**Comparative Analysis of MySQL and MongoDB in Developing a High-Concurrency Ticketing System: A Practical Implementation Study**

Blue text on a white background

Description automatically generated

**Gabriel Alberto Avina Solares**

Being a thesis presented for the award of

*Masters Software Solutions Architecture*

Supervisor: Cormac Keogh

Technological University Dublin, Tallaght Campus

School of Enterprise Computing and Digital Transformation

December 2024

**DECLARATION**

I hereby certify that the material, which I now submit for assessment on the programmes of study leading to the award of Master of Science, is entirely my own work and has not been taken from the work of others except to the extent that such work has been cited and acknowledged within the text of my own work. No portion of the work contained in this thesis has been submitted in support of an application for another degree or qualification to this or any other institution.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of Candidate Date

**ABSTRACT**

**ACKNOWLEDGEMENTS**

**To Silvia, Karla, and Andrea**

**To my friends and supervisor**

**A mis abuelos sláinte mhaith…**

Table of Contents

[**LIST OF TABLES AND FIGURES** 1](#_Toc183615406)

[**ABBREVIATIONS** 2](#_Toc183615407)

[**INTRODUCTION** 3](#_Toc183615408)

[Background and context 3](#_Toc183615409)

[Problem statement 3](#_Toc183615410)

[Research objectives 4](#_Toc183615411)

[Research questions 4](#_Toc183615412)

[Scope and limitations 5](#_Toc183615413)

[Significance of the Study 5](#_Toc183615414)

[1 Review of Relevant Literature 6](#_Toc183615415)

[1.1 Database management systems overview 6](#_Toc183615416)

[1.2 Relational Databases 6](#_Toc183615417)

[1.2.1 Overview and key features 6](#_Toc183615418)

[1.2.2 MySQL 6](#_Toc183615419)

[1.3 NoSQL Databases 6](#_Toc183615420)

[1.3.1 Overview and key features 6](#_Toc183615421)

[1.3.2 Document-oriented databases 9](#_Toc183615422)

[1.4 Transactional Processing and Concurrency 12](#_Toc183615423)

[1.5 Comparative Studies 12](#_Toc183615424)

[1.5.1 Data models and Schema flexibility 12](#_Toc183615425)

[1.5.2 Scalability and performance 13](#_Toc183615426)

[1.5.3 Handling unstructured data 13](#_Toc183615427)

[1.5.4 Consistency and transactions 13](#_Toc183615428)

[1.5.5 Integration and migration challenges 14](#_Toc183615429)

[1.5.6 Benchmarking Cloud Serving Systems with YCSB (Cooper *et al.*, 2010) 14](#_Toc183615430)

[1.5.7 Databases vs NoSQL Databases (Stonebraker, 2010) 16](#_Toc183615431)

[2 Methodology 16](#_Toc183615432)

[2.1 Research design 16](#_Toc183615433)

[2.2 Experimental Framework 17](#_Toc183615434)

[2.2.1 Technical Architecture 17](#_Toc183615435)

[2.2.2 Transaction processing implementation 18](#_Toc183615436)

[2.3 Analysis Methods 18](#_Toc183615437)

[2.3.1 Quantitative 18](#_Toc183615438)

[2.3.2 Qualitative 19](#_Toc183615439)

[2.4 Testing Methodology 19](#_Toc183615440)

[2.4.1 Performance Test Scenarios 19](#_Toc183615441)

[2.4.2 Concurrency Tests cases 20](#_Toc183615442)

[2.4.3 Schema Modification tests 20](#_Toc183615443)

[2.5 Validation Strategy 20](#_Toc183615444)

[2.5.1 Experimental validation 20](#_Toc183615445)

[2.5.2 Data and Process validation 21](#_Toc183615446)

[3 Implementation 23](#_Toc183615447)

[3.1 System Architecture 24](#_Toc183615448)

[3.1.1 Data Access Layer 24](#_Toc183615449)

[3.1.2 Service Layer 24](#_Toc183615450)

[3.1.3 Simulation Framework 26](#_Toc183615451)

[3.1.4 Data initialization 27](#_Toc183615452)

[3.1.5 Domain Entities 28](#_Toc183615453)

[3.2 MySQL implementation 28](#_Toc183615454)

[3.2.1 System Architecture & Technical stack 31](#_Toc183615455)

[3.2.2 Data Model Design 31](#_Toc183615456)

[3.2.3 Transaction management 33](#_Toc183615457)

[3.3 MongoDB implementation 41](#_Toc183615458)

[3.3.1 System Architecture & Technical Stack 41](#_Toc183615459)

[3.3.2 Document Model Design 42](#_Toc183615460)

[3.3.3 Transaction management 49](#_Toc183615461)

[4 Results and discussion 53](#_Toc183615462)

[4.1 Performance Analysis 53](#_Toc183615463)

[4.1.1 Concurrency Test Results 53](#_Toc183615464)

[4.1.2 Transaction Processing Metrics 53](#_Toc183615465)

[4.2 Schema Analysis 53](#_Toc183615466)

[4.3 Data Structure Management 54](#_Toc183615467)

[4.4 Comparative analysis 55](#_Toc183615468)

[4.4.1 Transaction Management 55](#_Toc183615469)

[4.4.2 Schema Flexibility 55](#_Toc183615470)

[4.4.3 Development Experience 55](#_Toc183615471)

[4.5 Research Questions Assessment 55](#_Toc183615472)

[4.6 Implications 55](#_Toc183615473)

[4.7 Limitations of the research 55](#_Toc183615474)

[5 Conclusion 55](#_Toc183615475)

[5.1 Summary of findings 55](#_Toc183615476)

[5.2 Answers to Research questions 55](#_Toc183615477)

[5.3 Contributions 55](#_Toc183615478)

[5.4 Recommendations and future research 55](#_Toc183615479)

[6 References 55](#_Toc183615480)

[7 Appendices 60](#_Toc183615481)

[7.1 MySQL Schema Implementation 60](#_Toc183615482)

[7.2 Entity-Relationship Diagram 61](#_Toc183615483)

[7.3 MySQL Class Diagram 62](#_Toc183615484)

[7.4 Document Structure Diagram for MongoDB 63](#_Toc183615485)

[7.5 Ticket Purchase process diagram 64](#_Toc183615486)

[7.6 Sequence Diagram for Ticket Purchase Transaction 65](#_Toc183615487)

[7.7 Backend application diagram 66](#_Toc183615488)

[7.8 Deployment with technological stack 67](#_Toc183615489)

[7.9 Component Diagram 68](#_Toc183615490)

[7.10 MySQL MoSCOW 69](#_Toc183615491)

[7.11 MongoDB MoSCOW 70](#_Toc183615492)

[7.12 Code samples 71](#_Toc183615493)

[7.13 Test results 71](#_Toc183615494)

[7.14 Configuration files 71](#_Toc183615495)

[7.15 Architecture Diagram for Ticketing System using MongoDB 1](#_Toc183615496)

[7.16 Transaction Flow Protocol MongoDB 1](#_Toc183615497)

[7.17 Transaction Flow Protocol MongoDB 2](#_Toc183615498)

# **LIST OF TABLES AND FIGURES**

# **ABBREVIATIONS**

|  |  |
| --- | --- |
| Business Intelligence | BI |
|  | BERT |
|  | CRISP |
| Online analytical processing | OLAP |
| Extracting, transforming, and loading | ETL |
| Artificial Intelligence | AI |
| Internet of Things | IoT |
| User-generated context (), | UGC |
| Machine Learning | ML |
| Deep Learning | DL |
| Natural Language Processing | NLP |
| Linked Open Data | LOD |
| Structural Topic Modelling | STM |
| JavaScript object notation | JSON |
| binary-encoded serialization of JSON | BSON |
| Extensible Marup Language | XML |
| model-driven architecture | MDA |
| computation-independent model | CIM |
| Platform independent model | PIM |
| platform-specific model | PSM |
| platform independent data metamodel | PIDM |
|  | ORM |
|  | ODM |

# **INTRODUCTION**

## Background and context

*Background of the study refers to the context, circumstances, and history that led to the research problem or topic being studied. It provides the reader with a comprehensive understanding of the subject matter and the significance of the study.*

## Problem statement

The increasing complexity of data management in modern applications presents significant challenges in selecting the appropriate database system. Traditional relational databases like MySQL offer strong transactional integrity and enforce rigid schema designs, making them suitable for applications that require strict data consistency and complex transactions. (Patil *et al.*, 2017),

Conversely, NoSQL databases such as MongoDB have emerged to address the limitations of relational databases by offering flexible schemas, horizontal scalability, and efficient handling of large volumes of unstructured or semi-structured data (Sudiartha *et al.*, 2020). This makes them strong candidates for applications like property listing management systems, which deal with nested data structures, user-generated content and rapid evolving data requirements.

Despite the availability of both SQL and NoSQL solutions, selecting the most appropriate database system for specific application requirements remains a complex task (Capris *et al.*, 2022; Shareef, Sharif and Rashid, 2022; Yedilkhan *et al.*, 2023). Existing comparative studies often provide broad overviews without delving into the practical implications of database selection based on workload characteristics and application contexts (Győrödi *et al.*, 2015). While some research highlights the superior performance of NoSQL databases in handling data loads and dynamic datasets (Chang and Chua, 2019; Sudiartha *et al.*, 2020; Wodyk and Skublewska-Paszkowska, 2020), relational databases continue to be preferred in scenarios requiring strong data integrity and complex transactions.

Developers and organizations need detailed insights to make informed decisions when selecting a database system that balances transactional integrity, development agility, schema design flexibility, and the management of complex data relationships This research aims to address this need by implementing a ticketing system using both MySQL and MongoDB to critically analyse their transactional mechanisms, schema design patterns, and strategies for modelling complex data structures.

By simulating concurrent purchase attempts and managing intricate data relationships within the ticketing system, the study seeks to provide practical insights into the development experiences and challenges associated with each database.

## Research objectives

**Primary objective**

To critically analyse and compare the transactional behaviour, schema design, and management of nested data structures in MySQL and MongoDB through the implementation of a ticketing system.

**Secondary objectives**

1. To examine how MySQL handles transactional integrity and schema rigidity in the ticketing system, especially during simultaneous ticket purchase attempts, and its impact on managing nested data structures.
2. To explore MongoDB’s approach to transactional behaviour and schema flexibility in the same ticketing scenario, analysing how it manages concurrent transactions and the implications of its flexible schema design on modelling intricate data structures.
3. To compare the development experiences and challenges encountered when implementing the ticketing system in both MySQL and MongoDB, focusing on transaction management, agility versus rigidity of schema design, and the strategies for representing of complex structures in each database.

## Research questions

1. How does MySQL ensure transactional integrity in a high-concurrency ticketing system, and what challenges arise from its rigid schema when dealing with complex data relationships?
2. How does MongoDB handle transactions in a concurrent purchase scenario, and how does its flexible schema influence the modelling of intricate data structures?
3. What are the key differences in implementing transactional operations and data modelling between MySQL and MongoDB in the context of the ticketing system?

## Scope and limitations

## Significance of the Study

# Literature Review

## Database management systems overview

## Relational Databases

It would be nice to have the architecture, data models, strengths, and limitations`. ACID properties and their importance in data integrity

### Overview and key features

### MySQL

MySQL, renowned as the world's leading open-source relational database (Solid, 2020), is favored by tech giants like Facebook and Twitter for its reliability, performance, and user-friendly nature. It is predominantly used by professionals for web applications, and a variety of companies and software publishers offer MySQL. Depending on its function, MySQL can be either open-source or a proprietary product. It is compatible with multiple platforms, including AIX, BSDi, FreeBSD, HP-UX, Linux, Mac OS X, NetWare, NetBSD, OpenBSD, OS/2 Warp, SGI Irix, Solaris, SunOS, SCO OpenServer, SCO UnixWare, Tru64 Unix, and Windows. Additionally, databases in MySQL can be accessed using a range of programming languages such as C, C++, C#, Delphi/Kylix, Eiffel, Java, Perl, PHP, Python, Ruby, and Tcl. (Comparative Study Between the MySQL Relational Database and the MongoDB NoSQL Database)

## NoSQL Databases

### Overview and key features

The term NoSQL (Not Only SQL) was first used in 1998 by Carlo Strozzi as the name of his small RDBMS that did not use SQL for data manipulation. Starting in 2009, the term NoSQL is used for the growing number of distributed data management systems that abandoned the support of ACID transactions (Atomicity, Consistency, Isolation, Durability) which is a key principle of relational databases (Lith and Mattsson, 2010). . atomic operations across multiple documents and collections. Although historically NoSQL databases sacrificed strong consistency for availability and partition tolerance, following BASE (Basically Available, Soft state, Eventually consistent) principles

NoSQL databases have emerged as a pivotal solution for managing the vast and varied data generated in contemporary applications, particularly those involving unstructured and semi structured data. Unlike traditional RDBMS, NoSQL systems offer enhanced scalability, flexibility, and performance, making them well-suited for modern data intensive environments.

These databases are broadly categorized into four primary types on their data models: key-value stores, document-oriented databases, column-family stores, and graph databases. Each one is tailored to specific data management needs, offering unique features that address the limitations of relational databases (Kuznetsov and Poskonin, 2014).

* + 1. Key-Value Stores

Key-value stores are the simplest form of NoSQL databases, where data is stored as a collection of key-value pairs. Each key serves as a unique identifier for its corresponding value, which can be any arbitrary data type (Hecht and Jablonski, 2011; Amghar, Cherdal and Mouline, 2020). This model excels in scenarios requiring high-speed read and write operations, such as caching, session management, and real-time data processing.

The main characteristics of these databases are:

* + - * *Simplicity*: Key-value stores operates on a schema-less model where data is accessed by unique keys, making operations straightforward and efficient but is unsuitable for complex queries, as values are opaque to the system and require logic in the application layer to handle relationships between data (Kuznetsov and Poskonin, 2014).
      * *Performance:* Optimized for speed, particularly in in-memory store like Redis, which offers rapid data access (Amghar, Cherdal and Mouline, 2022).
      * *Scalability:* Easy scalable horizontally by distributing key-value pairs across multiple nodes.

Examples of Key-Value stores databases are Aerospike, Amazon DynamoDB, Memcached, Redis, Berkeley DB, etc.

* + 1. Document-Oriented Databases

According to Kuznetsov and Poskonin (2014) provide richer capabilities than key-value systems. The unit of storage is a document, and the format used is typically JSON, BSON or XML. This model provides indexing and aggregation capabilities, which are advantageous in handling unstructured data.

The main characteristics of these databases are:

* + - * *Schema flexibility:* Documents can have varying structures, accommodating diverse and evolving data models without requiring predefined schemas.
      * *Querying:* Support for complex queries, including indexing and aggregation, simplified data retrieval and analysis.
      * *Integration with analytical tools:* Compatibility with frameworks like Apache Spark enhances their utility in data-driven applications.

According to Amghar et al.(2022) document-oriented databases are particularly well-suited for applications requiring real-time analytics and flexible data models, such as content management systems and social media platforms.

Some examples of these databases are MongoDB, Couchbase, CouchDB, Realm, Google Cloud Firestore, etc.

* + 1. Column-Family Stores

Inspired by Google’s Bigtable, organize data into columns and rows, grouping related columns into families. This design optimizes storage efficiency and allows for high scalability across distributed systems (Kuznetsov and Poskonin, 2014; Amghar, Cherdal and Mouline, 2020).

The main characteristics of these databases are:

* + - * *Scalability:* Designed for horizontal scalability, handling massive datasets across multiple nodes with ease.
      * *Efficient Storage:* The sparse, multidimensional data model reduces storage inefficiencies by eliminating the need to store null values (ul Haque, Mahmood and Ikram, 2019).
      * *High Throughput:* Suitable for applications requiring high read and write throughput, such as event logging and large-scale analytics.

Examples of Column-family stores are Apache Cassandra, Apache HBase, ScyllaDB, etc.

Despite their strengths, column-family stores demand careful considerations of partitioning strategies to balance performance and flexibility, as highlighted by ul Haque et al. (2019).

* + 1. Graph Databases

Graph databases are designed to manage and query data with complex relationships, using nodes and edges to represent entities and their interconnections. This model is particularly effective for applications where understanding and traversing relationships is crucial. Thakare et al. (2023), Amghar et al., (2022), and Hecht and Jablonski (2011) provided a series of characteristics:

* + - * *Relationship management:* Efficiently handles multidimensional relationships, making them ideal for social networks, recommendation engines, and fraud detection.
      * *Flexible schema:* Supports dynamic and interconnected data structures without requiring a predefined schema.
      * *Advanced querying:* Enables sophisticated queries involving traversals and pattern matching across connected data.

Examples are Neo4j, JanusGraph, MemGraph, etc.

Graph databases, while powerful in managing interconnected data, face challenges in scalability, particularly in distributed environments where graph partitioning is inherently complex.

The exploration of various NoSQL database types reveals their strengths and applicability across different use cases. For the purposes of this research document-oriented databases were chosen, due to their versatility in balancing flexibility, performance and scalability. Their ability to handle complex, semi-structured data and integrate well with analytical tools, makes them ideal for real-time data processing and analytics in the MICE industry where managing diverse data sources is critical for enhancing customer experience and operational efficiency (ul Haque, Mahmood and Ikram, 2019; Amghar, Cherdal and Mouline, 2022).

### Document-oriented databases

Document-oriented databases, a subset of NoSQL databases, are designed to store, retrieve, and manage document-oriented information, typically in formats like JSON, BSON, or XML Unlike traditional relational databases that use tables with rows and columns, document-oriented store data as documents within collections, offering a flexible schema that allows for varying data structures within the same collection (Mok, 2021).

The core of these databases lies in their model and structure. Data is organized into documents, analogous to records, and collections, akin to tables. These documents support nested data structures, including arrays and sub-documents, enabling the representation of complex relationships, within a single entity (Sen and Mukherjee, 2024). The variety of supported data types- strings, numbers, dates, arrays, and sub-documents- enhances their capability to handle diverse data forms.

Schema design in document-oriented databases involves careful consideration of denormalization practices to optimize read performance. The choice between embedding data within documents and referencing other documents depends on factors like data volatility and access patterns. Embedding is often recommended when records are frequently queried together, improving read efficiency, while referencing is preferred for write-heavy workloads or when data is volatile (Imam *et al.*, 2018).

Querying and indexing mechanisms in these databases are powerful and flexible. They support ad-hoc queries, range queries, and aggregations, with indexing on fields within documents improving query performance. Full-text search capabilities further enhance the management of unstructured textual data however, managing schema variations without a rigid schema can complicate data analysis and integration. Gallinucci, Golfarelli, and Rizzi (2018) addressed this challenge by introducing “schema profiling”, a technique that captures and explains schema variations using decision trees, aiding in data quality assessment and business intelligence.

Advantages in managing unstructured data

Scalability and performance are inherent advantages of document-oriented databases. They are designed for horizontal scaling through sharding, distributing data across multiple servers to handle large volumes (Stonebraker, 2010). Load balancing and replication mechanisms enhance performance and availability. However, performance can vary based on the specific database system and workload. For instance, Carvalho, Sá and Bernardino (2023) found that MongoDB outperformed Couchbase and CouchDB in most tests, particularly in read-heavy operations, but may falter in scan-heavy workloads.

Diaz-Ordoñez, Rodríguez Baena and Yun-Casalilla (2023) highlighted the flexibility of these tools due to their dynamic schema, since allows for accommodating evolving data models without downtime or complex migrations, which is essential in industries where data structures frequently change. This flexibility supports rapid development, enabling developers to iterate quickly and adjust data structures as requirements evolve.

Handling diverse data types, particularly unstructured and semi-structured data, is a significant strength. Document databases efficiently store and process data like emails, social media content, multimedia files, and other unstructured formats common in various industries, including MICE, they naturally model complex, hierarchical, and nested data, simplifying data retrieval and manipulation (Sen and Mukherjee, 2024).

Another capability is the performance optimization. Since are tailored for high-performance read and write operations, essential for real-time data access. They support data aggregation and real-time analytics. Mok (2021) demonstrated how schema trees generated from conceptual models can optimize query retrieval times in MongoDB, improving performance in managing large and evolving datasets.

Challenges and limitations

Despite their advantages, these databases present challenges and limitations. According to Imam *et al.* (2018) data integrity concerns arise due to the lack of rigid schemas, potentially leading to inconsistencies. While the flexible schema allows for adaptability, it requires careful management to prevent data anomalies.

Complex transactions are another area of concern investigated by Sen and Mukherjee (2024). Although support for transactions has improved, it may not be as robust as in relational databases for complex, multi-document transactions, also managing write-heavy operations can be challenging, as their schema design focuses on read optimization and may not deeply engage with write-heavy workloads.

Integration with Existing systems

Interoperability is crucial for integrating document databases with existing applications, middleware, and data processing tools. APIs and connectors facilitate this integration, allowing for seamless data flow across systems. Data migration strategies, including ETL processes, are essential for transitioning from RDBMS to document-oriented databases (Li, 2010; Reinero, 2017).

Hybrid approaches are a suitable solution, Seghier and Kazar (2021) presented a polyglot persistence, involving multiple types of databases within the same application to leverage the strength of each, and can be used in combination to optimize performance across different workloads. In a microservices architecture, services can utilize the most appropriate data storage solutions, enhancing scalability and maintainability.

Case studies and Industry examples

Document-oriented databases offer advantages in managing unstructured and semi-structured data, providing flexibility, scalability, and performance optimization. They align well with the needs of industries dealing with complex and evolving data structures, such as the investigated in this proposal.

While challenges exist, particularly in data integrity, complex transactions, learning curve, etc, these can be addressed through careful schema design, validation mechanisms, and ongoing training and tool development.

## Transactional Processing and Concurrency

## Comparative Studies

The evolution of data management systems has led to significant discussions regarding the suitability of traditional RDBMS versus NoSQL databases for handling modern data demands. This section examines the differences between relational databases and document-oriented NoSQL databases, specifically on their architectures, data models, scalability, performance, and applicability in managing unstructured data.

### Data models and Schema flexibility

RDBMS structure data into predefined tables with fixed schemas, enforcing strict data integrity and relationships through primary and foreign keys (Stonebraker, 2010; Capris *et al.*, 2022). This rigidity ensures consistency but poses challenges when dealing with unstructured or semi-structured data, as schema alterations can be complex and time consuming (Li, 2010; Digittrix, 2023).

On the other hand, document-oriented NoSQL databases like MongoDB employ flexible, schema-less data models, storing data in JSON-like documents (Thakare *et al.*, 2023). This flexibility allows for the seamless integration of diverse and evolving data types without the need for predefined schemas. Sudiartha et al. (2020) highlighted that this adaptability is crucial for applications dealing with heterogeneous data, such as mobile-based tourist tracking systems.

### Scalability and performance

Relational databases traditionally scale vertically by improving hardware, which can lead to uprising costs and limitations (Sudiartha *et al.*, 2020; Digittrix, 2023). While, NoSQL are designed for horizontal scalability, distributing data across multiple servers or nodes, thus efficiently handling large volumes of data and high-traffic environments (Amghar, Cherdal and Mouline, 2022; Thakare *et al.*, 2023).

Performance comparisons have consistently shown that NoSQL databases often outperform relational databases in pecific workloads. (Capris *et al.*, 2022) conducted an empirical study comparing MySQL and MongoDB using the Yahoo! Cloud Serving Benchmark (YCSB). The findings indicated that MongoDB significantly outperformed MySQL across various workloads, particularly in write-intensive scenarios. This performance is attributed to MongoDB’s architecture, which avoids the overhead associated with schema validation and foreign key constraints inherent in SQL databases.

### Handling unstructured data

The schemas of RDBMS make storing unstructured data challenging, often requiring complex transformations or additional layers such as BLOBs (Li, 2010). According to Reinero (2017) Document-oriented databases natively support unstructured data, allowing for direct storage retrieval without extensive preprocessing.

### Consistency and transactions

RDBMS adhere to ACID properties, ensuring strict data consistency and reliability, which is essential for applications where data integrity is paramount (Stonebraker, 2010; Digittrix, 2023). On the other hand, NoSQL often follow BASE (Basically Available, Soft state, Eventual consistency) principles, prioritizing availability and partition tolerance over immediate consistency (Thakare *et al.*, 2023).

Stonebraker (2010) critiqued NoSQL databases for compromising on transactional integrity to achieve performance gains. HE argued that the performance benefits are not due to the abandonment of SQL but result from reducing overheads like logging and locking in traditional RDBMS. Digittrix (2023) echoes this sentiment, since MongoDB’s lack of mature multi-document ACID transactions may not be suitable for all scenarios, especially where strong consistency is required

### Integration and migration challenges

Transitioning from RDBMS to NoSQL systems involves significant challenges, including data migration and schema transformation. Li (2010) highlighted the complexities of transforming relational schemas into NoSQL formats like HBase, emphasizing the need for careful planning and potential operational overhead.

Candel, Sevilla Ruiz and García-Molina (2022) addressed this issue by proposing a unified metamodel to integrate relational and NoSQL databases, improving schema management and promoting polyglot persistence. Reinero (2017) suggested that understanding the fundamental differences in data modelling is crucial for a successful transition, as MongoDB requires a different approach to schema design compared to RDBMS.

The MICE industry deals with a range of unstructured data types, including customer feedback, social media content, and multimedia files. Document-oriented NoSQL databases like MongoDB offer the flexibility and scalability needed to manage this data effectively (Amghar, Cherdal and Mouline, 2022; Digittrix, 2023).

### Benchmarking Cloud Serving Systems with YCSB (Cooper *et al.*, 2010)

Cooper et al. introduced the Yahoo! Cloud Serving Benchmark (YCSB), a framework designed to evaluate the performance of cloud serving systems, particularly NoSQL databases. Addressing the lack of standardized tools for comparing cloud databases such as Cassandra, HBase, and sharded MySQL.

The findings showed variability between the systems evaluated, depending on the workload characteristics. Cassandra excelled in write-heavy scenarios due to its sequential disk-write optimization, while MySQL and PNUTS performed better in read-heavy workloads. Additionally, all systems demonstrated scalability, but the efficiency in handling increased loads varied across them. Cassandra and PNUTS exhibited better elasticity under stress compared to HBase, which had more erratic performance as the system scaled.

*Comparative analysis*

NoSQL databases such as Cassandra demonstrate superior performance in write-heavy environments, due to their sequential disk-write architecture. SQL databases, on the other hand, tend to focus on maintaining consistency and strong ACID (Atomicity, Consistency, Isolation, Durability) properties, which can increase response time in distributed settings.

Throuhput, which measures how efficiently a database can handle read/write ioeratuinsm was higher in NoSQL systems like Cassandra and HBase in write-heavy workloads. MySQL, while performing well in read-heavy workloads, encountered more difficulty in handling large-scale, distributed environments due to the overhead associated with maintaining ACID compliance. In terms of scalability, the NoSQL systems demonstrated superior horizontal scaling, which is more suitable for cloud applications, whereas SQL databases like Microsoft SQL Server often scale vertically, requiring more powerful individual servers rather than distributing the load across multiple nodes.

Transactional integrity is another critical aspect of comparison. SQL databases, such as Microsoft SQL Server, provide strong transactional integrity through ACID properties. In contrast, NoSQL systems like MongoDB and Cassandra often trade consistency for availability and partition tolerance, following the principles of CAO theorem. This trade-off allows NoSQL databases to achieve higher availability but whit eventual consistency, which can lead to temporary data inconsistencies during operations.

*Strengths and Weaknesses*

MySQL primary strengths are its mature ecosystem, strong transactional integrity, and efficient read operations. Its long-standing presence in the industry means it has robust tools for query optimization, data integrity, and complex queries. Additionally, MySQL’s ability t handle structured data with a well-defined schema makes it ideal for applications that require strong consistency.

However, MySQL struggled with horizontal scalability and write -intensive workloads. The overhead associated with ensuring ACID compliance, particularly in distributed environments, hindered its ability to handle large-scale, high-volume data environments like cloud-based applications.

On the other hand, in the experiment NoSQL systems resulted to be highly scalable, making them better for cloud envionments that required high availability and partition tolerance. They exceled in write-heavy workloads, where MySQL performance was inferior. However, noSQL systems trade off consistency for availability, and their lack of support for complex queries and multi-row transactions could be limited.

Future research should explore how SQL databases like SQL Server perform in more complex transactional settings compared to NoSQL systems.

### Databases vs NoSQL Databases (Stonebraker, 2010)

Stonebraker’s article evaluated the prevailing arguments that advocate for the adoption of NoSQL databases over traditional SQL (RDBMS) systems, particularly focusing on performance and scalability. The primary objective was to determine whether the performance benefits claimed by NoSQL proponents, such as key-value and document-based databases, are substantial enough to justify transitioning from established relational databases.

Stonebraker concluded that the performance advantages of NoSQL databases are often overstated and the result from the elimination of certain overheads inherent in traditional RDBMS, such as logging, locking, and buffer management (Harizopoulos *et al.*, 2008). He argues that these perceived gains do not stem from any inherent architectural superiority of NoSQL systems. Instead, modern SQL databases have incorporated features like automatic sharding and enhanced OLTP (Online Transaction Processing) performance, which address scalability and performance without compromising the ACID (Atomicity, Consistency, Isolation, Durability) properties essential for transactional integrity. Consequently, Stonebraker anticipates the emergence of high-performance, open-source SQL engines that combine the scalability of NoSQL systems with the robust ACID compliance of traditional SQL databases.

# Methodology

## Research design

This study adopted a comparative experimental research approach (Reisner, 1988), to examine how MySQL and MongoDB handle transactional behaviour, schema rigidity, and nested structures within a ticketing system context The experimental design enabled direct comparison of database performance under controlled conditions through systematic implementation and testing of identical functionality in both systems.

The selection of a comparative experimental methodology was informed in previous studies (Győrödi *et al.*, 2015; Patil *et al.*, 2017; Capris *et al.*, 2022; Stonebraker and Pavlo, 2024), which demonstrated the effectiveness of controlled testing in revealing performance differences between SQL and NoSQL databases. This approach enabled objective measurement of each database’s capabilities in managing concurrent transactions, maintaining data consistency, and handling data relationships.

This research is based on pragmatism, emphasizing practical outcomes and real-world applicability. By implementing actual systems rather than relying solely on theoretical analysis or simulations, the study aimed to produce findings that are directly relevant to practitioners facing similar database selection challenges. This aligned with the recommendations of Shareef, Sharif and Rashid (2022), who identified the need for studies that provide actionable insights into database performance in specific application scenarios.

The experimental framework encompassed three key areas of investigation:

1. Transactional behaviour in concurrent ticket booking scenarios
2. Impact of schema design approaches on system implementation
3. Management of nested data structures in booking records

The comparative analysis was structured through:

1. Controlled test scenarios that examined basic booking operations, transaction handling methods, and schema management approaches
2. Systematic comparison of data handling strategies, and transaction management mechanisms
3. Quantitative measurement of performance metrics and qualitative assessment of development experiences

This methodological framework enabled an examination of how each database system addressed the core research objectives through controlled experimentation and structured comparison.

## Experimental Framework

### Technical Architecture

All the experiments were implemented on a development workstation Windows 11 Home. The system utilized an Intel(R) Core (TM) i5-12500H Processor 12th Generation with 12 cores, 16gb ram, 2500Mhz and SSD for storage operations, 12 Core(s), 16 Logical Processor(s). Network connectivity was maintained through 500mb ethernet connection to minimize latency impacts on database operations

The software environment used MySQL Community Server 8.0. and MongoDB Community Service 8.0.3. Database management was assisted through MySQL Workbench 8.0.40 and MongoDB Compass 1.44.6 respectively. The development stack included OpenJDK 23.0.1 for core implementation, with Eclipse IDE 2023-09 serving as the primary development environment. The MongoDB shell, Mongosh 2.3.3, was used for direct interaction with the MongoDB database.

Maven 3.9.5 managed project dependencies and build automation. Version control was maintained through Git 2.42.0, with project artifacts stored in a private repository. Test data generation utilized Mockaroo’s for creating realistic user profiles.

The system architecture implemented domain-driven design principles (reference), separating concerns across:

* Entity definitions
* Data Access Objects (DAO)
* Service layer
* Test simulation framework

### Transaction processing implementation

## Analysis Methods

### Quantitative

The method employed to evaluate the performance characteristics and behavioural patterns of MySQL and MongoDB was collected with, instrumented service classes and automated testing scenarios. This approach aligned with the methodology proposed by Hellerstein et al. (2007) for database performance evaluation.

In terms of success rate, the formula utilized was:

*Success Rate = (Successful Bookings / Total Booking Attempts) × 100*

According to Gray and Reuter (2009)[1], transaction success rate directly correlates with system reliability and user experience in OLTP (Online Transaction Processing) systems. For ticketing systems specifically, Bernstein et al. (2019)[2] establish that a success rate above 95% is considered industry standard for high-concurrency booking systems.

The use of *AtomicInteger* ensured thread-safe counting in concurrent scenarios, following best practices outlined by Goetz et al. (2006) for Java concurrency.

*private final AtomicInteger successfulBookings = new AtomicInteger(0);*

*private final AtomicInteger failedBookings = new AtomicInteger(0);*

Consequently, the Average query time was taken in consideration to compare both databases performance. For ticketing systems, Zhao et al. (2020) recommend maintaining average query times below 100ms for optimal user experience.

*Average Query Time (ms) = Total Query Time / Total Queries*

This implementation followed the measurement methodology described in Simin Cai, Barbara Gallina, Dag Nyström & Cristina Seceleanu (2019), where timing provided more reliable metrics than individual query measurements.

For the concurrency metrics, an evaluation of transaction effectiveness was performed based on Kleppmann (2017), who said conflict rates in distributed systems proved insight into the effectiveness of concurrency control mechanisms. For ticketing systems, Chen et al. (2018) suggested that conflict rates should remain below 5% for acceptable performance.

*Conflict Rate = (Number of Concurrency Conflicts / Total Transactions) × 100*

The schema modification success rate indicated the adaptability of the database (Mark Lukas Möller, Stefanie Scherzinger, Meike Klettke & Uta Störl, 2020).

*Modification Success Rate = (Successful Modifications / Total Modification Attempts) × 100*

### Qualitative

## Testing Methodology

### Performance Test Scenarios

The performance testing focused primarily on quantitative measurement of database behaviour under controlled conditions. Specifically, test scenarios were structured to evaluate system performance across multiple dimensions.

In the first phase, transaction response time analysis provided baseline performance data. Specifically, measured query execution duration across various operation types. Additionally, transaction competition rates underwent continuous monitoring to assess system throughput under different load conditions.

### Concurrency Tests cases

Following the performance evaluation, the methodology examined database behaviour under simultaneous access patterns. Particularly, focused on transaction isolation and resource contention handling under conditions like production environments.

During the concurrent access testing, it was examined each database’s handling of simultaneous transactions, specifically, transaction processing and resource allocation.

In addition to performance metrics, data consistency formed a central component of concurrency testing. Under these circumstances, each system demonstrated distinct characteristics during high-concurrency operations.

### Schema Modification tests

The final phase focused on schema modification testing to assess structural adaptability. Initially, draw from Sadalage and Fowlers (2012) work on schema evolution patterns.

Structural adaptability began with basic schema modifications. Subsequently, the tests progressed to more complex scenarios involving relationship modifications and constraint evolution.

## Validation Strategy

### Experimental validation

#### Implementation of comparable booking systems in both MySQL and MongoDB

#### Controlled test environments to ensure fair comparison

#### Standardized metrics collection across both implementations

### Data and Process validation

Ensuring data integrity is fundamental for the reliable operation of any system, since it plays a central role in detecting and correcting errors, inconsistences, and inaccuracies within datasets. The primary types of data validation employed were format, and consistency validation.

Building upon the methodology proposed by Van der Loo (2020), validation was approached as a surjective function mapping datasets to Boolean values. This method implanted through explicit validation rules in both MySQL and MongoDB databases, allowing for systematic detection of data anomalies based on specific information requirements:

1. Single-Point Validations: Focused on individual data points, such as checking the status of a ticket.
2. Cross-Variable Validations: Examined relationships between different fields within a record to ensure logical consistency.
3. Cross-Record Validations: Assessed constraints across multiple records, for instance, booking limitations affecting several tickets.
4. Temporal Validations: Monitored changes in data over time, tracking the evolution of ticket statuses.

Both databases employed different approaches to enforce consistency. MySQL maintained consistency through atomic transactions using *EntityManager,* along with the application of pessimistic locking for concurrent access control. Data integrity constraints were also utilized including foreign key relationships to maintain referential integrity, unique constraints to prevent duplicate bookings, and check constraints to validate business rules.

In contrast, MongoDB application utilized *SessionFactory* to manage multi-document transactions that ensured atomicity across multiple collections. Optimistic concurrency control was implemented (in contrast to MySQL pessimistic locking) using version fields within documents. Finally, schema validation was enforced through JSON Schema definitions, specifying structure and data types for documents.

Validation concurrent operations was important to address the challenges posed by simultaneous access and modification of data by multiple users. Key considerations included handling race conditions, preventing deadlocks, and maintained consistency levels.

The schema validation process involved ensuring that changes to the database did not disrupt ongoing operations or compromise data integrity. Version tracking were employed to monitor changes over time, aiding in maintaining compatibility with different versions of the application. On the other hand, Mongo benefited from its flexible schema design, allowing dynamic updates without adversely affecting existing documents.

# Implementation

A screenshot of a computer

Description automatically generated

## System Architecture

The system architecture implemented a layered approach, adhering to Separation of Concerns principles (SoC) (Software Architect's Handbook by Joseph Ingeno Released August 2018) while easing comparative analysis between MySQL and MongoDB implementations. The architecture comprised three primary layers: data access, service, and simulation framework.

### Data Access Layer

The data access layer established the foundation for database operations through a set of Data Access Objects (DAOs). These objects encapsulated the database specific implementations while maintaining a consistent interface for both MySQL and MongoDB. The layer included specialized DAOs for managing bookings (*BookingDAO*), tickets (*TicketDAO*), users (*UserDAO*), events (*EventDAO*), and related entities. For MySQL, the implementation utilized JPA annotations and Hibernate persistence provider, while MongoDB implementation employed Morphia for object-document mapping.

The layer implemented the following components:

* Entity mappings (JPA annotations for MySQL, Morphia annotations for MongoDB)
* CRUD operations for all domain objects
* Transaction management (ACID properties for MySQL, session management for MongoDB)
* Connection pool management using HikariCP for MySQL and MongoDB’s native connection pooling

### Service Layer

The service layer implemented specialized transaction management strategies that addressed the distinct operational characteristics of both databases. This component was important for maintaining data consistency while enabling comparative analysis of the different transaction models.

MySQL implementation supported ACID transaction properties through JPA/Hibernate, employing a pessimistic locking strategy for concurrent ticket reservations (**references). T**his approach ensured strict data consistency by establishing explicit transaction boundaries at the database level (**references).** The system utilized connection pool optimization to manage transaction throughput, implementing a configurable pool size that balanced resource utilization with response time requirements (**references).**

On the other hand, MongoDB required a different architectural approach due to its distributed nature (**what is its nature)**.The system implemented multi-document transactions through client sessions, managing transaction state across multiple operations. This approach utilized optimistic concurrency control for ticket reservations, incorporating retry logic to handle conflict scenarios(**reference) (why optimistic lock).** This drawed the hypotheses that effective handling of concurrent operations will be supported while maintaining data consistency, albeit with different performance characteristics compared to MySQL **(references)**.

In relation to transaction coordination, the booking coordination mechanism implemented a two-phase ticket reservation process (**references)**. During the first phase, the system verified resource availability and established preliminary locks. The second phase either confirmed the booking through a commit operation or released resources through a rollback procedure (**why? What is rollback? References).** The decision of this approach was to manage concucurrent access patterns while preventing race conditions in high-load scenarios (**we need references, more studies that demonstrated this).**

State management formed a critical component of the transaction architecture. The system tracked transitions through distinct states. AVAILABLE, marked the beginning of a transaction, RESERVED indicating active resource allocation, CONFIRMED represented successful completion, and FAILED denoted unsuccessful transactions requiring compensation. Each state transition triggered specific validation rules and consistency checks (**what are the consistency checks),** ensuring data integrity throughout the booking lifecycle.

Moving forward the concurrency control mechanism (**references, what is it)** implemented different strategies for each database. MySQL utilized database-level locks with configurable timeout parameters (**references)**, for duplicate bookings prevention. While Mongo implementation employed a combination of document-level locks and version control mechanisms, implementing retry logic when conflicts occurred (**references)(**why this approach).

Resource management was implemented with a queue mechanism(**what queue)(references)** that regulated incoming booking requests based on available system resources. This approach prevented system degradation under heavy load while maintaining fair request processing order(**references, what is system degradation, under what conditions it is present).** The implementation included monitoring mechanisms that tracked queue length and processing times, enabling dynamic adjustment of concurrency parameters (**references).**

**Error handling and recovery**

### Simulation Framework

The simulation framework facilitated the comparative analysis through structured test scenarios, designed to evaluate database performance and behaviour under controlled conditions. This layer comprised:

1. Scenario specific implementations:

Initially, the framework implemented four primary test scenaros, each designed to evaluate specific database characteristics. For example, the transaction consistency testing scenario for MySQL examined ACID compliance under concurrent operations, ultimately implementing controlled transaction conflicts to measure system response (**references).**

Next, the schema modification impact assessment scenario evaluated MySQL’s behavior during structural changes under load. Similarly, this implementation monitored system performance during table alterations, measuring transaction success rates and response times while modifications occurred (**references).** Furthermore, the scenario included the addition/modification of columns, index creation???, and constraint modifications during active booking operations.

For MongoDB, however, the nested data handling scenario evaluated document database capabilities in managing complex data structures. In general, the configurations, therefore measuring query performance and data manipulation efficiency(**is this true???).**

Likewise, the concurrent operation conflict resolution scenario for MongoDB examined the database’s handling of simultaneous booking attempts.

1. Metrics collection mechanisms

In the meantime, the framework implemented comprehensive metrics collection through instrumented service layers. Specifically, response time measurement captured.

Furthermore, transaction success and failure tracking implemtned detailed event logging, capturing the cause of each failure and the ystem state at the time of occurrence. Consequently, this mechanism enabled post-execution analysis of failure patterns and system behavior under stress conditions (**HOW? REFERENCES).**

1. Test data management

Ultimately, the framework implemented controlled test data population through the initialization framework. For example, initial data population followed predefined patterns ensuring consistent starting conditions across test iterations (**references).** In addition, the implementation include ddata verification mechanisms that validated test data integrity before and after each test execution (**how?? References).**

The architecture employed dependency injection patterns (DI) (**references**) to maintain loose coupling between components, enabling isolated testing of database specific characteristics while maintaining consistent application behaviour across both implementations. The separation of concerns improved the collection of comparative metrics without compromising the integrity of the core business logic (**references**).

### Data initialization

The system’s initialization architecture implemented a structured approach to populate test data across both database implementations. This design addressed the challenge of maintaining consistent test data while accommodating the different data models of MySQL and MongoDB.

The initialization framework resolved three key aspects:

1. Managing different referential integrity approaches between MySQL and MongoDB
2. Ensuring consistent test data population across both databases
3. Maintaining data validity for comparative testing

The framework was structured in three main components:

1. *DataInitializer* class served as the central coordinator, managing the initialization sequence and ensuring proper dependency resolution. This approach was necessary to handle the relationships between entities, particularly in maintaining referential integrity in MySQL while allowing for MongoDB’s more flexible document relationships.
2. *Entity* initialization implemented specialized initializers for each domain entity:
   1. Base data initializers (*GenreInitializer, PerformerInitializer, VenueInitializer)*
   2. Event structure initializers *(EventInitializer, TicketCategoryInitializer)*
   3. Transaction data initializers (*TicketInitializer, UserInitializer, BookingInitializer)*

The initialization sequence followed a predetermined order based on entity relationships, ensuring that dependant data was created only after its prerequisites were established.

### Domain Entities

The system implemented a comprehensive set of domain entities that adopted the core business objects. These entities established the data structure for both database implementations while accommodating their distinct data modelling approaches:

*Event* entity managed performance event data, maintaining relationships with performers, venues, and ticket categories. The implementation supported multiple performers per event and complex ticket category hierarchies.

*Ticket* entity handled individual ticket information, incorporating status management, category association, and booking relationships. The design supported atomic operations (**references**) for concurrent booking scenarios.

*Booking* entity managed transaction data, maintaining relationships between users, events, and tickets. The implementation incorporated status tracking and payment processing requirements.

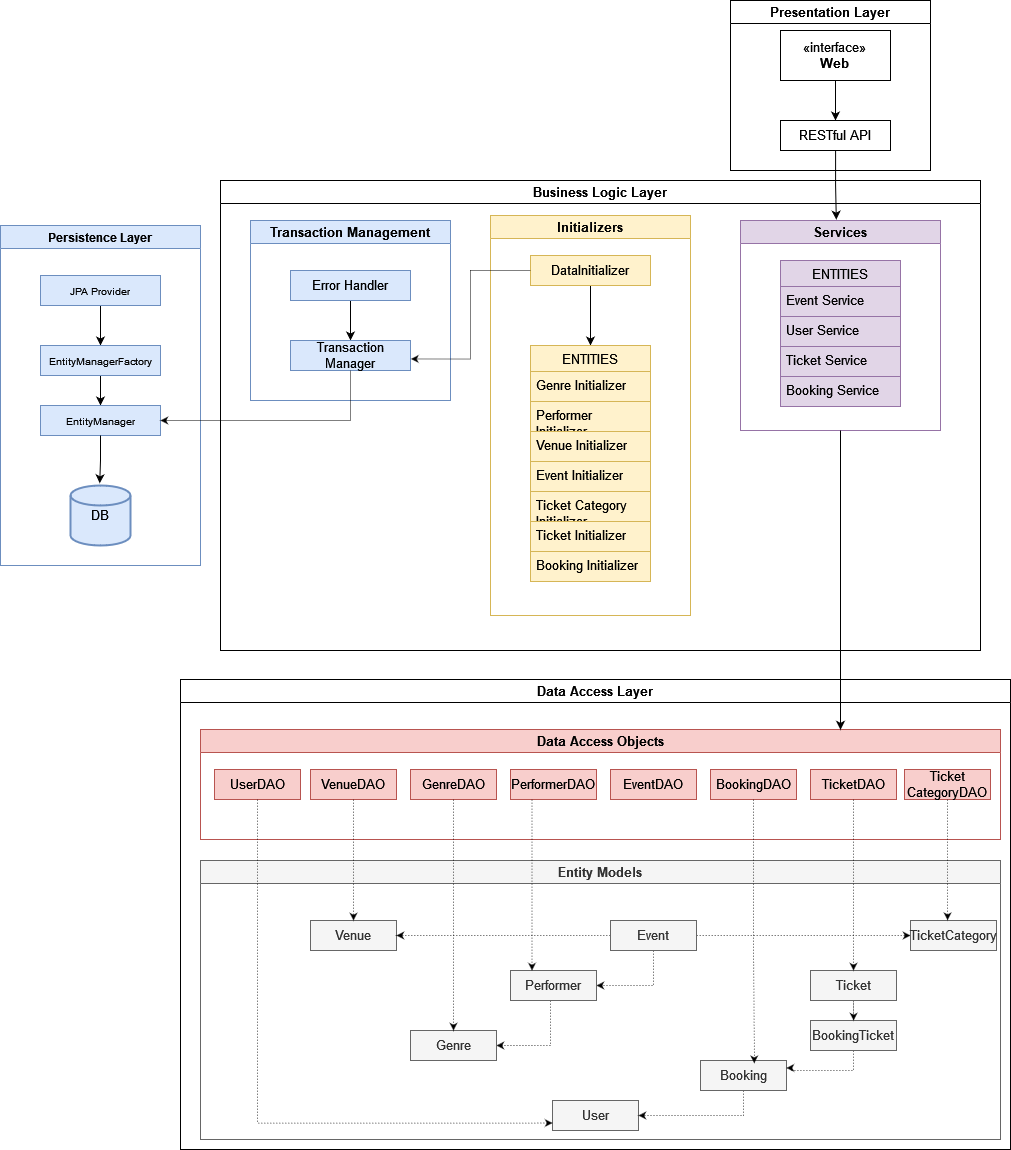
*User* entity handled customer information and booking history, supporting authentication and profile management requirements.

Supporting entities included:

* *Genre:* Managed musical style classifications
* *Performer:* Handled artist and performer information
* *Venue:* Managed location and capacity data
* *TicketCategory:* Implemented pricing and seating classification

The entity design accommodated both relational and document-based models for a consistent business logic across implementations through flexible relationships mapping strategies (**references**), the appropriate use of normalization in MySQL implementation, strategic denormalization in MongoDB implementation, and consistent identity management across both platforms (**references)**

## MySQL implementation



Create a New Maven Project, and configure the dependencies in pom.xml (annex??). JPA dependency, hibernate for JPA implementation with MySQL, MySQL connector, Mongo Java Driver, Junit for Testing, and Log4j for logging. JPA is designed for relational databases, thus integrating with MongoDb requires

Create a Maven project, and selelect the “java-quick-archetype”.

Add a groupid and an artifactid. The version is 0.0.1-SNAPSHOT,

Setting up pom.xml:

Inside the project a pom.xml should be created. Inside the Hibernate and MySQL dependency has been included.

<dependencies>

<dependency>

<groupId>org.hibernate</groupId>

<artifactId>hibernate-core</artifactId>

<version>5.4.2.Final</version>

</dependency>

<dependency>

<groupId>mysql</groupId>

<artifactId>mysql-connector-java</artifactId>

<version>8.0.22</version>

</dependency>

</dependencies>

Setting up the persistence.xml, connecting it with the local host MySQL. A persistence unit defines the details that are required when you acquire an entity manager.(*Oracle® Fusion Middleware Understanding Oracle TopLink*, 2015)

Database Creation and schema definition

A new MySQL database named TicketSystem was created to host all the application data.

The database schema was defined using SQL “*CREATE TABLE”* statements, afterwards the tables for key entities: *genres, performers, venues, events, ticket\_category, tickets, users, bookings, and booking\_ticket.*

EacH table included the *“AUTO\_INCREMENT”* for unique identification, also constrainst like “*NOT NULL”* and “*UNIQUE”* were applied. Finally the schema (figure 1) as designed in third normal form, to reduce redundancy, and optimize query performance.

### System Architecture & Technical stack

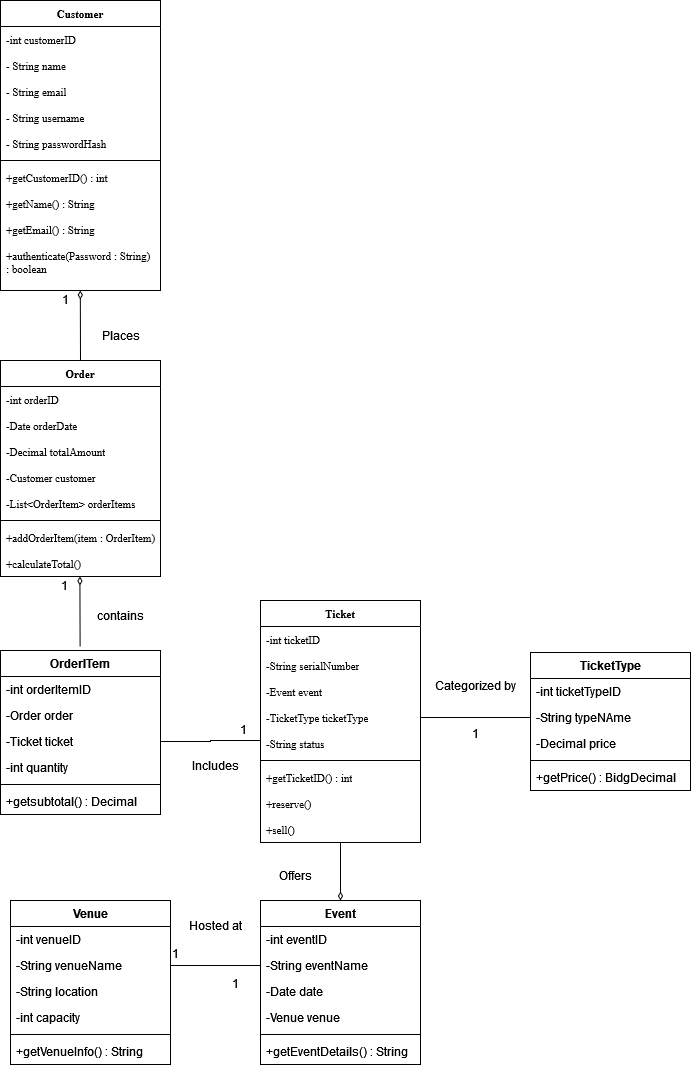
#### MySQL Setup

#### Connection configuration

### Data Model Design

#### Relational Schema Design

#### Transaction management



A screenshot of a computer

Description automatically generated

### Transaction management

 ACID implementation

 Locking mechanisms

 Isolation levels

 Error handling

Enhancing DAOs for Transaction Management and Locking

Before implementing the BookingService, we'll make slight enhancements to your DAOs to support pessimistic locking when reserving tickets. This ensures that when a ticket is being reserved, other transactions cannot modify it until the current transaction is complete. JPA specification defines three pessimistic lock modes that we’re going to discuss:

*findBySerialNumberWithLock* method retrieved a *Ticket* by its serial number and applied a *PESSIMISITC\_WRITE* lock. This for ensured that once a ticket is fetched for reservation, other transactions cannot modify or reserve the same ticket until the current transaction is completed (Kucharz and Valero Sanchez, 2018).

To maintain transactional integrity, all DAO operations within a service used the same *EntityManager*.

Implementing the *BookingService*

The *BookingService* class coordinated the booking process, ensuring that all steps were executed within a single transaction. If any step failed, the entire transaction would be rolled back to maintain data consistency.

**Explanation:**

* **Transactional Control:** Manages transactions manually using EntityTransaction.
* **Pessimistic Locking:** Utilizes the findBySerialNumberWithLock method to apply a PESSIMISTIC\_WRITE lock when reserving tickets.
* **Booking Status:** Directly sets the booking status to CONFIRMED upon successful reservation, bypassing the payment step.
* **Error Handling:** Ensures that any failure during the booking process triggers a transaction rollback, preventing partial data persistence.

 **Transactional Control:**

* The service begins a transaction using EntityManager to ensure that all booking operations are executed atomically.
* If any step fails, the transaction is rolled back to maintain data integrity.

 **Pessimistic Write Locking:**

* The reserveTickets method employs pessimistic locking by invoking ticketDAO.findAvailableTicketsWithLock(eventId, quantity, LockModeType.PESSIMISTIC\_WRITE), which retrieves available tickets with a PESSIMISTIC\_WRITE lock.
* This lock prevents other transactions from accessing the same tickets until the current transaction is completed, thereby avoiding overselling.

 **Data Access Objects (DAOs):**

* **BookingDAO:** Handles the creation and management of Booking entities.
* **TicketDAO:** Manages ticket retrieval and updates, including applying locks during reservation.
* **UserDAO:** Retrieves user information based on user ID.
* **EventDAO:** Fetches event details, including ticket categories and pricing.

 **Booking Process Flow:**

* **User and Event Retrieval:** The service first retrieves the user and event details to validate the booking request.
* **Availability Check:** It then checks if the desired number of tickets is available.
* **Ticket Reservation:** Reserved tickets are locked and marked as RESERVED to prevent concurrent bookings.
* **Price Calculation:** The total price is calculated based on the ticket categories.
* **Booking Creation:** A new Booking record is created and associated with the reserved tickets.
* **Ticket Status Update:** Tickets are updated to SOLD to finalize the booking.

 **Metrics Collection:**

* **successfulBookings:** Tracks the number of successful booking transactions.
* **failedBookings:** Monitors the number of failed booking attempts, aiding in performance and reliability analysis.

 **Error Handling:**

* Comprehensive error handling ensures that any exception during the booking process results in a transaction rollback, maintaining system consistency.
* Specific BookingException instances provide clear feedback on the nature of failures.

**Booking Simulation Class for JPA/MySQL-Based Ticketing System**

The BookingSimulation class is designed to simulate a high-concurrency environment where multiple users attempt to book tickets simultaneously. This simulation is intended to test the effectiveness of the PessimisticBookingService in handling concurrent booking requests without compromising data integrity or allowing ticket overselling.

**Key Objectives:**

1. **Simulate Concurrent Bookings:** Emulate 100 users, each attempting to book 2 tickets for the same event concurrently.
2. **Measure Performance Metrics:** Track the number of successful and failed bookings to evaluate system performance under load.
3. **Ensure Data Consistency:** Verify that the system accurately maintains inventory counts and prevents overselling.

**Class Overview:**

* **Initialization:** Sets up necessary components such as the EntityManager, DAOs, and the PessimisticBookingService.
* **Simulation Execution:** Utilizes a thread pool to manage and execute multiple booking tasks concurrently.
* **Metrics Collection:** Employs thread-safe counters (AtomicInteger) to track successful and failed bookings.
* **Reporting:** Outputs the results of the simulation, including total transactions, success rates, and failure counts.

**Implementation Details:**

1. **Thread Pool Management:** Uses ExecutorService with a fixed thread pool to manage concurrency efficiently.
2. **Task Definition:** Each booking task represents a user attempting to book a specified number of tickets.
3. **Synchronization:** Ensures thread-safe operations when updating shared metrics using AtomicInteger.
4. **Error Handling:** Captures and logs exceptions to prevent simulation crashes and accurately count failed bookings.
5. **Resource Cleanup:** Properly shuts down the thread pool and closes the EntityManager and EntityManagerFactory to release resources.

 Lock management

 Deadlock prevention

 Performance optimization

 Resource management

MySQL Connection Process

Establishing a connection to the MySQL database using JPA involved configuration and troubleshooting to ensure integration. Initially, the setup required defining the database schema with structured tables and relationships to maintain data integrity. Configuring the persistence.xml file was crucial, where specifying accurate connection properties, including the *JDBC URL*, user credentials, and the appropriate *Hibernate* dialect, was essential. A significant challenge encountered was a “*ZoneRulesExceptio*n” caused by an unrecognized time zone ID, which necessitated the addition of a valid “*serverTimezone*” parameter to the JDBC URL to align Java's time zone recognition with the MySQL server settings. Additionally, ensuring that all JPA entity classes were correctly annotated and registered within the persistence unit was important to prevent mapping errors. Another obstacle was rectifying method usage errors, such as substituting *createNamedQuery* with *createQuery* for executing ad-hoc JPQL queries, which resolved exceptions related to undefined named queries. These challenges highlighted the importance of precise configuration and validation of entity mappings.

* Used jakarta.persistence:jakarta.persistence-api
* Used org.hibernate:hibernate-core
* Used mysql:mysql-connector-java

Data population

The implementation utilized the Java Persistence API (JPA) alongside MySQL to initialize data across multiple interconnected tables.

The primary mechanism for data population was the *“DataInitializer”* class, using JPA’s Object-Relational Mapping (ORM) to manage entity relationships and data insertion. Data Access Object (DAO) classes were employed to handle CRUD (Create, Read, Update, Delete) operations for each entity. To prevent data duplication, the initializer utilized methods such as *“findByName()”* within each DAO to verify the existence of records prior to insertion. Additionally, SQL constructs like *“INSERT IGNORE”* were incorporated within native queries to bypass duplicate entries based on predefined unique constraints.

Best Practices to Address Thesis Objectives

Modularity:

Each initializer class handles a specific entity, promoting clean separation of concerns.

Maintainability:

Easier to update or modify data initialization logic for a specific entity without affecting others.

Scalability:

Adding new entities or modifying existing ones requires creating or updating only the relevant initializer class.

Efficiency:

Preloading existing records minimizes database queries, enhancing performance during data initialization.

Transactional Integrity:

Each entity’s data population is encapsulated within its own transaction, ensuring atomicity and reducing the risk of large-scale rollbacks.

CRUD Operations

**Why Not Work Directly with MongoDB Compass or MySQL:**

Instead of working directly with MongoDB Compass or MySQL command-line tools, I focused on backend development to create a controlled environment for testing and comparing the databases. Building the backend with Java and ORM tools allows for:

1. **Automated Data Handling:** Automating data initialization and transaction scenarios ensures consistent and repeatable testing conditions, which is crucial for accurate performance comparisons.
2. **Concurrency Simulation:** Implementing concurrency within the application provides a more realistic simulation of multiple users interacting with the system simultaneously.
3. **Scalability:** A backend application can be easily scaled and modified to test different scenarios without manual intervention, facilitating a comprehensive analysis.
4. **Integration with Both Databases:** Using an Object Relational Mapping framework like Hibernate for MySQL and a similar approach for MongoDB ensures that the comparison focuses on how each database handles similar operations within the same application context.

**Work Completed So Far:**

* **Database Schema Development:**
  + **MySQL:**
    - Designed a normalized relational schema with tables for genres, performers, venues, events, tickets, users, bookings, and their relationships.
    - Enforced data integrity using foreign keys, unique constraints, and NOT NULL constraints.

**Java Implementation:**

* Developed a Java application using Hibernate for MySQL to initialize and populate the database.
* Defined entity classes, Data Access Objects (DAOs), and an initializer to manage data insertion and ensure data integrity.
* Implemented transaction management to maintain consistency in a concurrent environment.

1. Initialization and Data population :
2. Parameter Type mismatch error
3. Enum mapping issue with TicketStatus
4. Infinite Loop during ticket generation

**Duplicate Entry Errors:** Implemented checks to prevent insertion of tickets with duplicate serial numbers, ensuring uniqueness and data integrity.

* Eclipse thrown an *SQLInteritytCOnstaintViolationException* , coming from a duplicate entry in the *serial\_number* field of the tickets table. The error was in the insertion of data, or attempting to insert tickets with serial numbers that already existed in the database.
* *INSERT INTO tickets (event\_id, purchase\_date, rownumber, seat\_number, section, serial\_number, status, ticket\_category\_id)*
* *VALUES (…)*
* *ERROR: Duplicate entry 'ROVI0001' for key 'tickets.serial\_number'*

**Parameter Type Mismatches:** Corrected issues in JPQL queries by ensuring the correct data types are used, which improved query reliability.

* *IllegalArgumentException* occurred because an Event object was passed where an Integer event ID was expected in the findByEventId method of TicketCategoryDAO.

**Enum Mapping Issues:** Created a custom converter to manage differences between database values and Java enums, resolving mapping errors and aligning database records with application logic.

* The system couldn't map database values to Java *enum* constants due to case sensitivity differences (AVAILABLE in Java vs. available in the database), leading to an *IllegalArgumentException*.

**Infinite Loop in Ticket Generation:** Introduced logical limits to prevent infinite loops during ticket creation, enhancing the stability of the data initialization process.

* The seatNumber kept incrementing without bounds. So the solution was to introduce limits for rows and seats within each ticket category and reset the seatNumber for each category.

**Booking-Ticket Associations:** Ensured that associations between bookings and tickets are correctly created and persisted, establishing accurate relationships within the database.

**Transaction Management:** Improved transaction handling to maintain data integrity in a concurrent environment, which is crucial for a high-concurrency system.

try {

em.getTransaction().begin();

// Perform database operations

em.getTransaction().commit();

} catch (Exception e) {

em.getTransaction().rollback();

throw e;

}

## MongoDB implementation

### System Architecture & Technical Stack

#### MongoDB setup

Configuring MongoDB involved establishing a comprehensive development environment and managing technical dependencies. Transitioning from a JPA-based architecture required adjustments to project configurations and dependency management. Essential components included:

1. MongoDB Community Edition with network service configuration
2. MongoDB Compass for database visualization and management
3. MongoDB Shell (mongosh) for command-line operations
4. MongoDB Driver Sync (version 4.5.1) for database connectivity
5. Morphia as the Object-Document Mapping (ODM) (version 2.2.6)

The selection of Morphia as the ODM solution was based on its compatibility with MongoDB’s document model, replacing the traditional JPA implementation that was not inherently designed for NoSQL databases. This choice enabled a more natural mapping between Java objects and MongoDB documents.

#### Connection configuration

Database initialization and connection management were centralized within a *DataInitializer* class, which served as the central component for establishing and managing MongoDB connections. This class utilized Morphia ODM and implemented an approach to database connectivity and data management.

Key implementation aspects included:

1. Establishing the MongoDB connection using *MongoClients.create()* with a local connection string *(“mongodb://localhost:27017”)*
2. Creating a Morphia *Datastore* to manage the “*ticketsystem*” database through *datastore.ensureIndexes()* (Kumar, 2019)
3. Managing the connection lifecycle with the *MongoClient* instance
4. Executing resource cleanup through a dedicated *close()* method
5. Implementing connection validation via data verification routines

public DataInitializer() {

mongoClient = MongoClients.create("mongodb://localhost:27017");

datastore = Morphia.createDatastore(mongoClient, "ticketsystem");

datastore.getMapper().mapPackage("dev.morphia.example");

datastore.ensureIndexes();

}

### Document Model Design

A combination of embedding and referencing strategies to maintain data consistency across collections (Figure 1):

1. Genres
   1. Stored information about genres linked to performers
   2. Enforced a unique name field with indexing to ensure rapid queries and prevent duplicates
2. Performers
   1. Contained details, with each performer referencing *genre\_id* from the genres collection
   2. Unique indexes on *name* and *genre\_id* for data integrity and query efficiency
3. Venues
   1. Hold venue details, with filds like *name, address, type,* and *capacity.*
   2. The schema enforces uniqueness on venue names and requires non-negative capacity values.
4. Events
   1. Encapsulated event details, including embedded ticket categories with attributes like *description, price, start\_date,* and *area*
   2. The schema also used compound indexes to prevent duplicate events and optimize lookups by performer or venue
   3. Embedded ticket categories to simplify data retrieval by tightly coupling them with their events
5. Users
   1. Represented system users with unique *username* and *email* fields
   2. Validated email formats and indexed fields to enhance performance and avoid duplicated entries
6. Booking
   1. Recorded user ticket purchases, including *user\_id, delivery\_email, status,* and *tickets*
   2. Referenced the *tickets* collection in the tickets array to reduce redundancy and enable associations between bookings and tickets
   3. Indexed *user\_id* and *status* supported lookups by user and booking state
7. Tickets
   1. Stored individual ticket details, such as *serial\_number, event\_id, section,* and *status.*
   2. Ensured each ticket had a distinct serial number through unique indexes, while compound indexes on *event\_id* and *status* enabled quick retrieval of available tickets
   3. Denormalized the ticket category to simplify queries

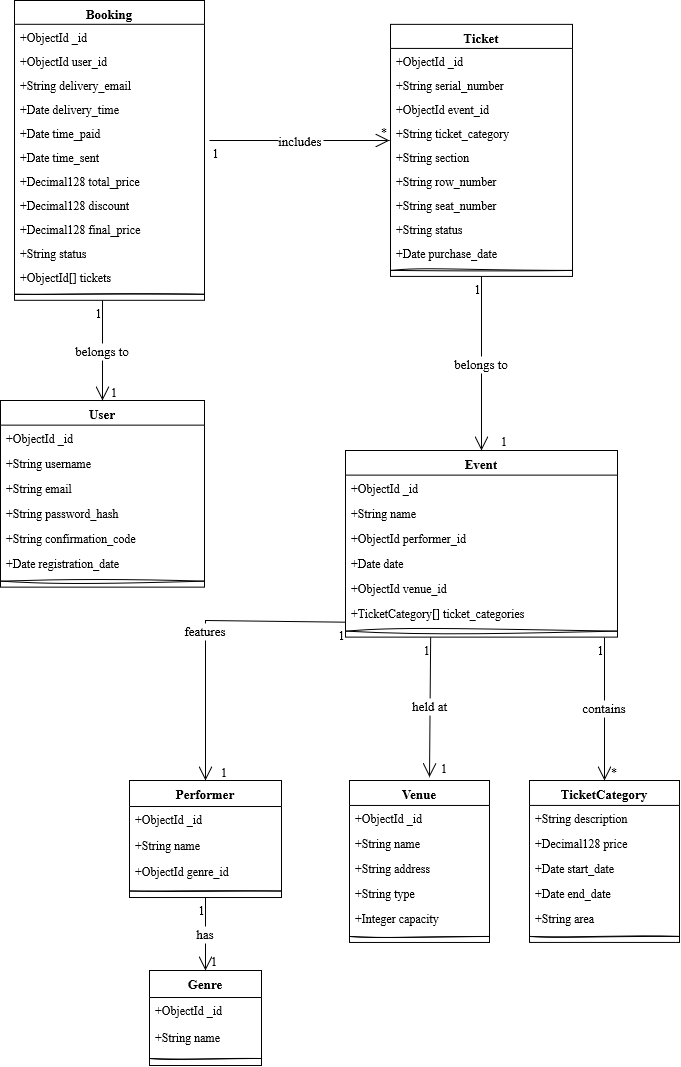


Figure 1. MongoDB System Class Diagram

#### Flexibility vs Rigidity Impact

Adopting a schema-less architecture (Figure 2) allowed each document within a collection to possess a distinct structure. The flexibility accommodated evolving application requirements without necessitating extensive schema migrations. In the context of the ticketing system, embedding *TicketCategory* documents with *Event* documents optimized read performance by minimizing join operations.

A diagram of a function

Description automatically generated

Figure 2. DB Schema Flexible using MongoDB

Nevertheless, the flexible schema introduced challenges related to data consistency and redundancy. Maintaining consistent data structures across documents required meticulous application-level validations. For instance, ensuring that all *User* documents contained necessary fields such as *passwordHash* and *confirmationCode* necessitated validation logic within the application layer.

Example:

@Entity("users")

public class User {

@Id

private ObjectId id;

@Property("user\_name") //maps to username

@Indexed(options = @dev.morphia.annotations.IndexOptions(unique = true, name = "userName\_idx"))

private String userName;

@Property("email")

@Indexed(options = @dev.morphia.annotations.IndexOptions(unique=true, name = "email\_idx"))

private String email;

@Property("password\_hash")

private String passwordHash;

@Property("confirmation\_code")

private String confirmationCode;

@Property("registration\_date")

private Date registrationDate;

Additionally, denormalization, while beneficial for read performance, increased the potential for data redundancy. Embedding related data could lead to duplicated information if not managed appropriately, thereby increasing storage requirements and complicating update operations. For example, if a *TicketCategory* description needed to be updated, the change would have to be propagated to every embedded instance within *Event* documents to maintain consistency.

@Entity("ticketCategories")

public class TicketCategory {

//@Id

//private Object id;

@Property("description")

private String description;

@Property("price")

private BigDecimal price;

@Property("start\_date")

private Date startDate;

@Property("end\_date")

private Date endDate;

@Property("area")

private String area; // Seating area description

Challenges encountered during implementation, particularly in embedding ticket categories within event documents and preventing data duplication. These were resolved by using Morphia annotations. Replacing the deprecated @*Embedded* with *@Reference* without an *@Id* field for embedded documents (Chooly, 2019). The collection for ticket categories was changed from a *List* to a *Set,* and equality checks were implemented to ensure uniqueness. These adjustments not only resolved immediate issues but also enhanced the system’s ability to handle dynamic data structures without sacrificing consistency.

@Entity("events")

@Indexes({

@Index(fields = {@Field("event\_name"), @Field("date")}, options = @IndexOptions(unique = true, name = "name\_date\_idx")),

@Index(fields = @Field("performer"), options = @IndexOptions(name = "performer\_idx")),

@Index(fields = @Field("venue"), options = @IndexOptions(name = "venue\_idx"))

})

public class Event {

@Id

private ObjectId id;

@Property("event\_name")

private String name;

@Reference(lazy = true)//References the performer document, loaded lazily

private Set<Performer> performers;

@Property("date")

private Date date; //date of the event

@Reference(lazy = true)

private Venue venue; //References the venue document

//("ticketCategories") //embeds a list of TicketCategory documents within the Event document

@Reference("ticketCategories")

private Set<TicketCategory> ticketCategories;

The schema proposed offered significant flexibility by enabling embedding of related documents, such as *TicketCategory* within *Event* documents to optimize performance. However, this flexibility introduced challenges, including the need for a strict application-level validations to ensure consistency and the risk of redundancy inherent in denormalized designs.

#### Data Structure Management

Document-oriented model facilitated the natural representation of complex data structures within individual documents. Implementing unique indexes across various collections prevented duplicate entries and ensured the uniqueness of critical fields, thereby enforcing data integrity at the database level and eliminating the possibility of duplicate records that could compromise reliability.

Example:

@Entity("bookings") // Maps to the 'bookings' collection

@Indexes({

@Index(fields = @Field("user\_id"), options = @IndexOptions(name = "user\_id\_idx")),

@Index(fields = @Field("status"), options = @IndexOptions(name = "status\_idx})

public class Booking {

@Id

private ObjectId id;

@Property("user\_id")

private ObjectId userId;

// \*\*Add event\_id to Booking\*\*

@Property("event\_id")

private ObjectId eventId;

@Property("delivery\_email")

private String deliveryEmail;

@Property("delivery\_time")

private Date deliveryTime;

@Property("time\_paid")

private Date timePaid;

@Property("status")

private String status;

@Property("tickets")

private List<ObjectId> tickets;

The *Booking* entity referenced both *User* and *Ticket* documents, maintaining normalization by avoiding the embedding of extensive ticket arrays within a booking. *@Indexed annotations* ensured that fields maintained unique values. Attempting to insert duplicate values in these fields would result in a database error, thereby preserving data uniqueness.

#### Relationship Handling

One-to-Many relationships were implemented through embedding or referencing dependent entities. For example, a *User* could have multiple *Booking* documents, established via a one-to-many relationships where each booking referenced its associated user.

Many-to-many relationships managed using referencing strategies. The relationship between *Performer* and *Genre* entities was handled by referencing genres within performer documents or vice versa, depending on access patterns.

Implementation techniques included:

* Morphia annotations:
  + *@Reference:* Simplified referencing documents without embedding, maintaining loose coupling between entities
* Data integrity maintenance:
  + Application-level validation ensured referential integrity through application logic due to the lack of enforced foreign key constraints in MongoDB
  + Indexing creation on referenced fields to optimize join-like operations

Example:

@Entity("events")

public class Event {

@Id

private ObjectId id;

private String name;

private Date date;

@Reference

private Venue venue; // Referenced for scalability

@Reference

private List<TicketCategory> ticketCategories; //

}

The *Event* entity referenced *TicketCategory* and *Venue* using @*Reference* annotation, enabling the retrieval of related data without embedding entire performer or venue documents within each event.

### Transaction management

Transaction management represented an important component of the system, particularly in handling concurrent booking operations while maintaining data consistency. While traditional NoSQL systems often prioritize eventual consistency following the CAP theorem (Brewer, 2012). However, MongoDB’s introduction of multi-document ACID transactions starting from version 4.0 (O’Grady, 2020) marked a significant shift toward stronger consistency guarantees. This implementation bridged the gap between NoSQL flexibility and traditional RDBMS reliability, following the theoretical framework established by Pritchett (2008) regarding eventually consistent systems

A hybrid approach balanced ACID compliance with distributed system performance, aligning with Stonebraker's (2010) observations on the necessity of maintaining transactional integrity in specific domains. The structured booking service coordinated multiple database operations within atomic units, implementing what Gray and Lamport (2006) termed a “distributed transaction commit protocol”, modified for document-oriented databases displayed in Figure 3.

A diagram of a process

Description automatically generated

Figure 3. Transaction Flow

#### Core Transaction component

Transaction management was executed through the *BookingService* class, following the Repository Pattern (Fowler, 2002). This separation of transaction logic from data access concerns enabled a clean transaction structure, where the *BookingService* acted as the primary transaction coordinator, managing multiple Data Access Objects (DAOs) and maintaining transaction metrics.

// Primary transaction coordinator

public class BookingService {

private final BookingDAO bookingDAO;

private final TicketDAO ticketDAO;

private final UserDAO userDAO;

private final EventDAO eventDAO;

private final Datastore datastore;

// Metrics tracking

private AtomicInteger successfulBookings = new AtomicInteger(0);

private AtomicInteger failedBookings = new AtomicInteger(0);

}

Key characteristics included:

1. Resource encapsulation
   1. Immutable and separated DAO references promoted modularity
   2. Centralized transaction within the *BookingService*
2. State management
   1. Atomic operations tracked successful and failed bookings using *AtomicInteger*

#### Transaction protocol

A multi-phase commit pattern, aligned with the ACID transaction protocol (Faraj, 2022) was implemented, involving distinct phases of resource verification, allocation, state update, and commit. This approach ensured atomicity and consistency throughout the transaction lifecycle.

Session-based transaction processing managed the lifecycle. During session initialization, the system acquired the necessary resources, enabling transaction boundaries to encapsulate all operations within a single transactional context. Java’s try-with-resources construct aided the atomic cleanup of resources.

The state management component utilized MongoDB’s *ClientSession* to maintain transaction integrity. Also, the employment of atomic counters (*successfulBookings and failedBookings)* to track transaction outcomes, providing metrics for monitoring booking attempts (Salander, 2022). All operations were confined within their respective session scopes, which maintained isolation and consistency of the transaction data.

The transaction workflow was structured through distinct phase separation.

1. Resource verification
   1. Validated the event’s existence and checked available ticket quantities
2. Resource allocation
   1. Attempted to book the requested number of available tickets
3. State update
   1. Created a new booking in the system
4. Commit phase
   1. Committed all changes to the database

public boolean bookTickets(ObjectId userId, ObjectId eventId, int quantity) {

try (ClientSession session = datastore.startSession()) {

session.startTransaction();

try {

// Phase 1: Resource Verification

Event event = eventDAO.findById(eventId);

long availableTickets = ticketDAO.countAvailableTickets(eventId);

// Phase 2: Resource Allocation

List<Ticket> bookedTickets = ticketDAO.bookAvailableTickets(

session, eventId, quantity);

// Phase 3: State Update

Booking booking = new Booking(/\* ... \*/);

bookingDAO.create(booking);

// Phase 4: Commit

session.commitTransaction();

successfulBookings.incrementAndGet();

return true;

} catch (Exception e) {

// Rollback Phase

session.abortTransaction();

failedBookings.incrementAndGet();

return false;

}

}

}

Error handling utilized a try-catch mechanism defining error boundaries, employing transaction abort functionality to ensure automatic rollback of all changes in case of failures (Bernstein and Newcomer, 2009). This mechanism guaranteed that the system-maintained consistency by reverting to its previous state when errors occurred, while also incrementing the *failedBookings* counter for monitoring purposes.

#### Core Concurrency component

The *BookingSimulation* class (Appendix 17) embodied the work done by Seppälä (2024) helping management of concurrent booking operations through a fixed thread pool and task queuing mechanisms.

This class depended on instances of *BookingDAO, UserDAO, EventDAO, TicketDAO,* and *Datastore,* object from the Morphia library, enabling interaction with the underlying data storage.

**The simulation architecture employed a fixed-thread executin model with configurable parameters to facilitate systematic performance evaluations across various load scenarios. The configuration included constants such as *NUM\_USERS* set to 1000, *MAC\_TICKETS\_PER\_USER* limited to 1, and *THREAD\_POOL\_SIZE* optimized to 10 based on references recommendation for balancing resource utilization and thread management overhead. Additionally, *SIMULATION\_TIMEOUT\_MINUTES* was established at 1 minute, and *BATCH\_SIZE* was configured to 100.** These parameters ensured that the simulation could adapt to different concurrency levels while maintaining controlled conditions.

The simulation followed a structured execution flow encompassing resource initialization, concurrent task management, load generation, and performance monitoring. The *runSimulation* method served as the main entry point for executing the booking simulation, accepting three parameters: the *eventide* for which bookings were to be simulated, the number of users attempting bookings, and the maximum number of tickets each user could attempt to book.

A fixed-thread pool managed task submission rates and included shutdown procedures, which prevented resource exhaustion and optimized resource utilization. Concurrent access was handled through MongoDB’s transaction capabilities, ensuring consistency during multiple booking attempts. Additionally, exception handling was employed to manage transaction rollbacks and verify system state consistency after failures, thereby preventing data inconsistencies and maintaining system stability.

#### Resource manager implementation

The *TicketDAO* class functioned as the primary resource manager, controlling the booking of available tickets. A point for highlighting is that this approach employed a resource locking strategy to maintain data integrity as suggested King (2024).

Each ticket’s state was updated individually, adhering to a consistent status update where transitions were uniformly applied across operations, maintaining consistency.

In relation to King's (2024) approach, a resource locking strategy was implemented with an optimistic concurrency control strategy to handle simultaneous transactions, in which the status-based availability was determined based on the ticket status, preventing overbooking.

.filter(Filters.and(

Filters.eq("event\_id", eventId),

Filters.eq("status", "available")

))

Transactions were encapsulated within *ClientSession* blocked to ensure atomicity during operations, such as ticket bookings. The implementation of the ticket booking functionality in the system incorporated an atomic approach to ensure reliable operations during high-concurrency scenarios. The method *bookAvailableTickets* encapsulated the logic for finding and updating ticket statuses within the system, and it primary task was to retrieve tickets matching specific criteria (those linked to a given event ID and marked as “available”). This retrieval was performed using MongoDB filters within a *datastore.find()* operation, ensuring that only eligible tickets were selected (Chooly, 2024). The method iteratively processed the required quantity of tickets, updating their status to “booked” and recording the purchase date before saving the changes to the database. By using *datastore.save()* for each ticket, the system ensured that updates were atomic, reducing the risk where multiple concurrent transactions might modify the same tickets.

# Results and discussion

## Performance Analysis

### Concurrency Test Results

### Transaction Processing Metrics

## Schema Analysis

|  |  |  |
| --- | --- | --- |
| **Aspect** | **MySQL (genres Table)** | **MongoDB (genres Collection)** |
| **Schema Structure** | Relational table with fixed columns and data types. | Document-based with JSON Schema validation. |
| **Primary Key** | genre\_id (INT, AUTO\_INCREMENT). | \_id (ObjectId, automatically generated). |
| **Unique Constraints** | genre\_name is unique via SQL constraint. | name is unique via MongoDB unique index. |
| **Data Integrity** | Enforced through foreign keys and constraints. | Enforced via JSON Schema and application-level validations. |
| **Flexibility** | Rigid schema; altering structure requires migrations. | Flexible schema; easy to add new fields without migration. |
| **Indexing** | Indexed through primary and unique keys. | Indexed using MongoDB's createIndex with uniqueness. |
| **Relationships** | Managed through foreign keys in related tables. | Managed via referencing (genre\_id as ObjectId). |
| **Query Complexity** | Simple queries for single table; joins for relations. | Single collection queries; aggregation needed for relations. |

## Data Structure Management

## Comparative analysis

### Transaction Management

A white paper with black text

Description automatically generated

# Transaction Management Implementation Analysis: MongoDB vs MySQL

## Core Transaction Components

### MongoDB Implementation

- Utilizes ClientSession for transaction boundaries

- Employs document-level locking through WiredTiger storage engine

- Implements atomic operations with multi-document ACID guarantees

- Manages transaction state through session-based protocol

### MySQL Implementation

- Uses JPA EntityTransaction for transaction management

- Implements row-level locking with PESSIMISTIC\_WRITE

- Relies on traditional RDBMS ACID properties

- Manages transaction state through explicit begin/commit cycles

## Concurrency Control Mechanisms

### MongoDB Approach

1. \*\*Session Management\*\*

- ClientSession ensures transaction isolation

- Atomic operations for status updates

- Optimistic concurrency control with version checks

2. \*\*Resource Locking\*\*

```java

Filters.and(

Filters.eq("event\_id", eventId),

Filters.eq("status", "available")

)

```

- Document-level filters for atomic operations

- Status-based availability checking

- Atomic updates for ticket status changes

### MySQL Approach

1. \*\*Lock Management\*\*

```java

.setLockMode(LockModeType.PESSIMISTIC\_WRITE)

.getSingleResult();

```

- Pessimistic locking strategy

- Row-level locking for concurrent access control

- Transaction isolation through explicit locks

## Implementation Differences

### Transaction Protocol

1. \*\*MongoDB\*\*

- Multi-phase commit with session scope

- Atomic operations within ClientSession

- Rollback through session abort

2. \*\*MySQL\*\*

- Traditional two-phase commit

- ACID compliance through JPA

- Automatic rollback on exception

### Error Handling

1. \*\*MongoDB\*\*

- Session-based error boundaries

- Explicit transaction abort

- Atomic counter tracking for metrics

2. \*\*MySQL\*\*

- Exception-based rollback

- JPA transaction management

- EntityTransaction state tracking

## Performance Considerations

### MongoDB Optimizations

- Document-level atomic operations

- Reduced lock contention through optimistic control

- Session pooling for connection management

### MySQL Optimizations

- Pessimistic locking for high-contention scenarios

- Index-based lock acquisition

- Connection pool management through JPA

## Implementation Trade-offs

### MongoDB

Advantages:

- Flexible document schema

- Atomic operations for simple updates

- Scalable distributed transactions

Limitations:

- Transaction size limitations

- Session timeout constraints

- Complex multi-document transactions

### MySQL

Advantages:

- Mature transaction management

- Strong ACID guarantees

- Predictable locking behavior

Limitations:

- Lock contention in high concurrency

- Schema rigidity

- Scalability constraints

### Schema Flexibility

### Development Experience

## Research Questions Assessment

## Implications

## Limitations of the research

# Conclusion

## Summary of findings

**MongoDB Advantages:**

1. **Atomic Operations**: MongoDB's document model allows for atomic operations within a single document
2. **Session-based Transactions**: Supports multi-document transactions with better performance characteristics
3. **Flexible Consistency Models**: Can adjust write concerns based on requirements

**MySQL Challenges:**

1. **Lock Contention**: Heavy use of pessimistic locking affects concurrency
2. **Transaction Overhead**: ACID compliance introduces additional overhead
3. **Schema Rigidity**: Fixed schema impacts transaction flexibility

**Schema Flexibility**

**MongoDB Advantages:**

1. **Dynamic Schema**: Can modify document structure without downtime
2. **Partial Updates**: Supports updating only specific documents
3. **No Lock Requirements**: Schema modifications don't block operations

**MySQL Challenges:**

1. **Table Locks**: Schema modifications require table-level locks
2. **Downtime Risk**: Modifications can block access to tables
3. **All-or-Nothing**: Changes affect all rows

**Architectural Implications**

1. **6.1 Data Modeling**

* MongoDB: Flexible document model allows for embedded documents and arrays
* MySQL: Normalized relational model requires multiple joins

1. **6.2 Scalability**

* MongoDB: Horizontal scaling through sharding
* MySQL: Vertical scaling with some horizontal capabilities

1. **6.3 Maintenance**

* MongoDB: Rolling updates and schema changes
* MySQL: Requires careful planning for schema modifications

**Recommendations**

**Use Case Specific Recommendations**

1. **High Concurrency Booking Systems**:
   * MongoDB advantages in document-level locking
   * Better handling of peak loads
2. **Schema Evolution**:
   * MongoDB's flexible schema better suited for rapid changes
   * Lower maintenance overhead
3. **Transaction Requirements**:
   * MySQL for strict ACID compliance
   * MongoDB for better performance with relaxed consistency

**Implementation Strategies**

1. **Hybrid Approach**:
   * Use MongoDB for ticket inventory and bookings
   * MySQL for user management and financial transactions
2. **Performance Optimization**:
   * Implement proper connection pooling
   * Use appropriate indexing strategies
   * Optimize batch operations

## Answers to Research questions

## Contributions

## Recommendations and future research

# References

Amghar, S., Cherdal, S. and Mouline, S. (2020) ‘Data Integration and NoSQL Systems: A State of the Art’, in *Proceedings of the 4th International Conference on Big Data and Internet of Things*. *International Conference on Big Data and Internet of Things*, New York, NY, USA: Association for Computing Machinery (BDIoT ’19), pp. 1–6. Available at: https://doi.org/10.1145/3372938.3372954.

Amghar, S., Cherdal, S. and Mouline, S. (2022) ‘Storing, preprocessing and analyzing tweets: finding the suitable noSQL system’, *International Journal of Computers and Applications*, 44(6), pp. 586–595. Available at: https://doi.org/10.1080/1206212X.2020.1846946.

Babic, T. (2019) *A Data Model for Online Concert Ticket Sales*, *Vertabelo Data Modeler*. Available at: https://vertabelo.com/blog/a-data-model-for-online-concert-ticket-sales/ (Accessed: 19 October 2024).

Bernstein, P.A. and Newcomer, E. (2009) ‘Chapter 1 - Introduction’, in P.A. Bernstein and E. Newcomer (eds) *Principles of Transaction Processing (Second Edition)*. San Francisco: Morgan Kaufmann (The Morgan Kaufmann Series in Data Management Systems), pp. 1–29. Available at: https://doi.org/10.1016/B978-1-55860-623-4.00001-9.

Brewer, E. (2012) ‘CAP twelve years later: How the “rules” have changed’, *Computer*, 45(2), pp. 23–29. Available at: https://doi.org/10.1109/MC.2012.37.

Candel, C.J.F., Sevilla Ruiz, D. and García-Molina, J.J. (2022) ‘A unified metamodel for NoSQL and relational databases’, *Information Systems*, 104, p. 101898. Available at: https://doi.org/10.1016/j.is.2021.101898.

Capris, T. *et al.* (2022) ‘Comparison of SQL and NoSQL databases with different workloads: MongoDB vs MySQL evaluation’, in *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*. *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, IEEE, pp. 214–218. Available at: https://doi.org/10.1109/ICDABI56818.2022.10041513.

Carvalho, I., Sá, F. and Bernardino, J. (2023) ‘Performance Evaluation of NoSQL Document Databases: Couchbase, CouchDB, and MongoDB’, *Algorithms*, 16(2), p. 78. Available at: https://doi.org/10.3390/a16020078.

Chang, M.-L.E. and Chua, H.N. (2019) ‘SQL and NoSQL Database Comparison’, in K. Arai, S. Kapoor, and R. Bhatia (eds) *Advances in Information and Communication Networks*. *FICC 2018*, Cham: Springer International Publishing, pp. 294–310. Available at: https://doi.org/10.1007/978-3-030-03402-3\_20.

Chooly, E. (2019) *Version 1.5 Migration Guide · Issue #1337 · MorphiaOrg/morphia*, *GitHub*. Available at: https://github.com/MorphiaOrg/morphia/issues/1337 (Accessed: 16 November 2024).

Chooly, E. (2024) ‘MorphiaOrg/morphia’. MorphiaOrg. Available at: https://github.com/MorphiaOrg/morphia (Accessed: 16 November 2024).

Cooper, B.F. *et al.* (2010) ‘Benchmarking cloud serving systems with YCSB’, in *Proceedings of the 1st ACM symposium on Cloud computing*. *1st ACM symposium on Cloud computing (SoCC ’10)*, New York, NY, USA: Association for Computing Machinery (SoCC ’10), pp. 143–154. Available at: https://doi.org/10.1145/1807128.1807152.

Diaz-Ordoñez, M., Rodríguez Baena, D.S. and Yun-Casalilla, B. (2023) ‘A new approach for the construction of historical databases—NoSQL Document-oriented databases: the example of AtlantoCracies’, *Digital Scholarship in the Humanities*, 38(3), pp. 1014–1032. Available at: https://doi.org/10.1093/llc/fqad033.

Digittrix, N. (2023) ‘The Battle of Databases: MySQL vs. MongoDB’, *Medium*, 18 September. Available at: https://medium.com/@nikita.digittrix/the-battle-of-databases-mysql-vs-mongodb-ca57164b8386 (Accessed: 22 September 2024).

Faraj, H.A. (2022) ‘Moving RDBMS to NoSQL Paradigms’, in *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*. *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA: IEEE, pp. 684–689. Available at: https://doi.org/10.1109/CSCI58124.2022.00125.

Fowler, M. (2002) *Patterns of Enterprise Application Architecture*. Wesley. Available at: https://dl.ebooksworld.ir/motoman/Patterns%20of%20Enterprise%20Application%20Architecture.pdf.

Gallinucci, E., Golfarelli, M. and Rizzi, S. (2018) ‘Schema profiling of document-oriented databases’, *Information Systems*, 75, pp. 13–25. Available at: https://doi.org/10.1016/j.is.2018.02.007.

Gray, J. and Lamport, L. (2006) ‘Consensus on transaction commit’, *ACM Trans. Database Syst.*, 31(1), pp. 133–160. Available at: https://doi.org/10.1145/1132863.1132867.

Győrödi, C. *et al.* (2015) ‘A comparative study: MongoDB vs. MySQL’, in *2015 13th International Conference on Engineering of Modern Electric Systems (EMES)*. *2015 13th International Conference on Engineering of Modern Electric Systems (EMES)*, pp. 1–6. Available at: https://doi.org/10.1109/EMES.2015.7158433.

ul Haque, A., Mahmood, T. and Ikram, N. (2019) ‘Performance Comparison of State of Art NoSql Technologies Using Apache Spark’, in K. Arai, S. Kapoor, and R. Bhatia (eds) *Intelligent Systems and Applications*. *Intelligent Systems and Applications. IntelliSys 2018*, Cham: Springer International Publishing, pp. 563–576. Available at: https://doi.org/10.1007/978-3-030-01057-7\_44.

Harizopoulos, S. *et al.* (2008) ‘OLTP through the looking glass, and what we found there’, in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. New York, NY, USA: Association for Computing Machinery (SIGMOD ’08), pp. 981–992. Available at: https://doi.org/10.1145/1376616.1376713.

Hecht, R. and Jablonski, S. (2011) ‘NoSQL evaluation: A use case oriented survey’, in *2011 International Conference on Cloud and Service Computing*. *2011 International Conference on Cloud and Service Computing*, Hong Kong, China: IEEE, pp. 336–341. Available at: https://doi.org/10.1109/CSC.2011.6138544.

Imam, A.A. *et al.* (2018) ‘Data Modeling Guidelines for NoSQL Document-Store Databases’, *International Journal of Advanced Computer Science and Applications (ijacsa)*, 9(10). Available at: https://doi.org/10.14569/IJACSA.2018.091066.

King, E. (2024) *Hello Interview | System Design in a Hurry*, *Hello Interview*. Available at: https://www.hellointerview.com/learn/system-design/answer-keys/ticketmaster (Accessed: 23 October 2024).

Kucharz, P. and Valero Sanchez, J. (2018) *Pessimistic Locking in JPA | Baeldung*. Available at: https://www.baeldung.com/jpa-pessimistic-locking (Accessed: 17 November 2024).

Kumar, C. (2019) *Introduction to Morphia - Java ODM for MongoDB | Baeldung*, *Baeldung*. Available at: https://www.baeldung.com/mongodb-morphia (Accessed: 16 November 2024).

Kuznetsov, S.D. and Poskonin, A.V. (2014) ‘NoSQL data management systems’, *Programming and Computer Software*, 40(6), pp. 323–332. Available at: https://doi.org/10.1134/S0361768814060152.

Li, C. (2010) ‘Transforming relational database into HBase: A case study’, in *2010 IEEE International Conference on Software Engineering and Service Sciences*. *2010 IEEE International Conference on Software Engineering and Service Sciences*, pp. 683–687. Available at: https://doi.org/10.1109/ICSESS.2010.5552465.

Lith, A. and Mattsson, J. (2010) *Investigating storage solutions for large data - A comparison of well performing and scalable data storage solutions for real time extraction and batch insertion of data*. CHALMERS UNIVERSITY OF TECHNOLOGY. Available at: https://www.semanticscholar.org/paper/Investigating-storage-solutions-for-large-data-A-of-Lith-Mattsson/b8f6b9d79c75e66c3b2f5034fe8172fd24cc0d13 (Accessed: 28 August 2024).

Mok, W.Y. (2021) ‘A Logical Database Design Methodology for MongoDB NoSQL Databases’, in *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore, Singapore: IEEE, pp. 1451–1455. Available at: https://doi.org/10.1109/IEEM50564.2021.9673004.

O’Grady, S. (2020) *MongoDB ACID Transactions Whitepaper*, *MongoDB*. Available at: https://www.mongodb.com/resources/products/capabilities/mongodb-multi-document-acid-transactions (Accessed: 16 November 2024).

*Oracle® Fusion Middleware Understanding Oracle TopLink* (2015). Available at: https://docs.oracle.com/middleware/1221/toplink/concepts/app\_dev.htm#OTLCG94288 (Accessed: 27 October 2024).

Patil, M.M. *et al.* (2017) ‘A qualitative analysis of the performance of MongoDB vs MySQL database based on insertion and retriewal operations using a web/android application to explore load balancing — Sharding in MongoDB and its advantages’, in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 325–330. Available at: https://doi.org/10.1109/I-SMAC.2017.8058365.

Pritchett, D. (2008) ‘BASE: An Acid Alternative: In partitioned databases, trading some consistency for availability can lead to dramatic improvements in scalability.’, *Queue*, 6(3), pp. 48–55. Available at: https://doi.org/10.1145/1394127.1394128.

Reinero, B. (2017) *Transitioning from Relational Databases to MongoDB - Data Models | MongoDB Blog*, *MongoDB*. Available at: https://www.mongodb.com/blog/post/transitioning-from-relational-databases-to-mongodb (Accessed: 22 September 2024).

Reisner, P. (1988) ‘Query Languages1’, in M. Helander (ed.) *Handbook of Human-Computer Interaction*. Amsterdam: North-Holland, pp. 257–280. Available at: https://doi.org/10.1016/B978-0-444-70536-5.50017-8.

Salander, A.-K. (2022) *Using atomic counters in the Enhanced DynamoDB AWS SDK for Java 2.x client | AWS Developer Tools Blog*, *AWS Amazon*. Available at: https://aws.amazon.com/blogs/developer/using-atomic-counters-in-the-enhanced-dynamodb-aws-sdk-for-java-2-x-client/ (Accessed: 17 November 2024).

Sen, P.S. and Mukherjee, N. (2024) ‘An ontology-based approach to designing a NoSQL database for semi-structured and unstructured health data’, *Cluster Computing*, 27(1), pp. 959–976. Available at: https://doi.org/10.1007/s10586-023-03995-y.

Seppälä, I. (2024) ‘Java Design Patterns’. Available at: https://github.com/iluwatar/java-design-patterns (Accessed: 17 November 2024).

Shareef, T.H., Sharif, K.H. and Rashid, B.N. (2022) ‘A Survey of Comparison Different Cloud Database Performance: SQL and NoSQL’, *Passer Journal of Basic and Applied Sciences*, 4(1), pp. 45–57. Available at: https://doi.org/10.24271/psr.2022.301247.1104.

Stonebraker, M. (2010) ‘SQL databases v. NoSQL databases’, *Commun. ACM*, 53(4), pp. 10–11. Available at: https://doi.org/10.1145/1721654.1721659.

Stonebraker, M. and Pavlo, A. (2024) ‘What Goes Around Comes Around... And Around...’, in *SIGMOD*. *2024 ACM SIGMOD/PODS Conference*, Santiago, Chile: ACM (2), pp. 21–37. Available at: https://doi.org/10.1145/3685980.3685984.

Sudiartha, I.K.G. *et al.* (2020) ‘Data Structure Comparison Between MySql Relational Database and Firebase Database NoSql on Mobile Based Tourist Tracking Application’, *Journal of Physics: Conference Series*, 1569(3), p. 032092. Available at: https://doi.org/10.1088/1742-6596/1569/3/032092.

Thakare, A. *et al.* (2023) ‘NoSQL Databases: Modern Data Systems for Big Data Analytics - Features, Categorization and Comparison’, *International Journal of Electrical and Computer Engineering Systems*, 14(2), pp. 207–216. Available at: https://doi.org/10.32985/ijeces.14.2.10.

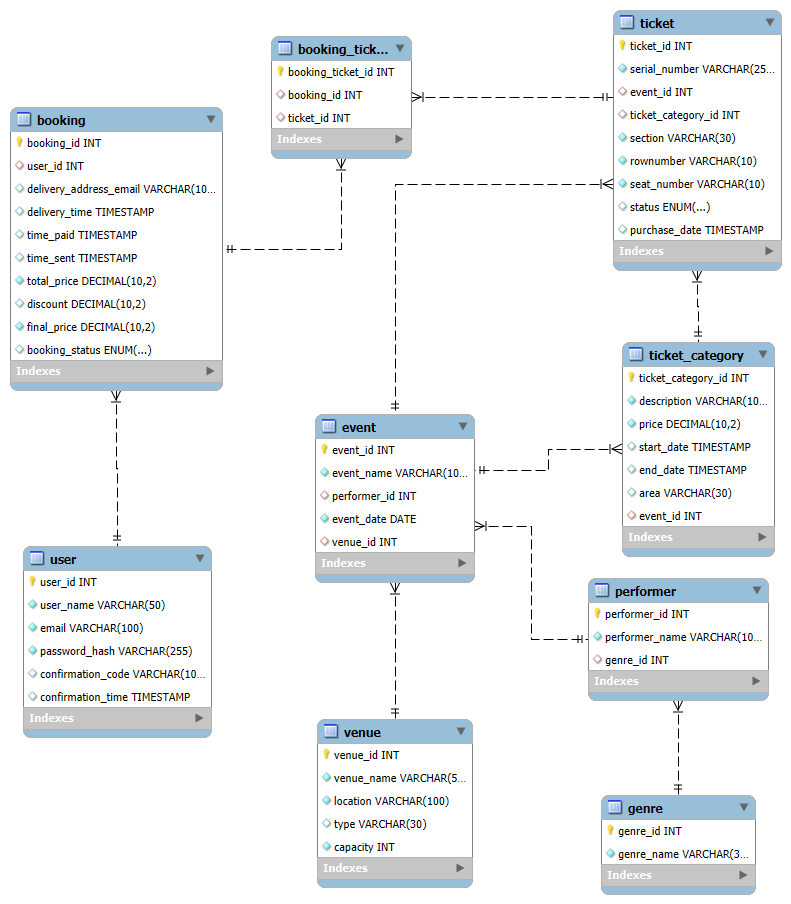
Wodyk, R. and Skublewska-Paszkowska, M. (2020) ‘Performance comparison of relational databases SQL Server, MySQL and PostgreSQL using a web application and the Laravel framework’, *Journal of Computer Sciences Institute*, 17, pp. 358–364. Available at: https://doi.org/10.35784/jcsi.2279.

Yedilkhan, D. *et al.* (2023) ‘Performance Analysis of Scaling NoSQL vs SQL: A Comparative Study of MongoDB, Cassandra, and PostgreSQL’, in *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*. *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, pp. 479–483. Available at: https://doi.org/10.1109/SIST58284.2023.10223568.

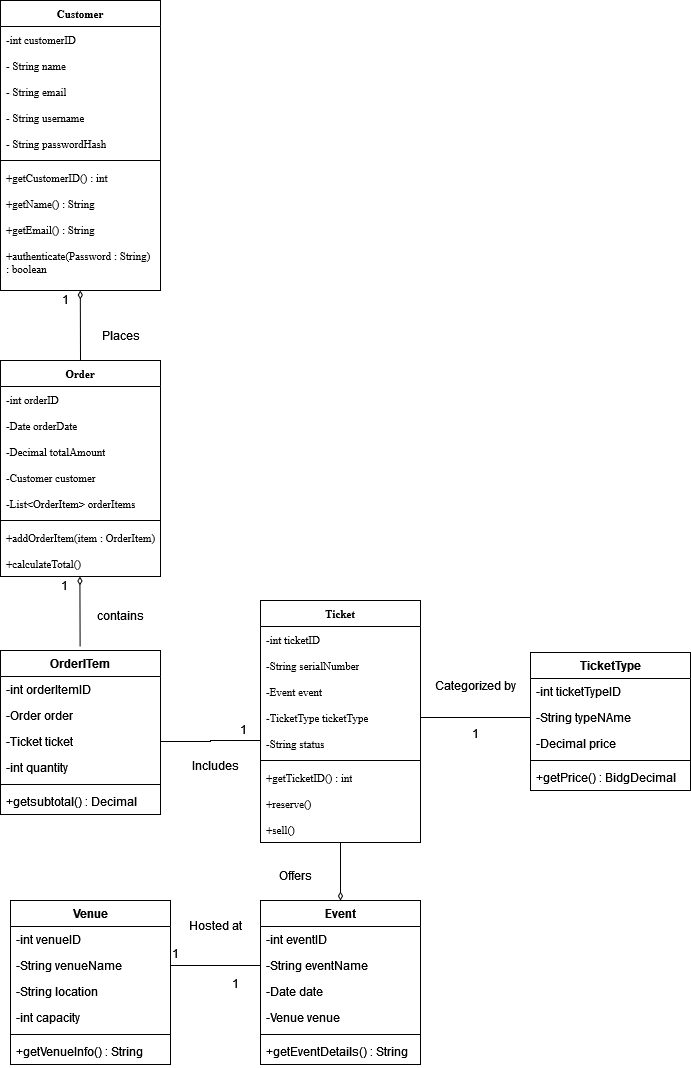
# Appendices

## MySQL Schema Implementation

## Entity-Relationship Diagram



## MySQL Class Diagram



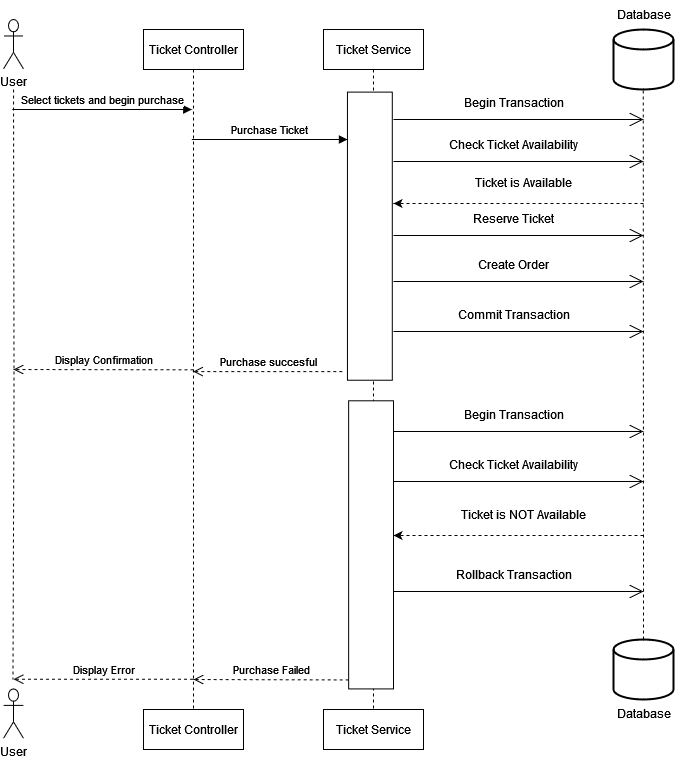
## Document Structure Diagram for MongoDB

## Ticket Purchase process diagram

A diagram of a customer service

Description automatically generated

## Sequence Diagram for Ticket Purchase Transaction



## Backend application diagram

## Deployment with technological stack

## Component Diagram

Depicts the high-level components of the system and their interactions

## MySQL MoSCOW

A table of informational text

Description automatically generated with medium confidence

## MongoDB MoSCOW

A table of informational text

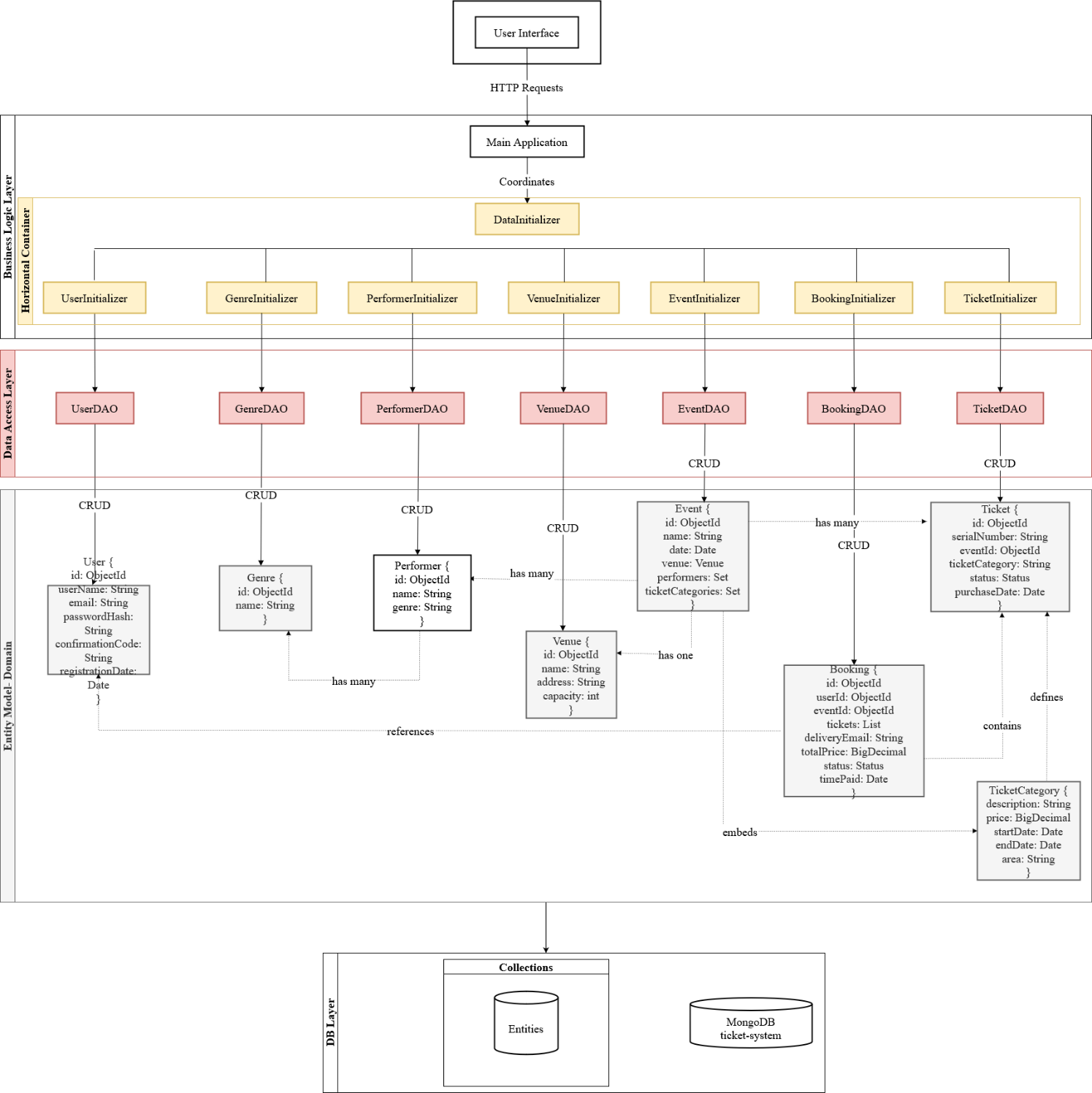
Description automatically generated with medium confidence

## Code samples

## Test results

## Configuration files

## Architecture Diagram for Ticketing System using MongoDB



## Transaction Flow Protocol MongoDB

A close-up of a diagram

Description automatically generated

## Transaction Flow Protocol MongoDB

public class BookingSimulation {

private final BookingService bookingService;

private final UserDAO userDAO;

private final EventDAO eventDAO;

private final TicketDAO ticketDAO;

private final BookingDAO bookingDAO;

private final Datastore datastore;

/\*\*

\* Runs the booking simulation.

\*/

public void runSimulation(ObjectId eventId, int numUsers, int maxTicketsPerUser) {

ExecutorService executor = Executors.newFixedThreadPool(1); // Adjust thread pool size as needed

List<Callable<Boolean>> tasks = new ArrayList<>();

// Retrieve all users to simulate booking attempts

List<User> users = userDAO.findAll();

// If not enough users

if (users.size() < numUsers) {

System.out.println("Not enough users in the system. Please add more users for the simulation.");

executor.shutdown();

return;

}

Random random = new Random();

for (int i = 0; i < numUsers; i++) {

final User user = users.get(random.nextInt(users.size()));

final int ticketsToBook = random.nextInt(maxTicketsPerUser) + 1; // 1 to maxTicketsPerUser

Callable<Boolean> task = () -> {

return bookingService.bookTickets(user.getId(), eventId, ticketsToBook);

};

tasks.add(task);

}

try {

List<Future<Boolean>> results = executor.invokeAll(tasks);

// Wait for all tasks to complete

executor.shutdown();

executor.awaitTermination(1, TimeUnit.MILLISECONDS);

// Output results

System.out.println("Simulation completed.");

System.out.println("Total booking attempts: " + numUsers);

System.out.println("Successful bookings: " + bookingService.getSuccessfulBookings());

System.out.println("Failed bookings: " + bookingService.getFailedBookings());

// Verify no overselling

long totalBookedTickets = bookingDAO.findAll().stream()

.filter(booking -> booking.getEventId().equals(eventId))

.mapToInt(booking -> booking.getTickets().size())

.sum();

long totalTickets = ticketDAO.countAvailableTickets(eventId) + totalBookedTickets;

System.out.println("Total tickets available before booking: " + totalTickets);

System.out.println("Total tickets booked: " + totalBookedTickets);

System.out.println("Tickets remaining: " + ticketDAO.countAvailableTickets(eventId));

} catch (InterruptedException e) {

System.err.println("Simulation interrupted: " + e.getMessage());

e.printStackTrace();

}

}

}