Graph-Based Database Management System Use Cases for Retail Titans - Based on Walmart and eBay Utilizing Neo4j

Yi-Hsueh Yang
Heinz College, Carnegie Mellon University
Master of Science in Public Policy and Management – Data Analytics
Pittsburgh, PA, United States
yihsuehy@andrew.cmu.edu

***Abstract*—With database management systems appearing in succession in response to the exponential growth in data and information, the level of compatibility between the database and customers' pain points greatly speaks for the performance of companies. NoSQL database systems are bringing revolutionary changes into the data storage world, with key-value pair, document-base, column-family-based, and graph-based these four types of storage methods, many business problems are gradually conquered by utilizing them. This research report aims to break down the process and upshot for a solution in each of the companies after transforming from a relational database to Neo4j. The ultimate goal is to bring out a thorough discussion and analysis of how this new technology works for these companies and whether there may be some potential hazards that existed for them.**

Keywords—***Walmart, eBay, Neo4j, NoSQL, relational database***

I. INTRODUCTION

This research concentrates on the use cases of two retail titans - Walmart Inc. and eBay Inc., referred to hereinafter as "Walmart" and "eBay", in an identical NoSQL database management system - Neo4j in detail. The background, the methodology, the effectiveness of using, and the comparison of these two use cases on one database management tool are discussed. The purpose of this research is to dive into the methodologies and the pros and cons of these two use cases and provide a general overview of how the non-traditional approach help solves problem evolved in recent year. Moreover, meanwhile looking at the outstanding parts of the technology, and how the breakthrough dominates the current e-commerce market, I also want to evaluate whether the downsides of the technique are going to be viewed only as a temporary solution as if the market and people's demand continue to grow. In short, I'm trying to understand whether the solution of using Neo4j as their backend database entirely solves the problem of growing users without any concern in the future. Ease of Use

II. TYPE OF BUSINESS

A. Walmart

Walmart is an internationally-known retail company that was established in 1962. As a family-owned business, Walmart has become the largest private company with 2.2 million employees, owning more than 10k stores worldwide and collecting about 36 million dollars from its US stores every day. Besides operating a chain of supercenters, Walmart has an enormous e-commerce service that facilitates online ordering and shipping. There are currently 10 websites operated by Walmart to service online retailing to its customers. For handling emerging digital customers, Walmart has built a data center called Area 71 in Jane, Missouri. The data center aids the retail titan to store terabytes of data and conduct big data analysis for its products and customers.

B. eBay

Unlike Walmart, eBay is purely a company that focuses one hundred percent on e-commerce, it has neither supercenters nor physical stores, it performs as a marketplace allowing sellers and buyers to trade. Originating in 1995, in San Jose, California, eBay gained its ground after the doc.com foam, it is now to be said one of the five largest e-commerce platforms in the world. The business model at eBay is mainly charging selling fees from the sellers that are putting up goods to sell. There are two main parts of the fee, one charge whenever you start your business on eBay and list over 250 items a month, and another charge 10% to 15% of the final whenever a trade happened. For increasing revenue, eBay tries its best to deliver goods to their customer within 3 days to meet customers' needs. However, challenges coming up with more and more e-commerce platforms are emerging, a way to stay competitive is to reduce the delivery time from seller to buyer, that is the time when eBay is constantly improving their delivery, which results in options like 2-day shipping and 1-day shipping appears.

III. Data use and previous Storage Method

A. Walmart

Making the website sit still passively isn't a huge problem, but it is not ideal to expect a big leap in revenue. Therefore, Walmart rapidly introduced the recommendation system on its website, trying to increase its selling performance by recommending products and stimulating people to buy more while browsing its website. Initially, Walmart utilizes the batch process real-time online recommendation method which is a way to recommend products by throwing batches of data into the recommendation calculation model to generate products with a higher potential of being chosen. The batch method had its impact, however, accompanied by the increasing amount of customer visits and requests sent, the drawbacks gradually emerge: the system is experiencing more and more latency, and some manpower has to be assigned to do the analysis work.

For understanding the cause of the problem, we can dive deeper into the batch processing method. The concept of the batch processing method works in two ways, one is it processes at a given time, and another is it processes when the amount of data has reached its threshold. The data are stored in a conventional relational database which data analysts or data scientists extracted data from whenever the premise mentioned above is met. This leaves a huge risk to the system whenever the visits to the website are huge of a sudden, the webpage might not be able to react and provide a high-accuracy recommendation in real-time. The failure of the system is expected to be more serious if the website is constantly meeting its maximum capacity of visits, and the company might start to lose its client and resulting in a loss of sales and revenue. Therefore, to prevent the potential hazard, Walmart shifted its focus to the NoSQL database management system, referred to hereinafter as "DBMS", Neo4j in particular, to transform the way they had their data stored and improve the recommendation system.

B. eBay

Shutl, a UK-based company was acquired by eBay and renamed "eBay Now", facilitating the shipping procedure for eBay in the UK initially but expanding its service territory to the US in 2014. Shutl acting like a coordinator between customers, couriers, and suppliers. They use data like product information, inventory information of suppliers, couriers' availability, and address from both ends to come up with a better delivery strategy and timing. Promising customers get their goods in time, ensuring a sufficient amount of couriers, and keeping orders running for suppliers to sell more, what Shutl is doing, is the most important work in the system, therefore, optimization is considerably important. However, all the data are stored in MySQL DBMS initially. The relational database can handle it at first but not until when the number of requests greatly increased, the problem of the relational database was then unveiled. Joins are too time-consuming to operate, especially when the query grows larger and larger. Volker Pacher, a senior developer at eBay, even stated that the slower query they have, about 15 minutes, is longer than the fastest delivery they offer. These reasons are propelling eBay to find a new way to restructure their data and they turn to the graph-based DBMS - Neo4j.

IV. Use Cases

Neo4j is a graph-based NoSQL DBMS where there are two most important features: node and edge. Nodes are entities that can hold key-value pairs and be tagged with labels while edges are usually called relationships which provide a clear connection between two existing nodes. With these two features, Neo4j is known to be the solution to handling complex connection problems. Companies value three characteristics when choosing DBMS: intuitiveness, speed, and agility.

A. Walmart

The design itself for Neo4j is straightforward with the term of node and edge(relationship), it helps the user easily understand the relationship within the linked list of entities. The design itself also avoids excessive joins than relational databases to get relationships between 2 entities. Other kinds of NoSQL databases can also greatly lessen the number of joins in the query, however, those have to go through the searching process of using indexes which inherently makes Neo4j a better solution in acquiring relationships. As with the other NoSql DBMSs, the ability to alter data dynamically is never a problem, the effect towards it is substantially smaller than relational DBMS. Neo4j can even remain its status available for computing while altering the graph which makes it extra efficient in proving real-time calculation results.

Walmart uses Neo4j to greatly reduce the processing time and make real-time recommendations smoothly while eBay uses it to perform more efficiently in its logistics and implement same-day delivery to its customers. The detailed method of how Neo4j helps them would be organized for the final report. The transition of data, how those are important, what the query language Cypher looks like, and why it helps a lot in this aspect will be covered. The trade-off of how and why Walmart is choosing availability over consistency while eBay does the opposite to value consistency more over availability is going to be deeply dived into.

B. eBay

V. Comparison

A. Similarties

The bullet points are the main points that will be discussed in this comparison session, they will be revised and completed in the final report.

- Both are transformed due to facing a high quantity of growth and requests coming up.
- Neo4j are both helping them to speed up with queries and avoid joins by using Cypher to substitute traditional SQL.
- Attributes of data do not change often for retail industry, only adding and deleting is encountered, so Neo4j can suit their needs best.

B. *Differences*

- usage: recommendation system vs same-day delivery system
- Relationships are utilizing in different ways: feature relationship between products v.s. Resources relationship among sellers, customers and couriers
- CAP theorem: AP vs CP

C. *Drawbacks*

- No protection - no data encryption
- Not sharable - risky, all data on one database server
- Upper limit of stored nodes - tens of billions
- *Personal thought: the structure of the recommendation system built today always uses one's previous activities as a reference, however in this way, people don't get the chance to learn a new concept or know new things unless they try to search for them on their own so will the system recommend that information. It is hard for people to step out of their comfort zone or be educated passively by getting the result from the recommendation. It does no good for a person's well-being and even harms their curiosity.*

D. *Argument*

The argument is going to be provided here to state that other e-retailers should use Neo4j, a graph-based database to manage the network, although there are some major drawbacks as mentioned above, it is still likely that it is a decent solution if those drawbacks are handled wisely. It can bring incredible efficiency to companies' performance and enhance their service to the customers.

VI. Conclusion

References

1. https://www.projectpro.io/article/how-big-data-analysis-helped-increase-walmarts-sales-turnover/109#toc-1
2. https://commons.erau.edu/cgi/viewcontent.cgi?article=1401&context=adfsl
3. https://www.ecommercenext.org/kb/what-is-ebay/
4. https://go.neo4j.com/rs/710-RRC-335/images/wp_sca_neo4j.pdf
5. https://vlomni.com/real-time-vs-batch-data-integration/
6. https://www.bmc.com/blogs/batch-processing-stream-processing-real-time/
7. https://neo4j.com/blog/walmart-neo4j-competitive-advantage/#:~:text=Walmart%20is%20now%20using%20Neo4j,IT%20team%20based%20in%20Brazil.
8. https://ieeexplore.ieee.org/abstract/document/8301622
9. https://dist.neo4j.com/wp-content/uploads/Neo4j_CS_eBay.pdf
10. https://scholarworks.calstate.edu/downloads/fn106z26t
11. https://neo4j.com/docs/getting-started/current/get-started-with-neo4j/graph-database/
12. https://www.sciencedirect.com/topics/computer-science/neo4j#:~:text=Neo4j%20has%20some%20upper%20bound,relationships%20in%20a%20single%20graph.&text=No%20security%20is%20provided%20at,there%20is%20no%20data%20encryption.&text=Security%20auditing%20is%20not%20available%20in%20Neo4j.