evolutio
Empowering secure AI

IBM
Platinum Partner

Infraestructura diseñada para la era de la IA

Cómo construir una base tecnológica segura, resiliente y escalable para llevar la Inteligencia Artificial a producción





Índice

01 De experimentar con IA a operar con IA

02 Cinco retos para construir IA empresarial a escala

03 No toda la IA necesita la misma infraestructura

04 La arquitectura para la IA empresarial: dato, computación e inteligencia

05 Cómo funcionan juntas estas capacidades

06 De la tecnología al negocio: dónde empezar

07 Evolutio + IBM: de la infraestructura al servicio de IA

08 Una hoja de ruta para pasar a producción

09 Conclusiones: la IA necesita una infraestructura a su altura

01 De experimentar con IA a operar con IA

Durante los últimos años, la conversación sobre **Inteligencia Artificial** ha estado **muy vinculada a la experimentación**: pruebas de concepto, pilotos, asistentes internos, modelos generativos aplicados a tareas concretas... Las organizaciones han usado este periodo para entender qué puede aportar la IA, identificar oportunidades y comprobar su impacto antes de asumir compromisos mayores.

Pero el escenario está cambiando.

La cuestión ya no es únicamente qué puede hacer la Inteligencia Artificial, sino **cómo incorporarla de forma real y sostenida a la actividad de una organización**. Cuando un piloto deja de ser una prueba aislada y empieza a formar parte de un proceso de negocio, cambian también las exigencias que lo rodean.

La IA deja de ser solo una cuestión de modelos y algoritmos para convertirse en una cuestión de **datos, arquitectura, seguridad y operación**.

Y es, precisamente en ese paso **de experimentar a operar**, donde muchas organizaciones empiezan a enfrentarse a preguntas que no estaban presentes en las primeras fases de adopción.



La IA entra en una nueva etapa

Una **prueba de concepto** puede funcionar con un volumen limitado de datos, usuarios y consultas. Se puede ejecutar durante unas semanas, con una arquitectura diseñada específicamente para demostrar su viabilidad, e incluso puede tolerar determinados niveles de intervención manual o de complejidad operativa que serían difíciles de mantener en un servicio empresarial. **La producción plantea otro escenario.**

Cuando una **aplicación de IA** comienza a utilizarse de manera habitual, necesita **responder de forma consistente, integrarse con otros sistemas y estar disponible cuando el negocio la necesita**. Puede tener que procesar grandes volúmenes de información, atender a múltiples usuarios simultáneamente o generar respuestas a partir de datos que se actualizan continuamente.

Además, su impacto deja de estar limitado al equipo que desarrolló el piloto. Una recomendación generada por IA puede **influir en una decisión comercial**; un agente puede interactuar con sistemas corporativos; un modelo puede participar en procesos financieros, operativos o de atención al cliente.

La IA empieza así a formar parte de la cadena de valor de la organización. Y esto cambia la conversación tecnológica.

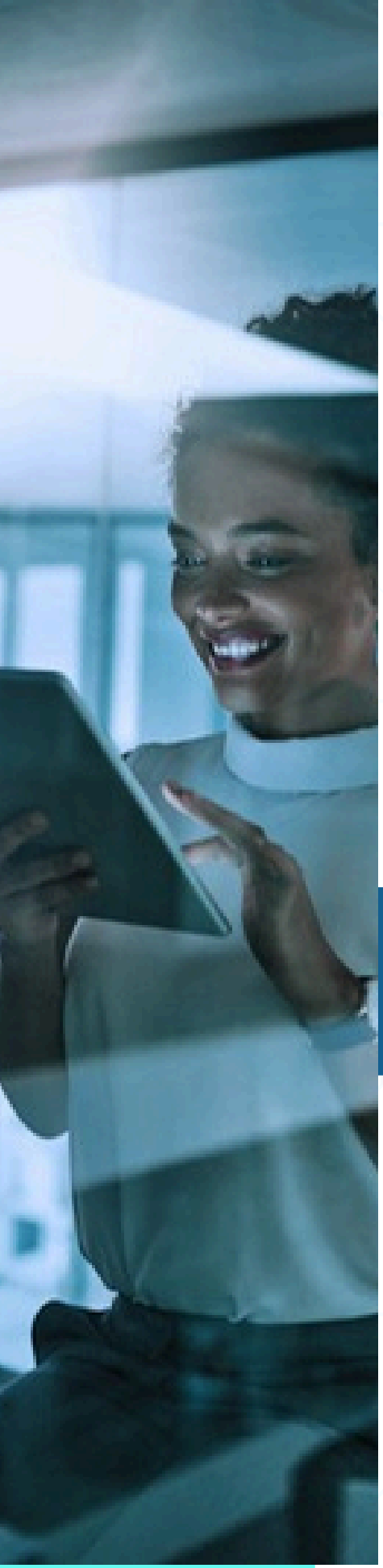
El **rendimiento** deja de medirse únicamente por la capacidad del modelo; también depende de la infraestructura que permite alimentarlo, ejecutarlo y conectarlo con el resto del ecosistema.

La **disponibilidad** deja de ser deseable para convertirse en un requisito.

La **seguridad** ya no afecta únicamente a los sistemas que rodean al modelo, sino también a **los datos, los prompts, el contexto y las interacciones** que se producen alrededor de él.

La propia **infraestructura** pasa, por tanto, a formar parte de la estrategia de IA. Y los datos disponibles apuntan en esa dirección: Gartner previó que, al finalizar 2025, **al menos el 30 % de los proyectos de IA generativa serían abandonados después de la prueba de concepto**, debido, entre otros factores, a una calidad de datos insuficiente, controles de riesgo inadecuados, costes crecientes o falta de claridad sobre el valor de negocio.

Y la conclusión es relevante: **el reto no termina cuando el modelo funciona.**



El verdadero reto comienza cuando la IA se conecta al dato empresarial

La **IA empresarial adquiere buena parte de su valor cuando deja de trabajar únicamente con información genérica** y empieza a utilizar el conocimiento propio de la organización.

Contratos, expedientes, información financiera, documentación interna, propiedad intelectual, datos de clientes o información procedente de sistemas core pueden convertirse en fuentes de valor para nuevos casos de uso. Pero **cuanto más relevante es el dato para el negocio, mayor es también la responsabilidad** asociada a su tratamiento.

Aquí aparece una cuestión fundamental: **no todos los datos pueden tratarse de la misma manera, ni todas las cargas de IA necesitan ejecutarse en el mismo lugar.**

Una aplicación que trabaja con información pública tiene unas necesidades muy diferentes de otra que consulta documentación confidencial mediante **RAG (generación aumentada por recuperación)**, una arquitectura que permite a los modelos recuperar información relevante de fuentes de conocimiento antes de generar una respuesta. Ambas difieren, a su vez, de una carga de IA conectada directamente con un sistema transaccional crítico.

Cambian los requisitos de seguridad, rapidez de respuesta, disponibilidad, integración y gobierno.



Por eso, llevar la IA a producción no consiste simplemente en poner un modelo a funcionar. Implica decidir **cómo se relacionará ese modelo con los datos, dónde se procesará la información, qué sistemas tendrá capacidad de consultar y qué garantías deben existir alrededor de todo ese flujo.**

El punto de inflexión llega cuando la IA **empieza a trabajar con el conocimiento que realmente diferencia a una organización**. Es ahí donde reside buena parte de su valor, pero también donde aparece el riesgo.

La rapidez de respuesta deja de ser una métrica puramente técnica **para convertirse en experiencia de usuario**. La disponibilidad deja de ser una característica deseable cuando una caída **puede afectar directamente** a una operación.

Y la protección de la información deja de limitarse al almacenamiento: también hay que proteger **el contexto que se proporciona al modelo, los prompts, las respuestas y las interacciones con otros sistemas.**

La infraestructura deja de ser un elemento neutro y pasa a formar parte del modelo de confianza de la IA.

Esta realidad cobra especial relevancia a medida que las organizaciones **avanzan hacia arquitecturas híbridas**. La posibilidad de combinar infraestructura on-premise, cloud privada y cloud pública ofrece flexibilidad para elegir dónde ejecutar cada carga, pero también introduce una nueva capa de complejidad que debe gestionarse de forma coherente.

Y es aquí donde la conversación deja de ser exclusivamente tecnológica **para convertirse en estratégica**.

Las nuevas preguntas para CIOs y CTOs

Esta **transición hacia una IA operativa** obliga a los responsables de tecnología a **reformular algunas de las preguntas** que se planteaban durante la fase de experimentación.

¿Dónde debe ejecutarse cada carga de IA?

No donde resulte más sencillo desplegarla, sino **donde mejor encaje con el dato, la criticidad, la latencia y los requisitos regulatorios**.

Cuando la IA accede al conocimiento empresarial, la decisión afecta directamente a **la seguridad, la privacidad, la propiedad intelectual y la soberanía del dat**o.

¿Qué información puede salir de la organización y cuál debe permanecer bajo control?

¿Cómo escalar desde un primer caso de uso hasta múltiples servicios de IA sin multiplicar la complejidad y los costes?

Lo que empieza como una aplicación puede convertirse en múltiples modelos, agentes y fuentes de información. Por eso, necesita **una base tecnológica común y un modelo de operación coherente.**

A mayor impacto de la IA, mayor necesidad de **supervisar su funcionamiento, asegurar la continuidad y poder demostrar cómo se utilizan los datos, modelos y resultados**.

¿Cómo garantizar continuidad, trazabilidad y gobierno cuando la IA participa en procesos reales de negocio?

Responder a estas cuestiones no es un ejercicio teórico: es el paso previo para construir una arquitectura capaz de acompañar la siguiente etapa de la IA. Porque cuando la Inteligencia Artificial deja de ser un experimento y empieza a formar parte del negocio, **la infraestructura deja de ser el soporte de la iniciativa y se convierte en una parte esencial de ella**.

02 Cinco retos para construir IA empresarial a escala

Pasar de un primer caso de uso **a una IA integrada en el negocio** implica mucho más que aumentar la capacidad de procesamiento. A medida que crecen los volúmenes de datos, el número de usuarios y la variedad de aplicaciones, **también lo hacen las exigencias sobre la arquitectura** que las soporta.

Hay **cinco dimensiones** especialmente relevantes para entender este desafío:

Rendimiento **Seguridad** **Resiliencia** **Gobierno** **Eficiencia**

No son cuestiones independientes, sino que **se condicionan entre sí** y, juntas, determinan hasta qué punto una organización está preparada para llevar la IA a una escala empresarial. Veamos un poco más sobre estas cinco dimensiones:



Rendimiento

Acercar la capacidad de proceso al dato

La IA moderna es, en buena medida, un problema de **datos en movimiento**. Las arquitecturas basadas en RAG, la preparación de documentos para que puedan ser consultados por los modelos y la generación de respuestas en tiempo real necesitan procesar grandes volúmenes de información con rapidez y capacidad suficiente. Cuando el dato y la capacidad de proceso están demasiado alejados, el movimiento de información puede convertirse en un cuello de botella.

Además, el volumen de datos continúa creciendo. Ya lo proyectaba el proyecto Data Age 2025 de IDC, previendo que el Global Datasphere **alcanzaría los 175 zettabytes en 2025**, frente a los 33 zettabytes de 2018, una evolución que refleja la magnitud de información que las empresas necesitan almacenar, procesar y convertir en conocimiento.

En este contexto, **acercar la capacidad de procesamiento al dato deja de ser una optimización puntual para convertirse en un principio de diseño.**

Una arquitectura RAG, por ejemplo, debe recuperar información de diferentes fuentes, preparar esa información para poder localizarla y proporcionar al modelo el contexto relevante antes de generar una respuesta. Por eso, no se trata únicamente de disponer de más capacidad, sino de **situarla donde pueda aportar mayor valor.**

Una infraestructura preparada para IA **debe mantener un rendimiento consistente** incluso cuando se pasa de la experimentación al uso intensivo, respondiendo a picos de demanda sin convertirse en el límite de la aplicación.

Cuando la IA interactúa directamente con empleados, clientes o procesos operativos, **una respuesta lenta deja de ser un problema técnico y puede convertirse en un problema de negocio.**

Rendimiento significa conseguir que los datos lleguen a la IA con la rapidez que necesita y que la arquitectura pueda acompañar su crecimiento.

Proteger mucho más que el dato

La incorporación de IA amplía el perímetro de **información que una organización debe proteger**.

Ya no hablamos únicamente de bases de datos o sistemas corporativos: en una aplicación de IA pueden intervenir los datos utilizados para generar una respuesta, prompts introducidos por los usuarios, contexto recuperado de fuentes internas, modelos utilizados y las propias interacciones con la aplicación.

Cada uno de estos elementos **puede contener información sensible** o revelar conocimiento sobre el funcionamiento interno de la organización.

Por eso, **la seguridad debe abordarse de extremo a extremo**: desde el almacenamiento hasta el procesamiento y la utilización de la información.

La protección en reposo y en tránsito sigue siendo fundamental, pero los escenarios de IA hacen necesario prestar atención también a **los datos en uso**, precisamente en el momento en que están siendo procesados.

Aquí cobra relevancia la **computación confidencial**, que permite crear entornos aislados para proteger información y aplicaciones durante el procesamiento. **IBM** plantea esta tecnología para escenarios en los que la confidencialidad y la integridad deben mantenerse durante todo el ciclo de vida del dato, incluido el uso de IA.

Esto cambia la forma de abordar determinados casos de uso.

En lugar de preguntarnos únicamente cómo llevar el dato hasta el modelo, podemos plantearnos **cómo acercar la capacidad de procesamiento al dato sin perder las garantías de protección necesarias**.

La IA empresarial necesita, por tanto, **seguridad alrededor del dato, pero también alrededor del contexto** que permite convertirlo en inteligencia.

Cuando la IA se convierte en un servicio crítico

Durante una prueba piloto, una interrupción puede ser un inconveniente. Cuando una aplicación de IA forma parte de un proceso empresarial, puede convertirse en **una interrupción del propio negocio.**

Las organizaciones que incorporan IA a **procesos críticos** necesitan garantizar la continuidad del servicio, detectar y responder ante incidentes y recuperar la operación ante posibles fallos. Esta resiliencia debe contemplar tanto los incidentes tecnológicos como las amenazas de ciberseguridad, incluido el ransomware, capaz de comprometer los datos y plataformas de los que dependen las aplicaciones de IA.

Y ante esto, los datos de IBM ayudan a dimensionar el problema: su informe Cost of a Data Breach Report 2024 situó el coste medio global de una brecha de datos en **4,88 millones de dólares**, un 10 % más que el año anterior.

Además, el 40 % de las brechas analizadas involucraron **datos distribuidos entre diferentes entornos** (cloud pública, cloud privada y on-premise) y tardaron una media de 283 días en ser identificadas y contenidas.

En arquitecturas híbridas, **cuantos más entornos intervienen, más importante se vuelve contar con visibilidad y con mecanismos coherentes de protección y de recuperación.** Por eso, las copias protegidas, los mecanismos de recuperación y las estrategias de continuidad deben formar parte de la arquitectura desde el principio.

IBM FlashSystem, por ejemplo, incorpora capacidades de copias inmutables y aisladas y funciones de detección de anomalías orientadas a identificar amenazas como el ransomware.

Una IA preparada para producción debe estar diseñada no solo para funcionar cuando todo va bien, sino también para **seguir funcionando, o para recuperarse rápidamente, cuando algo falla.**

Saber qué hace la IA y poder demostrarlo

A medida que la IA adquiere un papel más relevante en la organización, **aumenta también la necesidad de supervisarla**.

El **gobierno de la IA** implica conocer qué modelos y aplicaciones están en uso, con qué datos trabajan, qué procesos soportan y qué controles se aplican sobre ellos. También requiere identificar riesgos, establecer límites y mantener la trazabilidad de su utilización.

Esta necesidad ya no responde únicamente a buenas prácticas: desde el 2 de agosto de 2026 se deben aplicar ciertas **obligaciones de transparencia** previstas por el AI Act, entre ellas informar a los usuarios de cuando están interactuando directamente con determinados sistemas de IA y establecer mecanismos de etiquetado o mención en ciertos contenidos que hayan sido generados o incluso manipulados mediante IA.

Aunque no todos los sistemas estarán sujetos a las mismas obligaciones, el mensaje es claro: **la transparencia y la capacidad de demostrar cómo se utiliza la IA adquieren cada vez más peso.**

Esto cobra especial relevancia en arquitecturas basadas en agentes, donde una aplicación puede consultar diferentes fuentes, utilizar herramientas o desencadenar acciones. **Gobernar la IA requiere mantener la visibilidad y el control sobre estas interacciones a lo largo de su ciclo de vida.**

Para ello, las organizaciones necesitan capacidades que permitan supervisar modelos y activos de IA, evaluar riesgos y rendimiento y mantener la información necesaria para la trazabilidad y la auditoría. **IBM watsonx.governance proporciona estas capacidades, facilitando un gobierno continuo de la IA.**

El gobierno no debería aparecer al final del proyecto. Debe formar parte de la arquitectura desde el principio.

Eficiencia

Crecer sin sobredimensionar

El primer caso de uso de IA rara vez es el último. Una vez se demuestra el valor de una aplicación, **aparecen nuevas oportunidades**: otros departamentos quieren incorporar asistentes, diferentes procesos pueden beneficiarse de agentes y nuevas fuentes de información pueden alimentar aplicaciones existentes.

El reto consiste en que ese crecimiento no se traduzca de forma automática en una proliferación de tecnologías, plataformas y entornos **difíciles de gestionar**.

Una arquitectura preparada para la IA debe permitir **consolidar capacidades, reutilizar recursos y adaptar la capacidad a las necesidades reales de cada carga**.

La **eficiencia** adquiere además una dimensión adicional. Las cargas de IA pueden incrementar las necesidades de computación y almacenamiento, por lo que optimizar recursos deja de ser únicamente una cuestión económica. El **consumo energético y el uso eficiente de la capacidad disponible** forman parte cada vez más de las decisiones tecnológicas.

Escalar no significa simplemente añadir más recursos, sino que **significa crecer de forma ordenada, manteniendo bajo control la complejidad operativa, los costes y el consumo.**

De cinco retos a una decisión arquitectónica

Rendimiento, seguridad, resiliencia, gobierno y eficiencia **no pueden resolverse de manera aislada**. La arquitectura debe diseñarse **en función de las características de cada carga y del dato con el que trabaja**, teniendo en cuenta su sensibilidad, criticidad y necesidades de rendimiento.

No todas las cargas tienen por qué ejecutarse en el mismo entorno.

03 No toda la IA necesita la misma infraestructura

Una de las primeras decisiones que debe abordar una organización cuando empieza a escalar su estrategia de IA es aceptar que **no todas las cargas deben tratarse de la misma manera**.

Una **herramienta generativa** utilizada para resumir información pública no plantea las mismas exigencias que un sistema RAG que **consulta documentación intern**a. Y una aplicación que utiliza información confidencial tampoco tiene los mismos requisitos que una carga de IA conectada directamente con sistemas core o procesos transaccionales.

Sin embargo, es fácil caer en la tentación de buscar una única arquitectura que sirva para todo. Pero, seamos sinceros: **la realidad empresarial es más compleja.**

La infraestructura debe adaptarse a la carga de trabajo y, sobre todo, **al dato que existe detrás de ella.** La sensibilidad de la información, la criticidad del proceso, la rapidez de respuesta necesaria, las obligaciones regulatorias y las necesidades de integración son **factores que deberían participar en la decisión** sobre dónde y cómo ejecutar cada aplicación.



El dato debe decidir dónde se ejecuta la IA

El punto de partida no debería ser preguntarse dónde resulta más sencillo desplegar un modelo, sino qué necesita proteger y **qué necesita conseguir la aplicación que utiliza ese modelo.**

Una aplicación que trabaja exclusivamente con información pública puede priorizar rapidez y flexibilidad. En cambio, cuando intervienen datos personales, propiedad intelectual o información estratégica, **la protección de la información adquiere mayor importancia**.

Lo mismo ocurre con la criticidad: no es igual generar un resumen que participar en una operación financiera, consultar información de un sistema core o intervenir en un proceso que debe permanecer disponible de forma continua.

También importa **la rapidez con la que necesitamos obtener una respuesta**. Algunos casos de uso pueden permitir que la información viaje entre diferentes sistemas antes de ser procesada.

Otros necesitan que los datos **se procesen cerca** de donde se generan para responder de forma más ágil.

Y junto a todo ello está la regulación: las obligaciones relacionadas con privacidad, la protección de datos, soberanía o con determinados sectores pueden condicionar dónde se almacenan y procesan determinadas informaciones.

Por eso, la decisión arquitectónica debería considerar conjuntamente:

- **Sensibilidad del dato.**
- **Criticidad del proceso.**
- **Rapidez de respuesta necesaria.**
- **Requisitos regulatorios.**
- **Necesidades de integración.**

Intentar uniformizarlo todo **puede llevar a dos errores:** sobredimensionar cada carga para responder al nivel más exigente o aplicar a los datos más sensibles una infraestructura pensada para la experimentación.

La estrategia más eficiente no consiste en elegir una infraestructura para toda la IA, sino en decidir qué infraestructura necesita cada tipo de IA.

Tres niveles para una arquitectura de IA

Para gestionar esta diversidad con criterio, proponemos clasificar las cargas en **tres grandes niveles o carriles arquitectónicos**, según la sensibilidad del dato, la criticidad del proceso y las exigencias de la aplicación.



Carril estándar

Destinado a casos de uso de productividad general que trabajan con **datos públicos o no sensibles**, como generación de contenidos, búsqueda en fuentes abiertas, asistentes de uso común o pruebas de concepto.

En estos escenarios, la arquitectura debe priorizar **rapidez, flexibilidad y facilidad de adopción**, permitiendo experimentar y escalar nuevas aplicaciones de IA de forma ágil.



Carril confidencial

Orientado a aplicaciones que trabajan con **datos personales, propiedad intelectual o información empresarial sensible**, como los sistemas RAG sobre bases de conocimiento corporativas.

Aquí cobran especial importancia **la protección del dato, el control de accesos y el aislamiento de las cargas**, para integrar la IA con el ecosistema empresarial sin comprometer la información que utiliza.



Carril crítico

Reservado para cargas vinculadas a **sistemas core, información altamente sensible o procesos transaccionales** donde la disponibilidad y la rapidez de respuesta son determinantes.

En estos escenarios, resulta especialmente relevante **acercar la capacidad de procesamiento al dato**, reduciendo movimientos innecesarios de información y manteniendo la IA dentro de un entorno altamente controlado.

Cuanto mayor es la criticidad del dato y del proceso, mayor debe ser el nivel de control sobre dónde y cómo se ejecuta la IA.

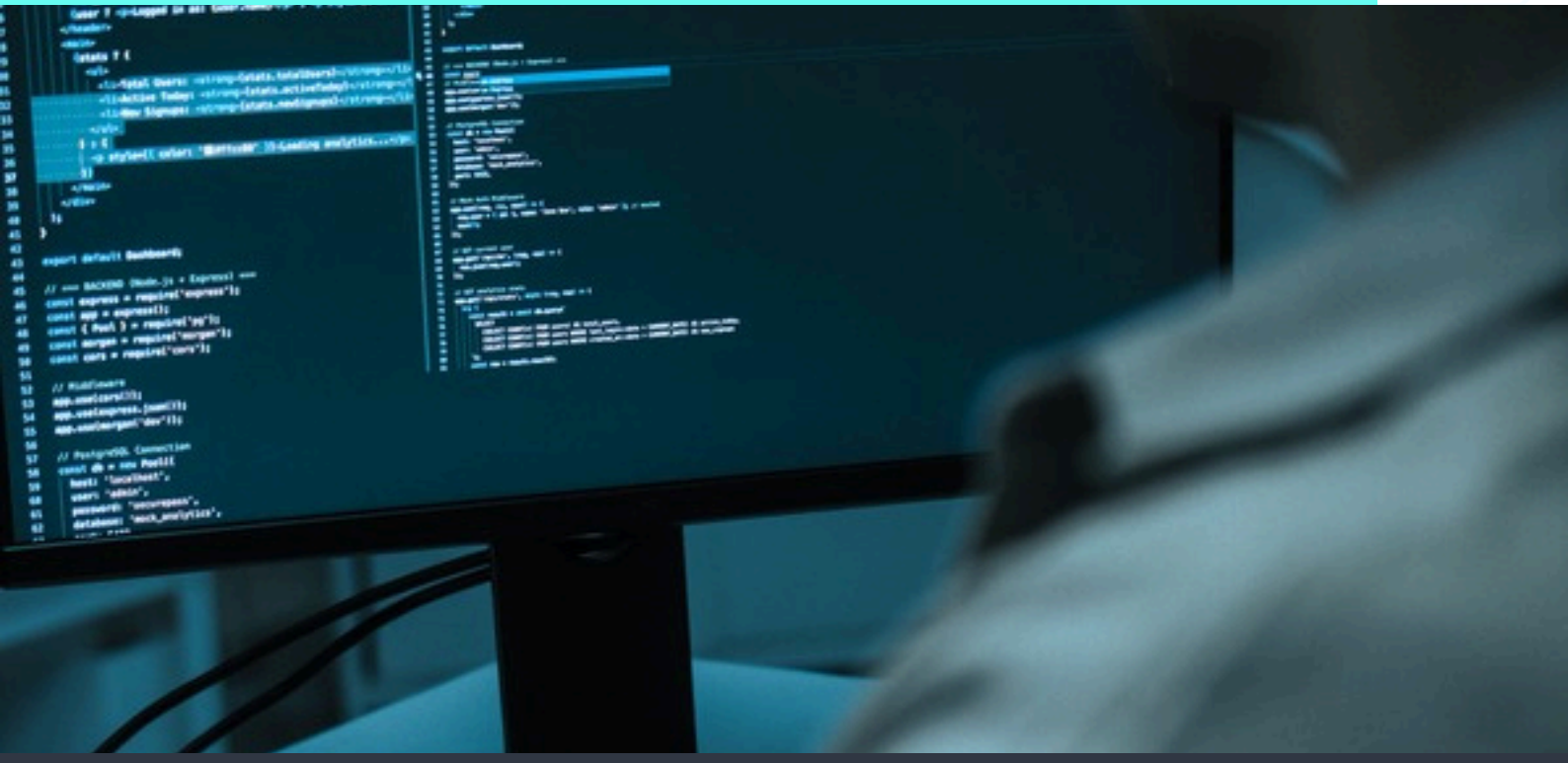
Arquitectura híbrida: ejecutar cada carga donde tenga sentido

Una arquitectura híbrida permite combinar distintos entornos de infraestructura en función de las necesidades de cada carga. Las organizaciones pueden **combinar on-premise, cloud privada y cloud pública**, asignando cada carga al entorno que mejor responde a sus requisitos.

Una organización puede utilizar servicios cloud para determinados escenarios de **experimentación** o de **productividad**, mantener aplicaciones que trabajan con información sensible en un entorno privado y reservar infraestructura propia para cargas críticas que necesitan mayor proximidad a los sistemas y datos corporativos.

La arquitectura híbrida permite adaptar la decisión tecnológica a las características del dato y de cada carga, pero para que esta flexibilidad no se convierta en fragmentación, los diferentes entornos deben poder integrarse y gestionarse bajo unos principios comunes.

Los datos no siempre se mueven, sino que en muchos casos, es la capacidad de IA la que debe acercarse a ellos. Este principio guía la arquitectura que veremos a continuación.



04 La arquitectura para la IA empresarial: dato, computación e inteligencia

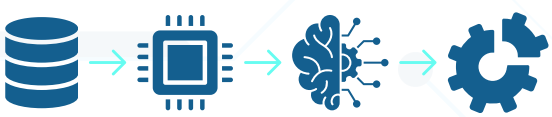
Si **el dato determina dónde tiene sentido ejecutar una carga de IA**, el siguiente paso es entender qué arquitectura permite hacerlo de forma coherente. Una infraestructura preparada para Inteligencia Artificial no debería contemplarse como una suma de productos independientes, sino que **almacenamiento**, **computación**, **plataformas de IA**, **aplicaciones**, **seguridad** y **operación** forman parte de un mismo sistema.

La calidad de las piezas importa, pero la calidad del conjunto importa todavía más.

Una infraestructura diseñada como un sistema

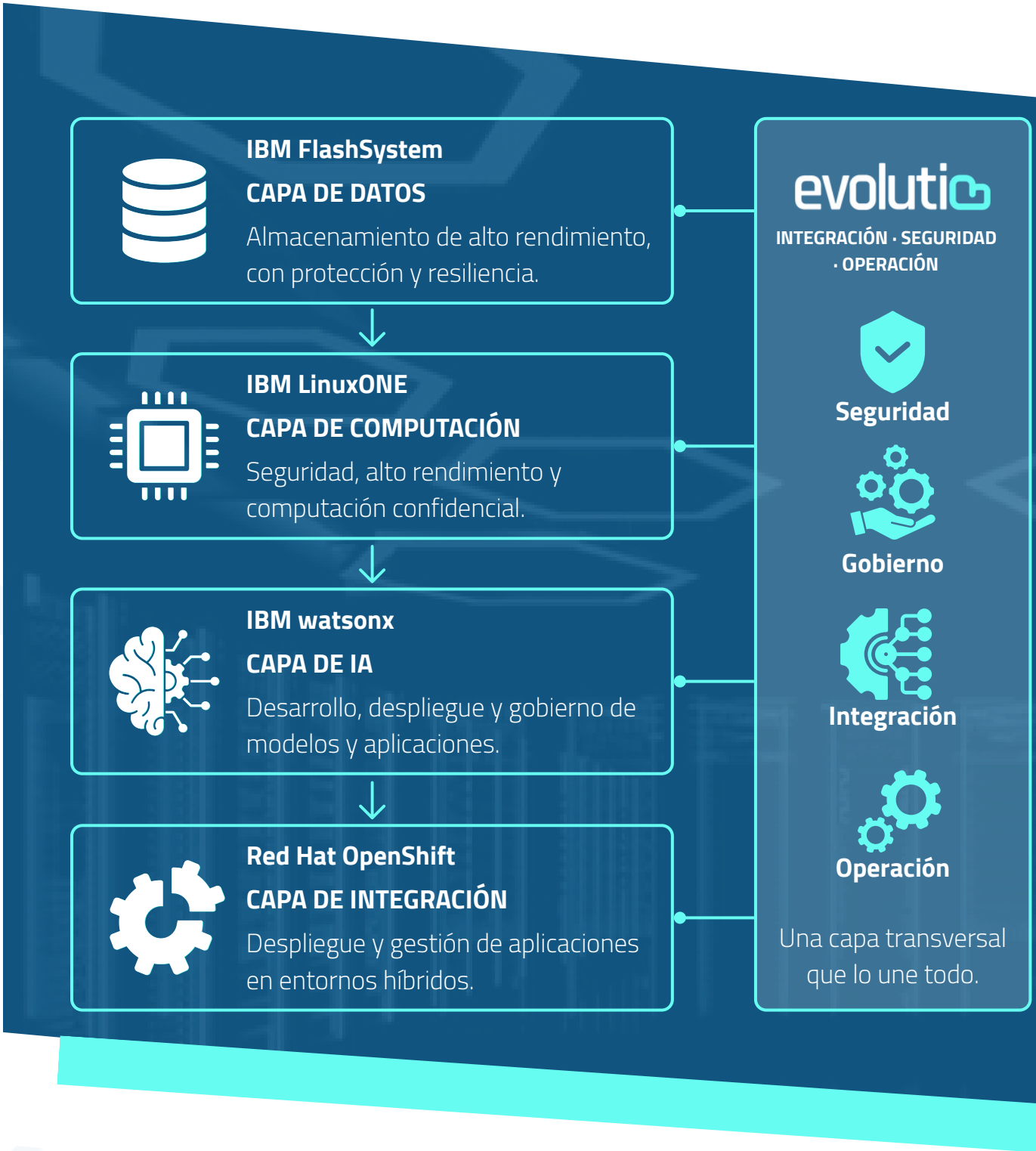
Podemos entender esta arquitectura a través de **cuatro grandes capas:**

- El **dato** es el punto de partida, es la materia prima que permite fundamentar modelos y aplicaciones en el conocimiento real de la organización.
- La **computación** proporciona la capacidad necesaria para procesar esa información y ejecutar las cargas de trabajo. Pero **su ubicación también importa**: en determinados casos será más eficiente llevar la capacidad hacia el dato que mover grandes cantidades de información hacia otro entorno.
- Por encima se sitúa la **plataforma de IA**, donde se desarrollan, despliegan y gobiernan modelos, aplicaciones, RAG y agentes.
- Y, envolviendo todo ello, encontramos la **integración y la operación**, que permiten conectar la arquitectura con los sistemas existentes y garantizar que el conjunto pueda evolucionar.



De las capas a una arquitectura integrada

Una **infraestructura preparada para IA** debe conectar dato, computación, plataforma e integración en un mismo sistema.





IBM FlashSystem: una capa de datos preparada para la IA

El almacenamiento adquiere un papel especialmente relevante **cuando la IA trabaja con grandes volúmenes de información** o necesita consultar continuamente datos empresariales.

Las cargas asociadas a **RAG, preparación de documentos e inferencia** pueden requerir un acceso frecuente al dato, haciendo que **la rapidez de acceso y la capacidad de procesamiento** sean factores relevantes para el rendimiento de la aplicación.

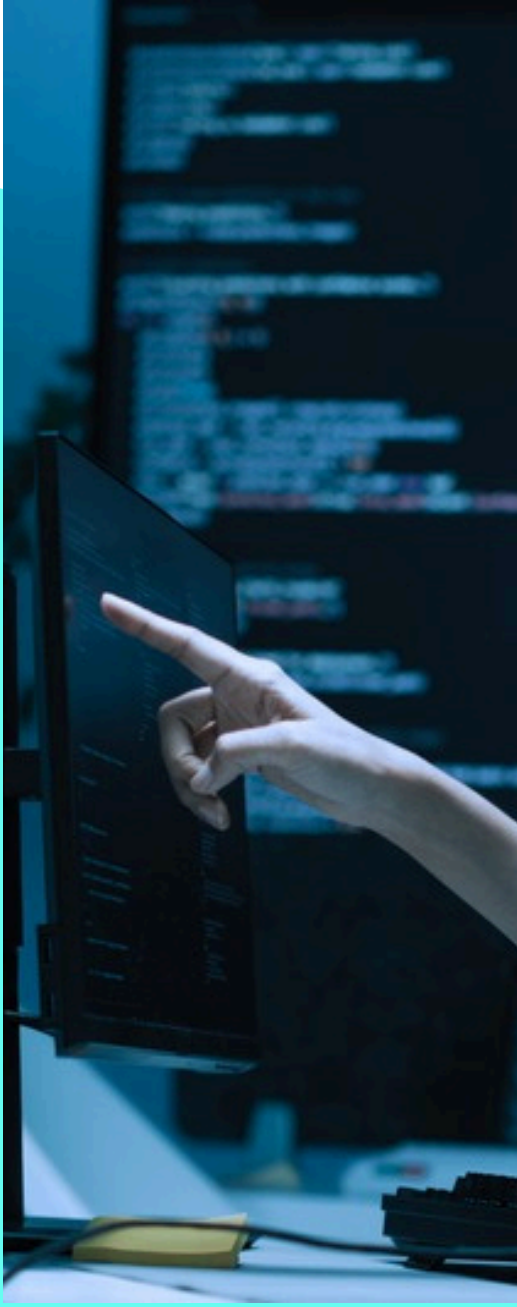
IBM FlashSystem está diseñado para **combinar rendimiento, gestión del dato y resiliencia**, incorporando capacidades de protección como copias inmutables y mecanismos de detección de amenazas.

Esta última capacidad resulta especialmente relevante en un contexto de ciberresiliencia. **IBM FlashSystem utiliza modelos de machine learning** para identificar anomalías asociadas, entre otros comportamientos, al ransomware, y dispone de copias protegidas e inmutables que ayudan a preservar puntos de recuperación.

Esto permite abordar el almacenamiento desde una perspectiva diferente. **No se trata simplemente de guardar información, sino de garantizar que el dato que alimenta la IA esté disponible, protegido y recuperable cuando el negocio lo necesite.**

En una arquitectura de IA, por tanto, **el almacenamiento deja de ser una capa pasiva para convertirse en una pieza activa de rendimiento y resiliencia.**

IBM LinuxONE: IA junto al dato crítico



Cuando la IA trabaja con información especialmente sensible o participa en procesos críticos para el negocio, **dónde se procesan los datos cobra todavía más importancia.**

IBM LinuxONE está diseñado para cargas empresariales que necesitan **altos niveles de seguridad, disponibilidad y capacidad de crecimiento**. La plataforma combina estas capacidades con funcionalidades de IA y mecanismos avanzados para proteger la información durante su procesamiento.

Una de sus posibilidades más relevantes en determinados escenarios es **ejecutar la IA cerca de los datos. IBM** denomina a este enfoque **Transactional AI** y permite incorporar capacidades de inferencia directamente en sistemas de transacciones, reduciendo el tiempo y los movimientos innecesarios de información.



La lógica es sencilla: en determinados casos, **en lugar de llevar el dato hacia la IA, tiene más sentido llevar la IA hacia el dato.**

A esto se suma la **computación confidencial**. **IBM Secure Execution for Linux** permite crear entornos aislados en los que **los datos y las aplicaciones permanecen protegidos mientras están siendo procesados.**

Este enfoque resulta especialmente interesante cuando la IA necesita trabajar con información sensible que **no debería salir innecesariamente del entorno corporativo.**

IBM watsonx: de los modelos a la IA empresarial gobernada

La infraestructura **proporciona la base**, pero una organización necesita también las herramientas que permitan **convertir esa capacidad tecnológica en aplicaciones de IA útiles y controladas.**

IBM watsonx.ai proporciona un **entorno para desarrollar y desplegar soluciones de IA,** trabajar con diferentes modelos y construir aplicaciones basadas en RAG y otras arquitecturas. Además, ofrece **flexibilidad para adaptarse a distintos modelos y formas de despliegue**.

Esta flexibilidad es importante porque **no existe un único modelo adecuado para todos los casos de uso.** La elección dependerá de factores como la tarea, el nivel de precisión necesario, los costes, el riesgo o dónde se necesite ejecutar la solución.

En el caso de RAG, **IBM watsonx.ai** permite conectar los modelos con las fuentes de conocimiento de la organización para generar respuestas basadas en **información relevante y actualizada.**

Pero desarrollar una aplicación es solo una parte del proceso. Cuando la IA llega a producción, también es necesario **saber cómo se comporta, identificar sus riesgos y mantener el control sobre ella a medida que evoluciona.**

Aquí entra **IBM watsonx.governance,** que proporciona capacidades para gestionar y supervisar los activos de IA, desde la evaluación de los modelos hasta el seguimiento de su rendimiento, riesgos y cumplimiento.

Esta capa cobra todavía más importancia con las arquitecturas agentivas. Los agentes pueden consultar información, utilizar herramientas y ejecutar acciones, por lo que el control debe abarcar también **qué hacen, cómo lo hacen y con qué información trabajan**.

La IA empresarial no termina cuando el modelo responde. Empieza realmente cuando podemos operar ese modelo con garantías.

Una arquitectura abierta para entornos híbridos



La realidad tecnológica de muchas organizaciones es híbrida. Aplicaciones existentes, sistemas propios, servicios cloud y nuevas cargas de IA deben **convivir sin que incorporar nuevas tecnologías implique reconstruir todo el entorno desde cero.**

En este contexto, las plataformas abiertas y basadas en contenedores adquieren especial relevancia.

Red Hat OpenShift proporciona una capa común para desplegar aplicaciones y cargas de trabajo en diferentes entornos, facilitando su integración con las aplicaciones y fuentes de datos que ya existen.

El objetivo no es eliminar la diversidad tecnológica, sino hacerla gestionable.

Una **arquitectura abierta** permite evolucionar modelos, aplicaciones y entornos de ejecución sin que cada cambio implique empezar de nuevo. Y facilita que las nuevas cargas de IA puedan incorporarse progresivamente **sin romper aquello que ya sostiene el negocio.**



05 Cómo funcionan juntas estas capacidades

Una vez analizadas las diferentes piezas, conviene evitar **entenderlas como productos independientes**. El verdadero valor aparece cuando estas capacidades se integran dentro de un flujo completo:



Cada etapa necesita apoyarse en la anterior.



Del dato al servicio de IA

Imaginemos una organización que quiere crear un asistente **capaz de responder preguntas sobre documentación interna**.

El proceso empieza en el dato: contratos, procedimientos, normativa, manuales o documentación corporativa. Esa información debe **almacenarse y prepararse** para poder ser consultada de manera eficiente.

A continuación, una **arquitectura RAG** puede recuperar la información relevante y **proporcionar ese contexto** al modelo. El modelo genera entonces una respuesta que llega a una aplicación o asistente utilizado por los empleados.

Pero el flujo no termina ahí, ya que la organización necesita **controlar quién puede acceder a la información**, conocer cómo funciona la aplicación, monitorizar su comportamiento y **establecer mecanismos de gobierno**.

La Inteligencia Artificial es el resultado de un sistema, no de un único componente.

La arquitectura debe evaluarse como un conjunto: un almacenamiento rápido sin recuperación ante incidentes deja expuesto el dato, un modelo potente sin gobierno deja a la organización sin trazabilidad, y una aplicación brillante sin operación difícilmente podrá mantenerse en producción.

Una arquitectura de referencia Evolutio + IBM

En este modelo, **IBM FlashSystem** actúa como capa de datos; **IBM LinuxONE**, como infraestructura para cargas críticas; **IBM watsonx**, como plataforma para desarrollar y gobernar las capacidades de IA; y **Red Hat OpenShift,** como capa abierta para desplegar aplicaciones y cargas en entornos híbridos.

Evolutio aporta la capa de **integración, diseño, seguridad y operación** necesaria para convertir estas capacidades en una arquitectura adaptada al entorno concreto de cada organización.

La importancia está precisamente en esa integración.

Las organizaciones no necesitan únicamente tecnología. Necesitan **una arquitectura que encaje con sus aplicaciones, sus datos, sus políticas de seguridad y su modelo operativo**.

El resultado: IA empresarial con control

Cuando las diferentes capas están diseñadas conjuntamente, **cada carga puede ejecutarse donde tiene sentido.**

Los **datos críticos** pueden permanecer próximos a la capacidad de procesamiento que los necesita. Las **cargas** pueden integrarse con aplicaciones corporativas sin crear silos aislados. La **información sensible** puede mantenerse bajo los controles adecuados y **el ciclo completo puede quedar sujeto a mecanismos de gobierno y observabilidad**.

El resultado no es simplemente una infraestructura más potente, sino **una arquitectura capaz de acompañar a la IA desde el dato hasta el servicio.**

Y precisamente porque puede soportar diferentes escenarios, aparece una pregunta más concreta: **¿por dónde empezar?**

06 De la tecnología al negocio: dónde empezar

Una **infraestructura preparada para IA** puede habilitar numerosos casos de uso. Pero eso no significa que una organización deba abordarlos todos al mismo tiempo.

Uno de los **principales riesgos de una estrategia de IA** es empezar por la tecnología en lugar de hacerlo por una necesidad real del negocio.

> La pregunta no debería ser "¿qué modelo podemos desplegar?", sino "**¿qué problema merece la pena resolver y qué necesitamos para hacerlo de forma segura y medible?**"

Qué caracteriza un buen primer caso de uso

Un buen punto de partida suele combinar cuatro elementos: **impacto medible, datos disponibles, un proceso repetible y una necesidad real de seguridad, disponibilidad o trazabilidad.** Esto permite evitar dos extremos.

Por un lado, **casos demasiado pequeños** que difícilmente justificarán una industrialización. Por otro, **proyectos excesivamente ambiciosos** que intentan transformar procesos críticos antes de haber validado el modelo operativo.

El primer caso de uso debería servir también para aprender: sobre los datos, sobre la arquitectura, sobre los controles necesarios y sobre la forma en que los usuarios interactúan con la IA.

La prueba de concepto deja entonces de ser únicamente una demostración tecnológica.
Y se convierte en el primer paso hacia un servicio empresarial.

RAG documental sobre información sensible

Uno de los escenarios con mayor potencial para muchas organizaciones es utilizar RAG sobre documentación propia.

Contratos, expedientes, normativa, procedimientos y conocimiento interno **contienen información que puede resultar muy valiosa** cuando se hace accesible **mediante lenguaje natural**.

En lugar de obligar a los usuarios a localizar manualmente información entre grandes volúmenes documentales, **una aplicación puede recuperar el contenido relevante** y utilizarlo como contexto para generar una respuesta.

El valor es doble, puesto que resuelve una necesidad real y, al mismo tiempo, pone a prueba varias de las capacidades más exigentes de la arquitectura: **rendimiento, protección de la información y trazabilidad.**

Es, por tanto, un caso especialmente interesante para comenzar.

Copilotos y agentes para empleados

El siguiente paso puede ser llevar esa inteligencia **directamente al trabajo diario**.

Un copiloto puede ayudar a consultar procedimientos, resumir documentación o preparar respuestas. **Un agente puede ir más allá** y utilizar herramientas para ejecutar determinadas acciones en diferentes sistemas. La diferencia es importante.

Mientras un asistente principalmente proporciona información, **un agente puede convertirse en parte activa del proceso.**

Esto abre oportunidades relevantes para **mejorar productividad y reducir tareas repetitivas**, pero también **eleva las exigencias de gobierno**. Cuantas más acciones puede realizar una aplicación, más importante resulta controlar sus permisos, sus fuentes de información y sus interacciones con otros sistemas.

Analítica e inferencia crítica

No todos los casos de uso tienen que basarse en IA generativa. La Inteligencia Artificial puede intervenir en **detección de fraude y riesgo**, mantenimiento predictivo, control de calidad o priorización operativa.

En estos escenarios, la capacidad de analizar datos y generar resultados en el momento adecuado puede tener **un impacto directo sobre procesos críticos**.

Aquí vuelven a cobrar protagonismo **la rapidez de respuesta, la disponibilidad y la proximidad al dato.**



Cuando una inferencia forma parte de una operación empresarial, **la arquitectura que la soporta pasa a ser tan relevante como el propio modelo.**

Casos sectoriales

Aunque las tecnologías sean comunes, las oportunidades concretas cambian según el sector.

- En **banca y seguros**, pueden aparecer casos de uso relacionados con fraude, scoring, gestión de siniestros, reclamaciones o asistentes para gestores.
- En **industria y logística**, la IA puede aplicarse al mantenimiento predictivo, la planificación, la ingeniería o la protección del conocimiento y la propiedad intelectual.
- En **salud y Administraciones Públicas,** el tratamiento de expedientes y documentación puede beneficiarse de RAG y asistentes inteligentes, siempre dentro de requisitos especialmente exigentes de seguridad, privacidad y gobierno.
- En **retail y media**, las aplicaciones pueden orientarse hacia una atención omnicanal, recomendación, generación de contenidos, previsión de demanda e identificación de insights.

El denominador común es siempre el mismo: **el caso de uso determina las necesidades de la arquitectura, y no al contrario.**

07 Evolutio + IBM: de la infraestructura al servicio de IA

Llevar una aplicación desde un entorno de prueba hasta producción **implica mucho más que desplegarla.** Es necesario integrarla con las aplicaciones existentes, definir los controles de seguridad, establecer mecanismos de observabilidad y garantizar que el servicio **puede mantenerse y evolucionar**.

La **seguridad** debe formar parte del diseño desde el principio, y lo mismo ocurre con la **observabilidad** y el **gobierno**.

Industrializar la IA significa convertir una capacidad tecnológica en un servicio empresarial gestionable. Y ahí es donde la experiencia de integración adquiere valor.



El papel de IBM

IBM aporta las capacidades tecnológicas que permiten construir las diferentes capas de la solución.

IBM FlashSystem proporciona almacenamiento empresarial orientado al **rendimiento**, la **protección** y la **resiliencia**.

IBM LinuxONE aporta una plataforma para cargas Linux de misión crítica, con capacidades avanzadas de **seguridad**, **computación confidencial** y ejecución de **IA próxima al dato**.

IBM watsonx proporciona capacidades para desarrollar y **desplegar soluciones de IA**, trabajar con **modelos** y **RAG** y **gobernar activos** durante su ciclo de vida.

Y las tecnologías de **Red Hat** aportan una capa abierta para **desplegar cargas en entornos híbridos**.

IBM aporta, en suma, las piezas tecnológicas.
El reto consiste en ensamblarlas de forma coherente.



evolutio | IBM Platinum Partner



El papel de Evolutio como integrador

L**a tecnología necesita adaptarse a la realidad de cada organización.** No todas tienen los mismos sistemas, los mismos datos, los mismos requisitos regulatorios ni el mismo nivel de madurez en IA.

Por eso, el papel de Evolutio se sitúa precisamente en esa capa de integración.

En Evolutio diseñamos, integramos, protegemos y operamos la arquitectura que permite convertir las capacidades de IBM en un servicio de IA adaptado a cada organización.

Esto implica **acompañar el proyecto** desde la identificación de oportunidades y la clasificación de los datos hasta el diseño arquitectónico, la integración con sistemas existentes, la ciberseguridad, la observabilidad y la operación.

El objetivo es evitar que la IA se convierta en **un nuevo silo tecnológico.**

Un modelo end-to-end

El recorrido puede resumirse en **cinco etapas**:

ASSESSMENT
Identificar oportunidades, datos y requisitos.

ARQUITECTURA
Diseñar la base tecnológica adecuada.

POC
Validar el caso de uso y su valor.

INDUSTRIALIZACIÓN
Preparar la solución para producción.

OPERACIÓN
Monitorizar, optimizar y evolucionar.

Primero se identifica **dónde puede aportar valor la IA** y qué requisitos existen, después se diseña la arquitectura adecuada. La **prueba de concepto** permite validar el valor y las métricas, la **industrialización** convierte ese caso en un servicio preparado para producción, y la operación permite **monitorizarlo**, **optimizarlo** y **evolucionarlo**.

Este enfoque permite que la infraestructura no se diseñe únicamente para el primer caso de uso, sino **pensando desde el principio en cómo crecerá la estrategia de IA de la organización.**



08 Una hoja de ruta para pasar a producción

Una **estrategia de IA empresarial** necesita avanzar de manera **progresiva**. No todos los casos de uso requieren la misma arquitectura ni todos están preparados para llegar directamente a producción.

Establecer una hoja de ruta permite **reducir riesgos, demostrar valor y tomar las decisiones tecnológicas en el momento adecuado.**

Discovery: identificar dónde aportar valor

Todo comienza por **comprender el negocio**. ¿Qué objetivos se quieren conseguir? ¿Qué equipos están implicados? ¿Qué procesos presentan oportunidades de mejora? ¿Qué datos están disponibles?

El resultado no debería ser todavía una solución tecnológica, sino una **selección fundamentada de oportunidades**, priorizadas por impacto y viabilidad.

La tentación en esta fase es saltar directamente a la tecnología, así que es importante tener disciplina y **pararse a entender el valor y la disponibilidad del dato** y, más adelante, evitar pilotos estériles.

Assessment: decidir el modelo de despliegue

Una vez identificado el caso de uso, llega el momento de **analizar sus requisitos**. La sensibilidad del dato, la criticidad del proceso, la arquitectura existente, los riesgos y las necesidades de seguridad y cumplimiento **determinarán qué entorno resulta más adecuado**.

Es aquí donde **se decide en qué carril** (estándar, confidencial o crítico) encaja la carga y qué combinación de cloud pública, privada y on-premise resulta adecuada.

La arquitectura debe responder al caso de uso, y no al contrario.

POC: demostrar valor con métricas

La prueba de concepto debe responder a una pregunta concreta: **¿funciona y aporta el valor esperado?**

Puede tratarse de un RAG documental, un copiloto, un agente o una aplicación de IA para un proceso crítico. Pero en todos los casos **conviene definir previamente qué métricas determinarán el éxito**.

La **precisión de las respuestas** puede ser relevante en un RAG; el tiempo ahorrado, en un copiloto; la **tasa de automatización**, en un agente; y la rapidez de respuesta o capacidad de proceso, en una aplicación crítica.

El objetivo no es demostrar que la IA puede hacer algo, **es demostrar que merece la pena convertirlo en un servicio.**

Industrialización: convertir el caso de uso en servicio

Una vez validado el valor, comienza una fase que suele requerir más atención de la que recibe durante la experimentación: la arquitectura de producción debe estar definida, el modelo operativo debe establecer responsabilidades, la seguridad y el gobierno deben estar incorporados, y, por supuesto, **debe existir un plan para desplegar el servicio de forma controlada.**

Aquí es donde la **infraestructura** adquiere todo su sentido: aquello que funcionaba en un entorno limitado debe poder soportar usuarios reales, datos reales, integraciones reales y condiciones reales de operación. **La industrialización es lo que separa un piloto exitoso de un servicio empresarial sostenible.**

Operación y evolución continua

La puesta en producción no es el final: los modelos cambian, los datos evolucionan, los usuarios encuentran nuevas necesidades y aparecen nuevos casos de uso. Por eso, la IA necesita **monitorización, optimización, continuidad y gobierno continuo.**

Además, la **experiencia adquirida** con el primer caso de uso puede utilizarse para acelerar los siguientes. La arquitectura, los procesos y los mecanismos de operación desarrollados dejan de servir únicamente a una aplicación y **pasan a convertirse en capacidades reutilizables para la estrategia global de IA.**

La **hoja de ruta** queda así definida:



Un recorrido que permite avanzar con **ambición**, pero también con el nivel de **rigor** necesario para llevar la Inteligencia Artificial a procesos empresariales reales.

09 Conclusiones: la IA necesita una infraestructura a su altura

La **siguiente etapa de la Inteligencia Artificial** empresarial no consiste únicamente en incorporar modelos más avanzados.

Consiste en crear las condiciones para que esos modelos puedan trabajar de forma **segura, eficiente y fiable sobre los datos y procesos que realmente importan al negocio.** La diferencia entre experimentar con IA y operar con IA está precisamente ahí.

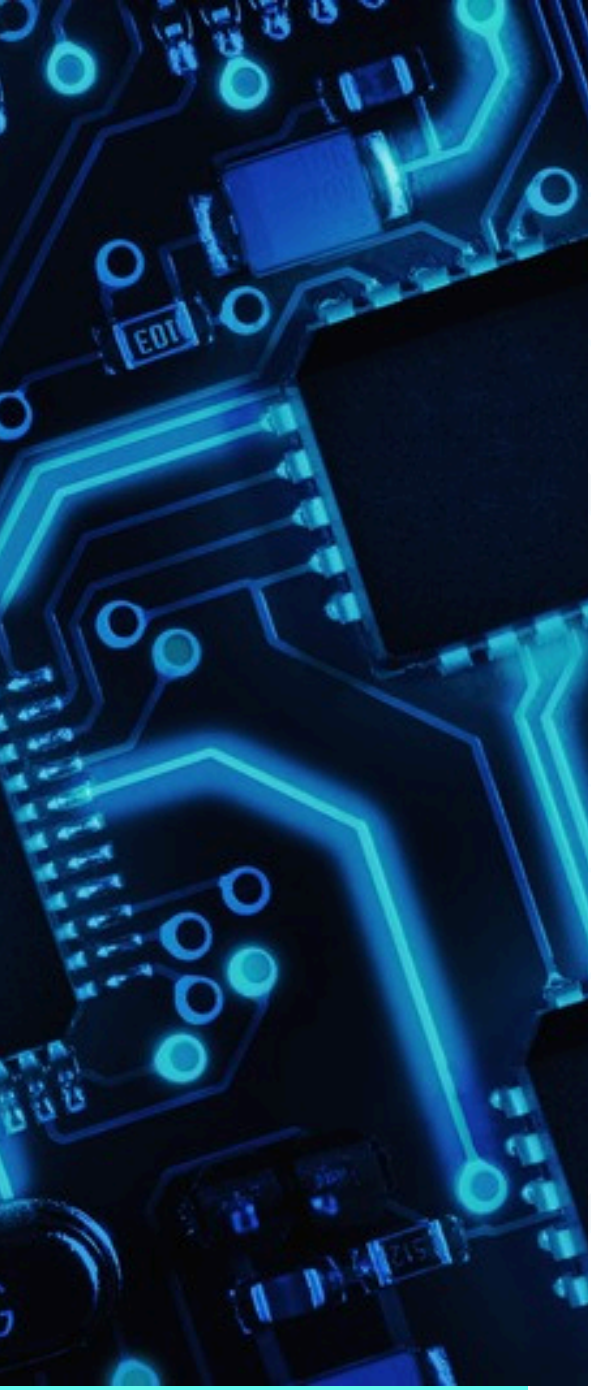
Una prueba puede demostrar el potencial de una tecnología. Pero convertir ese potencial en valor sostenido exige resolver cuestiones de **rendimiento**, **disponibilidad**, **seguridad**, **resiliencia**, **gobierno** y **eficiencia**.

Y ninguna de ellas puede abordarse de manera aislada.

El dato debe determinar **dónde tiene sentido ejecutar cada carga.** La capacidad de procesamiento debe situarse donde pueda aportar mayor valor. Los modelos y aplicaciones deben poder evolucionar dentro de un marco de gobierno. Y todo el conjunto debe integrarse con los sistemas que ya sostienen el negocio.

Por eso, **la infraestructura deja de ser simplemente el soporte de la IA y pasa a formar parte de su estrategia.**





Una arquitectura basada en capacidades como **IBM FlashSystem**, **IBM LinuxONE** y **IBM watsonx** permite abordar diferentes necesidades dentro de un mismo ecosistema: desde la protección y disponibilidad del dato hasta la ejecución de cargas críticas, pasando por el desarrollo, despliegue y gobierno de aplicaciones de IA.

Sobre esta base, **la experiencia de Evolutio** en diseño, integración, seguridad y operación permite llevar estas capacidades a escenarios empresariales concretos, conectándolas con los sistemas, datos y necesidades de cada organización.

El objetivo final no es tener más IA, e**s tener una IA capaz de trabajar allí donde realmente puede aportar valor, con las garantías que exige cada dato y cada proceso.**

Porque la IA que importa no es la que se demuestra en un piloto. Es la que **opera cada día, de forma segura, eficiente y trazable, sobre los procesos que sostienen el negocio.**

IA útil, escalable y gobernada, con una infraestructura a su altura.

La IA no necesita una infraestructura única. Necesita una infraestructura capaz de ejecutar cada carga donde tenga sentido.

En Evolutio le ayudamos a convertir el potencial de la IA en capacidades empresariales reales, seguras, escalables y preparadas para evolucionar.

evolutio.com 91 270 6000 info@evolutio.com