

#CrecimientoProfesional  
#AprendeConLosPioneros

Online |  Perú



CAPACITACIÓN  
PROFESIONAL

# SESIÓN I

Docente: Victor Gutierrez  
Data Architect

## REGLAS



Se requiere **puntualidad** para un mejor desarrollo del curso.



Para una mayor concentración **mantener silenciado el micrófono** durante la sesión.



Las preguntas se realizarán **a través del chat** y en caso de que lo requieran **podrán activar el micrófono**.



Realizar las actividades y/o tareas encomendadas en **los plazos determinados**.



**Identificarse** en la sala Zoom con el primer nombre y primer apellido.

# Agenda

- Análisis del nuevo entorno de negocios.
- La importancia de la gestión de la información para la toma de decisiones empresariales.
- Necesidades de información en los diversos niveles de las organizaciones.
- BI Tradicional vs BI Moderno.
- Las empresas y la necesidad de contar con estrategias de gestión y almacenamiento de datos.
- Revisión de casos de éxito de implementación de soluciones de Inteligencia de Negocios en empresas.

## Análisis del nuevo entorno de negocios.

El Análisis de Negocios es la práctica de desarrollar cambios en una empresa mediante la definición de necesidades (oportunidades y / o problemas) y la recomendación de soluciones que aporten valor

- Wal-Mart usa grandes cantidades de datos y análisis para dominar la industria.
- Amazon se basa en la inteligencia de negocios para llevar a cabo cambios comerciales



# La importancia de la gestión de la información para la toma de decisiones empresariales.

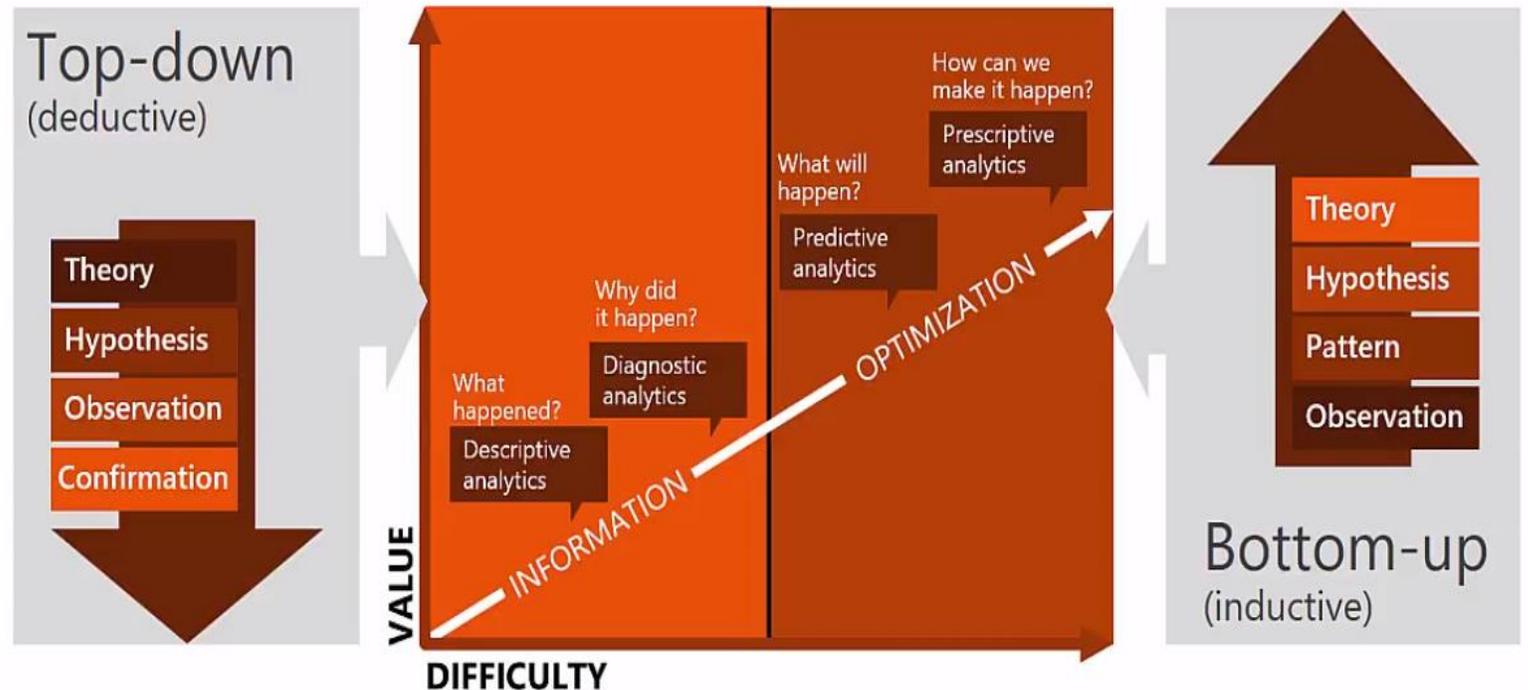
- Acopiar y obtener información de nuestros sistemas o fuentes externas
- Identificar las métricas o indicadores que permitan evaluar las operaciones de negocio
- Presentar evaluar y corregir los cambios en los procesos.



# La importancia de la gestión de la información para la toma de decisiones empresariales.

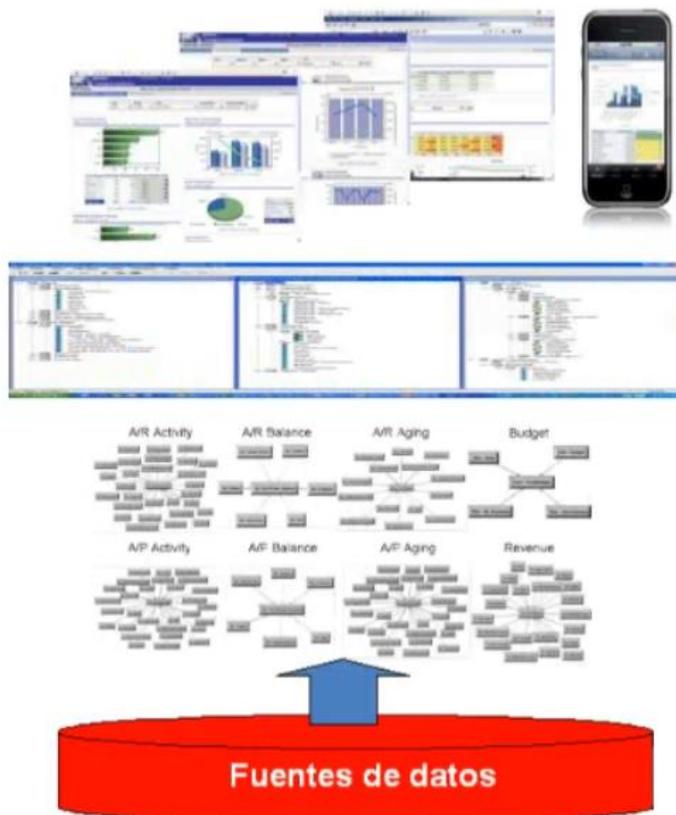
- La toma de decisiones en la empresa se fundamenta en el *aprovechamiento del conocimiento*
- Las tecnologías de información resultan ser una herramienta muy valiosa que permite recolectar, procesar y almacenar datos de la misma operación del negocio en el día a día

Two approaches to information management for analytics: Top-down and bottom-up



# Necesidades de información en los diversos niveles de las organizaciones.

- 4 Librería de **Cuadros de mando, Inform** y **alertas** para cubrir los requerimientos todas las áreas de negocio.
- 3 **Metadato** que incluye cálculos y métricas de la gestión de la empresa. **Disponible para consultas Ad-hoc.**
- 2 **Data Warehouse** con esquemas en estrella o copo de nieve, que contienen información agregada y significativa.
- 1 **ETL** que extrae los datos de las tablas operacionales de distintas fuentes, y las carga en el esquema de DW



## BI Tradicional vs BI Moderno.

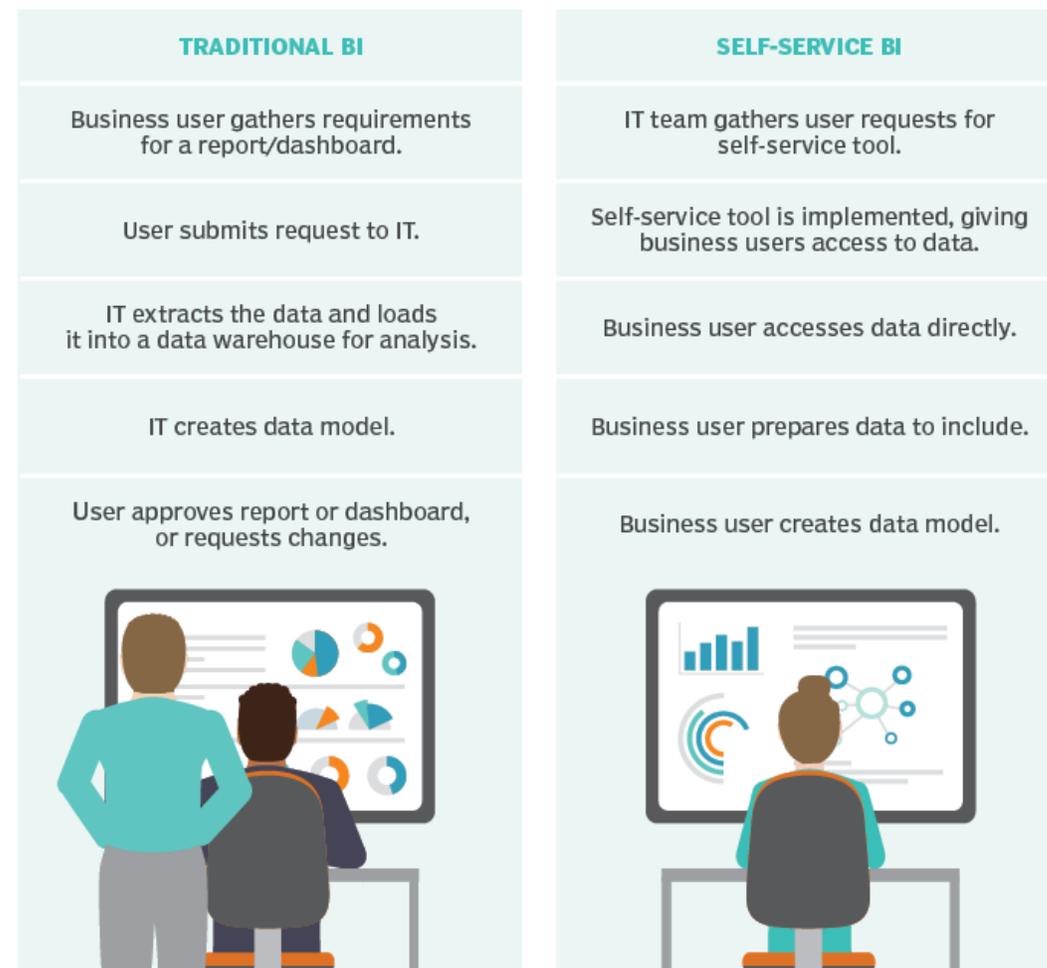
### Disponibilidad de los datos

Los usuarios de negocio requieren acceso inmediato para combinarlos con fuentes externas; para llegar a este objetivos se presentan dos escenarios:

**Desarrollo Experto:** Participación de un usuario experto en la aplicación; se utilizan formatos, diseños y gobierno de datos.

**Auto Servicio:** El mismo usuario dispone de acceso a los datos tanto de aplicación y externos para realizar sus propias combinaciones.

## Traditional BI vs. self-service BI



# Las empresas y la necesidad de contar con estrategias de gestión y almacenamiento de datos.

## ¿Qué son los datos oscuros?

Los datos oscuros son todos los datos desconocidos y sin explotar en toda la organización, generados por los sistemas, los dispositivos y las interacciones.

*El cincuenta y cinco por ciento de los datos de una organización son "oscuros": sin cuantificar y sin explotar - de acuerdo con el nuevo global investigación de TRUE Global Intelligence, patrocinada por Splunk Los 1.300 líderes empresariales encuestados reconocen esos datos son clave para el éxito, ahora y en el futuro, aún muy pocos dicen que sus organizaciones pueden aprovechar con éxito el valor de todos sus datos. O encuéntralo*

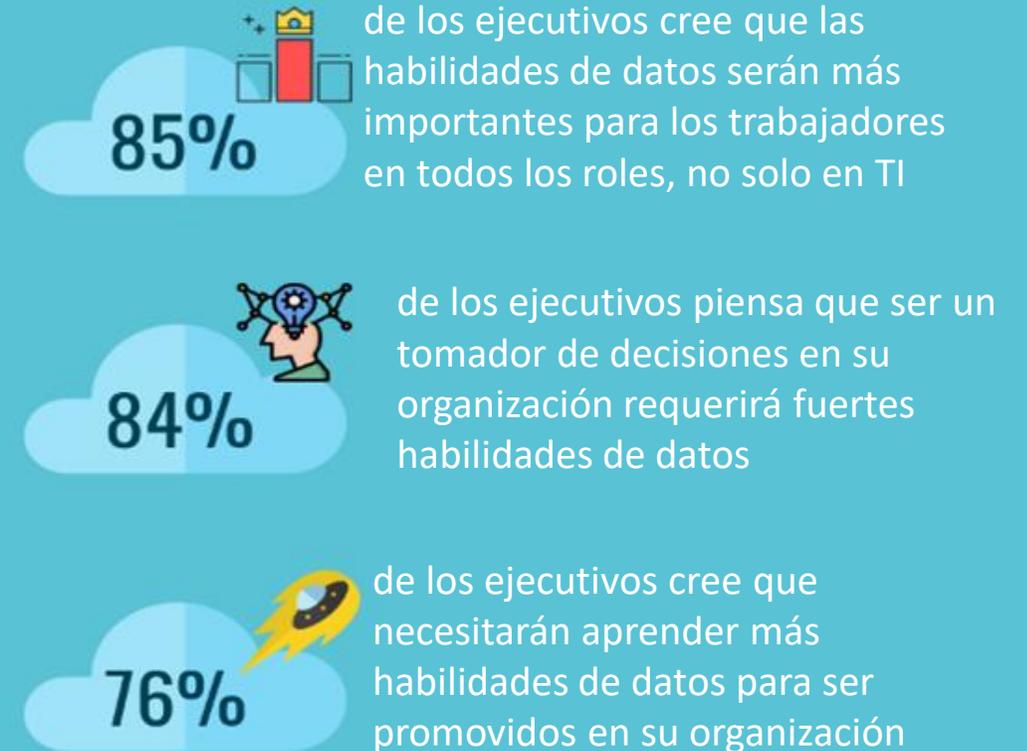


## Las empresas y la necesidad de contar con estrategias de gestión y almacenamiento de datos.

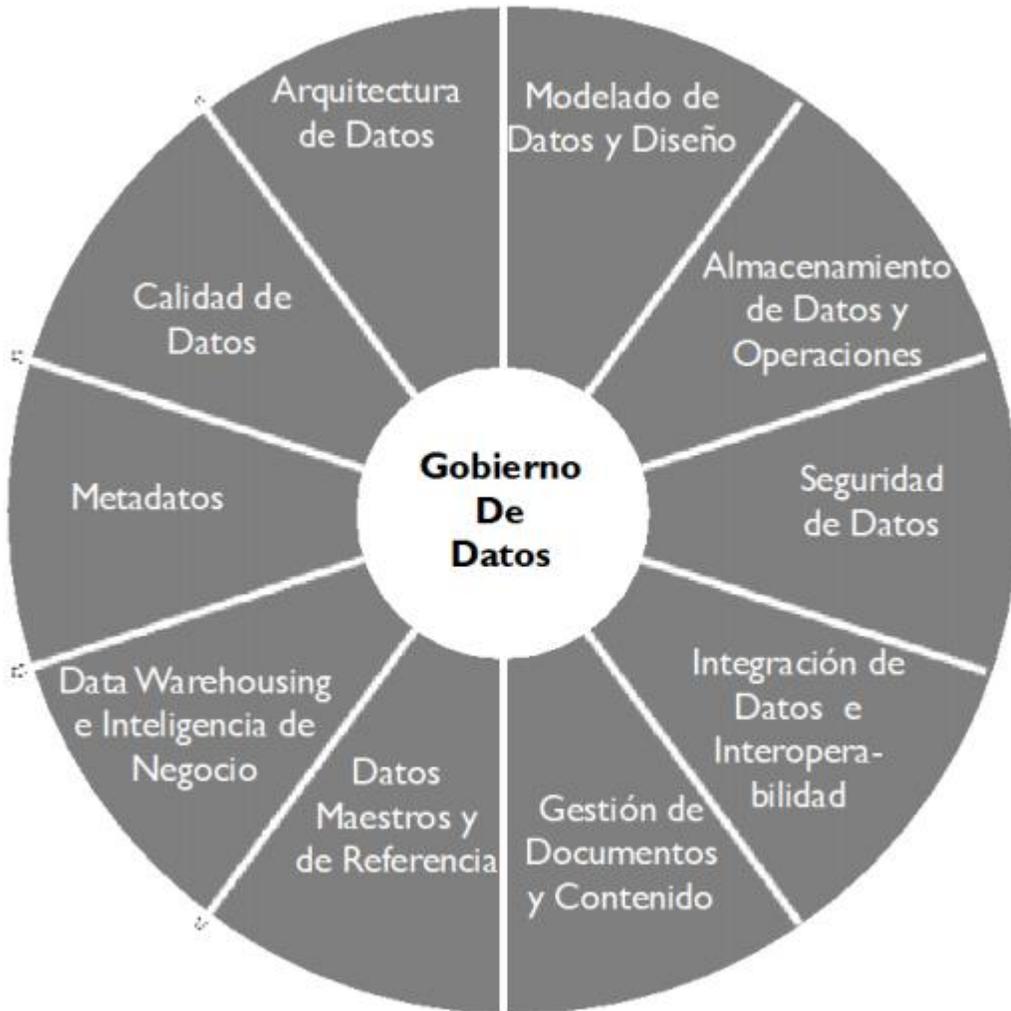
### Lo que impide que las empresas recuperen datos oscuros



### Los datos son la clave del éxito corporativo y profesional...

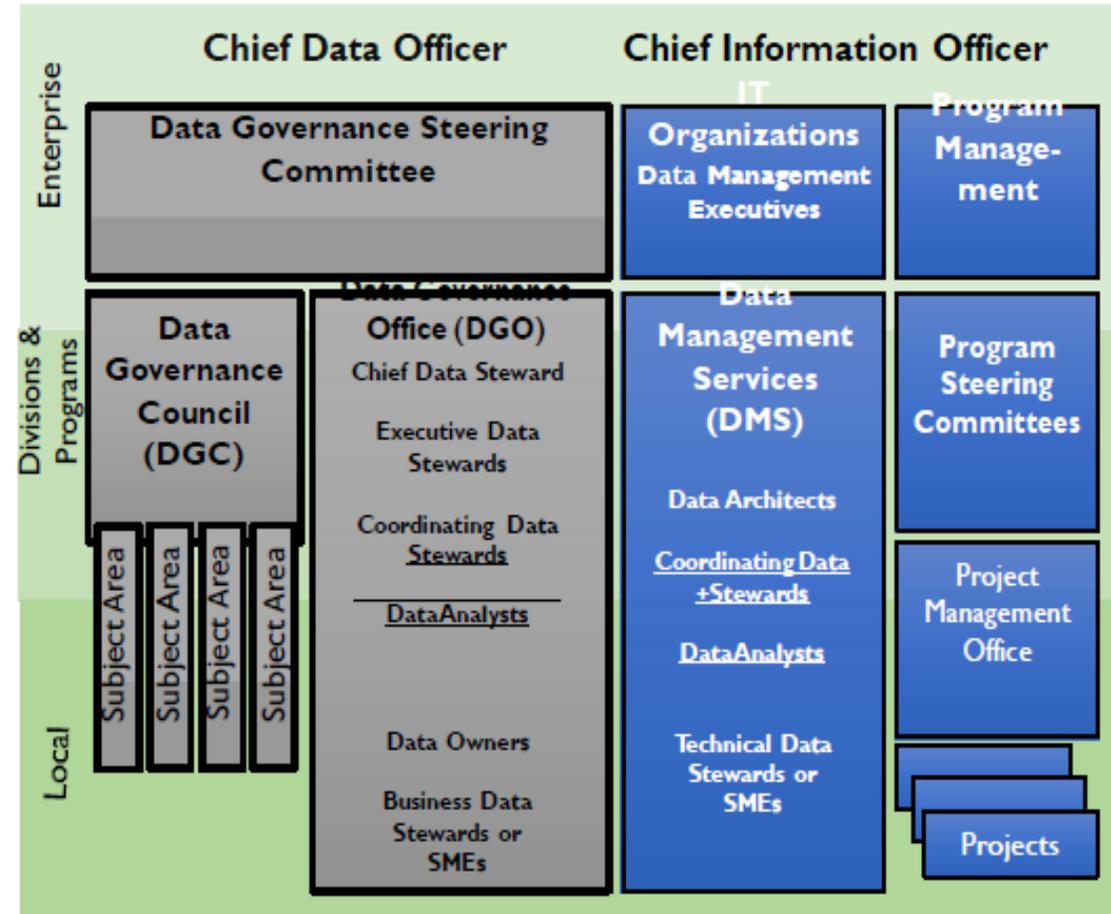


# Las empresas y la necesidad de contar con estrategias de gestión y almacenamiento de datos.

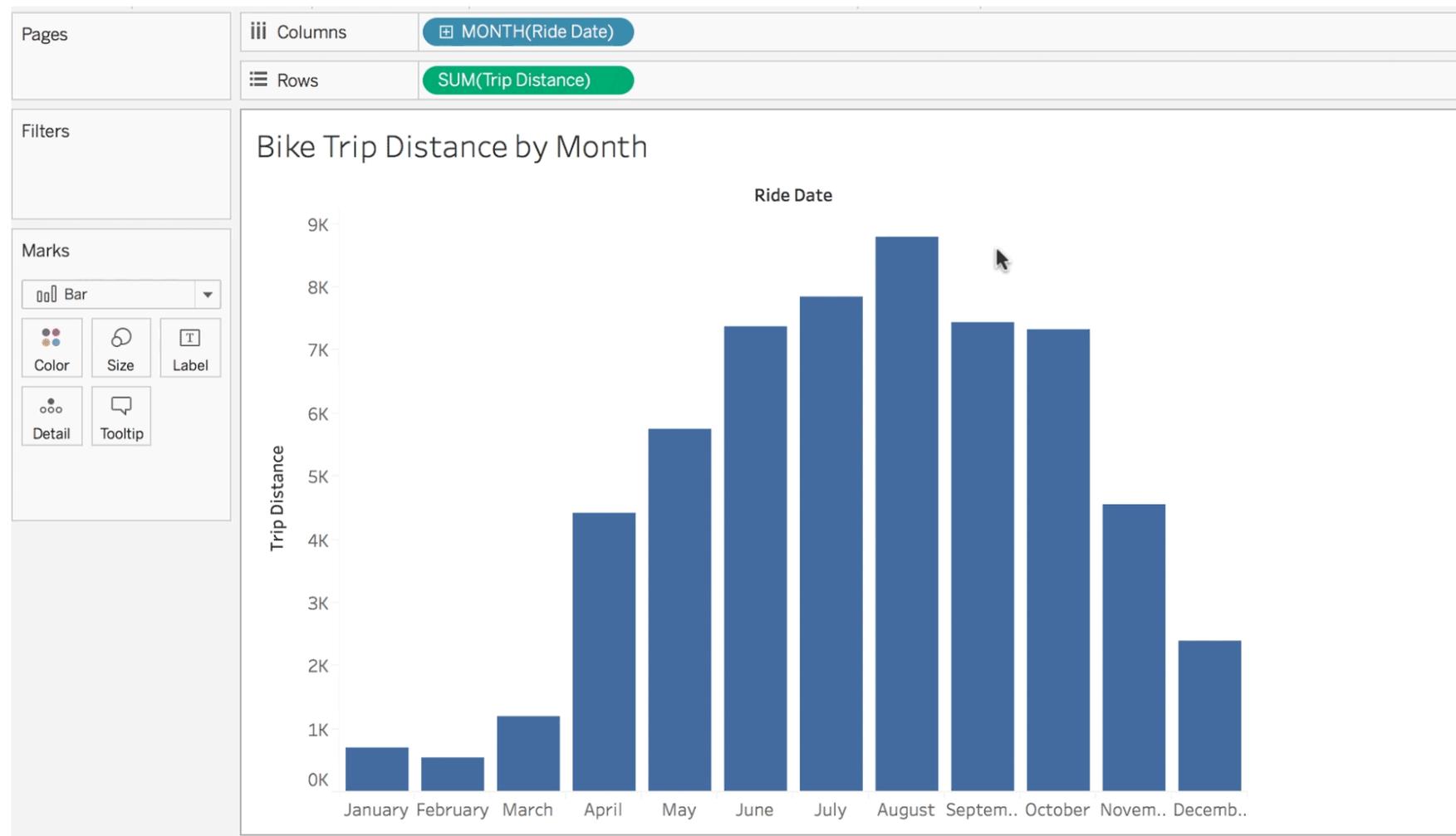


## Roles y Funciones / Data Governance Office DGO

Todo es propiedad de la organización de gestión de datos.



# Las empresas y la necesidad de contar con estrategias de gestión y almacenamiento de datos.



## Revisión de casos de éxito de implementación de soluciones de Inteligencia de Negocios en empresas.

- Reducción de fraudes y agilización de precios dinámicos
- Procesamiento de data y pronósticos de ventas
- Gestión de inventario
- Previsión de ventas
- Marketing personalizado

# Agenda

- Definición y arquitectura de bases de datos.
- Modelo lógico conceptual.
- Metodología Business Intelligence.
- OLTP vs OLAP.
- Modelamiento multidimensional.
- Introducción al Data Ware House (DWH).
- Desarrollo de la metodología para la construcción de un DWH.
- Introducción a los datalakes.
- Revisión de los principales softwares de manejo de bases de datos y de sistemas de DWH.

# Definición y arquitectura de bases de datos.

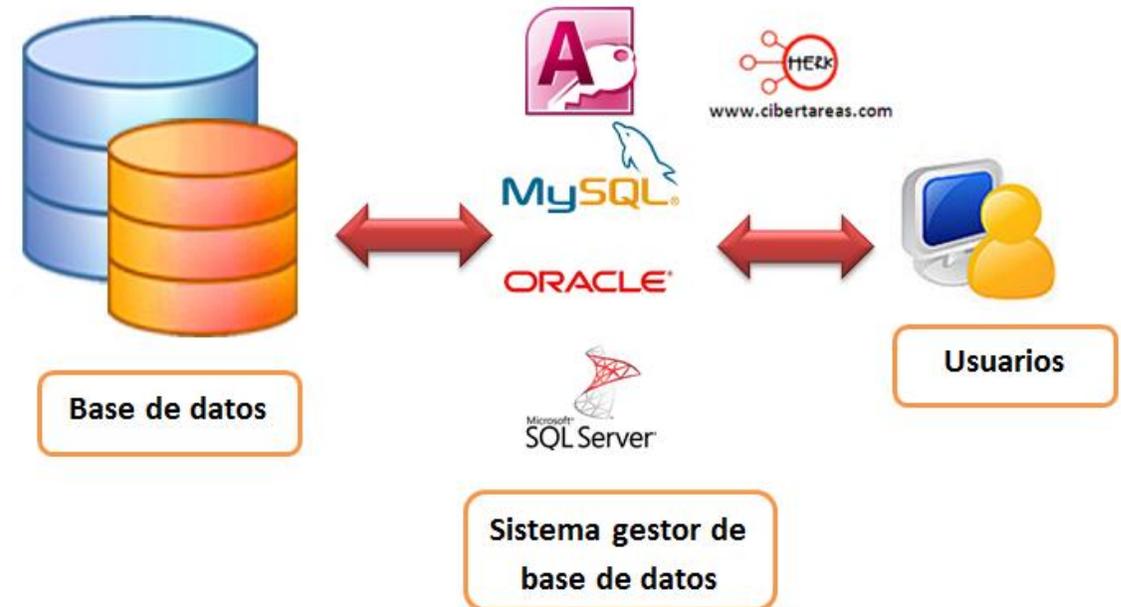
## Definición de base de datos

Se define una base de datos como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.

## Características

Entre las principales características de los sistemas de base de datos podemos mencionar:

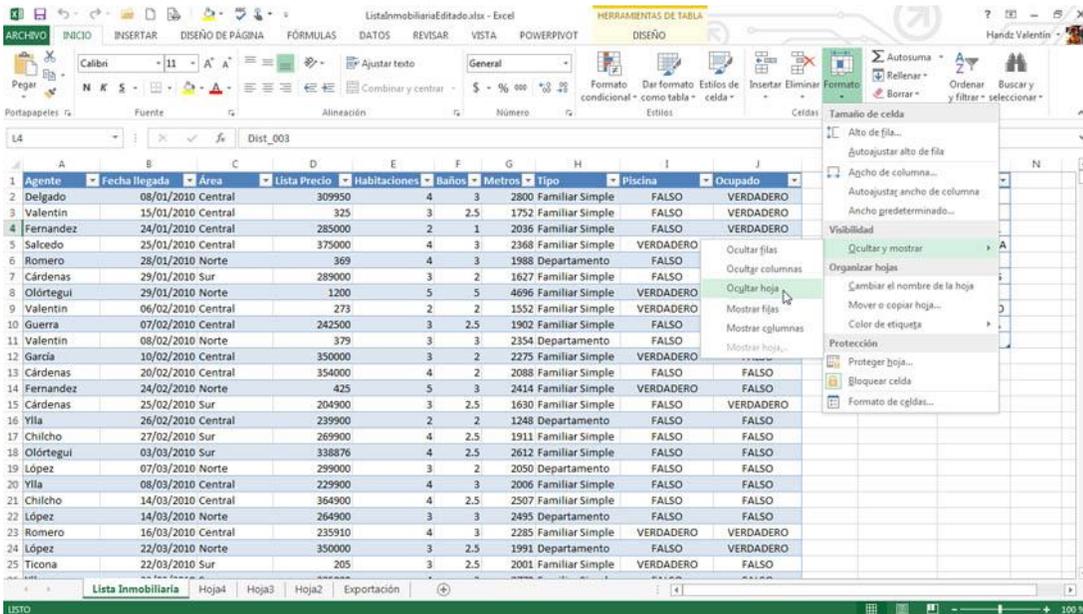
- Acceso concurrente por parte de múltiples usuarios.
- Integridad de los datos.
- Consultas complejas optimizadas.
- Seguridad de acceso y auditoría.
- Acceso a través de lenguajes de programación estándar (Transact-SQL)



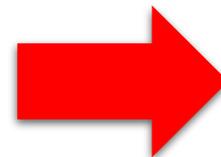
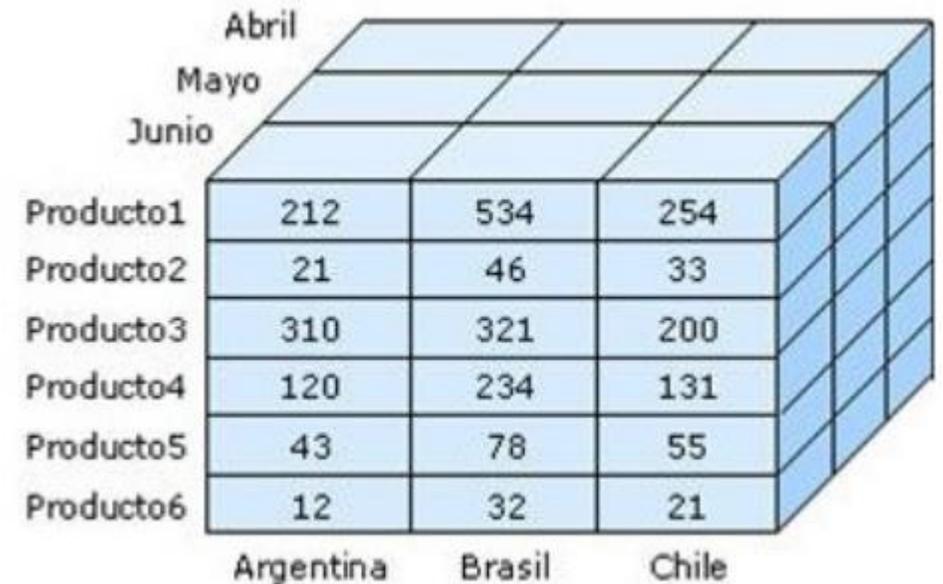
# Definición y arquitectura de bases de datos.

Una base de datos, a fin de ordenar la información de manera lógica, posee un orden que debe ser cumplido para acceder a la información de manera coherente. Cada base de datos contiene una o más tablas, que cumplen la función de contener los campos.

*Como podemos responder diferentes preguntas de lo que esta pasando:  
¿ Cuanto vendimos?, ¿ Que vendimos ?, ¿ A quienes le vendimos?*

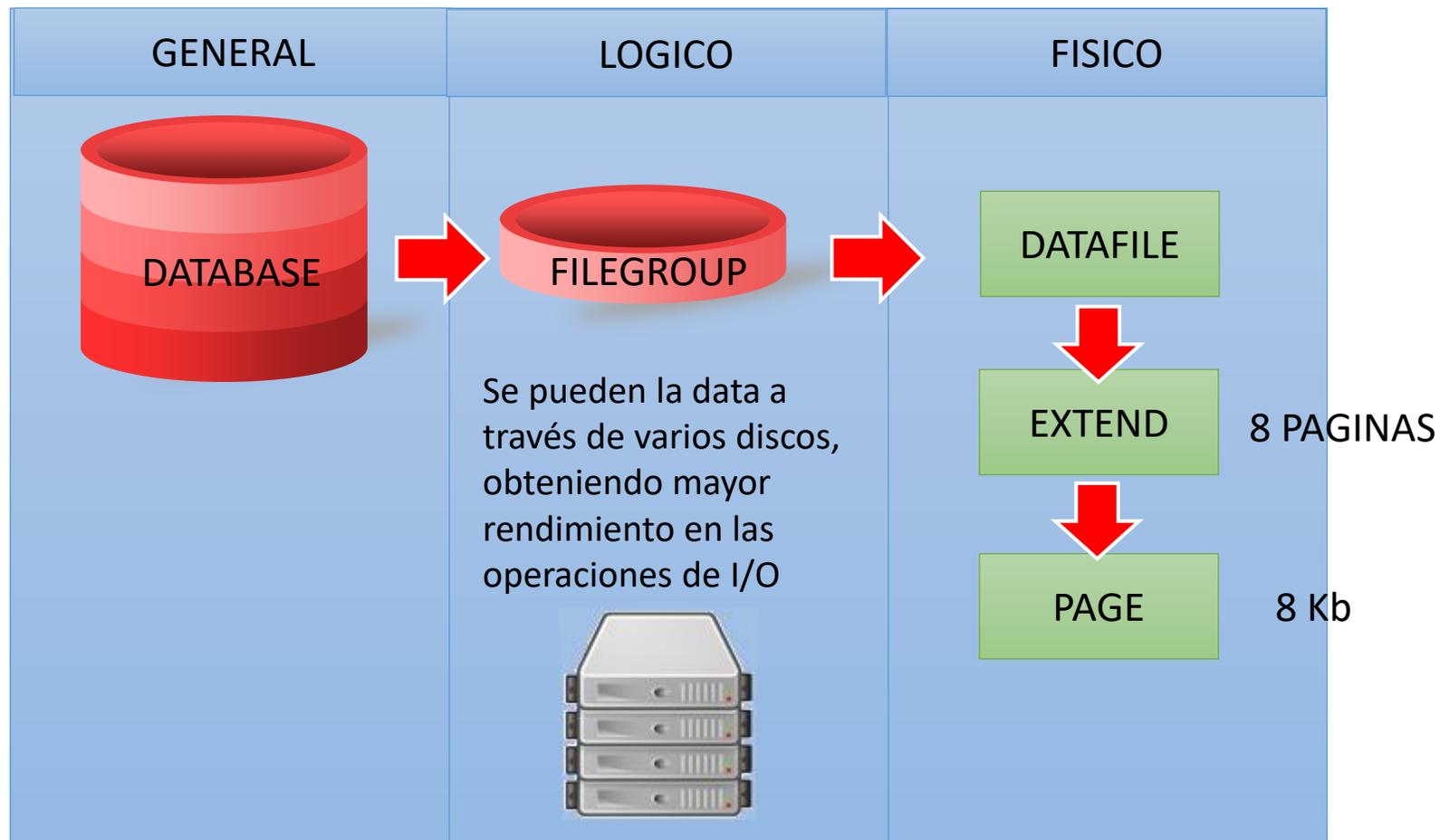


Agente	Fecha llegada	Área	Lista Precio	Habitaciones	Baños	Metros	Tipo	Piscina	Ocupado
Delgado	08/01/2010	Central	309950	4	3	2800	Familiar Simple	FALSO	VERDADERO
Valentin	15/01/2010	Central	325	3	2.5	1752	Familiar Simple	FALSO	VERDADERO
Fernandez	24/01/2010	Central	285000	2	1	2036	Familiar Simple	FALSO	VERDADERO
Salcedo	25/01/2010	Central	375000	4	3	2368	Familiar Simple	VERDADERO	
Romero	28/01/2010	Norte	369	4	3	1988	Departamento	FALSO	
Cárdenas	29/01/2010	Sur	289000	3	2	1627	Familiar Simple	FALSO	
Olórtegu	29/01/2010	Norte	1200	5	5	4696	Familiar Simple	VERDADERO	
Valentin	06/02/2010	Central	273	2	2	1552	Familiar Simple	VERDADERO	
Guerra	07/02/2010	Central	242500	3	2.5	1902	Familiar Simple	FALSO	
Valentin	08/02/2010	Norte	379	3	3	2354	Departamento	FALSO	
García	10/02/2010	Central	350000	3	2	2275	Familiar Simple	VERDADERO	
Cárdenas	20/02/2010	Central	354000	4	2	2088	Familiar Simple	FALSO	FALSO
Fernandez	24/02/2010	Norte	425	5	3	2414	Familiar Simple	VERDADERO	FALSO
Cárdenas	25/02/2010	Sur	204900	3	2.5	1630	Familiar Simple	FALSO	VERDADERO
Ylla	26/02/2010	Central	239900	2	2	1248	Departamento	FALSO	
Chilcho	27/02/2010	Sur	269900	4	2.5	1911	Familiar Simple	FALSO	FALSO
Olórtegu	03/03/2010	Sur	338876	4	2.5	2612	Familiar Simple	FALSO	FALSO
López	07/03/2010	Norte	299000	3	2	2050	Departamento	FALSO	FALSO
Ylla	08/03/2010	Central	229900	4	3	2006	Familiar Simple	FALSO	FALSO
Chilcho	14/03/2010	Central	364900	4	2.5	2507	Familiar Simple	FALSO	FALSO
López	14/03/2010	Norte	264900	3	3	2485	Departamento	FALSO	FALSO
Romero	16/03/2010	Central	235910	4	3	2285	Familiar Simple	VERDADERO	VERDADERO
López	22/03/2010	Norte	350000	3	2.5	1991	Departamento	FALSO	VERDADERO
Ticona	22/03/2010	Sur	205	3	2.5	2001	Familiar Simple	VERDADERO	FALSO

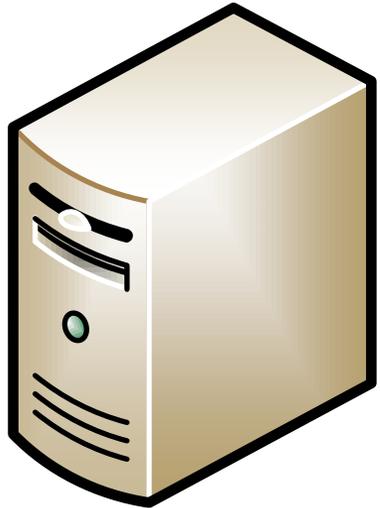
	Abril	Mayo	Junio
Producto1	212	534	254
Producto2	21	46	33
Producto3	310	321	200
Producto4	120	234	131
Producto5	43	78	55
Producto6	12	32	21
	Argentina	Brasil	Chile

# Definición y arquitectura de bases de datos.



# Definición y arquitectura de bases de datos.

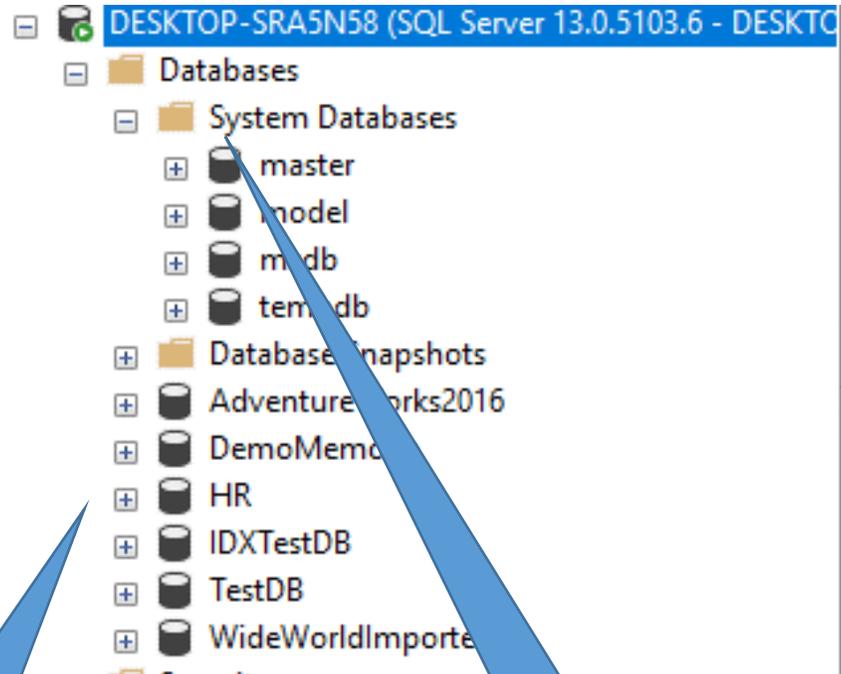
Servidor Host



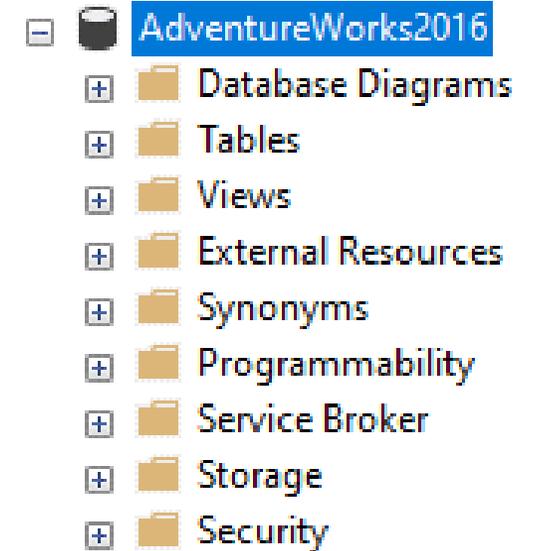
Servidor SQL Server



Base de datos



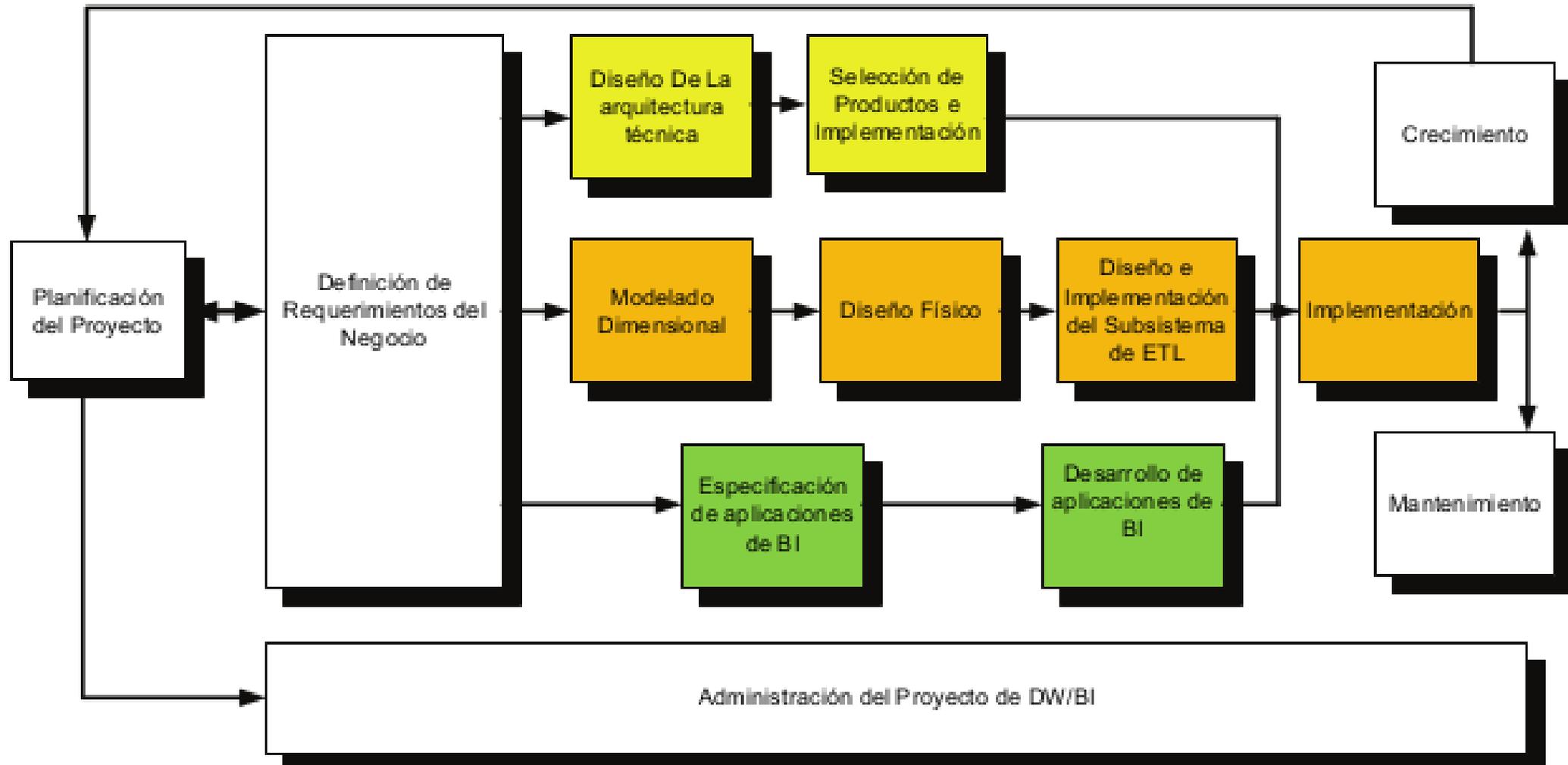
Organización de objetos Base de datos



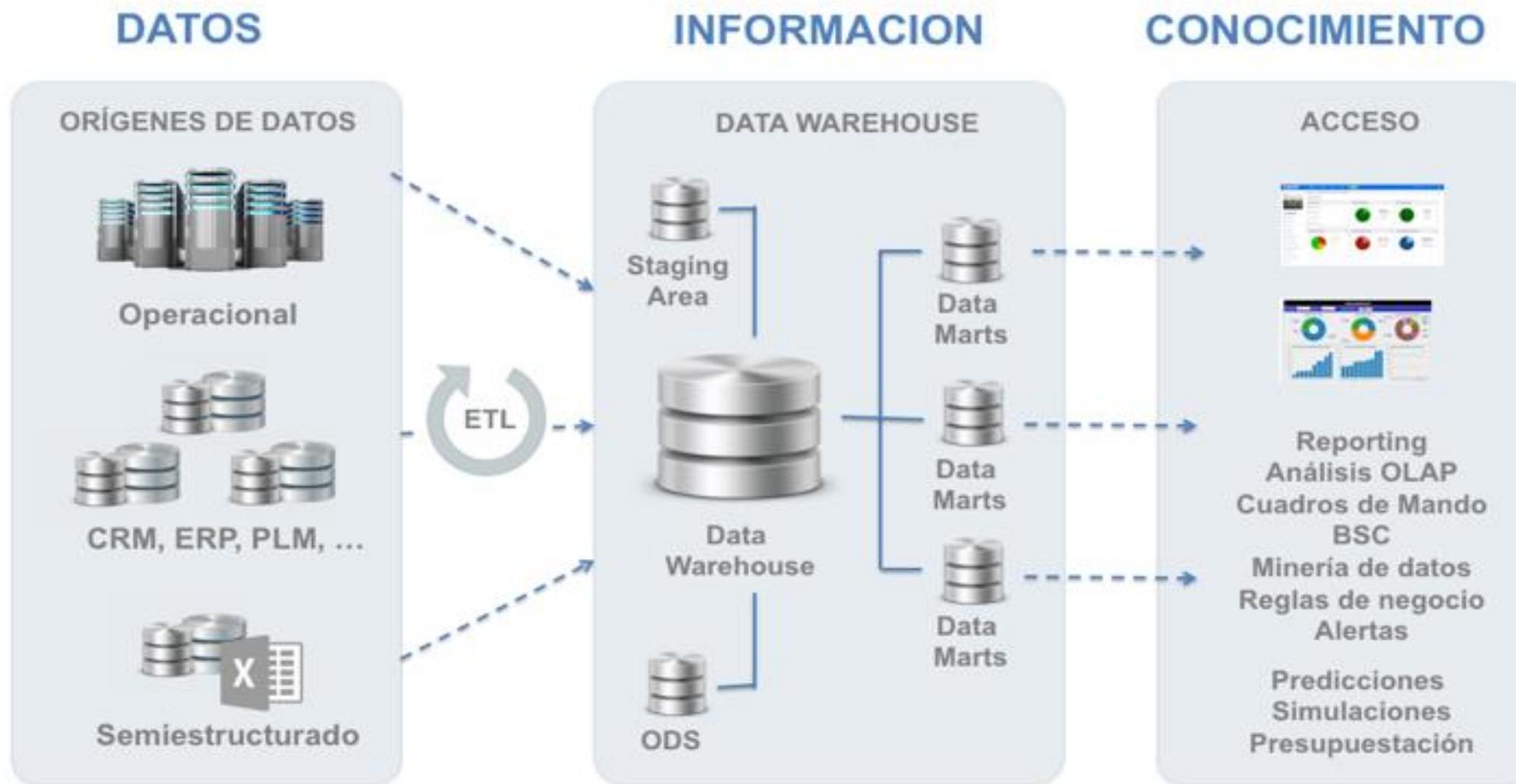
Base de datos de usuarios

Base de datos del sistema

# Metodología Business Intelligence.



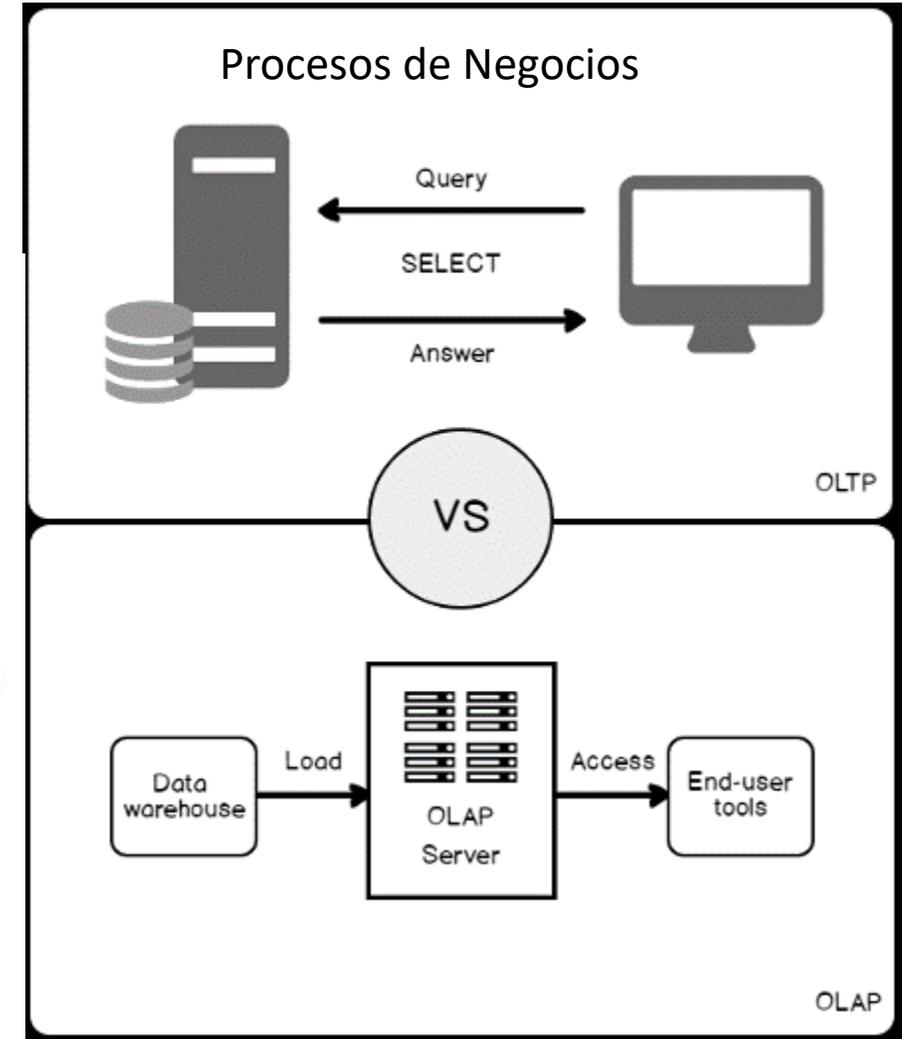
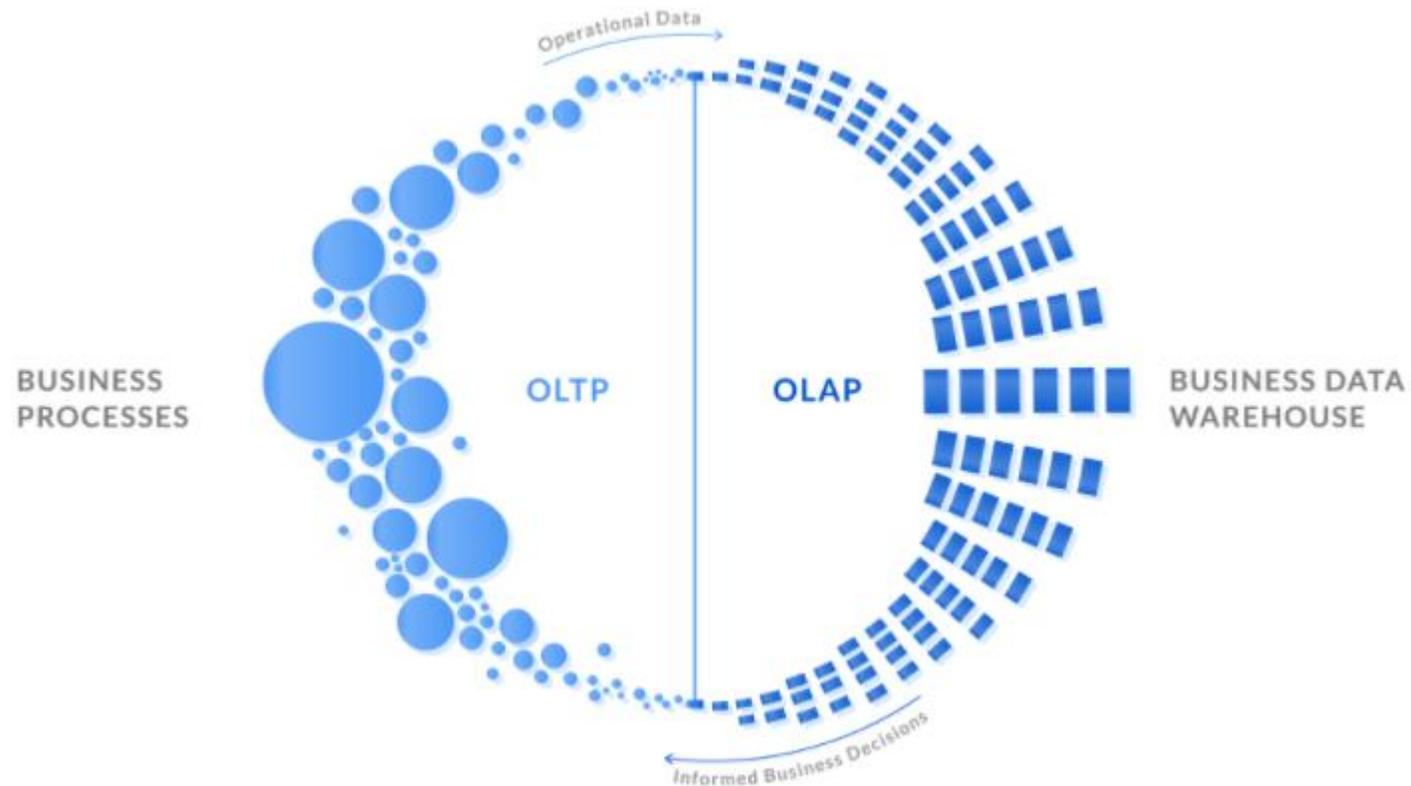
# Metodología Business Intelligence.



# OLTP vs OLAP

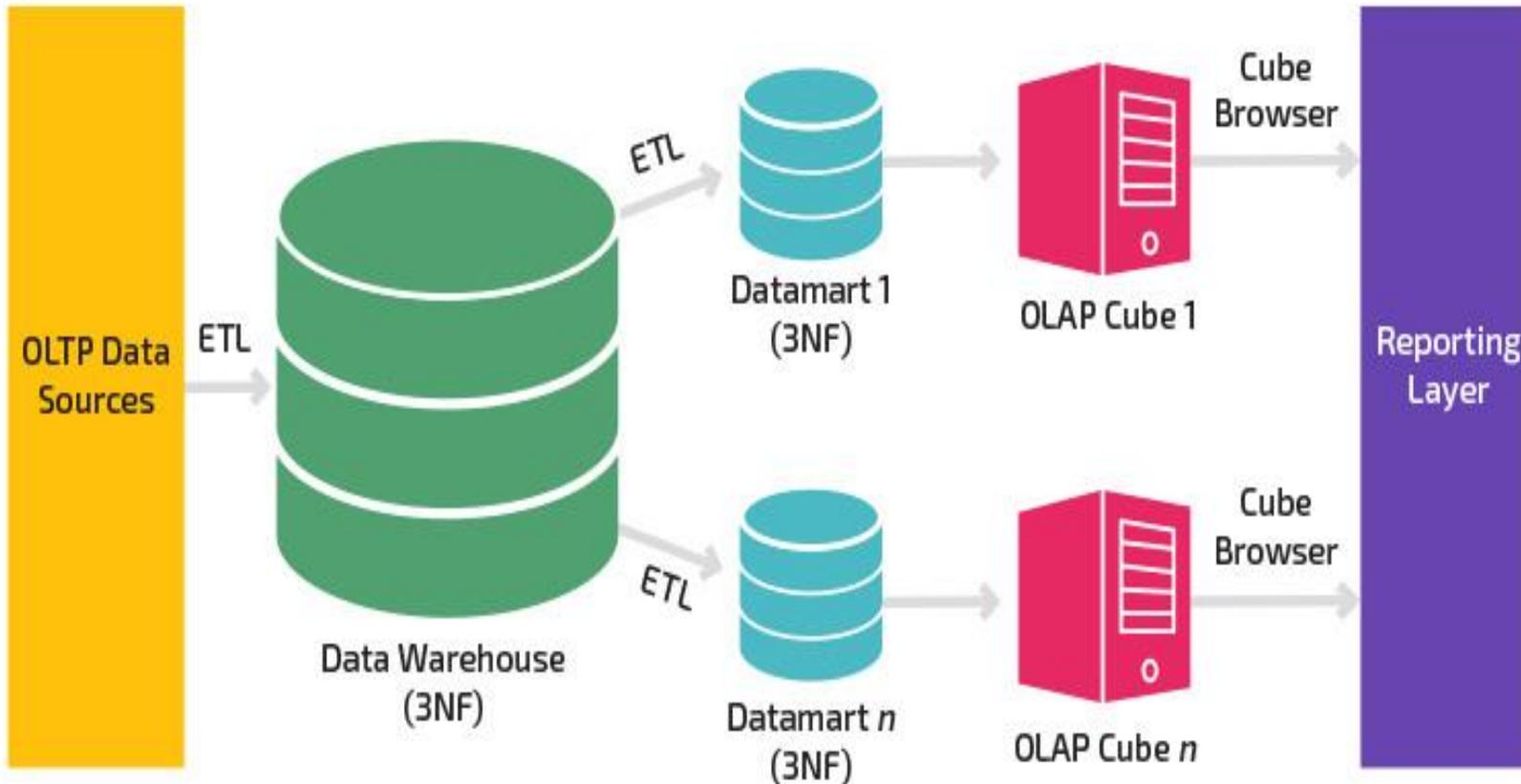
**OLTP** se utiliza para almacenar y administrar datos para las operaciones diarias.

**OLAP** se utiliza para analizar datos y tomar decisiones.



# Modelamiento multidimensional.

## Inmon Model



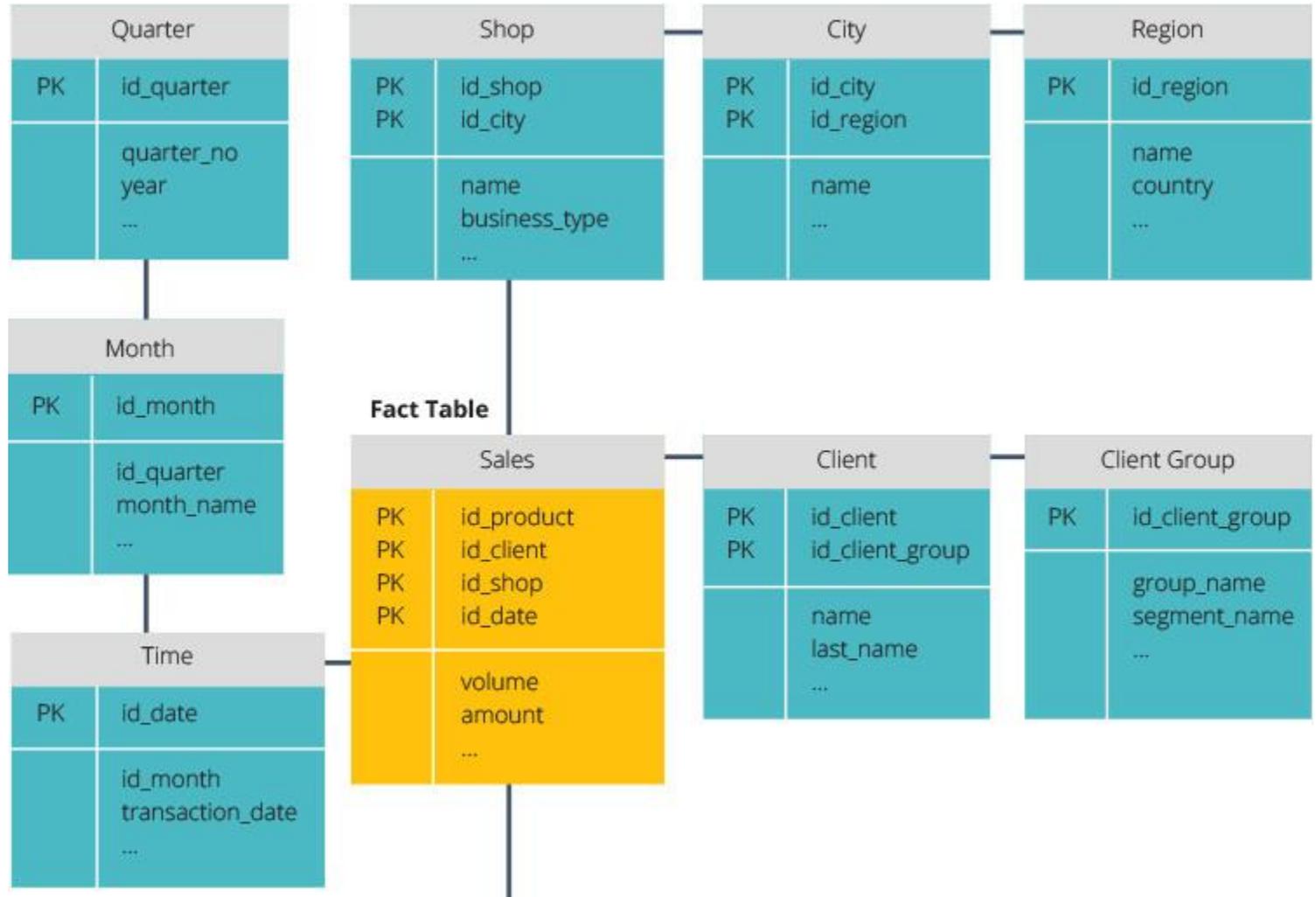
El enfoque de Bill Inmon favorece un diseño descendente en el que el DW es el depósito de datos centralizado y el componente más importante de los sistemas de datos de una organización.

El enfoque de Inmon primero construye el modelo centralizado de datos corporativos, y el almacén de datos se considera la representación física de este modelo. Los almacenes de datos dimensionales relacionados con líneas comerciales específicas se pueden crear desde el almacén de datos cuando se necesitan.

# Modelamiento multidimensional.

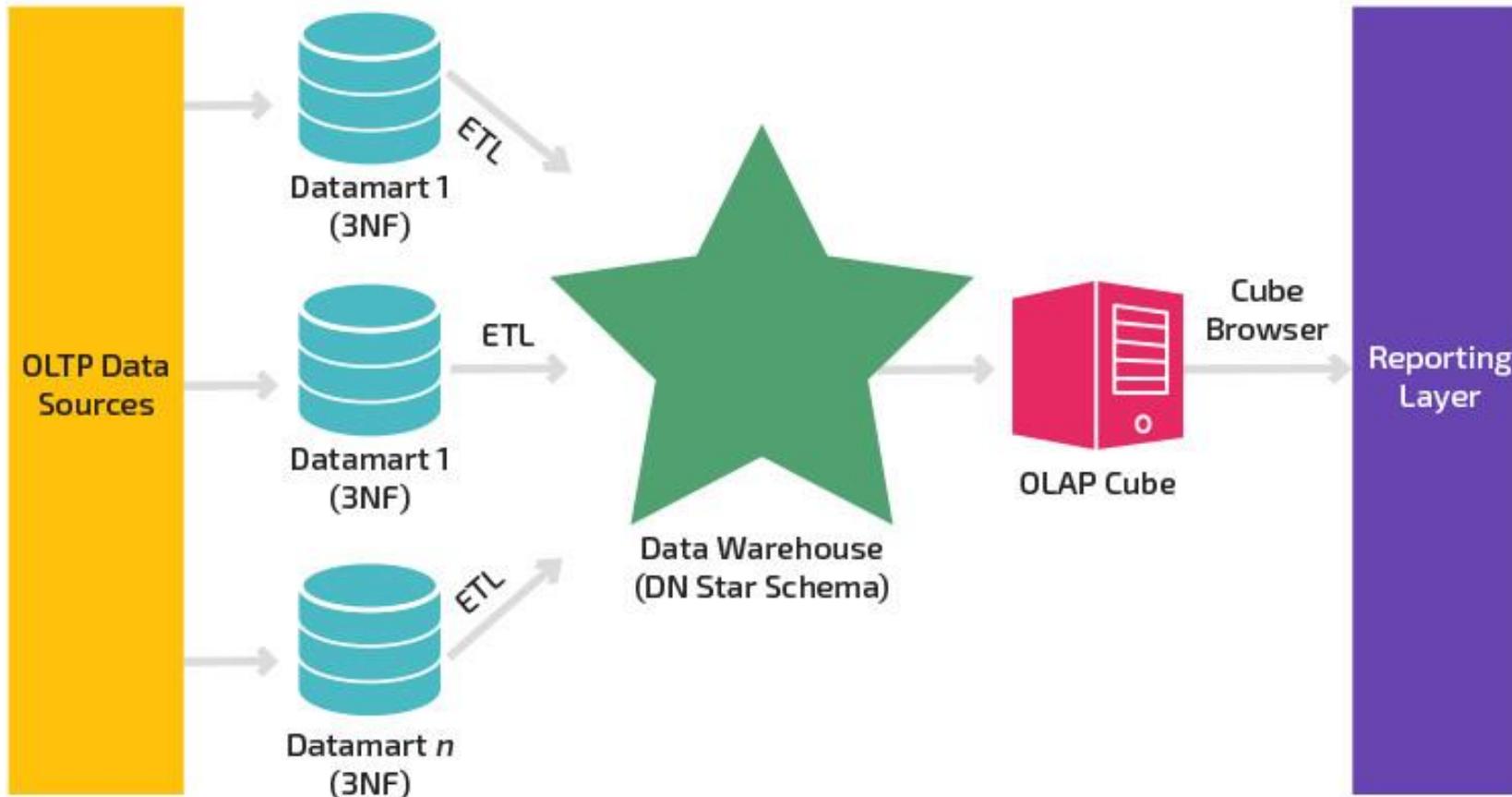
## El esquema de copo de nieve

Divide la tabla de hechos en una serie de tablas de dimensiones normalizadas. La normalización crea más tablas de dimensiones con múltiples combinaciones y reduce los problemas de integridad de datos. Sin embargo, las consultas son más desafiantes usando el esquema de copo de nieve, porque las consultas deben profundizar más para acceder a los datos relevantes.



# Modelamiento multidimensional.

## Kimball Model

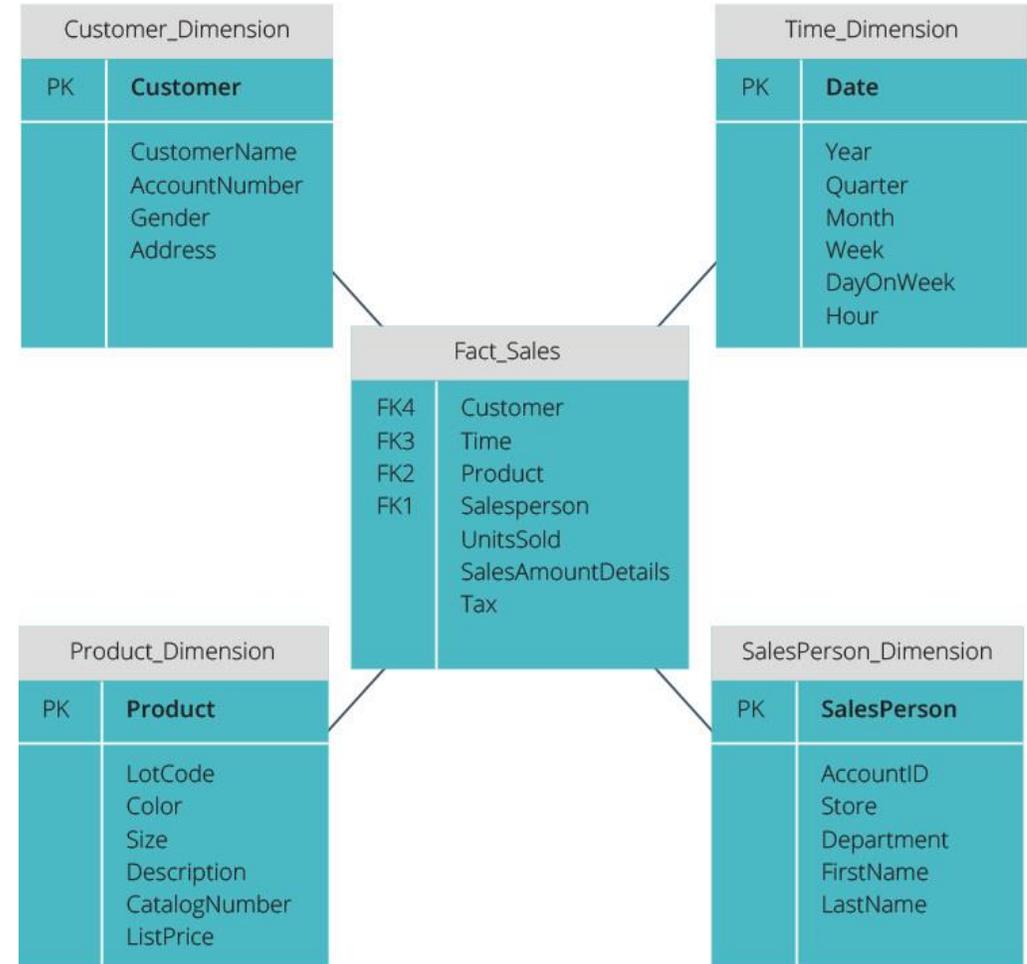


El diseño DW de Ralph Kimball comienza con los procesos comerciales más importantes. En este enfoque, una organización crea DM que agregan datos relevantes en torno a áreas específicas del tema. El DW es la combinación de los DM individuales de la organización.

# Modelamiento multidimensional.

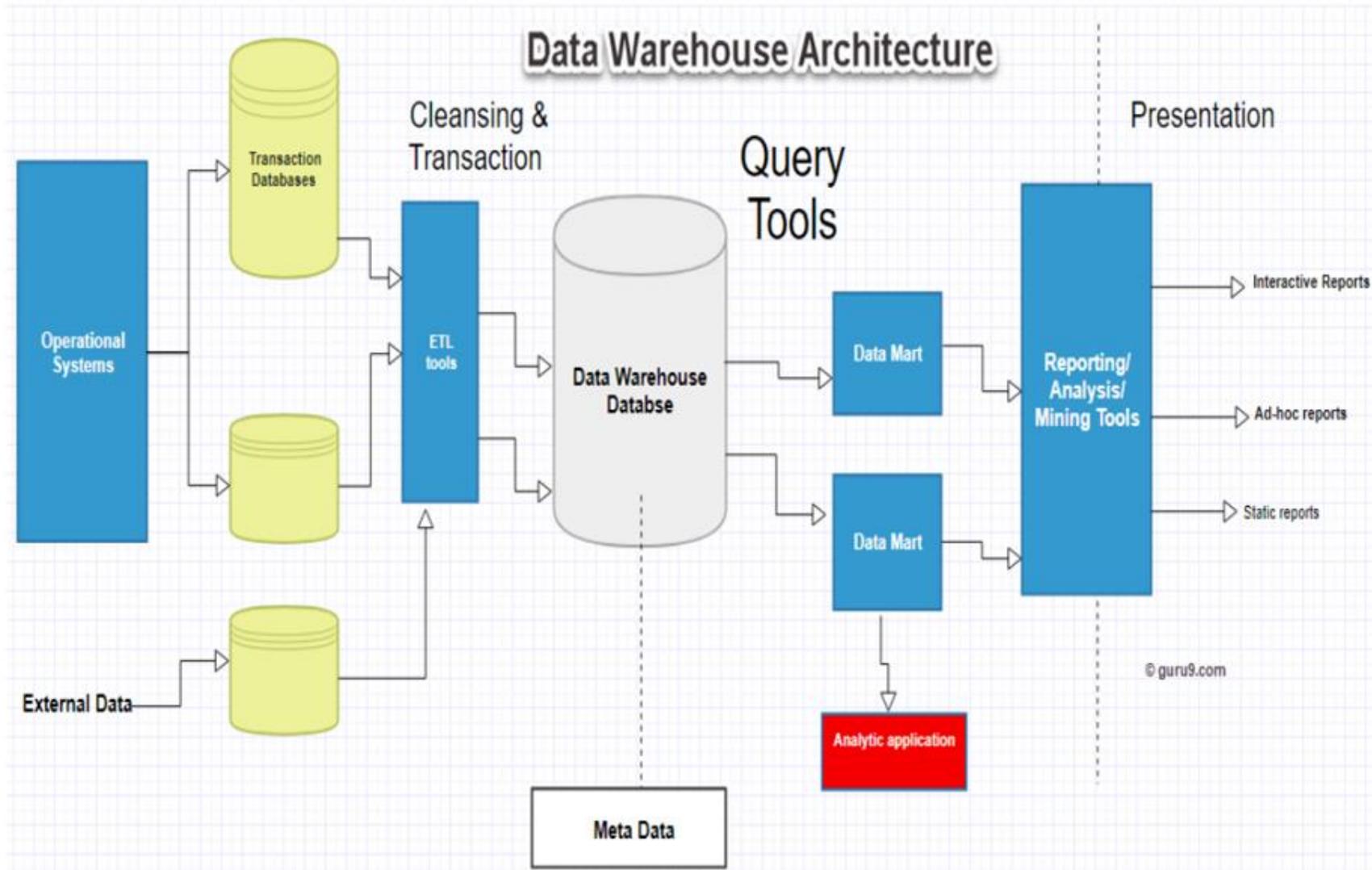
## El esquema en estrella

Toma la información de la tabla de hechos y la divide en tablas de dimensiones desnormalizadas. El énfasis para el esquema en estrella está en la velocidad de consulta. Solo se necesita una combinación para vincular tablas de hechos a cada dimensión, por lo que consultar cada tabla es fácil.



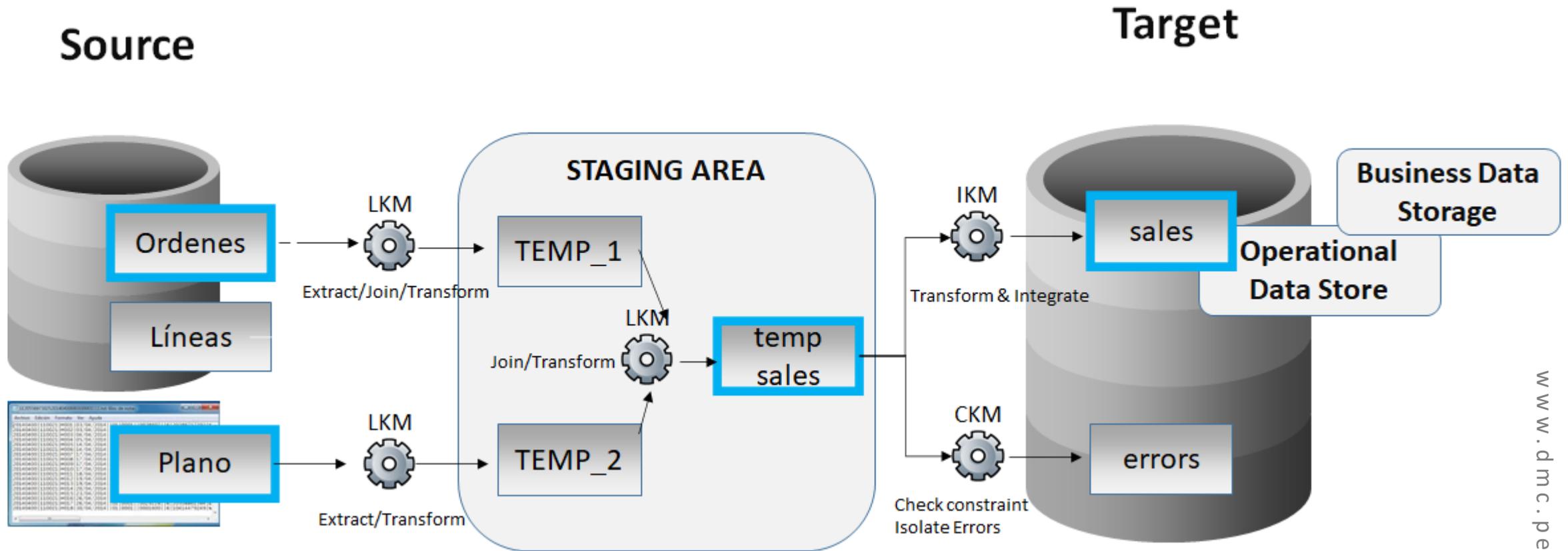
Legend: PK = primary key | FK = foreign key

# Introducción al Data Ware House (DWH)



# Introducción al Data Ware House (DWH)

Generalmente está diseñado para contener datos de bajo nivel o atómicos (indivisibles) (como transacciones y precios) con un historial limitado que se captura "en tiempo real" o "casi en tiempo real"

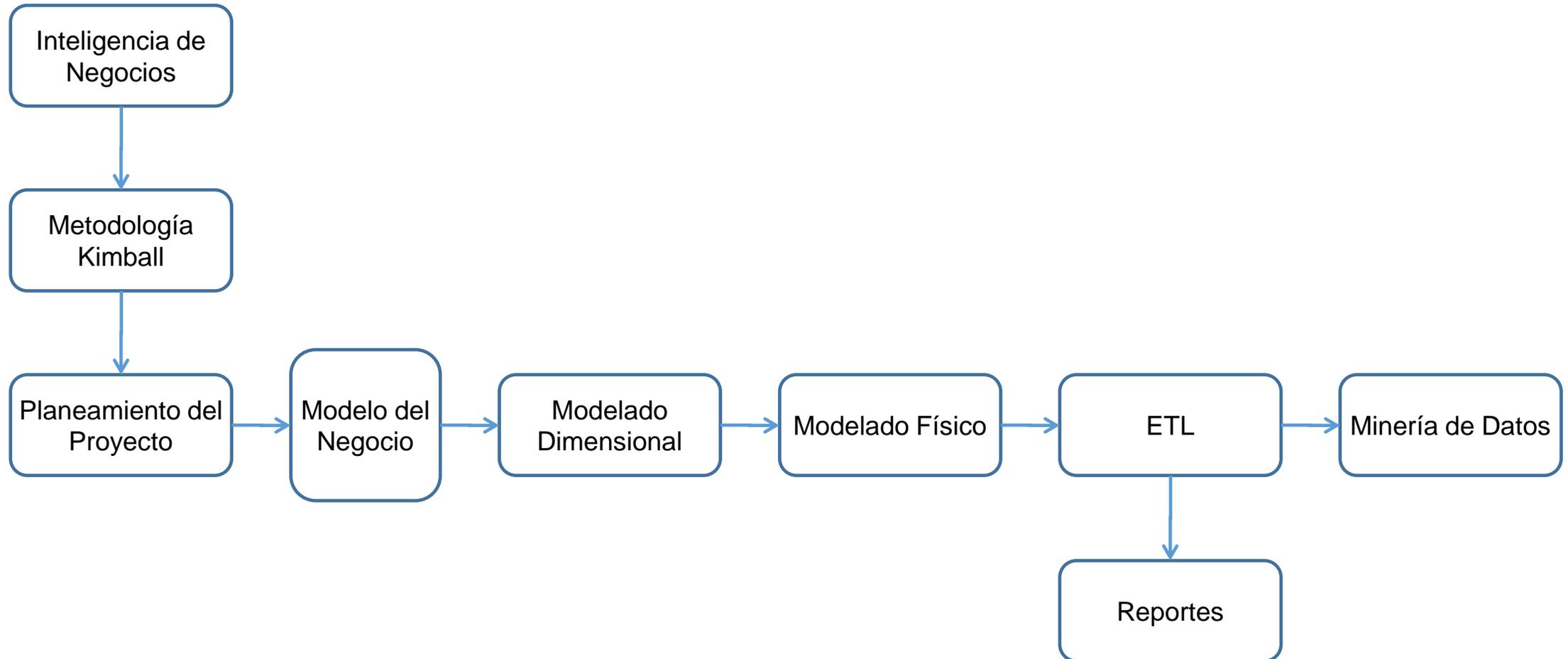


# Desarrollo de la metodología para la construcción de un DWH.

## Metodología Kimball

- Planificación del proyecto
- Requerimientos del Negocio
- Línea tecnológica
  - Arquitectura tecnológica
  - Selección e instalación de productos
- Línea de datos
  - Modelo dimensional
  - Modelo físico
  - ETL
- Línea de aplicación del BI
  - Diseño del BI
  - Desarrollo del BI
- Despliegue
  - Despliegue
  - Crecimiento
  - Mantenimiento

# Desarrollo de la metodología para la construcción de un DWH.



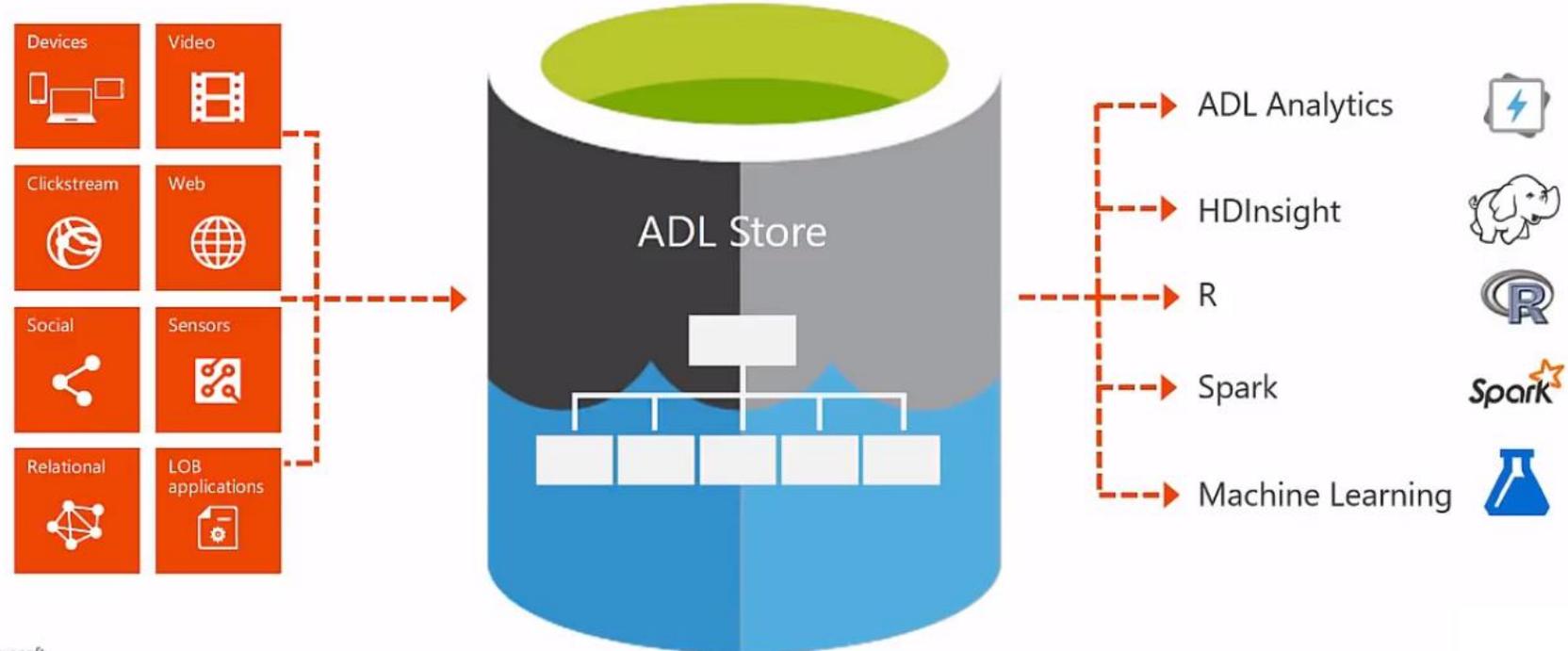
# Introducción a los datalakes.

Echemos un vistazo al concepto de lago de datos. A diferencia de las bases de datos tradicionales, un lago de datos almacena datos en su formato sin procesar. Por lo general, es un único depósito para todos los datos.

Incluye las copias sin procesar de los datos fuente y transformados. Un Datalake puede contener datos estructurados de bases de datos relacionales (una tabla), datos semiestructurados (CSV, JSON, registros, etc.), datos no estructurados (como correos electrónicos, documentos y PDF) y datos binarios (imágenes, audio y video).

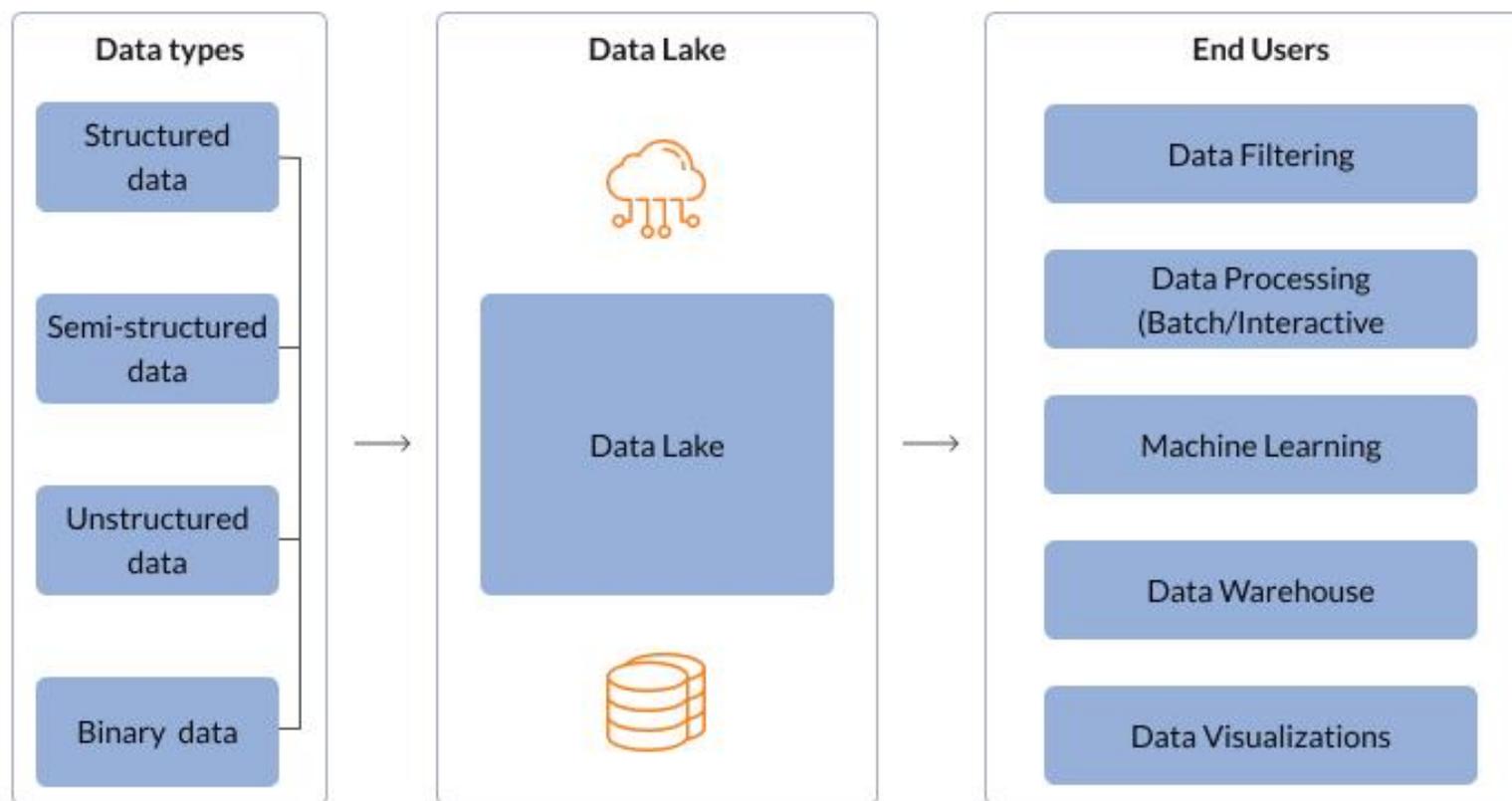
## What is Azure Data Lake (ADL) Store?

A highly scalable, distributed, parallel file system in the cloud specifically designed to work with multiple analytic frameworks



# Introducción a los datalakes.

## Data lake architecture (example)



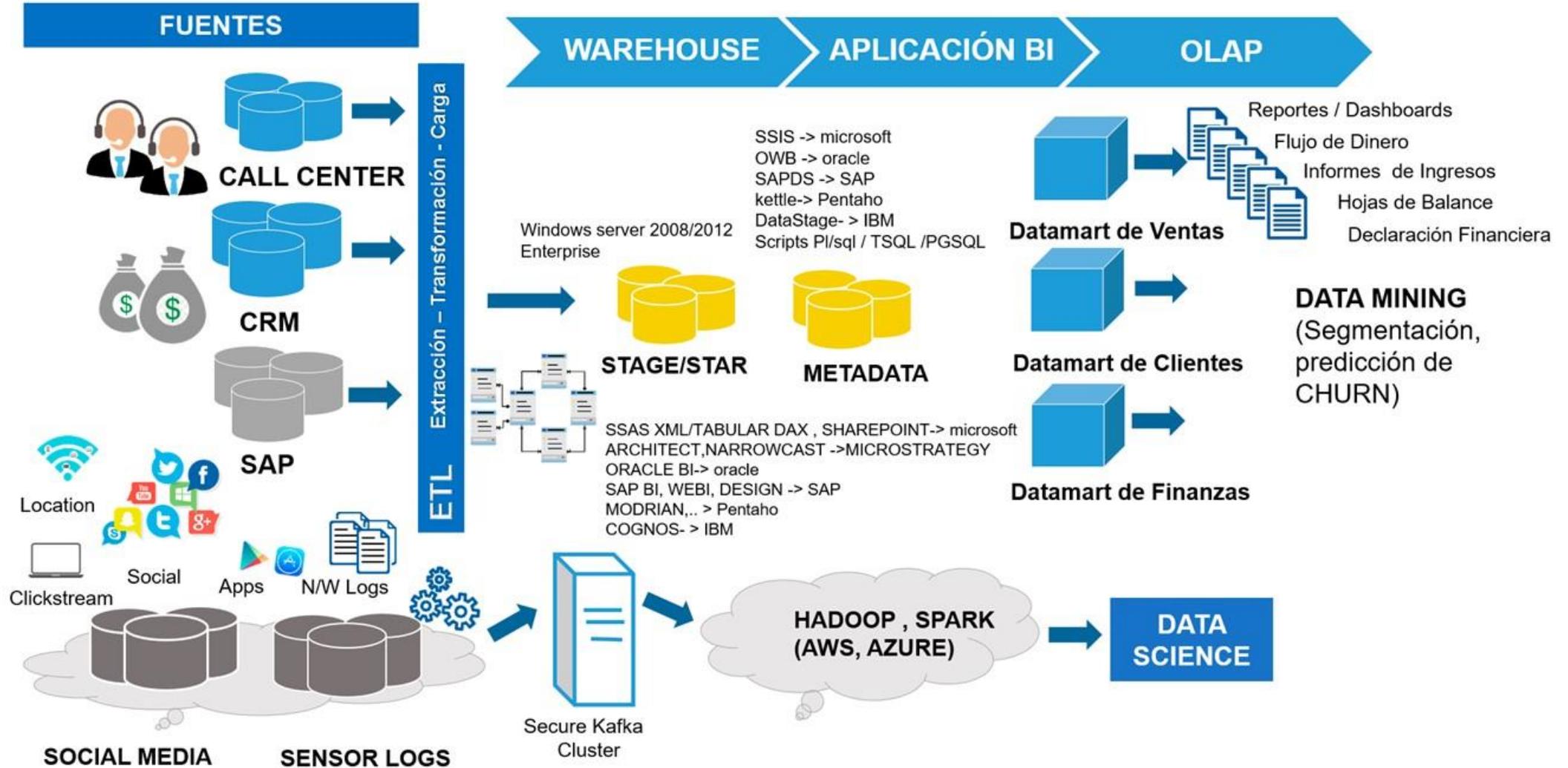
This chart offers an example of a structure in terms of the sources, business application, and analytics

## Introducción a los datalakes.

### Data lake pros and cons

- + Unlimited scalability
- + Data from diverse sources is stored in its raw format
- + Flexibility
- + Support of advanced algorithms
- + Excellent for integration with ML and IoT solutions
- Increased storage and computes costs
- Lack of insights from the previous findings
- Chances of data integrity loss

# Revisión de los principales softwares de manejo de bases de datos y de sistemas de DWH.





CAPACITACIÓN  
PROFESIONAL