Project Report: Database Design for an ABC Electronics Retailer

Group 3 - Student IDs: 12695125, 12695120, 12691429

Deciphering Big Data

University of Essex Online

10 June 2024

Table of Contents

Introduction 3

Logical Design 3

Proposed Database Model 3

Data Items/Entities and Attributes 3

Table 1. Database Entities and Attributes 4

Relationships and Associations 4

Data Types and Formats 5

Table 2. Data Types 5

Choosing a Database Management System (DBMS) 5

Data Management Pipeline Process 6

Data Capturing and Source 6

Methods for Cleaning Data 6

Data Cleaning Process Stages 6

Normalization 7

Critical Evaluation 7

Data Wrangling 7

Methodologies and Tools 7

Conclusion 8

References 8

Appendices 10

Appendix 1. Operational Workflow of ABC Electronics 10

Introduction

Our team, consisting of software consultants and developers, has been hired by ABC Electronics, a mid-sized retailer that specialises in electronics and home appliances, to design and construct a logical database. ABC Electronics operate through two sales streams; physical stores and an online store. As shown in **Appendix 1**, while these are two different methods for customers, the processes are similar: browsing, selection, and purchase of products. The store checks inventory, orders more if necessary, confirms payment, and fulfils the order. For a physical store, this means immediate handover of purchased items, whereas the online store primarily utilizes delivery. All these processes provide feedback and data for analysis to aid decision-making. The objective of this project is to develop a comprehensive database system that facilitates user access, enhances data management, and expedites the processes of data retrieval and modification. Optimal inventory utilisation, real-time consumer engagement, and well-informed decision-making processes are all enhanced by effective data management in retail, which uses big data analytics (Kameswari et al., 2024 and Hossain et al., 2024). Taipalus (2024) asserts that managing substantial amounts of transactional data and delivering useful insights require a well-structured database architecture. The logical architecture, suggested database model, database management system (DBMS) choice, and data management pipeline procedure are all described in this study.

Logical Design

Proposed Database Model

As there are many interconnected processes, we suggest utilising a relational database approach for ABC Electronics' database. The requirements of the customer depend on this model's capability for complicated queries and guarantee of data integrity.

Data Items/Entities and Attributes

Defining important entities and their characteristics are a critical component of the database's logical architecture. Based on the workflow described in **Appendix 1**, we have identified the principal entities and their attributes in **Table 1** below.

Table 1. Database Entities and Attributes

Entity	Attributes
Customer	CustomerID*, FirstName, LastName, Email, PhoneNumber, Address, JoinDate
Product	ProductID*, ProductName, Category, Brand, Price, StockQuantity, SupplierID**
Order	OrderID*, OrderDate, CustomerID**, EmployeeID**, TotalAmount
OrderDetails	OrderDetailsID*, OrderID**, ProductID**, Quantity, UnitPrice, DeliveryRequested, DeliveryID**, CompletionStatus, CompletionDate
Delivery	DeliveryID*, DeliveryDate, DeliveryAddress, DeliveryPostcode, EmployeeID**, CompletionStatus
StockOrders	StockOrderID*, StockOrderDate, ProductID**, EmployeeID**, Quantity, TotalAmount
Supplier	SupplierID*, SupplierName, ContactName, PhoneNumber, Email
Employee	EmployeeID*, FirstName, LastName, Email, PhoneNumber, Position, HireDate, Salary
Payments	PaymentID*, OrderID**, StockOrderID**,EmployeeID**, PaymentAmount, PaymentDate, Purpose
Date	DateID*, Date, Day, Month, Year, Quarter

**Primary Key, **Foreign Key*

Relationships and Associations

There are a number of these entities that are related to each other. These are primarily One (1) to Many (M) relationships, although there are also a couple of One (1) to One(1) relationships. The relationships between these entities are listed below:

- Customer (1) → (M) Order
- Product (1) → (M) OrderDetails, StockOrders
- Order (1) → (M) OrderDetails
- Order (1) → (1) Payments
- OrderDetails (1) → (M) Delivery
- Employee (1) → (M) Order, Payments, Delivery, StockOrders

- Supplier (1) → (M) Product
- StockOrders (1) → (1) Payments
- Date (1) → (M) Order, OrderDetails, Customer, Employee, Delivery, StockOrders, Payments

Data Types and Formats

The database will use a variety of data types including, Numeric, String, Boolean, and Date. **Table 2** details the data types for each field.

Table 2. Data Types

Data Type	Field
Integer	CustomerID, ProductID, OrderID, OrderDetailsID, DeliveryID, StockOrderID, SupplierID, EmployeeID, PaymentID, DateID, Quantity, Day, Month, Quarter, Year
String	FirstName, LastName, Email, Address, ProductName, Category, Brand, DeliveryAddress, DeliveryPostcode, SupplierName, ContactName, Position, Purpose, PhoneNumber
Decimal	StockQuantity, Price, TotalAmount, UnitPrice, Salary, PaymentAmount
Boolean	DeliveryRequested, CompletionStatus
Date	JoinDate, OrderDate, HireDate, DeliveryDate, PaymentDate, StockOrderDate, CompletionDate, Date

Choosing a Database Management System (DBMS)

We suggest using MySQL as the DBMS in light of ABC Electronics' needs for user access, data processing, and storage for the following reasons (Mirayala, 2024 and Ma & Wang, 2024):

1. **Scalability**: As ABC Electronics expands, MySQL can grow to accommodate higher data volumes.
2. **Performance**: It provides excellent performance for workloads involving a lot of reading, which is advantageous for retail operations.
3. **Security**: To safeguard confidential customer and transactional data, MySQL has strong security measures.
4. **Cost-Effectiveness**: MySQL is an open-source database management system that is affordable and has a sizable community behind it.

Data Management Pipeline Process

Data Capturing and Source

The primary data capture sources are online orders, inventory management, point-of-sale systems, existing operational databases, and third-party vendors. Extraction, Transformation, and Loading (ETL) procedures will be used to incorporate each data source into the database. In order to minimise the probability for errors, the data capture should minimise the use of free-text where possible, for example, by using drop down boxes or postcode lookups to find the full address.

Methods for Cleaning Data

To guarantee data accuracy and quality, data cleansing is essential. The below methodologies shall be executed: (Cuzzocrea et al., 2011).

1. **Data Validation**: Ensuring that data complies with specified requirements and formats.
2. **Managing Missing Values**: Applying context-based imputation or deletion strategies.
3. **Eliminating Duplicates**: To preserve data integrity, duplicate records must be found and eliminated.
4. **Standardization and Normalization**: In order to better understand data trends, standardisation and normalisation include computing new values using current data values and eliminating outliers to provide uniform data formats across all records.

Data Cleaning Process Stages

The stages carried out during the data cleaning process include:

1. **Initial Assessment**: Evaluating the raw data to understand its structure and identify obvious issues.
2. **Data Profiling**: Analysing data patterns, distributions, and relationships.
3. **Cleaning and Transformation**: Applying the cleaning techniques mentioned above.
4. **Validation**: Ensuring that the cleaned data meets the quality standards required for accurate analysis.

5. **Documentation**: Keeping detailed records of the cleaning process for transparency and reproducibility.

Normalization

Data integrity and anomaly prevention depend on normalisation. The standard forms listed below will be used:

- **First Normal Form (1NF)**: Guarantees that every row is distinct and every column has atomic values.
- **The Second Normal Form (2NF)**: Verifies that every non-key characteristic depends entirely on the primary key for functionality.
- **Third Normal Form (3NF)**: Removes transitive dependencies by guaranteeing that all non-key characteristics are only reliant on the primary key.

Critical Evaluation

Data Wrangling

Preparing and converting unprocessed data into a format that may be used is known as data wrangling. Managing a variety of data sources, resolving inconsistent or missing data, and maintaining data integrity are major problems. These difficulties can be lessened by making good use of technologies like SQL queries, Tableau, and Python's Pandas module (Panda & Patra, 2015).

Methodologies and Tools

For this project, the following methodologies and tools were adopted:

- **ETL Tools**: Talend or Apache NiFi for extracting, transforming, and loading data.
- **Data Cleaning**: Python libraries (Pandas, NumPy) for data manipulation and cleaning.
- **Database Management**: MySQL for efficient data storage and retrieval.
- **Data Visualization**: Tableau for presenting data insights visually.

Conclusion

This report provides ABC Electronics with a detailed strategy on how to design and construct a logical database. We want to improve the client's data management skills by utilising a relational database model and putting strong data management practices in place, guaranteeing effective data modification, access, and storage. Our suggested MySQL solution satisfies the client's requirements for cost-effectiveness, scalability, performance, and security. By recording employee actions and offering a strong framework for date-based analysis and reporting, the addition of the Employee and Date columns will significantly improve the database.

References

Kameswari, P., Ramesh, V., Ishwarappa Bhavikatti, V., Santhoshi Gondesi, G., et al. 2024. Analyzing the role of big data and its effects on the retail industry. *Web Intelligence*. 22 (2), pp.1-19. DOI: https://doi.org/10.3233/WEB-230027

Hossain, M.A., Akter, S., Yanamandram, V. and Strong, C. 2024. Navigating the platform economy: Crafting a customer analytics capability instrument. *Journal of Business Research*, 170. DOI: https://doi.org/10.1016/j.jbusres.2023.114260

Cuzzocrea, A., Song, I.-Y., & Davis, K. C. 2011. Analytics over large-scale multidimensional data: the big data revolution! *Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP*, 101-104. DOI: https://doi.org/10.1145/2064676.2064695

Panda, M., & Patra, M. R. 2015. Data wrangling: Techniques and challenges of big data. *International Journal of Applied Research*, 1(10), 997-1001. DOI: https://doi.org/10.5441/002/edbt.2016.44

Taipalus, T. 2024. Database management system performance comparisons: A systematic literature review. *Journal of Systems and Software*, 208. DOI: https://doi.org/10.1016/j.jss.2023.111872

Mirayala, N.K. 2024. Evolving trends in open-source RDBMS: Performance, scalability and security insights. *International Journal of Science and Technology*, 12(2). DOI: https://doi.org/10.21275/SR24126224648

Ma, X. and Wang, Z. 2024. Computer security technology in E-commerce platform business model construction. *Heliyon*, 10(7), e28571. DOI: https://doi.org/10.1016/j.heliyon.2024.e28571

Appendices

Appendix 1. Operational Workflow of ABC Electronics