

Database Systems

Module code: ACCA7002

University of Wales Trinity Saint David

Sriskantharaja Mithushan (2232358)

Table of Contents

Abstract.....	4
1. Introduction.....	4
1.1 Background of ArtVista.....	4
1.2 Purpose of the Database Solution	4
1.3 Objectives of the Report	5
2. Overview of ArtVista's Goals	5
2.1 Expansion Plans:	5
2.2 Collaboration with WalesArtHub	5
2.3 Need for Diverse Digital Artworks.....	6
3. Comparison of NoSQL Database Types.....	6
3.1 Document Database	6
3.1.1 Use Case.....	6
3.1.2 Features	7
3.1.3 Advantages and Disadvantages:	7
3.2 Graph Database	7
3.2.1 Use Case:	7
3.2.2 Features:	8
3.2.3 Advantages and Disadvantages:	8
3.3 Key-Value Database	8
3.3.1 Use Case.....	8
3.3.2 Features	8
3.3.3 Advantages and Disadvantages:	8
3.4 Wide Column Database	10
3.4.1 Use Case.....	10
3.4.2 Features:	10
3.4.3 Advantages and Disadvantages:	10
3.5 Comparison Against Relational Database	10
4. When to Choose NoSQL vs Relational Database.....	11
4.1 Scenarios Favouring NoSQL	11
4.1.1 Dynamic Schema and Flexible Data Models:.....	11
4.1.2 Scalability and High Throughput:.....	11
4.1.3 Handling Unstructured or Semi-Structured Data:.....	12
4.2.1 Well-Defined and Structured Data:	12
4.2.2 Complex Queries and Joins:	12
4.2.3 Transactions and Data Integrity:	13
4.3 Hybrid Approach Considerations	13

4.3.1 Flexible Data Storage:.....	13
4.3.2 Scalability and Performance:	13
4.3.3 Migration Flexibility:.....	13
5. Recommendation and Justification	14
5.1 Chosen Database Solution	14
Relational Database (SQL):	14
Document Database (NoSQL):	14
5.2 Factors Influencing the Decision	14
5.2.1 Data Organization and Integrity:	14
5.2.2 Consideration of Data Volume and Query Complexity:.....	15
5.2.3 Scalability and Development Ease:	15
5.3 Data Organization and Integrity.....	15
5.3.1 Relational Database (SQL):	15
5.3.2 Document Database (NoSQL):	15
5.4 Consideration of Data Volume and Query Complexity.....	16
5.4.1 Relational Database (SQL):	16
5.4.2 Document Database (NoSQL):	16
5.5 Scalability and Development Ease.....	16
5.5.1 Relational Database (SQL):	16
5.5.2 Document Database (NoSQL):	16
5.6 Deployment Strategies	16
5.6.1 Relational Database (SQL):	16
5.6.2 Document Database (NoSQL):	16
6. Data Security and Backup.....	16
Reference:	18

ArtVista Online Art Marketplace:

Database Solution Report

Abstract:

This report explores the database solution options for ArtVista, an expanding digital art marketplace based in Swansea. Considering the company's goals and requirements, we analyze the advantages of NoSQL databases (Document, Graph, Key-Value, and Wide Column) compared to traditional relational databases. The report covers scenarios favoring NoSQL, the specific needs of ArtVista, and provides a detailed recommendation, considering factors like data organization, integrity, volume, query complexity, scalability, and development ease. Deployment strategies, data security, and backup importance are discussed, leading to a conclusive recommendation aligning with ArtVista's project goals [1].

1. Introduction

1.1 Background of ArtVista

ArtVista emerged as a response to the growing digital art movement, aiming to create a vibrant online marketplace for artists and enthusiasts. The platform was founded in Swansea with a mission to bridge the gap between digital artists and a global audience. Since its inception, ArtVista has successfully cultivated a community that celebrates digital art, and it now stands at the forefront of the digital art scene.

1.2 Purpose of the Database Solution

The exponential growth experienced by ArtVista necessitates a sophisticated database solution to manage the increasing volume of data seamlessly. The purpose of implementing a robust database is to enhance operational efficiency, optimize user experience, and facilitate the platform's expansion plans. By

centralizing and organizing data effectively, ArtVista aims to lay the foundation for scalability and sustained growth.

1.3 Objectives of the Report

This report is designed to serve as a comprehensive guide for ArtVista in choosing the most suitable database solution. The key objectives include providing an in-depth analysis of NoSQL databases, specifically Document, Graph, Key-Value, and Wide Column, in comparison to traditional relational databases. Additionally, the report aims to elucidate scenarios where the characteristics of NoSQL databases are advantageous, leading to a thorough understanding of when to choose them over traditional relational databases. The final objective is to furnish ArtVista with a tangible demonstration of both database approaches to aid in making an informed decision.

2. Overview of ArtVista's Goals

2.1 Expansion Plans:

ArtVista's ambitious expansion plans involve reaching a wider audience not only within the UK but also on a global scale. To accommodate this growth, the selected database solution must be scalable, ensuring that the platform can seamlessly handle increased user activity, diverse digital art collections, and an expanding network of artists.

2.2 Collaboration with WalesArtHub

The collaboration with WalesArtHub underscores ArtVista's commitment to partnering with reputable art entities. For this collaboration to thrive, the database solution must support smooth integration with partner studios and artists. This integration is crucial for maintaining data consistency, facilitating collaborative efforts, and ensuring a unified and coherent experience for users across the platform.

2.3 Need for Diverse Digital Artworks

Recognizing the importance of catering to varied artistic tastes, ArtVista is focused on providing a diverse range of digital artworks. The selected database solution should offer robust categorization and tagging capabilities to enable efficient organization and retrieval of artworks. This not only enhances user exploration but also aligns with the platform's goal of becoming a comprehensive repository of digital art.

2.4 Platform Features (Exploration, Purchases, Support)

ArtVista's platform distinguishes itself by offering users a rich and immersive experience. The chosen database solution is expected to seamlessly support features such as artwork exploration, secure and efficient purchase transactions, and responsive customer support. This requires a database that can effectively manage user accounts, transaction records, and the extensive catalog of digital artworks available on the platform.

3. Comparison of NoSQL Database Types

In this comprehensive comparison, we delve into the distinctive features, use cases, advantages, and disadvantages of various NoSQL database types, shedding light on their applicability for ArtVista's specific requirements. Additionally, we touch upon a brief comparison against traditional relational databases.

3.1 Document Database

3.1.1 Use Case

Document databases, such as MongoDB, are well-suited for scenarios where data structures are dynamic and may evolve over time. In ArtVista, where digital

artworks can have varying attributes, a document database could efficiently handle this diversity [2].

3.1.2 Features

These databases store data in flexible, JSON-like documents, allowing for nested structures and dynamic schemas. For ArtVista, this means the ability to represent each digital artwork with its unique set of attributes without the constraints of a fixed table structure.

3.1.3 Advantages and Disadvantages:

Advantages:

- The flexibility of schema accommodates changes in data requirements without a predefined structure.
- Efficient for storing and retrieving complex, nested data structures, providing a natural representation for digital artworks.
- Ideal for scenarios where data does not neatly fit into traditional tabular structures[2].

Disadvantages:

- The lack of strict schema enforcement may lead to inconsistent data if not managed carefully.
- May not perform as well as relational databases for complex queries involving multiple joins.

3.2 Graph Database

3.2.1 Use Case:

Graph databases, exemplified by Neo4j, excel in scenarios where relationships between data entities are complex and interconnected. In ArtVista, a graph database could aptly represent relationships between artists, artworks, and user interactions[2].

3.2.2 Features:

Data is stored as nodes and edges, facilitating efficient representation of relationships. This is advantageous for ArtVista's scenario, where understanding the connections and dependencies between different entities is crucial.

3.2.3 Advantages and Disadvantages:

Advantages:

- Efficient for traversing and querying relationships in interconnected data, making it suitable for representing artistic collaborations and user interactions.
- Well-suited for scenarios where the focus is on understanding and analyzing relationships.

Disadvantages:

- May not be the best choice for scenarios where relationships are less prominent or straightforward.
- Managing and maintaining relationships can become complex as the graph grows in size.

3.3 Key-Value Database

3.3.1 Use Case

Key-Value databases, like Redis or DynamoDB, are ideal for scenarios where rapid data retrieval based on unique keys is essential. In ArtVista, a Key-Value database could efficiently manage user profiles or provide quick access to specific artwork details[3].

3.3.2 Features

These databases store data as pairs of keys and values, offering extremely fast access to data. This simplicity is beneficial for scenarios where quick and direct access to specific pieces of information is crucial for user interactions.

3.3.3 Advantages and Disadvantages:

Advantages:

- Extremely fast read and write operations, making them suitable for scenarios with high read and write throughput.
- Scalable for scenarios where simple data retrieval tasks are frequent, such as caching or indexing.

Disadvantages:

- Limited in handling complex queries that involve multiple criteria.
- May not be optimal for scenarios where data structures evolve and change frequently.

3.4 Wide Column Database

3.4.1 Use Case

Wide Column databases, exemplified by Apache Cassandra, are designed for scenarios where data needs to be distributed across multiple nodes efficiently. In ArtVista, a Wide Column database could be employed for managing transactional data and ensuring high availability [4].

3.4.2 Features:

Data is organized in columns rather than rows, optimizing queries that involve retrieving specific columns of data. This structure is particularly advantageous for scenarios with a high volume of transactions, such as tracking purchases on the platform.

3.4.3 Advantages and Disadvantages:

Advantages:

- Efficient for read and write-intensive workloads, making them suitable for high-transaction scenarios.
- Scales horizontally by distributing data across nodes, ensuring high availability and fault tolerance.
- Well-suited for scenarios with large volumes of data and frequent updates.

Disadvantages:

- May not be optimal for scenarios with complex, relational queries involving multiple joins.
- Requires careful schema design to ensure optimal performance.

3.5 Comparison Against Relational Database

Relational Database Advantages:

- Strict schema enforces data integrity, ensuring that data conforms to predefined rules.
- Well-suited for scenarios with complex relationships and structured data, providing a standardized and organized approach.

- Support for ACID properties ensures transactional consistency, making them reliable for critical applications.

Relational Database Disadvantages:

- Less flexible in accommodating changes in data structure, which can be a limitation in dynamic and evolving environments.
- May face scalability challenges with large volumes of unstructured data, making them less suitable for scenarios with rapid data growth.
- Complex queries involving multiple joins can impact performance, especially in scenarios with intricate relationships.

4. When to Choose NoSQL vs Relational Database

In this section, we will explore scenarios that favor the selection of NoSQL databases and those where relational databases may be more suitable. Additionally, we will consider the potential benefits of adopting a hybrid approach that combines elements of both database types [5].

4.1 Scenarios Favouring NoSQL

4.1.1 Dynamic Schema and Flexible Data Models:

NoSQL databases are well-suited for scenarios where the data structure is dynamic and can evolve over time. In ArtVista, where the attributes of digital artworks may vary, a Document Database, for instance, provides the flexibility to store diverse and evolving data without a predefined schema. This is advantageous when dealing with artists introducing new attributes to their artworks or when supporting different types of digital art formats.

4.1.2 Scalability and High Throughput:

When the application demands high scalability and efficient handling of a large volume of concurrent read and write operations, NoSQL databases, such as Key-Value or Wide Column stores, shine. ArtVista, with its plans for global expansion

and increasing user engagement, could benefit from the horizontal scalability offered by NoSQL databases, ensuring robust performance even during peak usage periods.

4.1.3 Handling Unstructured or Semi-Structured Data:

NoSQL databases are particularly effective in managing unstructured or semi-structured data. In ArtVista, where digital artworks may have varying attributes, relationships, and metadata, a Graph Database could be valuable for representing complex relationships between artists, artworks, and users. This enables efficient querying and traversal of interconnected data, enhancing the user experience in exploring and discovering related artworks [5].

4.2 Scenarios Favoring Relational Database

4.2.1 Well-Defined and Structured Data:

Relational databases are ideal for scenarios where the data structure is well-defined, stable, and adheres to a strict schema. In ArtVista, aspects like user profiles, transaction records, and well-defined relationships between entities can be efficiently managed using a relational database. The structure and relationships are clearly defined, ensuring data integrity and consistency.

4.2.2 Complex Queries and Joins:

In scenarios where complex queries involving multiple joins are prevalent, relational databases excel. ArtVista's requirement for detailed analytics, such as understanding user preferences based on artwork attributes, or analyzing transaction patterns, may benefit from the powerful querying capabilities of relational databases. Complex reporting and analytics tasks can be more efficiently handled through SQL queries that involve joins and aggregations.

4.2.3 Transactions and Data Integrity:

Relational databases are well-known for supporting ACID (Atomicity, Consistency, Isolation, Durability) properties, ensuring transactional integrity. In scenarios where the consistency of transactions is paramount, such as recording and processing purchase transactions on ArtVista, a relational database provides the necessary guarantees, preventing data corruption or inconsistencies [6].

4.3 Hybrid Approach Considerations

4.3.1 Flexible Data Storage:

A hybrid approach combines the strengths of both NoSQL and relational databases. ArtVista could adopt a hybrid model where structured data with well-defined relationships, such as user profiles and transaction records, are stored in a relational database. Simultaneously, dynamic or unstructured data, like diverse digital artwork attributes, could be stored in a NoSQL database. This approach ensures flexibility in managing different types of data with optimal performance.

4.3.2 Scalability and Performance:

A hybrid approach allows ArtVista to leverage the scalability of NoSQL databases for parts of the system that require high throughput, while still benefiting from the structured nature and ACID properties of a relational database where transactional consistency is crucial. This approach strikes a balance between performance and data integrity, catering to the diverse requirements of an online art marketplace.

4.3.3 Migration Flexibility:

Adopting a hybrid approach provides flexibility in migration. ArtVista can gradually transition specific components or functionalities between database types based on evolving needs. This incremental approach minimizes disruption

to the existing system while allowing for continuous optimization as the platform grows.

5. Recommendation and Justification

In this section, we present a comprehensive recommendation for ArtVista's database solution, focusing on a hybrid approach that integrates both a relational database (SQL) and a NoSQL document database. The justification considers critical factors such as data organization, integrity, volume, query complexity, scalability, development ease, and deployment strategies.

5.1 Chosen Database Solution

Relational Database (SQL):

We recommend employing a relational database to manage structured and well-defined data, emphasizing data integrity and supporting complex queries. This includes user profiles, transactions, and relationships [7].

Document Database (NoSQL):

For the dynamic and diverse nature of digital artworks, a document database is recommended. This NoSQL solution provides flexibility in handling evolving data structures and diverse attributes associated with each artwork.

5.2 Factors Influencing the Decision

5.2.1 Data Organization and Integrity:

The hybrid model ensures that the relational database maintains the organization and integrity of structured data. Meanwhile, the document database accommodates the varied attributes of digital artworks without compromising data integrity.

5.2.2 Consideration of Data Volume and Query Complexity:

The hybrid approach allows for a balanced handling of data volume and query complexity. Relational databases excel in managing large volumes of structured data and handling complex queries, ensuring optimal performance for user-related tasks. Simultaneously, the NoSQL document database efficiently manages unstructured or semi-structured data, supporting the diverse attributes associated with digital artworks.

5.2.3 Scalability and Development Ease:

The NoSQL document database contributes to scalability, particularly in managing the dynamic digital artwork catalog. Meanwhile, the relational database ensures scalability for user profiles and transactions. The hybrid model provides development ease, allowing for flexibility in evolving data structures without compromising on the benefits of a structured relational approach.

5.3 Data Organization and Integrity

5.3.1 Relational Database (SQL):

- *User Profiles:* Storing user details such as name, address, and contact information.
- *Transactions:* Recording purchase details, ensuring ACID properties for transactional integrity.
- *Relationships:* Managing well-defined relationships between users, artworks, and transactions.

5.3.2 Document Database (NoSQL):

- *Digital Artworks:* Storing diverse attributes associated with each artwork, allowing for flexibility and evolution.

5.4 Consideration of Data Volume and Query Complexity

5.4.1 Relational Database (SQL):

- Efficient for managing structured data with well-defined relationships.
- Complex queries involving multiple joins can be handled effectively.

5.4.2 Document Database (NoSQL):

- Well-suited for scenarios with high data diversity and evolving structures.
- Efficient for queries specific to digital artworks and their dynamic attributes.

5.5 Scalability and Development Ease

5.5.1 Relational Database (SQL):

- Ensures transactional consistency.
- Supports structured data scalability.

5.5.2 Document Database (NoSQL):

- Provides horizontal scalability for dynamic and diverse digital artwork data.
- Allows for agile development with evolving data structures.

5.6 Deployment Strategies

5.6.1 Relational Database (SQL):

- Deploy using a cloud-based relational database service for scalability.
- Utilize indexing and query optimization for performance.

5.6.2 Document Database (NoSQL):

- Leverage cloud-based NoSQL database services for scalability.
- Optimize schema design for efficient querying.

6. Data Security and Backup

In the intricate landscape of an online art marketplace like ArtVista, the significance of robust data security and backup strategies cannot be overstated. Not only is customer trust paramount, but the protection of sensitive user

information and digital art assets is fundamental to the platform's integrity. A breach in security could not only compromise individual privacy but also lead to severe reputational damage for ArtVista [8].

To fortify its security posture, ArtVista should implement a multi-faceted approach. Encryption, both for data in transit and at rest, provides a baseline for safeguarding sensitive information. Access control mechanisms, coupled with robust authentication procedures, ensure that only authorized individuals have access to critical data. Regular security audits and penetration testing serve as proactive measures, identifying vulnerabilities before they can be exploited by malicious actors.

In tandem with security measures, a resilient backup strategy is imperative. Automated backups, performed regularly, serve as a safety net against data loss due to system failures or accidental deletions. Storing backups in geographically redundant locations adds an extra layer of protection, mitigating risks associated with localized disasters [9]. Versioned backups and regular testing of the restoration process enhance the reliability of the backup system, ensuring that ArtVista can swiftly recover from any unforeseen incidents.

In conclusion, ArtVista's strategic adoption of a hybrid database solution, integrating both a relational database and a NoSQL document database, provides a nuanced and tailored approach to address the diverse and evolving needs of its online art marketplace. This hybrid model ensures optimal data organization, integrity, and scalability, striking a balance between the structured nature of user-related data and the dynamic attributes of digital artworks. Furthermore, the emphasis on robust data security and backup strategies underscores ArtVista's commitment to safeguarding user information and valuable digital assets, contributing to the platform's credibility and long-term success in the ever-expanding digital art landscape.

Reference:

[1]

K. Sahatqija, J. Ajdari, X. Zenuni, B. Raufi, and F. Ismaili, "Comparison between relational and NOSQL databases," *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018, doi: <https://doi.org/10.23919/mipro.2018.8400041>.

[2]

A. Gupta, S. Tyagi, N. Panwar, S. Sachdeva, and U. Saxena, "NoSQL databases: Critical analysis and comparison," *IEEE Xplore*, Oct. 01, 2017. <https://ieeexplore.ieee.org/document/8284494>

[3]

A. B. M. Moniruzzaman and S. A. Hossain, "NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison," *arXiv:1307.0191 [cs]*, Jun. 2013, Available: <https://arxiv.org/abs/1307.0191>

[4]

S. Venkatraman, K. F. Kaspi Samuel, and R. Venkatraman, "SQL Versus NoSQL Movement with Big Data Analytics," *International Journal of Information Technology and Computer Science*, vol. 8, no. 12, pp. 59–66, Dec. 2016, doi: <https://doi.org/10.5815/ijitcs.2016.12.07>.

[5]

Jing Han, Haihong E, Guan Le, and Jian Du, "Survey on NoSQL database," *2011 6th International Conference on Pervasive Computing and Applications*, Oct. 2011, doi: <https://doi.org/10.1109/icpca.2011.6106531>.

[6]

P. P. Srivastava, S. Goyal, and A. Kumar, "Analysis of various NoSql database," *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, Oct. 2015, doi: <https://doi.org/10.1109/icgciot.2015.7380523>.

[7]

M. A. Kausar and M. Nasar, "SQL Versus NoSQL Databases to assess their appropriateness for Big Data Application.," *Recent Patents on Computer Science*, vol. 12, Oct. 2019, doi: <https://doi.org/10.2174/2213275912666191028111632>.

[8]

M. S. Kumar and P. J, “Comparison of NoSQL Database and Traditional Database-An emphatic analysis,” *JOIV : International Journal on Informatics Visualization*, vol. 2, no. 2, p. 51, Mar. 2018, doi: <https://doi.org/10.30630/joiv.2.2.58>.

[9]

M.-L. E. Chang and H. N. Chua, “SQL and NoSQL Database Comparison,” *Advances in Intelligent Systems and Computing*, pp. 294–310, Dec. 2018, doi: https://doi.org/10.1007/978-3-030-03402-3_20.