Bachelor's thesis
Computer Science
June 17, 2022

Analysis of the recent uptake and impact of NoSQL databases in companies

The practices, concept and challenges of NoSQL

Linnea Gullmak

Dept. of Software Engineering
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden

This thesis is submitted to the Department of Software Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the bachelor degree in the science of software engineering. The thesis is equivalent to 20 weeks of full-time studies.

Contact Information:
Author(s):
Linnea Gullmak
E-mail: ligm19@student.bth.se

University advisor:
Dr. Nauman bin Ali
Dept. of Software Engineering

Dept. of Software Engineering
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

Context: Data is at the heart of any information system. Choosing the appropriate database and its operation is a major decision for any company and choosing from the pool of different options can can feel overwhelming. In this thesis we take a look at the main factors to consider when making your decision, to help you with the whole process. This thesis will explore the selection, prioritization and considerations when choosing a database. It is aimed at exploring the recent uptake and impact of NoSQL in companies and analyze the results of the literature and empirical study.

Aim and Objectives: Our aim is to investigate the recent uptake and continued use of NoSQL databases in software development companies. It is imperative to know how companies are choosing to adopt the right technology for their application. The objective is to provide instructions for companies on how to choose the right DB for their needs and what to consider.

Method: Interviews are conducted to find out the process/approach that practitioners employ when choosing the database technology. Then an analysis of the considerations and their priority is conducted using a questionnaire. The focus is on the considerations, meaning factors to consider when choosing a database.

Results: The result of the interviews show that infrastructure is the most essential consideration when choosing a DB, and the survey questionnaire show that consistency is the most essential consideration.

Conclusions: The result suggests that there are several essential considerations when choosing a database. Furthermore, we conclude that the challenges of adopting NoSQL technology may be the following: only provide eventual consistency, which can impact availability and performance, reliability, the challenges of transitioning, keeping track, lacking data integrity, handling of complex queries, and security and privacy risks.

Keywords: considerations, NoSQL, uptake, companies

Acknowledgement

My Bachelor's thesis has been an interesting and learning experience. I would like to thank my boyfriend, who has supported me during this thesis. My warm thanks are also extended to my lovely cats for their patience. Much love, appreciation and thanks to my parents, who has supported me and my aspirations all my life. I would especially like to express my thanks and gratitude to my father and his very valuable connections and for his moral support.

My warm thanks to the University, Blekinge Institute of Technology, the university that gave me the opportunity to conduct my thesis, my reviewer and the examiner.

I would like to express my heartfelt gratitude to my supervisor Dr. Ali for his help and guidance during the process of my thesis. His motivation and encouragement during my research is highly appreciated. This work would have not been possible without his knowledge, valuable suggestions and discussions.

Contents

Chapter 1

Introduction

The main purpose of this thesis is to investigate the considerations when companies adopt NoSQL. This also provides insights into the recent uptake of NoSQL in the software industry. We will further explore the challenges of adoption [1] and long-term use of NoSQL in companies.

If your business is growing, keeping track of the increasing amount of data can be tricky. Having the right database for your purpose can help you manage all your business-critical data centrally, safely and securely and boost your chances of success.

Using the right database for your needs if you manage staff records can save you time and money. It can automate routine jobs and speed up the processing of data such as hours, leave, benefits, payroll, etc. This may leave you more time to focus on growing your business.

Researchers like Maté [2] and Chatzipetrou [3] have studied different strategies and procedures, which could be applied when choosing a database. They believe prioritization is a procedure of principal importance in decision making. In Software Engineering it is encountered in cases where multiple attributes have to be considered in order to take a decision. However, opinions are subjective and may vary greatly when different people try to prioritize independently. These factors have led to the adoption of voting schemes where stakeholders express their relative preferences for certain attributes in a systematic and controlled manner [3].

Researchers like Chatzipetrou [3] focus on the investigation of what matters the most to industry practitioners during component selection. They believe component-based software engineering is a common approach to develop and evolve contemporary software systems. Another existing study [4], which focus on evaluating different database systems for the purpose of selecting the most suitable DB, analyzed performance. "NoSQL DBs show superior performance compared to SQL DBs, demonstrating that NoSQL DBs are more appropriate for processing large amounts of data"

So, what is the thought process and strategies used when choosing a DB? Could performance and cost be major factors? This thesis will further investigate the factors and analyze the essential considerations when choosing a DB.

The origin of the problem is that selecting a DB can be a major concern for businesses. There are many database management systems available on the market and choosing the one can be overwhelming. However, if companies know the essential considerations then it might help them decide.

The main goal of this thesis is to investigate companies recent uptake of NoSQL. The focus is on the selection and prioritization of considerations. The enumeration and prioritization will reveal what the main concern is.

Key terminology

Relational databases

Relational databases, also known as SQL databases.

Efficiency

Efficiency, meaning when you carry out the correct tasks in the right way, with the least waste of time and effort.

NoSQL

NoSQL, refers to non relational databases.

DBs

DBs stands for databases. A database is an organized collection of structured information, or data, typically stored electronically in a computer system.

Big Data

Big Data is a collection of data that is huge in volume, yet growing over time.

Cost

The cost can both be the price, and time and effort.

Impact

The impact is the effect or influence that NoSQL has on a situation or person.

Schema

A schema defines the structure of a database, most often for a relational database.

Scalability

Scalability is the ability to expand or contract the capacity of system resources in order to support the changing usage of the application.

Availability

If a database is available, then the data is easy to access. Any condition that renders the resource inaccessible causes the opposite of availability, aka unavailability.

Agile

Agility is the power of moving quickly and easily. An agile database can be defined as nimble and effective.

RDMBS

A RDMBS "Relational Database Management System" is a collection of programs and capabilities that allows you to create, update, administer and otherwise interact with a relational database.

1.1 Research questions

In this section 1.1, we present the research questions. The research questions are framed to fill the gaps in the research area. The two research questions and motivation behind these questions are:

1.1.1 Research question 1

What are the essential considerations when choosing a database?

Motivation

What should be considered when choosing the DB? The choice of a DB should not be taken lightly because the DB is an important part of the application. Companies are most likely storing sensitive data, which should be safely and securely stored and handled. Database systems are very important to your business because they communicate information related to your sales transactions, product inventory, customer profiles and marketing activities. This question is for those

who want to know how to choose and what to consider or for those who want to know some strategies for choosing a DB.

1.1.2 Research question 2

What can be the challenges of adopting NoSQL technology?

Motivation

NoSQL databases may offer many benefits over traditional relational technology including a more flexible data model, horizontal scalability, and superior performance. But there might be certain challenges, which comes with these benefits. As a company you might want to know the challenges before deciding on a DB and as a developer you want the process of developing to go smoothly. These challenges are therefore important to recognize when choosing to adopt and use NoSQL.

1.2 Background

1.2.1 What is NoSQL

NoSQL databases are non-relational and can accommodate unstructured data [5]. They are not a replacement for SQL DBs, but rather compliments it and both technologies can coexist. The NoSQL technology can also be described [6] as a different approach to data storage and access, when compared to Relational Database Management Systems (RDBMS). "NoSQL DBs is often referred to as “Not Only SQL” to highlight the fact that most systems do not leave the relational model completely."

NoSQL or “Not only SQL” is a type of database that can accommodate a wide variety of different data types and models [7], while holding different formats such as key-value, documents, columns, and graphs. "NoSQL is schema-less and more flexible to adapt to changes".

However, researchers like Oliveira [8] can not give a solid definition of what NoSQL is. "NoSQL is more understood as a movement that proposes non-relational database solutions, which do not use the SQL language. Thus, the term NoSQL is often interpreted as "Not Only SQL"".

NoSQL data model

Hassan [9] studied the advantages and the limitations of relational databases, the NoSQL data model, types of NOSQL data stores, characteristics of each data

store and advantages and disadvantages of NoSQL over RDBMS. The author´s purpose is to help the interest users take a review of the different database model solutions, which can serve as a base for selecting the proper database model, and satisfy their application requirements. Parts of the structure is borrowed and some interesting considerations are used in this study.

Characteristics of NoSQL

NoSQL explained by Chen and Lee [10]:

Non-relational

NoSQL databases do not use the relational database model, and they do not support SQL join operations. Therefore, the related data needs to be stored together to improve the speed of data access.

Distributed

Data in NoSQL databases is usually stored in different servers and the locations of the stored data are managed by metadata.

Open-source

Most NoSQL databases are open source and free to download.

Horizontally scalable

Horizontally scalable, meaning you can increase or decrease multiple servers to meet the data processing capacity of NoSQL databases.

Schema-free

NoSQL databases do not need a schema, meaning NoSQL databases can flexibly add data.

Easy replication support

NoSQL databases mostly support master-slave replication or peer-to-peer replication, making it easier for NoSQL databases to ensure high availability.

Simple API

The NoSQL database provides APIs for network delivery, data collection, etc. for programmers to use, so that programmers do not need to design additional programs to make writing programs easier.

1.2.2 The purpose of NoSQL

SQL DBs are primarily called Relational DBs (RDBMS) [11]. The term SQL stands for "Structured Query Language". Until a decade ago there was only SQL, distributed, sometimes replicated, and fully consistent DBs. But then, web and cloud applications emerged and they needed to deal with complex big data. The purpose of NoSQL was to address this rising data.

The traditional RDBMS support the so called ACID properties, which means the data is strongly consistent. The ACID model stand for Atomicity, Consistency, Isolation and Durability [8]. Atomicity, meaning the transaction must be executed in its entirety or not to be executed at all. Consistency, meaning if a transaction runs entirely from start to finish, without interference from other transactions, it should take the database from one consistent state to another. Isolation, meaning the execution of a transaction must not be interfered with by any other transaction running at the same time. Durability, meaning the changes applied to the database must be persistent in the database and changes must not be lost due to any failure. Please see Figure 7.4. NoSQL DBs do not support the ACID properties, instead they follow the BASE properties. Please see Figure 7.3. The BASE properties can further be described as follows:

Basically Available

The DB system can execute and always provide services. Some parts of the DB system may have partial failures and the rest of the DB system can continue to operate. Some NoSQL DBs typically keep several copies of specific data on different servers, which allows the DB system to respond to all queries even if few of the servers fail. In other words, basically Available means an application should work basically all the time.

Soft state

The DB system does not require a state of strong consistency. This means that no matter which replication of a certain data is updated, all later reading operations of the data must be able to obtain the latest information. In other words, soft state means an application need to be consistent all the time.

Eventual consistency

Eventually consistent means the consistency will be achieved once all writes are propagated to all nodes [8]. In other words, the DB system needs to meet the consistency requirement after a certain time. Sometimes the DB may be in an inconsistent state. For example, some NoSQL DBs keep multiple copies of certain data on multiple servers. However, these may be inconsistent during a time. This

may happen when a copy of the data is updated while the other copies continue to have data from the old version. The replication mechanism in the NoSQL DB system will eventually update replicas to be consistent. You could also say that eventual consistency happens when an operation is confirmed without checking all nodes [12]. A temporary inconsistency between redundant servers (nodes) is accepted, but eventually the system will reach a consistent state. Simply explained, eventual consistency means an application should be in some known state eventually.

Scaling

NoSQL DBs can scale horizontally, while relational DBs can scale vertically [13]. Horizontal scaling refers to adding additional nodes by adding more servers or instances. Vertical scaling describes adding more power to your current machines, using larger ram or a stronger processor. Please see Figure 7.5

The four main types/models of NoSQL databases

It is said [14] that NoSQL DBs can accommodate a wide variety of data models or types, including key-value, document, column and graph formats. Please see Figure 7.6.

Key-value

A key-value DB is a combination of two main attributes, key and value. You could describe key-value as a DB using a hash table where a unique key and a pointer to a particular item of the data are [13]. Key-value stores provide high query processing speed, good performance and options for large storage. Redis is one example of an advanced and open-source key-value store [14].

Document

The document store databases store the data in the form of documents, which consist of two main attributes, key and document. Documents resemble the rows of relational databases, but are more flexible due to them being schema-less [14]. Furthermore, the structure of the documents do not need to be formally specified beforehand, because it is usually the application programs that verify rules about the structure of a document.

Column-based

Column-based databases store the data in column oriented tables, unlike relational databases that store the data in a more row oriented way [13]. The data is stored in cells grouped in the columns of the data rather than as rows of data.

Columns are logically grouped into the column families. The column families can contain a virtually unlimited number of columns that can be created at runtime or the definition of a schema. So, reading and writing data is done by using the columns instead of rows.

Graph-based

Graph databases are based on the graph theory concept, meaning they store data in the form of graphs. A flexible graphical representation is used, which help to address scalability concerns [13]. Graph structures are used with edges, nodes, and properties, providing index-free adjacency. The graph contain information about the properties related to nodes. Neo4j is one example of a graph database [14].

Here are some NoSQL DBs that store in document, key-value, column and graph:

- DBs using document: MongoDB, CouchDB
- DBs using key-value stores: SimpleDB, Redis, Riak, Dynamo, Voldemort
- DBs using column-oriented: Cassandra, DynamoDB, HBase
- DB using graph: Neo4j

1.2.3 DB vs DBMS

DB stands for database, and a database can be described as a collection of information that is organized so that it can be easily accessed, managed and updated. DBMS stands for Database Management System, and it is a system software for creating and managing databases. The DBMS provides users and programmers with a systematic way to create, retrieve, update and manage data.

Here is a quote that might explain this further [8]:
"A database is created and maintained through a database management system (DBMS), a computer program that helps maintain and use data sets that compose the databases."

Chapter 2

Research method

This part of the document explain the research method chosen to answer the research questions. The motivation for choosing the research method and the applied strategies are mentioned. Firstly, a literature review is made and then a handful of interviews are conducted. Lastly, a survey questionnaire is sent to different employees in companies.

2.1 Literature review

The literature review is intended to collect considerations and their relative importance in the context of database technology selection decisions.

For practical reasons there is a focused search in Scopus using the following search string: (TITLE ("nosql") AND TITLE-ABS-KEY ("choosing" OR "selecting")).

The purpose of a literature review is to gain an understanding of the existing research and present that knowledge. Also, the purpose is to help build knowledge in this field. A literature review is done by gathering the required, relevant and related literature materials. Sources concerning the subject that the research is describing is used. After checking if the source is relevant, the concentration is given to the actual content.

The aim of the literature review is to collect essential considerations using Scopus, a selection criteria, and data extraction. The selection criteria is used for selecting relevant papers from search results in the literature review and the data extraction form is used to find the right data to collect from the selected papers. Table 2.1 below is an example of data extraction. There are three examples of different sources with ID 1-3. ID 1 mentions Cassandra, the reason for choosing Cassandra, which is cost, and what kind of cost, which is operational cost, in this case, and if the factor has a low or high priority when choosing the DB. ID 2 mentions MySQL, where the factor for choosing the DB is the ability, specifically query performance. However the factor has low priority, meaning it might not be

ID	Database	Factor	Definition	Priority
1	Cassandra(NoSQL)	cost	operational	high
2	MySQL(SQL)	ability	query performance	low
3	MongoDB(NoSQL)	scaling	horizontal	high

Table 2.1: Data extraction

an essential consideration. The last source with ID 3 mentions MongoDB where scaling is a factor for the choosing the DB, and specifically horizontal scaling. Here we can see that the priority is high, meaning that the factor might be an essential consideration.

2.2 Interviews

Interviews are an important data gathering technique involving verbal communication between the researcher and the participant. For the interviews the participants are asked to select the most important considerations from a list. Next, they got to prioritize their selections using the 100 dollar method.

The 100 dollar method is a prioritization method that can be used where participants can prioritize items by spending an imaginary 100 dollar. The participant may choose to give all 100 dollars to one single consideration, or the person may distribute the the money more evenly. This method is one way to reveal which considerations are the most important. Please see Figure 7.2. Prioritization is a procedure of principal importance in decision making. In Software Engineering it is encountered in cases where multiple attributes have to be considered in order to make a decision. The purpose of the interviews is to help explain and explore the opinions, behavior, experiences, phenomenon, etc of the respondents.

The answers gathered are then used and analyzed in the research to locate useful expertise. However, it's important to remember the answers collected will not represent development around the world or through the whole industry. My goal with the empirical research is to at least get some different perspectives and enough information to analyze and draw conclusions.

2.2.1 Target population, size of sample and sampling strategies

Sample size is a statistical concept that involves determining the number of observations used to estimate the variability that should be included in a statistical sample.

The confidence interval or margin of error, is the plus-or-minus figure, which is an estimated range of likely values for a population parameter, for example, 40 ± 2 or 40 ± 5 percent. Taking the commonly used 95 percent confidence level as an example, if the same population were sampled multiple times, and interval estimates made on each occasion, in approximately 95 percent of the cases, the true population parameter would be contained within the interval. Some factors that affect the width of a confidence interval include size of the sample, confidence level, and variability within the sample.

The confidence level is a percentage measure of how sure you can be that your findings are accurate. If your sample size conforms to the 95 percent confidence level, then 95 people in every 100 in your total population would answer the same way.

Size of sample is determined by using an online sample size calculator (fieldwork-assistance.co.uk/what-we-do/sample-size-calculator). Firstly, a calculation is made to find out how many interviews should be selected, where the population size is 100, which I believe is as a reasonable population size. A confidence level of 95 percent is chosen, a confidence interval of 50, with a population of 100. The calculator show a sample size of 4.

Then, a calculation is made to find the so called confidence interval, which means finding out how robust the data is. We still have a confidence level of 95 percent, a sample size of 4, which is calculated above, the population is still 100, and the percentage is 50. The calculator show a confidence interval of 48.25.

To find out the sample size for the survey questionnaire another online calculator is used (calculator.net/sample-size-calculator.html). The confidence level is 95 percent, the margin of error is 30 percent, the population proportion is 50 percent, and the population is 100. The calculator show a sample size of 10.

Then, a calculation is made regarding the confidence interval of the survey. Again, the confidence level is 95 percent, the sample size is 10, which is calculated above, the population proportion is 50 percent, and the population is 100. This means, in this case, there is a 95 percent chance that the real value is within ±29.55 percent of the surveyed value.

The target population is people in an organization that have had to decide on a database. The sampling strategy is snowball sampling, purposeful sampling and convenience sampling. Snowball sampling is when research participants help recruit future subjects for a study. Please see Figure 7.1 This method is used to find participant for the survey. 2 people got the link to the survey, please see the link in the appendix, then they would send the link to other employees in the

company, and now there are 10 people in total, which has answered the survey. Purposeful sampling and convenience sampling is used to find participant to be interviewed. These sampling strategies are used because they are drawn from a source that is conveniently accessible to the researcher and because it is easier for the researcher to rely on their own judgment when choosing members of the population to participate.

2.2.2 Three broad categories that are taken into consideration

Demographics about the respondents

Statistical data about the characteristics of the respondents (company size, the technology used, product domain, size and nature of data they collect, store, and process etc). This data is used to identify the characteristics of the respondents like their job and responsibilities in the company and the company itself, which they work for, for example, size of the company. This is noted, because it can affect the result of the research.

Considerations when choosing a DB technology

To collect essential considerations for choosing a specific database. The intent is to analyze the considerations and their priority. This criteria is used by having a focused search during the literature search, during the interviews and the with the survey.

Importance/prioritization of the considerations

Prioritizing considerations to understand the relative importance of factors/considerations. It is important to remember that the focus of this research is on the selection and prioritization of considerations. The enumeration and prioritization will reveal what the main concerns are. The idea, for example, is to use the 100 dollar method to find out the importance or the prioritization of the considerations.

2.3 Survey questionnaire

The result of the survey is part of the result and analysis phase. The survey answers had to be managed and constructed in a systematic manner in order to minimize the researcher´s efforts and time. In other words, the questions and answers had to be managed and controlled. The purpose of the survey questionnaire is to gather statistical information about the attributes, attitudes, or actions of a population by a structured set of questions.

Chapter 3

Result and analysis

This section is about the results from the literature, interviews and survey and their analysis.

3.1 Literature review

Search results

When using the following search string in Scopus : (TITLE ("nosql") AND TITLE-ABS-KEY ("select")) it presented 32 document results. At least 3 documents were chosen. When using the following search string in Scopus : (TITLE ("nosql") AND TITLE-ABS-KEY ("choosing")) it presented 29 document results. At least 1 document is chosen. When using the following search string in Scopus : (TITLE ("nosql") AND TITLE-ABS-KEY ("challenges")) it presented 174 document results. At least 3 documents were chosen. When using the following search string in Scopus : (TITLE ("nosql") AND TITLE-ABS-KEY ("adopting")) it presented 10 document results. At least 2 documents were chosen.

3.1.1 What are the essential considerations when choosing a database?

Choosing a DB solution should not be made on the basis of random choice or personal whim, but rather based on criteria, application characteristics, users, and data [8].

Hassan [9] studied factors to consider, with the purpose to help interested users to take a review of the different database model solutions, concluded that there are many factors that should be taken into consideration when choosing a DB. The main considerations found are the type, the amount of data, the schema characteristics, the cost, the transactions amount and how frequently they are called.

A NoSQL DB is considered schemaless because it does not require a rigid, predefined schema, like a relational database do, and NoSQL DBs can in this way add data more efficiently and flexibly.

Chen and Lee [10] agree that there are several factors to be considered in order to select an appropriate database. These factors include the data model, access patterns, queries and non-functional requirements, including data access, performance, replication, partition and scalability.

However, Chen and Lee are not alone in believing the data model is an essential consideration when choosing a DB. Researchers like Oliveira [8] say: "characterizing a system's data model is essential when choosing a DB". Also, Choosing which logical data model to use can affect the performance of operations.

The purpose of the application is a criteria for choosing a proper database model. Therefore, it is extremely important that we always identify the requirements of each application, as this is an essential issue. When choosing a database you generally want to account for the needs of the enterprise, as well as the popularity and the feedback about each database.

It could be said that: "It is important to select a suitable database for a certain enterprise because this decision may affect the performance of the operations" [10]. Achieving good performance may very much depend on how you deploy your infrastructure, for example, dedicated or cloud, in memory or on disk or multi data-center deployment. So, the conclusion could be that infrastructure is an important consideration when searching for a suitable database.

If the data is documents, unstructured or semi-structured with advanced query features, or if the data is schema-less or the schema is continuously changing while the consistency is preferred over availability, then using NoSQL might be the appropriate choice. Based on this, the conclusion is that consistency and availability are essential considerations when choosing a DB.

Ceresnak and Kvet [13] studied query performance in relational and non-relation DBs, and made a test for query performance of DBs. "The result for NoSQL DBs, Mongo and Cassandra show a faster query performance rate". Researchers like Kanchan [6] conclude from an extensive empirical analysis that NoSQL outperforms SQL based systems in terms of basic read and write operations. "But SQL based systems are better if queries on the dataset mainly involve aggregation operations"

According to one source [2] NoSQL technologies have become a common component in many information systems and software applications, focusing on per-

formance and enabling scalable processing of large volumes of structured and unstructured data [2]. "Unfortunately most developers consider security an afterthought, putting at risk personal data of individuals and potentially causing severe economic losses as well as reputation crisis". However, other researchers like Sicari [14] have studied security and privacy for NoSQL databases believe "Security and privacy represent critical requirements in databases, in general, and for NoSQL databases, in particular. This is due to them being non-relational databases, and frequently based on sharding, meaning that data is distributed over multiple servers". Also, some relevant factors the authors [14] bring up are: authentication, authorization, access control, privacy, policy enforcement, integrity and confidentiality. Furthermore, other researchers [15] [16] have conducted research regarding security. "Security and privacy represent critical requirements in databases, in general, and for NoSQL databases, in particular" [14]. This could mean that access to data, data integrity and security is essential considerations when choosing a DB.

Bhogal and Choksi have researched NoSQL [5] and believe the popularity of NoSQL systems can be caused by their efficiency in handling unstructured data and backing up effective design schemes that give the system users efficient flexibility and scalability. Furthermore, authors [3] researching component selecting in software engineering is claiming that companies often need to face challenges with growing complexity, limits on scalability, increasing data volumes, data security and data management. This may suggest that the need for businesses and organizations to innovate means they have to stay agile and continue operating at any scale.

Chatzipetrou [3] explored what matters the most to industry practitioners during component selection. The result show cost is considered the most important attribute during the selection. It is said [3] that practitioners who work on more mature products, more than 15 years, value non-functional attributes. However, the authors claim cost is still their first priority.

SQL and NoSQL databases are designed to fulfill very different needs [17]: RDBMSs provide a high level of functionality whereas NoSQL databases excel on the non-functional side, through scalability and availability. So, if the schema is continuously changing while the consistency is preferred over availability, then using NoSQL might be the more appropriate choice. But if the data is structured or extremely relational while high availability is preferred over consistency, then using NoSQL might not be the most appropriate choice.

Researchers like Roy, Hubara and Maté studied the selection of DBs and they believe [18] you can select the database based on trends, or by selecting based knowledge, or by analyzing data. These strategies can further be described [2] as

follows:

Agenda-based

Agenda-based strategy means that the selection is based on trends and the strong desire to learn something new. Many new technologies emerge daily and some of them tend to gain much attention. "When exposed to new technology, developers may tend to adopt it mainly because of the attention it gets"

Knowledge-based

Knowledge-based strategy means that the selection is based on personal or organizational knowledge or experience with previously used databases.

Exploration-based

Exploration-based strategy means that the selection is based on analyzing a project's data, goals and finding the best DBMSs that fit the project needs. You could say: "This approach is the most optimal and would possibly result in a fitting technology for the project's overall needs (data, functional, non-functional, etc)". Unfortunately, since it takes more time and resources, it is not often used.

3.1.2 What can be the challenges of adopting NoSQL technology?

In October 2021, the top database, according to DBEngines Ranking was Oracle. Hassan [9] conclude that some NoSQL DBs can be found among the top 10. After checking DBEngines Ranking it show that this year, 2022, Oracle is still at the top of the list. MongoDB is placed as number 5 and Redis as number 6.

Adopting the BASE model can increase availability at the expense of consistency [8]. "Adopting the BASE model of transactional properties promotes an increase in availability at the expense of relaxation of consistency, which implies a decrease in the delay but rises the possibility of occurrence of inconsistencies". So, with distributed NoSQL databases there may be a trade-off between consistency and availability.

It is said that: "Some databases are built to guarantee strong consistency and serializability (ACID), while others favour availability (BASE)" [17]. NoSQL systems have eventual consistency and the reason is due basically all of them being distributed. The challange with this is that with fully distributed systems there might be difficult to maintain strict consistency. In other words, the availability and the performance benefits at the cost of consistency and this means NoSQL

databases can be less reliable than relational databases since they compromise reliability for performance.

Different consistency levels for read operations can be adopted, and such levels may distinctly affect system behaviour. So in other words, the choice of the consistency level can impact availability and performance of storage systems [12].

"There are large differences among NoSQL DBs" [17]. Riak and Cassandra, for example, can be configured to fulfill many non-functional requirements, but are only eventually consistent and do not feature many functional capabilities apart from data analytics and, in case of Cassandra, conditional updates. MongoDB and HBase, on the other hand, offer stronger consistency but do not maintain read and write availability during partitions.

It is said [5] that some of the main advantages of RDBMS are flexibility, simplicity, ease of data retrieval and data integrity. However, in comparison to relational databases, NoSQL databases do not have the same distinct properties or data integrity [5]. This is due to the fact that relational databases support the ACID constraints (atomicity, consistency, isolation, durability), each of these four qualities guaranteeing stability, security, and contribute to the ability of a transaction to ensure data integrity.

Whilst NoSQL is great for dealing with many difficulties surrounding and unstructured data, it is however limited in several key areas [5]. “Those who do not understand the lessons from previous generation systems are doomed to repeat their mistakes" is a quote, which the authors [5] believe may be justified with NoSQL having barriers that include consistency and reliable standards. Since NoSQL data is held in partitions, either availability or consistency cannot be 100 percent guaranteed. According to the the authors [5], this can have an impact on the ability to handle complex queries.

It is said [14] that apart from scalability and performance, security and privacy requirements represent one of the most difficult challenges NoSQL databases face nowadays. NoSQL databases were not initially designed and integrated with security and privacy-related functionalities, such as data encryption, authentication, and authorization mechanisms.

Furthermore, the highly distributed nature of NoSQL DBs further pose security risks. The main security risks encountered by NoSQL databases are related to non encrypted data storage, unauthorized exposure of data and backups, or replicated data, and insecure communication over the network [14].

There are many database management systems available on the market and choos-

ID	Factor	Reference
1	The type	[9]
2	The schema	[9]
3	Transactions	[9]
4	Consistency	[9]
5	Speed	[9]
6	Cost	[3]
7	Data model	[10]
8	Non-functional requirements	[10]
9	Data access	[10]
10	Replication	[10]
11	Partition	[10]
12	Query performance	[13]
13	Security	[15, 16]
14	Availability	[5]
15	Scalability	[5]
16	Infrastructure	All references listed above

Table 3.1: Factors to consider

ing the one can be overwhelming. This can also make it hard to keep track of where specific NoSQL DBs excel, where they fail or even where they differ, as implementation details change quickly and feature sets evolve over time [17].

3.1.3 List of essential considerations/factors

Here are the enumerated essential considerations from the literature review presented. They are listed in no particular order, and there are 16 considerations in total. Please see Table 3.1. The factors are further explained below.

The type

You need to consider what type of data the database can handle and how. Maybe think about what type of data you can store and use with your potential database.

The schema

The schema can have an impact on how efficiently the DB runs and how quickly you can retrieve information from the DB, and NoSQL DBs do not generally have a schema in the way relational DBs do.

Consistency

keeping the data consistent becomes even more important as more sources feed into the database. Therefore, consistency rules are very important and the ability to define these should be considered when choosing a new DBMS.

Speed

Speed of data access or query processing speed.

Cost

It could be wise to make sure your decision is based on the software being fit for the purpose. It could be a costly mistake to adopt a DB and invest time and resources, only to find that it does not fit your needs.

Data model and transactions

There is a major difference regarding the transactions support between RDBMSs and NoSQL DBs [6]. RDBMSs offer the ACID (Atomicity, Consistency, Isolation, Durability) model for transactions, assuring all transactions are properly completed and data remains consistent. However, NoSQL stores provide BASE (Basically, Available, Soft-State, Eventually Consistent) model to provide high scalability, availability, and performance compromising strict consistency.

Scalability

A database might fit your needs now, but when data and demands are growing then you may require your new database to grow with your needs. Most database solutions allow you to add additional capacity, however, you might want to consider how this happens.

Non-functional requirements

Non-functional requirements include several considerations, including security, capacity, performance, reliability, maintainability and usability. These are requirements that could be underestimated or taken for granted, but might have an impact on the cost and resources in the context of choosing a DB.

Security

Does your system need to control the user access and session, and store data in a secure location and format? Does it require a secure communication channel for the data?

capacity

How many users does the system need to handle? How much data can the system store and for how long?

Performance

What about the performance? Performance could generally be described as a time expectation. This is also important to consider when talking about non-functional requirements.

Reliability

Is it necessary to ensure and notify about the system transactions and processing? Having a system log will increase the time and effort. This in turn can have a significant impact on the system.

Maintainability

How and for how long will the system be maintained? For how long is system meant to be up and running?

Data access

The ability to access the data in the DB.

Replication

Replication is the ability provide a consistent copy of data across all the database nodes to increase the availability and reliability of the data.

Partition

Database partitioning is when you split the data into separate tables or nodes and is usually done for manageability, performance or availability reasons. It is popular in DBMSs, where each partition may be spread over multiple nodes, where users at the node perform local transactions on the partition.

Security

Do you have sensitive data and personal information, which must be stored securely to adhere to regulations and be protected from loss or theft?

Infrastructure

The hardware and networking infrastructure to take care of provisioning, scalability, resiliency, failover, restoration and backup.

3.2 Result of the interviews and survey questionnaire

The enumerated essential considerations gathered from the literature review can be interpreted in different ways and it could be argued that many considerations are the same. The intention is to give participants from the interviews and survey as many choices as possible and then ask them for their interpretation of the chosen considerations.

The survey questionnaire

For the survey, when asked if they considered infrastructure a factor, only four respondents would Tend to disagree. Please see Figure 7.9. So, the conclusion is that not everyone consider infrastructure or that the infrastructure is interpreted differently.

The survey answers reveal that consistency, meaning that only valid data will be written to the database, is the most essential consideration when choosing a database. Please see Figure 7.8. This is the only statement were no respondents disagreed. All ten respondents more or less agree that consistency is an important consideration when choosing a database.

The survey included information about the purpose, intent and consent agreement. The questions are written in statements and the answer options are Agree, Tend to agree, Tend to disagree, and Disagree.

The interviews

The result of the interviews show that the most essential consideration is infrastructure. Please see Figure 7.9. Another essential consideration mentioned several times in the interviews is security. Please see Figure 7.7. Only one respondent would answer Tend to disagree when asked if they would consider security. The other respondents would consider security when choosing a database.

Two of interviewees (id 1 and id 4) work for the same company, which is a big company with millions of users. Another participant (id 2) also works for big company with many users. However, one participant (id 3) has their own smaller

company, with much less users, but where this person can make all the decisions regarding the database. Two of the interviewees (id 1 and id 2) worked as a developer, one regular and one senior, and the other two (id 3 and id 4) worked as a an architect, although as different kinds of architects.

Id	Factor	Definition	amount
1	infrastructure	indestructible, network, servers, internet provider	50
2	infrastructure	indestructible, all hardware, network etc.	50
2	security	operations, security experts, access etc	30
3	type	using the right tool and guarantee consistency	30
4	type	consistency	30
1	data model	the customer´s perception and business idea	20
2	speed	cost of expert who can configure and maintain	20
3	transactions	guarantee data integrity, including consistency	20
3	security	data access, infrastructure and hosting	20
1	security	indestructible and correct (true) data	20
4	transactions	consistency, storage and handling	20

Table 3.2: The highest rated considerations from the 100 dollar method. The table show the amount of dollars four different people with id 1-4 would spend on the factors/considerations out of 100 dollars. For example, two interviewees spend 50 dollar out of 100 dollar on infrastructure.

Id	Role	Database	Reason
1	senior developer	Cassandra	license cost
1	senior developer	Cassandra	availability
1	senior developer	Cassandra	scalability
2	developer	PostgreSQL	configuration
2	developer	PostgreSQL	memory
2	developer	PostgreSQL	resources
3	architect	PostgreSQL	data access
3	architect	PostgreSQL	consistency
3	architect	PostgreSQL	performance
4	product architect	PostgreSQL	transactions
4	product architect	PostgreSQL	performance

Table 3.3: Brief description of the roles of the interviewees and their choice of database and the reasons for the choice.

Chapter 4

Discussion

4.1 The ethical, societal and sustainability aspect

In this research the aspect of ethical, societal and sustainability are relevant to this work in terms of the adoption of NoSQL in the companies and its impact. The ethical aspect is privacy and security. The choice of the database can impact sensitive data that is stored daily in the databases, making the privacy problem more serious while raising essential security issues. Choosing the wrong database can impact the application and the handling of Big data, which in term can affect the overall sustainability of the company. Storing Big data has become a crucial part of many applications in all sectors of human life. Big data, meaning large, hard-to-manage volumes of data.

4.2 What are the essential considerations when choosing a database?

Factors such as infrastructure, security and consistency is mentioned in the literature. Please see Table 3.1. Many of the interviewees chose these factors and talked about their impact in companies. Please see Table 3.2. Therefore these factors are considered to be essential considerations. However, It could be argued that infrastructure include both security and consistency, which many of the interviewees pointed out. It could also be argued that businesses and organizations cannot create value out of data without having the proper infrastructure. This would include the entire support system required to process, store, transfer, and safeguard data.

The fact that in this study about NoSQL, 70 percent of respondents does not actually use NoSQL is worthy of note. Maybe it is due to the fact that NOSQL can't guarantee the ACID properties. Or this could be the case due to financial transactions, etc. Transactions are available in NoSQL, however they are not as well developed as they are in relational databases.

Researcher like Maté [2] have studied the security of NoSQL and believe most developers consider security as an afterthought. "Putting at risk personal data of individuals and potentially causing severe economic losses as well as reputation crisis." However, there are several other researchers who have conducted research regarding security [15], [16]. This strongly suggests that most want to have efficient and concurrent access to data, ensure data integrity and security and protect against failures and unauthorized access. Furthermore, the result of the interviews conducted show that security is considered when choosing a DB. It could therefore be argued that security is or at least should be an essential consideration when choosing a DB.

4.3 What can be the challenges of adopting NoSQL technology?

Relational DBs have been around for several decades now. It is not difficult finding developers who have the knowledge and experience in writing SQL queries, normalizing schemas and ensuring ACID properties. However, this is not the case with NoSQL where knowledge, expertise, and experience might be more scarce. Having insufficient knowledge and expertise could lead to the wrong design decisions in using NoSQL where it doesn't fit in. Another common challange and sometimes costly mistake in adopting NoSQL might happen when developers design the NoSQL DB adhering to the fundamentals of relational DBs.

NOSQL has so called eventual consistency [3]. This means that the data, which is written to the database, is not immediately propagated to all of the nodes in the system. If a project needs to scale, due to a lack of performance or disc space, then a switch to NOSQL can be the solution. However, lets not forget that a switch to NoSQL will in most cases come with some work and may take time.

Chapter 5

Conclusion

The conclusions are discussed in this section.

In this paper, interviews are conducted with employees who work in the field. They were presented with 16 essential considerations and chose the ones they thought are the most essential. Please see Table 3.1, which show the list of all considerations gathered from the literature review.

5.1 What are the essential considerations when choosing a database?

Several factors/considerations

Many factors should be taken into consideration when choosing a database [10]. Hassan [9] also studied considerations and he believe some of the main considerations are the type, the schema characteristics and the cost. Chen and Lee [10] believe the data model, data access, performance, replication, partition and scalability are essential considerations. Please see Table 3.1, which show a list of considerations gathered from the literature review. The result of the interviews, using the 100 dollar method, also show that there are several factors to consider, like the cost and data access. According to the interviews the most essential consideration when choosing a DB is the infrastructure. However, this includes the hardware, network, consistency, security and other operational costs. Please see Table 3.2.

Consistency

The result of the survey questionnaire show that consistency, meaning that the data is safely and accurately stored in an indestructible manner, is the most essential consideration. Please see Figure 7.8.

Strategy

Having an appropriate strategy might be essential when choosing a database. The authors of [2] "Improving security in nosql document databases through model driven modernization" believe developers can adopt one of three strategies when choosing a database. These strategies are: agenda-based, knowledge-based and exploration-based. Furthermore, it is important to select a suitable database for a certain enterprise because this decision may affect the performance of the operations [10].

Query performance

Query performance can be an essential consideration. This conclusion is based on the fact that experiments regarding query performance have been conducted several times by researchers [6] [13], many showing faster read and write performance for NoSQL DBs compared to relational DBs.

Security

Some authors studied NoSQL and security and conclude that most developers consider security as an afterthought [2]. However, there are several other researchers who have conducted research regarding security [15], [16]. Also, "security and privacy represent critical requirements in databases, in general, and for NoSQL databases, in particular" [14]. As mentioned, thereult of the interviews, using the 100 dollar method, show that the infrastructure, including consistency and security, is considered to be the most essential consideration when choosing a DB. Please see Table 3.2. So the conclusion could be that security is an essential consideration.

If your business is growing, keeping track of the increasing amount of data can be tricky. Having the right database for your purpose can help you manage all your business-critical data centrally, safely and securely and boost your chances of success.

Flexibility and Scalability

Scalability can be an essential consideration when choosing a DB. The authors Bhogal and Choksi [5] believe the popularity of NoSQL systems can be caused by their efficiency in handling unstructured data and backing up effective design schemes to give the system users efficient flexibility and scalability.

Cost

Some authors like Chatzipetrou investigated what matters most to industry practitioners during component selection [3]. Their study considered cost to be the

most important consideration when choosing a DB. The result of the survey questionnaire show that consistency, meaning that the data is safely and accurately stored in an indestructible manner, is the most essential consideration. Please see Figure 7.8. Furthermore, the result of the interviews show that cost is seen as an essential consideration.

Predictability

So the results gathered using the literature review, interviews and survey questionnaire could suggest that predictability, meaning the correct data is safely and accurately stored in an indestructible manner, is the most important consideration when choosing a DB.

5.2 What can be the challenges of adopting NoSQL technology?

This research question could be explored further and more efficiently. Perhaps there is not enough information gathered to properly answer this question. However, here are some possible challenges presented.

Eventual consistency

In conclusion, NoSQL technology may provide eventual consistency [10]. This means that a temporary inconsistency between redundant servers (nodes) is accepted. However, the system will reach a consistent state eventually.

The result of the interviews show that consistency can be an essential consideration when choosing a DB, so having eventual consistency might not be ideal. Please see Table 3.2.

Less reliable

By having eventual consistency or inconsistency it can have an impact the availability and the performance of the operations [12]. This is why NoSQL technology can be less reliable compared to traditional DBs like SQL.

Trade-off

Adopting the BASE model can increase availability at the expense of consistency [8]. So, in other words, there might be a trade-off between consistency and availability when adopting NoSQL technology.

Cost

There may be a cost when transitioning to NoSQL [5]. The result of the interviews show that the hardware, network and other operational costs are essential considerations when choosing a DB. Please see Table 3.2. So the overall operational cost can be a business concern when adopting NoSQL technology.

Lacking data integrity

"NoSQL technology may be lacking data integrity compared to relational DBs and using NoSQL can impact the ability to handle complex queries" [5]. The result of the survey questionnaire show that consistency, meaning that the data is stored safely and securely with integrity, is an essential consideration when choosing a DB. Please see Figure 7.7.

Security and privacy risks

There may be security and privacy risks when choosing adopting NoSQL technology [14]. The result of the interviews and the survey questionnaire show that security is an essential consideration when choosing a DB, so the aspect of handling the security and privacy risks might be a challenge when choosing to adopt NoSQL technology. Please see Table 3.2 and Figure 7.7.

Lack of knowledge

It might feel overwhelming or confusing when choosing to adopt NoSQL because there are so many different NoSQL DBs to choose from. So, when choosing to adopt NoSQL technology it can be tricky to keep track of where NoSQL DBs excel, where they fail and where they differ [17].

Chapter 6

Validity threats

The structure of this chapter is based on the guidelines for conducting and reporting case study research in software engineering [19]. Validity threats for this thesis are identified and listed here. The different types of threats to validity in this research are:

6.1 Internal threats

Internal threats can be caused by the treatment of the subjects in the research and can happen without the researcher's knowledge. To not be properly prepared when communicating with subjects can result in treating the subjects differently. The results may vary depending on how the questions are asked and interpreted. This is why the interview protocol is used. Please see Appendix 1 regarding the structure of the interview protocol. The protocol is used to ensure the same questions and the same treatment is held for all the interviewees. In this way the internal threats are controlled.

Furthermore, 4/6 employees could be interviewed. This means that 2 of the interviewees dropped out. One dropped out due to moving and the other due to lack of time. So, one internal threats is the risk of dropouts. In turn, the result may be based with sample of only the people who chose not to leave or due to them having something in common, such as higher motivation or time.

The interviewer may guide participants into acting in some type of way through the research method they use. This might result in participants acting in ways different from how they usually would. Also, something to note is that at the end of the interview the interviewees are not as focused as they were in the beginning. Maybe the interview protocol for this study is too long or organized in an insufficient way? Furthermore, choosing the participants at random or in a manner in which they are not representative of the population could threaten the validity of the study.

A past event may directly or indirectly influence the result of research through

the participants. Unrelated events like having a bad day or being stressed can influence the outcome of the study. The outcomes of the study can also be influenced by their knowledge and experiences. They most likely have different experiences working with databases and may have certain preferences. Speaking of experiences, the factors/considerations are not explained to the subjects to instead let them interpret and explain their interpretations. The researcher should only observe and listen, and not willingly or unwillingly manipulate or sway the subject in any particular way.

6.2 External threats

There are three identified threats to external validity because there are three ways the interviewer could be wrong wrong. These threats are the people, places or times chosen. Critics may argue that the results of the study are due to the unusual type of people included in the study, or that it might only be accurate because of the unusual place the researcher did the study in. Or, they might suggest that the study is conducted in a peculiar time.

There could be a situational threat, for example, time of day, location, noise, researcher characteristics, and how many measures are used, which may affect the usefulness and validity of findings. So an external threat could be that the results may only be applied in certain cases. Maybe all the interviewees come from the same place and can only represent a certain place, and perhaps the result of the study cannot be assumed to be the same all over the world.

6.3 Construct validity

A construct validity threat in this research could be that the results from the literature review are insufficient and not suitable for the study. Data required for the case study is gathered from different sources in the form of a literature review, interviews and a survey. This is done in order to achieve data triangulation i.e. gathering data from different sources which would allow analysis in different aspects from the thought process to different perspectives.

6.4 Reliability

It is said that the research will be reliable when the researcher is able to reduce the biases in the study [19]. The reliability is therefor achieved by efficiently reducing all possible biases. In this research the intent is to not use systematic biases and instead use organized and systematic search procedures. In this way threats can be identified and the research will become more reliable.

Chapter 7

Future work

In future work, further evaluation could be made of the decision-making process to support different levels of decisions, i.e. strategic decisions, tactical decisions and selection between different services. Furthermore, there could be research done regarding different business scenarios that involve diverse business models, assets and company characteristics. The aim could be to explore different short- and long-term consequences of each choice of databases. Maybe there should some form guidelines that software business practitioners may use when considering various options.

Future work can explore the combination of different data models for different scenarios and discuss the possible improvements these combinations would have, as well as the challenges of managing more than one database for each application. If and possibly when is one database enough, is a question, which could be explored in future work. The thought process of someone who has full responsibility of choosing a database versus someone who is partly responsible could also be explored in future work.

Research regarding different NoSQL use case scenarios, as well as when NoSQL technology is most useful could be further explored in future work. Many organizations today use more than one database system to meet their different requirements. Future work could explain why organizations have multiple databases and the advantages of doing so.

References

[1] N. B. Ali, "Is effectiveness sufficient to choose an intervention?: Considering resource use in empirical software engineering," in *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM 2016, Ciudad Real, Spain, September 8-9, 2016*. ACM, 2016, pp. 54:1–54:6. [Online]. Available: https://doi.org/10.1145/2961111.2962631

[2] A. Maté, J. Peral, J. Trujillo, C. Blanco, D. García-Saiz, and E. Fernández-Medina, "Improving security in nosql document databases through model-driven modernization," *Knowledge and Information Systems*, vol. 63, no. 8, pp. 2209–2230, 2021, cited By :1. [Online]. Available: www.scopus.com

[3] P. Chatzipetrou, E. Alégroth, E. Papatheocharous, M. Borg, T. Gorschek, and K. Wnuk, "Component selection in software engineering - which attributes are the most important in the decision process?" in *Proceedings - 44th Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2018*, 2018, pp. 198–205, cited By :7. [Online]. Available: www.scopus.com

[4] J. Antas, R. R. Silva, and J. Bernardino, "Assessment of sql and nosql systems to store and mine covid-19 data," *Computers*, vol. 11, no. 2, 2022. [Online]. Available: www.scopus.com

[5] J. Bhogal and I. Choksi, "Handling big data using nosql," in *Proceedings - IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, WAINA 2015*, 2015, pp. 393–398, cited By :43. [Online]. Available: www.scopus.com

[6] S. Kanchan, P. Kaur, and P. Apoorva, "Empirical evaluation of nosql and relational database systems," *Recent Advances in Computer Science and Communications*, vol. 14, no. 8, pp. 2637–2650, 2021. [Online]. Available: www.scopus.com

[7] O. Abahussain and A. Alqaddoumi, "Dbms, nosql and securing data: The relationship and the recommendation," in *2020 International Conference on*

Innovation and Intelligence for Informatics, Computing and Technologies, 3ICT 2020, 2020. [Online]. Available: www.scopus.com

[8] V. Oliveira, M. Pessoa, F. Junqueira, and P. Miyagi, "Sql and nosql databases in the context of industry 4.0," *Machines*, vol. 10, no. 1, p. 20, 12 2021. [Online]. Available: www.scopus.com

[9] M. A. Hassan, "Relational and nosql databases: The appropriate database model choice," in *2021 22nd International Arab Conference on Information Technology, ACIT 2021*, 2021. [Online]. Available: www.scopus.com

[10] J. Chen and W. Lee, "An introduction of nosql databases based on their categories and application industries," *Algorithms*, vol. 12, no. 5, 2019, cited By :16. [Online]. Available: www.scopus.com

[11] P. Valduriez, R. Jimenez-Peris, and M. T. Özsu, *Distributed Database Systems: The Case for NewSQL*, ser. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2021, vol. 12670 LNCS, cited By :2. [Online]. Available: www.scopus.com

[12] C. Gomes, E. Tavares, M. N. O. Junior, and B. Nogueira, "Cloud storage availability and performance assessment: a study based on nosql dbms," *Journal of Supercomputing*, vol. 78, no. 2, pp. 2819–2839, 2022, cited By :2. [Online]. Available: www.scopus.com

[13] R. Čerešňák and M. Kvet, "Comparison of query performance in relational a non-relation databases," in *Transportation Research Procedia*, vol. 40, 2019, pp. 170–177, cited By :14. [Online]. Available: www.scopus.com

[14] S. Sicari, A. Rizzardi, and A. Coen-Porisini, "Security&privacy issues and challenges in nosql databases," *Computer Networks*, vol. 206, 2022. [Online]. Available: www.scopus.com

[15] G. Vonitsanos, E. Dritsas, A. Kanavos, P. Mylonas, and S. Sioutas, "Security and privacy solutions associated with nosql data stores," in *SMAP 2020 - 15th International Workshop on Semantic and Social Media Adaptation and Personalization*, 2020. [Online]. Available: www.scopus.com

[16] N. Singh, A. Shyam, S. R. Swamy, and P. B. Honnavalli, *Differential Privacy in NoSQL Systems*, ser. Lecture Notes in Networks and Systems, 2021, vol. 290. [Online]. Available: www.scopus.com

[17] F. Gessert, W. Wingerath, S. Friedrich, and N. Ritter, "Nosql database systems: a survey and decision guidance," *Computer Science - Research and Development*, vol. 32, no. 3-4, pp. 353–365, 2017, cited By :59. [Online]. Available: www.scopus.com

[18] N. Roy-Hubara, P. Shoval, and A. Sturm, "Selecting databases for polyglot persistence applications," *Data and Knowledge Engineering*, vol. 137, 2022. [Online]. Available: www.scopus.com

[19] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," vol. 14, 2009.

Appendix 1

Interview protocol

Introduction
1. Why the respondent should participate
2. Inform consent
3. Inform the sampling strategy
4. How the information will be stored and used

Demographics (information about the respondents that may help analyze their answer)
1. Experience
2. Education
3. Role and responsibility

Checking if they are appropriate to respond to the rest of the interview
Topic-specific questions:
1. Have you been involved in choosing a Database technology for your company?
In case of a positive answer to the previous question, you may ask the following:
2. How often is such a decision taken?
3. Who else is involved in that decision?
4. How is choosing a database any different from selecting any other off-the-shelf component?
Think back to a time when you were involved in choosing a database technology for you company...
5. Why did the company need to choose a database technology?
6. Was there anyone else, besides you, involved in the decision?
7. Which database technologies were considered and why?
8. Which database technology was chosen and why?

A clear description of how the respondent should apply the 100 dollar method.

Wrap up and thank you (is there something they would like to add and where they can expect to see the results).
Member checking. Is the information accurately understood?

Appendix 2

Factors

Here are explanations of the factors gathered from the 100 dollar method. Please tee Table 3.1.

Infrastructure

Infrastructure is described as all hardware and network, servers, internet provider, etc. It is also described as all things needed to make sure the data is destructible and that the data is safety stored in the database.

Security

Security, meaning paying for security experts, that manage operations and access to data in a safe manner. Achieving security is also described as making sure the data is indestructible and correct (true).

Type

Considering the type, means using the right tools to guarantee consistency.

Data model

The data model is described as the customer´s perception and business idea.

Speed

This is the speed of experts to configure and maintain the database.

Transactions

This is seen as maintaining consistency, storage and handling of data. Transactions is a factor, which is defined as guaranteeing data integrity and consistency.

Appendix 3

Link to the survey questionnaire

https://freeonlinesurveys.com/app#/1594630/build

Appendix 4

Figures

Figure 7.1: Snowball sampling

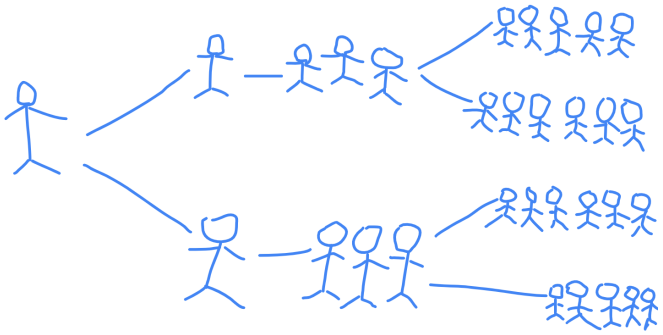

Figure 7.2: The 100 dollar method

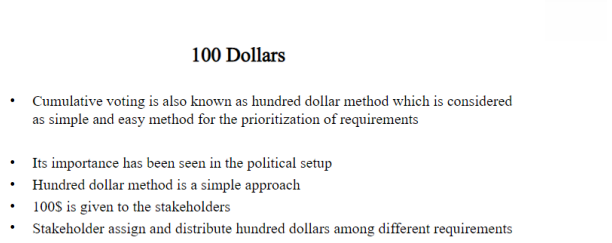

Figure 7.3: BASE

Figure 7.4: The ACID model

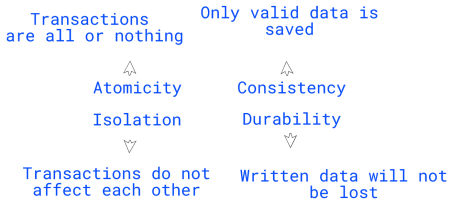

Figure 7.5: Scaling

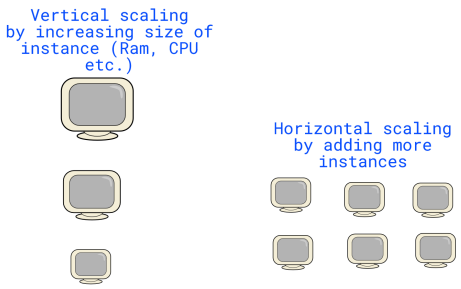

Figure 7.6: NoSQL

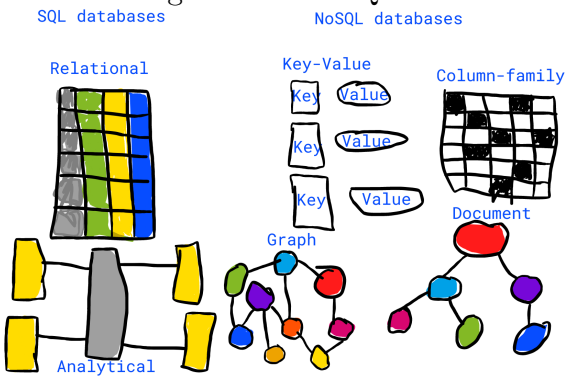

Figure 7.7: security

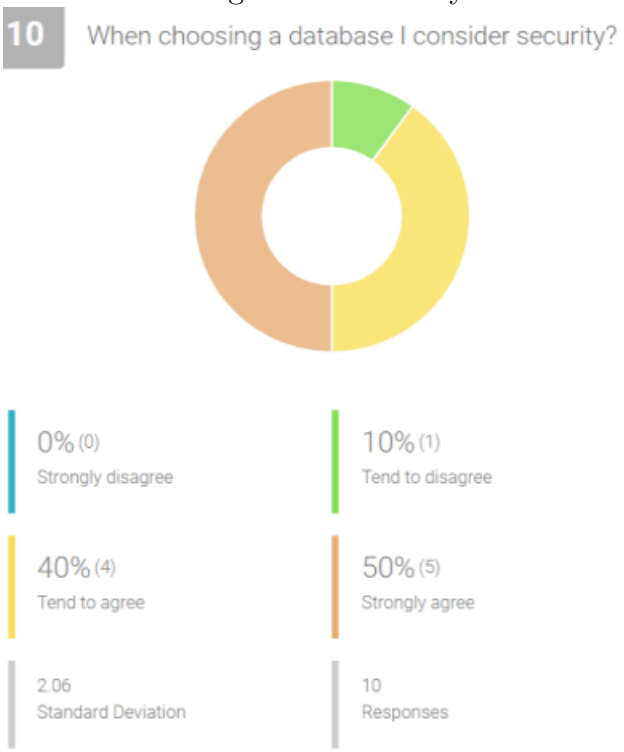

Figure 7.8: consistency

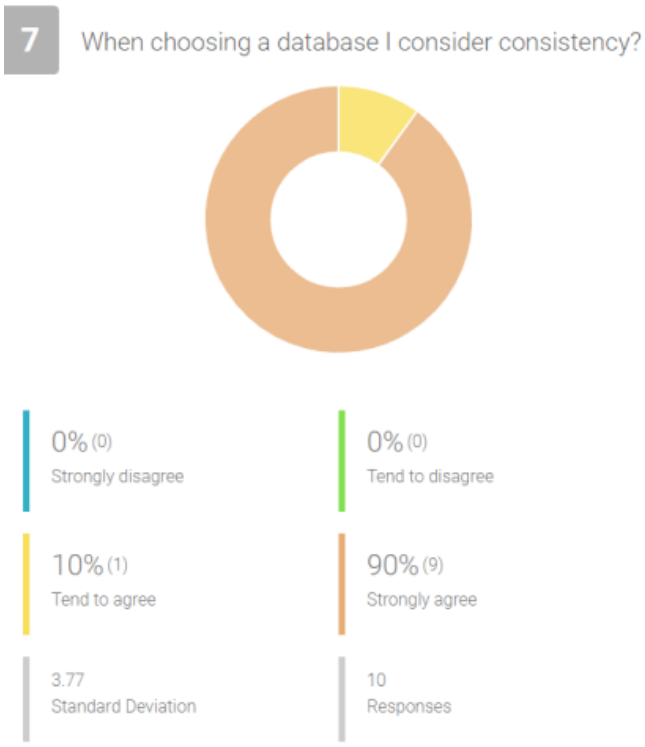

Figure 7.9: Infrastructure