

Report

Group 14

Report

Part 1: Database Design and Implementation

1.1 E-R Diagram Design

At the initial creation, our group developed first Entity-Relationship (E-R) diagram with nine entities including Supplier, Product, Ads, Category, Inventory, Customer, Order, Payment, and Shipping – connected by eleven relationships. Each entity contains a primary key for unique identification of each instance in the entities, related attributes, and a foreign key as needed. However, through the deeper examination when converting the diagram into the schema, it was found that there were several errors in the initial E-R diagram, making the database incompatible with the third normal form (3NF). The following adjustments - such as adding and removing entities, relationships, and attributes - were done to solved the problem.

Firstly, we decided that Order, Category, and Ads should be remove. The Order entity was more appropriate to be a relationship between the Customer entity and Product entity to prevent redundant functionality and simplify the diagram. The Category entity was eliminated and changed to an attribute under the Product entity to made the synthetic data we generated be more realistic and reasonable. Ads were removed due to the complexity of generating and maintaining such data. (CAN ADD MORE REASON)

Secondly, we detected some looping relationships which created redundancy in the data storage of the database. For example, there is a relationship between Supplier that supplies Product, a relationship between Supplier that supplies Inventory, and a relationship of Inventory that has Product. By eliminating the relationship between Supplier and Inventory and utilising the `PRODUCT_ID` as a linkage, redundancy was decreased. Similar relationship issues also occurred around Customers, Payment, and Shipping. We decided to modify those relationships resulting in a reduction from eleven to five total relationships.

Thirdly, we made changes in some certain attributes. We had spotted three non-atomic attributes including `SHIPMENT_ADDRESS`, `BILLING_ADDRESS`, and `DESCRIPTION`. For the address, more attributes for city, country, and zip code were added. The `DESCRIPTION`

attribute was deleted since we already had `PRODUCT_CATEGORY` attribute which has a similar function of describing the product. Moreover, we removed the `PRODUCT_ID` from the Supplier entity as the `SUPPLIER_ID` attribute was already existed in the Product entity. This enables us to use Supplier as a starting point of the schema creation since it is the entity without the foreign key. Furthermore, an `INVENTORY_ID` was created to solve the lack of a primary key in Inventory entity.

Lastly, some modifications were made to the relationship types. The Supplier-Product relationship was changed from N:M to N:1, taking into account the possibility that a single product may be provided by several providers. Similar to this, the Inventory-Product relationship was changed from a 1:1 to a N:1 considering the possibility of storing the same product across multiple shelves depending on the product quantity. Lastly, the Customer-Payment relationship was redefined as 1:N to account for situations when a single customer may make several payments.

After all of these adjustments, a final E-R diagram that could accommodate a database in 3NF was produced. In order to maintain data integrity and reduce redundancy, the resultant diagram consists of six entities: Supplier, Product, Inventory, Customer, Payment, and Shipping. These entities are related by five relationships.

1.2 SQL Database Schema Creation

The initial schema was created based on the initial E-R Diagram containing nine tables according to the first nine entities our group identified. However, the subsequent investigation revealed that the database lacked normalization in form of 3NF. This resulted in redundancies and dependency issues as described in the E-R Diagram Design phase.

In order to rectify this, we took care of unnecessary entities, redundant relationships, and non-atomic attributes. After this process, the final database schema consisting of six tables - Supplier, Product, Inventory, Customer, Payment, and Shipping – according to the final E-R diagram was created. With utilisation of SQL code, all attributes were added as the columns of the tables. The primary key and foreign key of each entity were also determined. Additionally, we also carefully specified the data type of each attribute or each column of the table- for example, `SUPPLIER_NAME` should be text, `PRICE` should be numeric, and `CUSTOMER_BIRTHDAY` should be date. This is important as it helps in organizing and managing the data while maintaining consistency and integrity.