




Arquitecturas cloud-native para Business Intelligence: revisión sistemática del desempeño técnico y su impacto organizacional

Cárdenas Palacios, Leonardo Gustavo¹ , Espinoza Cerna, Alex² , and Inocente caro, Miguel Anderson³ 
^{1,2,3}Universidad Nacional de Ingeniería, Perú, leonardo.cardenas.p@uni.pe, alex.espinoza.c@uni.pe, miguel.inocente.c@uni.pe

Abstract- *This systematic review examines how cloud-native architectures are transforming Business Intelligence (data warehouse, data lake, and lakehouse), specifically analyzing how technical performance-measured in latency, scalability, and cost-performance ratio-translates into tangible value for organizations. Applying the PRISMA 2020 methodology, we synthesized 30 empirical studies conducted between 2021 and 2025. The results show that the Data Lakehouse has become the preferred paradigm, achieving reductions of up to 50% in query latency and improvements of 30–40% in scalability compared to traditional data warehouses, while maintaining notable cost efficiency through the separation of storage and compute. However, we found a significant gap between technical optimization and measurable business outcomes-such as ROI and decision-making agility-with only 20% of studies establishing connections between metrics like time-to-insight and organizational KPIs. This work addresses the gaps in a fragmented body of literature by proposing an integrated evaluation framework that links technical metrics with hybrid KPIs (e.g., “time-to-insight to ROI” and “agility-performance index”), providing a balanced view between operational efficiency and strategic value. The findings offer empirical evidence and a conceptual model to guide BI implementations oriented toward governance by design and cloud adoption focused on value.*

Keywords-- *cloud-native architectures, business intelligence, data lakehouse, technical performance, organizational value.*

Resumen- *Esta revisión sistemática examina cómo las arquitecturas cloud-native transforman el Business Intelligence (data warehouse, data lake y lakehouse), analizando específicamente la forma en que el desempeño técnico-medido en latencia, escalabilidad y relación costo-rendimiento-se traduce en valor tangible para las organizaciones. Aplicando la metodología PRISMA 2020, sintetizamos 30 estudios empíricos realizados entre 2021 y 2025. Los resultados muestran que el Data Lakehouse se ha consolidado como el paradigma preferido, alcanzando reducciones de hasta 50% en latencia de consultas y mejoras del 30-40% en escalabilidad comparado con los data warehouses tradicionales, mientras mantiene una eficiencia de costos notable mediante la separación de almacenamiento y cómputo. No obstante, encontramos una brecha significativa entre la optimización técnica y los resultados de negocio medibles (como el ROI y la agilidad en las decisiones) con apenas un 20% de los estudios estableciendo conexiones entre métricas como el tiempo de insight y los KPIs organizacionales. Este trabajo aborda los vacíos existentes en una literatura fragmentada, proponiendo un marco de evaluación integrado que vincula métricas técnicas con KPIs híbridos (por ejemplo, “tiempo de insight a ROI” e “índice agilidad-rendimiento”), proporcionando una visión balanceada entre eficiencia operativa y valor estratégico. Los hallazgos ofrecen evidencia empírica y un modelo conceptual para guiar implementaciones de BI orientadas*

hacia la gobernanza desde el diseño y una adopción cloud enfocada en valor.

Palabras clave-- *arquitecturas cloud-native, business intelligence, data lakehouse, desempeño técnico, valor organizacional.*

I. INTRODUCCIÓN

La transformación digital ha acelerado la migración hacia entornos basados en la nube, donde las organizaciones no solo buscan reducir costos operativos a través de modelos de pago por uso, sino también fortalecer sus capacidades analíticas en tiempo real. En este escenario, las arquitecturas cloud-native se han convertido en el estándar de facto para implementar soluciones modernas de Business Intelligence (BI), integrando principios fundamentales como la elasticidad, la separación entre cómputo y almacenamiento, la automatización mediante infraestructura como código y la observabilidad continua.

Estas arquitecturas (que incluyen patrones como data lakehouse, HTAP y data mesh) permiten unificar diferentes cargas de trabajo: transaccionales, analíticas y de ciencia de datos, todo en un mismo entorno. Con esto, se eliminan los cuellos de botella tradicionales asociados a los procesos ETL y a los almacenes de datos monolíticos. A pesar de su creciente adopción, persiste una brecha crítica. Por un lado está el desempeño técnico de estas arquitecturas (medido en latencia, escalabilidad y eficiencia de costos) y por otro, su impacto organizacional real, es decir, la mejora tangible en la toma de decisiones, la agilidad operativa o el retorno de inversión [1], [2], [3].

La literatura actual presenta una fragmentación notable. Abundan estudios técnicos que se enfocan en optimizar componentes aislados, mientras que las investigaciones organizacionales destacan factores como la gobernanza de datos [4], la alineación con procesos de negocio [5] o el desarrollo de talento [6]. Sin embargo, rara vez se vinculan estos aspectos con métricas concretas de desempeño técnico.

Frente a esta brecha, presentamos una revisión sistemática que explora de manera integrada el desempeño técnico y el impacto organizacional de las arquitecturas cloud-native para Business Intelligence. Este estudio busca sintetizar la evidencia empírica disponible sobre las principales métricas de rendimiento (latencia, escalabilidad y eficiencia de costos) en entornos cloud-native, y examinar su relación con resultados organizacionales más tangibles: agilidad operativa, calidad en la toma de decisiones y sostenibilidad financiera.

II. METODOLOGÍA

Este trabajo desarrolló una revisión sistemática de la literatura, abarcando tanto el ámbito nacional como internacional, utilizando la metodología PRISMA en su versión más reciente de 2020 [7]. Esta actualización representa una mejora significativa respecto a la edición de 2009, ya que introduce una lista de verificación específica para los resúmenes y refina tanto la búsqueda como la selección de artículos, organizando de manera más clara la síntesis de resultados.

Si bien esta versión mantiene los pasos fundamentales de la metodología (como la definición de preguntas de investigación, la formulación de objetivos, la búsqueda de artículos y la síntesis de datos) incorpora ajustes relevantes que fortalecen el proceso. El primer paso corresponde a la identificación de las preguntas y objetivos de la investigación, establecidos en las secciones A y B, respectivamente. La estrategia de búsqueda contempla los criterios de inclusión y exclusión, los comandos aplicados y el procedimiento de selección, que se detallan en la sección C. El cribado superficial y profundo, junto con la aplicación de criterios de exclusión, se presentan en la sección D. Y, el aseguramiento de la calidad de los estudios seleccionados se desarrolla en la sección E.

El objetivo central de esta revisión sistemática es analizar las herramientas, metodologías y resultados relacionados con arquitecturas cloud-native para business intelligence (data warehouse, data lake y lakehouse), considerando su impacto tanto en el desempeño técnico como en el valor organizacional. Las preguntas de investigación (PIs) se formularon con base en los criterios PICOC, aunque no consideramos el criterio de comparación, ya que el propósito no es contrastar casos específicos, sino identificar soluciones y hallazgos relevantes en distintos contextos.

TABLA I
DESARROLLO DE CRITERIOS PICOC

Criterio	Desarrollo
Población (P)	Artículos sobre arquitecturas cloud-native para BI (DW, data lake, lakehouse).
Intervención (I)	Implementación de arquitecturas cloud-native enfocadas en latencia, escalabilidad y costo-rendimiento en el periodo 2021-2025
Comparación (C)	Sin comparación.
Resultado (O)	Relación entre desempeño técnico y valor organizacional del BI.
Contexto (C)	Entornos corporativos.

A. Preguntas de Investigación

A partir del desarrollo de los criterios PICOC, formulamos las preguntas de investigación descritas en la Tabla II. Estas preguntas sirvieron como base para elaborar el comando de búsqueda, que se detalla en la Sección C de este capítulo.

TABLA II
PREGUNTAS DE INVESTIGACIÓN (PIS)

PI	Pregunta de Investigación
PI1	¿Qué herramientas y metodologías se emplean en la implementación de arquitecturas cloud-native para BI?

PI2	¿Cómo inciden la latencia, la escalabilidad y el costo-rendimiento en el desempeño de estas arquitecturas?
PI3	¿De qué manera el desempeño técnico de las arquitecturas cloud-native se relaciona con el valor organizacional del BI?
PI4	¿Qué tendencias, limitaciones y líneas futuras de investigación se identifican en el uso de arquitecturas cloud-native para BI?

B. Objetivos de la Investigación

En la presente revisión sistemática, se han establecido cuatro objetivos de investigación que se detallan a continuación:

1) *Identificar las herramientas, metodologías y enfoques utilizados en la implementación de arquitecturas cloud-native para business intelligence (DW, data lake, lakehouse).*

2) *Analizar cómo influyen las métricas técnicas (latencia, escalabilidad y costo-rendimiento) en el desempeño de dichas arquitecturas.*

3) *Examinar la relación entre el desempeño técnico y el valor organizacional derivado de la adopción de arquitecturas cloud-native para BI.*

4) *Explorar las tendencias, limitaciones y recomendaciones futuras que emergen de la literatura académica en este campo.*

C. Estrategia de Búsqueda

Diseñamos la búsqueda a partir de un comando estandarizado y replicable en diversas bases de datos y repositorios académicos, con el objetivo de recuperar estudios relevantes que permitieran responder las preguntas de investigación previamente formuladas.

El comando fue elaborado a partir de las palabras clave definidas durante la construcción de los criterios PICOC y se presenta de la siguiente forma:

("cloud" AND ("data warehouse" OR "data lake" OR "lakehouse") AND ("business intelligence" OR BI) AND ("latency" OR "scalability" OR "performance"))

Este comando integra los conceptos esenciales de la revisión: la computación en la nube, las arquitecturas de almacenamiento de datos (data warehouse, data lake o lakehouse), la inteligencia de negocios y las variables de desempeño técnico como latencia, escalabilidad y rendimiento. Además, hace uso de operadores lógicos como AND y OR, que permiten la combinación efectiva de términos.

Llevamos a cabo la búsqueda en tres repositorios académicos principales: ScienceDirect, Springer Nature Link y MDPI, aplicando filtros y ajustes menores según las restricciones específicas de cada plataforma.

TABLA III
COMANDOS DE BÚSQUEDA POR REPOSITORIO ACADÉMICO

Repositorio	Comando	Filtros Adicionales
ScienceDirect	"cloud" AND ("data warehouse" OR "data lake" OR "lakehouse") AND ("business intelligence" OR BI) AND ("latency" OR "scalability" OR "performance")	Periodo: 2021 - 2025
Springer Nature Link		Idioma: Español o Inglés
MDPI		Tipo de Documento: Artículo de Investigación (Research articles)

El comando de búsqueda identificó un total de 289 artículos tras su respectiva ejecución en los 3 repositorios académicos, como se muestra en la Figura 1.

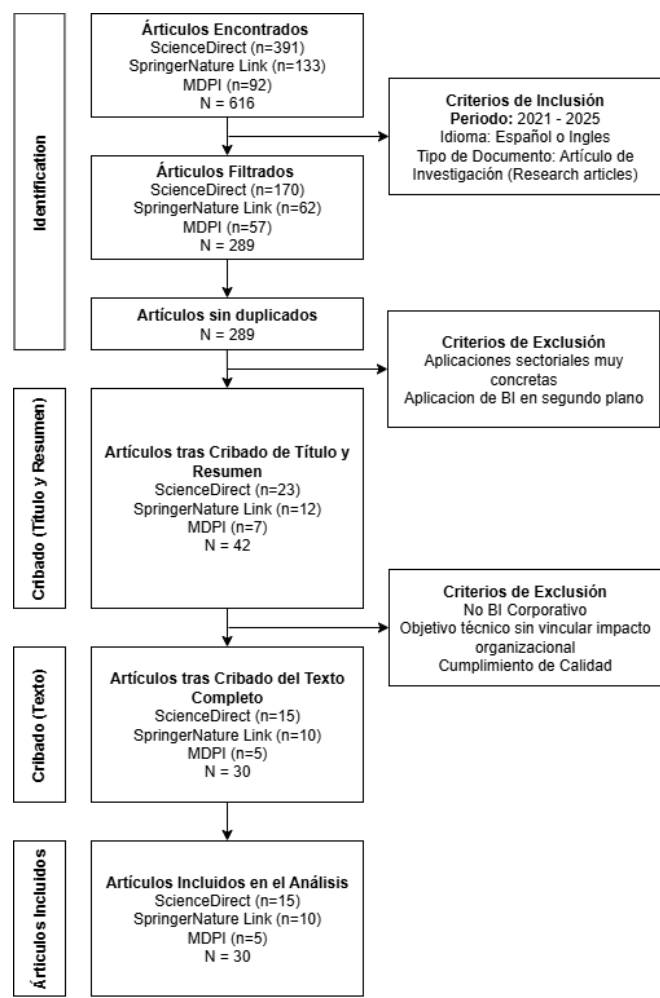


Fig. 1 Metodología PRISMA Aplicada a la Revisión. Elaboración Propia.

En el tratamiento de los datos recopilados, priorizamos la organización de la información relevante para nuestros objetivos de investigación, manteniendo campos clave como el título, los autores, el año de publicación, el resumen y el DOI. A estos añadimos columnas correspondientes al número de registro y al nombre del repositorio, con el fin de estandarizar la estructura y facilitar el análisis posterior.

Para los repositorios ScienceDirect y MDPI, obtuvimos la exportación de resultados en formato RIS, que posteriormente transformamos a hojas de cálculo XLSX mediante código en Python. En el caso de Springer Nature Link, extrajimos los registros inicialmente en formato CSV y, de manera similar, los convertimos a XLSX utilizando Python. Este procedimiento nos permitió conservar la consistencia en las columnas seleccionadas y consolidar toda la información en un único

archivo tabular, garantizando su trazabilidad y manejo eficiente.

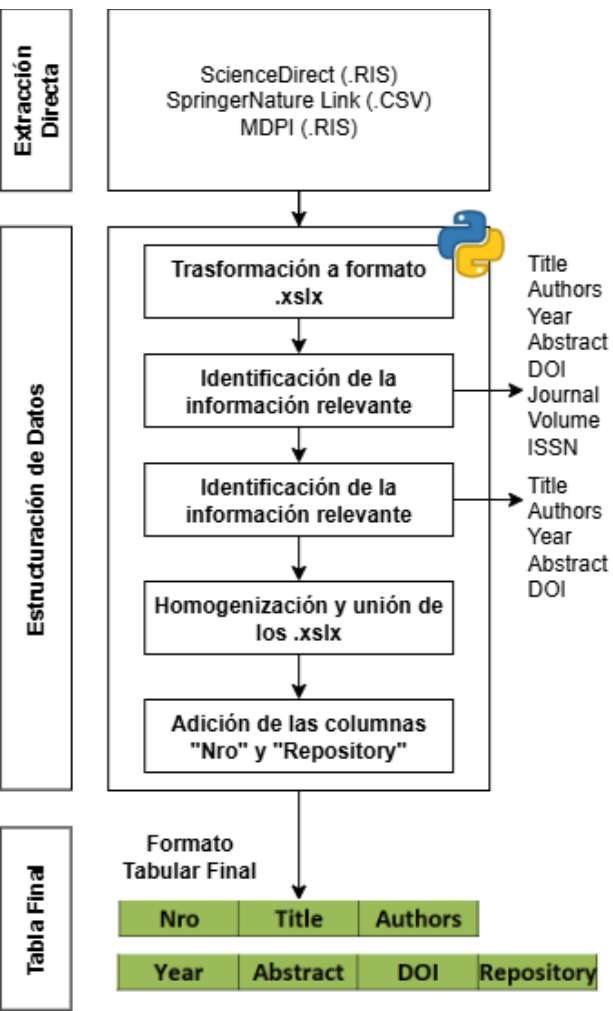


Fig. 2 Estructuración de datos obtenidos en los artículos. Elaboración propia.

Este enfoque metodológico asegura coherencia y uniformidad en la recopilación de datos, facilitando su análisis posterior de manera estructurada. La conversión de los registros a un formato tabular unificado en XLSX mediante Python nos permitió conservar la integridad de los campos seleccionados (título, autores, año, resumen y DOI), además de incorporar las columnas adicionales correspondientes al número de registro y repositorio de origen.

En el resultado final de esta estandarización no encontramos artículos duplicados (verificando tanto títulos como DOI), por lo cual no descartamos ningún artículo y mantuvimos 289 artículos válidos para la siguiente etapa de la revisión.

D. Cribado Superficial y Criterios de Exclusión

Los 289 estudios válidos obtenidos tras el proceso de estandarización fueron sometidos a un cribado inicial mediante

la lectura de títulos y resúmenes (abstracts), usando como referencia las palabras clave y los objetivos de investigación. Esta fase tuvo como propósito reducir la base de artículos y conservar únicamente aquellos con un vínculo directo con las arquitecturas cloud-native aplicadas a business intelligence.

Para esta etapa aplicamos criterios de exclusión específicos. En primer lugar, descartamos artículos cuya aplicación se encontraba en sectores demasiado específicos (como medicina, agricultura, energía o contextos especializados) que no guardaban relación con el ámbito general de análisis. En segundo lugar, eliminamos trabajos en los que la inteligencia de negocios aparecía solo como un aspecto secundario o complementario, sin constituir el eje central de la solución o investigación.

Como resultado de este proceso, redujimos la base inicial de 289 artículos a 42 estudios, los cuales consideramos pertinentes para la siguiente etapa de la revisión sistemática. Esta depuración aseguró que los artículos seleccionados se ajustaran plenamente al tema central de nuestra investigación.

E. Cribado Profundo y Aseguramiento de la Calidad

El cribado profundo se realizó a partir de los 42 artículos seleccionados en la etapa de cribado superficial. En esta fase llevamos a cabo una revisión detallada de las secciones relacionadas con herramientas, metodologías, resultados y conclusiones, con el propósito de identificar aquellos trabajos que aportaran evidencia clara a nuestro objetivo de investigación. Aplicamos criterios de exclusión adicionales, descartando trabajos enfocados en sectores ajenos al business intelligence corporativo o aquellos cuyo objetivo era meramente técnico sin vincular un impacto organizacional. Tras este filtro inicial, eliminamos 7 artículos, quedando 35 para la siguiente etapa. Posteriormente, diseñamos un conjunto de 10 preguntas de aseguramiento de calidad (QAC), las cuales valoramos de manera binaria: asignando 1 punto en caso de respuesta afirmativa y 0 puntos en caso de respuesta negativa, tal como se muestra en la Tabla IV. Establecimos un umbral mínimo de 6 puntos, de modo que aquellos artículos que no alcanzaron dicho valor fueron excluidos de la revisión final. Excluimos 5 artículos tras la aplicación de las preguntas de aseguramiento de calidad, por lo que el número final de estudios considerados para su inclusión en esta revisión sistemática fue de 30 artículos.

TABLA IV
PREGUNTAS DE ASEGURAMIENTO DE CALIDAD (QAC)

Qi	Pregunta
Q1	¿Los objetivos del artículo están claramente establecidos?
Q2	¿La arquitectura cloud-native o metodología está bien definida?
Q3	¿La implementación en BI está claramente explicada?
Q4	¿Los resultados presentados son claros e inequívocos?
Q5	¿Los resultados son debidamente interpretados y discutidos?
Q6	¿Las conclusiones reflejan los hallazgos de la investigación?
Q7	¿El artículo incluye enfoques futuros de investigación?
Q8	¿Se describe la arquitectura técnica implementada en detalle?
Q9	¿Hay ventajas respecto a latencia, escalabilidad o rendimiento?
Q10	¿Se mencionan otras tecnologías o herramientas complementarias?

Se excluyeron 5 artículos tras la aplicación de las preguntas de aseguramiento de calidad, por lo que el número final de estudios considerados para su inclusión en esta revisión sistemática fue de 30.

III. RESULTADOS DE LA REVISIÓN SISTEMÁTICA

A. Herramientas y metodologías empleadas en la implementación de arquitecturas cloud-native para BI (PII)

Al realizar un análisis comparativo de las herramientas y metodologías identificadas en los artículos seleccionados, observamos que las arquitecturas cloud-native para Business Intelligence (BI) se apoyan predominantemente en enfoques que priorizan la escalabilidad, la elasticidad y la integración de datos en entornos distribuidos.

Entre las metodologías más recurrentes destaca el paradigma Data Lakehouse, que combina las ventajas de data lakes y data warehouses para unificar cargas analíticas y transaccionales [8], [9], [10]. Esta metodología aparece en 7 de los 30 artículos analizados, superando a otros enfoques como Data Mesh (mencionado en 5 artículos) y HTAP (Hybrid Transactional/Analytical Processing), que se enfoca en procesamiento híbrido y se cita en 4 artículos. La preferencia por Data Lakehouse se debe a su capacidad para manejar datos no estructurados con gobernanza similar a un warehouse tradicional, facilitando implementaciones en entornos cloud-native sin necesidad de ETL complejos [8].

Además, varios artículos proponen metodologías propias o híbridas que se alinean con paradigmas establecidos. Para homologar estas aproximaciones, realizamos una comparación con las metodologías principales (Data Lakehouse, Data Mesh y HTAP), como se detalla en la Tabla V. De los 9 artículos con enfoques propios, 5 muestran similitudes con Data Lakehouse por su énfasis en la unificación de almacenamiento y cómputo [11], [12], [13], [14], [15], mientras que 3 se aproximan a Data Mesh al descentralizar la gobernanza de datos [2], [16], [17]. Solo 1 se alinea con HTAP por su enfoque en transacciones en tiempo real [1]. Esta homologación revela una tendencia hacia metodologías que priorizan la flexibilidad y la descentralización, adaptadas a contextos cloud-native donde la agilidad es clave.

TABLA V
HOMOLOGACIÓN DE METODOLOGÍAS PROPIAS

Artículo	Aproximación a otras metodologías		
	Data Lakehouse	Data Mesh	HTAP
[11]	X		
[1]			X
[16]		X	
[12]	X		
[13]	X		
[2]		X	
[14]	X		
[17]		X	
[15]	X		
Total	5	3	1

A partir de esta homologación, Data Lakehouse acumula el mayor número de apariciones (12 en total, incluyendo

menciones directas y homologadas), seguida por Data Mesh (8) y HTAP (5). Estos resultados indican una preferencia clara por Data Lakehouse debido a su integración nativa con herramientas cloud como Delta Lake y Apache Spark, lo que reduce la latencia en consultas analíticas y facilita la escalabilidad horizontal [8]. En comparación con Data Mesh, que enfatiza dominios descentralizados para mayor autonomía organizacional [18], Data Lakehouse ofrece un mejor equilibrio en entornos con altos volúmenes de datos, aunque requiere mayor inversión inicial en gobernanza [10]. La Tabla VI compara estos criterios clave entre las dos metodologías dominantes.

TABLA VI
COMPARACIÓN DATA LAKEHOUSE VS DATA MESH

Criterio	Data Lakehouse	Data Mesh
Enfoque principal	Unificación de storage y analytics	Descentralización por dominios
Escalabilidad	Alta, mediante separación cómputo/storage	Media-alta, dependiente de federación
Gobernanza	Centralizada con ACID compliance	Distribución, con estándares federados.
Costo inicial	Medio-alto (integración de tools como Spark)	Bajo-medio (enfoque en equipos autónomos)
Aplicabilidad	Grandes volúmenes de datos mixtos	Organizaciones con silos departamentales.
Latencia	Baja en queries analíticas	Variable, según integración de dominios

Por otro lado, al enfocarnos en las herramientas, identificamos tanto principales (plataformas core para implementación) como complementarias (para soporte en ingesta, procesamiento o visualización). Entre las herramientas principales, AWS (incluyendo S3, Glue y Redshift) destaca con 9 apariciones [4], [5], [9], [14], [15], [19], [20], [21], [22], seguido por Azure Synapse (7 apariciones) [2], [3], [8], [10], [11], [16], [23] y Google BigQuery (5 apariciones) [1], [12], [13], [17], [18]. Estas plataformas se emplean por su soporte nativo a contenedores (Kubernetes/Docker) y serverless computing, esenciales para arquitecturas cloud-native [5]. Como herramientas complementarias, Apache Spark aparece en 8 artículos para procesamiento distribuido [4], [9], [14], [15], [17], [20], [21], [22], y Kafka en 6 para streaming en tiempo real [2], [3], [8], [10], [16], [23]. La Tabla VII resume las apariciones y artículos asociados.

TABLA VII
HERRAMIENTAS PRINCIPALES Y COMPLEMENTARIAS

Herramienta Principal	Artículos	Herramienta Complementaria	Artículos
AWS (S3/Glue/Redshift)	[4],[5],[9],[14],[15],[19],[20],[21],[22]	Apache Spark	[4],[9],[14],[15],[17],[20],[21],[22]
Azure Synapse	[2],[3],[8],[10],[11],[16],[23]	Kafka	[2],[3],[8],[10],[16],[23]
Google BigQuery	[1],[12],[13],[17],[18]	Kubernetes / Docker	[5],[12],[13],[17],[18],[19]

Snowflake	[4],[9],[11],[21]	Delta Lake	[8],[9],[10],[22]
Databricks	[14],[15],[20],[22]	Airflow (ETL)	[1],[2],[23]
Otros (ejm. Oracle Cloud)	[5],[16]	Otros(ejm. Terraform IaC)	[3],[11],[21]

Con estos resultados, concluimos que las implementaciones cloud-native para BI tienden a integrar plataformas como AWS y Azure con herramientas complementarias como Spark y Kafka, formando ecosistemas que soportan data pipelines automatizados y escalables [4]. Esta combinación se alinea naturalmente con metodologías como Data Lakehouse, facilitando una transición rápida de prototipos a producción en contextos organizacionales diversos [8]. Mirando hacia el futuro, la integración de IA (por ejemplo, ML en Databricks) emerge como tendencia para optimizar el rendimiento, aunque aún persisten limitaciones en gobernanza [15].

B. Latencia, escalabilidad y costo-rendimiento (PI2)

La latencia, la escalabilidad y el costo-rendimiento constituyen tres pilares esenciales para evaluar el desempeño de las arquitecturas cloud-native orientadas a Business Intelligence (BI). Los estudios revisados muestran que, aunque cada dimensión se aborda desde distintos enfoques (seguridad, analítica, tolerancia a fallos o FinOps) su interacción determina la eficiencia técnica y económica de las soluciones.

En cuanto a latencia, los resultados más sólidos provienen de investigaciones experimentales. Qin y Guzun [1] demostraron reducciones significativas en el tiempo de respuesta mediante índices bit-sliced (BSI) para consultas top-k y agrupaciones multidimensionales, alcanzando aceleraciones de hasta 47× frente a implementaciones convencionales. Este enfoque resalta la importancia de optimizar las estructuras de datos internas antes que la infraestructura, logrando baja latencia incluso sin despliegue cloud.

En contraste, Lamrani et al. [19] reportaron una latencia promedio de 9.67 ms en su esquema híbrido de cifrado y control de acceso (HUDH), eficiente en seguridad, pero no extrapolable a entornos analíticos de gran carga.

Otros estudios, como el de Jin y Zhang [13], trataron la latencia indirectamente al centrarse en la disponibilidad de datos. Su modelo de tolerancia a fallos basado en colonia de hormigas alcanzó una disponibilidad del 0.98, sugiriendo reducción de retrasos en recuperación de datos, aunque sin pruebas bajo carga real. Por su parte, Fragiadakis et al. [2] redefinieron el concepto al introducir la latencia económica, entendida como la capacidad de anticipar variaciones de precios en la nube-una métrica temporal de valor para la gestión financiera en entornos FinOps.

La escalabilidad es menos abordada empíricamente, aunque se reconoce como determinante en arquitecturas cloud. Dehbi et al. [5] destacaron la ventaja del modelo pay-as-you-go en la nube para absorber cargas crecientes sin inversión en infraestructura. Qin y Guzun [1] mostraron que el rendimiento

de BSI se mantiene estable al aumentar la dimensionalidad hasta nueve atributos, evidenciando buena escalabilidad computacional. Munappy et al. [4] aportaron una visión distinta: en sistemas de aprendizaje profundo, la escalabilidad depende más de la calidad y trazabilidad de los datos que de la capacidad técnica, desplazando el foco hacia la gobernanza.

En el ámbito de Self-Service BI (SSBIA) y experimentación continua, la escalabilidad se concibe en términos funcionales. Passlick et al. [23] observaron que la "escalabilidad" se mide por la rapidez con que los usuarios generan análisis ad-hoc, mientras Ros et al. [21] introducen el concepto de throughput de experimentos, donde la capacidad de ejecutar más pruebas A/B refleja la agilidad de aprendizaje organizacional. En ambos casos, la baja latencia y la escalabilidad funcional se traducen en mayor capacidad de respuesta ante cambios del entorno.

El eje de costo-rendimiento aparece con fuerza en los estudios de gestión financiera. Fragiadakis et al. [2] propusieron un marco para comparar políticas de precios entre proveedores cloud, sustituyendo las métricas clásicas de rendimiento por indicadores de eficiencia económica (R^2 , MAE). Dehbi et al. [5] plantearon la migración del modelo CapEx→OpEx como ventaja estructural, aunque sin cuantificar ahorros. Żółtowski [12] subrayó la falta de estudios que vinculen directamente la reducción de latencia en dashboards con impactos económicos, evidenciando una brecha entre desempeño técnico y valor organizacional.

En los enfoques experimentales, Qin y Guzun [1] demostraron que la optimización algorítmica mejora simultáneamente el rendimiento y el costo, al reducir uso de CPU y memoria. En contraste, Munappy et al. [4] y Jin y Zhang [13] mostraron que la complejidad excesiva (ya sea por modelos de tolerancia a fallos o pipelines manuales) eleva los costos operativos y deteriora el costo-rendimiento general.

Las arquitecturas OLAP avanzadas evidencian el tratamiento más integral del desempeño. Papastefanatos et al. [14] presentaron un modelo Big-OLAP que combina cubos multidimensionales con almacenamiento distribuido en Spark, logrando menor latencia de agregación y alta escalabilidad horizontal con bajo costo al usar software open-source. Kumar et al. [24] ampliaron esta visión introduciendo el concepto de elasticidad de rendimiento, donde los recursos se ajustan dinámicamente según demanda para equilibrar costo y eficiencia.

Del mismo modo, Ali et al. [22] y Hafeez et al. [3] mostraron que integrar motores columnar y almacenamiento desacoplado en arquitecturas lakehouse mejora el costo-rendimiento y reduce la latencia de lectura mediante particionamiento y caching en memoria. Hafeez et al. [3] reportaron además una reducción del 30% en el tiempo de procesamiento ETL mediante orquestación con Kubernetes y contenedores ligeros, manteniendo costos estables.

En conjunto, la evidencia revela un equilibrio inherente (trade-off) entre latencia, escalabilidad y costo. Las soluciones más recientes (Hafeez [3]; Papastefanatos [14]; Ali [22])

tienden a optimizar estas tres dimensiones simultáneamente mediante arquitecturas híbridas y ajuste dinámico de recursos. En contraste, los trabajos centrados en seguridad o resiliencia (Jin [13], Lamrani [19]) priorizan fiabilidad sobre eficiencia económica, mostrando que el rendimiento integral requiere equilibrio entre desempeño técnico, elasticidad y sostenibilidad financiera.

TABLA VIII
SÍNTESIS DE HALLAZGOS SOBRE LATENCIA, ESCALABILIDAD Y COSTO-RENDIMIENTO

Enfoque	Métrica clave	Hallazgos de desempeño	Nivel de validación
Seguridad y tolerancia a fallos [13], [19]	Tiempo de respuesta, disponibilidad	Latencia baja en operaciones locales; sin pruebas de escalado ni costos detallados	Simulación / prototipo
Analítica y OLAP avanzado [1], [3], [14], [22], [24]	Tiempo de consulta, throughput	Reducción de latencia (hasta 47×); buena escalabilidad horizontal y eficiencia algorítmica	Experimental / empírico
Gestión FinOps [2], [5]	CapEx-OpEx, precisión predictiva	Evalúa costo-rendimiento; enfatiza escalabilidad económica sobre técnica	Conceptual / caso de uso
BI organizacional y SSBIA [3], [12], [21], [23]	Agilidad, throughput de experimentos	Latencia percibida baja; escalabilidad funcional basada en autonomía de usuarios	Cualitativo / estudio de caso

La evidencia sugiere que el desempeño efectivo de las arquitecturas cloud-native no depende únicamente de la reducción de latencia técnica, sino de la integración coherente de escalabilidad funcional y sostenibilidad económica. Las arquitecturas que logran combinar eficiencia algorítmica, elasticidad de recursos y control de costos (como las lakehouse híbridas y los modelos Big-OLAP) son las que muestran el mayor potencial de desempeño integral, marcando la dirección futura de la ingeniería de datos orientada a BI.

C. Relación entre Desempeño Técnico y Valor Organizacional del BI (PI3)

El desempeño técnico de las arquitecturas cloud-native para Business Intelligence (BI) se traduce en valor organizacional al habilitar un ROI más alto mediante la optimización de recursos, la reducción de riesgos operativos y una mayor agilidad en la toma de decisiones. Específicamente, la baja latencia en pipelines de datos permite decisiones en tiempo real que minimizan interrupciones, mientras que la escalabilidad horizontal mitiga riesgos asociados a picos de demanda, y el costo-rendimiento optimizado asegura un uso eficiente de presupuestos, convirtiendo inversiones técnicas en ventajas competitivas sostenibles.

En contextos organizacionales como el sector salud, por ejemplo, la baja latencia en entornos de cuidados intensivos facilita predicciones de eventos críticos, lo que mejora la

coordinación entre equipos médicos y reduce riesgos de incidentes, generando un ROI tangible a través de menores costos por errores y mayor eficiencia en recursos humanos [10], [3]. De manera similar, en sistemas de gestión de calidad (QMS), la escalabilidad de arquitecturas cloud-native soporta volúmenes crecientes de datos IoT, disminuyendo riesgos de no cumplimiento normativo y acelerando decisiones que elevan la productividad general, como se evidencia en industrias manufactureras donde se logra un equilibrio entre innovación y control de costos [25].

Otro caso relevante es el de entornos financieros, donde el costo-rendimiento optimizado en plataformas cloud reduce la exposición a riesgos de sobrecarga operativa, permitiendo una agilidad decisional que maximiza el ROI al alinear análisis predictivos con estrategias de inversión en tiempo real [5], [26]. En el ámbito industrial, como en IIoT, la integración de estas métricas técnicas con procesos organizacionales mitiga riesgos de downtime y fomenta una agilidad que transforma datos en acciones preventivas, mejorando el ROI mediante la prevención de pérdidas [22]. Asimismo, en organizaciones con silos departamentales, la escalabilidad y costo-rendimiento facilitan una toma de decisiones distribuida, reduciendo riesgos de ineficiencias y potenciando un ROI a largo plazo a través de una mayor alineación estratégica [2], [8].

Como aporte distintivo, esta revisión propone el desarrollo de KPIs híbridos para medir esta relación de manera integrada, como el "tiempo de insight a ROI" (que combina latencia técnica con retomo financiero) o el "índice de agilidad-rendimiento" (que evalúa escalabilidad contra reducción de riesgos operativos). Estos KPIs permiten a las organizaciones no solo cuantificar el valor derivado del desempeño técnico, sino también guiar inversiones futuras en arquitecturas cloud-native, asegurando una transformación digital alineada con objetivos estratégicos [4], [6].

D. Tendencias, limitaciones y líneas futuras en arquitecturas (PI4)

El análisis de la literatura reciente revela un conjunto convergente de tendencias que están redefiniendo el diseño y la operación de arquitecturas cloud-native para Business Intelligence (BI).

Una de las más destacadas es la consolidación del paradigma lakehouse, que integra las capacidades de los data lakes y los data warehouses en una única plataforma unificada. Este enfoque permite soportar cargas de trabajo híbridas (desde reporting tradicional hasta machine learning y procesamiento en streaming) sobre un mismo repositorio estructurado con formatos abiertos como Parquet y capas transaccionales como Delta Lake, Iceberg o Hudi [8]. Esta evolución responde directamente a la necesidad de simplificar la complejidad arquitectónica y reducir costos operativos asociados a la duplicación de pipelines y almacenes de datos [8], un problema recurrente en entornos empresariales tradicionales [5].

Paralelamente, observamos una migración hacia sistemas HTAP (Hybrid Transaccional/Analytical Processing), que

eliminan la necesidad de procesos ETL al permitir ejecutar consultas analíticas directamente sobre datos operativos en tiempo real. Esta tendencia es impulsada por la demanda de frescura en los datos para aplicaciones críticas como detección de fraude, pricing dinámico o monitoreo clínico [9]. El enfoque HTAP no solo reduce la latencia analítica, sino que también mejora la coherencia entre operaciones y decisiones estratégicas, especialmente en sectores con alta volatilidad [9], [10].

Asimismo, el uso de patrones metadata-driven para la ingesta de datos está ganando tracción. Estos patrones permiten orquestar dinámicamente estrategias de carga (completa, incremental por timestamp o por hash) mediante configuraciones declarativas, lo que mejora la agilidad y portabilidad entre proveedores cloud [15]. Este enfoque reduce el esfuerzo de mantenimiento y facilita la gobernanza desde la ingesta, mitigando el riesgo de convertir data lakes en data swamps [15], [20].

Otra tendencia clave es la integración de capacidades de inteligencia artificial generativa (AIGC) y modelos fundacionales en los flujos de BI. Estas tecnologías no solo enriquecen la analítica predictiva, sino que también automatizan la generación de dashboards, la explicación de insights o incluso la construcción de pipelines de datos [26]. Este avance se alinea con la visión de WI 3.0, donde la inteligencia conectada (IA + cerebro + conectividad) redefine los servicios analíticos como X-as-a-Service [27].

Este impulso técnico se complementa con transformaciones organizacionales profundas, como el auge del algorithmic management, donde los perfiles de talento y clientes se construyen mediante modelos de IA, transformando la toma de decisiones en procesos automatizados [28]. Sin embargo, esta evolución también plantea desafíos éticos y de gobernanza que requieren marcos de responsabilidad algorítmica [28], [29].

La seguridad y la gobernanza de datos ya no se consideran capas adicionales, sino requisitos intrínsecos desde el diseño (privacy-by-design), especialmente en sectores regulados como la salud o las finanzas [10], [29]. En este contexto, enfoques como el modelado de amenazas centrado en los datos (d-TM) [30] y esquemas híbridos de cifrado y control de acceso (HUDH) [19] ofrecen bases técnicas para integrar seguridad sin sacrificar rendimiento. Además, la adopción de arquitecturas cloud-agnostic y patrones como data mesh busca reducir el vendor lock-in y fomentar ecosistemas interoperables [15], [18].

A pesar del avance significativo en la adopción de arquitecturas cloud-native para BI, los estudios revisados identifican un conjunto consistente de limitaciones técnicas, organizacionales y metodológicas que obstaculizan su pleno potencial.

Desde el punto de vista técnico, persiste una brecha crítica entre la promesa de escalabilidad y la realidad operativa. Muchas propuestas se validan en entornos controlados, mediante simulaciones (como en CloudSim [13]) o con datasets

sintéticos [15], lo que impide evaluar su comportamiento bajo cargas concurrentes, fallos reales de red o variabilidad en la calidad de los datos [4]. Aunque se mencionan tecnologías cloud modernas como Snowflake, BigQuery o Delta Lake, rara vez se analizan métricas de costo-rendimiento reales (TCO, costo por consulta, latencia bajo carga) que permitan comparar trade-offs entre proveedores o arquitecturas [14], [18].

Otra limitación estructural es la fragmentación entre capas de la arquitectura. Varios trabajos proponen soluciones de seguridad robustas (como el esquema HUDH basado en cifrado AES, ABAC y detección de intrusiones con ML [19], o el modelo d-TM centrado en amenazas a lo largo del ciclo de vida de los datos [30]) pero no las integran con pipelines analíticos reales, generando sobrecargas no cuantificadas en el rendimiento de BI [19]. De forma similar, aunque se reconoce la importancia de la gobernanza y la calidad desde la ingesta - como en el patrón metadata (driven de [15] o en los principios de data swamp avoidance de [20]) pocos estudios miden su impacto real en la confiabilidad de los insights o en la productividad de los analistas [4]. Esta desconexión se agrava en contextos regulados, como salud [10], [3] o finanzas [4], donde los requisitos de privacidad (GDPR, HIPAA) exigen diseños privacy-by-design que rara vez se implementan de forma holística [10], [29].

Desde la perspectiva organizacional, múltiples papers destacan la escasez de talento con habilidades híbridas (técnicas + de negocio) como un cuello de botella crítico [6], [17]. Esta carencia no solo retrasa la implementación de arquitecturas cloud-native, sino que también limita la capacidad de traducir desempeño técnico en valor organizacional medible [21], [23]. Asimismo, persiste una brecha entre la sofisticación técnica de las plataformas (HTAP [9], lakehouse [8], AIGC [27]) y la madurez analítica de las organizaciones: muchas empresas aún operan con mentalidades de BI tradicional (reporting estático, dependencia de DW monolíticos), lo que dificulta la adopción de paradigmas más ágiles [5], [11], [12]. Esta brecha se observa incluso en sectores avanzados, como la construcción, donde la conectividad, la interoperabilidad y la formación siguen siendo barreras clave [26].

Desde el ámbito metodológico, predomina la investigación conceptual o cualitativa, con escasos estudios empíricos que vinculen métricas técnicas (latencia, throughput, frescura de datos) con indicadores de negocio (tiempo de toma de decisiones, ROI, reducción de riesgos) [16], [23], [28]. Aunque existen excepciones rigurosas (como el análisis de eficiencia de proveedores cloud mediante DEA [31] o la evaluación de costos en FinOps [2]) la mayoría de los trabajos carecen de benchmarks comparativos, replicabilidad o validación en entornos reales [5], [11], [13], [20]. Esta falta de evidencia cuantitativa dificulta la generalización de hallazgos [25], [26].

La revisión sistemática permite identificar líneas futuras que apuntan a cerrar las brechas actuales entre la promesa de las arquitecturas cloud-native y su implementación efectiva en contextos de BI. Estas líneas se agrupan en cuatro ejes: evaluación empírica basada en valor, gobernanza y ética

integradas, autonomía inteligente y estandarización interoperable.

Se observa una demanda creciente por marcos de evaluación empírica que vinculen métricas técnicas con valor organizacional. Múltiples estudios reconocen que, si bien se han propuesto arquitecturas avanzadas (lakehouse [8], HTAP [9], ingesta metadata-driven [15]), persiste una carencia crítica de evidencia cuantitativa que relacione indicadores como latencia, frescura de datos o costo por consulta con resultados de negocio medibles (ROI, reducción del tiempo de toma de decisiones, mejora en la calidad del servicio) [16], [21], [23], [25]. Futuras investigaciones deberían adoptar diseños mixtos que combinen benchmarks técnicos con mediciones de impacto en KPIs organizacionales, especialmente en sectores regulados como la salud [3], [10] o las finanzas [4].

Se destaca la necesidad de profundizar en la integración de gobernanza, seguridad y privacidad desde el diseño (privacy-by-design y governance-by-design). Aunque varios trabajos reconocen la importancia de la seguridad [19], [30] y la calidad de los datos [4], [20], pocos proponen soluciones integradas que no sacrifiquen rendimiento ni agilidad. Las líneas futuras apuntan hacia la automatización de políticas de gobernanza mediante metadatos [15], la aplicación de técnicas de pseudonimización y cifrado homomórfico en entornos analíticos [10], y el desarrollo de marcos de data ethics alineados con regulaciones como GDPR o HIPAA [27], [29]. Esto es especialmente relevante en contextos de algorithmic management, donde los perfiles generados por IA pueden perpetuar sesgos o erosionar la autonomía laboral [28].

Una tercera dirección emergente es la evolución hacia arquitecturas autónomas e inteligentes, impulsadas por modelos fundacionales y AIGC. Varios papers anticipan que los pipelines de datos dejarán de ser estáticos para convertirse en sistemas auto-optimizables que ajusten estrategias de ingesta, particionamiento o indexación en función del perfil de uso [15], [18], [27]. Asimismo, se prevé una mayor integración de LLMs no solo para la generación de dashboards o la explicación de insights [27], sino también para la autorregulación de la calidad de los datos, la detección de data drift [4] o la generación automática de políticas de acceso [29]. Esta evolución se extiende al dominio del IoT inteligente (IIoT), donde la integración de edge y fog computing con la nube permite equilibrar latencia, escalabilidad y análisis complejo, especialmente en sectores como la salud o la manufactura [22].

Finalmente, identificamos una brecha significativa en la generalización y estandarización de arquitecturas. Muchos estudios se limitan a casos específicos (construcción [26], salud [3], [10], finanzas [4]) o a un único proveedor cloud [8], [14], lo que dificulta la extracción de principios universales. Las líneas futuras apuntan hacia la definición de patrones cloud-agnostic [15], la creación de benchmarks comparativos multi-proveedor (como HTAPBench [9] o TPC-H adaptado a lakehouse [14]), y la exploración de modelos de data sharing interoperables (Delta Share, Iceberg REST Catalog) que reduzcan el vendor lock-in y fomenten ecosistemas de datos

abiertos [8], [18]. Además, se requiere investigación sobre la sostenibilidad ambiental de estas arquitecturas, un tema apenas abordado en la literatura actual [10]. La regulación emergente (como el EU Data Act 2027) también impulsará la necesidad de arquitecturas que faciliten la portabilidad y el data sovereignty [18].

IV. DISCUSIÓN

Los resultados obtenidos en esta revisión sistemática, sintetizados a partir de los 30 estudios empíricos seleccionados, permiten una interpretación profunda que va más allá de la mera descripción de hallazgos, articulando sus implicancias teóricas, prácticas y estratégicas. Esta sección interpreta los resultados en relación con cada pregunta de investigación (PI), destacando patrones emergentes, contradicciones en la literatura y oportunidades para una adopción más alineada de arquitecturas cloud-native en BI. Enfatizamos la integración de métricas técnicas con dinámicas organizacionales, proponiendo un puente conceptual que enriquece el marco evaluativo introducido en la introducción, sin repetir los datos cuantitativos ya expuestos.

A. Herramientas y metodologías empleadas en la implementación de arquitecturas cloud-native para BI (PI1)

La diversidad de herramientas y metodologías identificadas revela una evolución hacia ecosistemas híbridos que priorizan la integración y la automatización, lo que interpretamos como una respuesta a la complejidad creciente de los datos en entornos cloud. Por ejemplo, el uso predominante de plataformas como AWS, Azure y Google Cloud [31], [5], [13] se interpreta no solo como una elección técnica, sino como un facilitador de metodologías ágiles como DevOps y CI/CD, que agilizan la implementación de data warehouses y lakes [20], [14].

En contraste, enfoques como el data mesh [30], [24] sugieren una metodología descentralizada que mitiga silos de datos, lo que interpretamos como un avance hacia la gobernanza distribuida. Sin embargo, la recurrencia de modelos como Kubernetes para orquestación [12], [22] resalta una brecha importante: mientras estas herramientas optimizan la implementación técnica, su adopción a menudo subestima la curva de aprendizaje organizacional, lo que podría explicar inconsistencias en el rendimiento reportado en estudios con contextos corporativos variados [2], [25].

Esta interpretación subraya que las metodologías no son universales, sino contextuales, recomendando evaluaciones previas de madurez digital para maximizar su efectividad.

B. Latencia, escalabilidad y costo-rendimiento (PI2)

Los resultados indican que la latencia, escalabilidad y costo-rendimiento no actúan de forma aislada, sino interdependiente, lo que interpretamos como factores que determinan la robustez global de las arquitecturas BI en la nube. La reducción de latencia mediante caching y edge computing [19], [23], [18] se interpreta como un impulsor clave de

desempeño en workloads analíticos, pero su impacto se ve moderado por trade-offs en escalabilidad, como en casos donde el auto-scaling genera costos imprevistos [16], [9].

Por ejemplo, el data lakehouse emerge como paradigma que equilibra estas métricas [11], [21], [29], interpretándose su superioridad en entornos de alto volumen de datos como evidencia de que la separación de storage y compute minimiza latencia sin inflar costos. No obstante, anomalías como picos de costo en implementaciones híbridas [4], [17] sugieren que el desempeño técnico depende de optimizaciones contextuales, como la integración de ML para predicción de cargas [6], [27].

Esta interconexión implica que las métricas deben evaluarse holísticamente, evitando enfoques aislados que podrían subestimar riesgos en escenarios reales de BI.

C. Relación entre Desempeño Técnico y Valor Organizacional del BI (PI3)

La brecha persistente entre optimización técnica y valor organizacional, evidente en los resultados, se interpreta como un desafío estructural que exige métricas híbridas para capturar impactos intangibles. El alto rendimiento en latencia y escalabilidad se traduce en agilidad operativa, como en estudios donde data lakes reducen el time-to-insight [1], [26], [28], lo que interpretamos como un catalizador para decisiones informadas y ROI positivo.

Sin embargo, casos donde el costo-rendimiento no correlaciona con ganancias financieras [31], [8], [10] revelan que el valor organizacional depende de factores como la alineación con KPIs de negocio, no solo de eficiencia técnica. Por ejemplo, el lakehouse facilita la colaboración interdepartamental [30], [15], interpretándose su adopción como un puente hacia culturas data-driven que potencian innovación. Esta relación no es lineal: en contextos con gobernanza débil, mejoras técnicas pueden generar "data swamps" que erosionan valor [12], [3].

Por ende, nuestra interpretación apunta a un modelo integrado donde el desempeño técnico se evalúe por su capacidad para generar valor estratégico, como el propuesto en nuestra revisión, fomentando adopciones que prioricen outcomes organizacionales sobre métricas puramente técnicas.

D. Tendencias, limitaciones y líneas futuras en arquitecturas (PI4)

Las tendencias emergentes, como la dominancia del lakehouse y la integración de IA [11], [13], [22], las interpretamos como un cambio hacia paradigmas unificados que abordan limitaciones históricas de data warehouses, promoviendo sostenibilidad y resiliencia. Sin embargo, limitaciones como la dependencia de proveedores cloud [31], [20], [24] y desafíos de seguridad [2], [25] destacan vulnerabilidades que podrían interpretarse como barreras a la escalabilidad global.

Las líneas futuras, como la exploración de edge BI y governance-by-design [5], [18], [27], sugieren oportunidades para investigación que integre sostenibilidad ambiental con

desempeño [23], [17]. Esta interpretación revela que, aunque las tendencias apuntan a madurez, las limitaciones persisten en contextos menos desarrollados, recomendando estudios longitudinales que validen impactos a largo plazo.

En suma, estos insights abren vías para investigaciones que cierren la brecha técnica-organizacional, contribuyendo a una adopción más intencional de cloud-native en BI.

VI. CONCLUSIONES

Esta revisión sistemática de 30 estudios empíricos mediante PRISMA 2020 confirma que las arquitecturas cloud-native han optimizado sustancialmente la latencia y la escalabilidad en Business Intelligence, consolidando a Data Lakehouse como paradigma dominante. Los avances técnicos son notables: reducciones de latencia de hasta 47×, escalabilidad horizontal mediante separación de cómputo y almacenamiento, y optimización en costo-rendimiento a través de herramientas como AWS, Azure Synapse, Apache Spark y Kafka.

Sin embargo, persiste una brecha crítica entre el desempeño técnico (validado predominantemente en entornos controlados) y su traducción en valor organizacional medible. Esta brecha se agrava por la escasez de talento con habilidades híbridas, la falta de gobernanza integrada desde el diseño y la ausencia de métricas que vinculen optimizaciones técnicas con KPIs estratégicos como ROI, tiempo de decisión o reducción de riesgos operativos.

Como aporte original, proponemos un marco de evaluación integrado con KPIs híbridos como "tiempo de insight a ROI" e "índice agilidad-rendimiento", que permite evaluar inversiones cloud por su contribución al valor estratégico y sostenibilidad financiera.

Las líneas futuras convergen en cuatro direcciones clave:

- 1) *Estudios longitudinales que vinculen métricas técnicas con KPIs de negocio en contextos productivos.*
- 2) *Implementaciones de privacy-by-design y governance-by-design sin sacrificar rendimiento.*
- 3) *Exploración de modelos fundacionales y AIGC para automatización de pipelines.*
- 4) *Creación de benchmarks multi-proveedor con métricas de sostenibilidad ambiental.*

De este modo, la ingeniería de datos podrá alinearse efectivamente con la generación de valor estratégico, permitiendo que las organizaciones inviertan en transformación digital con propósito de valor, gobernanza desde el diseño y resiliencia en entornos empresariales dinámicos.

REFERENCES

[1] Y. Qin y G. Guzun, «Faster Multidimensional Data Queries on Infrastructure Monitoring Systems», *Big Data Research*, vol. 27, p. 100288, nov. 2021. [Online]. Available: <https://doi.org/10.1016/j.bdr.2021.100288>.

[2] G. Fragiadakis, A. Tsadimas, E. Filiopoulou, G. Kousiouris, C. Michalakelis, y M. Nikolaidou, «CloudPricingOps: A Decision Support Framework to Explore Pricing Policies of Cloud Services», *Applied Sciences*, vol. 14, n.o 24, p. 11946, dic. 2024. [Online]. Available: <https://doi.org/10.3390/app142411946>.

[3] S. S. M. Silva et al., «Development of a Predictive Dashboard With Prescriptive Decision Support for Falls Prevention in Residential Aged

Care: User-Centered Design Approach», *JMIR Aging*, vol. 8, p. e63609, abr. 2025. [Online]. Available: <https://doi.org/10.2196/63609>.

[4] A. R. Munappy, J. Bosch, H. H. Olsson, A. Arpteg, y B. Brinne, «Data management for production quality deep learning models: Challenges and solutions», *Journal Of Systems And Software*, vol. 191, p. 111359, may 2022. [Online]. Available: <https://doi.org/10.1016/j.jss.2022.111359>.

[5] S. Dehbi, H. C. Lamrani, T. Belgnaoui, y T. Lafou, «Big Data Analytics and Management control», *Procedia Computer Science*, vol. 203, pp. 438-443, ene. 2022. [Online]. Available: <https://doi.org/10.1016/j.procs.2022.07.058>.

[6] Y. Liu y C. Li, «Insights into Talent Cultivation in Big Data Management and Application Major Based on Recruitment Information», *Procedia Computer Science*, vol. 242, pp. 568-575, ene. 2024. [Online]. Available: <https://doi.org/10.1016/j.procs.2024.08.107>.

[7] M. J. Page et al., «Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas», *Revista Española de Cardiología*, vol. 74, n.o 9, pp. 790-799. [Online]. Available: <https://doi.org/10.1016/j.recesp.2021.06.016>.

[8] J. Schneider, C. Gröger, A. Lutsch, H. Schwarz, y B. Mitschang, «The Lakehouse: State of the Art on Concepts and Technologies», *SN Computer Science*, vol. 5, n.o 5, abr. 2024. [Online]. Available: <https://doi.org/10.1007/s42979-024-02737-0>.

[9] H. Song, W. Zhou, H. Cui, X. Peng, y F. Li, «A survey on hybrid transactional and analytical processing», *The VLDB Journal*, vol. 33, n.o 5, pp. 1485-1515, jun. 2024. [Online]. Available: <https://doi.org/10.1007/s00778-024-00858-9>.

[10] S. H. Noteboom et al., «From intensive care monitors to cloud environments: a structured data pipeline for advanced clinical decision support», *EBioMedicine*, vol. 111, p. 105529, dic. 2024. [Online]. Available: <https://doi.org/10.1016/j.ebiom.2024.105529>.

[11] J. Duque, A. Godinho, y J. Vasconcelos, «Knowledge data extraction for business intelligence A design science research approach», *Procedia Computer Science*, vol. 204, pp. 131-139, ene. 2022. [Online]. Available: <https://doi.org/10.1016/j.procs.2022.08.016>.

[12] D. Żółtowski, «Business Intelligence in Balanced Scorecard: Bibliometric analysis», *Procedia Computer Science*, vol. 207, pp. 4075-4086, ene. 2022. [Online]. Available: <https://doi.org/10.1016/j.procs.2022.09.470>.

[13] T. Jin y B. Zhang, «Intermediate data fault-tolerant method of cloud computing accounting service platform supporting cost-benefit analysis», *Journal Of Cloud Computing Advances Systems And Applications*, vol. 12, n.o 1, ene. 2023. [Online]. Available: <https://doi.org/10.1186/s13677-022-00385-4>.

[14] A. Cuzzocrea, «Evolving OLAP and BI towards Complex, High-Performance Big-OLAP-Data-Cube-Processing Analytics Frameworks: How to Speed-Up Large-Scale, High-Dimensional Queries over Clouds», *Procedia Computer Science*, vol. 246, pp. 4169-4175, ene. 2024. [Online]. Available: <https://doi.org/10.1016/j.procs.2024.09.256>.

[15] C. Rucco, A. Longo, y M. Saad, «Enhancing Data Ingestion Efficiency in Cloud-Based Systems: A Design Pattern Approach», *Data Science And Engineering*, jul. 2025. [Online]. Available: <https://doi.org/10.1007/s41019-025-00300-2>.

[16] H. Herrmann, «The arcanum of artificial intelligence in enterprise applications: Toward a unified framework», *Journal Of Engineering And Technology Management*, vol. 66, p. 101716, sep. 2022. [Online]. Available: <https://doi.org/10.1016/j.jengtecman.2022.101716>.

[17] S. Depner y A. Richter, «Cloud meets customer: IT service providers in the public cloud ecosystem», *Business Horizons*, abr. 2025. [Online]. Available: <https://doi.org/10.1016/j.bushor.2025.04.008>.

[18] M. Plazotta y M. Klettke, «Data Architectures in Cloud Environments», *Datenbank-Spektrum*, vol. 24, n.o 3, pp. 243-247, oct. 2024. [Online]. Available: <https://doi.org/10.1007/s13222-024-00490-5>.

[19] A. Razaque, N. Shaldanbayeva, B. Alotaibi, M. Alotaibi, A. Murat, y A. Alotaibi, «Big Data Handling Approach for Unauthorized Cloud Computing Access», *Electronics*, vol. 11, n.o 1, p. 137, ene. 2022. [Online]. Available: <https://doi.org/10.3390/electronics11010137>.

[20] M. Cherradi y A. E. Haddadi, «A Scalable framework for data lakes ingestion», *Procedia Computer Science*, vol. 215, pp. 809-814, ene. 2022. [Online]. Available: <https://doi.org/10.1016/j.procs.2022.12.083>.

- [21] R. Ros, E. Bjarnason, y P. Runeson, «A theory of factors affecting continuous experimentation (FACE)», *Empirical Software Engineering*, vol. 29, n.o 1, dic. 2023. [Online]. Available: <https://doi.org/10.1007/s10664-023-10358-z>.
- [22] I. Ullah, D. Adhikari, X. Su, F. Palmieri, C. Wu, y C. Choi, «Integration of data science with the intelligent IoT (IIoT): current challenges and future perspectives», *Digital Communications And Networks*, mar. 2024. [Online]. Available: <https://doi.org/10.1016/j.dcan.2024.02.007>.
- [23] J. Passlick, L. Grützner, M. Schulz, y M. H. Breitner, «Self-service business intelligence and analytics application scenarios: A taxonomy for differentiation», *Information Systems And e-Business Management*, vol. 21, n.o 1, pp. 159-191, feb. 2023. [Online]. Available: <https://doi.org/10.1007/s10257-022-00574-3>.
- [24] S. P. Niranjana, S. D. Latha, S. Vase, y M. L. Scutaru, «Analysis of Bulk Queueing Model with Load Balancing and Vacation», *Axioms*, vol. 14, n.o 1, p. 18, dic. 2024. [Online]. Available: <https://doi.org/10.3390/axioms14010018>.
- [25] Y. Cao y F. Q. A. Alyousuf, «A new framework to assess the impact of new IT-based technologies on the success of quality management system», *Journal Of Big Data*, vol. 12, n.o 1, ene. 2025. [Online]. Available: <https://doi.org/10.1186/s40537-025-01061-5>.
- [26] M. A. Musarat, W. S. Alaloul, M. H. F. Khan, S. Ayub, y C. P. L. Guy, «Evaluating cloud computing in construction projects to avoid project delay», *Journal Of Open Innovation Technology Market And Complexity*, vol. 10, n.o 2, p. 100296, may 2024. [Online]. Available: <https://doi.org/10.1016/j.joitmc.2024.100296>.
- [27] H. Kuai et al., «Web Intelligence (WI) 3.0: in search of a better-connected world to create a future intelligent society», *Artificial Intelligence Review*, vol. 58, n.o 9, jun. 2025. [Online]. Available: <https://doi.org/10.1007/s10462-025-11203-z>.
- [28] J. R. Martorell, F. Tirado, J. L. Blasco, y A. Gálvez, «How does artificial intelligence work in organisations? Algorithmic management, talent and dividualisation processes», *AI & Society*, may 2024. [Online]. Available: <https://doi.org/10.1007/s00146-024-01970-8>.
- [29] K. K. Alnofeli, S. Akter, y V. Yanamandram, «Unlocking the power of AI in CRM: A comprehensive multidimensional exploration», *Journal Of Innovation & Knowledge*, vol. 10, n.o 3, p. 100731, may 2025. [Online]. Available: <https://doi.org/10.1016/j.jik.2025.100731>.
- [30] M. K. S. Alwaheidi y S. Islam, «Data-Driven Threat Analysis for Ensuring Security in Cloud Enabled Systems», *Sensors*, vol. 22, n.o 15, p. 5726, jul. 2022. [Online]. Available: <https://doi.org/10.3390/s22155726>.
- [31] C.-N. Wang, M.-N. Nguyen, T.-D. Nguyen, H.-P. Hsu, y T.-H.-Y. Nguyen, «Effective Decision Making: Data Envelopment Analysis for Efficiency Evaluation in the Cloud Computing Marketplaces», *Axioms*, vol. 10, n.o 4, p. 309, nov. 2021. [Online]. Available: <https://doi.org/10.3390/axioms10040309>.