



Clase 23. Curso SQL

# ***DATAWAREHOUSE Y BUSINESS INTELLIGENCE***

***RECUERDA PONER A GRABAR LA  
CLASE***





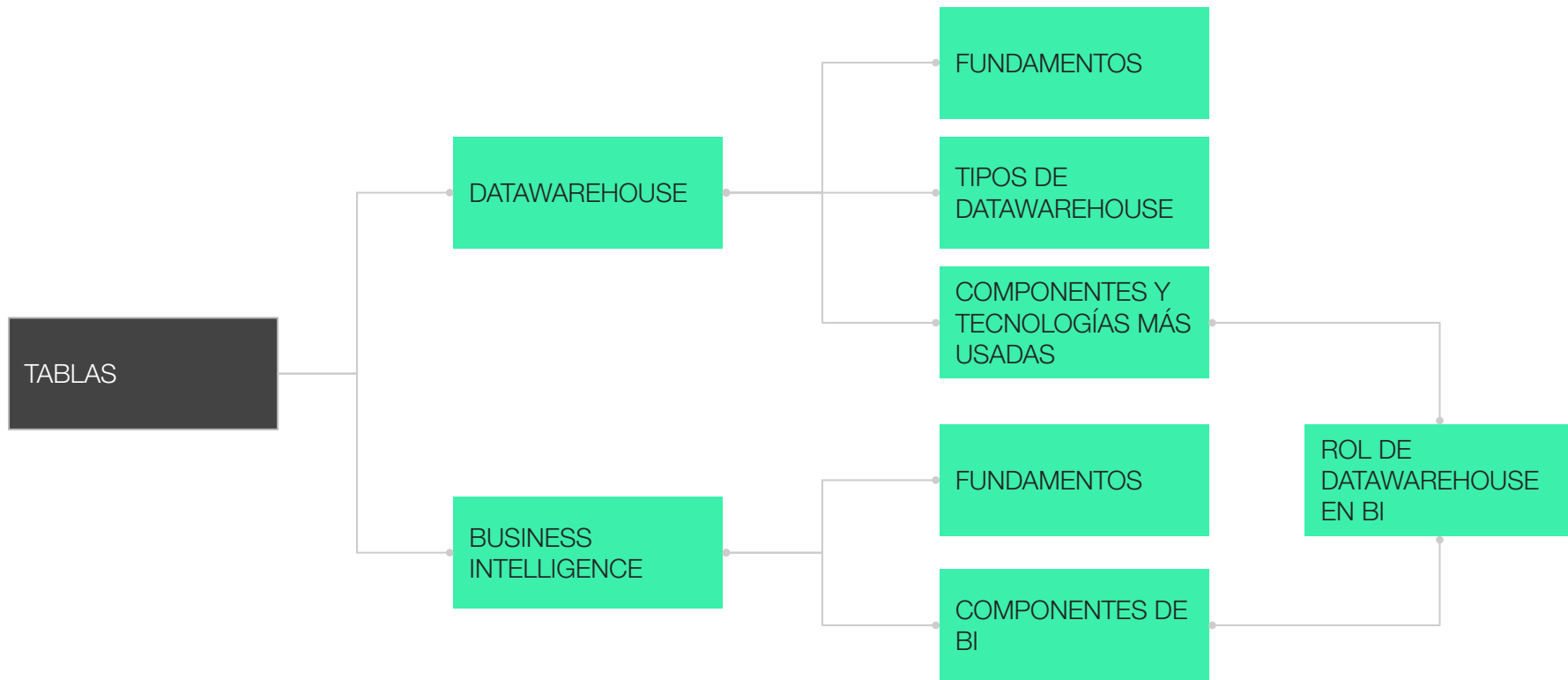
## ***OBJETIVOS DE LA CLASE***

- Identificar un Datawarehouse y sus componentes.
- Identificar los componentes del Business intelligence.
- Reconocer el rol de datawarehouse y business intelligence

# ***MAPA DE CONCEPTOS***

# MAPA DE CONCEPTOS CLASE 23

¡Para  
recordar!



# ***DATAWAREHOUSE***



En esta clase abordamos dos elementos fundamentales: **Datawarehouse** y **Business Intelligence**.

Ambos se complementan muy bien y trabajan junto a otras herramientas para que, cualquier negocio, alcance la excelencia de la mano de la inteligencia.

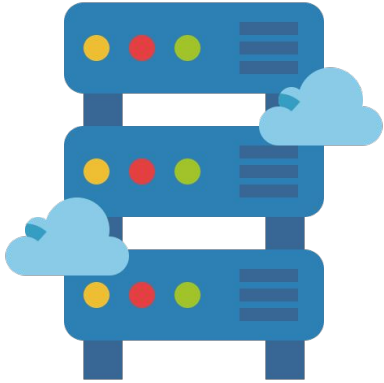
# ***DATAWAREHOUSE***

# ***FUNDAMENTOS***

***CODER HOUSE***



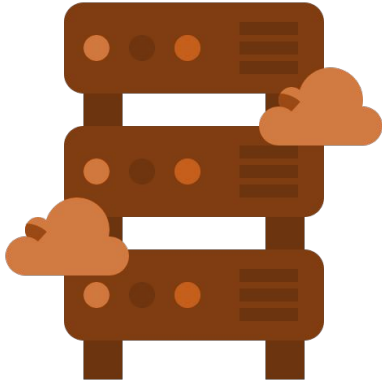
# ***DATAWAREHOUSE***



Denominamos a **Datawarehouse** (DWH) a un espacio común que concentra y almacena de forma estructurada toda la información obtenida a partir de diferentes fuentes de datos en una empresa.

Sistemas de gestión administrativa, de compras, contables, de rr.hh., CRM, ERP, entre otros, pueden ser propietarios o estar desarrollados con diferentes tecnologías...

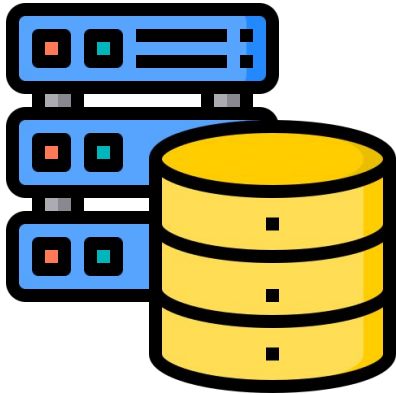
# ***DATAWAREHOUSE***



... y es allí donde el **Datawarehouse** brinda herramientas que permiten concentrar toda esta diversidad de información, en un único punto en común.

Además de preparar el terreno para cualquier Auditoría de Calidad, un Datawarehouse aporta también unificación de información para su análisis mediante BI.

# ***ESTRUCTURA DE ALMACENAMIENTO***



La estructura de un Datawarehouse se compone por diversos fragmentos derivados de éste, los cuales son conocidos como **Data Marts**.

A su vez, cada Data Mart está diseñado para almacenar la información bajo una topología conocida como **Modelo Estrella** o **Modelo SnowFlake**.

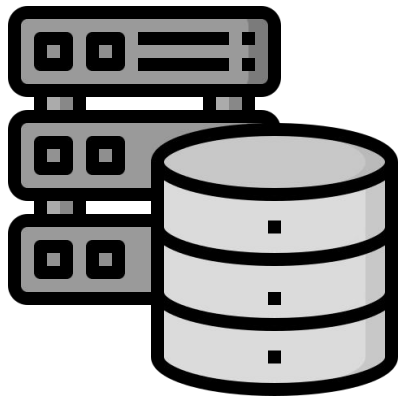
# ***ESTRUCTURA DE ALMACENAMIENTO***

El diseño de la estructura del Datawarehouse **nos permitirá**  
**explotar la información** acumulada en éste **para**  
**diferentes propósitos**, como ser:

- la **generación de reportes estadísticos**,
  - toma de **decisiones basadas en evidencia**,
  - **análisis de información mediante cubos OLAP**
  - y **minería de datos**,
- (entre otros tantos factores).

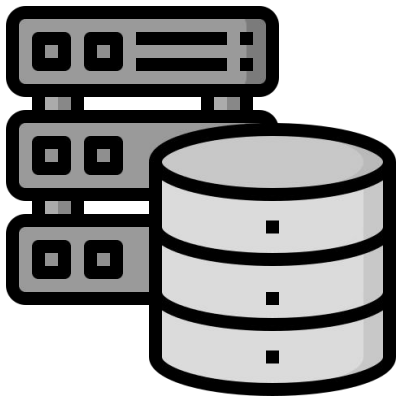
# ***OBJETIVOS DEL DATA MART***

# ***OBJETIVOS DEL DATA MART***



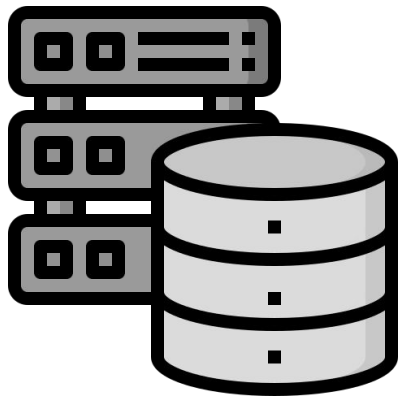
Como bien dijimos, un Data Mart es un ambiente aislado de una o más DB de una empresa y, generalmente, cada punta de la estrella o snowflake hace referencia a un área específica de la empresa (*ventas, Finanzas, Marketing, etcétera*).

# ***OBJETIVOS DEL DATA MART***



Y al estar cada una de estas áreas aisladas, concentrando la información que les compete, se hace mucho más fácil minar la misma para conseguir datos fehacientes que puedan ser aplicados mediante inteligencia de negocios.

# ***OBJETIVOS DEL DATA MART***



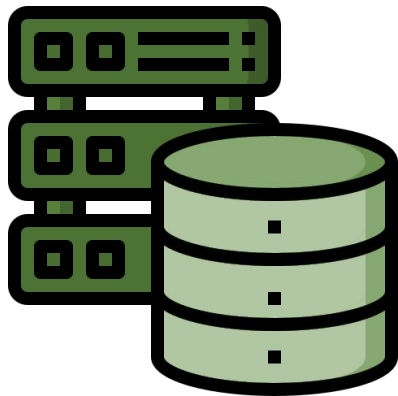
A su vez, podemos trabajar con datos estáticos, dado que en un entorno de producción la información es generada de forma constante.

Por lo tanto, nunca conseguiremos nivelar la información procesada, además de generar tráfico innecesario y posible ralentización del servicio en áreas productivas.



# ***TIPOS DE DATAWAREHOUSE***

# ***TIPOS DE DATAWAREHOUSES***



Existen tres tipos de **Datawarehouses** en el mercado corporativo:

- Datawarehouse empresarial
- Almacén de datos operativo
- Mercado de datos

# ***DATAWAREHOUSE EMPRESARIAL***



Un **Datawarehouse empresarial** se estructura a partir de DB centrales donde la información se organiza, clasifica y **usa para la toma de decisiones**. A su vez, **etiquetan y clasifican la información** para facilitar su acceso.

# ALMACÉN DE DATOS OPERATIVOS



Un **Almacén de datos operativos** (*ODS, por sus siglas en inglés*) es el escenario ideal para actividades diarias.

A diferencia de un datawarehouse empresarial, un ODS se actualiza en tiempo real y almacena datos específicos del segmento corporativo.

# ***MERCADO DE DATOS***

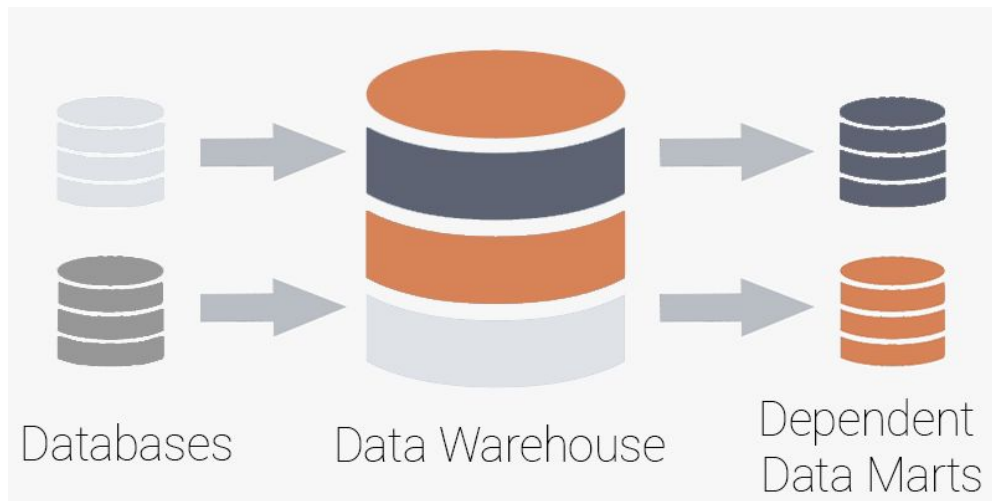


El **Mercado de datos** es parte del Datawarehouse, y está diseñado para dar soporte a los Departamentos o Equipos Empresariales, tomando la información real, depurándose y almacenando en el Datawarehouse para su posterior uso.

# ***COMPONENTES***

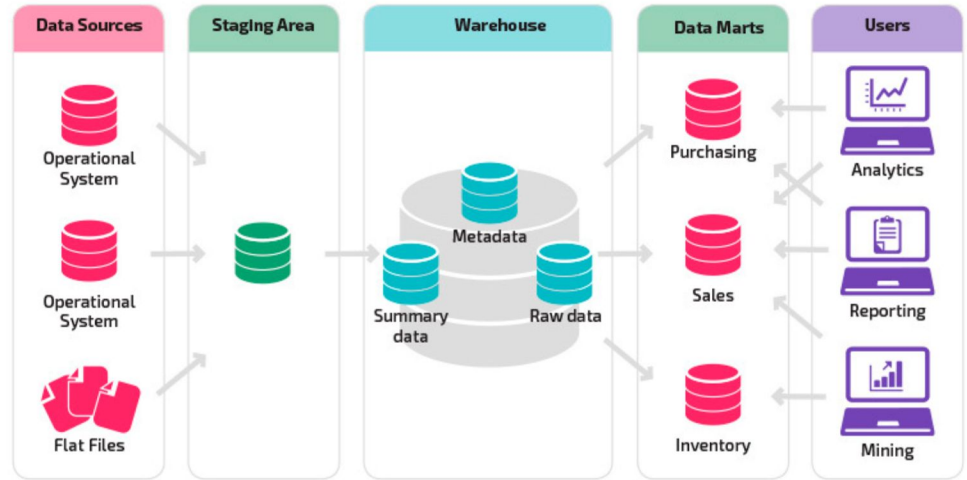
# ***DIAGRAMA DE INTERACCIÓN***

**De izquierda a derecha:** las DB productivas abastecen a un **Datawarehouse** quien, a su vez, provee información a cada **Data Mart** segmentado.



# DIAGRAMA GENERAL

Aquí, una visión corporativa de un ecosistema de múltiples fuentes de datos, un Datawarehouse, Data Marts y áreas de explotación y análisis de la información.





# ÁREA DE STAGING



Y, para extraer la información de los ambientes productivos y operativos, se utiliza el **área de Staging**.

La misma es un **área intermedia de almacenamiento** de datos, **usada para procesar** los mismos a **través del mecanismo de extracción, transformación y carga**.

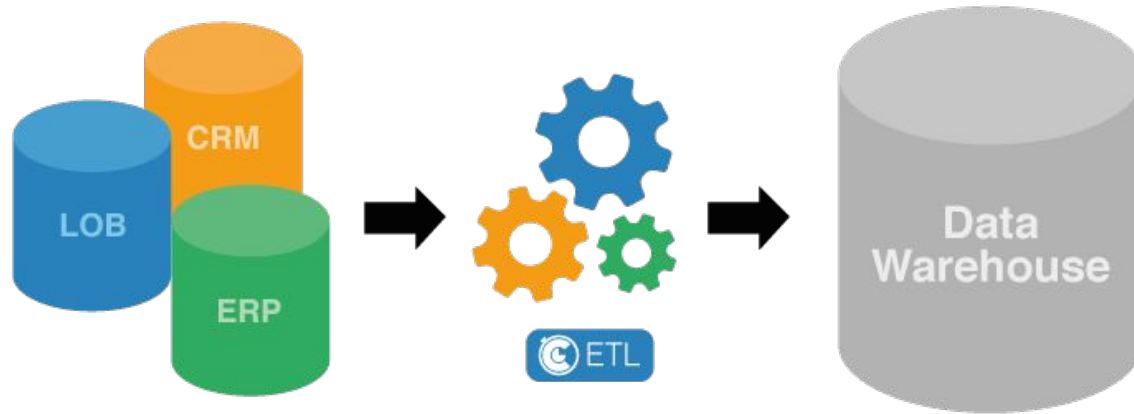
***ETL***

# ***ETL***

**ETL** proviene de la sigla (**E**xtract, **T**ransform and **L**oad).

**ETL** representa el proceso de transferir, dar formato, depurar y finalmente cargar los datos desde múltiples fuentes (las productivas y/u operativas) en un **DWH**.

# ETL



Como mencionamos anteriormente, el área de **Staging** se ocupa de contener los procesos y aplicaciones de ETL, para que la información llegue lo más depurada posible al DWH.

Y, tal como su nombre lo indica, ETL se divide en tres procesos diferentes: **Extracción, Transformación y Carga.**

*¡Veamos en detalle qué realiza **ETL** en cada una de sus etapas de cara a aportar información efectiva a cada uno de los procesos de negocio en una empresa!*

# ***ETL: EXTRACCIÓN***

# **ETL:** *EXTRACCIÓN*

Esta etapa se ocupa de **extraer los datos desde los sistemas originarios**; en este caso pueden ser tanto **bases de datos SQL, NoSQL**, como también **sistemas de archivos planos**.

En esta etapa, además de extraer los datos, ETL **realiza un chequeo de verificación** el cual garantiza que **la información cumple con las pautas o estructura esperada**.

# **ETL: EXTRACCIÓN**

Durante el proceso de extracción, ETL debe causar el mínimo impacto dentro del sistema de origen de los datos.

Si es mucho el volumen de información a extraer, se debe pensar en un proceso separado en ciclos cortos e iterativos a lo largo de la jornada operativa, para la obtención de la información.

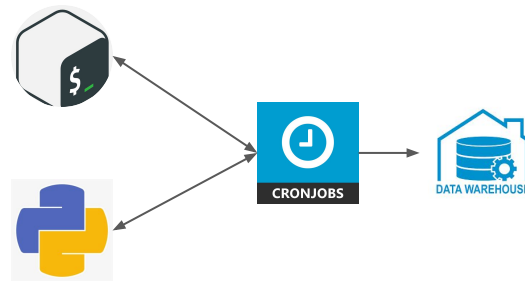


# **ETL: EXTRACCIÓN**



Los procesos cortos e iterativos pueden realizarse a través de archivos batch, programados en un servidor.

Existen varias opciones para realizar este trabajo. Lo más efectivo y usado hoy en día, son procesos programados, que se ejecutan mediante tareas de servidor.



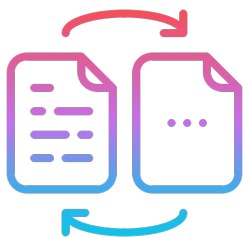
# ***ETL: TRANSFORMACIÓN***

# ***ETL: TRANSFORMACIÓN***

En esta otra fase, **determinados puntos de la información** son convertidos/simplificados en base a una serie de reglas de negocio, que **moldean los datos previo a cargarlos en los Data Marts**.

Aquí se aplican cambios en la fuente de datos, a través de simplificación basada en una convención establecida.

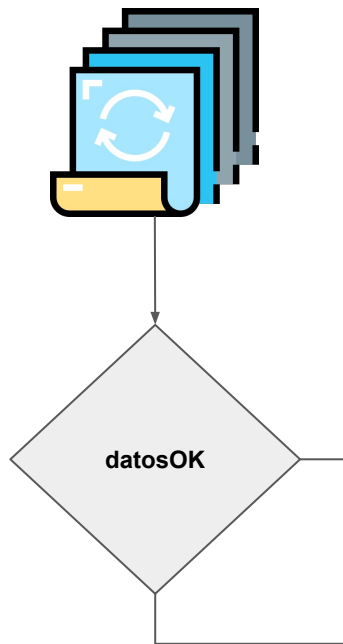
# **ETL:** TRANSFORMACIÓN



Veamos algunos ejemplos (*tentativos*) de simplificación de datos:

- “Hombre” o “Mujer” son convertidos a “H” o “M”.
- Dividir una columna en varias: “Juan Perez Lopez” en “nombre:Juan”, “Apellido1:Perez”, “Apellido2:Lopez”.
- 200 registros de Ventas generados en Marzo en “VentasTotales1Marzo: 200”.
- “True” o “False” son convertidos a “1” o “0”.

# ETL: TRANSFORMACIÓN



- **TRUE:** envío de datos al proceso **ETL: carga**
- **FALSE:** aplica políticas de tratamiento de excepciones (*rechazar la información completa, aplicar **null** en determinados campos, aplicar un valor centinela*)

Se validan los datos y se aplica una acción, según el resultado.

***ETL: CARGA***

***CODER HOUSE***

# **ETL: CARGA**

En esta última fase, se toman los datos provenientes de la transformación, y se cargan finalmente en los sistemas de destino: **Data Marts**.

De acuerdo al tipo de información que se va almacenando, puede que ésta se sobrescriba o que se agregue a información previamente procesada.

# **ETL: CARGA**

Lo que se estila a través de un DWH, es poder mantener un historial de información almacenado actualizado, ya que éste usualmente es consultado por áreas de Auditoría (*interna o externa*).

También **esta información es consultada cuando se realizan procesos de certificación** o recertificación de Normas de Trabajo, como ser (*ISO, IRAM, BPF, etcétera*).



# ***ETL: CARGA***

**Rolling:** representa un **proceso asociado a ETL que aporta niveles de granularidad por jerarquías a la información almacenada.**

La información se guarda de forma resumida en distintos niveles, que corresponden a agrupaciones de unidad de tiempo o jerárquico, según la magnitud de datos contenidos.

# ***¿CÓMO PROCESAR ETL?***



Python es un **lenguaje de programación muy potente** que cubre muchos campos de desarrollo de software, entre ellos, el procesamiento de la información.

**Cuenta con poderosos complementos que nos facilitan crear** con pocas líneas de código, **un programa que realice la entrada, procesamiento y almacenamiento** de datos.



# ETL: CARGA

En **Material Ampliado** te compartimos un video donde se muestra cómo *Automatizar Excel con Python*. Podrás apreciar que, con muy pocas líneas de código, logras trabajar las planillas de cálculo de forma flexible.

```
1  import openpyxl
2
3  excel_files = ['/Users/davidemerlin/Downloads/SampleData.xlsx', '/Users/davidemerlin/Downloads/SampleData2.xlsx']
4
5  for file in excel_files:
6      wb = openpyxl.load_workbook(file)
7      worksheet = wb["SalesOrders"]
8      worksheet['G46'] = '=AVERAGE(G3:G45)'
9      wb.save(file)
```



## ***EJEMPLO EN VIVO***

*Veamos un ejemplo de cómo se aplica la técnica ETL, para obtener información depurada y normalizada, previo a la aplicación de técnicas de BI*



***BREAK***

**¡5/10 MINUTOS Y VOLVEMOS!**

***¡Ya estás llegando al fin de la cursada!  
Recordá que luego de la corrección de tu proyecto  
final, se notificará por Slack y email si quedaste en el  
TOP10***

***No cuelgues, que tenés hasta 2 semanas desde que te  
notificamos para solicitar los beneficios.***

**¡5/10 MINUTOS Y VOLVEMOS!**

***CODER HOUSE***

# ***BUSINESS INTELLIGENCE***

# ***CONCEPTO GENERAL***



# ***BUSINESS INTELLIGENCE***

El desarrollo incesante de las nuevas tecnologías llevó al sector empresarial a experimentar cambios cíclicos constantes.

Entre estos, podemos destacar la comunicación continua con el resto del mundo, la necesidad del **Networking**, la **creación de Alianzas** y **Nuevos Tipos de Negocios**.

# ***BUSINESS INTELLIGENCE***

Para lograr todo esto, **las empresas dejaron de acumular datos por obligación** empresarial-fiscal, y comenzaron a explotar la información de los mismos con el fin de:

- mejorar su productividad
- diseñar nuevos modelos de negocio
- mejorar los existentes

# ***BUSINESS INTELLIGENCE***



Y para que una empresa pueda lograr lo determinado por este nuevo paradigma, lo que conocemos como **Toma de decisiones** debe realizarse con celeridad y precisión, de la mano de herramientas que permitan alcanzar estos objetivos.

# ***BUSINESS INTELLIGENCE***



Allí es donde **Business Intelligence** (o Inteligencia de Negocios), aporta su grano de arena al ayudar a las empresas a que conozcan a sus consumidores, a sí mismas o a sus socios, y apliquen técnicas de mejora continua con el afán de alcanzar sus metas.

# ***BUSINESS INTELLIGENCE***



De esta forma, las empresas aprovechan la **inteligencia de negocios** para crear modelos de datos los cuales les permitan configurar objetivos corporativos para volverse más acertadas al momento de tomar decisiones.

# ***COMPONENTES DE BI***

# ***COMPONENTES EN BI***

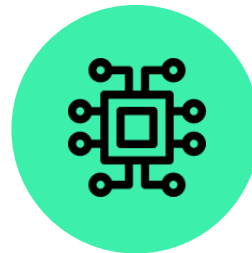
Business Intelligence está compuesto por tres pilares fundamentales:



Herramientas



Técnicas

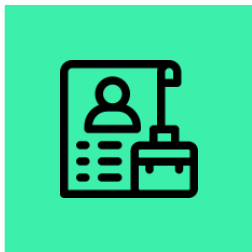


Tecnologías

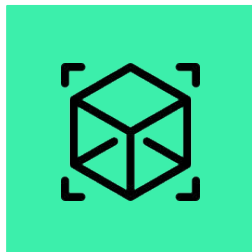
👉 Los tres pilares, en conjunción, brindan el conocimiento necesario aplicable a la toma eficiente de decisiones.

# ***COMPONENTES EN BI***

A su vez, las herramientas se clasifican por complejidad, bajo los segmentos:



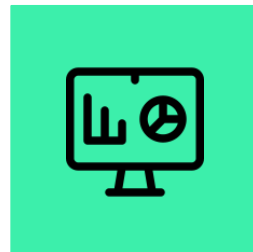
Consultas e  
Informes



Cubos OLAP



Data Mining



Tableros de  
Control



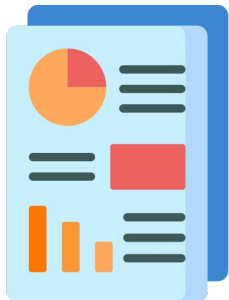
# ***COMPONENTES DE BI***



Gracias a estos componentes, **las métricas otorgan un mayor conocimiento de los factores que influyen en el modelo de negocio** de la organización.

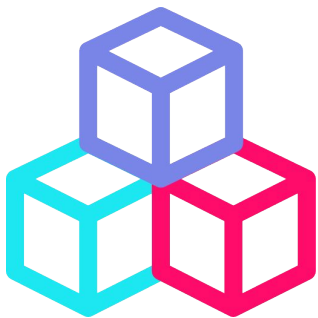
*Veamos a continuación, un detalle de los mismos...*

# ***CONSULTAS E INFORMES***



Presentan un reporte de manera predeterminada pudiendo, como usuario, filtrar la información de éste según parámetros pre-establecidos y decidir entre diferentes criterios de agrupación y ordenamiento de la misma.

# ***CUBOS OLAP***



Es una herramienta que **permite** a los usuarios **consultar información “*en tiempo real*”**, aplicando filtros similares a los que nos brinda una planilla de cálculo a través de las **Tablas Dinámicas**.

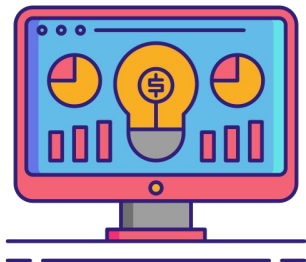
# DATA MINING



Haciendo uso de la “*Minería de Datos*”, **se identifican patrones que no son evidentes a través de los métodos tradicionales** de presentación de la información.

([ver](#) “*historia de pañales y la cerveza*”)

# TABLEROS DE CONTROL



Proponen el **monitoreo de los objetivos estratégicos de la empresa**, mediante las perspectivas (*económico-financieras, clientes, procesos internos, y desarrollo*). Representan la información a través de gráficos interactivos, tacómetros y tablas de datos.



# ***HERRAMIENTAS 2.0***

A tono con la tecnología moderna, encontramos otras herramientas, como ser:

- **Notificaciones Push:**

Emiten alertas según parametrizaciones previas que se activan cuando coinciden con valores establecidos en KPI's.

- **Modelado y predicción:**

Muestran conclusiones confiables de eventos futuros, mediante Estadística, Matemática, y Reconocimiento de patrones.

# ***ROL DE DATAWAREHOUSE EN BI***

# ***DATAWAREHOUSE EN BI***



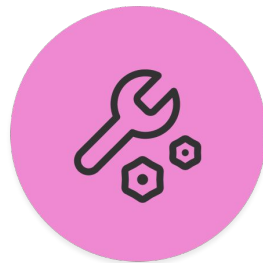
Como vimos a lo largo de los fundamentos de BI, la información que nutre y permite obtener estadísticas basadas en evidencias, es clave para cualquiera de los procesos y/o herramientas que una empresa implemente.



# ***DATAWAREHOUSE EN BI***



Y es así como un DWH se ocupa de aportar las fuentes de datos operacionales a través de sus almacenes de información. Estos registran todos los datos operativos de una empresa que luego serán los que soporten las decisiones estratégicas.



# ***DIAGRAMA DE IMPLEMENTACIÓN DE DWH Y BI***

Realizarás un diagrama en el que se pueda ilustrar la implementación de un DWH y procesos de BI

Tiempo estimado: 20 minutos

# DIAGRAMA DE IMPLEMENTACIÓN DE DW Y BI

Tiempo estimado: 20 minutos

Desafío  
generico

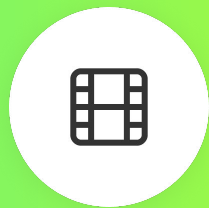


Realizarás un diagrama en el que se pueda ilustrar la implementación de un DWH con, al menos, 2 bases de datos de temáticas relacionadas, y los procesos de **business intelligence** que conforma a cada una de las DB.

Este diagrama puedes elaborarlo en:

- PowerPoint
- [Google Slides](#)
- [Lucid chart](#)

👉 Te compartimos algunos links de material relacionado a [Esquemas en Datawarehousing](#), y [Guía para construir un DWH](#).



***¿QUIERES SABER MÁS? TE DEJAMOS  
MATERIAL AMPLIADO DE LA CLASE***



- [¿Qué es un Datawarehouse?](#) | **Auribox Training**
- [Cómo crear un Datawarehouse](#) | **Josep Curto, Jordi Conesa**
- [Conectar Mysql y Python](#) | **Gerry Studios**
- [Leer archivos Excel en Python](#) | **Mundo Python**
- [Automatizar Excel con Python](#) | **Davide Merlin**

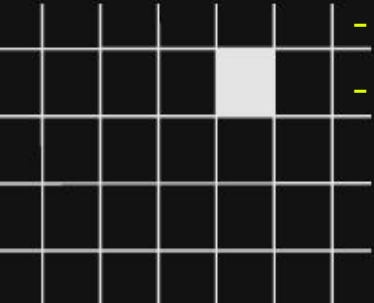
***¿PREGUNTAS?***





# ***¡MUCHAS GRACIAS!***

Resumen de lo visto en clase hoy:

- Fundamentos de Datawarehouse.
  - Componentes de un Datawarehouse.
  - Business Intelligence.
  - Componentes de BI.
  - Herramientas de BI.
- 



***OPINA Y VALORA ESTA CLASE***



***#DEMOCRATIZANDO LA EDUCACIÓN***

***CODER HOUSE***