic, transitive dependencies will be eliminated, and every non-primary or foreign key column will depend on the primary key of the table by splitting the raw data into multiple tables, such as a table for the region, a table for the country, a table for item types, and a table that holds all the sales record information.

**3.  Justify why a database solution will solve the identified business problem.**

This database solution will solve the business problem because it will allow the database to be scalable and consistent. By normalizing the table, this will make sure data anomalies will occur less frequently as there is only one place to update information instead of multiple different places. It will also ensure data integrity as required fields could be implemented so updates or inserts can’t happen unless those fields are being included. This solution will also help with scalability because new record types or regions can be added without changing the existing table that holds the sales record. Also, by splitting the data into multiple columns, queries can run faster since they will be smaller, which will also allow for faster joins or even lookups. Lastly, with a database implementation, security can be implemented as the server can be protected by a username and password, or even multifactor authentication. Audit logging can also be implemented and it can track when a table was last updated and by which user.

**4.  Explain how the business data will be used within the database solution.**

The business data will be used within the database solution because the tables that are being created are determined by the raw csv file that is provided. Multiple tables will be created to implement this solution, and the data from the csv file will be used to fill those tables. These tables will then be able to be joined with each other so that the business can still understand its sales records. By implementing the database solution, the business can also find all the regions or countries in which they are selling their products by looking at those individual tables. They could understand how many of each type of product they are selling, perform calculations, or aggregate calculations on the data to understand revenue, how many items sold, etc Lastly, multiple different queries can be run to understand more about business insights.

**B.  Create a logical data model for storing data in the database solution.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**C.  Describe the database objects and storage, identifying the file attributes within the database solution.**

There are a total of 4objects/tables: Country, Region, Item\_Type, and Sales\_Records.

Each table has a primary key, for Country, there is a country\_id, for Region there is a region\_id, for item\_type there is a item\_type\_id, and lastly, for sales\_records, the primary key is order\_id. The sales\_records table has foreign keys that join together with the other tables on the ids to get the country name, region name, and item\_type name.

The country table has 2 columns: Country\_ID which is an int and Country, which is a varchar(1000).

The region table has 2 columns: Region\_ID which is an int and Region, which is a varchar(1000)

The item\_type table also has 2 columns: Item\_Type\_ID which is an int and item\_type which is a varchar(1000)

Lastly, the sales\_records table has multiple columns:

Sales\_channel: varchar(500)

Order\_priority: varchar(5)

Order\_date: Date

Order\_ID: Int

Ship\_date: Date

Units\_sold: Int

Unit\_price: Decimal(50,2)

Unit\_cost: Decimal(50,2)

Total\_revenue: Decimal(50,2)

Total\_cost: Decimal(50,2)

Total\_profit: Decimal(50,2)

Region\_ID: Int

Country\_ID: Int

Item\_Type\_ID: Int

**D.  Discuss how the proposed database design addresses scalability concerns, including strategies that align with the chosen scenario.**

The proposed database design addresses scalabilityby making it easier to add more tables and information because the database design is in third normal form. It also allows for easier maintainability because there is less data that needs to be updated when tables are split out. Having the data in third normal form also eliminates data redundancy, which allows the database to be more efficient, which also helps the database structure grow in size. In addition, if there are changes that need to be made to the tables that contain additional information, such as country or region, instead of changing all the fields in the singular table with all the information, only the country table or region table would need to be updated.

**E.  Outline the privacy and security measures that should be implemented in the proposed database design.**

Any data should be protected, especially since it contains company information. So, privacy and security measures that should be implemented could mean that the Database requires a login and 2 factor authentication. That way only verified users can view the data, and in addition, each user could have a different set of permissions. For example, engineers could have write, read, and update permissions, while analysts may only have read permissions. By having a login, this ensures privacy and security as only verified users can view the data and it can’t be accessed by the public. Another measure that could be implemented is audit logging. This will ensure that all past operations are recorded, so if something goes wrong, it can be tracked to when and what caused the issue. Lastly, another measure that could be implemented is stress testing and database security testing (*What Is Database Security Testing - Complete Guide*, n.d.). This helps discover new ways that the database could be hacked and will allow for implementations to be in place before it becomes an issue.