SEVEN publicações acadêmicas

Patterns in data modeling: addressing common patterns and best practices

https://doi.org/10.56238/innovhealthknow-026

Arleson João Serrão Carvalho
Graduating in Software Engineering from UEPA.
E-mail: arlesonc85@gmail.com

Caroline Nunes Carr
Specialist in Professional and Technological Education, IFES and Professor at UEPA.
E-mail: carolinecarr@uepa.br

Emerson Leandro da Silva Silva
Graduating in Software Engineering from UEPA.
E-mail: leandroclash203040@gmail.com

Lucas Victor Menezes Pereira
Graduating in Software Engineering from UEPA.
E-mail: lucasvictormenezes0@gmail.com

Urbano da Luz Ferreira Junior
Graduating in Software Engineering from UEPA.
E-mail: ulfjunior@hotmail.com

ABSTRACT
This article addresses the importance of standardization in data modeling, as well as its impact on the quality of DBMS. Standardization encourages uniformity by following fixed guidelines, which facilitates comprehension and collaboration among teams. Additionally, we discuss how the normalization of primary keys and indexes enhances data integrity. Reliability is addressed, highlighting security, referential integrity, and traceability. Furthermore, the article under discussion covers best practices such as selecting the best data and adapting standardization to sources of questionable provenance, as inappropriate standardization can lead to errors, given that data might be incomplete precisely due to their origins from dubious sources. However, we acknowledge that selective standardization is crucial for complex data; we also address how standardization within data modeling contributes to a more efficient and accurate system. In summary, standardization in analysis reinforces confidence in the data. Nevertheless, it should be applied cautiously to preserve the original information.

Keywords: Data modeling, DBMS, standards and keys.

1 INTRODUCTION

In the vast arena of data modeling, it is imperative to adopt standards and best practices, especially when multiple hands shape a single project. The absence of standards can blur the line between effectiveness and inefficiency. Reflecting on the importance of standardization, Taiichi Ohno once remarked, "Where there is no standard, there can be no improvement."

Setting standards in data modeling is not just a matter of uniformity, but a necessity to create robust, high-performance systems. When there are pre-defined principles, the integrity of the information is preserved and enhanced, because all the data follows a common pattern. This, in turn, reduces the chances of ambiguities and conflicts during data analysis or interpretation.

In addition, by embracing standards and best practices, we are recognizing and avoiding common obstacles faced during data modeling. This translates into savings in time and resources, as well as a smoother learning curve for new team members. By recognizing the challenges that arise and

the standard solutions to address them, data modelers can focus their energy on more advanced innovations and optimizations.

The paper not only illuminates the challenges intrinsic to data modeling, but also suggests solutions based on proven design patterns. By diving into these concepts, we hope to continually facilitate and improve the intricate process of data modeling.

"You can't expect everything to be perfect, but you have to move forward in pursuit of continuous improvement."

Mário Sérgio Cortella, a Brazilian philosopher and educator.

2 THE IMPORTANCE OF STANDARDIZATION IN DATA MODELING

Incorporating standardization into data modeling is a crucial principle that goes beyond mere database structure design. It drives the quality and maintenance of database management systems (DBMS) more broadly, as well as driving efficiency and effectiveness. The essence of standardization lies in the robust application of guidelines, ensuring uniformity at every stage of the data lifecycle.

Moreover, when a set of rules and conventions are adopted, data models consequently become easier to understand and easier to predict. As a consequence, we have fewer double entendres which allows for better communication between the parties involved; Which generates a very important benefit: consistency in the design in the databases. In addition, standardization simplifies the regulatory maintenance of databases. Programmers have an interdependence of data elements (which decreases the risk of delays and errors). All this is only possible thanks to the coherent and super documented data structures, the updates and improvements that can be inserted in an agile and, above all, much more precise way.

Reasonableness is another solid foundation of standardization in data modeling. With this foundation, standardized data models can be applied to multiple projects, which leads to a decrease in time and resources. The standardization of the designs allows the structure to be reused, since it has already been evaluated previously. This makes it speed up development and ensures system quality more consistently. In addition, standardization has a direct influence on the performance and efficiency of databases. The application of practices such as standardization generates faster and more effective queries, which improves the end-user experience and increases the utilization of system resources.

Standardization in data modeling also plays a crucial role in ensuring data governance. Organizations establish a solid foundation for the management and control of their information assets through the definition and consistent application of standards. This causes the data to be treated in a better way. Another undeniable advantage of standardization is the facilitation of collaboration between professionals, since by adopting a standard, teams can work more efficiently, that is, standardization eliminates barriers to interpretation, which allows professionals from different areas to

extend the structure of the shared data. Additionally, standardization provides a solid basis for evolution over time. As companies/organizations grow, new requirements emerge, and standardized models make it easy to incorporate change without generating unwanted impacts. And consequently, it generates more flexible systems.

In addition, the implementation of standards in data modeling also has a positive effect on data quality. Given that when standardizing elements, the chance of errors decreases considerably; which results in more accurate and reliable data.

In summary, standardization in data modeling, in addition to influencing DBMS designs, also plays a key role in areas such as professional collaboration and systems evolution. By adopting this approach, companies/organizations form fixed foundations for continued success.

3 HOW STANDARDS IMPROVE DATA PERFORMANCE

Patterns in data modeling are recognized and tested solutions to recurring problems in this field, they are a way to ensure that information is stored in a consistent and standardized way, which can significantly improve database performance, helping to optimize data access, reducing redundancy and increasing the efficiency of queries. In addition, standards can help ensure data integrity by avoiding errors and inconsistencies, for example, a common standard is foreign key restriction, which ensures that related information in different tables is always synchronized.

Another common pattern is the use of indexes, which are data structures that allow you to quickly find specific information in the database, indexes can significantly improve query performance by allowing the database to quickly find the information it needs. In short, patterns are an important tool in data modeling, but how do Standards Improve Data Performance?

3.1 CONSISTENCY AND CLARITY

Standards establish consistent conventions and nomenclatures for data elements such as tables, columns, and relationships, this makes it easier for developers, administrators, and end users to understand the model, a clear and well-defined data structure reduces the possibility of misinterpretation, resulting in more accurate and efficient queries. Some common patterns include:

- **Consistent nomenclature**: Assign clear, descriptive names to model objects, such as tables, columns, primary and foreign keys. This helps to avoid confusion and ambiguity.

- **Capitalization conventions**: Use case consistently to differentiate object names. For example, use uppercase for table names and lowercase for column names.

- **Naming patterns**: Use consistent patterns for object names. For example, use the prefix "tbl" for tables and "fk" for foreign keys.

- **Formatting standards**: Use consistent patterns to format the template. For example, aligning objects vertically or horizontally.

- **Proper documentation**: Properly document the model with clear descriptions of the model objects, including tables, columns, primary and foreign keys.

3.2 NORMALIZATION

Normalization is an important process in data modeling that aims to improve the quality and efficiency of the database, it consists of applying a series of rules to eliminate redundancies and inconsistencies in the data, ensuring that each table contains only relevant information and that the relationships between the tables are clear and accurate, it has several normal forms, each with its own rules and goals. The first normal form (1FN) requires that each table have a unique primary key and that each column contain only atomic values, the second normal form (2FN) requires that each column depend completely on the primary key of the table, and the third normal form (3FN) requires that each column depend only on the primary key of the table, not on other columns. There are other normal forms, such as the fourth normal form (4FN) and the fifth normal form (5FN), but they are less common.

Normalization has several advantages, firstly, it helps to avoid unnecessary redundancies in the data, which can save storage space and improve database performance, secondly, it helps to ensure data integrity by avoiding inconsistencies and errors, thirdly, it makes it easier to add or modify data in the database without affecting other parts of the system.

However, normalization also has some disadvantages, firstly, it may be difficult to apply normalization rules correctly, especially in complex databases, secondly, it may be necessary to make multiple joins between tables to retrieve related information, which may affect database performance, and lastly, normalization may not be suitable for all use cases; In some cases, it may be more efficient to store redundant data to avoid excessive joins.

3.3 IDIDES AND PRIMARY KEYS

The standards recommend the use of indexes and primary keys to optimize the speed of data retrieval, indexes are structures that allow the quick location of records based on values of specific columns, while primary keys, ensure the uniqueness of each record in a table, speeding up the search and updating of data.

3.3 CONTROLLED STANDARDIZATION

Denormalization is the process of adding redundancy to a database to improve query performance and reduce the complexity of the data model, this is done by adding duplicate columns or helper tables that store information that already exists in other tables. Denormalization can be useful in situations where database performance is critical and redundancy does not impair data integrity, however, denormalization also has some disadvantages, firstly, it can lead to inconsistencies in the data if redundant information is not updated correctly, secondly, it can make the data model more complex and difficult to maintain, and thirdly, it can be difficult to determine when and where to apply normalization.

To minimize these problems, it is important to follow some best practices when applying denormalization:

- **Identify the most common** queries: Identify the queries that run most frequently and optimize the data model for those queries.
- **Add redundancy carefully**: Add only the information you need to improve query performance.
- **Maintain data integrity**: Ensure that redundant information is updated correctly when the underlying data changes.
- **Document properly**: Properly document the data model to make it easier for other team members to maintain and understand the model.

3.4 PARTITIONING AND ADVANCED INDEXING

Advanced indexing is a database optimization technique that allows you to improve the performance of queries against large tables, it is used to divide a large table into several smaller parts, called partitions, based on a specific criterion, such as the date or value of a column, this allows queries to run faster, because the database only needs to search the relevant partitions.

Advanced indexing is particularly useful in databases that need to handle large amounts of data, such as content management systems or big data analytics applications, and it can be used to improve the performance of queries against tables that are frequently updated.

Advanced indexing can be combined with other optimization techniques such as cluster indexing and bitmap indexing, clustered indexing involves physically organizing the data on the hard disk to minimize the number of reads required to retrieve the data, and bitmap indexing is a technique that uses bits to represent distinct values in a column and allows queries to run faster.

4 STANDARDS-RELATED RELIABILITY

Data reliability is a key factor for making assertive decisions. In this section, we will examine how adopting design patterns impacts the reliability of stored data. We will address issues such as referential integrity, consistency, and data validation. We will highlight how the correct implementation of standards reduces errors and avoids inconsistencies, increasing the reliability of the system as a whole.

It plays a crucial role in data modeling, directly influencing the credibility and usefulness of information systems. Strict implementation of design standards plays a vital role in ensuring the reliability of data. In this section, we will explore in depth how consistency in the application of standards contributes to system reliability, addressing notions of referential integrity, traceability, resilience to failure, protection of privacy, and data security.

Adriano Martins Antônio, in his article "Three fundamentals on information security that every professional needs to know", explores the foundations of information security that are essential for reliable practices in data modeling. In this article, Antônio highlights the importance of establishing solid standards to ensure the integrity and reliability of data (ANTONIO, 2017).

4.1 CONSISTENCY IN REFERENTIAL INTEGRITY AND DATA CONSISTENCY

Referential integrity is a crucial foundation for data reliability. By establishing links between tables through primary and foreign keys, design patterns ensure that connections between data remain uniform. For example, in a customer and order management system, referential integrity ensures that an order cannot be associated with a non-existent customer, preventing inconsistencies that could affect sales analysis and demand forecasts.

Antônio points out that referential integrity is one of the pillars of data reliability. Through the application of design patterns that establish relationships between tables through primary and foreign keys, it is possible to maintain coherent connections between the data, avoiding inconsistencies that could harm analysis and decision-making (ANTONIO, 2017).

Data consistency is also maintained through validation standards and constraints, which establish formatting and allowed values for specific fields. Consider an inventory management system where the quantity of products in stock should always be positive. By applying a validation pattern to this field, you ensure that invalid information, such as negative values, is prevented from being entered, avoiding misconceptions and problems resulting from incorrect data.

4.2 TRACEABILITY AND AUDITING

The traceability of the data is essential for reliability, allowing the reconstitution of the history of the changes made. Design patterns can involve creating audit logs that document every action in the

data, including the person responsible and the timing. This is particularly pertinent in regulatory and compliance contexts, such as the General Data Protection Law (LGPD), where it is necessary to document the handling of sensitive data.

The author also highlights the importance of traceability and auditing as a means of ensuring data reliability. Through the implementation of audit records that document the actions taken on the data, it is possible to identify and correct errors or manipulations, contributing to greater reliability and transparency (ANTONIO, 2017).

In addition, traceability provides an additional layer of reliability by making it possible to identify errors or malicious manipulations in the data. For example, if a discrepancy is detected between the recorded sales and the final financial figures, traceability allows you to identify exactly when and by whom the changes were made, assisting in the rectification and clarification of discrepancies.

4.3 RESILIENCE AND FAILURES AND RECOVERY

Data reliability is intrinsically linked to the system's ability to recover from failures. Design patterns can include sound backup, replication, and disaster recovery strategies. By defining standardized procedures for performing regular backups and maintaining replicas of data, it is ensured that in the event of hardware or software failure, information can be recovered without significant loss.

These standards are especially relevant in systems that operate in high-availability environments, such as financial services platforms. Imagine an online bank where a failure on the main server occurs. Continuous replication of data to a secondary server allows for an almost seamless transition for clients, ensuring reliable access and transactions.

4.4 SAFEGUARDING PRIVACY AND DATA SECURITY

The reliability of data is intrinsically associated with the security and privacy of the information stored. Design patterns include security mechanisms, such as encryption, that safeguard sensitive data from unauthorized access. For example, in an electronic medical records system, applying encryption ensures that patient information remains confidential even in the face of a security breach.

Authentication and access control are also critical to data reliability. Implementing strict authentication standards, such as the use of complex passwords and two-factor authentication, ensures that only authorized users can access the data. This reduces the risk of intrusions and security breaches while maintaining data integrity.

In addition, the author Antônio (2017) points out that data security is fundamental to the reliability of the system. The application of security standards, such as encryption and access control,

helps protect data from unauthorized access and privacy breaches while maintaining the integrity of information (ANTONIO, 2017).

Data reliability is a prerequisite for confident and well-informed decisions. Through the consistent adoption of design standards in data modeling, including maintaining referential integrity, implementing traceability, resilience to failure, and preserving data privacy and security, it is possible to create systems that inspire trust both internally and for external stakeholders. Rigorous application of these standards is a guarantee that the information handled remains accurate, consistent and reliable, laying a solid foundation for operational and strategic effectiveness.

5 GOOD PRACTICES IN DATA MODELING

When working on a complex project, it is relevant to consider the large amount of information that will be used during the process. This excessive volume results in the need to adopt good practices, which aim to facilitate and improve the organization of data. These conducts encompass both conceptual and practical stages.

As the project is developed, it becomes common to escape from the reality of the big picture. Therefore, it is necessary that there is good communication between software engineers and business analysts, so that the data model chosen is consistent with reality. For this to occur, it is necessary that both parties have full knowledge of the requirements of the company and seek to filter the data really needed, among all those available.

Filtering addresses complications from high volumes of data, which cause memory and performance issues. In many cases, only one section of data is able to resolve project backlogs without information overload. Therefore, selecting the most relevant data will maximize the operability of the database.

In addition, the relevance of choosing the data modeling technique should be emphasized. This choice should be based on the type of data you are working with. In each technique, there are specific benefits and harms. The key point is to figure out which data model offers the greatest benefits to your usage situation and where the downsides are mitigated.

After choosing the data model, make sure that it is designed so that it can be updated and modified with minimal difficulty. There will always be changes in the future, and it is important that the model has the flexibility to adapt to them. Taking this into consideration from the beginning will prevent future problems.

Finally, proper documentation of your data model is of paramount importance. With good documentation, other people will know what data is available and this will also make it easier to maintain and modify the data model in the future, where even new team members will be able to join

the team more quickly. After all, the current state of the data model will always be up to date through its documentation.

6 BALANCING BENEFITS AND CHALLENGES: THE IMPORTANCE OF SELECTIVE STANDARDIZATION IN DATA MODELING

Standardization within data modeling entails several benefits such as those cited throughout the article. However, we must be careful about what we should standardize, since data that come from heterogeneous, incomplete, unstructured or inconsistent sources should not be standardized.

As we delve deeper into environments where data comes from heterogeneous, incomplete, unstructured, or inconsistent sources, standardization can become a double-edged sword. Imposing patterns on heterogeneous data can result in loss of relevant information or even distortion of the original meanings. The rigid application of standards can also be impractical when the data is unstructured or incomplete since the lack of a predefined format makes it difficult to adapt to the imposed rules. That is, standardization becomes a challenge when the data is inconsistent. Forcing a uniformity in this sense can be harmful since the information from the original data can be of paramount importance.

Given this complex scenario, it is critical to take a selective approach to standardization. Careful evaluation of each dataset allows you to determine which elements can benefit from standardization and which need a more flexible approach. Selective standardization recognizes the importance of adapting to the reality of data and preventing distortion/editing of information, even if it means renouncing total uniformity.

7 METHODOLOGIES

7.1 SEARCH CLASSIFICATION

The research carried out in this article is exploratory and descriptive in nature. The goal is to analyze and discuss the importance of standardization in data modeling and its impact on the quality of database management systems. The research is conducted based on a comprehensive review of literature, theories, and concepts related to standardization, data modeling, and systems quality.

7.2 DATA COLLECTION PROCEDURE

The data for this research were collected through an extensive bibliographic survey. Reliable sources were used, including books, scholarly articles, and online resources from renowned experts and institutions in the field of data modeling, database management, and standardization. The information collected ranges from the importance of standardization in data modeling to the challenges faced when applying standards in environments with heterogeneous sources and complex data.

7.3 DATA ANALYSIS

Data analysis was performed through a qualitative approach, involving the critical interpretation of the information obtained. The data collected was organized into thematic categories, including the importance of design patterns in data modeling, the contribution of standards to the quality of database management systems, and the discussion of the challenges of standardization in complex scenarios.

7.4 INTERPRETATION OF RESULTS

From the data analysis, the results were interpreted in the light of the theories and concepts presented in the article. The main advantages of standardization in data modeling were identified, such as improving the quality of systems, referential integrity, data consistency and information security. In addition, the selective standardization approach was highlighted as a strategy to address the challenges related to heterogeneous sources and complex data.

8 CONCLUSIONS

In this article, we explore the importance of standardization in data modeling and its impact on the quality of database management systems, we begin with a contextualization on the need for standards and good practices in data modeling, highlighting the quote from Taiichi Ohno who highlights the importance of standards for continuous improvement, Through the application of design standards, it is possible to create more durable, efficient and reliable systems. In each section, we cover the different aspects related to standardization in data modeling, explore how standards contribute to the quality of systems, from the incorporation of principles to the application of practices such as normalization, advanced indexing, and controlled denormalization, each of these aspects plays a key role in optimizing performance, data consistency, and the ability to adapt to diverse scenarios.

Data reliability was a central point of discussion, where we looked at how design patterns contribute to referential integrity, data consistency, traceability, resilience to failure, privacy protection and data security, highlighting the need to follow rigorous standards to ensure data reliability at all stages of the lifecycle. However, we also recognize that standardization is not an absolute concept and can face challenges in environments with heterogeneous, incomplete, or inconsistent sources. We proposed the idea of selective standardization, where the application of standards must be carefully evaluated to avoid the loss of relevant information or the distortion of the original data. Selective standardization allows a balance between the search for uniformity and the need to preserve the integrity of information.

In short, standardization in data modeling is an essential element for building reliable and efficient systems. Through the judicious application of design standards and good practices, it is possible to create solid foundations for assertive decision-making and for the continued success of

companies and organizations. Selective standardization, when applied with discernment, can be the key to addressing the challenges that heterogeneous and complex data present, while preserving the reliability and usefulness of information.

REFERENCES

ANTONIO, Adriano Martins. Três fundamentos sobre segurança da informação que todo profissional precisa saber.

OLIVEIRA, R. A. (2012). Introdução a Bancos de Dados: Com Modelagem Dimensional. Cengage Learning.

SILVA, R. S. (2008). Bancos de Dados: Projeto e Implementação. Novatec Editora.

SOUZA, M. H. P., et al. (2014). Banco de Dados: Modelagem, Implementação e Práticas com MySQL. Elsevier Brasil.

TEIXEIRA, C. S. (2006). Bancos de Dados Relacionais: Teoria e Prática. Editora Campus.

https://www.devmedia.com.br/as-vantagens-do-uso-de-padronizacao-em-objetos-no-banco-de-dados/3221

https://www.devmedia.com.br/normalizacao-e-desnormalizacao-de-dados/24345

https://powerbi.microsoft.com/pt-br/what-is-data-modeling/

https://www.totvs.com/blog/negocios/modelagem-de-dados/

https://www.alura.com.br/artigos/normalizacao-banco-de-dados-estrutura

https://sarasanalytics.com/blog/data-modeling-best-practices/

https://revistas.ancib.org/index.php/tpbci/article/view/262/262

https://revistas.faa.edu.br/SaberDigital/article/view/1029/739

http://www.pmgacademy.com/pt/blog/artigos/tres-fundamentos-seguranca-da-informação