Machine Translated by Google

RCMOS – Multidisciplinary Scientific Journal The Knowledge. ISSN: 2675-9128. São Paulo-SP.

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

Data architectures for complex information: persistence, access structures, and context preservation.

Data architectures for complex information: persistence, access structures and context preservation
Data architectures for complex information: persistence, access structures and context preservation

Ricardo José Leal dos Santos is an

Electronic Engineer with a postgraduate degree in Systems Analysis and a master's degree in Computer Engineering, specializing in Geomatics. His professional career focuses particularly on data systems, database administration, and Business Intelligence, combined with academic activities and teaching.

ABSTRACT

The expansion of analytical applications, cloud services, and artificial intelligence systems has increased the dependence on infrastructures capable of preserving data in different formats, scales, and contexts. This article analyzes the relationship between persistence, access structures, analytical modeling, provenance, and geographic information, taking as its central problem the adequacy between the nature of the data and the architecture used to store and query it. The research is qualitative, bibliographic, and documentary in nature, based on traceable academic literature and official technical documentation. The analysis shows that complexity does not only stem from volume: concurrency, temporality, query patterns, physical distribution, specialized types, and the need to reconstruct transformations also condition architectural decisions. Spatial data are used as a representative case because they highlight the need for indexing methods and operators compatible with non-scalar relationships. The study concludes that BI and artificial intelligence environments depend on a prior chain of decisions regarding persistence, quality, metadata, and provenance, and that technological modernization does not eliminate database fundamentals, but only repositions them in new infrastructure layers.

Keywords: databases; data architecture; persistence; geographic information; indexing; Business Intelligence; provenance.

ABSTRACT The

expansion of analytical applications, cloud services, and artificial intelligence systems has increased dependence on infrastructures capable of preserving data across different formats, scales, and contexts. This article analyzes the relationship between persistence, access structures, analytical modeling, provenance, and geographic information, focusing on the fit between the nature of data and the architecture used to store and query it. The research adopts a qualitative bibliographic and documentary approach based on traceable academic literature and official technical documentation.

The analysis shows that complexity is not determined by volume alone: concurrency, temporality, query patterns, physical distribution, specialized data types, and the need to reconstruct transformations also shape architectural decisions. Spatial data are used as a representative case because they make explicit the need for indexing methods and operators compatible with non-scalar relationships. The study concludes that Business Intelligence and artificial intelligence environments depend on a prior chain of decisions concerning persistence, quality, metadata, and provenance, and that technological modernization does not eliminate database fundamentals but relocates them within new infrastructure layers.

Keywords: databases; data architecture; persistence; geographic information; indexing; Business Intelligence; provenance.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge. ISSN: 2675-9128. São Paulo-SP.

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

1 INTRODUCTION

Digital transformation has increased the volume of information produced by transactional systems.
devices, relationship platforms, mobile applications, and analytical environments. At the same time
Over time, the popularization of cloud computing and artificial intelligence has raised expectations of
that organizations can explore this data with lower latency and greater capacity to
processing. This expansion, however, does not eliminate the classic problems of engineering.
data. On the contrary, it makes the need to decide how the information will be presented more relevant.
represented, persisted, indexed, historicized and recovered. Stonebraker and Çetintemel (2005) already
They argued that different classes of applications impose sufficiently distinct requirements for
To question the idea of a single database architecture suitable for all scenarios. The architecture
The internal structure of DBMSs combines storage, query processing, transactions, and components.
shared, which prevents reducing the data problem to a single technological layer.
(HELLERSTEIN; STONEBRAKER; HAMILTON, 2007).
The complexity of an architecture cannot be measured solely by the number of records.
stored. A dataset may be small in size and still require processing.
sophisticated because it involves competition, history, spatial relationships, provenance, or consultations.
specialized. The PostgreSQL documentation (2026) highlights this diversity by offering
different indexing methods and proprietary mechanisms for concurrency control, processing
Transactional and physical storage. The presence of multiple structures in a mature system.
shows that performance and reliability depend on the relationship between usage patterns and how...
The data is organized. Fundamentals of modeling, query processing,
Storage, indexing, and transactions are treated as interconnected dimensions of administration.
of databases by Ramakrishnan and Gehrke (2003).
This problem becomes even more relevant when the data crosses the boundary between operation and
analysis. Kimball and Ross (2013) demonstrate that data warehouse environments require decisions
specific granularity, dimensional modeling, and historical processing, different from
priorities of a transactional system. Glavic (2021), in turn, shows that provenance is
It has established itself as an important area of research in databases because it allows us to explain the origin and
derivation of information. This article brings these fundamentals closer to geographic information,
domain that makes particularly visible the relationship between data type, access structure, and cost.
computational.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

2 METHODOLOGY

The research adopts a qualitative, exploratory, and bibliographic-documentary approach. They were...
Selected recognized academic references in the areas of databases and modeling.
dimensional, provenance and data management, complemented by official technical documentation of
PostgreSQL 18 and the National Institute of Standards and Technology's Research Data Framework.
The selection criteria prioritized sources with authorship, publisher, DOI, or institutional documentation.
traceable, in order to allow verification of the statements used in the development.
The analysis was organized into five dimensions: persistence and transactional processing; structures
Access; transition from operating systems to analytical environments; metadata and provenance;
and the processing of geographic information. Geomatics is not treated as an autonomous field in relation to
not to the central object, but as a domain that highlights problems of data storage and retrieval.
complex. This delimitation is consistent with the objective of discussing engineering fundamentals of
data, avoiding transforming the author's professional experience into scientific evidence.
The procedure consisted of comparing established concepts from the literature with mechanisms.
documented in contemporary systems. Dimensional modeling is discussed from the perspective of
Kimball and Ross (2013); the diversity of architectures is contextualized by Stonebraker and Çetintemel.
(2005); provenance is based on Glavic (2021); management and preservation principles are
related to the NIST Research Data Framework (HANISCH et al., 2024); and mechanisms of
Indexing, transactions, and storage are covered in the PostgreSQL 18 documentation.
(POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). No results are presented.
Neither original experiments nor quantitative generalizations that cannot be traced back to their sources.

3. Persistence, Transactions, and Competition

Persistence is often reduced to the idea of saving a value to disk, but banking systems
Data processors need to ensure broader properties. An application can execute multiple
Changes that only make sense when confirmed together. The transaction abstraction.
This allows treating this set as a logical unit, preventing faults from leaving intermediate states.
improperly exposed. In PostgreSQL, transactions and concurrency control mechanisms are
treated as core components of operational consistency (POSTGRESQL GLOBAL)
(DEVELOPMENT GROUP, 2026). The database systems literature deals with management
of storage and transaction processing as interdependent components of the architecture

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

of a DBMS (HELLERSTEIN; STONEBRAKER; HAMILTON, 2007; RAMAKRISHNAN;
GEHRKE, 2003).

The durability of a confirmed transaction does not mean that all modified pages
have been immediately rewritten in their final position. Banks use logging mechanisms.
to record changes in a way that is compatible with recovery after failures. This separation between
Logical confirmation and physical reorganization are essential to balance safety and performance.
Architecture, therefore, needs to be analyzed as a combination of abstractions offered to the application and
Internal mechanisms that materialize these guarantees.

Concurrency adds another level of difficulty. Multiple users and processes can read and
The system needs to prevent data from being modified simultaneously, and this overlapping data can be modified by creating different states.
incompatible with the application rules. Multi-version concurrency control, MVCC, used
Using PostgreSQL, it allows transactions to work with appropriate views of the data, reducing the
need to block reads and writes in all scenarios (POSTGRESQL GLOBAL)
(DEVELOPMENT GROUP, 2026). In distributed systems, the same consistency problems,
Competition and recovery now coexist with distribution, replication, and coordination among nodes.
(ÖZSU; VALDURIEZ, 2020).

The existence of different versions, however, generates maintenance and storage costs. Systems
They need to identify which versions can still be observed and which can be discarded or
reused. This point demonstrates that a logical property, such as isolation, has
physical consequences. In large environments, decisions about competition, transaction duration and
Maintenance can directly influence the behavior of the infrastructure.

It is also necessary to separate durability, backup, and history. Durability seeks to ensure the
Survival of a confirmed change in the face of certain failures; backup creates copies.
intended for recovery; historical data allows for the reconstruction of past states or events as part of the
Information model. An organization can have durable transactions and adequate backups, and
However, it is still not possible to answer what the value of an attribute was on a past date if that
temporality has not been preserved.

This distinction becomes important in analytical and regulated systems, in which the question
It's often not just about the current state, but how it came to be. (NIST Research)

The Data Framework includes activities related to storage, preservation, metadata, and management.
throughout the lifecycle, reinforcing that data reliability depends on processes that
They go beyond the initial recording (HANISCH et al., 2024).

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge. ISSN: 2675-9128. São Paulo-SP.

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

4. Access Structures and Complex Data

Indexes exist because locating records through full scans can become inadequate as the system progresses.
that tables grow or queries become more frequent. The PostgreSQL 18 documentation
It features B-tree, Hash, GiST, SP-GiST, GIN, and BRIN as built-in indexing methods, each
one with its own properties and operation classes (POSTGRESQL GLOBAL DEVELOPMENT
(GROUP, 2026). This diversity shows that indexing is not a generic optimization, but a
The choice is associated with the data type and query pattern. The classic database literature...
It presents index structures as central mechanisms to reduce the cost of access and support.
different query patterns (RAMAKRISHNAN; GEHRKE, 2003).
B-tree is a widely used structure for scalar data and equality operations.
sorting. PostgreSQL stands out for its suitability for concurrent applications that need to index.
scalar data. However, the documentation itself notes that non-scalar data may require...
methods such as GiST, SP-GiST, or GIN. An incorrect choice can produce space-consuming indexes.
and impose maintenance costs without efficiently resolving the predominant query.
GiST, Generalized Search Tree, functions as an extensible infrastructure for search schemes.
indexing. The official documentation explains that the method can serve as a basis for different
strategies and allows domain experts to define methods appropriate to custom types
(POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). This feature is especially
Relevant when relationships between objects cannot be represented solely by a linear order.
BRIN follows a different logic. Instead of maintaining a conventional entry for each line, it works
with summaries associated with block intervals, which is useful when values are correlated with the
Physical location. The savings in size can be significant in very large tables, although
Its efficiency depends on this correlation. The example demonstrates that the physical architecture needs to take into account
taking into account distribution and access patterns, not just high-level semantics.
Complex data can also include text, arrays, documents, time series, and structures.
spatial. Each category modifies the questions that the system needs to answer. The thesis of
Stonebraker and Çetintemel (2005) on the inadequacy of a single solution for all workloads
The following remains relevant: transactional, analytical, and specialized workloads have priority.
distinct from latency, update, read, and organization. Database theory also
This demonstrates that models and languages need to accompany different forms of representation and
information consultation (ABITEBOUL; HULL; VIANU, 1995).
Therefore, performance should not be treated as a simple consequence of adding hardware.

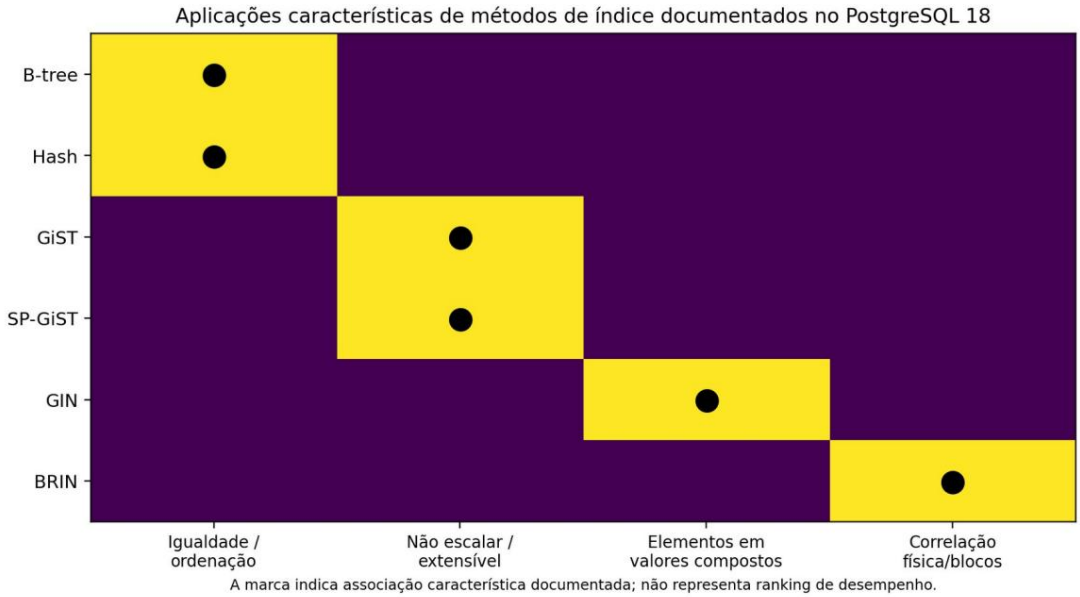
This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The Knowledge. ISSN: 2675-9128. São Paulo-SP.

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

Access structures, selectivity, distribution, concurrency, and storage format all play a role.
in the amount of work required to answer a query. An efficient architecture seeks
reduce unnecessary work without imposing maintenance costs that are incompatible with the standard of
written. Figure 1 summarizes, without assigning performance scores, some associations.
The differences between index methods and usage classes are documented by PostgreSQL.

Figure 1 – Typical applications of index methods documented in PostgreSQL 18



Source: author's elaboration based on PostgreSQL Global Development Group (2026), chapters 11 and 65. The figure is qualitative and does not express a performance comparison.

5. Geographic Information as a Data Problem

Geographic information offers a particularly clear case of complex data because it incorporates
Position, shape, and spatial relationships. A point, a line, or a polygon are not just sets.
of numbers. These objects can be used in questions about proximity, intersection, containment and
territorial distribution. The logic of the query, therefore, modifies the way the information should be presented.
represented and indexed. Guttman's classic work (1984) introduced the R-tree precisely for
enable efficient searching of spatial objects whose multidimensional extent is not well represented by
traditional indexing methods.
From a database perspective, the relevance of Geomatics lies precisely in this requirement for
Structures capable of persisting and retrieving spatial information. The PostgreSQL documentation includes
geometric types and extensible indexing methods, while the ecosystem of space databases
It expands these capabilities. The technical problem is not drawing a map, but enabling operations.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

that space-related tasks are executed in a consistent and efficient manner.

A query by identifier can locate a specific row; a spatial query can

It is necessary to discover all the objects that intersect with a given area. Without mechanisms

If appropriate, the system may need to compare a large number of objects. Frameworks like GiST

They provide a basis for search strategies suited to more complex relationships than sorting.

scalar (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). R-trees have been proposed for

to retrieve objects according to their spatial location, offering a dynamic structure of

search and update for geographic and computer-aided design applications (GUTTMAN,

1984).

Quality also takes on its own characteristics. Coordinates can be formally

filled in and still represent an incorrect location. Incompatible reference systems may

to produce inappropriate displacements or comparisons. A given spatial datum, therefore, can be

syntactically valid but semantically inappropriate for the intended analysis. This difference reinforces the

The idea that quality depends on usage. Batini and Scannapieco (2006) treat data quality as

a multidimensional problem, involving definitions, measurement and improvement strategies, which

It reinforces the idea that syntactic validity is not sufficient to guarantee appropriateness for use.

In analytical environments, the spatial dimension can be combined with business data, population,

infrastructure or operational events. This combination expands the potential for analysis, but increases

the need for metadata that explains origin, scale, reference system, period, and method.

production. Without this information, the visualization may convey accuracy that the source does not possess.

The relationship between databases and geomatics illustrates a broader principle of this article: the nature

Information shapes architecture. It's not about separating geographic data from the fundamentals of...

banks, but to show how a specialized domain requires that these fundamentals be

applied in a manner compatible with their properties.

6. FROM OPERATING SYSTEM TO BUSINESS INTELLIGENCE

Operating systems and analytical environments serve different needs. The first group

It prioritizes the recording of operations, updates, and transactional consistency. The second needs to go through

historical data, consolidating sources, and performing aggregations. This difference motivated data architectures.

Warehouses that reorganize data for consultation and decision-making. Chaudhuri and Dayal (1997)

They show that data warehousing and OLAP impose different requirements than those found in

Traditional transaction processing applications, including extraction, cleaning, loading, and modeling.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

multidimensional and efficient query processing.

Kimball and Ross (2013) structure dimensional modeling around facts, dimensions,
Granularity and business processes. The proposal aims to make analytical environments understandable.
and efficient for consultations, in contrast to often standardized operational models.
to reduce redundancy and preserve update consistency. This is not about declaring a model.
superior to the other, but recognizing different purposes.

The transition between environments requires extraction and transformation. ETL or ELT processes can...
Standardize values, integrate sources, and build analytical structures. Each transformation, however, creates...
a new step that needs to be documented. If the rule used to calculate an indicator changes, the
Results can change even when the observed phenomenon remains stable. The literature dates
warehousing already identified back-end tools for extraction, cleaning, and loading as
fundamental components of analytical architecture (CHAUDHURI; DAYAL, 1997).

This point brings Business Intelligence closer to governance and provenance. A dashboard is not just...
a visual layer. It represents the result of a chain that begins in source systems and
It goes through filters, conversions, joins, and business rules. The reliability of the visualization depends
the ability to explain this chain.

The historical experience of professionals who have worked in both bank administration and business intelligence.
This demonstrates that technical continuity. However, in the scientific field, the relationship must be supported.
based on verifiable principles. Kimball and Ross (2013) demonstrate that dimensional design,
Handling change and defining granularity are structural decisions, not simple choices.
of tool.

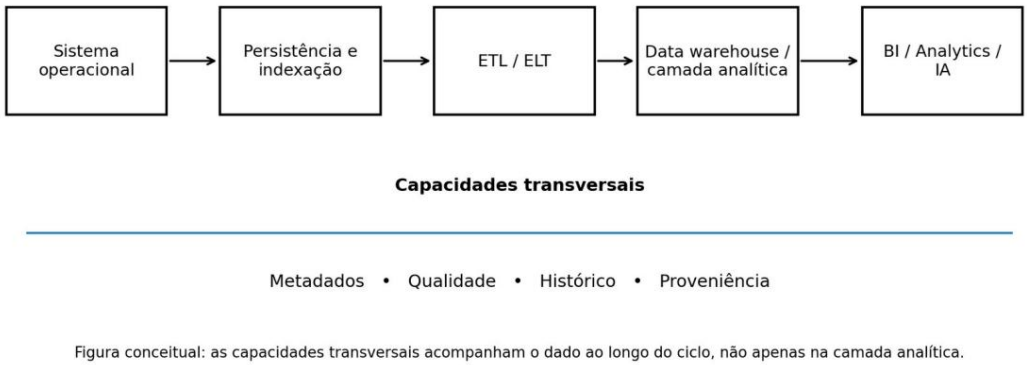
With the expansion of AI, this chain gains another consumer. Models can utilize data.
originating from operational and analytical environments, but they do not automatically correct errors.
inconsistencies, duplications, or lack of context. The greater the automation of the decision, the greater
The importance of understanding how the information was produced before it arrived at the model. Figure
2 summarizes this continuity between operational origin, persistence, transformation, and analytical environment.
and consumption, highlighting metadata, quality, history, and provenance as capabilities.
transversals. Lakehouse architectures were later proposed to combine storage.
open and directly accessible with traditional DBMS management and performance features
analytics, including transactions, versioning, auditing, indexing and optimization (ARMBRUST et
al., 2021).

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

Figure 2 – Conceptual data cycle between the operational environment and analytical use.



Source: author's elaboration based on Kimball and Ross (2013), Glavic (2021), Hanisch et al. (2024) and Wilkinson et al. (2016).

7. Metadata, Provenance, and Context Preservation

Glavic (2021) describes data provenance as a research area that seeks to represent and utilize Information about origin and derivation. In databases, provenance can help explain why a particular result exists and what inputs contributed to its production. This This capability is useful for auditing, debugging, reproducibility, and understanding pipelines. Metadata serves a complementary function. A field called value or status can be Technically accessible, yet still ambiguous to those unfamiliar with the system. Description. Domain, responsible party, unit, periodicity, and transformation rules transform a structure. Physical data in an interpretable asset. The NIST Research Data Framework emphasizes management practices that They track data throughout the life cycle (HANISCH et al., 2024). Quality and capacity Reuse also depends on descriptions that allow for an understanding of the meaning and the conditions of use of the data (BATINI; SCANNAPIECO, 2006; WILKINSON et al., 2016).

The FAIR principles also reinforce the importance of metadata and reuse. Wilkinson et al. (2016) propose that data be locatable, accessible, interoperable, and reusable, with emphasis in descriptions and identifiers that allow for their discovery and interpretation. Although formulated in In the context of scientific data management, these principles offer a useful reference for discussion. preservation of context.

Lineage is often used to describe the flow between sources, transformations, and destinations. In In an analytical environment, this representation can indicate which tables fed a given data.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

The indicator shows which processes modified the values. Provenance can provide a broader perspective.
Broad in scope and derivation. In practice, the necessary granularity depends on the risk and purpose.
Preserving context also has a temporal dimension. A value that is correct today may not be...
Representing the state that existed in the past. Data warehouses handle dimensional changes.
precisely because historical analyses need to decide whether attributes should reflect the current state or the
current state at the time of the event (KIMBALL; ROSS, 2013).
These mechanisms become even more relevant when data feeds into AI models.
Reproducing a result may require knowing the source version, extraction period, transformations.
and selection criteria. Without this record, the organization may be able to run the code again.
and yet not reconstruct the same set of information.

8 DISCUSSION

Analysis of the sources reveals that modernizing the platforms does not eliminate the
Fundamentals of databases. Managed services can automate replication, backup,
scalability and maintenance, but applications still depend on decisions regarding
Consistency, history, access structures, and quality. The shift in implementation to...
Cloud computing modifies operational responsibilities, not the need to understand the properties.
of the data. The architectural description by Hellerstein, Stonebraker, and Hamilton (2007) reinforces that
Scalability and reliability emerge from the combination of several internal subsystems, not from a single one.
unique technology.
It is also evident that volume is not a sufficient measure of complexity. Spatial data
They demonstrate how a relatively smaller set can require specialized operators and indices.
Similarly, an analytical foundation may require historical data and provenance even without achieving scale.
extreme. The architectural problem arises from the combination of the nature of the information, the pattern of
consultation and necessary guarantees. This conclusion is consistent with the data quality literature.
which differentiates multiple dimensions of suitability, and with the distributed banking literature, in which
Scale adds distribution and coordination problems (BATINI; SCANNAPIECO, 2006; ÖZSU; VALDURIEZ, 2020).

The literature analyzed supports the rejection of a universal architecture. Stonebraker and Çetintemel
(2005) formulated this criticism in relation to different classes of applications, and the documentation
Modern PostgreSQL embodies diversity through multiple indexing methods and
Internal mechanisms. The persistence of this variety reinforces that specialization remains relevant.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google
RCMOS – Multidisciplinary Scientific Journal The
Knowledge. ISSN: 2675-9128. São Paulo-SP.

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

Another result of the analysis is the continuity between operational databases and Business Intelligence.
Kimball and Ross (2013) show that analytical environments have their own modeling, but
They rely on data produced in previous systems. Therefore, BI does not replace business administration.
data; it expands the chain of responsibility regarding quality, history, and interpretation.
Provenance adds an essential layer to this chain. Glavic (2021) demonstrates that
Understanding origin and derivation is a well-established research problem. In AI scenarios, this...
The need tends to intensify because results may depend on datasets and
Transformations that are not visible to the end user. Traceability becomes a component of
trust.
Finally, the geographical information shows how an academic background in Geomatics can be
integrated into the data field without shifting the focus to cartography or interoperability.
Spatial storage and querying are legitimate database problems.
The framing is technically coherent and avoids giving undue weight to peripheral experiences.
when the professional core lies in data and the relationship between practice and academia.

9 CONCLUSION

Data engineering needs to be understood as a discipline that preserves the conditions necessary for...
Information remains usable as it traverses systems, technologies, and purposes. Persistence is a value.
This is just the beginning of the process. Architecture also needs to deal with competition, recovery,
History, indexing, quality, and interpretation.
The sources analyzed show that different workloads require different choices. The critique of
Stonebraker and Çetintemel (2005) argue that the idea of a single solution remains compatible with variety.
of structures observed in current platforms. Index methods, transactional models and
Analytical architectures address specific problems.
Geographic information reinforces this conclusion. Spatial relationships require representation and
Access mechanisms compatible with proximity, intersection, and containment. The case demonstrates that
The type of data and the type of question need to be considered together. Guttman's R-tree.
(1984) is a milestone in this specialization, as it was specifically designed for space search of
multidimensional objects.
Business Intelligence extends the lifecycle by reorganizing operational data for analysis.
The dimensional modeling described by Kimball and Ross (2013) shows that granularity, facts,
Dimensions and historical data are decisions that directly affect the interpretation of indicators.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

Provenance and metadata add explanatory power. Glavic (2021) shows that origin and derivation constitute a consolidated field of research, while the NIST Research Data The framework offers a management perspective throughout the lifecycle (HANISCH et al., 2024). These These elements are particularly important when multiple transformations separate the source from the result.

The FAIR principles of Wilkinson et al. (2016) also remind us that utility depends on conditions. of discovery, access, interpretation, and reuse. Although they originate in the context of data. scientific studies help reinforce the idea that storing something does not equate to making it understandable and reusable.

The expansion of cloud computing and artificial intelligence increases the available computing capacity, but This does not change that principle. Sophisticated models continue to depend on data whose origin, Structure and quality can be understood. Automating decisions based on a poorly defined foundation A structured approach can only accelerate the spread of inconsistencies.

It can be concluded, therefore, that robust architectures result from the adequacy between the nature of the information, Intended use and mechanisms for persistence and access. Technological evolution changes products and It scales, but preserves the need for solid foundations. For data, BI, and information environments. Geographically and through AI, trust continues to be built before the visible layer of the application.

REFERENCES

ABITEBOUL, Serge; HULL, Richard; VIANU, Victor. Foundations of Databases. Reading, MA: Addison-Wesley, 1995. ISBN 0-201-53771-0.

ARMBRUST, Michael; GHODSI, Ali; XIN, Reynold; ZAHARIA, Matei. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. In: CONFERENCE ON INNOVATIVE DATA SYSTEMS RESEARCH, 11., 2021. Proceedings [...].
CIDR, 2021.

BATINI, Carlo; SCANNAPIECO, Monica. Data Quality: Concepts, Methodologies and Techniques.
Berlin: Springer, 2006. DOI: 10.1007/3-540-33173-5.

CHAUDHURI, Surajit; DAYAL, Umeshwar. An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record, v. 26, no. 1, p. 65-74, 1997. DOI: 10.1145/248603.248616.

GLAVIC, Boris. Data Provenance: Origins, Applications, Algorithms, and Models. Foundations and Trends in Databases, v. 9, no. 3-4, p. 209-441, 2021. DOI: 10.1561/1900000068.

GUTTMAN, Antonin. R-Trees: A Dynamic Index Structure for Spatial Searching. In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 1984.
Proceedings [...]. New York: ACM, 1984. p. 47-57. DOI: 10.1145/602259.602266.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine Translated by Google

Year VII, v.2 2026 | Submission: July 15, 2026 | Accepted: July 20, 2026 | Publication: July 25, 2026

HANISCH, Robert J. et al. NIST Research Data Framework (RDaF): Version 2.0. Gaithersburg: National Institute of Standards and Technology, 2024. NIST SP 1500-18r2. DOI: 10.6028/NIST.SP.1500-18r2.

HELLERSTEIN, Joseph M.; STONEBRAKER, Michael; HAMILTON, James. Architecture of a Database System. Foundations and Trends in Databases, v. 1, no. 2, p. 141-259, 2007. DOI: 10.1561/1900000002.

KIMBALL, Ralph; ROSS, Margy. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd ed. Indianapolis: John Wiley & Sons, 2013. ISBN 978-1-118-53080-1.

ÖZSU, M. Tamer; VALDURIEZ, Patrick. Principles of Distributed Database Systems. 4th ed. Cham: Springer, 2020. DOI: 10.1007/978-3-030-26253-2.

POSTGRESQL GLOBAL DEVELOPMENT GROUP. PostgreSQL 18 Documentation. 2026. Available at: https://www.postgresql.org/docs/18/.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes. Database Management Systems. 3rd ed. New York: McGraw-Hill, 2003. ISBN 0-07-246563-8.

STONEBRAKER, Michael; ÇETINTEMEL, Uÿur. One Size Fits All: An Idea Whose Time Has Come and Gone. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING, 21., 2005. Proceedings [...]. IEEE, 2005. DOI: 10.1109/ICDE.2005.1.

WILKINSON, Mark D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, vol. 3, art. 160018, 2016. DOI: 10.1038/sdata.2016.18.

This article is published in open access under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.