

Más allá de un concepto, Data Fabric, un ejemplo práctico PARTE 2





Álvaro Enrique Chacón M.,
MBA, CISA, CDPSE, CSXF, CFC
Director Auditoría TI -BKF INTERNATIONAL

En el nuevo mundo digital, los datos son la esencia de las nuevas tecnologías. Su uso, cada vez más extendido, revela su importancia, más aún hoy en día, cuando la dinámica de los nuevos negocios, apoyada en un sofisticado manejo estratégico de los datos, desde el cual se asegure sea éste cada vez más completo y oportuno, no será de desconocer contribuye a una mayor competitividad y a la obtención de ventajas al identificar patrones, tendencias y comportamientos, que muy seguro son de gran interés para los CEO's de las empresas hoy conocer.

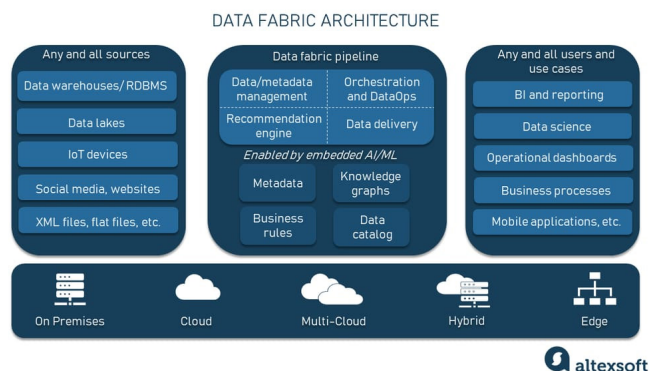
En el artículo anterior “¿Sabes que es el Data Fabric o Tejido de Datos? – Parte 1” hicimos referencia al concepto de Data Fabric o Tejido de Datos como igual se conoce en español, forma parte de la arquitectura no solo de datos, sino además de la organización, para proyectar eficiencia en el mejoramiento de sus procesos y la toma de nuevas decisiones. En otras palabras, se trata de un esquema que más allá de lo conceptual es netamente estratégico.

Son bases que operan en buena medida soportadas en API's (conectores basados en interfaces de programación para aplicaciones), desde los cuales, sin comprometer las fuentes de información, hace posible se pueda dar lectura a los datos y en línea si así fuera necesario, para construir nueva información y de mayor sentido para la toma de decisiones estratégicas y corporativas. Dijimos además que no se trata de grandes inversiones en nuevas tecnologías necesariamente, en esencia no es una herramienta más, pero sí una manera diferente de trabajar con los datos basado en un nuevo concepto de arquitectura que permitirá la unicidad, regulación y seguridad de los datos. Si no conoce nuestro artículo anterior, lo invito a que lo revise antes de continuar, (bkf.com.co - **item EL INDICADOR**) para que pueda aprovechar mucho mejor este contenido, el cómo puede ser reconocido al interior de su organización, indistintamente la plataforma de software que tenga instalado.

Para comprender cómo funcionaría un Data Fabric dentro de una organización, puede imaginarse como una capa inteligente de conexión y gobierno de datos que se sitúa entre las fuentes de información y los sistemas que consumen esos datos (analítica, aplicaciones, dashboards o automatización).

A continuación, se describe de forma simplificada su funcionamiento y los recursos informáticos que normalmente intervienen. No obstante, aunque se indican de una manera simple, de antemano vale reconocer que, para su aplicación y tratamiento, sí es requerido conocimientos en arquitectura y programación moderna; además de, contar con un buen conocimiento de los sistemas de información y procesos instalados del negocio, para identificar cómo se puede interactuar y generar esos conectores y gobierno en el tratamiento o manejo de los datos.

Permítame volver a compartir el diagrama que muestra la arquitectura y sus componentes:



1. Fuentes de datos (Data Sources)

Esta fuente de datos es con lo que cuenta un sistema de información por naturaleza, donde se almacenan los datos e información en su conjunto, que mueve la operación de negocio de la organización o empresa.

Ejemplos comunes:

- ERP (finanzas, contabilidad)
- CRM (clientes y ventas)
- Sistemas operacionales (producción, logística)
- Aplicaciones SaaS (marketing, RRHH)
- Bases de datos internas
- Archivos y documentos
- Plataformas en la nube

Vale recordar como se indicó en el artículo anterior, muchos de los sistemas tradicionales y arquitecturas en tecnología con que cuentan las organizaciones hoy en día, operan como silos, instalaciones donde cada uno opera y entrega resultados de manera aparte, no permitiendo la consolidación de información confiable para propósitos estratégicos.

2. Conectividad e integración (APIs y conectores)

Como ya dijimos, para que el Data Fabric funcione, se requiere un mecanismo de interconexión entre sistemas, para ello veamos cuáles serían algunos y su aplicabilidad.

APIs (Application Programming Interfaces)

Permiten que aplicaciones diferentes intercambien datos de forma estandarizada.

Ejemplo:

- El CRM envía datos de clientes al sistema de analítica.
- El ERP consulta inventarios en tiempo real desde el sistema logístico.

Conectores o pipelines de datos

Herramientas que extraen y sincronizan información.

Ejemplos de tecnologías usadas en la industria:

- Apache Kafka
- Talend
- Fivetran
- Azure Data Factory

Estas herramientas permiten crear flujos automáticos de datos.

3. Capa de almacenamiento y repositorio compartido

El Data Fabric no siempre requiere mover todos los datos desde la fuente, pero sí suele apoyarse en infraestructura de almacenamiento organizada.

Algunos componentes típicos:

Data Lake

Vale hacer una breve descripción de este concepto y tipo de repositorio de almacenamiento. Este trata de un repositorio centralizado para almacenar, procesar y asegurar grandes volúmenes de datos, guardando su formato original o nativo.

A diferencia de los almacenes de datos tradicionales (data warehouse o bodega de datos), permite gestionar información estructurada, semiestructurada y no estructurada (imágenes, videos, logs) sin que sea necesario se especifique un esquema previo.

Características:

- Datos estructurados y no estructurados
- Centralización
- Versatilidad
- Escalabilidad
- Almacenamiento histórico

Ejemplos:

- AWS S3
- Azure Data Lake
- Google Cloud Storage

Data Warehouse

Base optimizada de almacenamiento para un gran volumen de datos estructurados para análisis empresarial.

Ejemplos:

- Snowflake
- BigQuery
- Amazon Redshift

4. Capa de metadatos y catalogación

Esta es una de las piezas más importantes del Data Fabric, es aquí donde se registra:

- El origen de los datos
- Definiciones
- Relaciones entre tablas
- Responsables de los datos
- Reglas de calidad

por ello este componente se conoce como Catálogo de datos (Data Catalog)

Ejemplos de herramientas, algunas más conocidas que otras:

- Collibra
- Alation

- IBM Knowledge Catalog
- Apache Atlas

5. Gobierno y seguridad de los datos

El Data Fabric también incorpora controles de gobierno.

Esto incluye:

- Control de accesos
- Clasificación de datos sensibles
- Auditoría de uso
- Cumplimiento regulatorio
- Políticas de calidad de datos

Esto permite responder preguntas como:

- ¿Quién creó el dato?
- ¿Quién lo modificó?
- ¿Es confiable?
- ¿Puede ser usado por otras áreas?

6. Capa de consumo de datos

Finalmente, los datos integrados pueden ser utilizados por diferentes aplicaciones empresariales:

- Dashboards ejecutivos
- Business Intelligence
- Modelos de analítica avanzada
- Aplicaciones empresariales
- Automatización de procesos
- Sistemas de IA

Herramientas comunes:

- Power BI
- Tableau
- Looker
- aplicaciones corporativas

EJEMPLO SIMPLE DENTRO DE UNA EMPRESA

Supongamos una compañía comercial:

- El CRM registra clientes y ventas
- El ERP registra facturación
- El sistema de logística maneja inventarios

Ahora qué se podría hacer mediante la implementación de Data Fabric, demos un vistazo a este caso práctico:

1. APIs conectan los sistemas.
2. Los datos se sincronizan en un repositorio común.

3. El catálogo identifica cada dato.
4. Las reglas de gobierno validan calidad y seguridad.

Resultado:

Un gerente o CEO puede ver en un dashboard en tiempo real:
Ventas + Inventario + Rentabilidad sin esperar reportes manuales.

En resumen

El Data Fabric funciona como un ecosistema de integración inteligente que va a estar compuesto por:

- conectores y APIs
- repositorios de datos
- catálogo de metadatos
- reglas de gobierno
- herramientas de análisis

Todo esto permite que la organización tenga una visión unificada, confiable y gobernada de su información.

Es así el esquema ilustra las capas principales:

1. Fuentes de datos

ERP, CRM, SaaS, archivos, IoT.

2. Capa de integración

APIs, conectores y streaming de datos.

3. Núcleo Data Fabric

Metadatos, gobierno de datos y virtualización.

4. Almacenamiento y procesamiento

Data Lake y Data Warehouse.

5. Consumo de datos

BI, analítica, aplicaciones empresariales e IA.

Este artículo pretende dejar claro una de las tecnologías desde las cuales se soporta la arquitectura de los datos hoy en día, para que sea fuente principal en la interacción hacia soluciones más sofisticadas, donde su composición por capas tiene como propósito brindar el tratamiento para que sea confiable en la lectura y resultado hacia estrategias de negocio posibles y más amplias.

Referencias para mayor consulta:

Gartner, Inc. (2023). Top strategic technology trends. Gartner Research.

IBM. (2022). What is data fabric?. IBM Corporation. <https://www.ibm.com/topics/data-fabric>

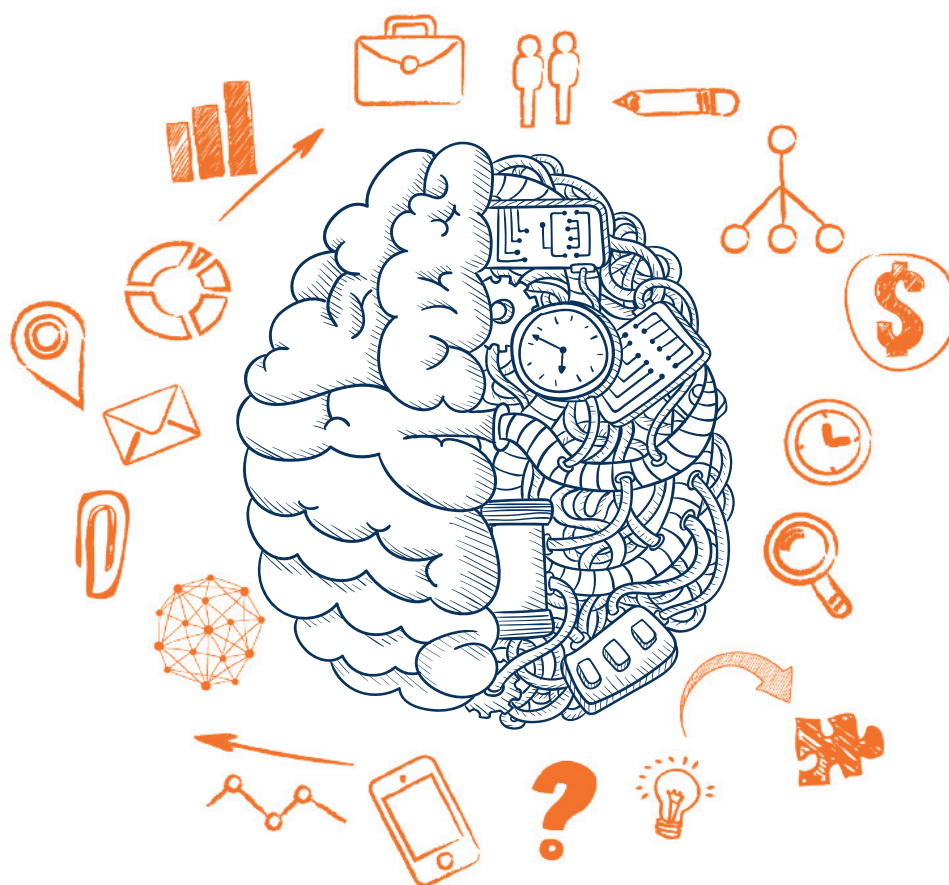


Más de

40 AÑOS

acompañando los Procesos de **expansión**
e **internacionalización** de la **Empresa**
Colombiana

ESTAMOS CONSTRUYENDO UNA
#menteBKF



Estamos en las redes sociales para compartir nuestro conocimiento y experiencia con toda la comunidad

Únete, participa, inspírate, comparte y alimenta tu #menteBKF