



# ANÁLISIS DE DATOS Y DATAWAREHOUSE

---

*Ing. Jairo Salazar*

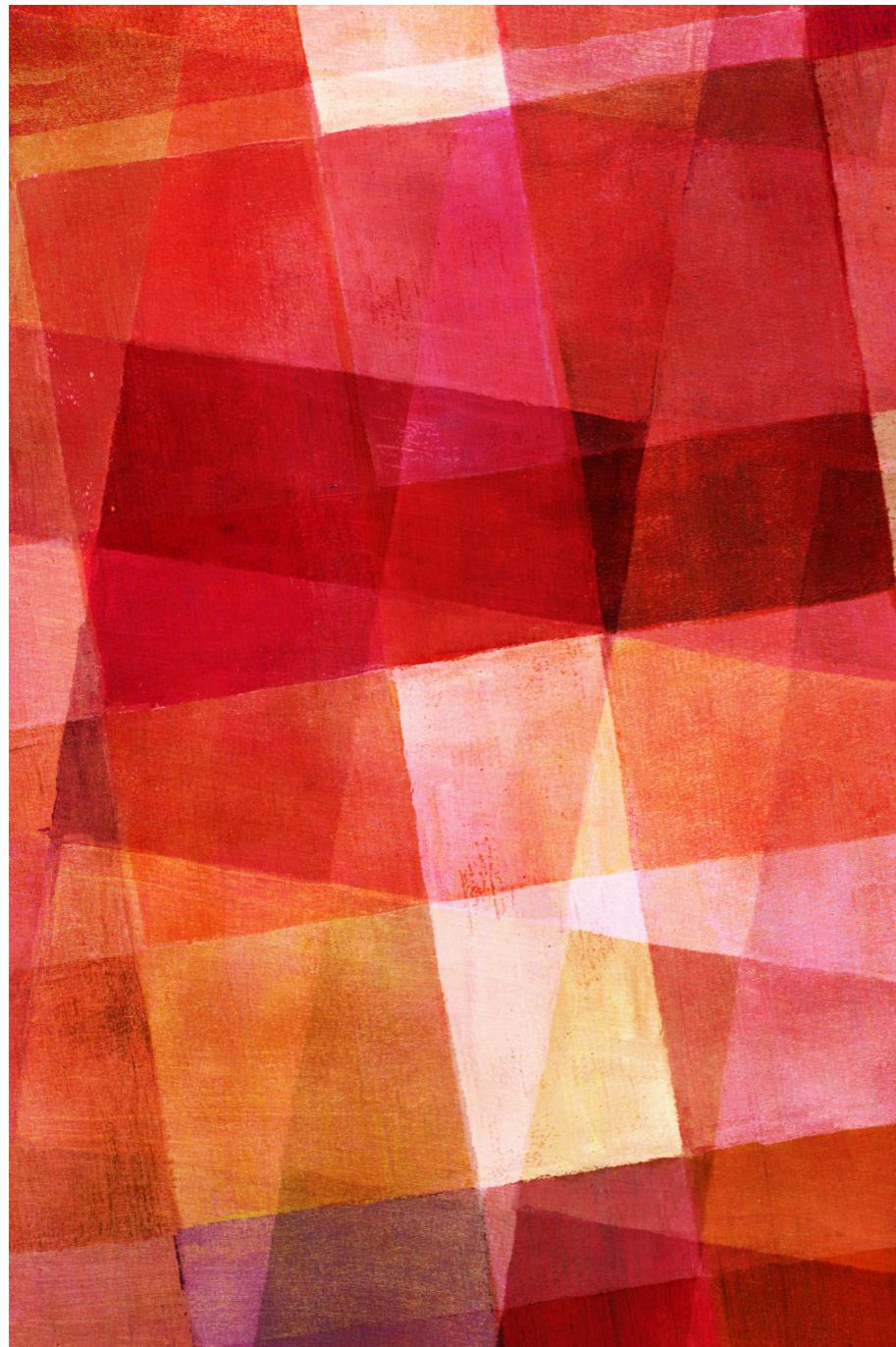
# AGENDA DE LA SEMANA

---

- **LUNES**
- Análisis de Datos
- Flujo de datos
- DW/BI
- **MIERCOLES**
- Diagrama Entidad/Relacion
- Docker

# ANÁLISIS DE DATOS

---



Las empresas generan datos constantemente. Estos pueden variar dependiendo de la industria, el modelo de negocio y los productos o servicios que cada corporación ofrece.

Ahora bien, no todas las empresas aprovechan sus datos al máximo, debido a la ausencia de un proceso de análisis de datos. Este último consiste, de acuerdo con la definición de Intel en: “*un proceso mediante el cual los datos sin procesar se convierten en conocimientos útiles con los que se podría actuar*”, es decir, tomar decisiones comerciales, estratégicas, financieras y productivas.

# FLUJO DE LOS DATOS

---

Para proceder con un análisis de datos, en primer lugar debemos identificar cada una de estas etapas en nuestros procesos empresariales, de manera que al final de este análisis tengamos los elementos necesarios para crear un mapa del flujo de datos. Estas etapas son las siguientes:

## Entrada de Datos

Por esto nos referimos a los puntos de captación de datos. ¿Cuántos hay en la empresa? ¿Cuáles son? ¿Están funcionando correctamente? ¿Qué información reúnen? Estas son algunas preguntas que deberías responder por lo que respecta a este importantísimo aspecto del flujo de datos, ya que es, por decirlo así, la alimentación de todo lo que haremos más adelante.

“En el caso de la mayoría de los datos no estructurados, la entrada constituye tanto el principio como el final del ciclo de vida de los datos”.

A esta información se le conoce como «datos oscuros», y aunque se introduce en las bases de datos rara vez se analiza y aprovecha; por lo que uno de los objetivos de crear un mapa de flujo de información es identificar qué datos podrían pertenecer a esta categoría.

# FLUJO DE DATOS

---

- Preparación de los datos
  - Una vez que han reconocido las fuentes de información, las empresas deben de realizar una etapa de limpieza y homologación que garantice la calidad de los datos. Estos datos deben ser depurados y transformados en información útil para que sean consultados fácilmente por los diferentes equipos de la empresa.
  - “Los datos pueden ser almacenados y clasificados según el grado de urgencia y utilidad, para que se pueda acceder a los datos más críticos lo más rápido posible”.
  - Análisis
  - Se refiere a las actividades enfocadas en procesar la información, de manera continua o por bloques. Para sacar el mayor provecho de ella en las diversas tomas de decisiones de la empresa.
-

# DECISIONES BASADAS EN DATOS

---

- Las decisiones comerciales y estratégicas de una compañía, hoy en día, no solo dependen de la experiencia de los gerentes, directores o inversionistas. Deben estar sustentadas en la realidad. Y para ello, los datos y su respectivo análisis son una fuente indispensable para la toma de mejores decisiones.
- “Una vez que los datos hayan sido introducidos, preparados y analizados, estarán listos para que se pueda actuar con ellos. La visualización de datos y la elaboración de informes ayudan a comunicar los resultados del análisis”.

WILEY

KIMBALL  
GROUP

# The Data Warehouse Toolkit

Third Edition

The Definitive Guide  
to Dimensional  
Modeling

Ralph Kimball  
Margy Ross



“

This asset is almost always used for two purposes: operational record keeping and analytical decision making.

*Ralph Kimball*

# SISTEMA DE DATAWAREHOUSE

---

Es una arquitectura de almacenamiento en donde los datos son organizados, limpiados, y estandarizados.

## **Integrado**

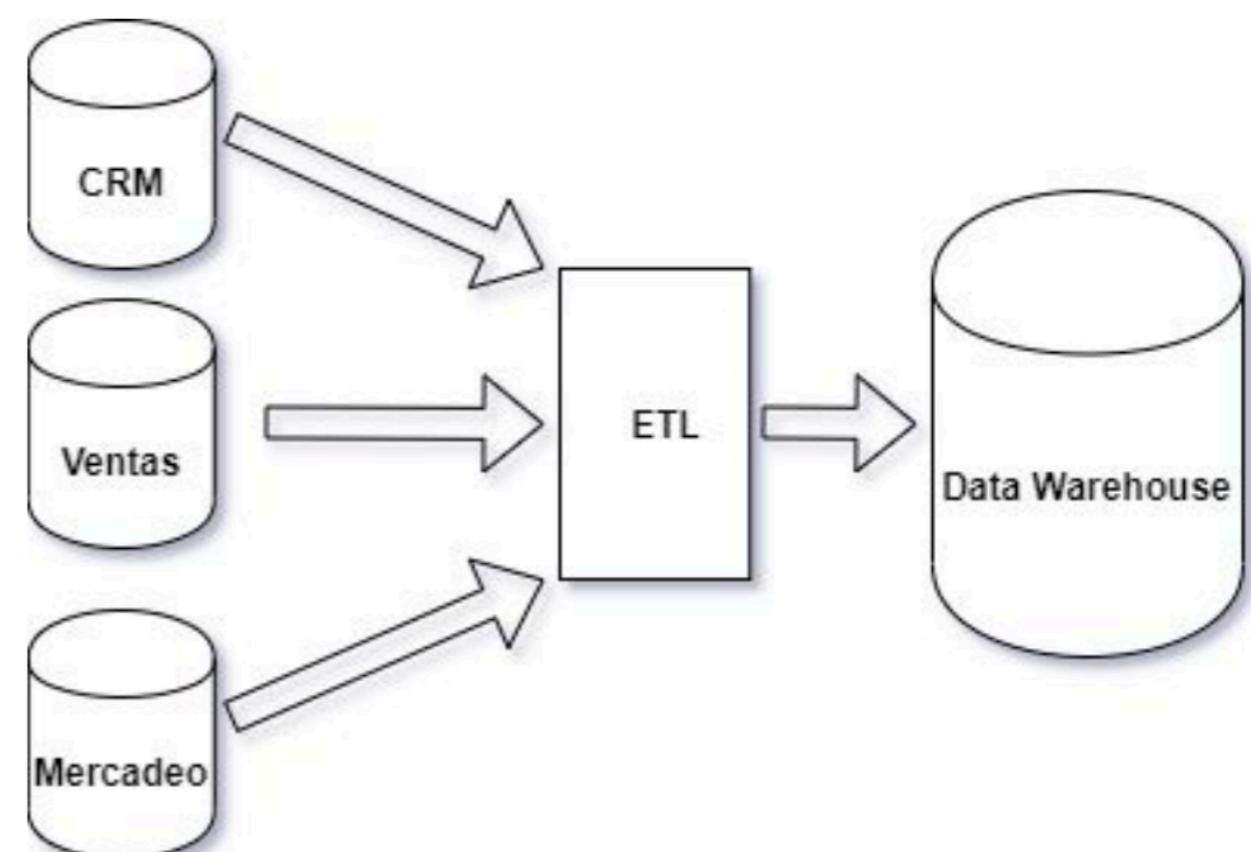
Fuentes de datos heterogéneas  
Limpieza de datos y conversión

## **Enfocado**

Ventas, Mercadeo, Recursos Humanos, etc.

## **Variaciones en el Tiempo**

Timestamps  
Datos históricos



# RAZONES PARA LA EXISTENCIA DE UN DATAWAREHOUSE

---

- Una sola versión de la verdad
- Desempeño

Simplicidad

Persistencia de datos

# CHECKLIST PARA PODER TENER UN DATAWAREHOUSE

---

Hay un grupo de personas utilizando el mismo dataset

Necesitamos tener datos limpios e integrados para la compañía

Necesitamos separar la estructura de datos de las fuentes transaccionales cambiantes

No queremos repetirnos



# BUSINESS INTELLIGENCE - BI

---

- Es el proceso de revelar insights esenciales del negocio al ejecutar modelos, métodos y algoritmos en el Data Warehouse para identificar patrones y similitudes en los datos con la finalidad de facilitar la toma de decisiones.
- **Convertir Datos en Información Accionable**

# OBJETIVOS DEL DW/BI

---

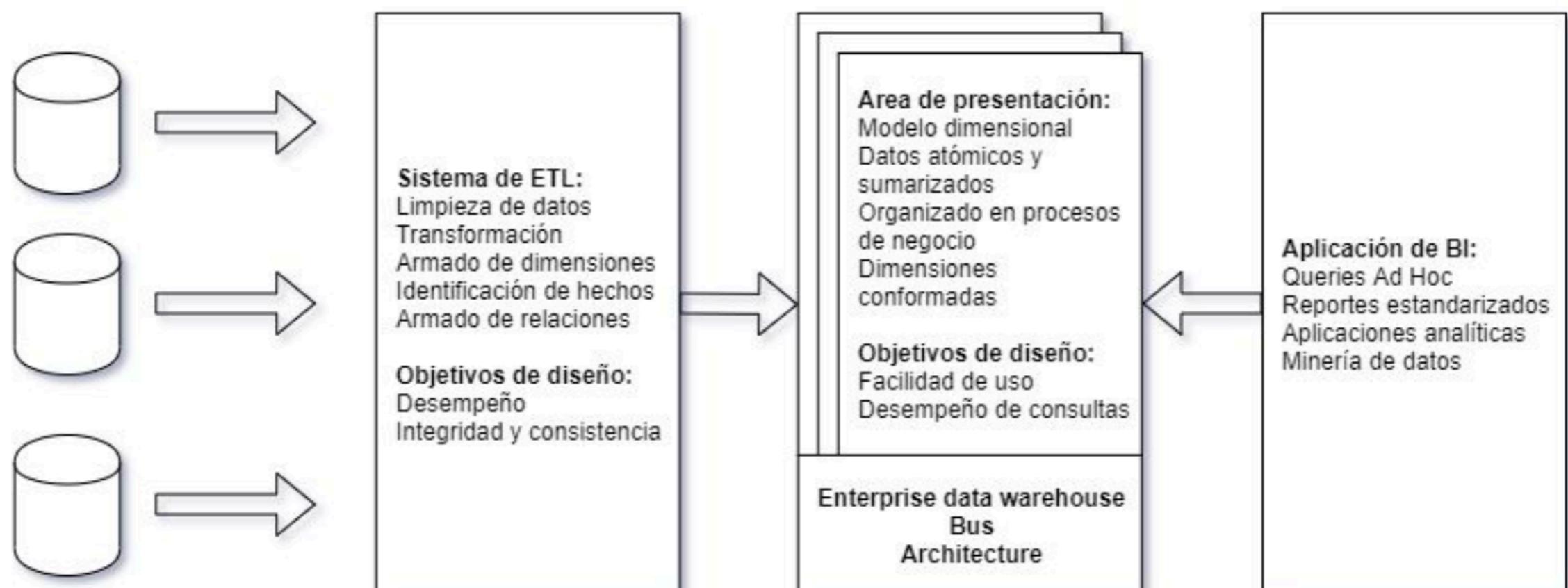
- El sistema debe de hacer accesible la información. **Simple y Rápido.**
- El sistema DW/BI debe de presentar la información consistentemente. **Una sola versión de la verdad.**
- El sistema DW/BI debe de adaptarse al cambio. **Robusto y Escalable.**
- El sistema DW/BI debe de presentar la información a tiempo. **Manejo de expectativas y compromiso.**
- El sistema DW/BI debe de proteger sus activos de información. **Seguridad.**
- El sistema DW/BI deve de ser la fuente mas confiable y definitiva de información para la toma de decisiones. **Sistema de soporte de decisiones.**

# OBJETIVOS DEL DW/BI

---

- El sistema DW/BI debe de ser aceptado para ser considerado exitoso.
- Debe de ser la fuente de información accionable, simple rápida y confiable.

Fuentes transaccionales ↔ Area de staging ↔ Area de Presentación ↔



**Arquitectura Kimball DW/BI**

# COMO ARQUITECTOS DE UN DATAWAREHOUSE/BI DEBEMOS

---

Crear la mejor solución técnica posible utilizando la tecnología adecuada.

Ganarnos la **confianza del patrocinador y usuarios finales**.

Suplir la demanda de información con modelos de Inteligencia de Negocios.

Convertirnos en expertos del negocio.

➤