

ACCA7004&5 - Report - 2222148

by Rania Abdalla

Submission date: 05-May-2023 08:32PM (UTC+0100)

Submission ID: 204985529

File name: 42277_Rania_Abdalla_ACCA70045_-_Report_-_2222148_488469_858699124.pdf (613.63K)

Word count: 987

Character count: 5860



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

University of Wales Trinity Saint David

DVLA: MSc Software Engineering (Web Development)

05/05/2023

Assignment 2

ACCA7004 & ACCA7005

Web Development & Web Services

By Rania Abdalla - 2222148

Table of Contents

<i>Introduction</i>	<i>3</i>
<i>Big Data</i>	<i>3</i>
<i>Document-based Databases and Big Data</i>	<i>4</i>
<i>Conclusion</i>	<i>5</i>
<i>References.....</i>	<i>6</i>
<i>Self-References</i>	<i>6</i>

Introduction

Traditional relational databases may struggle to handle the complexity of big data which led to the development of alternative database solutions, including document-based databases. This report will evaluate the document-based NoSQL database in handling big data which is characterised by its volume, velocity, and variety.

Big Data

Data is at the heart of every aspect of life, and in the world of the internet it is all about data being transmitted from one end to another. Traditionally data has been structured and stored in relational databases. Nowadays, with the rise in technology and web applications the amount of data has increased exponentially [1]. Not only has it increased in amount but also in the types of data. For example, structured data in the form of tables with rows and columns in a relational database might be used to store student information at a university with their personal details, course details, results etc. using foreign key relationships to eliminate redundant data. Each student in the university will have similar details to be stored making relational databases very efficient [2]. Or unstructured data if we think about a social media platform such as Facebook. Profiles on Facebook can have a variety of information about each user. Facebook has over two billion users according to Oberlo [6], so the amount of data is enormous, also not everyone is going to submit the exact same profile information and what about the types of data we would not only have user information but also data such as images, videos, posts which are types of unstructured data that need to be stored [2].

This large amount of unstructured and structured data is referred to as big data. Big data is characterised by the three Vs: Volume, Velocity, and Variety [2]. The volume refers to the amount of data that is produced every nanosecond [1], the velocity refers to the speed at which data is generated and collected, and variety refers to the types of data [2].

Relational databases are not well-designed to handle big data. If we refer to the Facebook example above, not every profile would provide the same information which means there is different details to be stored for each profile. If we used a relational database, we would end up with a lot of empty values in our tables. Moreover, the relationships that are created in relational databases would very quickly become complicated potentially making querying the database extremely slow and inefficient.

This is why NoSQL databases are becoming increasingly popular. NoSQL databases are very flexible as they allow for the storage of large amounts of structured, unstructured, and semi-structured data [2]. There are different types of NoSQL databases such as key-value, document, graph, and column-oriented databases.

Document-based Databases and Big Data

Document-based databases are a type of NoSQL database that use documents and collections to store data. A document is a record in a database that stores data in key-value pairs, typically in JSON format as demonstrated in figure 1 and a collection is a group of documents with similar content as seen on figure 2 [3] [4].

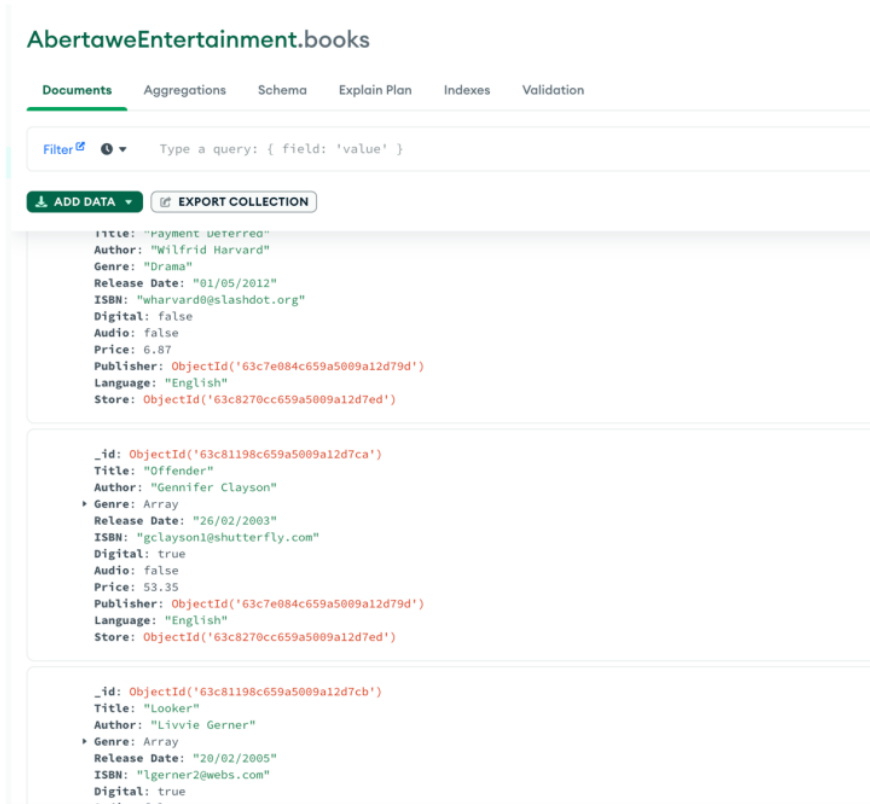


Figure1: Example of a document-based database document with key-value pairs data [7]

Collections

Create collection View   Sort by Collection Name

books	Storage size: 20.48 kB	Documents: 15	Avg. document size: 280.00 B	Indexes: 1	Total index size: 36.86 kB
Uncompressed data size: 1.08 kB					
	Storage size: 20.48 kB	Documents: 7	Avg. document size: 154.00 B	Indexes: 1	Total index size: 36.86 kB
publisher	Storage size: 20.48 kB	Documents: 5	Avg. document size: 83.00 B	Indexes: 1	Total index size: 20.48 kB
store	Storage size: 20.48 kB	Documents: 5	Avg. document size: 126.00 B	Indexes: 1	Total index size: 20.48 kB
users	Storage size: 20.48 kB	Documents: 21	Avg. document size: 181.00 B	Indexes: 1	Total index size: 36.86 kB

Figure 2: Example of a document-based database collection [7]

One of the key benefits of document-based databases is their flexibility in handling unstructured and semi-structured data. They are particularly well-suited for storing data that is constantly changing or has a complex structure, such as social media feeds, or product catalogues [3]. Document-based databases can also easily be scaled horizontally and vertically to handle large volumes of data [5].

Document-based databases have some drawbacks for example, they do not support Atomicity, Consistency, Isolation, and Durability (ACID) properties in the same way as relational databases as they prioritise flexibility and performance over strict transactions. However, some document-based databases like MongoDB support multi-document ACID transactions [3].

Conclusion

In conclusion, documents-based NoSQL databases are a good fit for handling big data, particularly in the use cases where the data is structured or unstructured. They offer flexibility and scalability which are important features when working with large volumes of data.

References

- [1] S. Kalid, A. Syed, A. Mohammad and M. N. Halgamuge, "Big-data NoSQL databases: A comparison and analysis of "Big-Table", "DynamoDB", and "Cassandra"," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 89-93, doi: 10.1109/ICBDA.2017.8078782.
- [2] F. Hamami, I. A. Dahlan, S. W. Prakosa and K. F. Somantri, "Big Data Analytics for Processing Real-time Unstructured Data from CCTV in Traffic Management," 2020 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 2020, pp. 1-5, doi: 10.1109/ICoDSA50139.2020.9212858.
- [3] <https://www.mongodb.com/document-databases>
- [4] L. Cazzanti, L. M. Millefiori and G. Arcieri, "A document-based data model for large scale computational maritime situational awareness," 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 2015, pp. 1350-1356, doi: 10.1109/BigData.2015.7363894.
- [5] A. K. Samanta, B. B. Sarkar and N. Chaki, "Query Performance Analysis of NoSQL and Big Data," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 2018, pp. 237-241, doi: 10.1109/ICRCICN.2018.8718712.
- [6] Oberlo – HOW MANY USERS DOES FACEBOOK HAVE? - Q1 2023 – Source: Meta <https://www.oberlo.com/statistics/how-many-users-does-facebook-have>

Self-References

- [7] Figure 1 & 2: From a MongoDB database I created for an assignment for the database module last semester.

FINAL GRADE

57 / 100

GENERAL COMMENTS

Instructor

Implementation

REST (30): 16.5

AJAX (20): 11.5

Code/Doc (10): 6.5

Features (20): 5

Report (20): 16.5

Feedback:

=====

You could have provided a little more detail of a specific NoSQL database such as MongoDB to illustrate your answer. But you have covered the main benefits and features quite well.

The basic concepts are understood, the implementation is a bit limited and there are a number of features that could have provided. Some user input validation. I wanted you to expand upon the concepts, such as use a new library.

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5
