**Task 2: Non-Relational Database Design and Implementation**

Krescens Kok

Department of Technology, Western Governors University

D597: Data Management

Sean Pettersen

2025 July 11

**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 and transactions, all in an effort to reach their goal of being sustainable. The solution will be to implement a non-relational database since the data is currently in semi-structured JSON files. 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.  Justify why a NoSQL database solution will solve the identified business problem.**

A NoSQL database solution is well-suited to solve the business problem because it uses a non-relational data structure. They are designed to be more flexible in the schemas, which allows the database to scale easily and adapt to different data formats. NoSQL databases don’t require predefined schemas, which is beneficial for semi-structured or document data, such as JSON files. The flexible schema property makes it easier to create and manage collections in the database. Additionally, NoSQL databases are horizontally scalable, which means that more servers can be added to handle higher traffic and higher data volumes (Smallcombe, 2025). Since the data files are in a JSON format for this scenario, a relational database would not be beneficial since relational databases are used for tabular, structured data.

**3.  Identify a NoSQL database type to solve the identified business problem.**

For this solution, MongoDB will be used because it is a great tool for document databases that supports flexible schemas for storing data (MongoDB, n.d.). MongoDB has a BSON data format that has been inspired by JSON, which allows the database to have objects in a collection that have different information. For example, EcoMart may have ingredients listed for multiple cosmetic items, but may not for others. This flexibility makes it easy to ingest and work with raw, semi-structured data.

MongoDB stores and represents data in a document format, enabling developers or users to access and manipulate the data using a variety of programming languages, such as Python, JavaScript, etc. Lastly, MongoDB allows users to insertMany and updateMany to insert or update multiple documents simultaneously. These functions allow for a significant performance boost when compared to batch writing in traditional databases (*Advantages of MongoDB*, n.d.). This is particularly useful for EcoMart because the company may need to insert multiple transactional data at once, or they may need to update their product list, and will have the ability to update multiple products at once.

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

A NoSQL database solution is appropriate because the 2 JSON files are very different in structure. The cosmetic JSON has information about cosmetic products, including the price, ingredients, etc, while the groceries JSON has information about the purchase date, product, and the member of the transaction. Each JSON file will be created as a collection in the database, where the grocery collection will be used by searching for the items that members have bought or finding the last time a member purchased an item. This can be helpful for identifying trends in customers or seeing which products customers usually buy together. The cosmetic collection will be used to search up product information and identify whether or not their prices are comparable to other stores, or to figure out which products have better ratings and decide whether they want to continue selling that product or not. By using MongoDB, it is simple to import the JSON files as collections, that will read the files in as document data.

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

MongoDB supports both vertical and horizontal scaling. Vertical scaling increases the processing power of a single server, whereas horizontal scaling or sharding adds additional servers to handle larger workloads. MongoDB also includes an auto-scaling feature to meet the needs of the database. In addition, MongoDB also supports replication and horizontal partitioning. Replication involves copying data across multiple nodes to ensure high availability and fault tolerance. If one of the nodes goes down, the database remains operational, making it reliable. Partitioning distributes data across different nodes within a cluster, improving performance and scalability by reducing the amount of data each node handles during read and write operations (*Database Scaling*, n.d.).

Aligning this with EcoMarts solution of creating a non-relational database, the more transactions EcoMart receives, it will be easier to scale both vertically and horizontally. If a query is taking too long to run, the processing power can be increased to decrease the time it takes to run the query, or, more servers can be added to handle the complex queries. In addition, since MongoDB supports replication, if a server goes out, EcoMarts data will still be safe because the data has been replicated on another node.

Another feature that MongoDB supports is indexing, which improves the speed and efficiency of queries. According to GeeksforGeeks, an index is a special data structure that stores a subset of data in a way that allows MongoDB to quickly locate documents in a collection. This allows MongoDB to only scan the indexes instead of the entire collection and find the relevant data once the index value has been found. Therefore, aggregates, filters, and order by are all optimized by an index since the data are automatically ordered by the field that the index is created on (GeeksforGeeks, 2025a).

Since MongoDB supports indexing, this will be beneficial to EcoMarts solution because columns like “member\_number” can be indexed to find transactions made by certain members. Likewise, with the cosmetics data, “label” can be indexed to find certain information faster, for example, how many items of each label are being sold.

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

MongoDB handles security by providing authentication, access control, encryption, and user management to protect data and restrict unauthorized access (Team, n.d.). Since the grocery collection has member data, it could be sensitive information that should not be released to the public. By adding authentication measures, this will protect the data and ensure that only users are able to access it. In addition, adding encryption to the member IDs could also ensure that the members are not identifiable.

**Part 2: Implementation:**

**D.  Implement the proposed database design in the WGU Virtual Lab environment by completing the following:**

**1.  Write script to create a database instance named “D597 Task 2” using the appropriate query language, based on your design in Part 1. Provide a screenshot showing the script and the database instance in the platform.**

**A black background with white text

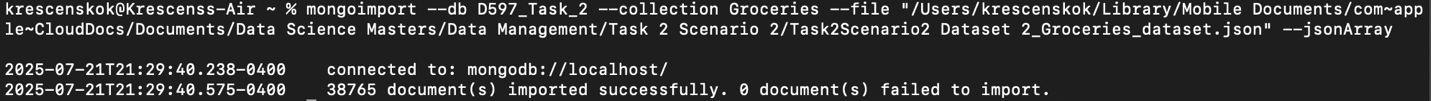
AI-generated content may be incorrect.**Using this script below will create the database if it does not already exist, but it won’t show up until there is data inserted into it or a collection is created.

**A screenshot of a computer

AI-generated content may be incorrect.**

**2.  Write script to insert or map the data records from the chosen scenario JSON files into the database instance. Provide a screenshot showing the script and the data correctly inserted or mapped into the database.**

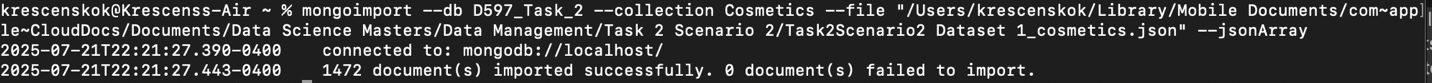
**Creating the Groceries collection and importing the data:**

****

**A screenshot of a computer

AI-generated content may be incorrect.**

**Creating the Cosmetics collection and importing the data:**

****

**A screenshot of a computer

AI-generated content may be incorrect.**

**3.  Write script for three queries to retrieve specific information from the database that will help to solve the identified business problem. Provide a screenshot showing the script for each query and each query successfully executed.**

**A. What are the top 20 member IDs with the most transactions?** This helps solve the business problem because EcoMart can target the top members with personalized, sustainable product recommendations. This will help increase the members’ loyalty and repeat purchases, which can help the company improve its sales and supply chain efficiency.

**A screen shot of a computer program

AI-generated content may be incorrect.** **Script:**

**Results:**

**A screen shot of a computer code

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.**

**B. Which top 5 items are purchased the most?** This helps answer the business problem because it finds the top 5 items that EcoMart is selling, which can help the company evaluate what customers are mostly buying, and if they are reaching their goal of fostering sustainability and environmental consciousness.

**Script:**

**A screen shot of a computer code

AI-generated content may be incorrect.**

**Results:**

**A screen shot of a computer code

AI-generated content may be incorrect.**

**C. Find how many items there are of each label type:** This helps solve the business problem because EcoMart can identify if they are selling too many products for the same label. If some labels have too many products, queries and filters could slow down, so EcoMart could decide if they should split out the items into different collections based on the label to continue scaling.

**A screenshot of a computer code

AI-generated content may be incorrect.** **Script:**

**Result:**

**A screen shot of a computer code

AI-generated content may be incorrect.**

**4.  Apply optimization techniques to improve the run time of your queries from part D3, providing output results via a screenshot.**

**Query A:**

A screen shot of a computer code

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.**Before Optimization:**

**After Optimization:** Optimizing with an index on “member\_number” allows MongoDB to process documents in sorted order by that field. The index allows MongoDB to scan the document in a sorted order instead of scanning through the document randomly. This means it can efficiently group documents by incrementally counting occurrences as it traverses the sorted index, rather than having to track all counts in memory simultaneously. This approach reduces memory usage and results in faster grouping performance. It is evident that the index helped optimize the query as the total time went from 75 milliseconds to 67 milliseconds.

**A screenshot of a computer

AI-generated content may be incorrect.**

A screen shot of a computer code

AI-generated content may be incorrect.

**Query B:**

**A screen shot of a computer code

AI-generated content may be incorrect.A screen shot of a computer code

AI-generated content may be incorrect.Before Optimization:**

**A screenshot of a computer

AI-generated content may be incorrect.After Optimization:** For this optimization, an index for “itemDescription” was also created on the Groceries collection. It is evident that the index helped optimize the query as the total time went from 81 milliseconds to 65 milliseconds.

**A screen shot of a computer code

AI-generated content may be incorrect.**

**Query C:**

**A screenshot of a computer code

AI-generated content may be incorrect.A screen shot of a computer code

AI-generated content may be incorrect.Before Optimization:**

**After Optimization:** For this optimization, an index for “Label” was also created on the Cosmetics collection. It is evident that the index helped optimize the query as the total time went from 28 milliseconds to 12 milliseconds.

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screen shot of a computer code

AI-generated content may be incorrect.**

**References**

*Advantages of MongoDB*. (n.d.). MongoDB. https://www.mongodb.com/resources/compare/advantages-of-mongodb

*Database scaling*. (n.d.). MongoDB. <https://www.mongodb.com/resources/basics/scaling>

GeeksforGeeks. (2025a, March 17). *Indexing in MongoDB*. GeeksforGeeks. https://www.geeksforgeeks.org/mongodb/indexing-in-mongodb/#

MongoDB. (n.d.). *Why use MongoDB and when to use it?* <https://www.mongodb.com/resources/products/fundamentals/why-use-mongodb>

Team, M. D. (n.d.). *Security*. Database Manual - MongoDB Docs. https://www.mongodb.com/docs/manual/security/

Smallcombe, M. (2025, July 11). *SQL vs NoSQL:  5 Critical Differences*. Integrate.io. https://www.integrate.io/blog/the-sql-vs-nosql-difference/