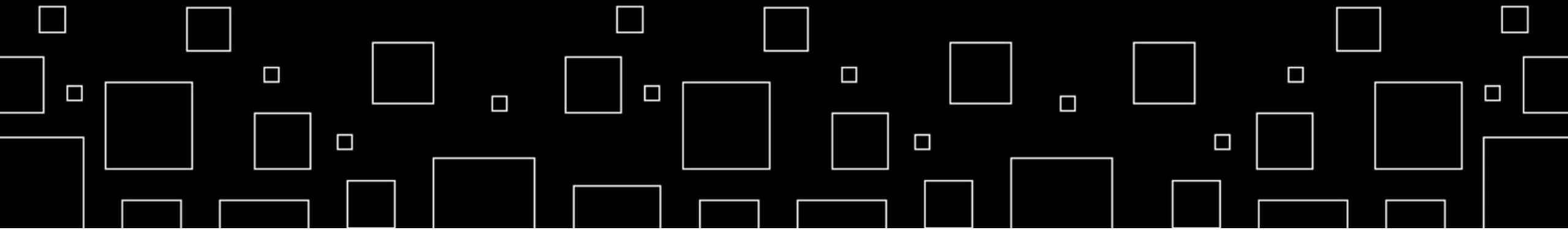




¡LES DAMOS LA BIENVENIDA!

Curso Data Analytics – Modulo 2





Noelia Soledad Brath

Educación ■

Ingeniería en Sistemas de Información - UTN FRBA

Laboral ■

Directora Comercial – 2020 hasta la actualidad

Directora de Servicios – 2019 - 2020

Líder de BI & Analytics – 2016 al 2018

Líder de Desarrollo – 2013 al 2016

Analista / Desarrolladora – 2011 - 2013

Curso Data Analytics



Módulo 1

2 clases

Introducción



Módulo 2

8 clases

Data Process



Módulo 3

8 clases

Data Viz



Módulo 4

6 clases

Ejercitación

Clases Módulo 2



- 1- Definición y Diseño Data Warehouse
- 2- Introducción de ETL
- 3- Lenguaje SQL: Sentencias Básicas
- 4- Lenguaje SQL: Sublenguaje DDL
- 5- Lenguaje SQL: Funciones de Agregación
- 6- Lenguaje SQL: Sublenguaje DML
- 7- Lenguaje SQL: Funciones / SP
- 8- Ejemplo de SP de Dims y Facts



PREGUNTAS DE LA CLASE ANTERIOR?

MATERIA D
Digital School

Clase 01 | Definición y Diseño DW

OBJETIVOS DE LA CLASE



Comprender que es una Base de Datos.



Entender que es un Data Warehouse y para que se construye. Diferencia con Data Marts

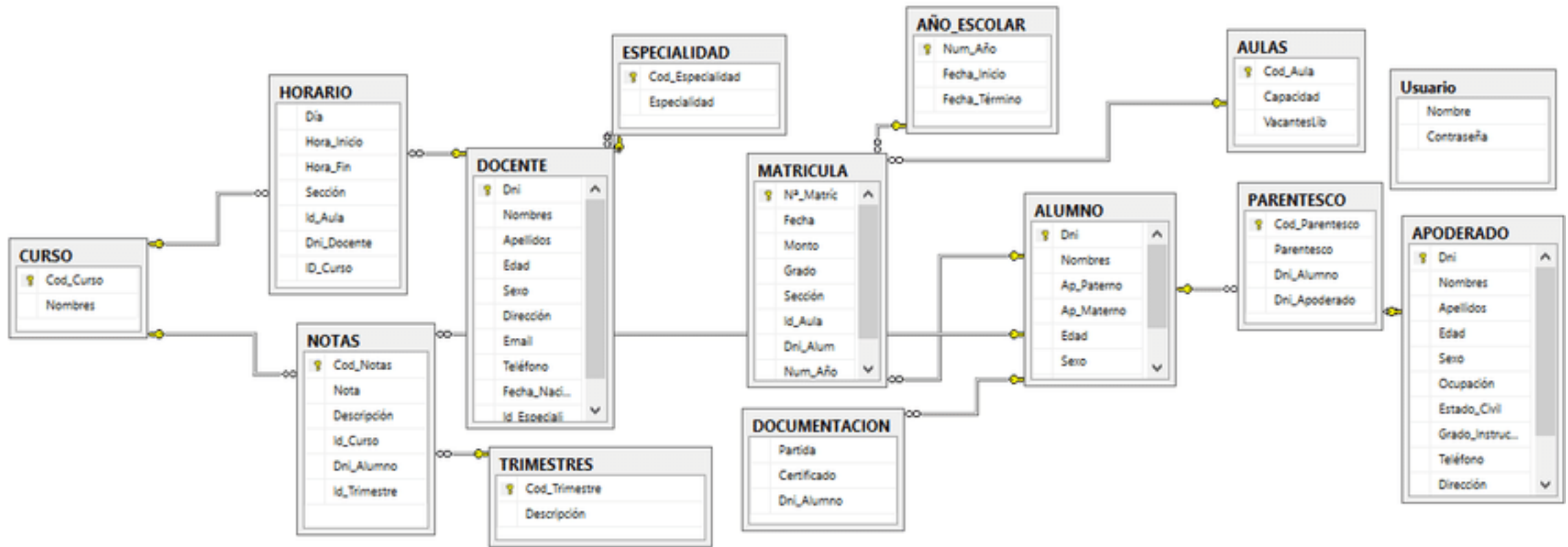


Aprender los conceptos y técnicas fundamentales del Modelado Dimensional y la diferencia con un Modelo Relacional



Crear una Base de Datos desde 0.

Base de Datos





Tipos de Bases de Datos



Según Variabilidad

Estáticas

Dinámicas



Según Contenido

Bibliográficas

Texto Completo

Directorio

Especializadas



Según el modelo

Jerárquicas

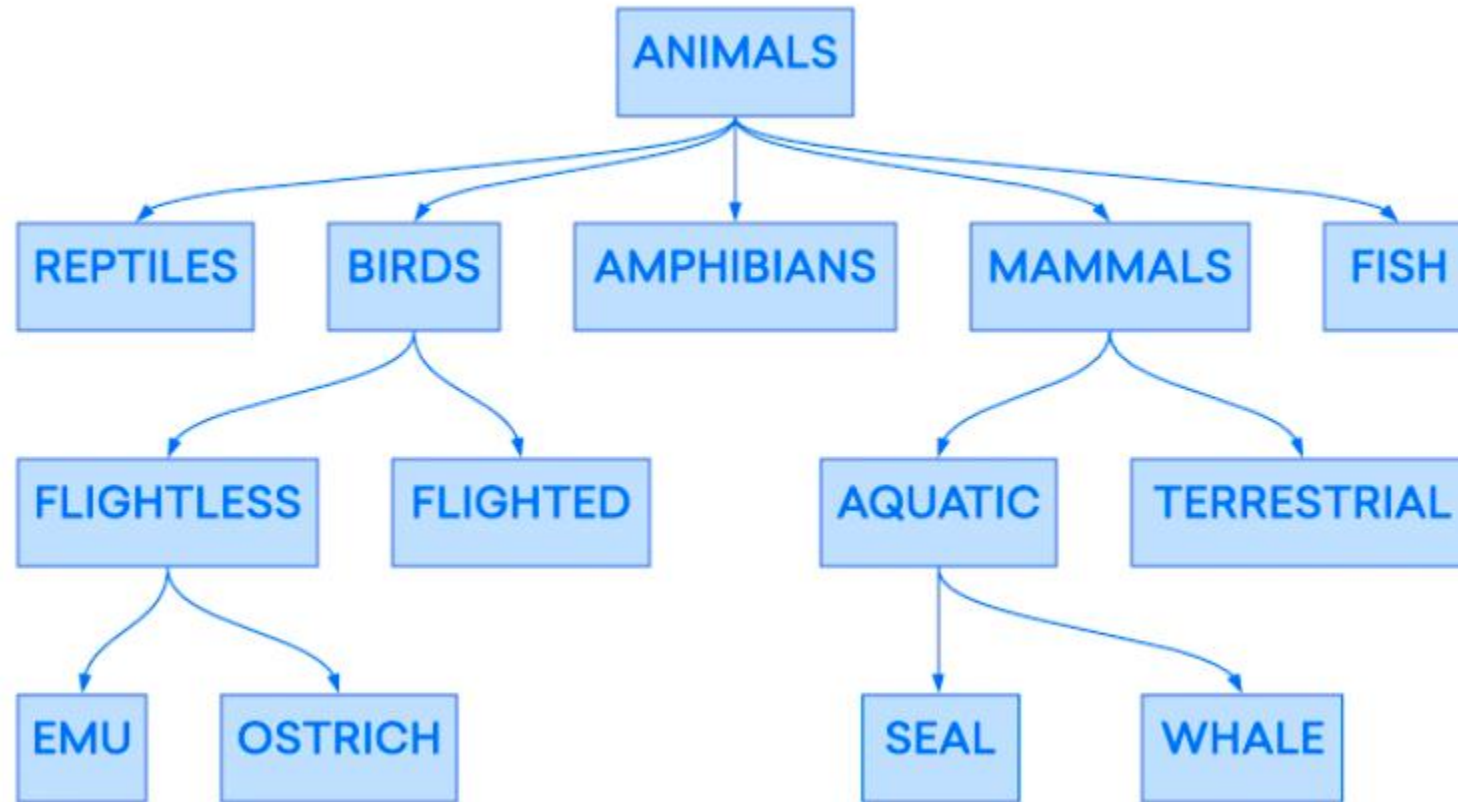
Red

Transaccionales

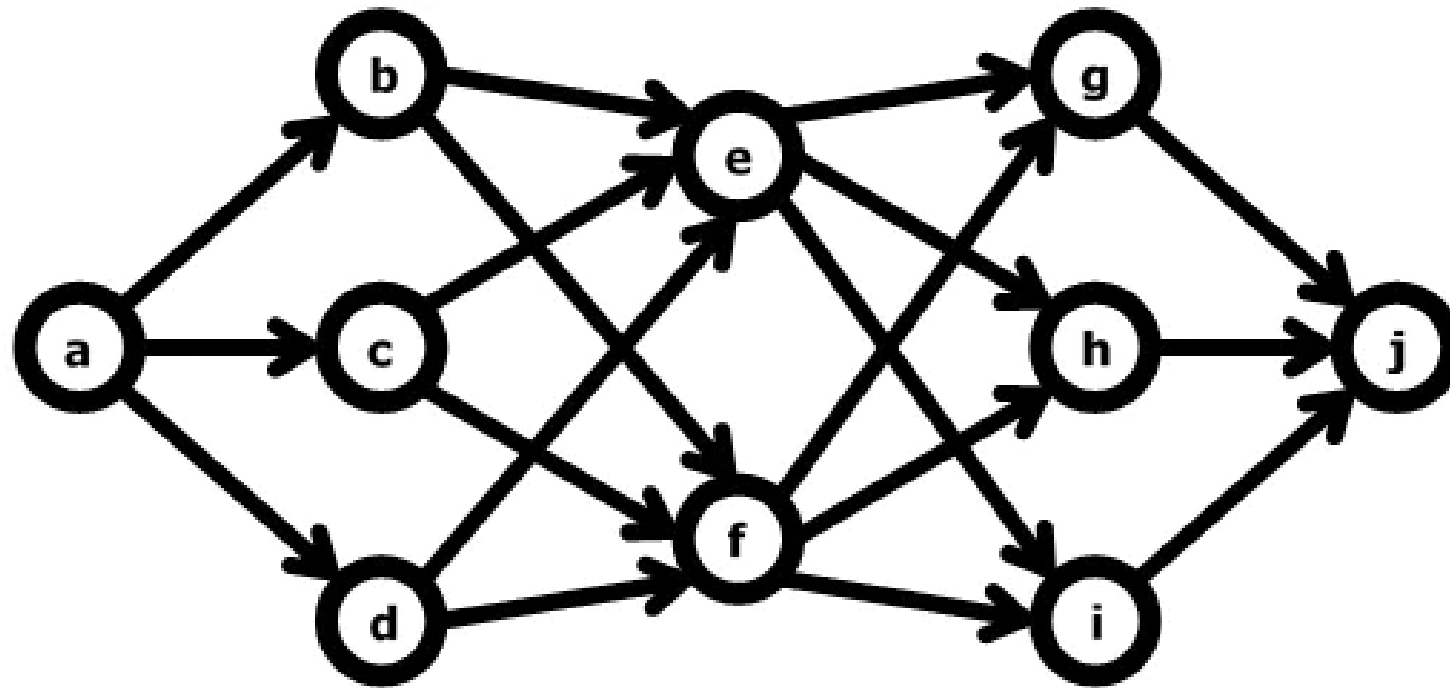
Relacionales

Multidimensionales

Bases de Datos Jerárquicas



Bases de Datos en Red



Bases de Datos Transaccionales



Bases de Datos Relacionales

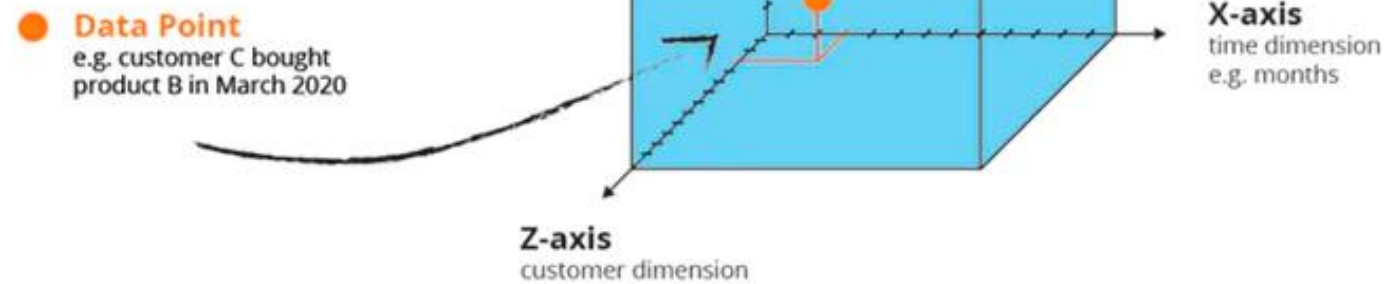
RELATION

TUPLE

ATTRIBUTE

Bases de Datos Multidimensionales

OLAP Cube with 3 dimensions



Tipo Relacional

Optimizar
Almacenamiento de
Información

Necesidades
Operacionales

Datos en general
desagregados.
Estructura relacional

Vs.

Propósito

Responden

Estructura

Tipo Multidimensional

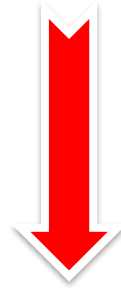
Optimizar Consulta de
Información

Necesidades
Informacionales

Datos en distintos
niveles de detalle y
agregación.

Visión
multidimensional

Data Warehouse



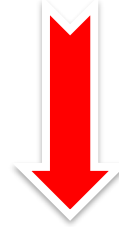
Colección de datos, para el soporte del proceso de toma de decisiones



**Orientada
al Negocio**



Integrada



**Variante
en el
Tiempo**

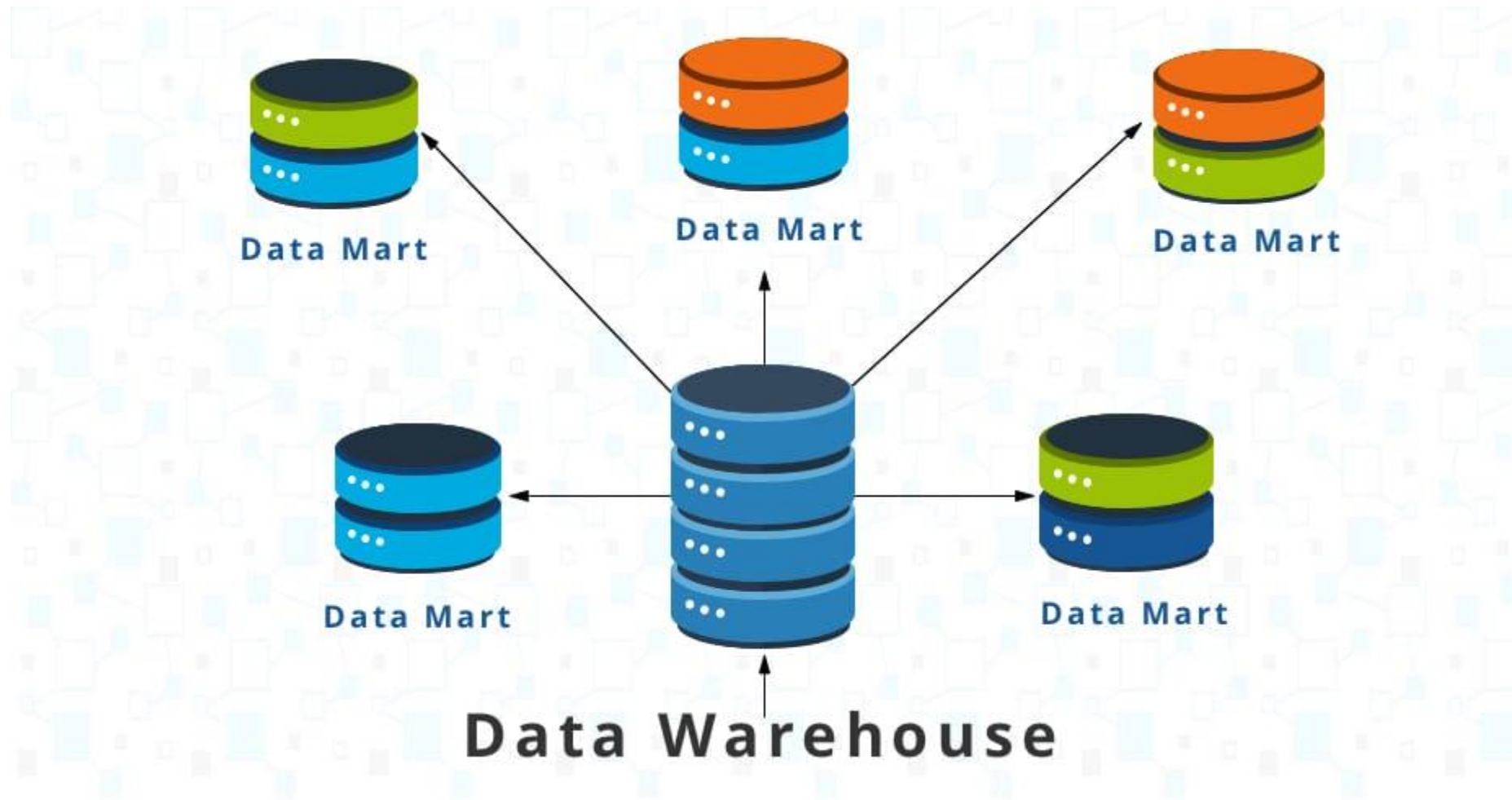


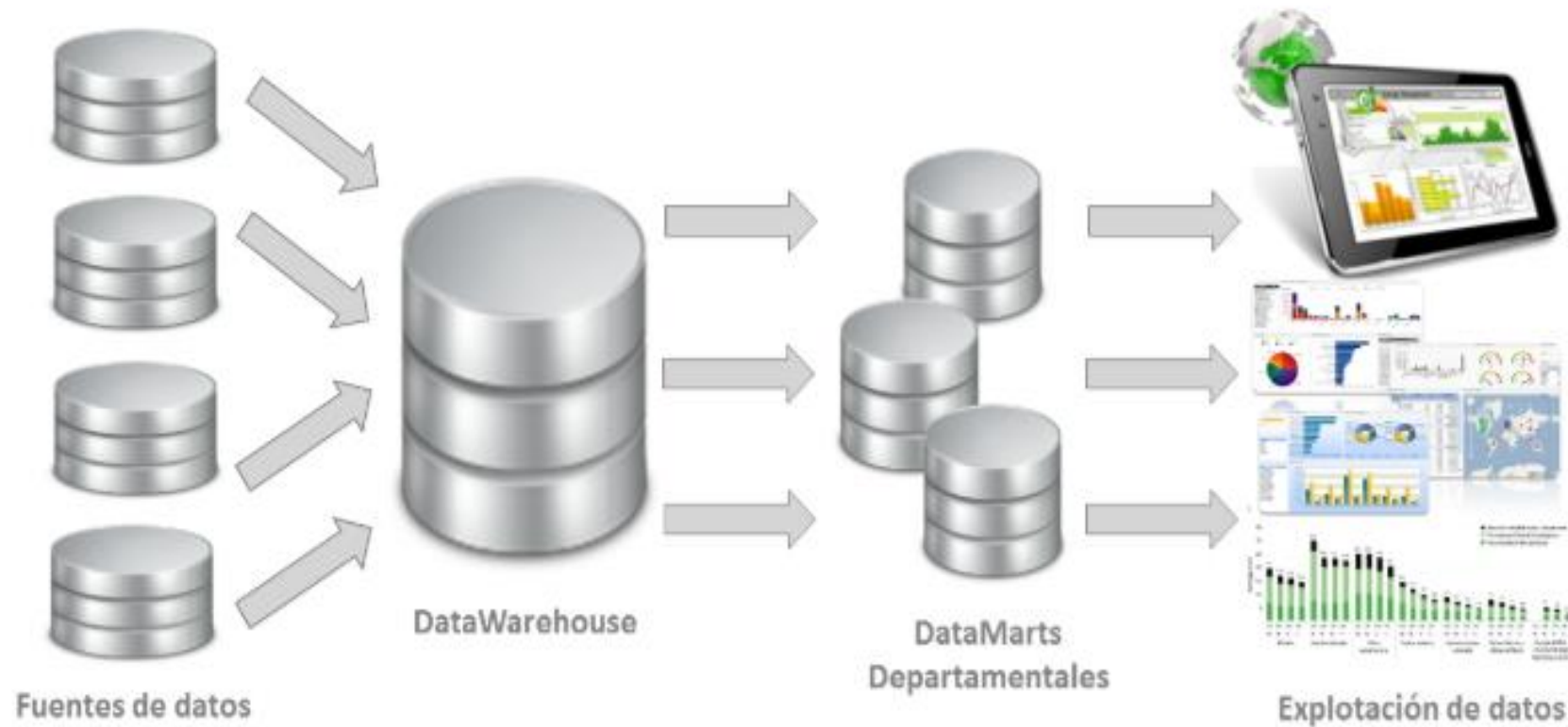
No volátil

OBJETIVOS DEL DATA WAREHOUSE



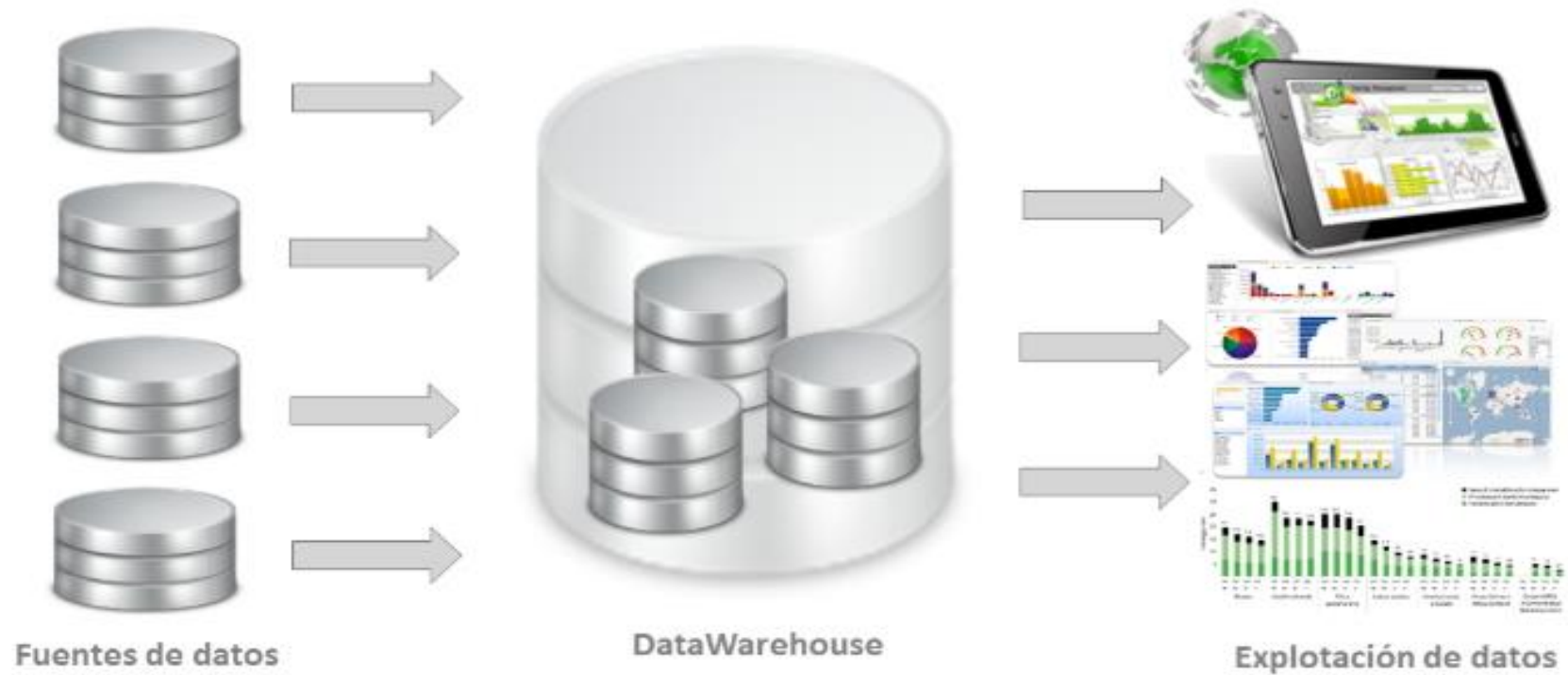
- Simplificar el acceso a los diversos datos de la organización (ERP, CRM, Call Center, etc.) integrándolos.
- Integrar y consolidar diferentes fuentes de datos (internas y/o externas), en una única plataforma sólida y centralizada.
- Hacer que la información de la organización sea consistente
- Adaptarse a los cambios
- Ayudar a entender qué está pasando, por qué está pasando, y predecir qué va a pasar





Enfoque Inmon (Top Down approach)

Referencia: <https://blog.bi-geek.com/arquitectura-enfoque-de-william-h-inmon/>



Enfoque Kimball (Bottom Up approach)

Referencia: <https://blog.bi-geek.com/arquitectura-el-enfoque-de-ralph-kimball/>

Modelos de Datawarehouse

	Inmon	Kimball
Presupuesto	Coste inicial alto	Coste inicial bajo
Plazos	Requiere más tiempo de desarrollo	Tiempo de desarrollo inferior
Expertise	Equipo con especialización alta	Equipo con especialización media
Alcance	Toda la compañía	Departamentos individuales
Mantenimiento	Fácil mantenimiento	Mantenimiento más complejo



BREAK

MATERIAD
Digital School

Dimensiones / Atributos

Son los valores cualitativos. Proporcionan descripciones a los hechos, aportando un contexto a los mismos.

Ej: Marca de coche, fecha, nombre concesionario, dirección de la empresa, nombre del colegio, etc

Hechos / Indicadores

Son las métricas, normalmente valores cuantitativos (numéricos) susceptibles de ser agregados.

Ej: La cantidad de ventas de coches de un concesionario, el rendimiento en euros de una empresa, el número de estudiantes de un colegio, etc.

A tener en cuenta de las Dimensiones

Grupo de atributos, separados y distintos unos de otros (no se comparten)

Generalmente tienen un número bajo de registros.

Cada registro puede contener un gran número de atributos.

Dan el contexto a los eventos :

¿Qué?, ¿Quién?, ¿Cuándo?, ¿Dónde?, ¿Cómo? y ¿Por qué?

Tienen una clave subrogada como PK.

Contienen redundancia.

Claves Subrogadas

Toda dimensión tiene una columna que actúa como clave única primaria (PK)

No tiene ningún sentido específico de negocio, es un valor numérico secuencial

Facilitan la integración de los diversos sistemas de la organización.

Facilitan el tracking de los cambios en los atributos.

Permiten manejar los nulos.

A tener en cuenta de las tablas de Hechos

Clave principal compuesta por los claves principales de las tablas de dimensiones

Registra medidas o métricas de un evento específico.
Ejemplo: cliente compra un geranio de maceta de 25cm en floristería mineral vegetal Lola a las 12:30 am del 10 de Octubre de 2027

Evita repetir de manera completa los atributos dimensionales.
En la TH sólo irá un ID de la dimensión

Se diseñan según el nivel de granularidad deseado, pudiendo registrar eventos a un gran nivel de atomicidad

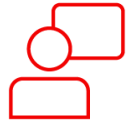
Diseñar un modelo Dimensional



Identificar un proceso de Negocio



Identificar Granularidad



Identificar Dimensiones



Identificar Hechos



Construir un esquema

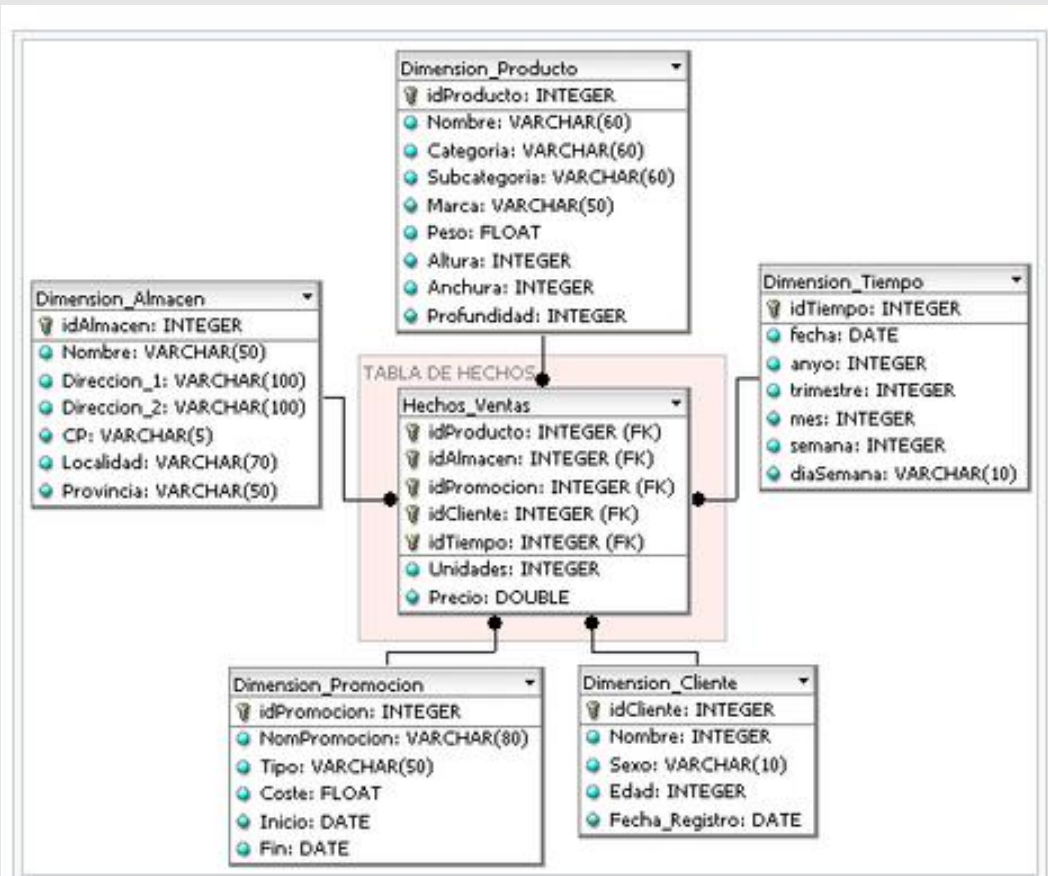
Modelos Dimensionales

Modelo Estrella

Una tabla de Hechos rodeada de las tablas Dimensiones.

Las dimensiones tendrán una clave primaria simple.

La tabla de Hechos la clave principal estará compuesta por las claves principales de las tablas dimensionales.



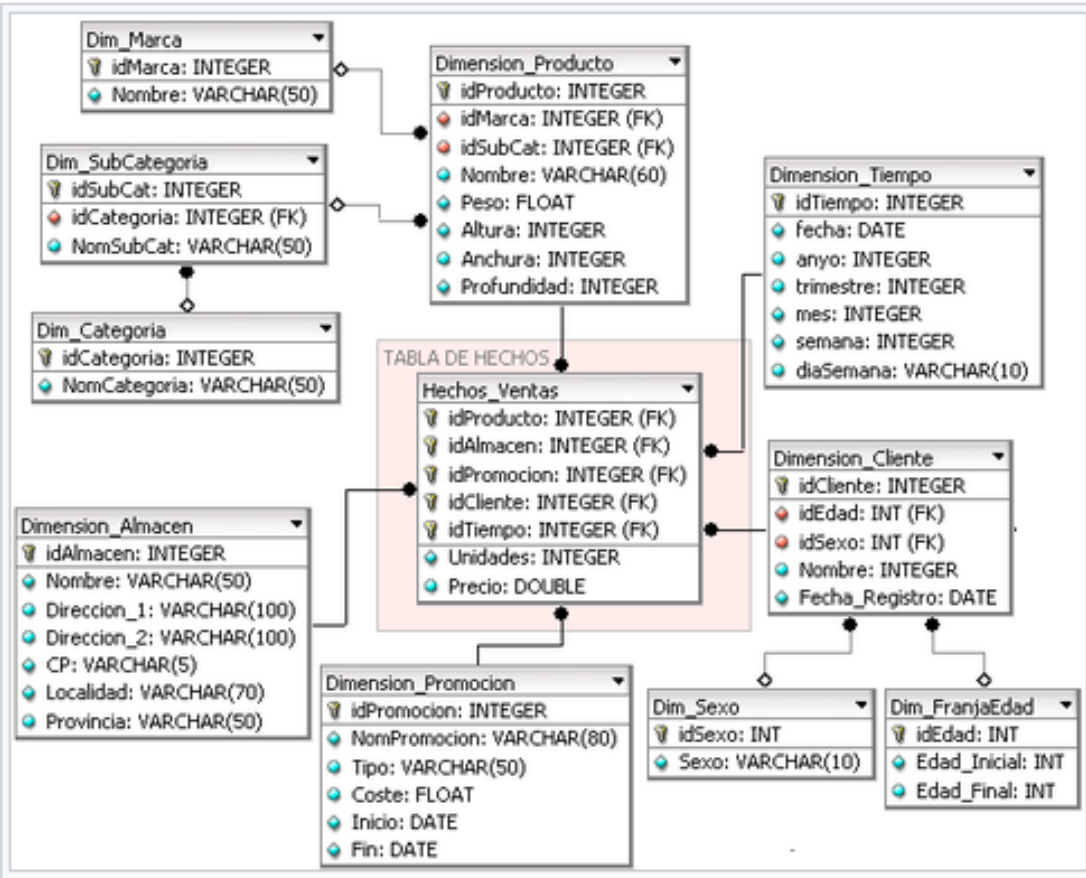
Modelos Dimensionales

Modelo Copo de Nieve

Estructura más compleja.

Dimensiones implementadas con más de una tabla.

Finalidad Normalizar las tablas y así reducir el espacio de almacenamiento al eliminar la redundancia de los datos



	Estrella	Copo de nieve
Mantenimiento	Tiene redundancia. Peor mantenimiento	No hay redundancia. Fácil mantenimiento
Facilidad de uso	Queries menos complejas. Fácil uso	Queries complejas. Difícil de entender
Rendimiento de las queries	Ejecuciones más rápidas	Más tiempo de ejecución debido a los cruces
Tipo de DWH	Data Mart	Data Warehouse
Joins	Bajo número de joins	Alto número de joins
Tablas de dimensión	Una tabla de dimensión por cada dimensión	Más de una tabla de dimensión por cada dimensión
Cuándo usarlo	Cuándo las tablas de dimensión tienen pocas filas	Cuándo las tablas de dimensión tienen un tamaño bastante elevado
Normalización / De-Normalización	Tablas de dimensiones y de hechos denormalizadas.	Tablas de dimensión normalizadas. Tablas de hechos denormalizadas.
Modelo de datos	Top-Down	Bottom-up

Modelo Estrella vs Copo de Nieve

Referencia: <https://blog.bi-geek.com/modelo-dimensional/>

Tips / Reglas para el modelo Dimensional

Cargar datos atómicos en estructuras dimensionales.

Construir modelos dimensionales en torno a los procesos empresariales.

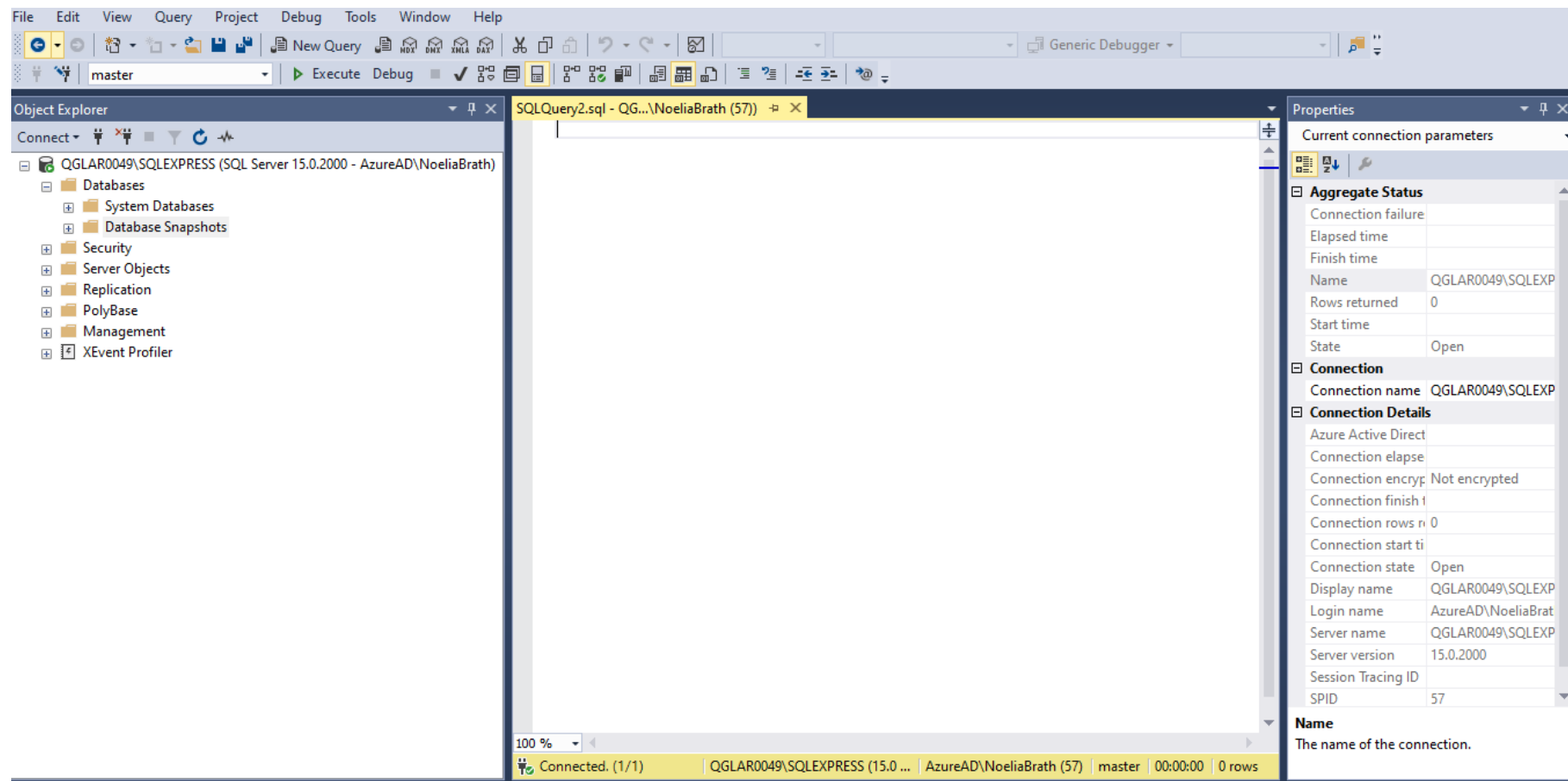
Asegurarse de que cada tabla de hechos tenga una tabla de dimensión de fecha asociada.

Asegurarse de que todos los hechos en una sola tabla de hechos estén en el mismo nivel de detalle.

Asegurarse de que las tablas de dimensiones utilicen una clave sustituta

Balancear continuamente los requisitos y las realidades para brindar una solución comercial que respalde su toma de decisiones

Creación de una Base de Datos





PREGUNTAS

MATERIA D
Digital School



**¡GRACIAS POR ESTUDIAR
CON NOSOTROS!**

