# Development Team Project: Project Report

## 1. Introduction:

This report contains an advised database solution for the inventory management system of an online bookstore. We detail how the data management pipeline is put together and which database options we choose to optimize the customer’s operations.

## 2. User Requirements:

Initially, we define the user requirements of the database. It will be used to undertake automatic stock changes once purchases or sales occur. Books can then be managed and tracked by genre, publisher, author, title, price and ISBN. Whenever business events lead to changes in the inventory, updates must be made using the database (Luo, 2019).

Orders will be tracked, keeping inventory up to date. The database also tracks the delivery status and invoice information. Payments and their related details are processed once the order is completed.

Additionally, the database will generate reports on inventory levels, turnover rates, best-selling stock, and sales trends, and integrate with accounting software to manage revenue, costs, and inventory value (Samaan, 2017).

## 3. Data Items and Relationships:

Next, we consider the data types and formats that run through the database and the records’ interrelations. To present the different data entries, their types and interrelations, we have plotted an Entity-Relationship-diagram using Canvas (2024) and best practices by Miro (n.d.).

IncomingOrderItem 
ISBN (PK) (varchar (255)) 
Purchase ID (FK) (integer (10)) 
Supplier ID (FK) (Integer (10) 
Quantity (Integer (10)) 
Supplier 
Supplier ID (PK) (integer (10)) 
Name (varchar (255)) 
Supplier contact narne (varchar (255)) 
Address (varchar (255)) 
Email (varchar (255)) 
Phone (unique) (varchar) 
Website (varchar (255)) 
IncomingOrder 
Purchase ID (PK) (integer (10)) 
Supplier ID (FK) (integer (10)) 
ISBN (FK) (varchar (255)) 
Date placed (datetime) 
Purchase status (varchar (255)) 
Outgoingorder 
Order ID (PK) (integer (10)) 
ISBN (FK) (varchar (255)) 
Customer ID (FK) (Integer (10)) 
Date placed (datetime) 
Shipping address (varchar (255)) 
Order status (varchar (255)) 
ISBN (PR) (varchar (255)) 
Title (varchar (255)) 
Author (varchar (255)) 
Publisher (varchar (255)) 
Publication Date (datetime) 
Genre (varchar (255)) 
Language (varchar (255)) 
Price (numeric (19. O)) 
Quantity in stock (integer (10) 
Ouantity on order (integer (10) 
ISBN (PK) (varchar (255)) 
Order ID (FK) (integer (10)) 
Customer D (FK) (Integer 
(10) 
Quantity (Integer (10)) 
Customer D (PK) (integer (10)) 
Name (varchar (255)) 
Address (varchar (255)) 
Email (varchar (255)) 
Phone (unique) (varchar) 
Password (varchar (255)) 

#### Graphic 1: ER-diagram of the proposed database design

The database is expected to manage data for different parts of the online bookstore’s business operations. The database manages data for books (ISBN, title, author, publisher, etc.) and tracks inventory (stock levels, incoming/outgoing orders). The diagram also shows the data types of each variable. E.g., it shows that the books’ titles are denoted as strings and that that the publication date is recorded in date format. The diagram presents the interrelations of different data types. We see, e.g., that the data on books is, via the ISBN, related to the data on the incoming order and the outgoing order item, illustrated by the connecting lines in the diagram representing the connecting keys.

Furthermore, the diagram shows the variables, data types and interrelations of all different parts of the inventory management process, i.e., books, incoming orders, incoming order items, outgoing orders, outgoing order items, suppliers and customers.

## 4. Chosen Database Type:

We selected a relational database, which is ideal for structured data with clear relationships and can be queried using SQL. SQL offers robust data integrity through constraints and transactions. Given the complexity of the bookstore's operations, a relational database will efficiently handle relationships and transactions, which ultimately led us to the choice of an SQL-based system like MySQL.

## 5. Data Cleaning Process:

The process of data cleaning ensures that the data records are up-to-date and accurate. As Lamb (2023) explains, this is especially important because purchase and supply decisions of the business are related to the stocking. We first check the incoming data for quality metrics like faults and missing data. Outliers should be considered with special care as it’s necessary to determine whether they hold an important meaning or not. Referencing Sarkar and Roychowdhury (2019), the wrongful decision to delete or not delete outliers would cause analytical problems. Furthermore, referencing Mertz (2021), special importance is allocated to developing sole access points to the data for the different stakeholders, which leads to our design of a single database.

## 6. Storage and Access Management Options:

We propose cloud storage for the database, which allows for unlimited scalability and a pay-as-you-go model (Brantner et al., 2008). Cloud storage is cost-effective, scaling according to demand, which is advantageous for an e-commerce business. Cloud providers like Google Cloud and Microsoft Azure offer flexible storage solutions (2024).

Storing data in the cloud affects identity access management. We recommend different security groups with varying access levels, making it easy to add or remove users while ensuring compliance with security standards, as discussed by Waters (2016).

## 7. Critical Evaluation of the Data Management Pipeline:

As discussed above, data capturing will be stored on the cloud due to its flexibility and scalability. However, other factors, such as the dependency on the cloud system provider, reliability, security, availability, and maintainability, also need to be considered (Song & Sohn, 2022). Furthermore, vendor lock-in is another factor that must be considered. If only proprietary services with low interoperability are used, it might be difficult to tackle issues of low performance and change providers if the cloud provider provides insufficient service. This must be mitigated, as poor service will ultimately reflect on the bookstore’s customers and the online bookstore business (Alhosban et al., 2024).

Another important factor in the data management pipeline is data capturing and compliance. Data, when stored, must comply with regulations like GDPR. Cloud service providers have servers in different locations, and the service must be scrutinized to ensure that clients' private data is processed correctly according to regional compliance (Alhosban et al., 2024). In addition, data cleaning is essential in the data management pipeline. The data-cleaning process is the most time-consuming part. Cleaning techniques like identification of missing data, handling of outliers, and removal of duplicate records are critical in the inventory management system. These processes could create a bottleneck if the cleaning is done manually. On the other hand, the possibility of automated data cleaning increases efficiency compared to manual methods (Valêncio et al., 2020). Furthermore, Peng et al. (2024) state that human labor costs can be deducted from the data cleaning process with the application of a deep reinforcement learning method that integrates the two processes in data cleaning, error detection and data repair, by understanding the best order of operations that cleans the error.

In conclusion, the proposed solutions aim to create a scalable, secure, and reliable database management system that meets the operational needs of the online bookstore and provides a positive customer experience.

## References:

* Alhosban, A., Pesingu, S. & Kalyanam, K. (2024) CVL: A Cloud Vendor Lock-In Prediction Framework. Mathematics (Basel) 12(3) 387-. DOI: <https://doi.org/10.3390/math12030387>
* Brantner, M. et al. (2008) Building a Database in the Cloud. Publications of the University of Leipzig. Available at: https://pubs.dbs.uni-leipzig.de/cloud/sites/cloud.pubs.dbs.uni-leipzig.de/files/Brantner2008BuildingaDatabaseintheCloud.pdf
* Canva (2024) Entity Relationship Team Whiteboard. Available at: https://www.canva.com/templates/EAFnnjJVQig-entity-relationship-diagram-team-whiteboard (Accessed: 01. September 2024).
* Google Cloud (2024) Google Cloud. Virtual machines for every workload. Available at: https://cloud.google.com/products/compute?hl=en#pricing (Accessed 06. September 2024)
* Lamb, D. (2023) digiteams-sa. The Importance of Clean Data for Effective Inventory Management. Available at: https://digiteams-sa.com/the-importance-of-clean-data-for-effective-inventory-management/#:~:text=Clean%20data%20can%20lead%20to%20cost%20savings%20by,inventory%20levels%20is%20critical%20to%20effective%20inventory%20management. (Accessed 06. September 2024)
* Luo, T. (2019) Traditional Book Stores Industry Reforming Based on the New Management System. Journal of Physics. Conference Series 1213(5): 52008-. DOI: <https://doi.org/10.1088/1742-6596/1213/5/052008>
* Mertz (2021) Cleaning Data for Effective Data Science: Doing the other 80% of the work with Python, R, and command-line tools. 1st ed. Packt Publishing.
* Microsoft Azure (2024) Microsoft Azure. Azure pricing. Available at: https://azure.microsoft.com/en-gb/pricing/ (Accessed 06. September 2024)
* Miro (n. d.) How to design a database schema in 7 steps. Available at: https://miro.com/diagramming/how-to-design-database-schema/#design-your-database-schema-in-miro (Accessed: 01. September 2024).
* Peng, J., Shen, D., Nie, T. & Kou, Y. (2024) RLclean: An unsupervised integrated data cleaning framework based on deep reinforcement learning. Information Sciences 682: 121281-. DOI: <https://doi.org/10.1016/j.ins.2024.121281>
* Samaan, S. S. (2017) Design and Implementation of a Pharmaceutical Inventory Database Management System. Ai-Khawarizmi Engineering Journal 13(1): 118–128. DOI: <https://doi.org/10.22153/kej.2017.08.008>
* Sarkar, T. & Roychowdhury, S. (2019) Data Wrangling with Python. 1st ed. Packt Publishing.
* Song, C. & Sohn, Y. (2022) The influence of dependability in cloud computing adoption. The Journal of Supercomputing 78(10): 12159–12201. DOI: <https://doi.org/10.1007/s11227-022-04346-1>
* Valêncio, C. R., Jardini, T., Martins, V. H. P., Colombini, A. C. & Fortes, M. Z. (2020) A system proposal for automated data cleaning environment. ITEGAM-JETIA 5(25). DOI: <https://doi.org/10.5935/jetia.v6i25.685>
* Waters, M. (2016) Working Paper. Evaluation of IAM as a Cloud Service. Available at: https://www.researchgate.net/profile/Michael\_Waters14/publication/311450332\_Evaluation\_of\_IAM\_as\_a\_Cloud\_Service/links/5846bfee08ae2d2175700892/Evaluation-of-IAM-as-a-Cloud-Service.pdf (Accessed 06. September 2024)