



Consumo y Serving de datos

Cómo los datos se convierten en valor

El arte del data serving

El **serving de datos** representa la diferencia fundamental entre simplemente almacenar información y convertirla en valor tangible para la organización.

No basta con tener datos bien estructurados en nuestro lakehouse o data warehouse si estos no llegan de manera efectiva a quienes los necesitan.

Mientras que el almacenamiento se enfoca en la persistencia y organización, el **serving** se concentra en **habilitar el uso práctico** de esos datos por diferentes usuarios y sistemas.



01

Business Intelligence (BI)

Dashboards y reportes para análisis visual

03

Reverse ETL

Integración con sistemas operacionales

02

APIs de Exposición

Acceso programático bajo demanda

04

GenBI con LLMs

Consultas en lenguaje natural

Conectividad universal: JDBC y ODBC

JDBC (Java Database Connectivity) y ODBC (Open Database Connectivity) son estándares fundamentales para la conectividad de datos. Actúan como adaptadores universales, permitiendo que una amplia gama de aplicaciones se conecte a diversas bases de datos sin depender de conectores específicos de cada proveedor. Esto abre el acceso a una gran variedad de herramientas de análisis y desarrollo.

Facilitan el acceso a datos para:

- **Gestores de bases de datos:** para consultas SQL puras y análisis ad-hoc.
- **Aplicaciones personalizadas:** integración directa en soluciones específicas.
- **Entornos de desarrollo:** creación de nuevas funcionalidades y microservicios.

En síntesis, JDBC y ODBC son la vía más común y neutral para acceder a datos, garantizando interoperabilidad y flexibilidad en cualquier ecosistema.

Business Intelligence: el consumo clásico

Las herramientas de BI como **Power BI** y **Metabase** representan la forma más tradicional y extendida de consumir datos. Sin embargo, su implementación efectiva requiere seguir principios arquitectónicos fundamentales para mantener la consistencia y calidad.



Principio clave

Nunca implementar lógica de negocio directamente en BI.

Las métricas complejas deben calcularse en la capa de transformación.



Mejores prácticas

Documentar exhaustivamente todas las métricas y dimensiones. Implementar pruebas de datos para garantizar su calidad y fiabilidad.

Ventajas del Enfoque

- Consistencia en definiciones de métricas
- Accesibilidad para usuarios no técnicos
- Mantenibilidad centralizada
- Versionado de lógica de negocio



Ejemplo práctico: El churn rate se calcula una vez en dbt usando SQL, y luego se consume desde Power BI, Metabase y cualquier otra herramienta, manteniendo coherencia total.

APIs de exposición: acceso programático

Las APIs representan el mecanismo más flexible para exponer datasets bajo demanda, permitiendo integración directa con aplicaciones, servicios web y herramientas de análisis programático. Su implementación requiere consideraciones específicas de arquitectura y seguridad.

1

Autenticación robusta

Implementar tokens JWT o API keys para controlar acceso granular a diferentes endpoints y datasets.

2

Paginación inteligente

Manejar grandes volúmenes de datos mediante limit/offset o cursor-based pagination para optimizar rendimiento.

3

Documentación completa

OpenAPI/Swagger para especificar endpoints, parámetros, respuestas y casos de uso esperados.

Reverse ETL: datos que impulsan acción

A diferencia de las APIs de exposición que esperan consultas, el **Reverse ETL** empuja proactivamente datos desde nuestro lakehouse hacia sistemas operacionales externos, cerrando el círculo entre análisis y acción empresarial.



Caso Real: Segmentación RFM → Mailchimp

Imaginemos que hemos segmentado a nuestros clientes por su comportamiento RFM (Recencia, Frecuencia, Valor Monetario) en nuestro lakehouse. En lugar de exportaciones manuales, el Reverse ETL:

1. Ejecuta automáticamente cada 24 horas
2. Extrae los clientes con su segmento RFM (ej. "Clientes VIP", "Clientes en Riesgo")
3. Los envía a Mailchimp (o similar) como audiencias o etiquetas dinámicas
4. Dispara campañas de marketing personalizadas según el segmento RFM

⚠ Diferencia clave: Los datos no solo informan decisiones, sino que **activan automáticamente** procesos de negocio.

Este enfoque transforma datos pasivos en inteligencia operacional activa, **maximizando el ROI** de nuestra infraestructura de datos.

La capa semántica: fuente única de verdad

La capa semántica actúa como **fuentes de verdad única** que garantiza consistencia absoluta entre BI, APIs, Reverse ETL y GenBI. Sin ella, cada canal de consumo podría interpretar los datos de manera diferente, creando confusión y decisiones erróneas.



Herramientas especializadas

- **Cube:** Modelado dimensional avanzado
- **MetricFlow (dbt):** Métricas versionadas
- **Looker**

✔ Una métrica definida una vez, consumida consistentemente por todos los canales.

GenBI y LLMs: la frontera del análisis

Los Large Language Models están revolucionando cómo los usuarios interactúan con datos, permitiendo consultas en lenguaje natural. Sin embargo, su efectividad depende críticamente de la calidad de metadata y semántica subyacente.

El poder del prompting

"¿Cuáles son los clientes VIP según su segmentación RFM?" se convierte automáticamente en SQL optimizado contra nuestro lakehouse.

Metadata como fundamento

El LLM necesita conocer esquemas, relaciones entre tablas, definiciones de métricas y contexto de negocio para generar consultas precisas.

El riesgo de la inconsistencia

Sin una capa semántica sólida, diferentes consultas pueden generar definiciones contradictorias de la misma métrica.

Ejemplo: GenBI sobre Segmentación RFM

Usuario: "Muéstrame cuántos clientes tenemos en cada segmento RFM."

GenBI: Interpreta automáticamente que necesita:

- Tabla de clientes con segmentación RFM
- Métrica de conteo de clientes
- Agrupación por segmento RFM (VIP, en Riesgo, etc.)
- Período temporal por defecto

Muéstrame cuántos clientes tenemos en cada segmento RFM.

Answer preparation steps 3 steps

Answer View SQL Chart Beta

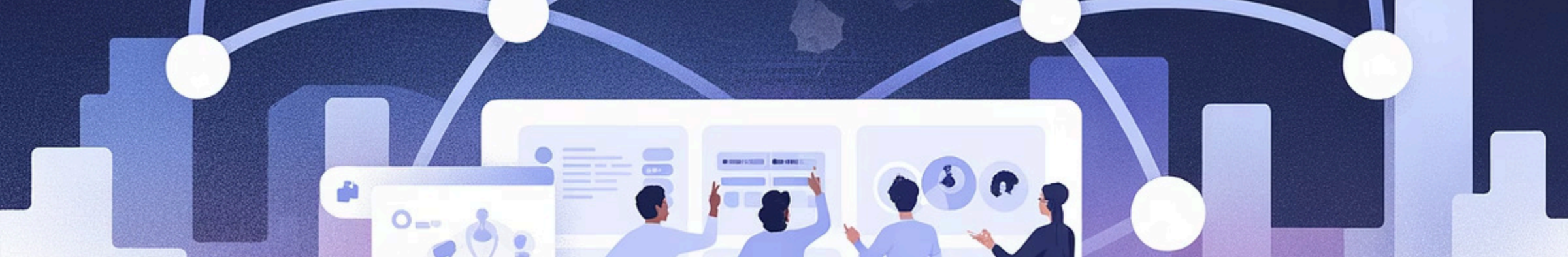
Adjust the answer

Claro, aquí tienes la cantidad de clientes en cada segmento RFM:

- Clientes Regulares: 748 clientes
- Clientes Dormidos: 624 clientes
- Nuevos Clientes: 49 clientes
- Clientes Potenciales: 418 clientes
- Necesitan Atención: 835 clientes
- En Riesgo: 773 clientes
- Clientes de Alto Valor: 23 clientes
- Clientes Leales: 420 clientes
- Campeones: 625 clientes

Hay más segmentos, pero estos son los principales.

⊗ **Atención crítica:** La calidad de las respuestas del GenBI es directamente proporcional a la riqueza de nuestra capa semántica y documentación de datos.



El serving es un ecosistema

El serving de datos no es una tecnología única, sino un **ecosistema coordinado** que permite diferentes formas de acceso para distintos perfiles de usuario, todos alimentándose de un mismo origen confiable.

Analistas de negocio

Consumen a través de dashboards interactivos y reportes automatizados

Desarrolladores

Integran mediante APIs REST/GraphQL en aplicaciones y servicios

Sistemas operacionales

Reciben datos automáticamente vía Reverse ETL para acción inmediata

Usuarios finales

Exploran datos conversacionalmente a través de interfaces GenBI

El valor real de cualquier pipeline de datos se mide en su capacidad de uso, no en su sofisticación técnica.

Los datos que no se consumen efectivamente son datos desperdiciados, independientemente de cuán perfectamente estén modelados o almacenados.