Aalto-yliopisto
Aalto-universitetet
Aalto University

Bachelor's Programme in 2024

How database normalization and schema versioning improve system efficiency, data integrity, and flexibility in changing business settings

Mateus Pykälistö

Bachelor's thesis
2024

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Mateus Pykälistö

Title of thesis **How database normalization and schema versioning improve system efficiency, data integrity, and flexibility in changing business settings**

Programme Bachelor's Programme in Business

Major Information and Service Management

Thesis supervisor Principal university lecturer Johanna Bragge

Thesis advisor(s) Johanna Bragge

Date 27.05.2024 **Number of pages** 29 **Language** English

Abstract

With an emphasis on major corporations like Nokia, this thesis investigates the effects of database normalization and schema versioning on system effectiveness, data integrity, and flexibility inside database management systems. The study investigates the implementation issues and successful techniques of these technologies through semi-structured interviews with database administration specialists.

The results show that sophisticated schema versioning and normalization greatly improve data management by enabling systems to adjust flexibly to new requirements without sacrificing efficiency. The study highlights several significant obstacles, including the difficulty in implementing schema modifications and the limitations imposed by company regulations. One of the highlighted effective solutions is the combination of schema management tools and version control systems to enhance traceability and rollback capabilities.

Keywords database normalization, schema versioning, schema evolution, Relational Database Management Systems

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Mateus Pykälistö

Työn nimi Miten tietokantojen normalisointi ja skeemaversiointi parantavat järjestelmän tehokkuutta, datan eheyttä ja joustavuutta muuttuvissa liiketoimintaympäristöissä

Koulutusohjelma Kauppatieteiden kandidaattiohjelma

Pääaine Tieto- ja palvelujohtaminen

Vastuuopettaja/valvoja Johtava yliopistonlehtori Johanna Bragge

Työn ohjaaja(t) Johanna Bragge

Päivämäärä 27.05.2024 **Sivumäärä** 29 **Kieli** Englanti

Tiivistelmä

Tässä opinnäytetyössä tutkitaan tietokantojen normalisoinnin ja skeemaversiointien vaikutuksia järjestelmien tehokkuuteen, datan eheyteen ja joustavuuteen tietokannanhallintajärjestelmissä, erityisesti suurissa yrityksissä kuten Nokia. Tutkimus selvittää näiden teknologioiden käyttöönottovaikeuksia ja onnistuneita menetelmiä teemahaastattelujen avulla tietokannanhallinnan asiantuntijoiden kanssa.

Tulokset osoittavat, että kehittyneet skeemaversioinnit ja normalisointi parantavat merkittävästi datanhallintaa mahdollistamalla järjestelmien joustavan sopeutumisen uusiin vaatimuksiin ilman tehokkuuden heikkenemistä. Tutkimus korostaa useita merkittäviä esteitä, mukaan lukien skeemamuutosten toteuttamisen vaikeudet ja yrityssääntöjen asettamat rajoitukset. Yksi esille tuoduista tehokkaista ratkaisuista on skeemanhallintatyökalujen ja versionhallintajärjestelmien yhdistäminen parantamaan jäljitettävyyttä ja palautuskykyä.

Avainsanat tietokantojen normalisointi, skeemaversiointi, skeeman evoluutio, relaatiotietokannanhallintajärjestelmät

Aalto-yliopisto
Aalto-universitetet
Aalto University

Table of contents

List of Figures and Tables

Glossary

AI	Artificial Inteligence
BCNF	Boyce-Codd Normal Form
BiDEL	Bidirectional Database Evolution Language
Database schema	Structure of a database
DBMS	Database Management Systems - is a software in modern computing, enabling data organization, storage, and retrieval
Denormalization	Adding redudancy to a database to increase performance
InVerda	A multi-schema-version database management system (MSVDB) for agile database development
MSVDB	Multi-schema-version database management - realizes co-existing schema versions within one database
NF	Normal Form
nNF	number Normal Form
Normailization	Is the process of effectively arranging data in the database to reduce redundancy and improve data integrity
OLTP	Online Transaction Processing – type of data processing that executes a concurrent amount of transactions e.g online banking
PRISM	Panta Rhei Information & Schema Manager
RDBMS	Relational Database Management Systems - organize data in tables that have unique identifiers that link to one other
Schema Drift	a situation in which schema evolution results in divergences that can make data integration and analytics more difficult

Schema evolution	Is the process of modifying a database's structure over time to meet new needs without interfering with ongoing data activities
Schema versioning	Allows changes to the database schema without jeopardizing the integrity of already-existing data
SSMS	SQL Server Management Studio

1 Introduction

Database management systems (DBMS) is a software which is pivotal in modern computing, enabling efficient and reliable data organization, storage, and retrieval (Chavan & Shaikh, 2022). This thesis explores schema evolution in relational database management systems (RDBMS), addressing the critical need for adaptability and longevity amid evolving business and technology landscapes. Relational database management systems are designed to organize data in tables that have unique identifiers that link to one other. They are based on the relational paradigm.

Schema evolution is the process of modifying a database's structure over time to meet new needs without interfering with ongoing data activities, such as transaction processing, data querying, and database maintenance. However, schema evolution presents a plethora of challenging considerations that affect broad aspects of a software system such as data integrity, system performance, as well as forward-and backward-version compatibility.

Adapting database schemas to accommodate new requirements while ensuring that the database is reliable, and that it can be efficiently manipulated through transactions presents numerous theoretical and practical challenges. These challenges are heightened by the tension between the benefits of database normalization (minimizing redundancy and ensuring data integrity) and the flexibility needed to support evolving requirements (Diederich & Milton, 1988).

In this setting, knowledge of schema versioning and normalization is crucial. Normalization is the process of effectively arranging data in the database to reduce redundancy and improve data integrity, whereas schema versioning allows changes to the database schema without jeopardizing the integrity of already-existing data. These ideas are fundamental to understanding how schema evolution can be in line with version control and normalization theories for database systems to preserve minimal redundancy, system performance, and data integrity.

The theoretical motivation stems from the essential role of DBMS in contemporary enterprises, focusing on the impact of schema versioning (De Castro et al., 1997) and normalization on system performance, redundancy, and data integrity (Chavan & Shaikh, 2022). These concepts are vital for understanding the compatibility of schema evolution with database system version control and normalization theories (Beeri et al., 1978).

The practical difficulties presented by RDBMS schema modifications became evident during my internship in Nokia, exposing a lack of comprehensive approaches for successful and efficient schema evolution. Therefore, from a managerial perspective, this research highlights the practical significance of database normalization and schema versioning. It examines how these practices can lead to cost reductions, improved data quality, and enhanced business adaptability (Ramu, 2023), while also considering their limitations. This dual perspective ensures a comprehensive exploration of schema evolution, aiming to develop more adaptable and efficient database systems.

This research is motivated by the need to address these challenges through a comprehensive examination of existing strategies and methodologies for schema evolution, with an emphasis on the role played by database normalization in the adaptability of schemas (Batini et al., 1986). Additionally, this research seeks to contribute to the development of more adaptable schemas by exploring new techniques and methodologies which can improve the state of database systems.

1.1 Research Objectives and Questions

The thesis sets out to accomplish the following objectives, through navigating the intricacies of schema evolution. The main objective is to identify and categorize the foremost challenges associated with schema evolution in Relational Database Management System (RDBMS) with specific emphasis on understanding the interaction between schema versioning and database normalization.

There are also subobjectives, from this main objective that the thesis will address, when discussing the process of schema evolution. One of these objectives is to explore the interaction of database normalization levels with schema evolution and investigate how normalization levels affect schema flexibility and manageability. Furthermore, the thesis will investigate new solutions to optimize the schema evolution process including the potential for automation and artificial inteligence (AI) driven strategies to enhance practices for schema evolution.

Leveraging data as a strategic asset requires database designs to be able to be efficiently adjusted in response to changing requirements without sacrificing data quality or performance. This thesis aims to shed light on advanced database management practices and promote the development of systems that can navigate the dynamic nature of today's technological and business environments through a thorough examination of the complexities surrounding schema evolution and the methodologies developed to address these challenges.

Research Questions:

1. How do schema versioning and normalization impact system efficiency and data integrity?
2. What are the practical challenges in schema evolution?
3. What effective strategies can be employed for managing schema versioning?
4. What are the limitations of current methods regarding schema versioning and normalization?
5. How can the integration of AI and automation technologies enhance schema versioning and normalization practices?

These questions aim to deepen the understanding of schema evolution's role in modern database systems, exploring both the theoretical frameworks and practical applications of these critical concepts.

1.2 Scope and Limitations of the Research

Although the primary emphasis of this study is relational database management systems, in which schema evolution is crucial, the conclusions and suggestions may also be helpful for non-relational databases and upcoming technological advancements. However since technology is changing so quickly, certain solutions would need to be modified in order to be applicable in other situations.

Moreover, it is important to notice that every business scenario is unique when creating a Relational Database Management System (RDBMS). This means that apart from the issues that will be discussed in this thesis, it is important to notice that there might be other issues related to each specific business that they will have to address when considering schema evolution through normilization and schema versioning.

1.3 Structure of the Research

In the first chapter the thesis introduces the literature on database management, schema evolution, and normalisation to open the thesis. It allows the reader to understand better the motivations and objectives of this thesis.

Chapter 2 moves on to a critical evaluation of the approaches used currently to manage schema change, through a literature review of existing studies that underly the theories of Relational Database Management Systems (RDBMS). The third chapter describes the methodology of the research, here interviews

are conducted with database specialists that provides us a theoretical discussion and offers a practical approach to addressing the obstacles associated with schema evolution.

The fourth chapter will analyse the answers from the database specialists in the interviews, allowing us to draw out the key challenges of schema evolution in RDBMS. Finally, an evalutation of the insights acquired from the interviews and prior studies will summarise the final conclusion discussed in this thesis. Also, it will offer suggestions for future database schema management research topics as well as useful advice for database practitioners.

2 Literature review

Examining database normalisation and schema versioning in relation to database management systems (DBMS) exposes a wide range of academic research and methodology. For example, using effective key words such as “database normalization” or “database schema evolution” in reliable database sources, which include Scopus, Google Scholar, and international conferences on Very Large Databases, reveals numerous scholarly articles and papers. This literature study acknowledges the significant technical changes affecting DBMS and pulls from seminal works and recent investigations.

Technical variability makes it difficult to directly apply findings everywhere, but the fundamentals of data integrity and schema management provide vital insights into efficient DBMS operation. Nevertheless, the underlying theories of data management and schema architecture provide invaluable insights. These realizations serve as the cornerstone around which contemporary DBMS are constructed and refined, guaranteeing that they efficiently satisfy the changing needs of users and applications.

2.1 Schema Evolution in DBMS

Schema evolution is essential for managing how databases adjust to new application requirements without affecting current data and system functionality. In order to maintain operational continuity and data integrity throughout the database lifecycle, Roddick's groundbreaking work highlights the necessity for intelligent systems that can dynamically handle schema modifications. It also alluded to the difficulties in integrating these solutions into current IT infrastructures without causing major issues (Roddick, 1995).

Furthermore, studies on schema evolution—such as those by De Castro, Grandi, and Scalas—highlight how important it is to effectively manage schema changes to accommodate queries for data that are both historical and prospective (De Castro et al., 1997). Also, the database schema integration approaches that Batini, Lenzerini, and Navathe covered illuminated the complexities of combining different schemas and recommended creative approaches to ensure seamless evolution and integration (Batini et al., 1986).

Automating complex schema evolution processes is a paradigm shift that has been demonstrated by the development of systems such as Panta Rhei Information & Schema Manager (PRISM). As a result of these systems' smooth integration with current database operations and substantial reduction of the manual workload related to schema modifications, there is an increasing demand for solutions that combine stability and agility (Curino

C. et al., 2008; Curino et al., 2013). Nonetheless, these systems frequently operate under the assumption of a static usage pattern, which might not accurately reflect how quickly corporate data is evolving.

Additionally, the coexistence of various schema versions within a same database is made possible by technologies like InVerDa and Bidirectional Database Evolution Language (BiDEL), which highlight a significant trend towards flexibility and allow for ongoing development and deployment in agile business environments (Herrmann et al., 2017, 2018). Although, it further calls for a closer investigation of the trade-offs between system complexity and schema flexibility.

Schema drift—a situation in which schema evolution results in divergences that can make data integration and analytics more difficult—and continuous maintenance expenses must be taken into consideration when analysing the strategic role that schema evolution plays in business flexibility. There is a clear research gap on cost-effective schema versioning solutions because the literature does not provide a thorough overview of controlling these costs and limiting hazards.

2.2 Database Normalization and Its Implications

Reducing data redundancy and enhancing data integrity have been made possible by database normalisation, which Codd primarily introduced (Codd, 1975). Diederich and Milton's historical development of normalisation theories, from first normal form (1NF) to Boyce-Codd Normal Form (BCNF), demonstrates how to strike a compromise between theoretical rigour and real-world application (Diederich & Milton, 1988). Additionally, it offers a strict framework that could not meet all analytical requirements. There may be an inconsistency between conventional normalization forms and current requirements due to the dynamic nature of data types and query requirements in modern systems.

Modern database applications are dynamic, which means that sophisticated normalisation methods are required. Automated normalisation methods, as proposed by Bahmani et al. (2008), greatly reduce manual labour and human error risk in large-scale databases by facilitating the efficient production of primary keys and normalised tables. However, this automation raises concerns about loss of flexibility when managing different kinds of data and overfitting in schemas designed for certain datasets.

Additionally, the application of normalisation in modern relational databases is improved by Köhler & Link's recent work, which expands on the concepts of normalisation to solve SQL-specific issues like null markers and duplicate

tuples (Köhler & Link, 2016). Modern database systems are significantly impacted by the developments in database normalisation, especially in terms of improving data integrity and operational effectiveness.

In big data applications, where denormalization may be preferable due to speed advantages, modern database systems must strike a balance between the need for performance optimization and normalization. More investigation into adaptive normalization strategies that may adapt dynamically to data usage patterns and query performance measurements is encouraged by this conflict between theoretical normalization concepts and real-world performance requirements.

2.3 Innovations in Schema Management

Significant advancements in database schema management have been made in the context of contemporary DBMS with the goal of improving performance and flexibility. The ongoing attempts to combine conventional database administration methods with cutting-edge optimisation frameworks are exemplified by Ramu's (2023) research on reconsidering schema normalisation inside optimisation methodologies. His method emphasises how dynamic database administration is and how systems must be able to adjust to shifting performance requirements.

The complexity of managing schema versions increases as databases play a more and bigger role in corporate operations. This challenge is addressed by the development of multi-schema-version database systems (MSVDB), which offer mechanisms that permit the coexistence and parallel evolution of multiple, potentially divergent schema versions. This improves the adaptability of databases to a wide range of quickly changing business needs (Herrmann et al., 2017, 2018).

Automation of activities connected to schemas is a major step towards streamlining operations and increasing productivity. Schema evolution is made easier by systems like PRISM and InVerDa, which also make sure that these modifications are smoothly incorporated into the current database frameworks to maintain consistency and integrity. Yet, relying too much on automation raises questions about the loss of human oversight in vital data processing procedures where a sophisticated grasp of business contexts and data linkages is essential.

In the future, the incorporation of cutting-edge technology like AI into schema management holds the potential to significantly transform this industry. These technologies have the potential to revolutionise how businesses handle and use their vital data resources by providing previously

unheard-of levels of automation, security, and efficiency. On the other hand, unanswered questions about data privacy, moral AI applications, and the viability of decentralized data management in conventional commercial settings are also raised by this combination.

3 Methodology

This study aims to assess how well database normalisation and schema versioning contribute to improved system efficiency, data integrity, and adaptability to future changes in business relationships and models. A thorough analysis of optimised and normalised database management system architectures will be part of this assessment, with an emphasis on how these approaches might change to meet changing business requirements.

The empirical part of the research will involve conducting in-depth semi-structured interviews with database experts, whom one will be a representative from Nokia. My previous internship at Nokia helped me with my choice of company, allowing me to get firsthand knowledge with the difficulties involved in database development, particularly with regard to schema versioning and normalisation. Understanding Nokia's strategy for handling these database difficulties is quite helpful since it provides a real-world viewpoint on putting theoretical database management ideas into practice at a company that operates at a global level with communication infrastructure.

3.1 Method

Four semi-structured interviews were carried out with database specialists that have managerial roles in a global technology company. This approach promotes detailed answers to delve into the nuances of schema evolution and normalisation. Since semi-structured interviews enables respondents to share in-depth personal experiences, obstacles they have faced, and solutions they have found for schema development and normalisation, this interview approach is especially well-suited for exploring intricate subjects like database management (Beeton, 2005). This strategy fits with the goal of the research, which is to collect detailed data that clarifies the intricacies of DBMS optimisation.

Semi-structured interviews were selected due to their ability to provide extensive exploration of the many nuances of database management, including a detailed discussion of the experts' own experiences. These talks greatly advance comprehension by offering insightful perspectives on the real-world difficulties and creative solutions used in schema evolution and normalisation. Furthermore, being on one-on-one sessions allows more space for them to share their annecdotes regarding database management, and for this reason I chose this method instead of a questionnaire, survey or focus group.

3.2 Data Collection and Analysis

The interviews took place remotely, via a video call. Additionally, recordings were made with the permission of the respondents, to eliminate the possibility of errors in data collection. In this case only one respondent agreed to share their information, and in total four database specialists where interviewed.

One of the interviewees is Rio Wibowa, who is an Product Manager in Nokia and has dealt with databases for more than 20 years. The data obtained from the interviews will be grouped thematically to identify the common issues and strategies implemented among the specialists.

3.3 Limitations

This section recognises the inherent limits of the selected techniques. Although expert insights from interviews are valuable, they may not fully capture all aspects of the DBMS optimisation difficulty due to individual perspectives. While providing empirical data, the interviews mostly captures the perspective of the work environement in which the interviewee deals with daily.

Therefore, it is important to note that the DBMS discussed in the interviews were mostly directed to smaller transactional databases which are used in a more daily bases within business operations. Nevertheless, the empirical data collected through interviews still aim to provide a thorough understanding of the ways in which schema versioning and database normalisation can lead to more effective, dependable, and flexible DBMS.

4 Results

Using semi-structured interviews with database administration specialists in a manager role at Nokia and at a global technology company, allows the empirical analysis to produce useful insights to the study topics that were first presented in the thesis. These results are shown in this part along with a discussion of their implications and the validity of the study's methodology. With the insights gained from the interviews, we will examine each topic in more detail in the sections that follow, focusing on the difficulties that are unique to it.

4.1 Impact of Schema Evolution and Normalization on System Performance and Data Integrity

It is clear from the interviews that Nokia's schema evolution and normalisation methods have a big impact on data integrity and system performance. Rio Wibowo outlined how, in order to reduce disruptions during schema updates, contemporary schema rules have moved from strict, hardcoded schemas to more flexible schema evolution solutions. This development helps the business adjust to shifting needs without incurring large costs or jeopardising data integrity (Rio Wibowo, 2024).

According to respondent 2, although large enterprises incorporates sophisticated third-party technologies to efficiently manage schema evolution, company policies often prohibit installing specific solutions, which might impede the effectiveness of managing schemas. The interviewee did, however, bring up the possibility of integrating GitHub with SQL Server Management Studio (SSMS), which would enhance the tracking of changes in SQL scripts as multiple users can work on the same Git repository (Respondent 2, 2024).

An additional viewpoint on the practical difficulties encountered in database management was provided by respondent 3. In particular, using integer IDs for systematisation inside the schema, which improves query efficiency by simplifying the handling of joins and other relational database operations, was highlighted (Respondent 3, 2024).

Respondent 4 stressed the advantages of calculating data only once for database storage, which reduces CPU usage and increases the flexibility of structure updates when data is centralized. Additionally, he pointed out that decreased susceptibility to evolutionary changes results from improved data integrity, however this can complicate joins when several sources of data are needed (Respondent 4, 2024).

4.2 Challenges in Schema Evolution

Rio Wibowo highlighted the difficulties associated with schema evolution, pointing out that prior to the adoption of more dynamic schema management tools, any schema change required extensive manual revisions and testing, which was time-consuming and error-prone (Rio Wibowo, 2024). This challenge has been addressed at Nokia by implementing schema versioning tools that enhance traceability and rollback capabilities, enabling more seamless updates and transitions.

From respondent 2 it is evident that the practical difficulties are associated with renaming database elements when changes are required across the database and renaming occurs late in the development process. The interviewee shared a personal anecdote about a simple typo in a column name that was discovered late, illustrating the difficulties and meticulousness involved in renaming (Respondent 2, 2024).

Respondent 3 spoke about the challenges posed by schema modifications, which include adding or changing primary keys inside a database and requiring significant system revisions all around (Respondent 3, 2024).

These issues were like the ones from respondent 4, that talked about dealing with complicated changes that were hard to handle and frequently made identifying possible problems more difficult. To address these, the interviewee made small adjustments to reduce concerns and conducted ongoing testing to make sure the changes did not cause new ones.

4.3 Effective Strategies for Managing Schema Versioning

The combination of version control with schema management tools was shown to be one of the most effective ways for managing schema versioning in a business environment that is changing quickly. Rio Wibowo talked on how tools that interface with version control systems like GitHub and SSMS are crucial for managing changes because they allow for traceability. Better tracking of the effects of schema changes and management of them have been made possible by this integration (Rio Wibowo, 2024).

Respondent 2 raised the possibility of integrating CoPilot-type features into tools like SSMS (Respondent 2, 2024). This might completely change the way schemas are managed by making the procedures more streamlined and unified.

Respondent 3 identified the creation of procedures that have rollback

methods that are easier to debug when a transaction log (see Table 1) is created. Furthermore, he emphasised how to effectively manage changes by using integer IDs and consistent naming rules, which lessens the strain on system performance and simplifies the maintenance of data integrity (Respondent 3, 2024).

Table 1: A Transaction Log Table example in a Database Management System (Fowler, 2016)

iRowID	sProcedureName	sType	sTable	sMessage	tBeginTime
1	pUpdateGrades	INSERT	tGrade	Inserted 15 rows in table tGrade	2023/01/10 20:30:80
2	pUpdateStudents	INSERT	tStudents	Inserted 2 rows in table tStudents	2023/01/10 20:31:50
3	pUpdateStudentGrades	INSERT	tStudentsGrade	Inserted 15 rows in table tStudents	2023/01/10 20:32:50
4	pUpdateTeachers	INSERT	tTeachers	Rollback foreign key constrain	2023/01/10 20:34:50

Respondent 4 also agreed with this and added that clear communication between colleagues that are also working in the database is extremely important. This minimizes the risk of creating queries with incorrect syntax, and by making a copy of the procedures instead of overwriting them ensures that the queries will not corrupt the procedures (Respondent 4, 2024).

4.4 Limitations of Current Normalization and Versioning Methodologies

Rio also talked about the shortcomings of existing approaches, especially with regard to normalisation techniques, which, although useful in Online Transaction Processing (OLTP) systems, would not be the best for analytical systems where denormalization could lead to fewer join operations and hence better performance. The interview highlighted the need for more flexible and adaptable normalisation techniques that can meet the unique needs of various kinds of database systems (Rio Wibowo, 2024).

The difficulties in standardising schema administration were underlined by Respondent 2, especially when corporate constraints restrict the tools available for efficient schema evolution and management (Respondent 2, 2024).

Respondent 3 talked about how hard it is to keep the flexibility needed to quickly adjust to shifting business needs inside of a strict normalisation framework, pointing to the necessity for more flexible and dynamic approaches (Respondent 3, 2024).

This was further emphasized by respondent 4, who suggested that in a dynamic business environment we must be aware of the security levels of a third-party tool when connecting it to a database (Respondent 4, 2024).

4.5 Future of Database Management Systems

During the interview with Rio Wibowo, we talked about how machine learning (ML) and artificial intelligence (AI) might be included into schema maintenance and normalisation in the future. He expressed hope that these technologies will further the automation and improvement of database management procedures (Rio Wibowo, 2024).

Respondent 2 predicted that AI will be relevant in code generation and optimisation, and proposed a major movement in database administration methods towards automation over the course of the next five years (Respondent 2, 2024).

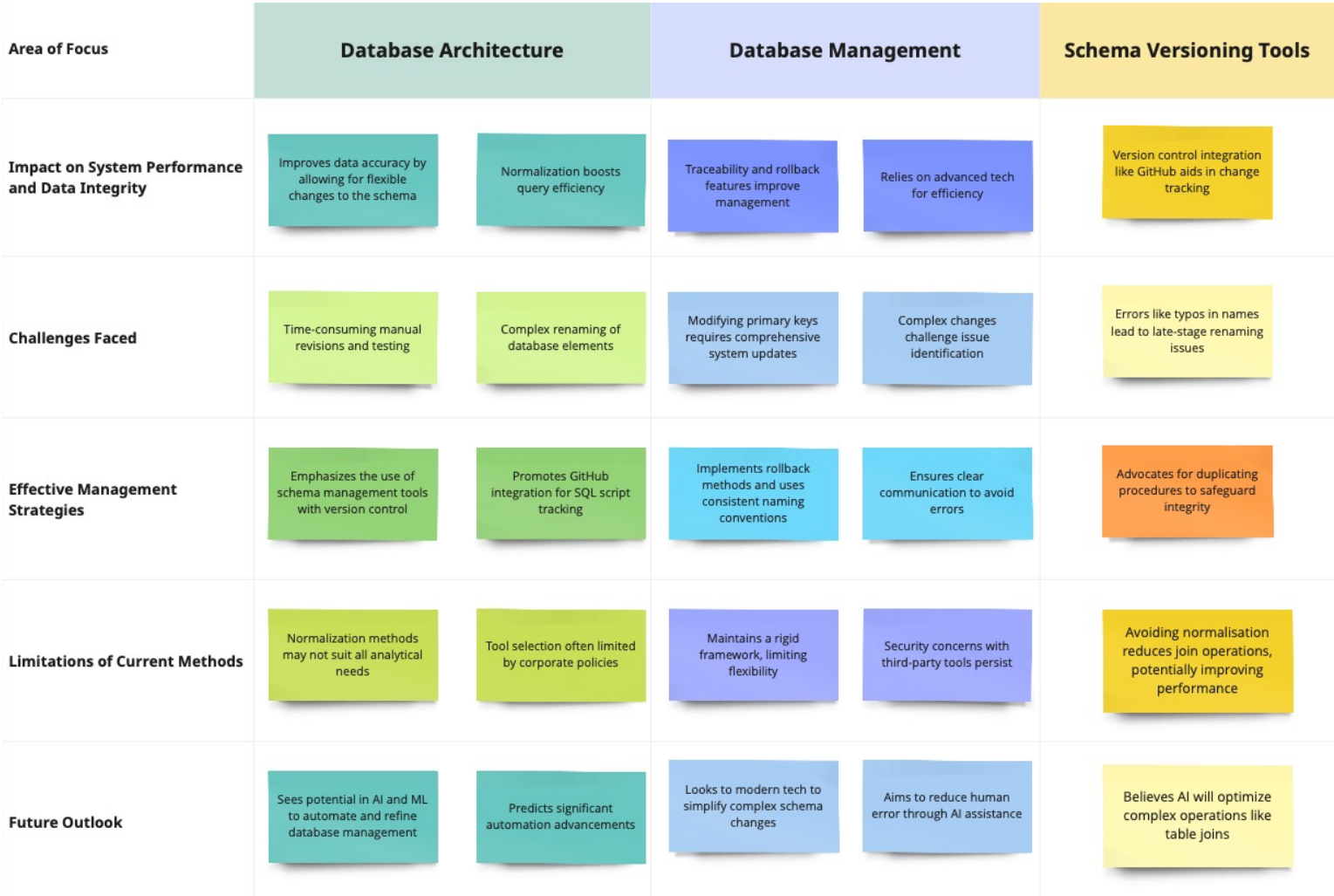
Respondent 3 also considered how modern technologies, including AI, might be integrated to manage complex schema modifications, potentially streamlining the procedure and lowering the chance of human error (Respondent 3, 2024). An example of this was suggested by respondent 4 that it could help with solving issues of joining tables where are the are multiple keys involved (Respondent 4, 2024).

4.6 Summary of Results

Figure 1 provides an organised overview of the important points of schema evolution and normalisation from the interviews, which is broken down into three categories: Schema Versioning Tools, Database Architecture, and Database Management. Each category's problems are grouped according to the main database management component from which they stem.

Many of the problems are cross-cutting, as the chart shows, meaning that solutions are frequently mutually reinforcing and advantageous to various facets of database administration. This underlines the possibility of advancements that could increase data integrity, system performance, and management effectiveness on a global scale.

Figure 1: Key aspects of schema evolution and normalization from the interviews.



5 Discussion and conclusions

The results illustrate numerous important insights into how database normalisation and schema versioning improve system efficiency, data integrity, and flexibility in changing business settings. These insights are based on semi-structured interviews with database administration specialists in a mangerial level at Nokia and at a global technology company. These findings complement and expand upon previous research, providing a thorough grasp of the difficulties and practical solutions in contemporary DBMS management.

The results of the interviews show that data integrity and system performance are much improved by Nokia's schema evolution and normalisation procedures. Rio Wibowo highlighted the move away from strict, hardcoded schemas and towards more adaptable schema evolution solutions, which minimise update disruptions and let the company adjust to new requirements without spending a lot of money (Rio Wibowo, 2024). This is consistent with the body of research indicating that flexible schema management can enhance the overall resilience of a system (Curino C. et al., 2008; Curino et al., 2013).

Although the usage of some tools may occasionally be restricted by company policy, Respondent 2 emphasised the need of advanced third-party technology in effectively controlling schema evolution (Respondent 2, 2024). The necessity of combining GitHub and SQL Server Management Studio (SSMS) to improve the tracking and management of schema modifications is shown by this real-world scenario (Herrmann et al., 2017, 2018).

The difficulties involved in evolving a schema, especially the laborious manual work necessary to make modifications, are similar to the difficulties described by Batini et al. (1986) in merging various schemas. These difficulties highlight the necessity of innovative and practical management techniques to guarantee smooth development and integration, demonstrating the continued applicability of fundamental database administration concepts in modern workflows.

Furthermore, it was seen that query efficiency was increased by using integer IDs for systematisation within the schema, as this made joins and other relational operations simpler. This useful realisation validates previous studies on the advantages of effective key management for improving database efficiency (Bahmani et al., 2008)

Significant obstacles to the evolution of schemas were also revealed by the interviewees, mainly the lengthy process of manual testing and changes prior

to the adoption of more dynamic schema management systems. This result supports previous work by (Roddick, 1995) and emphasises the need for tools that improve rollback and traceability in order to enable smooth updates and transitions.

The practical challenges of renaming database elements at the end of the development process and the attention to detail needed to prevent mistakes were also covered. These difficulties highlight the necessity of strict naming guidelines and version control to avoid problems and guarantee consistency throughout the database (De Castro et al., 1997)

The results also explore how AI-driven technologies such as CoPilot may be integrated into schema administration, pointing to a future trend where automation could greatly simplify and harmonize schema management procedures. This is consistent with research on AI's ability to improve database management effectiveness (Herrmann et al., 2018).

5.1 Implications to Research

This work contributes to the small body of research on the real-world uses of schema versioning and database normalisation in major corporations such as Nokia. It emphasises how crucial it is to integrate cutting-edge technologies with adaptable schema management tools in order to meet the changing demands of contemporary enterprises. Future studies should examine these solutions' long-term effects on data integrity and system performance, especially in settings with strict regulatory requirements and a variety of data sources.

The study also identifies gaps in the literature concerning the possibility of schema drift in constantly evolving databases and the cost-effectiveness of various schema management techniques. Filling in these gaps could lead to more profound understanding of how to make schema versioning procedures more efficient and maintain system performance.

5.2 Implications to Practice

The results highlight for practitioners how crucial it is to have dynamic, adaptable schema management solutions that work well with current systems. AI-driven tools and version control systems can improve traceability, lower mistakes, and expedite the schema evolution process. These tools should be taken into consideration by organisations in order to enhance their database administration procedures and better adjust to evolving business needs.

Furthermore, managing complicated schema updates and preserving data integrity depend heavily on clear communication and consistent naming standards. In order to guarantee that every member of the team is in agreement with optimal practices, practitioners should set up strong schema management protocols that include comprehensive documentation and frequent training.

5.3 Limitations and Future Research

While offering insightful information, this study is restricted by its small sample size and its concentration on a specific corporate setting, which may limit the generazilability used in different sectors or at different organisational sizes. Subsequent studies could broaden the scope of this analysis by embracing a more diverse range of viewpoints from various industries and potentially applying quantitative techniques to confirm and enhance the qualitative conclusions put forth.

In addition, although the study emphasises the advantages of AI and machine learning for schema management, it also poses concerns regarding data security and privacy. In order to create best practices for incorporating cutting-edge technology into schema management without jeopardising data security, future research should look more closely into these issues.

Furthermore, as proposed by Ramu (2023), a promising direction for future research is the combination of conventional normalisation techniques and state-of-the-art optimisation frameworks. It will be essential to investigate how these cutting-edge strategies might improve database systems' performance and adaptability, particularly as companies continue to manage massive and complicated data sets.

References

Bahmani A. H., Naghibzadeh, & Bahmani B. (2008). Automatic Database Normalization And Primary Key Generation. *Canadian Conference on Electrical and Computer Engineering*, 11–16.

Batini C., Lenzerini M., & Navathe S. B. (1986). A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys*, *18*(4), 324–364.

Beeton S. (2005). The case study in tourism research: A multi-method case study approach. *In B. W. Ritchie, P. Burns, & C. Palmer (Eds.), Tourism research methods: Integrating theory with practice* (pp. 37-48). CABI.

Beeri C., Bernstein P. A., & Goodman N.(1978). A sophisticate's introduction to database normalisation theory. *Very Large Data Bases (VLDB)*, *4*, 113–124.

Chavan H., & Shaikh S. (2022). *Introduction to DBMS: Designing and Implementing Databases from Scratch for Absolute Beginners*. BPB Online. www.bpbonline.com

Codd E. F. "Recent Investigations into Relational Data Base" in *Proc. 1974 Congress* (Stockholm, Sweden, 1974). New York, N.Y.: North-Holland (1974).

Curino C., Moon H. J., & Zaniolo C. (2008). Graceful Database Schema Evolution: the PRISM Workbench. *Proceedings of the VLDB Endowment*, *1*(1), 761–772. https://doi.org/10.14778/1453856.1453939

Curino C., Moon H. J., Deutsch A. & Zaniolo, C. (2013). Automating the database schema evolution process. *VLDB Journal*, *22*(1), 73–98. https://doi.org/10.1007/s00778-012-0302-x

De Castro, Grandi, & Scalas. (1997). Schema versioning for multitemporal relational databases. *Information Systems*, *22*(5), 249–290.

Diederich J., & Milton J. (1988). New Methods and Fast Algorithms for Database Normalization. *ACM Transactions on Database Systems*, *13*(3), 339–365.

Fowler K. (2016). *Data breach preparation and response : breaches are certain, impact is not*. Syngress.

Herrmann K., Voigt H., Behrend A., Rausch J., & Lehner W.(2017). Living in parallel realities - Co-existing schema versions with a bidirectional database evolution language. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, *Part F127746*, 1101–1116. https://doi.org/10.1145/3035918.3064046

Herrmann K., Voigt H., Pedersen T. B., & Lehner W. (2018). Multi-schema-version data management: data independence in the twenty-first century. *VLDB Journal*, *27*(4), 547–571. https://doi.org/10.1007/s00778-018-0508-7

Köhler H., & Link S. (2016). SQL schema design: Foundations, normal forms, and normalization. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, *26-June-2016*, 267–279. https://doi.org/10.1145/2882903.2915239

Ramu V. B. (2023). Optimizing Database Performance: Strategies for Efficient Query Execution and Resource Utilization. *International Journal of Computer*

Trends and Technology, *71*(7), 15–21.
https://doi.org/10.14445/22312803/ijctt-v71i7p103
Roddick J. F. (1995). A Survey of Schema Versioning Issues for Database Systems. *Information and Software Technology*, *37*(7), 383–393.

Interviews

Rio Wibowo - personal communication, April 8th 2024
Respondent 2 - personal communication, April 10th 2024
Respondent 3 - personal communication, April 12th 2024
Respondent 4 - personal communication, April 15th 2024

Appendix

Interview Questions

1. **Name:**

2. **Current position in the company**:

3. **Experience with databases and database projects**:

4. **How do schema evolution and database normalization practices impact system efficiency and data integrity in your experience?**

5. **Can you describe a challenge you faced related to schema evolution and how you addressed it?**

6. **What strategies have you found most effective for managing schema versioning in rapidly changing business environments?**

7. **In your opinion, what are the limitations of current normalization and versioning methodologies within DBMS?**

8. **How do you foresee the evolution of database management systems in terms of schema management and normalization in the next five years?**