Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

Arquitecturas de datos para información compleja: persistencia, estructuras de acceso y preservación del contexto.

Arquitecturas de datos para información compleja: persistencia, estructuras de acceso y preservación del contexto.
Arquitecturas de datos para información compleja: persistencia, estructuras de acceso y preservación del contexto.

Ricardo José Leal dos Santos es

ingeniero electrónico con un posgrado en análisis de sistemas y una maestría en ingeniería informática, con especialización en geomática. Su trayectoria profesional se centra particularmente en sistemas de datos, administración de bases de datos e inteligencia empresarial, combinada con actividades académicas y docentes.

RESUMEN

La expansión de las aplicaciones analíticas, los servicios en la nube y los sistemas de inteligencia artificial ha incrementado la dependencia de infraestructuras capaces de preservar datos en diferentes formatos, escalas y contextos. Este artículo analiza la relación entre persistencia, estructuras de acceso, modelado analítico, procedencia e información geográfica, tomando como problema central la adecuación entre la naturaleza de los datos y la arquitectura utilizada para almacenarlos y consultarlos. La investigación es de naturaleza cualitativa, bibliográfica y documental, basada en literatura académica verificable y documentación técnica oficial. El análisis muestra que la complejidad no solo proviene del volumen: la concurrencia, la temporalidad, los patrones de consulta, la distribución física, los tipos especializados y la necesidad de reconstruir transformaciones también condicionan las decisiones arquitectónicas. Los datos espaciales se utilizan como caso representativo porque resaltan la necesidad de métodos y operadores de indexación compatibles con relaciones no escalares. El estudio concluye que los entornos de BI e inteligencia artificial dependen de una cadena previa de decisiones con respecto a la persistencia, la calidad, los metadatos y la procedencia, y que la modernización tecnológica no elimina los fundamentos de las bases de datos, sino que solo los reubica en nuevas capas de infraestructura.

Palabras clave: bases de datos; arquitectura de datos; persistencia; información geográfica; indexación; inteligencia empresarial; procedencia.

RESUMEN La

expansión de las aplicaciones analíticas, los servicios en la nube y los sistemas de inteligencia artificial ha incrementado la dependencia de infraestructuras capaces de preservar datos en diferentes formatos, escalas y contextos. Este artículo analiza la relación entre persistencia, estructuras de acceso, modelado analítico, procedencia e información geográfica, centrándose en la adecuación entre la naturaleza de los datos y la arquitectura utilizada para almacenarlos y consultarlos. La investigación adopta un enfoque cualitativo, bibliográfico y documental, basado en literatura académica verificable y documentación técnica oficial.

El análisis demuestra que la complejidad no está determinada únicamente por el volumen: la concurrencia, la temporalidad, los patrones de consulta, la distribución física, los tipos de datos especializados y la necesidad de reconstruir transformaciones también influyen en las decisiones arquitectónicas. Se utilizan datos espaciales como caso representativo, ya que evidencian la necesidad de métodos y operadores de indexación compatibles con relaciones no escalares. El estudio concluye que los entornos de Business Intelligence e inteligencia artificial dependen de una cadena previa de decisiones relativas a la persistencia, la calidad, los metadatos y la procedencia, y que la modernización tecnológica no elimina los fundamentos de las bases de datos, sino que los reubica en nuevas capas de infraestructura.

Palabras clave: bases de datos; arquitectura de datos; persistencia; información geográfica; indexación; inteligencia empresarial; procedencia.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

1. INTRODUCCIÓN

La transformación digital ha aumentado el volumen de información producida por los sistemas transaccionales.
dispositivos, plataformas de relaciones, aplicaciones móviles y entornos analíticos. Al mismo tiempo
Con el tiempo, la popularización de la computación en la nube y la inteligencia artificial ha elevado las expectativas de
que las organizaciones puedan explorar estos datos con menor latencia y mayor capacidad para
procesamiento. Sin embargo, esta expansión no elimina los problemas clásicos de la ingeniería.
datos. Por el contrario, hace más relevante la necesidad de decidir cómo se presentará la información.
representado, persistido, indexado, historicizado y recuperado. Stonebraker y Çetintemel (2005) ya
Argumentaron que las diferentes clases de aplicaciones imponen requisitos suficientemente distintos para
Cuestionar la idea de una única arquitectura de base de datos adecuada para todos los escenarios. La arquitectura
La estructura interna de los sistemas de gestión de bases de datos (DBMS) combina almacenamiento, procesamiento de consultas, transacciones y componentes.
compartido, lo que impide reducir el problema de los datos a una única capa tecnológica.
(HELLERSTEIN; STONEBRAKER; HAMILTON, 2007).

La complejidad de una arquitectura no puede medirse únicamente por el número de registros.
almacenados. Un conjunto de datos puede ser pequeño y aun así requerir procesamiento.
Sofisticado porque implica competencia, historia, relaciones espaciales, procedencia o consultas.
especializado. La documentación de PostgreSQL (2026) destaca esta diversidad al ofrecer
diferentes métodos de indexación y mecanismos propietarios para el control de concurrencia, procesamiento
Almacenamiento transaccional y físico. La presencia de múltiples estructuras en un sistema maduro.
muestra que el rendimiento y la fiabilidad dependen de la relación entre los patrones de uso y cómo...
Los datos están organizados. Fundamentos del modelado, procesamiento de consultas,
El almacenamiento, la indexación y las transacciones se consideran dimensiones interconectadas de la administración.
de bases de datos por Ramakrishnan y Gehrke (2003).

Este problema se vuelve aún más relevante cuando los datos cruzan el límite entre operación y
análisis. Kimball y Ross (2013) demuestran que los entornos de almacenamiento de datos requieren decisiones
granularidad específica, modelado dimensional y procesamiento histórico, diferente de
prioridades de un sistema transaccional. Glavic (2021), a su vez, muestra que la procedencia es
Se ha consolidado como un área importante de investigación en bases de datos porque nos permite explicar el origen y
derivación de información. Este artículo acerca estos fundamentos a la información geográfica,
dominio que pone de manifiesto la relación entre el tipo de datos, la estructura de acceso y el coste.
computacional.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

2. METODOLOGÍA

La investigación adopta un enfoque cualitativo, exploratorio y bibliográfico-documental. Ellos fueron...
Referencias académicas reconocidas seleccionadas en las áreas de bases de datos y modelado.
dimensional, procedencia y gestión de datos, complementados con documentación técnica oficial de
PostgreSQL 18 y el marco de datos de investigación del Instituto Nacional de Estándares y Tecnología.
Los criterios de selección priorizaron las fuentes con autoría, editorial, DOI o documentación institucional.
rastreable, para permitir la verificación de las declaraciones utilizadas en el desarrollo.
El análisis se organizó en cinco dimensiones: persistencia y procesamiento transaccional; estructuras
Acceso; transición de sistemas operativos a entornos analíticos; metadatos y procedencia;
y el procesamiento de información geográfica. La geomática no se trata como un campo autónomo en relación con
no en relación con el objeto central, sino como un dominio que pone de relieve los problemas de almacenamiento y recuperación de datos.
complejo. Esta delimitación es coherente con el objetivo de discutir los fundamentos de ingeniería de
datos, evitando transformar la experiencia profesional del autor en evidencia científica.
El procedimiento consistió en comparar conceptos establecidos en la literatura con mecanismos.
documentado en sistemas contemporáneos. El modelado dimensional se analiza desde la perspectiva de
Kimball y Ross (2013); la diversidad de arquitecturas es contextualizada por Stonebraker y Çetintemel.
(2005); la procedencia se basa en Glavic (2021); los principios de gestión y conservación son
relacionado con el Marco de Datos de Investigación del NIST (HANISCH et al., 2024); y mecanismos de
La indexación, las transacciones y el almacenamiento se tratan en la documentación de PostgreSQL 18.
(GRUPO DE DESARROLLO GLOBAL DE POSTGRESQL, 2026). No se presentan resultados.
Ni experimentos originales ni generalizaciones cuantitativas que no puedan rastrearse hasta sus fuentes.

3. Persistencia, transacciones y competencia

La persistencia a menudo se reduce a la idea de guardar un valor en el disco, pero los sistemas bancarios
Los procesadores de datos deben garantizar propiedades más amplias. Una aplicación puede ejecutar múltiples
Cambios que solo tienen sentido cuando se confirman en conjunto. La abstracción de la transacción.
Esto permite tratar este conjunto como una unidad lógica, evitando que los fallos abandonen los estados intermedios.
expuestos incorrectamente. En PostgreSQL, las transacciones y los mecanismos de control de concurrencia son
tratados como componentes centrales de la consistencia operativa (POSTGRESQL GLOBAL)
(GRUPO DE DESARROLLO, 2026). La literatura sobre sistemas de bases de datos trata sobre la gestión.
del almacenamiento y el procesamiento de transacciones como componentes interdependientes de la arquitectura

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

de un DBMS (HELLERSTEIN; STONEBRAKER; HAMILTON, 2007; RAMAKRISHNAN;
GEHRKE, 2003).

La durabilidad de una transacción confirmada no significa que todas las páginas modificadas
han sido reescritos inmediatamente en su posición final. Los bancos utilizan mecanismos de registro.
registrar los cambios de una manera que sea compatible con la recuperación después de fallas. Esta separación entre
La confirmación lógica y la reorganización física son esenciales para equilibrar la seguridad y el rendimiento.
Por lo tanto, la arquitectura debe analizarse como una combinación de abstracciones ofrecidas a la aplicación y
Mecanismos internos que materializan estas garantías.

La concurrencia añade otro nivel de dificultad. Múltiples usuarios y procesos pueden leer y
El sistema debe evitar que los datos se modifiquen simultáneamente, y estos datos superpuestos pueden modificarse creando diferentes estados.
incompatible con las reglas de la aplicación. Se utiliza el control de concurrencia multiversión (MVCC).
Utilizando PostgreSQL, permite que las transacciones trabajen con vistas apropiadas de los datos, reduciendo la
Es necesario bloquear las lecturas y escrituras en todos los escenarios (POSTGRESQL GLOBAL).
(GRUPO DE DESARROLLO, 2026). En los sistemas distribuidos, los mismos problemas de consistencia,
La competencia y la recuperación coexisten ahora con la distribución, la replicación y la coordinación entre los nodos.
(ÖZSU; VALDURIEZ, 2020).

Sin embargo, la existencia de diferentes versiones genera costos de mantenimiento y almacenamiento. Sistemas
Necesitan identificar qué versiones aún se pueden observar y cuáles se pueden descartar o
reutilizado. Este punto demuestra que una propiedad lógica, como el aislamiento, tiene
consecuencias físicas. En entornos grandes, las decisiones sobre la competencia, la duración de las transacciones y
El mantenimiento puede influir directamente en el comportamiento de la infraestructura.

También es necesario separar durabilidad, copia de seguridad e historial. La durabilidad busca garantizar la
Supervivencia de un cambio confirmado frente a ciertos fallos; la copia de seguridad crea copias.
destinado a la recuperación; los datos históricos permiten la reconstrucción de estados o eventos pasados como parte de la
Modelo de información. Una organización puede tener transacciones duraderas y copias de seguridad adecuadas, y
Sin embargo, todavía no es posible responder cuál era el valor de un atributo en una fecha pasada si ese
La temporalidad no se ha conservado.

Esta distinción se vuelve importante en los sistemas analíticos y regulados, en los que la cuestión
A menudo, no se trata solo del estado actual, sino de cómo se llegó a él. (Investigación del NIST)
El Marco de Datos incluye actividades relacionadas con el almacenamiento, la preservación, los metadatos y la gestión.
a lo largo del ciclo de vida, reforzando que la confiabilidad de los datos depende de procesos que
Van más allá de la grabación inicial (HANISCH et al., 2024).

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

4. Estructuras de acceso y datos complejos

Los índices existen porque la localización de registros mediante escaneos completos puede resultar insuficiente a medida que el sistema avanza.
que las tablas crecen o las consultas se vuelven más frecuentes. La documentación de PostgreSQL 18
Incluye B-tree, Hash, GiST, SP-GiST, GIN y BRIN como métodos de indexación integrados, cada uno
una con sus propias propiedades y clases de operación (DESARROLLO GLOBAL DE POSTGRESQL)
(GRUPO, 2026). Esta diversidad muestra que la indexación no es una optimización genérica, sino una
La elección está asociada al tipo de datos y al patrón de consulta. La literatura clásica sobre bases de datos...
Presenta las estructuras de índices como mecanismos centrales para reducir el coste de acceso y soporte.
diferentes patrones de consulta (RAMAKRISHNAN; GEHRKE, 2003).

El árbol B es una estructura ampliamente utilizada para datos escalares y operaciones de igualdad.
ordenación. PostgreSQL destaca por su idoneidad para aplicaciones concurrentes que necesitan indexar.
datos escalares. Sin embargo, la propia documentación señala que los datos no escalares pueden requerir...
métodos como GiST, SP-GiST o GIN. Una elección incorrecta puede generar índices que consumen mucho espacio.
y que imponen costes de mantenimiento sin resolver eficazmente la consulta principal.

GiST, o Árbol de Búsqueda Generalizado, funciona como una infraestructura extensible para esquemas de búsqueda.
indexación. La documentación oficial explica que el método puede servir de base para diferentes
estrategias y permite a los expertos del dominio definir métodos apropiados para tipos personalizados
(GRUPO DE DESARROLLO GLOBAL DE POSTGRESQL, 2026). Esta característica es especialmente
Resulta relevante cuando las relaciones entre objetos no pueden representarse únicamente mediante un orden lineal.

BRIN sigue una lógica diferente. En lugar de mantener una entrada convencional para cada línea, funciona
con resúmenes asociados con intervalos de bloques, lo cual es útil cuando los valores están correlacionados con el
Ubicación física. El ahorro de espacio puede ser significativo en mesas muy grandes, aunque
Su eficiencia depende de esta correlación. El ejemplo demuestra que la arquitectura física debe tener en cuenta
teniendo en cuenta los patrones de distribución y acceso, no solo la semántica de alto nivel.

Los datos complejos también pueden incluir texto, matrices, documentos, series temporales y estructuras.
espacial. Cada categoría modifica las preguntas que el sistema necesita responder. La tesis de
Stonebraker y Çetintemel (2005) sobre la insuficiencia de una única solución para todas las cargas de trabajo.
Sigue siendo relevante lo siguiente: las cargas de trabajo transaccionales, analíticas y especializadas tienen prioridad.
distinto de la latencia, actualización, lectura y organización. La teoría de bases de datos también
Esto demuestra que los modelos y los lenguajes necesitan acompañar diferentes formas de representación y
consulta de información (ABITEBOUL; HULL; VIANU, 1995).

Por lo tanto, el rendimiento no debe considerarse una simple consecuencia de añadir hardware.

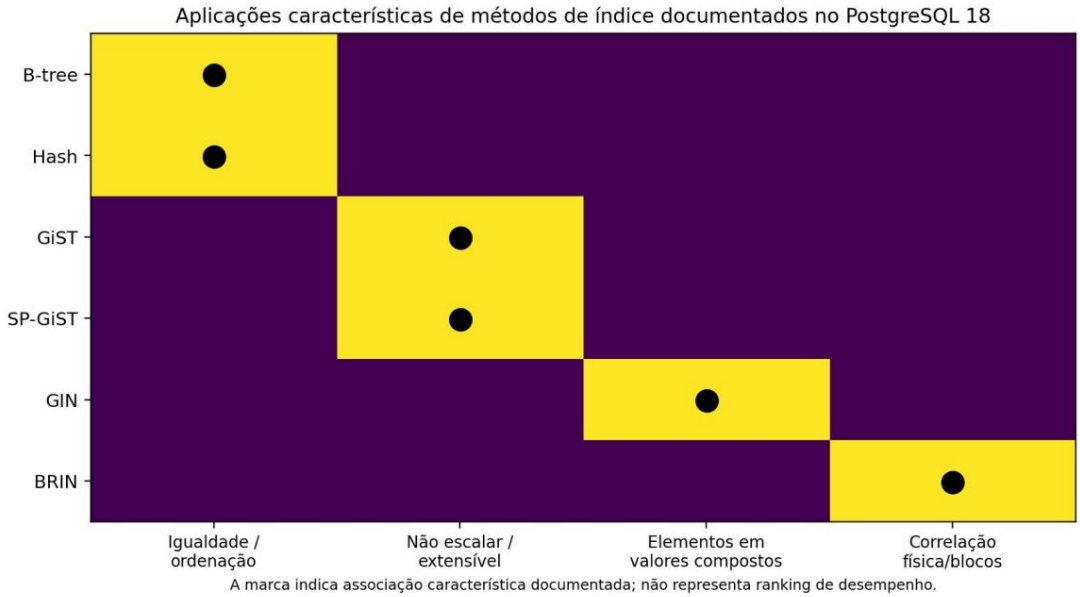
Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

Las estructuras de acceso, la selectividad, la distribución, la concurrencia y el formato de almacenamiento desempeñan un papel importante.
en la cantidad de trabajo necesario para responder a una consulta. Una arquitectura eficiente busca
reducir el trabajo innecesario sin imponer costos de mantenimiento que sean incompatibles con el estándar de
escrito. La Figura 1 resume, sin asignar puntuaciones de rendimiento, algunas asociaciones.
PostgreSQL documenta las diferencias entre los métodos de indexación y las clases de uso.

Figura 1 – Aplicaciones típicas de los métodos de indexación documentados en PostgreSQL 18



Fuente: Elaboración propia basada en PostgreSQL Global Development Group (2026), capítulos 11 y 65. La figura es cualitativa y no representa una comparación de rendimiento.

5. La información geográfica como problema de datos

La información geográfica ofrece un caso particularmente claro de datos complejos porque incorpora
Posición, forma y relaciones espaciales. Un punto, una línea o un polígono no son simplemente conjuntos.
de números. Estos objetos se pueden utilizar en preguntas sobre proximidad, intersección, contención y
distribución territorial. Por lo tanto, la lógica de la consulta modifica la forma en que se debe presentar la información.
representado e indexado. El trabajo clásico de Guttman (1984) introdujo el árbol R precisamente para
permitir búsquedas eficientes de objetos espaciales cuya extensión multidimensional no está bien representada por
métodos de indexación tradicionales.
Desde la perspectiva de las bases de datos, la relevancia de la Geomática radica precisamente en este requisito de
Estructuras capaces de persistir y recuperar información espacial. La documentación de PostgreSQL incluye
tipos geométricos y métodos de indexación extensibles, mientras que el ecosistema de bases de datos espaciales
Amplía estas capacidades. El problema técnico no reside en dibujar un mapa, sino en habilitar las operaciones.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

que las tareas relacionadas con el espacio se ejecuten de manera coherente y eficiente.

Una consulta por identificador puede localizar una fila específica; una consulta espacial puede

Es necesario descubrir todos los objetos que se intersecan con un área determinada. Sin mecanismos

Si procede, el sistema puede necesitar comparar un gran número de objetos. Marcos de trabajo como GiST

Proporcionan una base para estrategias de búsqueda adecuadas a relaciones más complejas que la simple clasificación.

escalar (GRUPO DE DESARROLLO GLOBAL DE POSTGRESQL, 2026). Se han propuesto árboles R para

recuperar objetos según su ubicación espacial, ofreciendo una estructura dinámica de

búsqueda y actualización para aplicaciones geográficas y de diseño asistido por computadora (GUTTMAN,

1984).

La calidad también adquiere sus propias características. Las coordenadas pueden ser formalmente

rellenado y aún así representa una ubicación incorrecta. Los sistemas de referencia incompatibles pueden

para producir desplazamientos o comparaciones inapropiadas. Por lo tanto, un datum espacial dado puede ser

sintácticamente válido pero semánticamente inapropiado para el análisis previsto. Esta diferencia refuerza la

La idea de que la calidad depende del uso. Batini y Scannapieco (2006) tratan la calidad de los datos como

un problema multidimensional, que involucra definiciones, medición y estrategias de mejora, que

Refuerza la idea de que la validez sintáctica no es suficiente para garantizar la idoneidad de su uso.

En entornos analíticos, la dimensión espacial se puede combinar con datos comerciales, población,

infraestructura o eventos operativos. Esta combinación amplía el potencial de análisis, pero aumenta

la necesidad de metadatos que expliquen el origen, la escala, el sistema de referencia, el período y el método.

producción. Sin esta información, la visualización puede transmitir una precisión que la fuente no posee.

La relación entre bases de datos y geomática ilustra un principio más amplio de este artículo: la naturaleza

La información da forma a la arquitectura. No se trata de separar los datos geográficos de los fundamentos de...

bancos, pero para mostrar cómo un dominio especializado requiere que estos fundamentos sean

aplicados de una manera compatible con sus propiedades.

6. DEL SISTEMA OPERATIVO A LA INTELIGENCIA EMPRESARIAL

Los sistemas operativos y los entornos analíticos satisfacen necesidades diferentes. El primer grupo

Prioriza el registro de operaciones, actualizaciones y consistencia transaccional. El segundo necesita pasar por

datos históricos, consolidación de fuentes y realización de agregaciones. Esta diferencia motivó el desarrollo de arquitecturas de datos.

Almacenes que reorganizan datos para consulta y toma de decisiones. Chaudhuri y Dayal (1997)

Demuestran que el almacenamiento de datos y OLAP imponen requisitos diferentes a los que se encuentran en

Aplicaciones tradicionales de procesamiento de transacciones, incluyendo extracción, limpieza, carga y modelado.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

Procesamiento de consultas multidimensional y eficiente.

Kimball y Ross (2013) estructuran el modelado dimensional en torno a hechos, dimensiones,
Granularidad y procesos de negocio. La propuesta busca hacer comprensibles los entornos analíticos.
y eficiente para las consultas, a diferencia de los modelos operativos a menudo estandarizados.
para reducir la redundancia y preservar la coherencia de las actualizaciones. No se trata de declarar un modelo.
superior al otro, pero reconociendo propósitos diferentes.

La transición entre entornos requiere extracción y transformación. Los procesos ETL o ELT pueden...
Estandarizar valores, integrar fuentes y construir estructuras analíticas. Sin embargo, cada transformación crea...
un nuevo paso que debe documentarse. Si la regla utilizada para calcular un indicador cambia,
Los resultados pueden cambiar incluso cuando el fenómeno observado permanece estable. La literatura data
El almacenamiento ya identificó herramientas de back-end para extracción, limpieza y carga como
componentes fundamentales de la arquitectura analítica (CHAUDHURI; DAYAL, 1997).

Este punto acerca la inteligencia empresarial a la gobernanza y la procedencia. Un panel de control no es solo...
una capa visual. Representa el resultado de una cadena que comienza en los sistemas fuente y
Pasa por filtros, conversiones, uniones y reglas de negocio. La fiabilidad de la visualización depende de
la capacidad de explicar esta cadena.

La experiencia histórica de profesionales que han trabajado tanto en administración bancaria como en inteligencia empresarial.
Esto demuestra la continuidad técnica. Sin embargo, en el ámbito científico, la relación debe ser respaldada.
basado en principios verificables. Kimball y Ross (2013) demuestran que el diseño dimensional,
Gestionar el cambio y definir el nivel de detalle son decisiones estructurales, no simples elecciones.
de herramienta.

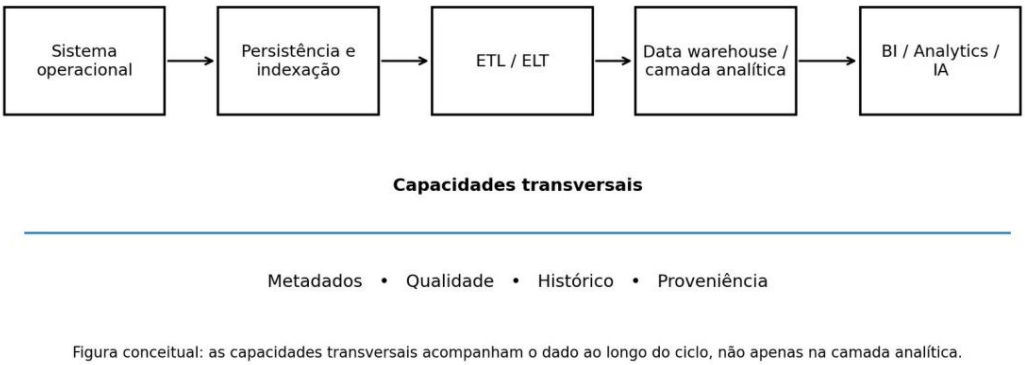
Con la expansión de la IA, esta cadena gana un nuevo consumidor. Los modelos pueden utilizar datos.
Se originan en entornos operativos y analíticos, pero no corrigen automáticamente los errores.
inconsistencias, duplicaciones o falta de contexto. Cuanto mayor sea la automatización de la decisión, mayor será la probabilidad de errores.
La importancia de comprender cómo se produjo la información antes de que llegara al modelo. Figura
El punto 2 resume esta continuidad entre el origen operacional, la persistencia, la transformación y el entorno analítico.
y el consumo, destacando los metadatos, la calidad, el historial y la procedencia como capacidades.
transversales. Posteriormente se propusieron arquitecturas de casas de lago para combinar almacenamiento.
Abierto y de acceso directo con funciones de gestión y rendimiento de sistemas de gestión de bases de datos tradicionales.
análisis, incluyendo transacciones, control de versiones, auditoría, indexación y optimización (ARMBRUST y
et al., 2021).

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

Figura 2 – Ciclo conceptual de datos entre el entorno operativo y el uso analítico.



Fuente: elaboración del autor basada en Kimball y Ross (2013), Glavic (2021), Hanisch et al. (2024) y Wilkinson et al. (2016).

7. Metadatos, procedencia y preservación del contexto

Glavic (2021) describe la procedencia de los datos como un área de investigación que busca representar y utilizar Información sobre origen y derivación. En las bases de datos, la procedencia puede ayudar a explicar por qué existe un resultado en particular y qué insumos contribuyeron a su producción.

Esta capacidad resulta útil para la auditoría, la depuración, la reproducibilidad y la comprensión de los flujos de trabajo.

Los metadatos cumplen una función complementaria. Un campo llamado valor o estado puede ser Técnicamente accesible, pero aún ambiguo para quienes no están familiarizados con el sistema. Descripción.

El dominio, la parte responsable, la unidad, la periodicidad y las reglas de transformación transforman una estructura. Datos físicos en un activo interpretable. El marco de datos de investigación del NIST enfatiza las prácticas de gestión que Realizan un seguimiento de los datos a lo largo del ciclo de vida (HANISCH et al., 2024). Calidad y capacidad La reutilización también depende de descripciones que permitan comprender el significado y el condiciones de uso de los datos (BATINI; SCANNAPIECO, 2006; WILKINSON et al., 2016).

Los principios FAIR también refuerzan la importancia de los metadatos y la reutilización. Wilkinson et al. (2016) proponen que los datos sean localizables, accesibles, interoperables y reutilizables, con énfasis en descripciones e identificadores que permiten su descubrimiento e interpretación. Aunque formulado en En el contexto de la gestión de datos científicos, estos principios ofrecen una referencia útil para el debate. preservación del contexto.

El linaje se utiliza a menudo para describir el flujo entre fuentes, transformaciones y destinos. En un entorno analítico, esta representación puede indicar qué tablas alimentaron un conjunto de datos determinado.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

El indicador muestra qué procesos modificaron los valores. La procedencia puede proporcionar una perspectiva más amplia.
Amplio alcance y origen. En la práctica, el nivel de detalle necesario depende del riesgo y del propósito.
Preservar el contexto también tiene una dimensión temporal. Un valor que es correcto hoy puede no serlo...
Representan el estado que existía en el pasado. Los almacenes de datos gestionan los cambios dimensionales.
precisamente porque los análisis históricos necesitan decidir si los atributos deben reflejar el estado actual o el
estado actual en el momento del evento (KIMBALL; ROSS, 2013).
Estos mecanismos cobran aún más relevancia cuando los datos se incorporan a los modelos de IA.
Para reproducir un resultado, puede ser necesario conocer la versión de la fuente, el período de extracción y las transformaciones.
y criterios de selección. Sin este registro, la organización podría volver a ejecutar el código.
y, sin embargo, no logran reconstruir el mismo conjunto de información.

8 DISCUSIÓN

El análisis de las fuentes revela que modernizar las plataformas no elimina la
Fundamentos de bases de datos. Los servicios gestionados pueden automatizar la replicación, la copia de seguridad,
escalabilidad y mantenimiento, pero las aplicaciones aún dependen de decisiones relacionadas con
Coherencia, historial, estructuras de acceso y calidad. El cambio en la implementación hacia...
La computación en la nube modifica las responsabilidades operativas, no la necesidad de comprender las propiedades.
de los datos. La descripción arquitectónica de Hellerstein, Stonebraker y Hamilton (2007) refuerza que
La escalabilidad y la fiabilidad surgen de la combinación de varios subsistemas internos, no de uno solo.
tecnología única.
También es evidente que el volumen no es una medida suficiente de complejidad. Datos espaciales
Demuestran cómo un conjunto relativamente pequeño puede requerir operadores e índices especializados.
Del mismo modo, una base analítica puede requerir datos históricos y procedencia incluso sin alcanzar una escala considerable.
extremo. El problema arquitectónico surge de la combinación de la naturaleza de la información, el patrón de
consulta y garantías necesarias. Esta conclusión es coherente con la literatura sobre calidad de datos.
que diferencia múltiples dimensiones de idoneidad, y con la literatura sobre banca distribuida, en la que
La escala añade problemas de distribución y coordinación (BATINI; SCANNAPIECO, 2006; ÖZSU; VALDURIEZ, 2020).

La literatura analizada respalda el rechazo de una arquitectura universal. Stonebraker y Çetintemel
(2005) formuló esta crítica en relación con diferentes clases de aplicaciones y la documentación.
El PostgreSQL moderno incorpora diversidad a través de múltiples métodos de indexación y
Mecanismos internos. La persistencia de esta variedad refuerza la idea de que la especialización sigue siendo relevante.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinar El Conocimiento. ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

Otro resultado del análisis es la continuidad entre las bases de datos operativas y la inteligencia empresarial.
Kimball y Ross (2013) muestran que los entornos analíticos tienen su propio modelado, pero
Dependen de los datos generados en sistemas anteriores. Por lo tanto, la inteligencia empresarial no reemplaza la administración de empresas.
datos; amplía la cadena de responsabilidad en cuanto a calidad, historial e interpretación.
La procedencia añade una capa esencial a esta cadena. Glavic (2021) demuestra que
Comprender el origen y la derivación es un problema de investigación bien establecido. En escenarios de IA, esto...
La necesidad tiende a intensificarse porque los resultados pueden depender de los conjuntos de datos y
Transformaciones que no son visibles para el usuario final. La trazabilidad se convierte en un componente de
confianza.
Finalmente, la información geográfica muestra cómo una formación académica en Geomática puede ser
integrado en el campo de datos sin desviar la atención hacia la cartografía o la interoperabilidad.
El almacenamiento espacial y las consultas son problemas legítimos de las bases de datos.
El planteamiento es técnicamente coherente y evita dar una importancia indebida a experiencias periféricas.
cuando el núcleo profesional reside en los datos y la relación entre la práctica y la academia.

9. CONCLUSIÓN

La ingeniería de datos debe entenderse como una disciplina que preserva las condiciones necesarias para...
La información sigue siendo útil a medida que atraviesa diferentes sistemas, tecnologías y propósitos. La persistencia es un valor.
Esto es solo el comienzo del proceso. La arquitectura también necesita lidiar con la competencia, la recuperación,
Historia, indexación, calidad e interpretación.
Las fuentes analizadas muestran que diferentes cargas de trabajo requieren diferentes elecciones. La crítica de
Stonebraker y Çetintemel (2005) sostienen que la idea de una única solución sigue siendo compatible con la variedad.
de estructuras observadas en las plataformas actuales. Métodos de indexación, modelos transaccionales y
Las arquitecturas analíticas abordan problemas específicos.
La información geográfica refuerza esta conclusión. Las relaciones espaciales requieren representación y
Mecanismos de acceso compatibles con proximidad, intersección y contención. El caso demuestra que
Es necesario considerar conjuntamente el tipo de datos y el tipo de pregunta. Árbol R de Guttman.
(1984) es un hito en esta especialización, ya que fue diseñado específicamente para la búsqueda espacial de
objetos multidimensionales.
La inteligencia empresarial amplía el ciclo de vida al reorganizar los datos operativos para su análisis.
El modelado dimensional descrito por Kimball y Ross (2013) muestra que la granularidad, los hechos,
Las dimensiones y los datos históricos son decisiones que afectan directamente a la interpretación de los indicadores.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

La procedencia y los metadatos añaden poder explicativo. Glavic (2021) muestra que el origen y la derivación constituyen un campo de investigación consolidado, mientras que los datos de investigación del NIST El marco ofrece una perspectiva de gestión a lo largo del ciclo de vida (HANISCH et al., 2024). Estos Estos elementos son particularmente importantes cuando múltiples transformaciones separan la fuente de la resultado.

Los principios FAIR de Wilkinson et al. (2016) también nos recuerdan que la utilidad depende de las condiciones. de descubrimiento, acceso, interpretación y reutilización. Aunque se originan en el contexto de los datos. Los estudios científicos ayudan a reforzar la idea de que almacenar algo no equivale a hacerlo comprensible y reutilizable.

La expansión de la computación en la nube y la inteligencia artificial aumenta la capacidad de computación disponible, pero Esto no cambia ese principio. Los modelos sofisticados siguen dependiendo de datos cuyo origen, La estructura y la calidad pueden entenderse. Automatizar decisiones basándose en una base mal definida. Un enfoque estructurado solo puede acelerar la propagación de inconsistencias.

Se puede concluir, por lo tanto, que las arquitecturas robustas resultan de la adecuación entre la naturaleza de la información, Uso previsto y mecanismos de persistencia y acceso. La evolución tecnológica cambia los productos y Es escalable, pero mantiene la necesidad de bases sólidas. Para entornos de datos, inteligencia empresarial e información.

Tanto desde el punto de vista geográfico como mediante la inteligencia artificial, la confianza se sigue construyendo antes de que aparezca la capa visible de la aplicación.

REFERENCIAS

ABITEBOUL, Serge; HULL, Richard; VIANU, Victor. Fundamentos de bases de datos. Reading, MA: Addison-Wesley, 1995. ISBN 0-201-53771-0.

ARMBRUST, Michael; GHODSI, Ali; XIN, Reynold; ZAHARIA, Matei. Lakehouse: Una nueva generación de plataformas abiertas que unifican el almacenamiento de datos y el análisis avanzado. En: CONFERENCIA SOBRE INVESTIGACIÓN EN SISTEMAS DE DATOS INNOVADORES, 11., 2021. Actas [...]. CIDR, 2021.

BATINI, Carlo; SCANNAPIECO, Monica. Calidad de los datos: conceptos, metodologías y técnicas. Berlín: Springer, 2006. DOI: 10.1007/3-540-33173-5.

CHAUDHURI, Surajit; DAYAL, Umeshwar. Una visión general del almacenamiento de datos y la tecnología OLAP. ACM SIGMOD Record, vol. 26, n.º 1, págs. 65-74, 1997. DOI: 10.1145/248603.248616.

GLAVIC, Boris. Procedencia de los datos: orígenes, aplicaciones, algoritmos y modelos. Foundations and Trends in Databases, vol. 9, n.º 3-4, págs. 209-441, 2021. DOI: 10.1561/1900000068.

GUTTMAN, Antonin. Árboles R: una estructura de índice dinámica para la búsqueda espacial. En: Conferencia Internacional ACM SIGMOD sobre Gestión de Datos, 1984. Actas [...]. Nueva York: ACM, 1984. págs. 47-57. DOI: 10.1145/602259.602266.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 2, 2026 | Recepción: 15 de julio de 2026 | Aceptación: 20 de julio de 2026 | Publicación: 25 de julio de 2026

HANISCH, Robert J. et al. Marco de datos de investigación del NIST (RDaF): Versión 2.0. Gaithersburg: Instituto Nacional de Estándares y Tecnología, 2024. NIST SP 1500-18r2. DOI: 10.6028/NIST.SP.1500-18r2.

HELLERSTEIN, Joseph M.; STONEBRAKER, Michael; HAMILTON, James. Arquitectura de un sistema de base de datos. Foundations and Trends in Databases, vol. 1, n.º 2, págs. 141-259, 2007. DOI: 10.1561/1900000002.

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

ÖZSU, M. Tamer; VALDURIEZ, Patrick. Principios de sistemas de bases de datos distribuidas. 4.ª ed. Cham: Springer, 2020. DOI: 10.1007/978-3-030-26253-2.

GRUPO DE DESARROLLO GLOBAL DE POSTGRESQL. Documentación de PostgreSQL 18. 2026. Disponible en: https://www.postgresql.org/docs/18/.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes. Sistemas de gestión de bases de datos. 3ª edición. Nueva York: McGraw-Hill, 2003. ISBN 0-07-246563-8.

STONEBRAKER, Michael; ÇETINTEMEL, Uğur. Talla única para todos: una idea cuyo tiempo ya pasó. En: CONFERENCIA INTERNACIONAL SOBRE INGENIERÍA DE DATOS, 21.ª, 2005. Actas [...]. IEEE, 2005. DOI: 10.1109/ICDE.2005.1.

WILKINSON, Mark D. et al. Los principios rectores FAIR para la gestión y administración de datos científicos. Scientific Data, vol. 3, art. 160018, 2016. DOI: 10.1038/sdata.2016.18.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.