



Denis Guido

Diseño e Implementación de Data Warehouse
con Sql Server 2019

INTRODUCCIÓN AL BUSINESS INTELLIGENCE



Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

¿CÓMO FUNCIONA BUSINESS INTELLIGENCE?



- **Apoyo en toma de decisiones**
- **Accesibilidad a la información**
- **Orientación al usuario**

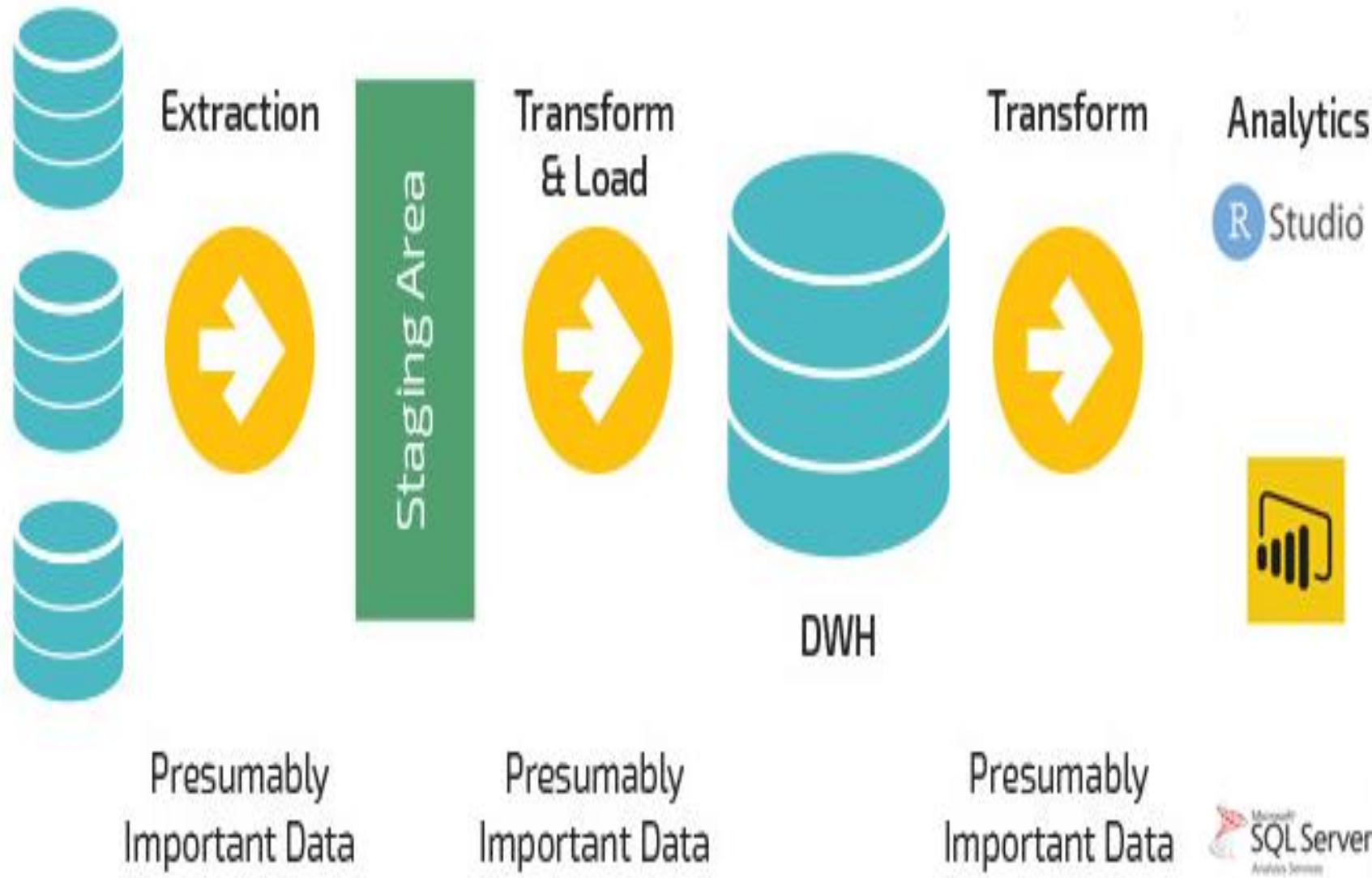
PRINCIPALES PRODUCTOS



- Cuadros de mando Integral
- Sistema de soporte a la decisión
- Sistema de información ejecutiva



PRINCIPALES COMPONENTES DE DATOS



- **Datamart**
- **Datawarehouse**

MODELADO

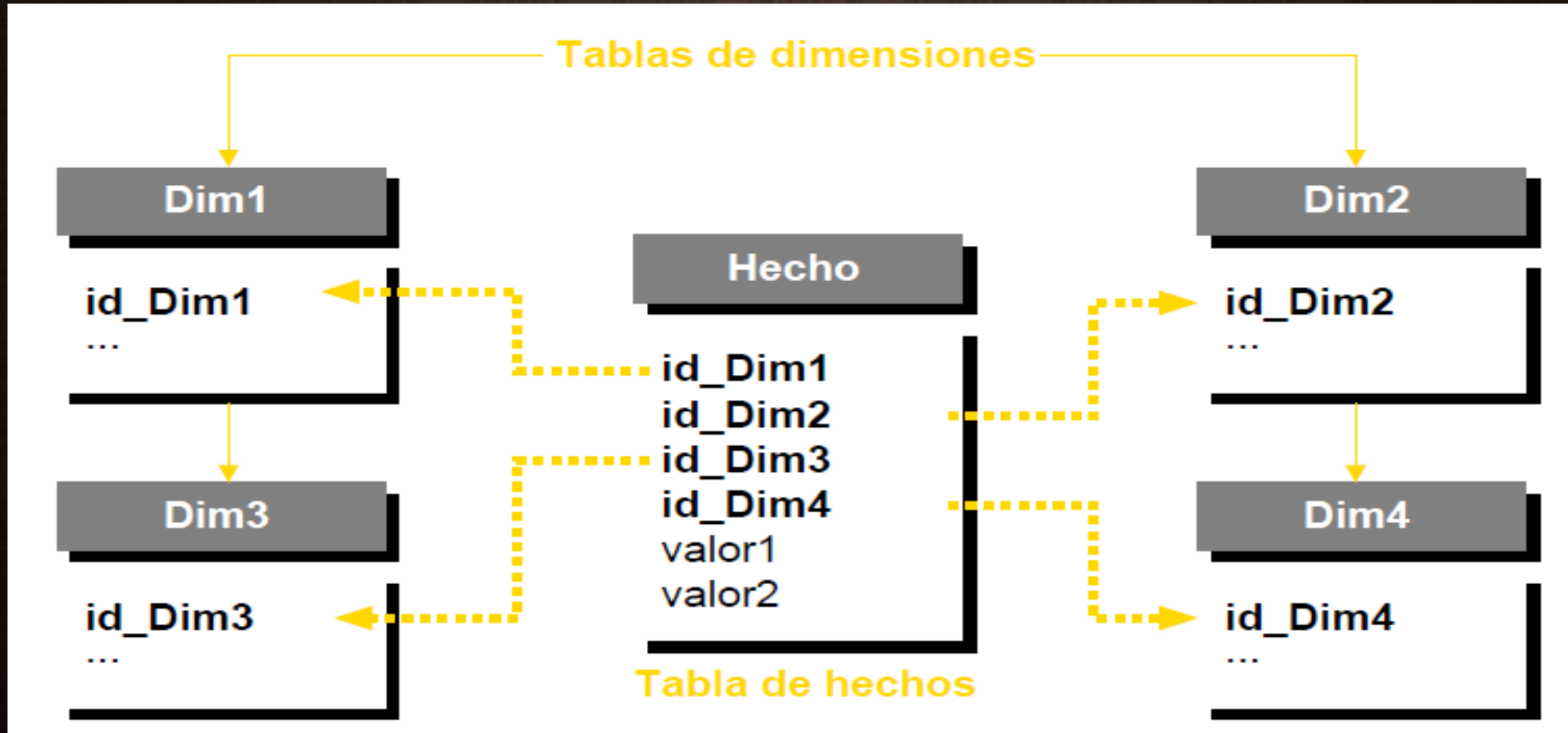


Esquema en estrella: formado por una tabla de hechos con una única tabla para cada dimensión.

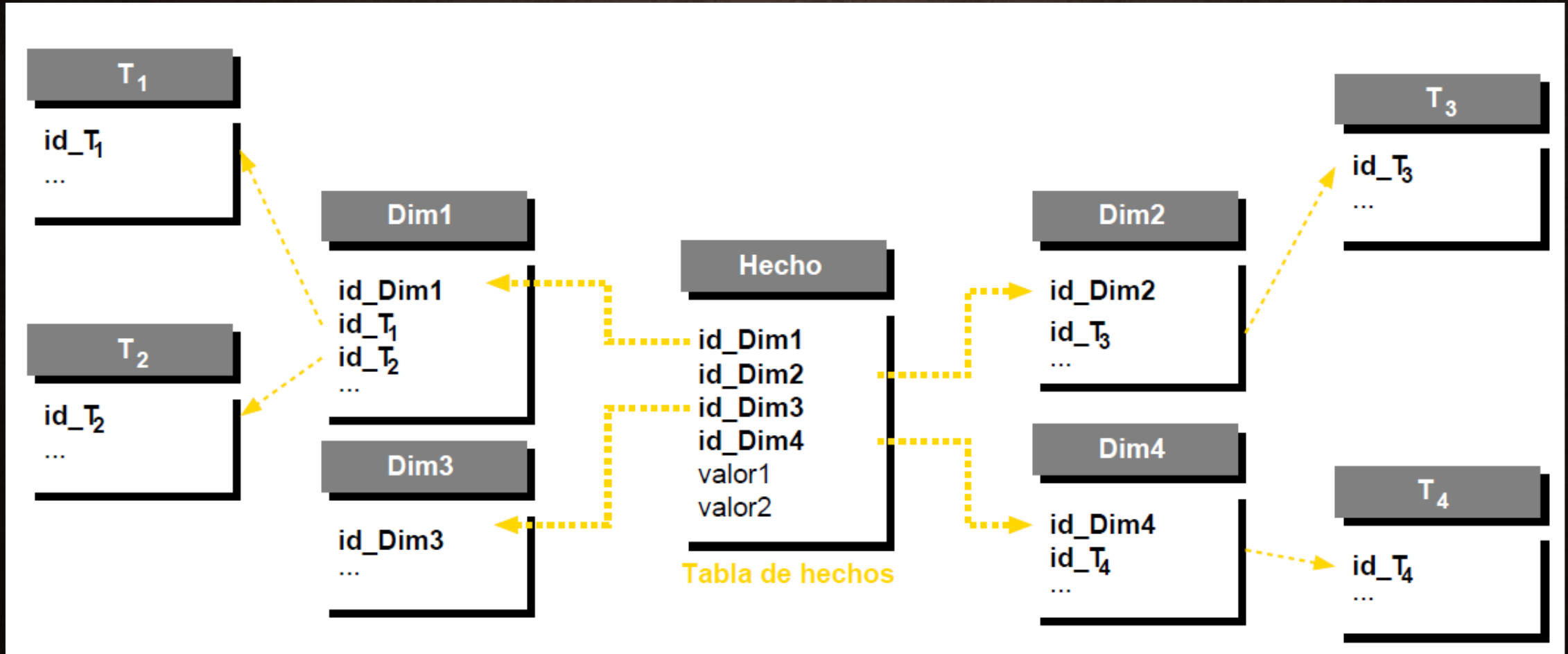
Esquema en copos: es una variante del esquema de estrella en el que las tablas dimensionales de este último se organizan jerárquicamente mediante su normalización.

Constelación de hechos: es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones.

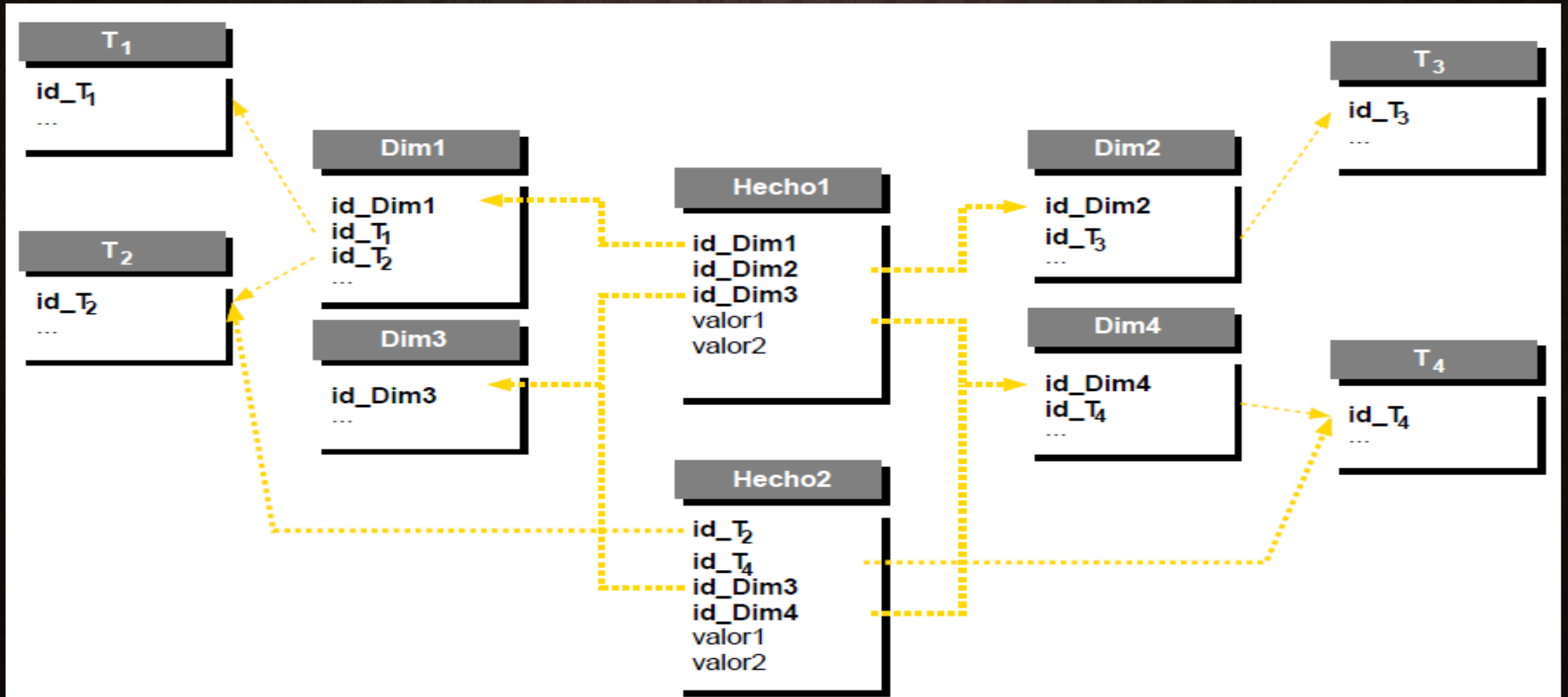
ESQUEMA DE ESTRELLA



ESQUEMA DE COPOS



CONSTELACIÓN DE HECHOS



ESQUEMA DE ESTRELLA A FONDO



El esquema en estrella es un enfoque de modelado maduro ampliamente adoptado por los almacenes de datos relacionales. Requiere que los modeladores clasifiquen sus tablas de modelos como *dimensión* o *hecho*

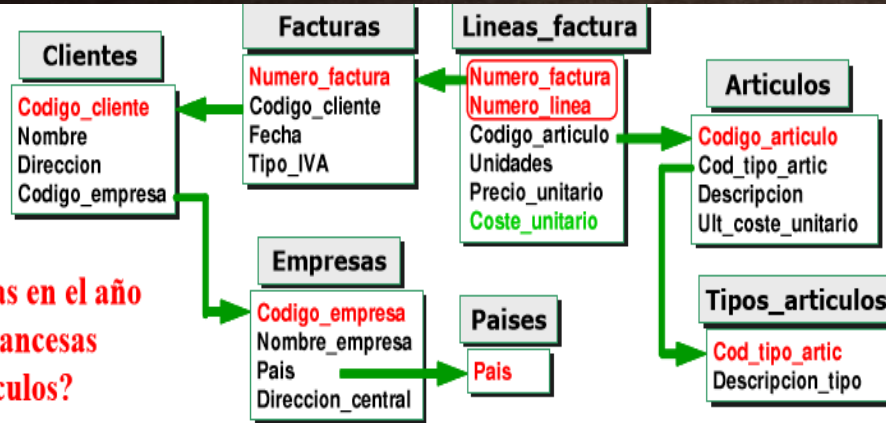
El diseño de esquema en estrella y muchos conceptos relacionados son muy relevantes para desarrollar modelos de Power BI que estén optimizados para el rendimiento y la usabilidad.

Ralph Kimball es autor en materia de almacenamiento de datos e inteligencia de negocios.

ESQUEMA DE ESTRELLA A FONDO



Sistema Transaccional



¿Beneficio de ventas en el año
2004 a empresas francesas
según tipos de artículos?

```
SELECT SUM(Unidades*(Precio_unitario - Coste_unitario)), Cod_tipo_artic
FROM Empresas INNER JOIN
  (Clientes INNER JOIN
    (Facturas INNER JOIN
      (Lineas_factura INNER JOIN Articulos
        ON Lineas_factura.Codigo_articulo=Articulos.Codigo_articulo)
      ON Facturas.Numero_factura=Lineas_factura.Numero_factura)
    ON Clientes.Codigo_cliente=Factura.Codigo_cliente)
  ON Empresas.Codigo_empresa=Clientes.Código_empresa
WHERE Fecha BETWEEN '1/1/2004' AND '31/12/2004' AND Pais='Francia'
GROUP BY Cod_tipo_artic
```

Modelo de Estrella

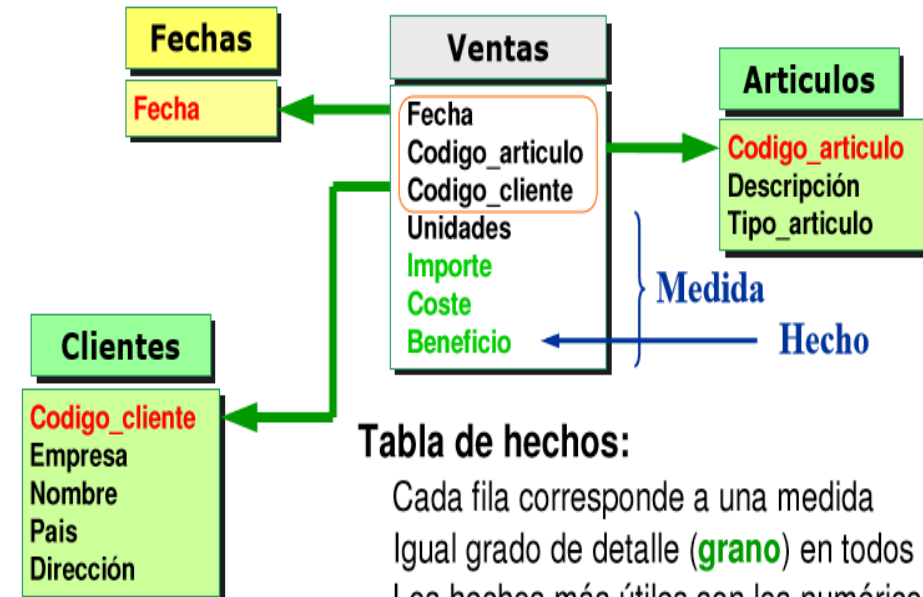


Tabla de hechos:

Cada fila corresponde a una medida
Igual grado de detalle (**grano**) en todos los hechos
Los hechos más útiles son los numéricos y aditivos

Tablas de dimensión:

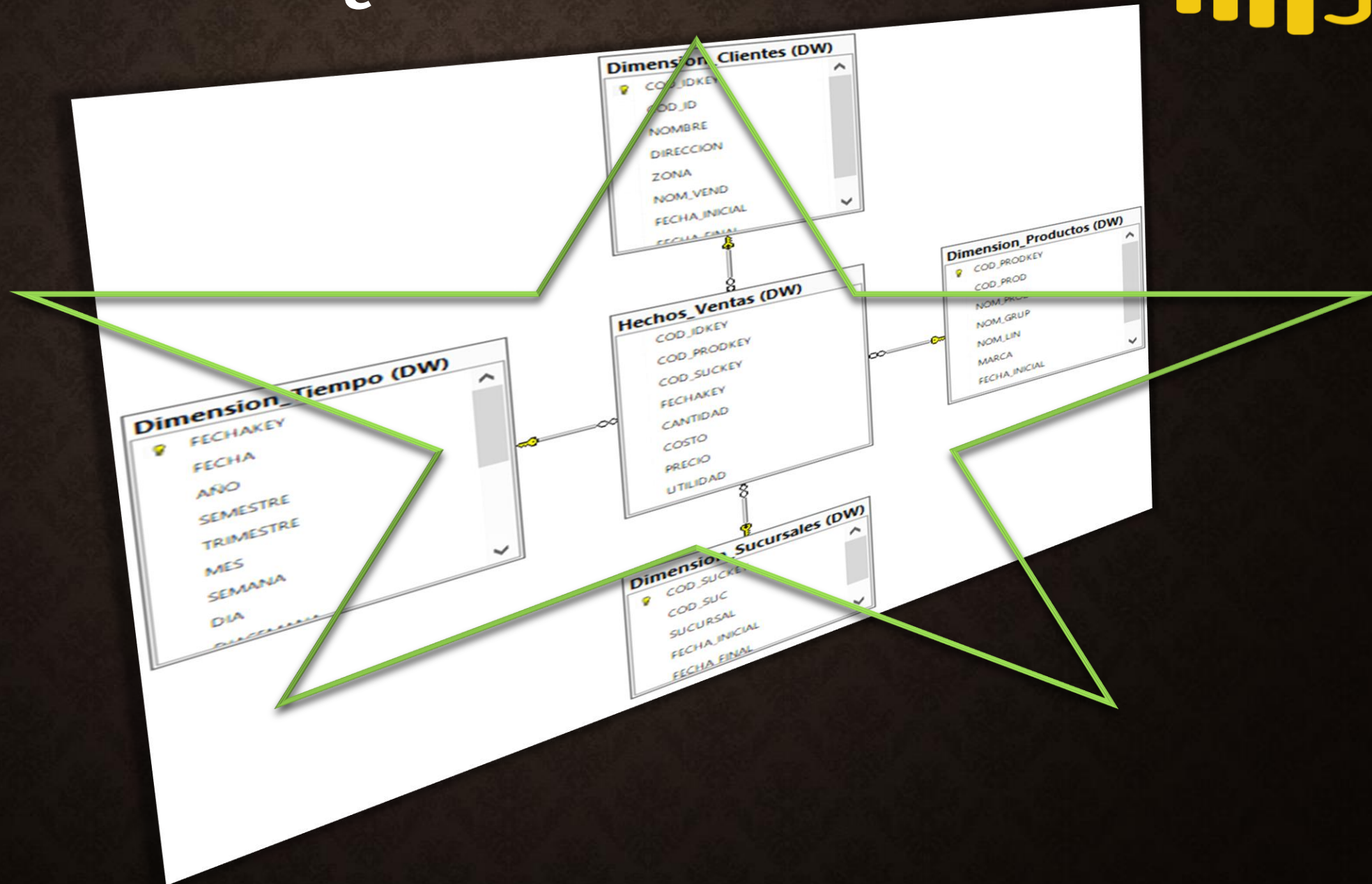
Contienen descriptores textuales
Son los puntos de entrada en la tabla de hechos

Transformación de datos

Desnormalización

Aún no es el modelo dimensional

EJEMPLO DE ESQUEMA DE ESTRELLA



GRUPOS DE ARCHIVOS DE BASE DE DATOS (FILEGROUPS)



Los objetos y los archivos de la base de datos se pueden agrupar en grupos de archivos para fines de asignación y administración. Hay dos tipos de grupos de archivos:

- **Primario** El grupo de archivos primario contiene el archivo de datos primario y cualquier otro archivo no asignado específicamente a otro grupo de archivos.
- **Grupos de archivos definidos por el usuario** son todos los grupos de archivos que se especifican utilizando la palabra clave **FILEGROUP** en una instrucción **CREATE DATABASE** o **ALTER DATABASE**.
- Los archivos de registro (**LOG**) nunca forman parte de un grupo de archivos. El espacio de registro se gestiona por separado del espacio de datos.

ARCHIVOS Y GRUPOS DE ARCHIVOS PARA EL ALMACÉN DE DATOS (FILEGROUPS)



Al diseñar la base de datos física,

- Separe hechos y dimensiones en diferentes grupos de archivos.
- Separe tablas de hechos grandes en diferentes grupos de archivos

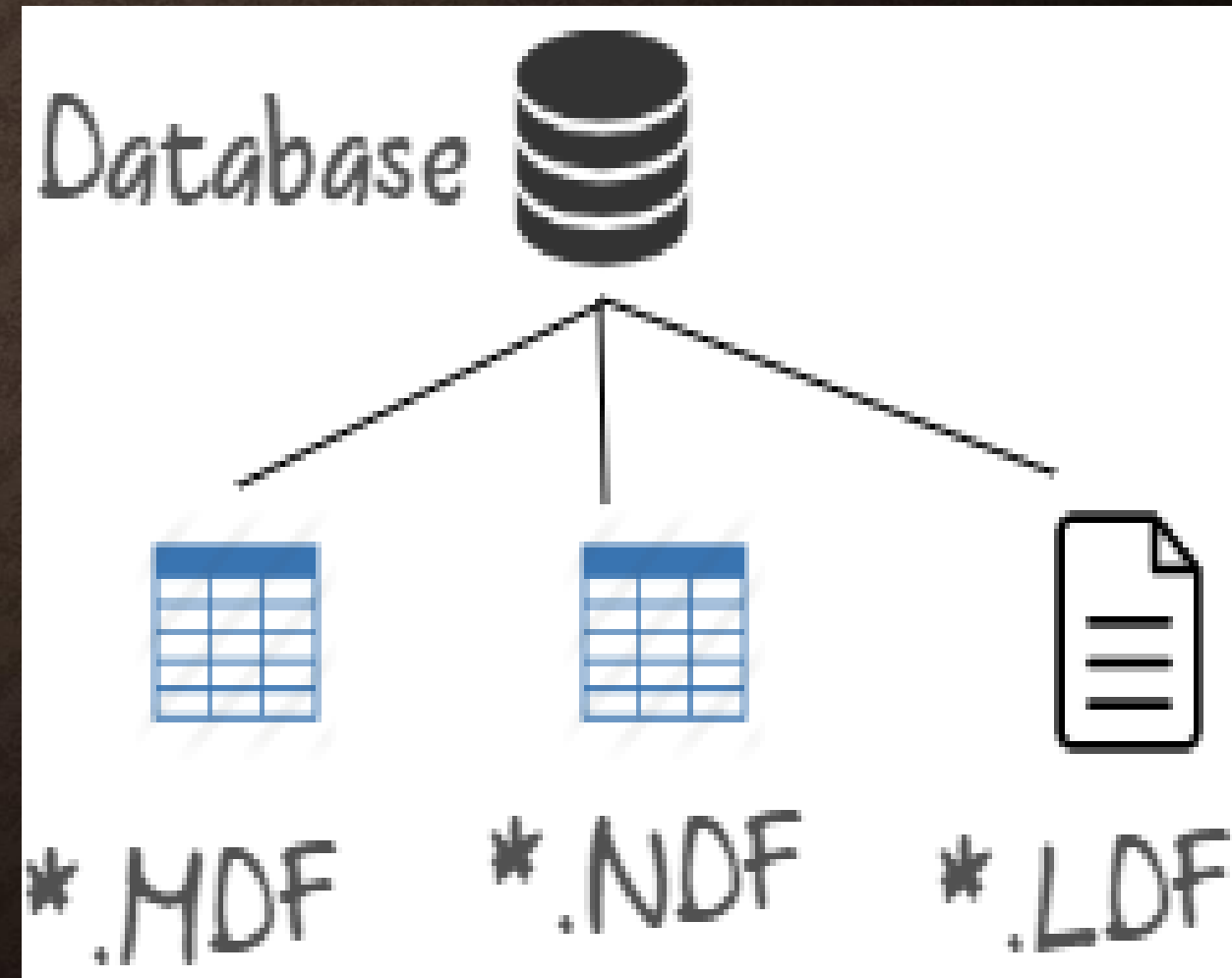
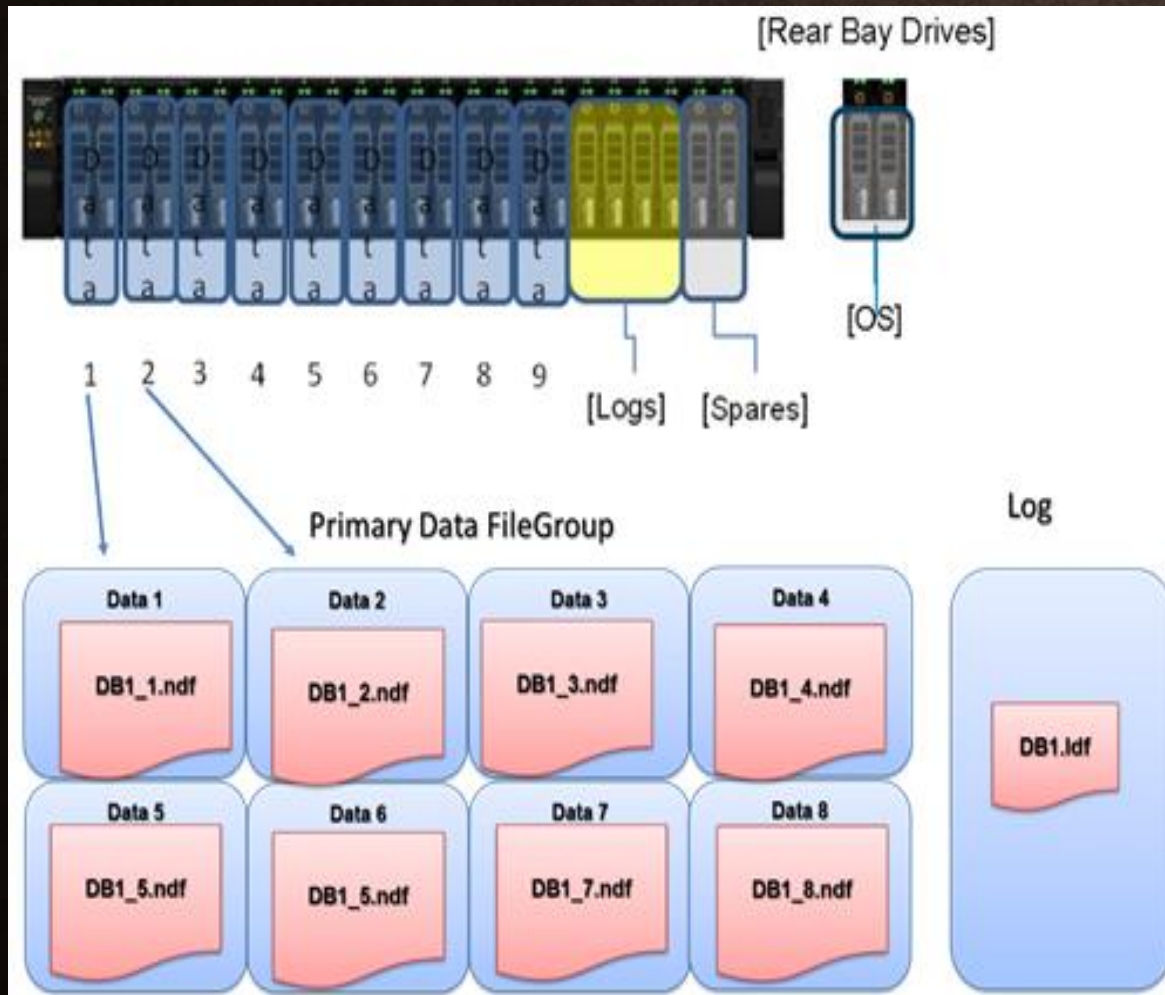
Ventajas

- Escritura rápida
- Lectura rápida

Desventajas

- Más archivos para mantener

ARCHIVOS Y GRUPOS DE ARCHIVOS PARA EL ALMACÉN DE DATOS (FILEGROUPS)



PARTICIONES



Los datos de tablas e índices particionados se dividen en unidades que se pueden distribuir en más de un grupo de archivos en una base de datos. Los datos se dividen horizontalmente, de modo que los grupos de filas se asignan a particiones individuales.

Componentes de particiones

Partición Función

- Esquema de partición
- Columna de partición
- Índice alineado
- Índice no alineado
- Eliminación de partición

PARTICIONES



All 10 Million Rows

ID	Date
1	2013-01-01
2	2013-05-20
10000000	2013-12-15

ID	Date
1	2013-01-01
240	2013-01-02
x	2013-01-31

ID	Date
5400000	2013-02-01
24256	2013-02-02
x	2013-02-28

ID	Date
876567	2013-12-01
30	2013-12-02
x	2013-12-31



Jan Rows

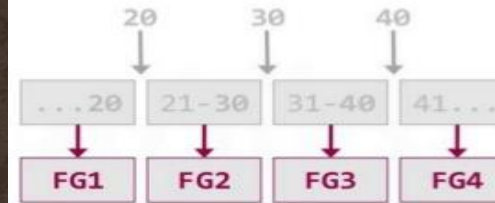


Feb Rows

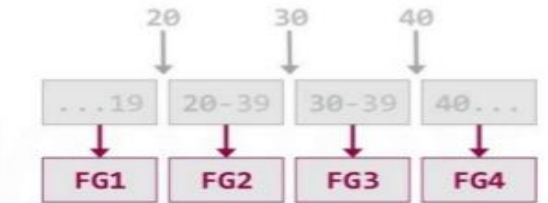


Dec Rows

```
CREATE PARTITION SCHEME
psLeft AS PARTITION pfLeft
TO (FG1, FG2, FG3, FG4);
```

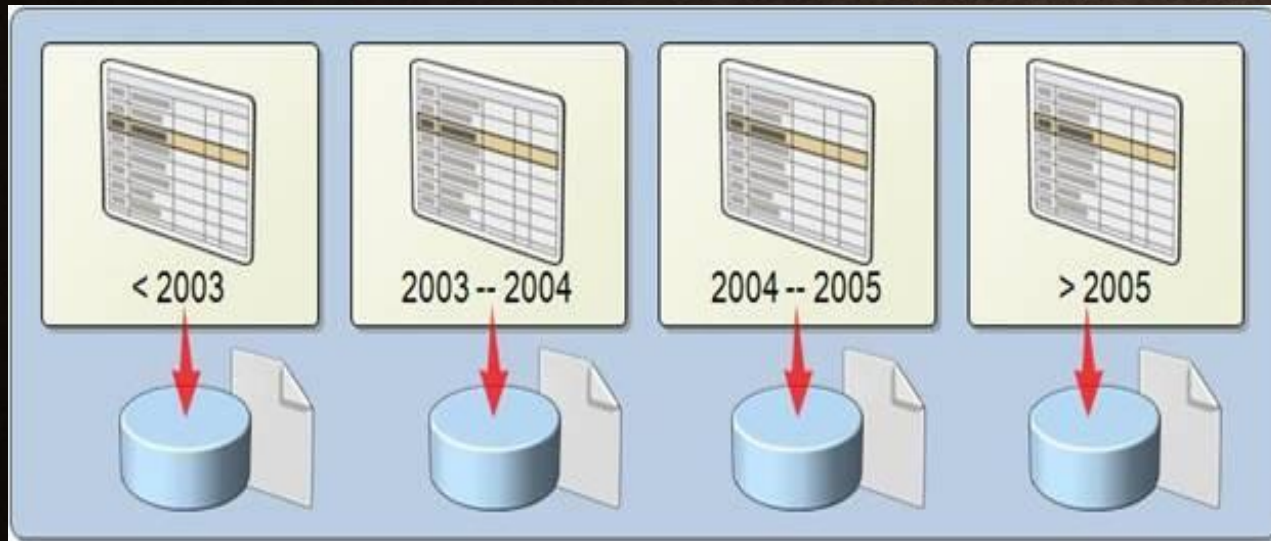


```
CREATE PARTITION SCHEME
psRight AS PARTITION pfRight
TO (FG1, FG2, FG3, FG4);
```



In scheme, using function we are sending values to file group

```
CREATE PARTITION SCHEME psOrderDateRange
AS
PARTITION pfOrderDateRange TO (FG1,FG2,FG3,[PRIMARY])
-- FG1: <2009 (all records before year 2009)
-- FG2: 2009 (from 2009/01/01 to 2009/12/31)
-- FG3: 2010 (from 2010/01/01 to 2010/12/31)
-- FG3: >2010 (all records after 31/12/2010 starting on 2011/01/01)
```



2012-01-01
2012-12-31
2013-01-01
2013-12-31
2014-01-01
2014-12-31
2015-01-01
2015-12-31

FILEGROUP1
(Read-Only)

FILEGROUP2
(Read-Only)

FILEGROUP3

FILEGROUP4



TABLAS OPTIMIZADAS EN MEMORIA EN SQL SERVER



Las Tabla de memoria optimizada, inician en SQL Server 2014, es simplemente una tabla que tiene dos copias, una en la memoria activa y otra duradera en el disco, ya sea que incluya datos o solo un esquema. Como la memoria se vacía al reiniciar SQL Services, SQL Server mantiene una copia física de la tabla que es recuperable. Aunque hay dos copias de la tabla, la copia de la memoria es completamente transparente y está oculta para usted.

¿CUÁL ES EL BENEFICIO ADICIONAL DE USAR ESTAS TABLAS EN MEMORIA?



Según Microsoft, el motor utiliza un enfoque optimista para esto, lo que significa que no coloca bloqueos en ninguna versión de filas de datos actualizadas, que es muy diferente a las tablas normales. Es este mecanismo el que reduce la contención y permite que las transacciones se procesen exponencialmente más rápido. En lugar de bloqueos, In-Memory usa Versiones de fila, manteniendo la fila original hasta que se confirma la transacción.

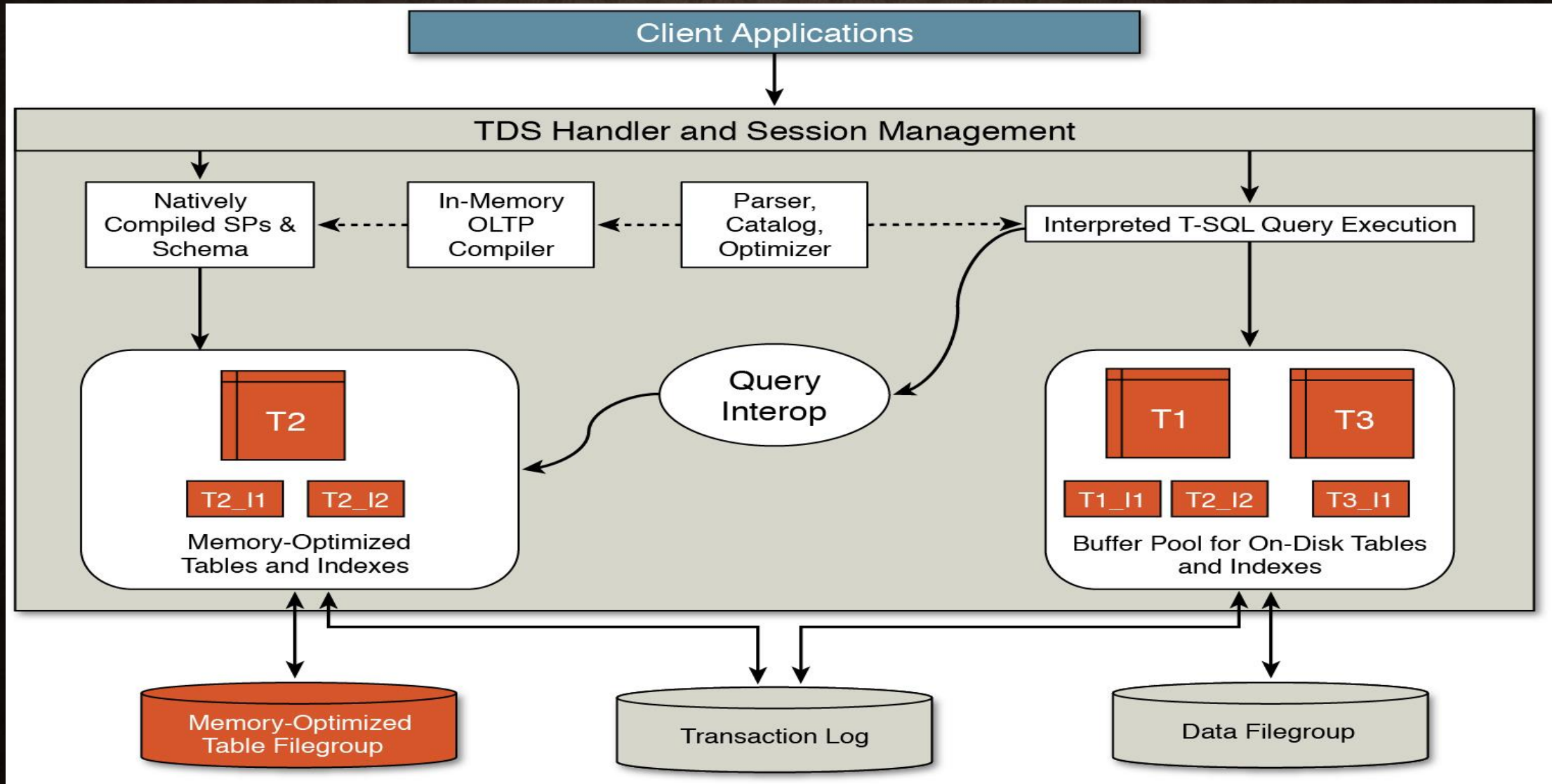
¿CUÁL ES EL BENEFICIO ADICIONAL DE USAR ESTAS TABLAS EN MEMORIA?



Esto permite que otras transacciones lean la fila original, al tiempo que actualiza la nueva versión de la fila. La versión estructurada en memoria no tiene páginas y está optimizada para la velocidad dentro de la memoria activa, lo que tiene un impacto significativo en el rendimiento según las cargas de trabajo.

Finalmente, otro beneficio adicional es la opción DURABILIDAD, El uso de SCHEMA_ONLY puede ser una excelente manera de evitar el uso de tablas # TEMP y agregar una forma más eficiente de procesar datos temporales, especialmente con tablas más grandes.

¿CUÁL ES EL BENEFICIO ADICIONAL DE USAR ESTAS TABLAS EN MEMORIA?





INDICES

Tipos de índices:

- Clustered
- Non-Clustered
- ColumnStore

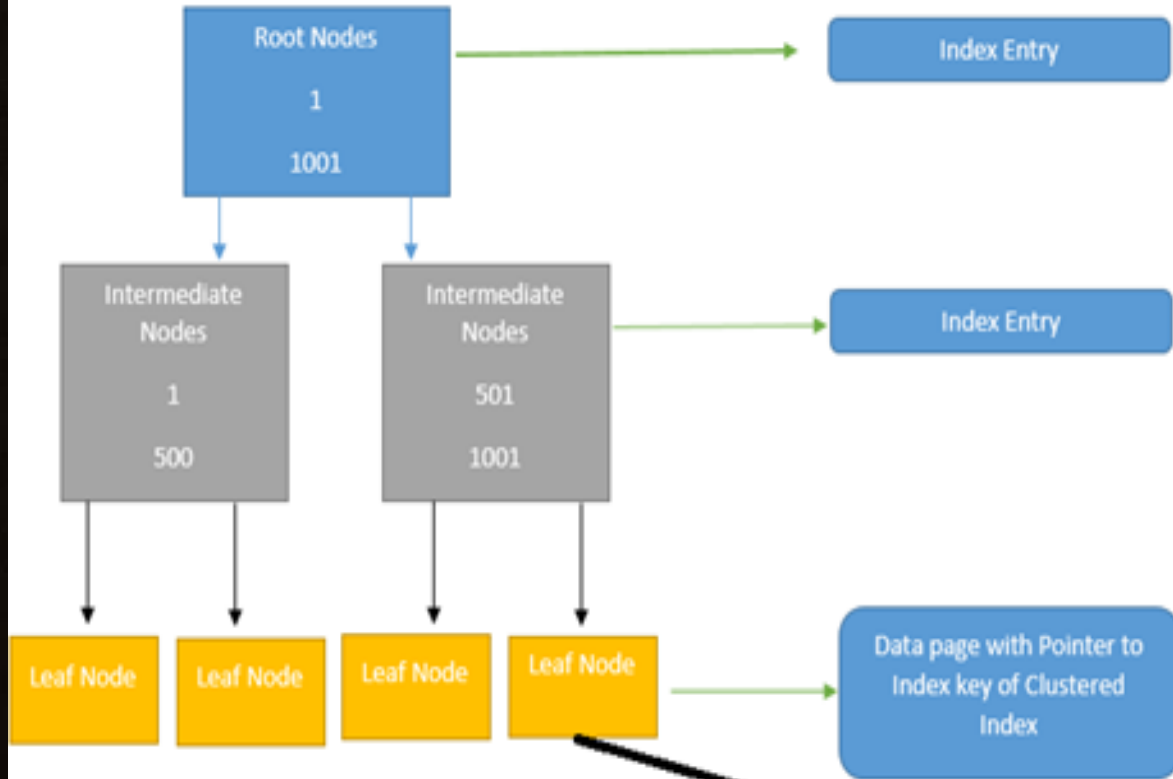
¿Deben los índices estar en las mismas unidades que los datos?

Respuesta: no, si desea aumentar el rendimiento. Esto no significa solo letras de unidad separadas, sino unidades físicas. La forma de separar los índices y los datos es mediante la asignación de los índices en diferentes grupos de archivos.

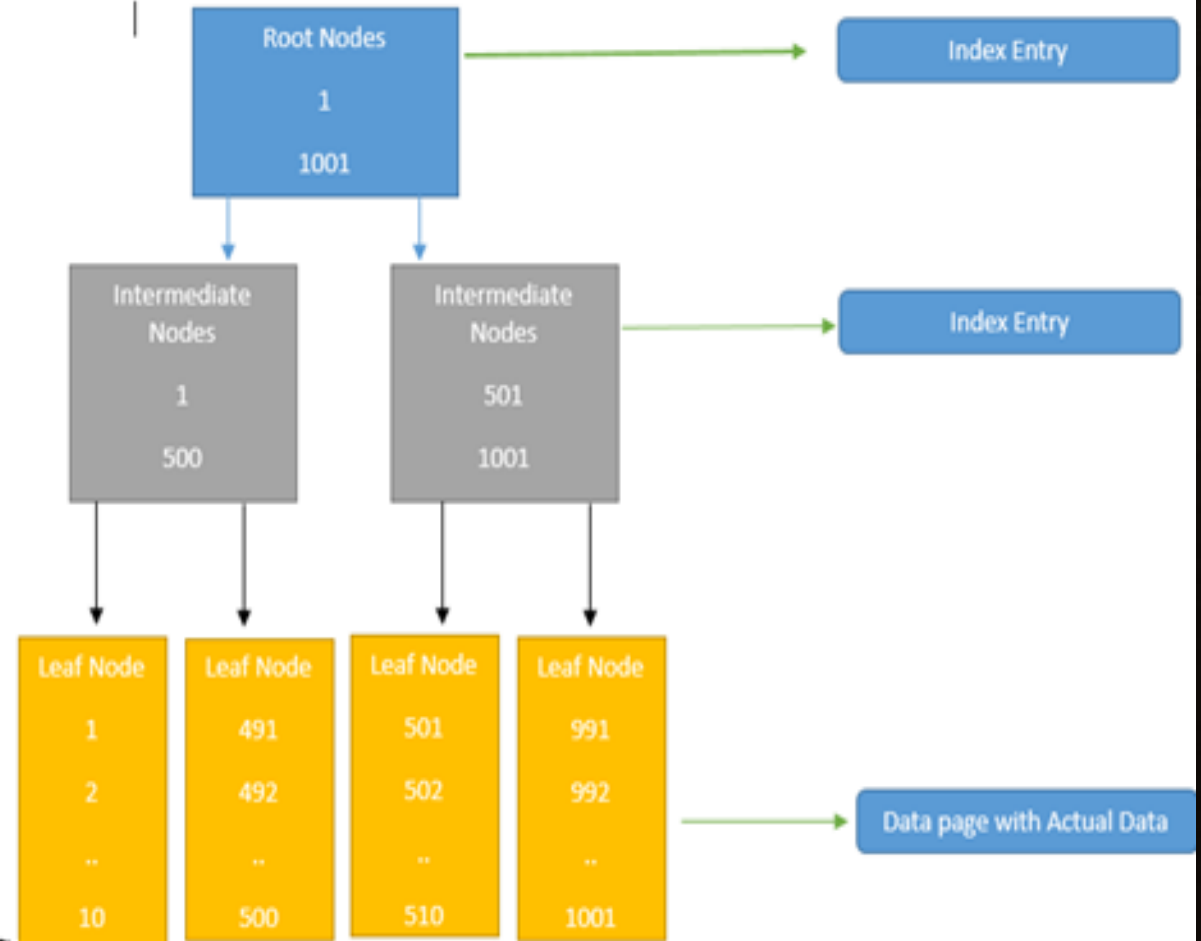
INDICES



Non-Clustered Index Structure



Clustered Index Structure



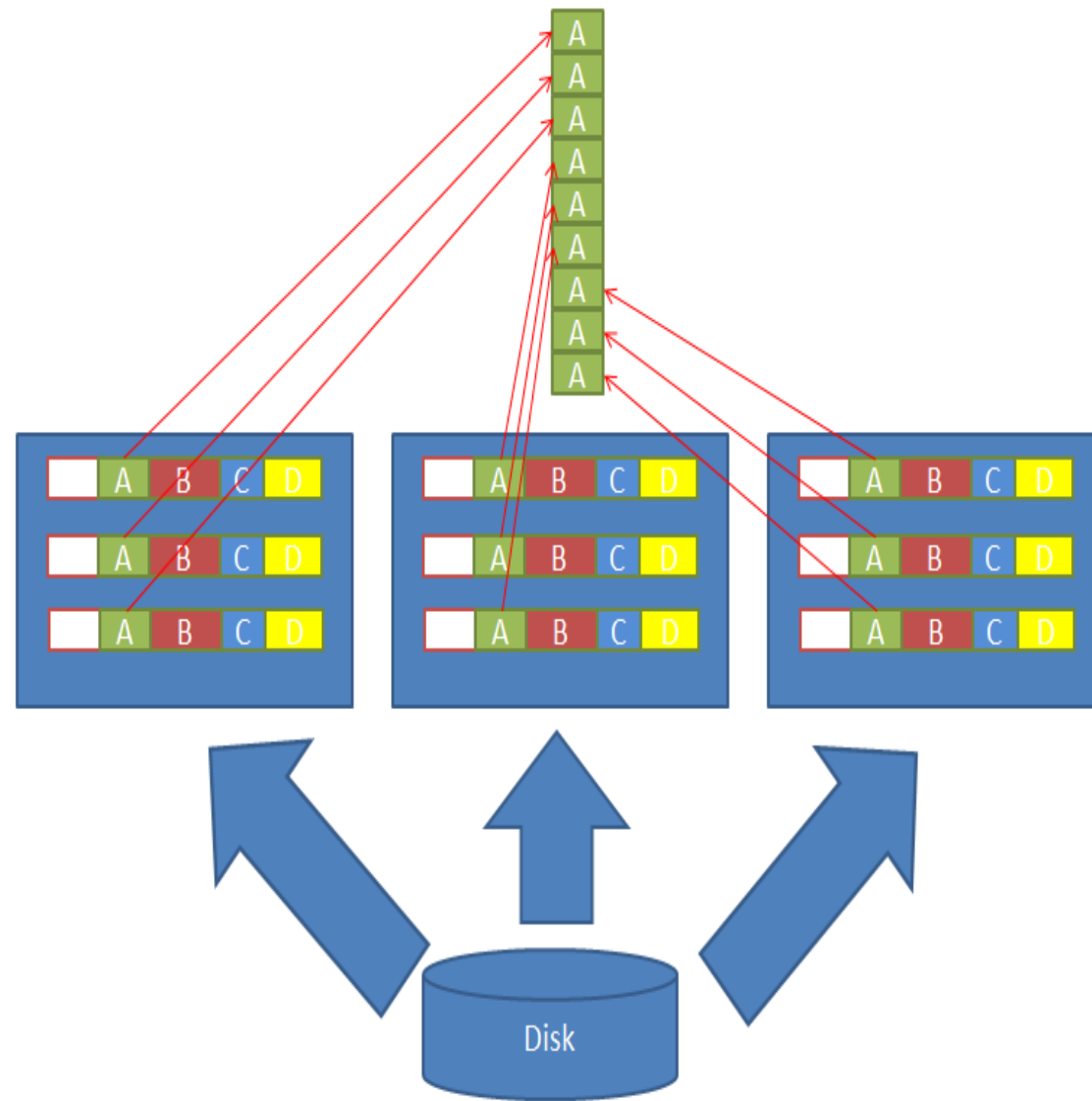
Leaf node of Non-Clustered index
Points to Clustered Index's Leaf
Node

INDICES



Column Store Index

	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Page 9	Page 10
Row 1	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row 2	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row 3	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row 4	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row 5	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row 6	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row 7	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row 8	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
.....	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Row n	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10



COMPRESIÓN DE BASE DE DATOS



Puede usar la función de compresión de datos para ayudar a comprimir los datos dentro de una base de datos y para ayudar a reducir el tamaño de la base de datos.

Compresión de datos a nivel de fila

Compresión de datos a nivel de página

Aunque se requiere CPU adicional, el hardware más nuevo tuvo una gran mejora en la CPU / Memoria, pero los husillos / discos no son mucho más rápidos.



DIMENSIONES LENTAMENTE CAMBIANTES (SCD 1 Y 2)

Una SCD Dimension son dimensiones en las que sus atributos cambian a lo largo del tiempo.

Dimensiones Tipo 1: Se sobrescriben los valores

Es el tipo de carga más básico y el más habitual. En este caso, se sobrescriben los valores existentes en el maestro con los nuevos valores.

COD_SUCKEY	COD_SUC	SUCURSAL
1	01	SUCURSAL 01
COD_SUCKEY	COD_SUC	SUCURSAL
1	01	CENTRO DE DISTRIBUCION

DIMENSIONES LENTAMENTE CAMBIANTES (SCD 1 Y 2)



Dimensiones Tipo 2: Se guarda la historia de cambios

En los casos que sí que nos interesa guardar la historia se utiliza la estrategia conocida como "SCD Tipo 2". En este caso, la tabla de dimensión incluye los campos de *fecha inicial* y *fecha final*. Estas fechas nos permiten determinar en que estado estaba la dimensión en cualquier fecha del calendario, el registro actual se representa por la fecha final con el valor de 31/12/999,

COD_SUCKEY	COD_SUC	SUCURSAL	FECHA_INICIAL	FECHA_FINAL
1	01	SUCURSAL 01	22/07/2020	31/12/9999
COD_SUCKEY	COD_SUC	SUCURSAL	FECHA_INICIAL	FECHA_FINAL
1	01	SUCURSAL DISTRIBUCION	22/07/2020	25/07/2020
2	01	CENTRO DE DISTRIBUCION	25/07/2020	31/12/9999

ETL & BI



El Business Intelligence trata de extraer los datos de la empresa a través de distintas fuentes mediante herramientas o técnicas llamadas ETL (extraer, transformar y cargar), luego de los datos ser pasados por este proceso se continúa con una carga de almacén de datos.

Finalmente, todo este análisis deberá permitir incrementar el nivel administrativo, financiero y la toma de decisiones en cuanto a las acciones de la empresa.

UNA SOLUCIÓN DE BI EFECTIVA PERMITE



Observar: *¿Qué está pasando?*

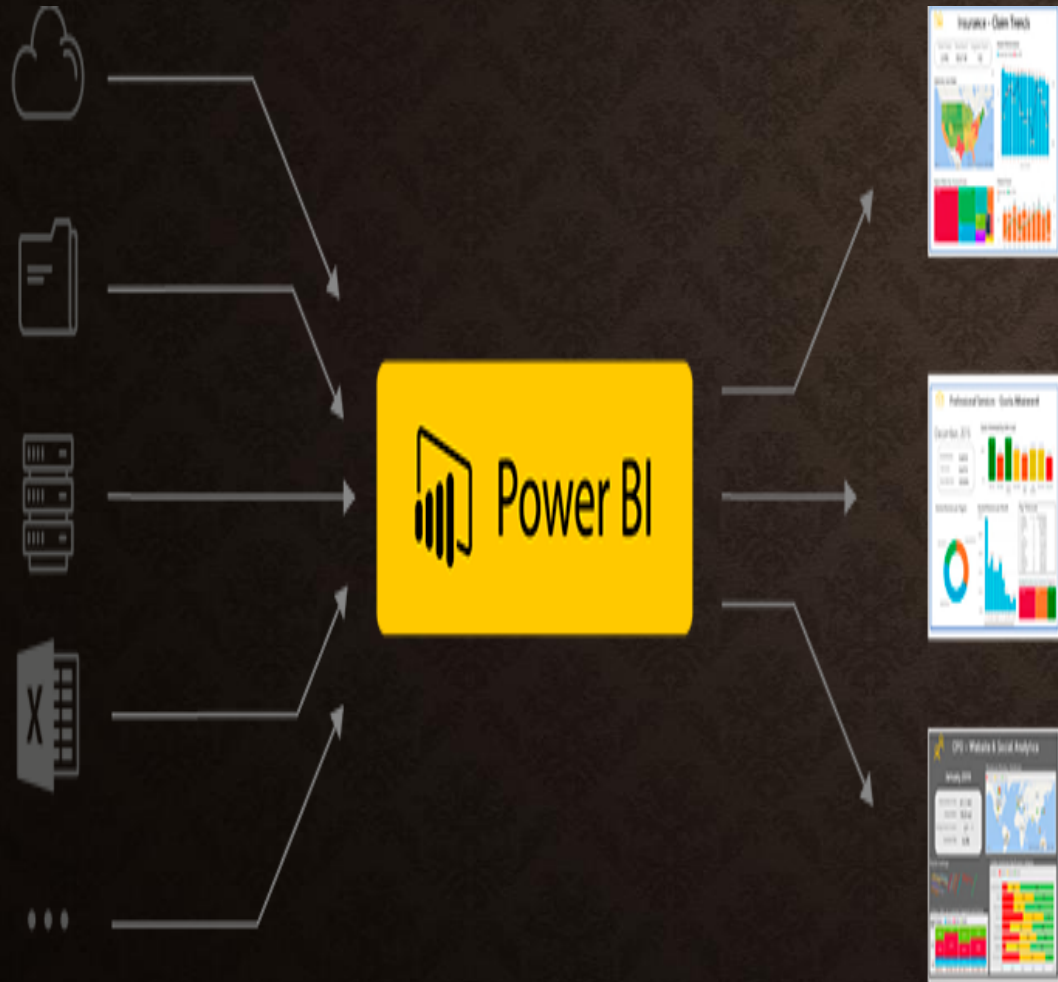
Comprender: *¿Por qué ocurre?*

Predecir: *¿Qué ocurriría?*

Colaborar: *¿Qué debe hacer el equipo?*

Decidir: *¿Qué camino debe seguir la empresa?*

¿ES REALMENTE NECESARIO PARA UN NEGOCIO?



Todo lo que pueda ser de utilidad para conocer una empresa en profundidad como datos relativos a ventas, compras, inversiones, tiempos, etc. son aquellos datos que el Business Intelligence recopilará y analizará para dar resultados claves con el fin de seguir mejorando.