



TENDENCIAS EN TECNOLOGÍA DE DATOS 2026



03

ARQUITECTURAS
MODERNAS DE DATOS

06

MULTI-CLOUD Y CLOUD HÍBRIDO

10

IA GENERATIVA APLICADA A
DATOS EMPRESARIALES

15

MLOPS Y LLMOPS

18

DATOS EN TIEMPO REAL Y
ARQUITECTURAS EVENT-DRIVEN



ARQUITECTURAS MODERNAS DE DATOS:

LAKEHOUSE, FABRIC Y MESH



2026 es el año en que Lakehouse, Data Fabric y Data Mesh se consolidan como el estándar de referencia para empresas que quieren una plataforma de datos realmente lista para IA.

En 2026, la discusión sobre arquitecturas de datos ya no va de conceptos básicos, sino de qué combinaciones permiten escalar IA y analítica sin hundir a la organización en complejidad y costo. Lakehouse, Data Fabric y Data Mesh dejan de ser “experimentos” y se convierten en el lenguaje común de los roadmaps de datos.

Tres señales claras:

Lakehouse como “plataforma base” de datos e IA

El gasto global en big data y analytics se proyecta en alrededor de 420 mil millones de dólares para 2026, lo que refleja la necesidad de arquitecturas que soporten más volumen, más casos de uso y más modelos sin colapsar. En ese contexto, análisis recientes posicionan el Data Lakehouse como la evolución natural del stack “DW + data lake”, al ofrecer un backbone único para BI, ML y GenAI sobre formatos de tabla abiertos.

Data Fabric como capa para un mundo multi-cloud e híbrido

El mercado global de Data Fabric se valoró en unos 2,545.5 millones de dólares en 2024, con una tasa de crecimiento anual compuesta estimada del 18.37% hasta 2032. Esta cifra refleja que la apuesta del mercado no es “moverlo todo a un único lugar”, sino conectar y gobernar datos distribuidos en varias nubes, on-prem y SaaS.

Data Mesh como respuesta organizacional al cuello de botella central

Un estudio dedicado a Data Mesh indica que alrededor del 84% de las organizaciones encuestadas ya lo han implementado total o parcialmente, con objetivos primarios de mejorar la calidad de datos (64%) y reforzar la gobernanza (58%). Esto confirma que la tendencia no es solo tecnológica, sino de modelo operativo: dominios responsables de sus data products, sobre una plataforma común.



Cómo encajan Lakehouse, Fabric y Mesh en un roadmap realista

Un patrón coherente para 2024–2027, alineado con lo que se observa en el mercado, podría ser:

CONSOLIDAR UN BACKBONE LAKEHOUSE AI-READY

- Reducir silos y almacenes redundantes.
- Basarse en formatos abiertos y soportar tanto batch como streaming.
- Alinear esta capa con los principales casos de BI y ML del negocio.

INTRODUCIR UNA CAPA DE DATA FABRIC DONDE MIGRAR NO TIENE SENTIDO

- Conectar sistemas core legados, SaaS críticos y datos con restricciones de residencia.
- Usar metadata activa para automatizar gobierno e integración siempre que sea posible.

EVOLUCIONAR HACIA DATA MESH EN DOMINIOS CLAVE

- Arrancar con 1–2 dominios de alto impacto, con data products, SLAs y owners claros.
- Escalar el modelo a más dominios según madurez y resultados





MULTI-CLOUD Y CLOUD HÍBRIDO: FLEXIBILIDAD SIN LOCK-IN



2026 es el año en que multi-cloud e híbrido dejan de ser "casos especiales" y se convierten en el modelo operativo estándar para plataformas de datos empresariales.

Si sigues asumiendo que *"los datos vivirán en una sola nube"* es casi tan realista como pensar que toda la organización correrá en un único ERP. La realidad es **multi-cloud e híbrida por diseño**, no por accidente. Este capítulo explora cómo las plataformas de datos modernas se adaptan a esta realidad, qué está impulsando la adopción de estrategias distribuidas y dónde están los verdaderos puntos de fricción.

Tres señales de mercado:

Multi-cloud como norma, no como excepción

Análisis recientes indican que entre 85% y 89% de las organizaciones grandes ya operan con estrategias multi-cloud hacia 2026, impulsadas por necesidades de resiliencia, poder de negociación con proveedores y cumplimiento regulatorio.

Híbrido como configuración de equilibrio entre control y flexibilidad

Un reporte de 2025 proyecta que para 2027, 90% de las organizaciones habrán adoptado una estrategia de nube híbrida. Este modelo combina infraestructura privada (para datos sensibles, core banking, sistemas regulados) con nube pública (para elasticidad, innovación y nuevos productos digitales).

FinOps como disciplina obligatoria

Un análisis sobre FinOps 2026 muestra que 72% de las empresas globales excedieron sus presupuestos cloud en el último año fiscal, lo que está acelerando la adopción de prácticas de optimización y control financiero cloud. Esto se agrava con las cargas intensivas en GPU de IA generativa, donde sin gobernanza de costos se disparan los gastos rápidamente.

Multi - cloud

Beneficios

- Resiliencia: workloads críticos en 2+ nubes.
- Optimización: cada workload en el proveedor con mejor costo/rendimiento.
- Menos lock-in: open table formats + motores open-source + APIs estándar.

Retos

- Complejidad operativa (identidades, redes, seguridad, monitoreo, costos).
- Networking y egress caros + latencias.
- Gobernanza fragmentada sin Fabric/catalogo unificado.

La clave

- Funciona solo con buena gobernanza, observabilidad y control de costos.

Híbrido

Idea

- Combina on-prem/cloud privada + nube pública.
- Target state en sectores regulados (banca, salud, sector público).

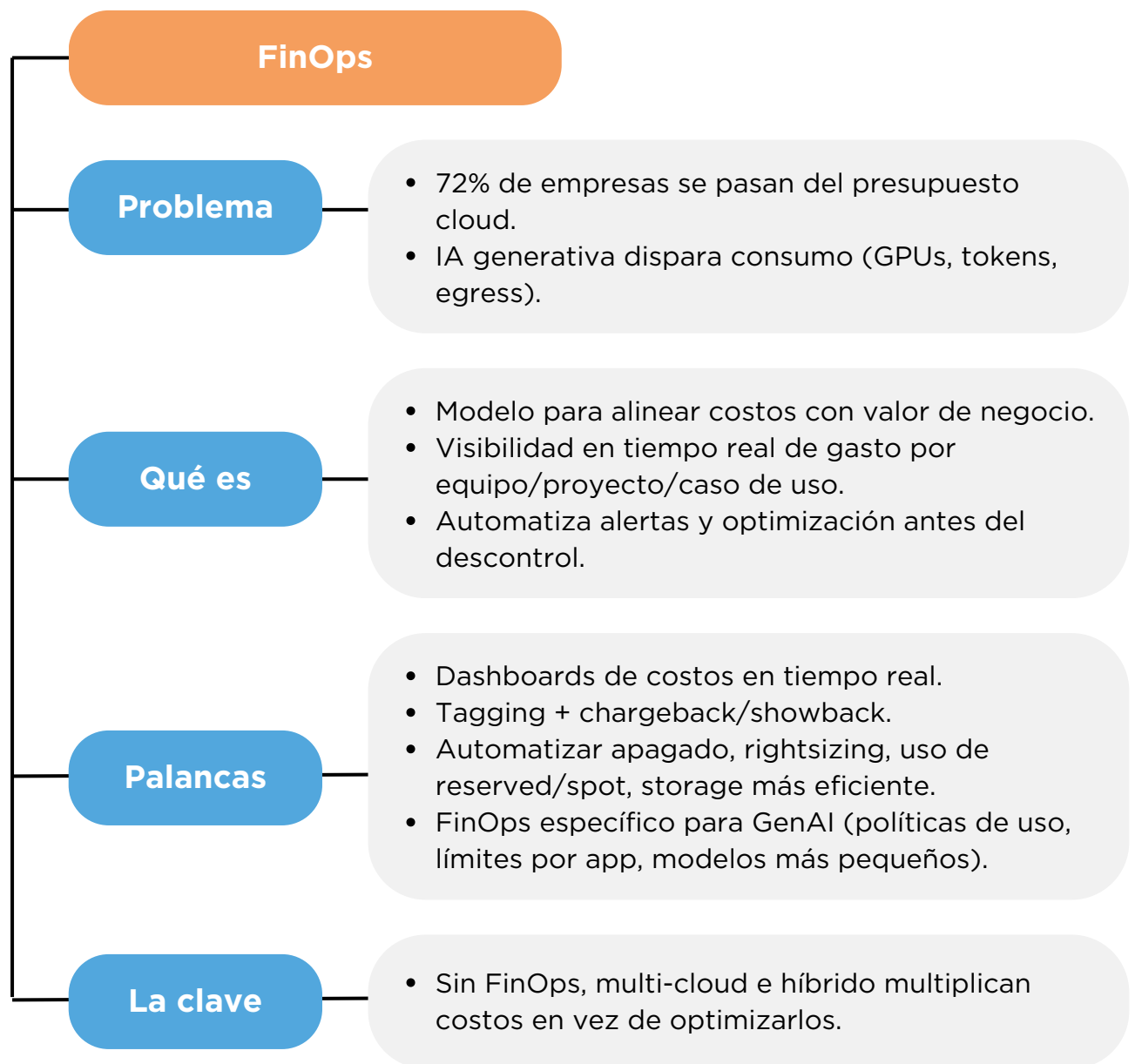
Patrón típico

- Core y datos regulados → on-prem / privado.
- Digital, analítica, ML y GenAI → nube pública.
- Integración → redes privadas dedicadas + APIs gobernadas.

La clave

- No es “mitad y mitad”, es una arquitectura deliberada: regula + innova.







IA GENERATIVA APLICADA A DATOS EMPRESARIALES



2026 es el año en que GenAI pasa de ser un campo experimental a una capa operativa esperada en plataformas de datos e IA, integrada en workflows de BI, analítica y decisiones de negocio.

Tres movimientos clave:

De pilotos a "enterprise operating layer"

Deloitte y otros analistas describen que las organizaciones que ya dominaban cloud y datos están moviendo GenAI desde pilotos aislados hacia una "enterprise operating layer" que se integra en procesos y productos core. Esto significa que GenAI pasa de "tenemos un chatbot de IA" a "nuestros usuarios de negocio interactúan con datos en lenguaje natural y modelos ejecutan acciones sin intervención humana".

RAG y vector databases como arquitectura dominante

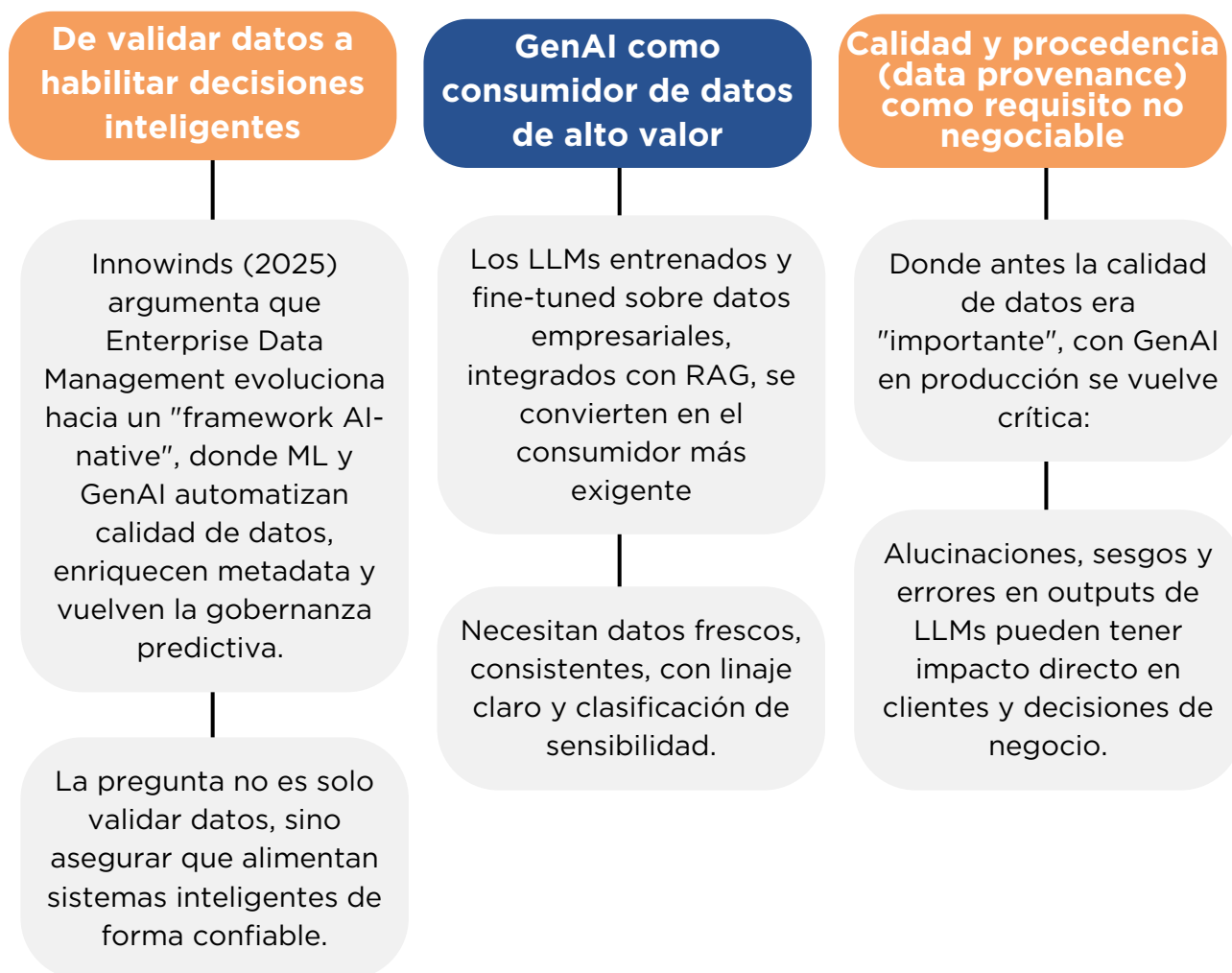
Un análisis de BrollyAI indica que para 2026, RAG será la arquitectura dominante en sistemas de IA empresarial, con tiempos de respuesta típicos entre 10 y 100 ms y capacidad de soportar millones de documentos. Según proyecciones de Gartner, más del 30% de las empresas adoptarán vector databases para enriquecer modelos fundacionales con datos empresariales y contexto propio hacia 2026.

Agentic AI en casos de uso productivos

Los agentes de IA (sistemas que no solo responden, sino que planifican y actúan sobre sistemas empresariales) pasan de prototipos a casos con impacto real. Bernard Marr identifica agentes autónomos como una de las tendencias dominantes de 2026, junto a generación sintética de datos y modelos de dominio específico para industrias reguladas.

EL CAMBIO DE CONVERSACIÓN EN ENTERPRISE DATA MANAGEMENT

Lo que caracteriza 2026 en el contexto de GenAI y datos es un cambio conceptual profundo:



COPILOTS DE DATOS EN EL CONTEXTO 2026

Lo que caracteriza 2026 en el contexto de GenAI y datos es un cambio conceptual profundo:

Consultar datos en lenguaje natural

"¿cuál fue la tasa de churn el mes pasado por región?"

Generar dashboards y reportes dinámicos sin escribir SQL.

Explicar métricas y trends en contexto

¿por qué bajaron las ventas?"

Sugerir acciones

"basado en estos datos, deberías revisar el inventario en la región Norte"

RAG: EL PATRÓN QUE HACE GENAI CONFIABLE SOBRE DATOS EMPRESARIALES

RAG (Retrieval-Augmented Generation) es la arquitectura técnica que más tracción está ganando para integrar LLMs con datos empresariales.

RAG conecta LLMs con datos empresariales: antes de responder, recupera documentos relevantes y así baja alucinaciones.

Ventajas vs fine-tuning: datos frescos, trazabilidad de la fuente y mucho menor costo.

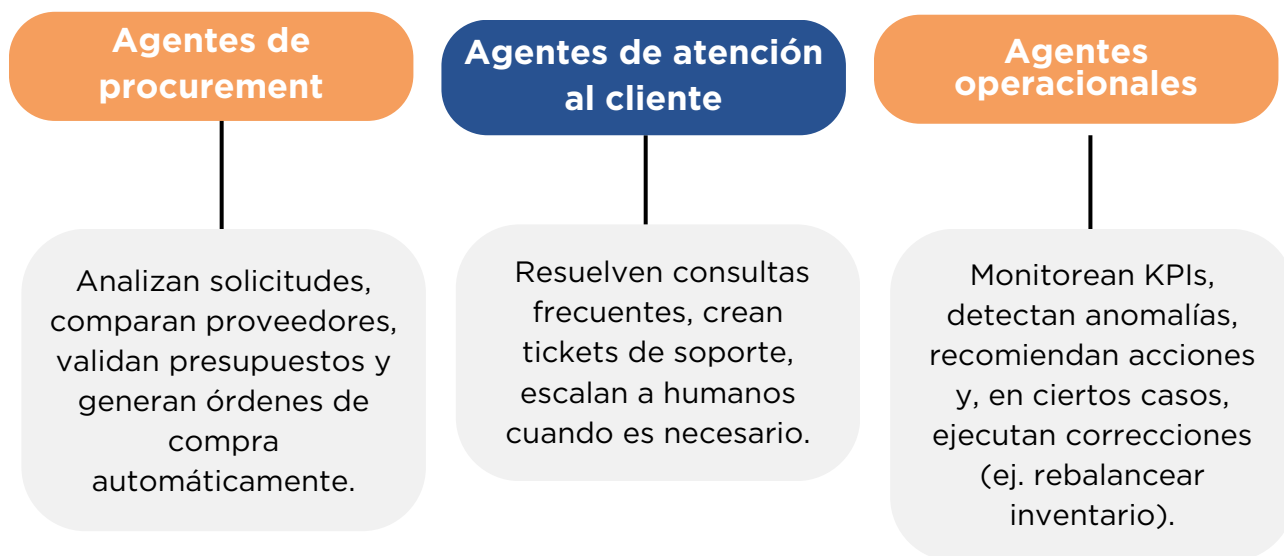
Vector databases guardan embeddings y son el "cerebro" de RAG para buscar contexto similar.

Madurez: las empresas pasan de RAG simple (top-k) → RAG híbrido (vectores + filtros) → RAG avanzado (re-ranking, multi-hop, graph-based).



AGENTIC AI: AUTOMATIZACIÓN QUE ACTÚA SOBRE SISTEMAS EMPRESARIALES

Mientras RAG responde preguntas sobre datos, Agentic AI va más allá: sistemas que planifican, llaman herramientas y ejecutan acciones sobre sistemas de negocio.





MLOPS Y LLMOPS: INDUSTRIALIZANDO LA IA



2026 es el año en que MLOps deja de ser "buena práctica opcional" y se convierte en requisito para capturar ROI medible de inversiones en datos e IA.

Tres movimientos clave:

Automatización completa del pipeline como norma

Análisis de tendencias de IA para 2026 enfatizan que las empresas que operacionalizan IA con pipelines de datos, gobernanza y MLOps adecuados obtienen ROI medible en cuestión de trimestres, mientras que las que se quedan en pilotos no logran retornos sostenibles. VisualPath subraya que en 2026 la automatización completa de pipelines (data → training → testing → deployment → monitoring) y el uso de herramientas MLOps con IA integrada serán estándar, no excepción.

Observabilidad autónoma y AIOps

Predicciones sobre observabilidad para 2026 señalan un cambio hacia "autonomous observability", donde al menos 40% de grandes organizaciones cloud-native tendrán sistemas que detectan y remedian incidentes de bajo riesgo automáticamente. La métrica evoluciona de "Mean Time to Resolution (MTTR)" a "Mean Time to Autonomy (MTTA)"— qué tan rápido los sistemas se auto-reparan sin intervención humana.

LLMOps como disciplina diferenciada de MLOps

Mientras MLOps es bien conocido, LLMOps (prácticas específicas para LLMs y GenAI) está consolidándose como disciplina separada, con herramientas, métricas y workflows propios. Esto refleja que los LLMs tienen ciclos de vida distintos (actualizaciones más frecuentes de APIs, mayor variabilidad en outputs, gestión de tokens) que requieren enfoques diferentes.

En síntesis:

- Automatización: pipelines end-to-end sin intervención manual.
- Observabilidad: visibilidad continua de datos, modelos y LLMs en producción.
- LLMOps específico: prácticas tailored para ciclos de vida de LLMs, no solo aplicar MLOps + logs.

EVOLUCIÓN HACIA "AUTONOMOUS OBSERVABILITY"

La tendencia para 2026 es que sistemas de observabilidad no solo detecten problemas, sino que también los remedien:

AUTO-HEALING

Si se detecta un pipeline roto, el sistema intenta repararlo automáticamente (ej. relanzar ingesta, limpiar datos malformados).

01

AUTO-SCALING

Si la latencia de un modelo degrada por carga, se escala automáticamente.

02

AUTO-ALERTING AJUSTABLE

El sistema aprende qué alertas son ruido y cuáles importan, ajustando umbrales dinámicamente.

03

AUTOMATIZACIÓN Y MLOPS CON IA INTEGRADA

Una meta-tendencia interesante: MLOps tools mismas están usando IA:

Herramientas MLOps predictivas

Plataformas que usan ML para predecir cuándo un modelo va a degradar, sugerir retraining, recomendar ajustes de hiperparámetros.

Pipelines self-healing

Cuando se detecta drift, el sistema automáticamente lanza reentrenamiento, valida el nuevo modelo y lo despliega si pasa thresholds.

Feature store con recomendaciones

Catálogos de features que sugieren qué features usar para nuevos modelos basado en problemas similares pasados.

Esta meta-aplicación de IA a MLOps es una tendencia 2026 que amplifica el impacto de automatización.





DATOS EN TIEMPO REAL Y ARQUITECTURAS EVENT- DRIVEN



En 2026, la pregunta sobre datos en tiempo real ya no es "¿lo necesitamos?" sino "¿dónde aún estamos haciendo batch que debería ser streaming?".

Tres movimientos clave:

Real-time como nuevo default

Informes sobre el landscape de datos 2026 describen explícitamente que los datos en tiempo real y streaming pasan de nicho a expectativa mainstream, ya que usuarios y sistemas esperan respuestas inmediatas. El panorama de datos de 2026 subraya que la analítica se está moviendo hacia la capa de streaming, donde ocurre procesamiento continuo en lugar de batch diario o semanal.

Data Streaming Platform (DSP) como categoría emergente

El concepto de Data Streaming Platform se consolida como nueva categoría de infraestructura que unifica conectividad, procesamiento, almacenamiento y governance de datos en movimiento, basada en tecnologías como Apache Kafka y Apache Flink. Esto marca un cambio: el streaming no es solo "otro ETL", sino un backbone estratégico de la plataforma de datos moderna.

Streaming + Lakehouse como patrón unificado

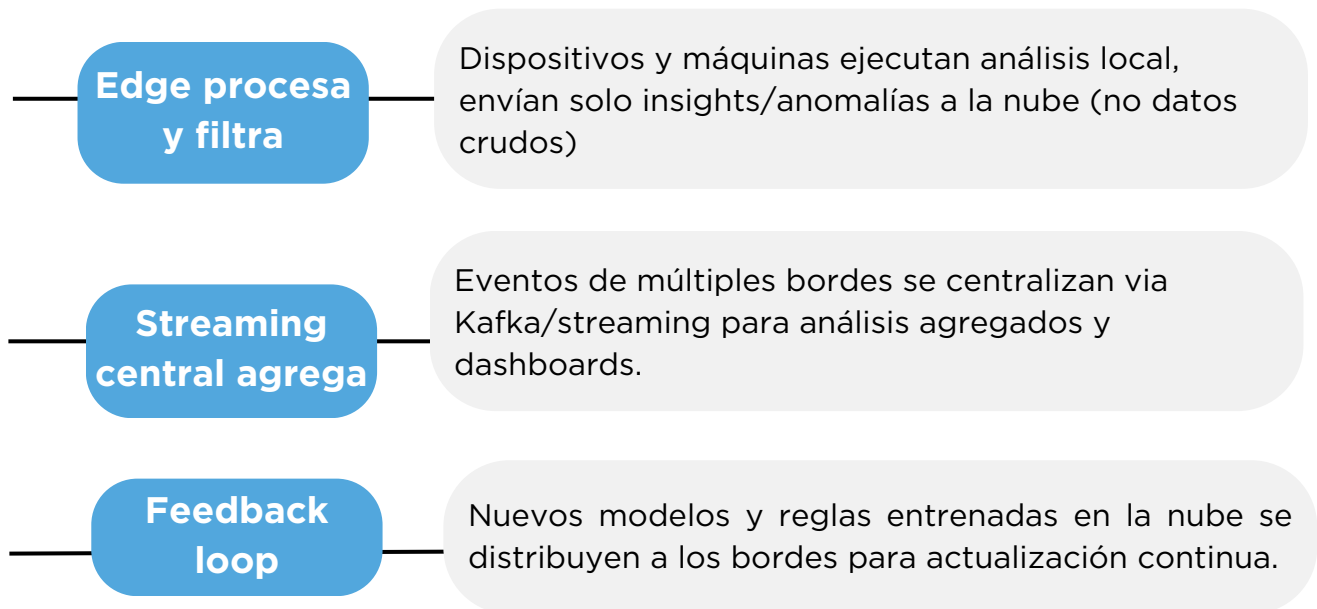
Tendencias 2026 describen la convergencia de streaming con formatos de lakehouse (Iceberg, Delta): permite que un mismo backbone analítico sirva tanto cargas de tiempo real como análisis histórico sin duplicación. Esto es crítico porque elimina la necesidad de mantener "un sistema para streaming" y "otro para batch".

Edge AI como complemento indispensable

En paralelo, edge computing para IA y procesamiento de datos crece acelerado, impulsado por IoT, manufactura, retail y automotriz. Empresas implementan Edge AI para reducir latencia, mejorar privacidad y habilitar casos donde la nube no llega.

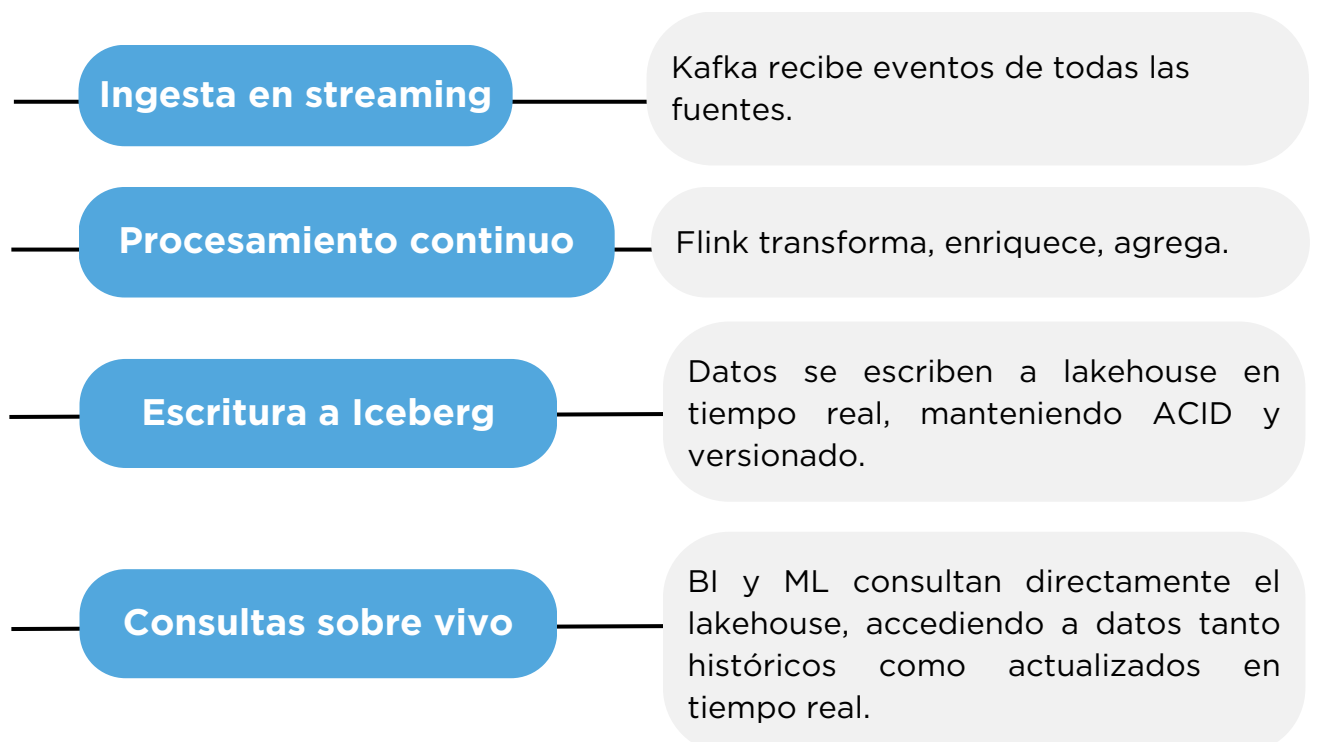
COMBINANDO STREAMING Y EDGE: EL PATRÓN 2026

La tendencia más interesante es la convergencia de streaming en el borde con análisis en la nube



STREAMING + LAKEHOUSE: UNIFICANDO BATCH Y REAL-TIME

Kai Waehner destaca explícitamente que 2026 es el año del "streaming lakehouse": arquitectura donde streaming (Kafka/Flink) y lakehouse (Iceberg/Delta) se integran para unificar datos en movimiento con almacenamiento analítico.



En 2026, la conversación sobre datos ya no es “¿qué tecnología probamos?”, sino cómo orquestamos Lakehouse, Data Fabric, Data Mesh, multi-cloud/híbrido, GenAI y datos en tiempo real para crear un backbone realmente listo para IA. Las organizaciones que capturen valor no serán las que tengan más herramientas, sino las que conviertan esa inversión masiva en big data y analytics en arquitecturas claras, operaciones disciplinadas (MLOps/LLMOps) y observabilidad autónoma que detecte y repare problemas antes de que peguen en el negocio.

El mensaje de fondo es simple: no se trata de hacer más pilotos, sino de industrializar lo que ya sabemos que funciona: Lakehouse como base, Fabric y Mesh para gobernar lo distribuido, RAG y vector databases para volver GenAI confiable, y streaming+edge para que el tiempo real sea la norma, no la excepción.

En los próximos 3 meses, define con tu equipo:

- ¿Cuál será tu backbone de datos para BI, ML y GenAI?
- ¿Qué 2–3 casos de uso priorizarás para demostrar valor medible?
- ¿Qué brechas de MLOps/LLMOps y observabilidad debes cerrar para pasar de pruebas a operación estable?



www.enki.mx



[ENKI](#)