

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/378440740>

Canva's NoSQL Database : MongoDB's Replication and Sharding in Content Management

Article · February 2024

CITATION

1

READS

140

1 author:



Rania Adel

2 PUBLICATIONS 1 CITATION

SEE PROFILE

Canva's NoSQL Database : MongoDB's Replication and Sharding in Content Management.

By: Rania Adel Mohamed

In the fast-paced world of online content creation where user experience is paramount, Canva, a Sydney-based company, has emerged as a global leader by delivering exceptional design services to over 10 billion designs and processing an impressive 30,000 document requests per second. The key to their success lies in the efficient management and retrieval of vast amounts of data. To achieve this, Canva uses MongoDB, a NoSQL database because of its scalability, flexibility, and reliability. MongoDB stores large volumes of big data in JSON documents, MongoDB also supports Sharding and Replication distribution models which are techniques to distribute data across multiple servers (machines) for horizontal scaling and availability. In this report, we analyze how Canva uses MongoDB's data model and distribution model (replication and sharding) to handle its content management needs, and what are the benefits and challenges of using MongoDB for this use case. The report is structured as follows: Section 1, shows an overview about Canva and its content management, Section 2, shows the adopted NoSQL data model (MongoDB) and its features, Section 3, shows the distribution model (sharding and Replication), section 4, shows The operational challenges, Section 5 shows other aspects that may highlight the success of Content management in Canva.

1- Overview

The process of producing, arranging, storing, and distributing various kinds of content, including text, photos, videos, audio, and more, is called content management. Businesses may use content management to accomplish a range of

objectives, including marketing, branding, consumer interaction, education, and entertainment. The quality and consistency of the content , as well as the user experience and happiness, may all be improved with content management. Content management is essential for online platforms like Canva, which enable users to design and share their own visual content. Canva's content management use case involves the following business and technical requirements:

Business requirements:

- Provide a user-friendly and intuitive interface for users to create and edit their own content.
- Offer a wide range of templates, elements, fonts, and photos for users to choose from and customize
- Support different types of content, such as logos, flyers, posters, videos, and more
- Enable users to save, export, and share their content in various formats and sizes
- Ensure the quality, security, and performance of the content and the platform
- Scale up the platform to handle the growing demand and traffic

Technical requirements:

- Use a flexible and scalable database system to store and manage the content and the user data .
- Implement a caching and distribution system to deliver the content faster and reduce the server load .
- Optimize the content for SEO and social media .
- Integrate with third-party services and APIs, such as Google Drive, Dropbox, Facebook, and Instagram .
- Monitor and analyze the content usage and performance metrics.
- Test and update the platform regularly to fix bugs and improve features.

Using NoSQL MongoDB, which provides more performance, scalability, and flexibility than relational databases, was one of Canva's primary technological considerations for its content management use case.

2- The adopted NoSQL data model

MongoDB is an open source document oriented database , schema-less NoSQL database allowing flexibility in data representation . It stores data in documents in a type of JSON format called BSON(Binary encoded JSON format) , this surpasses JSON in speed . Documents are grouped together in collections. It is not necessary for all of the documents in a collection to have precisely the same

set of fields. We refer to this as a "flexible schema.". This facilitates faster creation and iteration as well as more flexibility and agility in data modeling. The documents are created using the fields. Fields are key-value pairs in the documents. MongoDB allows developers to store nested documents within a document , this allows to create complex relationships between data.

Key characteristics Of MongoDB :

Dynamic Schema

Dynamic schema in MongoDB means that can add or remove fields from documents in database without needing predefined fixed structure ,This adaptability facilitates seamless adjustments to evolving data needs without disrupting the application.

Scalability

MongoDB is designed for horizontal scalability , it can scale out by sharding that distributing data across multiple servers to handle large amount of data.

Replication

MongoDB supports replication through replica sets providing data redundancy and high availability by maintaining multiple copies of data across different servers , Automatic failover ensures continuous operation in case of node failures.

Indexing

MongoDB supports the creation of indexes on fields which can improve query performance .

Query Language

MongoDB provides language with various operators and supports complex queries , It allows for filtering, sorting, and projecting data in a flexible manner.

Geospatial Capabilities:

MongoDB includes built-in support for geospatial data and allows for efficient storage and querying of location-based information. This is particularly useful for applications with geospatial requirements.

Security Features:

MongoDB offers various security features, including authentication, authorization, and encryption. These features help protect data integrity and restrict unauthorized access.

MongoDB is a popular database for content management systems because it stores and deliver any kind of content including social media , text , images , videos and audios MongoDB's flexible schema and data model make it easy to catalog multiple content types with diverse metadata, and to add new features and attributes without impacting performance or availability. MongoDB also scales automatically and transparently, allowing you to handle large volumes of content and global audiences .So , how Canva fits MongoDB in its content management . Canva's ever more complex designs and documents types have been a natural match for MongoDB 's flexible document model from its inception, A major factor of Canva's success is the flexibility of MongoDB provided and all new features are added to Canva platform handled by MongoDB. The Document Service, supported by MongoDB Atlas, allows users to open, create, edit, and remove any Canva design. Each and every user depends on the Document Service; without it, Canva users are unable to design. By using MongoDB document architecture Canva can store complex and diverse data in a simple manner without having worry about data standards or schema upgrades , MongoDB's robust operators and indexes for document manipulation further simplify data querying and updating.

3- The distribution model (replication and sharding)

The term distribution models describe how data is split up among various nodes inside a database system , they affect performance , availability and scalability of the database system. To manage the enormous scale and high availability of canva 's content management , Replication and Sharding are two distribution models that MongoDb provides to improve database speed and reliability and these distribution models also used for horizontal scaling. Both of two models use the concepts node(single server that stores data) and cluster (collection of multiple nodes).

- **Replication**

is the process of creating and maintaining multiple copies of the same data across different servers, or nodes., each node is a server that is physically separated , as a result the system is highly available and fault tolerant , Additionally, client requests may be routed and dispersed among nodes once they are all operational, preventing overloading on a single node. As a result, the database can process more queries simultaneously. MongoDB uses a replication model , which consist of replica sets , only one primary node and several secondary nodes in a cluster . The primary node or primary replica is responsible for receiving and applying write operations , while the secondary nodes or secondary replica replicate the data from the primary node and serve read operations . users can read from any of the nodes . If the primary node

fails, one of the secondary nodes can take over as the new primary node, ensuring fault tolerance and high availability. If the primary server fails, a secondary will hold the new primary. By using replication Canva can ensure security and quality of content and provide recovery and backup options.

- **Sharding**

Data distribution and load balancing are achieved through the technique of sharding which divides data into smaller chunks among several nodes called Shards based on shard key since several nodes may handle read and write requests concurrently, sharding further enhances the database system's scalability and speed. . Canva can add or remove shards as needed, depending on the data volume and workload. Canva can also balance the data distribution and load across the shards, using MongoDB Atlas, which is a cloud service that manages MongoDB clusters for them.

4- The rationale of adopting the chosen NoSQL data model and distribution model

The main reason why Canva chose MongoDB as its NoSQL database is because of its flexibility and scalability. MongoDB's document model and distribution models allow Canva to store and process its data in a way that suits its content management needs, without compromising on performance or reliability.

Some of the benefits of using MongoDB for content management are:

- Canva is able to store many types of content with different attributes and metadata because of MongoDB flexible schema design and document data format, all without the need for complex joins and schema migrations.
- Canva can conduct a variety of operations on its content, including full-text search, geographic searches, aggregations, and analytics, because of MongoDB's extensive query language and indexing capabilities.
- Canva can extend its database horizontally and provide high availability without sacrificing performance or consistency because of MongoDB's replication and sharding. This allows Canva to adapt to the changing needs and growth of its content, as well as optimize its resource utilization and cost efficiency.
- Using MongoDB Atlas the fully managed cloud service for MongoDB Canva can install and maintain its database with ease and security because of MongoDB's cloud-native and developer-friendly capabilities.

5- The operational challenges

Despite the advantages of using MongoDB for content management, Canva also faces some operational challenges that may backfire the application of NoSQL technology. Some of these challenges are:

Data consistency: MongoDB follows a consistency model called eventual consistency, which means that the data may not be immediately consistent across all the nodes or shards, but will eventually converge to the same state. This may cause some issues for Canva, such as stale reads, write conflicts, or data loss, especially in scenarios where strong consistency is required, such as transactions or real-time collaboration.

Data modeling: MongoDB's document model and schema flexibility also pose some challenges for Canva, such as choosing the right data structure, sharding key, and index strategy, as well as avoiding data duplication, fragmentation, or corruption. These decisions have significant impacts on the performance, scalability, and maintainability of the database system, and require careful planning and testing.

Data security: MongoDB's distribution models also introduce some risks for Canva, such as data breaches, unauthorized access, or malicious attacks, especially in a cloud environment. Canva needs to ensure that its data is protected and encrypted, both in transit and at rest, as well as implement proper authentication, authorization, and auditing mechanisms, to prevent any data leakage or compromise.

Data migration: MongoDB's replication and sharding models also entail some challenges for Canva, such as data migration, synchronization, and backup, especially when scaling or upgrading the database system. Canva needs to ensure that its data is transferred and replicated correctly and efficiently, without causing any downtime, data loss, or inconsistency.

6- Other aspects that may highlight the success of content management

Despite the challenges of using MongoDB for content management, Canva has also achieved some success factors that demonstrate the potential and value of NoSQL technology. Some of these factors are:

User satisfaction: Canva has attracted and retained millions of users around the world, who use its platform to create and share their designs and presentations. Canva has received positive feedback and reviews from its users, who appreciate its ease of use, functionality, and quality of its content. Canva has also won several awards and recognition for its innovation and excellence in content creation and management.

Business growth: Canva has also grown and expanded its business, both in terms of revenue and market share. Canva has generated millions of dollars in revenue, and has become one of the most valuable startups in the world. Canva has also diversified and extended its product offerings, such as Canva Pro, Canva for Education, Canva for Enterprise, and Canva for Nonprofits, to cater to different segments and needs of its customers.

Technical excellence: Canva has also demonstrated its technical excellence and leadership, both in terms of its database system and its content management system. Canva has successfully scaled and optimized its MongoDB clusters, using MongoDB Atlas and other tools and techniques, to handle its massive and dynamic data. Canva has also developed and deployed its content management system, using various technologies and frameworks, such as React, TypeScript, GraphQL, AWS, and Kubernetes, to deliver its content to its users.

7- Conclusion

In conclusion, Canva's strategic adoption of MongoDB as its NoSQL database has been instrumental in positioning the company as a global leader in online content management. MongoDB's dynamic schema, scalability, and support for distribution models like replication and sharding have allowed Canva to efficiently manage and retrieve vast amounts of data, catering to over 10 billion designs and processing 30,000 document requests per second. The symbiosis between Canva and MongoDB is evident in the platform's ability to address both business and technical requirements, offering a user-friendly interface, diverse content options, and scalability to meet growing demand.

The flexibility of MongoDB's document model has empowered Canva to store diverse content types with varying attributes, providing users with a rich and customizable experience. Replication ensures high availability and data redundancy, contributing to the security and quality of Canva's content. Sharding further enhances scalability, allowing Canva to adapt to changing needs efficiently.

