

CONCEPTOS SOBRE DATA WAREHOUSE

DATA WAREHOUSE

ANDRES SANCHEZ

LUIS FELIPE VELASCO TAO

UNIVERSIDAD DE SAN BUENAVENTURA

BOGOTÁ

11 DE AGOSTO

2021


 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	Conceptos sobre Data Warehouse	2021 - 2
---	---	-----------------------------------	----------

Tabla de contenido

DATA WAREHOUSE.....	4
Definición	4
Características.....	4
Uso	6
Tipos	7
Data Warehouse Empresariales (EDW).....	7
Data Warehouse Operativos (ODW)	8
Data Mart	9
Data Lake.....	10
Data Warehouse Locales	11
Data Warehouse Virtuales o basados en la nube	11
Arquitectura de Data Warehouse.....	12
Fuentes de Datos	12
Consolidación.....	13
Almacenamiento	13
Acceso	14
Explotación.....	15
Modelos de Data Warehouse	15
Modelo de estrella	17
Modelo de copo de nieve	18
Modelo de galaxia o estelar	18
Metodologías de Data Warehouse	19
Kimball e Inmon	19
COMPARACIÓN DE DATA WAREHOUSE Y DATA LAKE	22
BUSINESS INTELLIGENCE.....	24
TIPOS DE DATOS	25
Datos Estructurados	25
Datos No estructurados	26
Datos Semiestructurados.....	27



 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
--	---	---	-----------------

TABLA DE ILUSTRACIONES	29
BIBLIOGRAFIA	30

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

DATA WAREHOUSE

En la actualidad, el acceso a los datos es una necesidad que cualquier empresa debe suplir, dichos datos pueden ser de gran volumen y por ende se requiere de algún mecanismo que pueda permitir el almacenamiento y la consulta de los datos de forma eficiente, aquí es donde aparece el concepto de Data Warehouse.

Definición


Recordemos como funcionaban las empresas anteriormente con relación al manejo de información: grandes bodegas llenas de carpetas con documentos organizados alfabéticamente con la información de sus clientes, proveedores y demás actores o elementos que hagan parte de su actividad comercial. Las empresas se veían regularmente limitadas en tiempo e información ya que solo podían tener acceso a los documentos que se encuentren en cierto lugar, sin poder tener acceso a datos que otras empresas o organizaciones puedan tener, por lo que la información que pudieran rescatar en base a los datos obtenidos era reducida y muy limitada. (Bustamante-Granda et al., 2018; IBM, 2020; Sulmont, 2020; Zeus Data Talk, 2021)

Ahora todo ese mundo lleno de bodegas con miles de papeles migro al mundo digital, en lugar de carpetas de datos nos encontraremos con colecciones de bases de datos que permitan la extracción de la información requerida para las actividades de la empresa. En base a esto, un Data Warehouse o almacén de datos se puede considerar como una base de datos corporativa o una colección de bases de datos por medio de las cuales una o varias organizaciones pueden obtener información de distintas fuentes mucho más fácil, también se puede definir como un sistema el cual el cual permite el acceso de datos de distintas fuentes de forma centralizada, la función principal de un Data Warehouse, en base a estas definiciones, es la consulta de datos dentro de grandes volúmenes de información.

Características

En base a (Díaz Villanueva, 2015; Mendez et al., 2004) se pueden definir las siguientes características de un Data Warehouse o almacén de datos:

- Un Data Warehouse se compone de distintas fuentes de datos o en otras palabras, fuentes de datos heterogéneas, entre las cuales pueden encontrarse bases de datos relacionales, no relacionales, sistemas de ficheros o archivos.
- Las fuentes de datos que integran a un Data Warehouse pueden ser bases de datos operacionales (bases de datos las cuales están encargadas de tener registro de las actividades de la empresa, permitiendo la inserción, actualización, eliminación y búsqueda de datos), utilizando los datos que contengan dichas bases de datos en un momento determinado.


 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	Conceptos sobre Data Warehouse	2021 - 2
--	---	---	-----------------

- Las operaciones que se pueden realizar en un Data Warehouse son la carga de los datos iniciales y la consulta de datos (no se permite la actualización de los datos)
- Para acceder a distintas fuentes de datos se utilizan herramientas ETL (Extract – Transform – Load / Extracción – Transformación – Carga) las cuales permitan la limpieza, filtración y transformación de los datos.

Integrado	Variante en el tiempo	No Volatil	Orientado a sucesos
<ul style="list-style-type: none"> • Los datos deben estar almacenados bajo sistemas de medida comunes, esto con el fin de facilitar su análisis 	<ul style="list-style-type: none"> • Los datos almacenados son de distintas fechas o momentos, sin permitir su modificación, esto con el fin de que la información sea autentica y posteriormente analizable. 	<ul style="list-style-type: none"> • Se permite solo el ingreso de nuevos datos, sin requerir la eliminación de anteriores, permitiendo solo la consulta de los datos. 	<ul style="list-style-type: none"> • Los datos se organizan con base a temas esto con el fin de facilitar su comprensión y acceso para los usuarios.

Ilustración 1 Principales características de un almacén de datos

- La principal función de un Data Warehouse será permitir el análisis de datos históricos requeridos para la toma de decisiones de una organización.
- Son medios de almacenamiento no volátiles, es decir, todos los datos almacenados no sufren de modificaciones o eliminación (las eliminaciones que se realizan se dan con el fin de depurar la información, ya sea por la antigüedad de esta o porque ya no se requiere por parte de la organización).
- Un Data Warehouse debe permitir almacenar datos por sucesos de tiempo, por ejemplo, permitirle a una empresa almacenar los datos relacionados con las finanzas de los meses de enero y febrero, esto con el fin de permitir recatar información con el fin de obtener conocimiento de vital importancia en la organización.
- La actualización de los datos almacenados en un Data Warehouse se realiza por medio de políticas apegadas a las necesidades de la organización.


 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

- Un Data Warehouse debe ser transparente para sus usuarios, es decir, ellos deben percibir el sistema como un solo elemento, sin tener conocimiento de la estructura, ubicación o distribución de los datos.
- Un Data Warehouse debe ser escalable, permitiendo añadir las fuentes de datos que requiera la organización y soportando las consultas a estas nuevas fuentes.
- Los almacenes de datos deben tener un soporte multiusuario, de modo tal varios usuarios puedan realizar consultas sobre el almacén y ninguno tendrá problemas al obtener la información que requiera.
- Una de las ventajas en el uso de Data Warehouse es la mejora en el rendimiento para la creación de informes, esto gracias a que se tendrá acceso a grandes volúmenes de información, condesándolos y estructurándolos como o requiera la organización.

Uso

El propósito general de un Data Warehouse será el permitirle a una organización el análisis de grandes volúmenes de datos con el fin de permitirles tomar buenas decisiones, esto lo podemos ver aplicado en el sector financiero el cual hace uso de esta tecnología para el análisis de riesgos, del desempeño de la organización o del mercado, y esto se replica en cualquier organización que requiera realizar análisis o obtener información desde fuentes de gran volumen, alejándose de las actividades de inserción o modificación de los datos, y centrándose en la capacidad de realizar consultas sobre distintas fuentes de datos. (Bustamante-Granda et al., 2018; De Souza, 2019)

Otro uso muy importante en la actualidad para un Data Warehouse será la minería de datos, la cual se encarga de la extracción de información desde grandes fuentes de información, dicha información no es visible a simple vista por lo que se requieren de ciertos procesos, entre los cuales se encuentra la selección de los datos a valorar, la obtención de propiedades de los datos (tendencias, patrones, valores atípicos, etc.) y la aplicación de técnicas de clasificación y valoración sobre los datos, con el fin de satisfacer las necesidades del usuario final de la información explotada. La minería de datos es usada por distintas áreas, desde la salud, el comercio, la astronomía o las ciencias sociales, regularmente para rescatar información requerida en la planeación o predicción de eventos futuros (por ejemplo, actualmente debido a los volúmenes de datos que se manejan sobre el COVID 19 se requiere de la minería de datos para poder administrar de forma optima el comportamiento de la pandemia). (Bustamante-Granda et al., 2018; De Souza, 2019)

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------


Otros campos de la ingeniería de sistemas que se alimenta de los datos almacenados en Data Warehouse son la inteligencia artificial y el Machine Learning, dos apartados que en los últimos años se han posicionado al frente debido a la automatización y digitalización de actividades que regularmente eran trabajos manuales realizados por el ser humano. Por un lado, la inteligencia artificial requiere del análisis de grandes volúmenes de información para poder tomar la decisión más adecuada frente a un evento actual, sin olvidar que el mismo Data Warehouse puede tener integrado para su operación inteligencia artificial, esto con el fin de poder realizar un análisis de los datos mucho óptimo. Por otro lado, el Machine Learning se sigue apoyando en la inteligencia artificial y en el análisis de grandes volúmenes de datos, pero todo esto con un fin predictivo, es decir, cualquier maquina o dispositivo de Machine Learning siempre estará en búsqueda de información de la cual pueda aprender, prepararse para eventos futuros y mejorar de forma autónoma, sin ninguna intervención humana. (Bustamante-Granda et al., 2018; De Souza, 2019)

Tipos

Un Data Warehouse se puede clasificar principalmente por características como su uso o ubicación, tal cual como nos lo expresa (Díaz Villanueva, 2015; Grupo Atico34, 2021; Mendez et al., 2004; Mendoza, 2020), a continuación, se describirán algunos tipos de Data Warehouse:

Data Warehouse Empresariales (EDW)

Este tipo de almacén de datos es comúnmente conocido como la base de datos central o clave de una organización o empresa, la cual se le cargan datos en determinadas fechas y por medio de la cual una empresa podrá realizar análisis de grandes volúmenes de datos obtenidos de distintas organizaciones o sectores de la misma organización, todo esto con el fin de analizar la información rescatada y obtener cierto conocimiento que le permita a la empresa tomar las decisiones más optimas para su operación, aquí podemos ver actividades tan importantes como la evaluación y detección temprana de riesgos de las cuales puede depender toda una organización.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

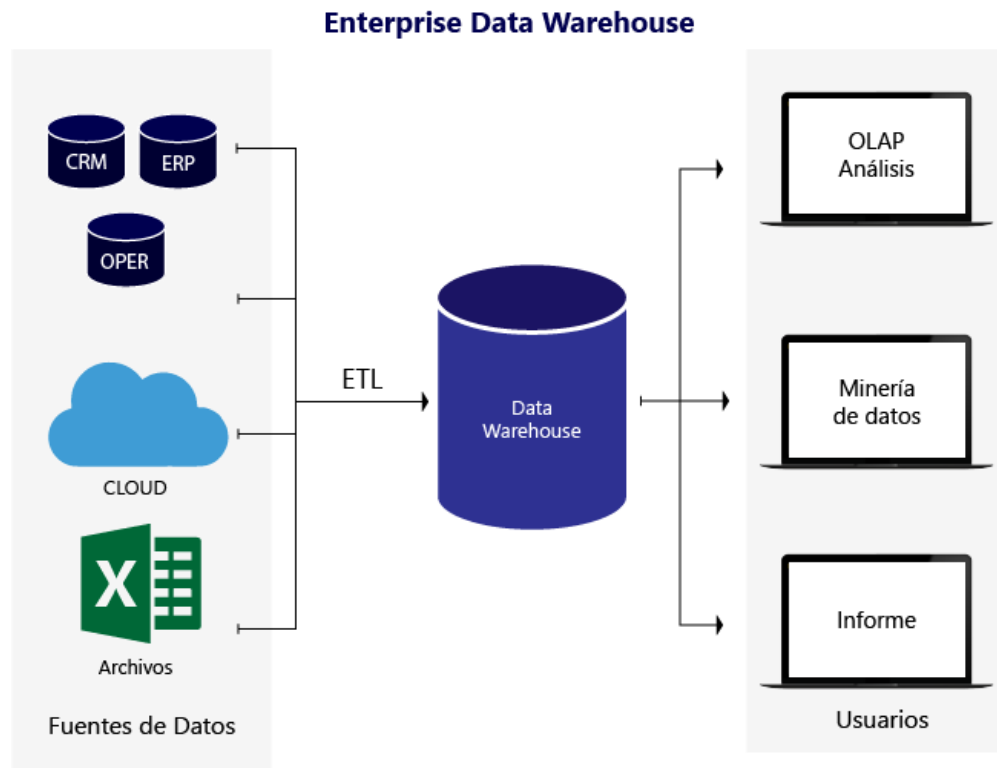



Ilustración 2 Flujo de los datos e información de un Data Warehouse empresarial

Data Warehouse Operativos (ODW)

Como se definió anteriormente, los EDW tienen una carga de datos periódica, pero estos a su vez se alimentan de otros Data Warehouses, tales como los operativos, los cuales no tienen fechas en donde se defina la carga de los datos, sino que estos se van incluyendo en tiempo real de operación del negocio, esto lo podemos ver como el ejemplo práctico de una empresa dedicada a la venta, esta requiere que se almacenen directamente datos relevantes como productos recibidos, vendidos, empleados, novedades, documentos recibidos, y sin fin de datos que se requieren indexar/almacenar de forma inmediata, para posteriormente ser una fuente de los EDW.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

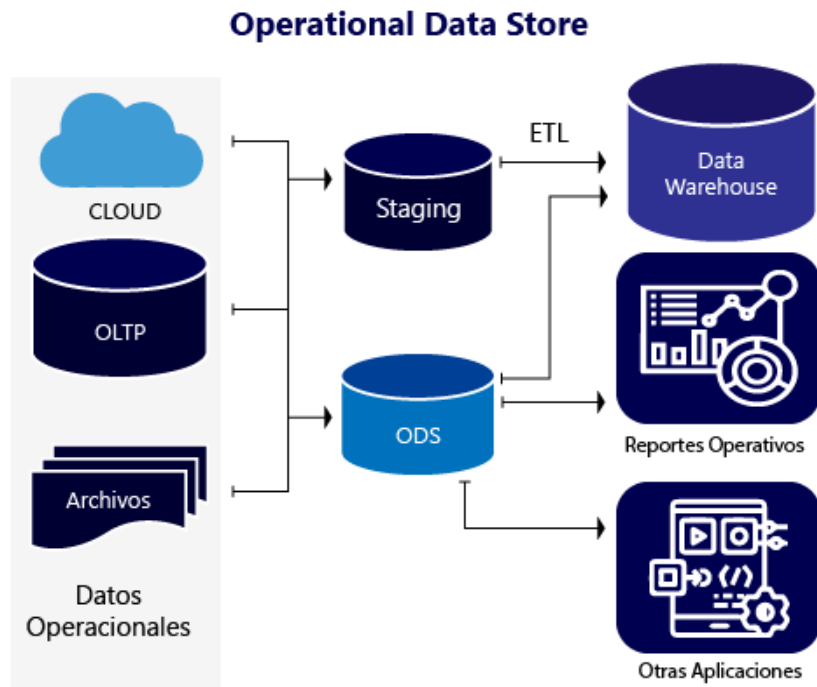



Ilustración 3 Flujo de datos para un Data Warehouse Operativo

Data Mart

Esta es una subcategoría de Data Warehouse (también conocida como bases de información departamental), enfocada en un área en específico de la organización, como por ejemplo, en una empresa se pueden tener Data Marts para los sectores de ventas, producción o empleados, cada uno de estos sectores, recopilando los datos directamente de las fuentes, permitiendo almacenar los datos de forma estructurada y de esta forma realizar análisis que le faciliten a la empresa obtener información de vital importancia, como por ejemplo, detectar problemas en el desarrollo de las tareas del sector y de esta forma tomar decisiones que ayuden a mejorar al sector del cual se alimenta el Data Mart, es resumidas palabras, el objetivo de un Data Mart es permitir a la organización un análisis mucho más detallado de sectores críticos o de vital importancia para el negocio

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

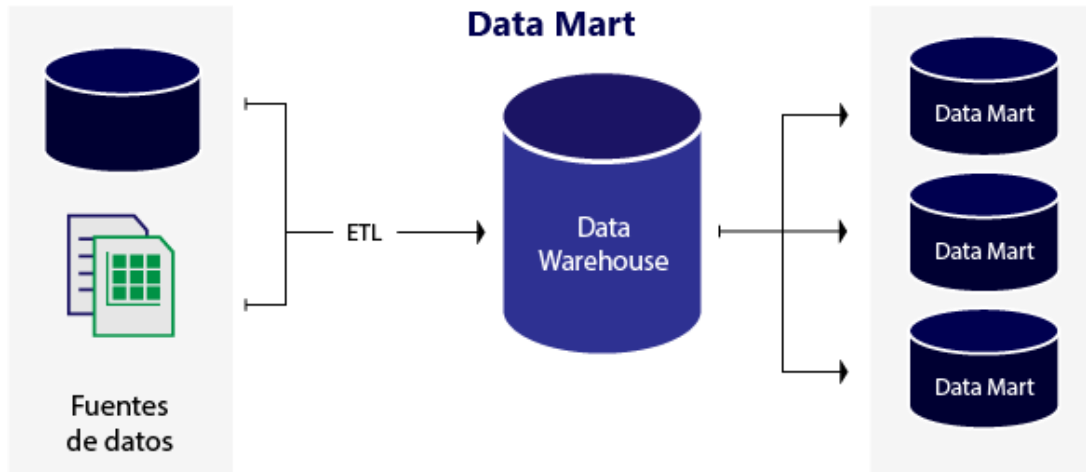



Ilustración 4 Flujo de los datos para la creación de un Data Mart con una Metodología Top-Down

Data Lake

El nombre de este tipo de almacén de datos nos puede dar una idea de como funcionan. Específicamente son un almacén de datos sin ninguna estructura u organización, obteniendo datos estructurados, no estructurados, sirviendo como una fuente de información para otros dispositivos, o para tareas muy específicas. Al contener datos sin ningún orden u organización las tareas de análisis son muy difíciles de realizar, por lo que estos almacenes son simplemente repositorios dedicados simplemente al almacenamiento y consultas de bajo nivel.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

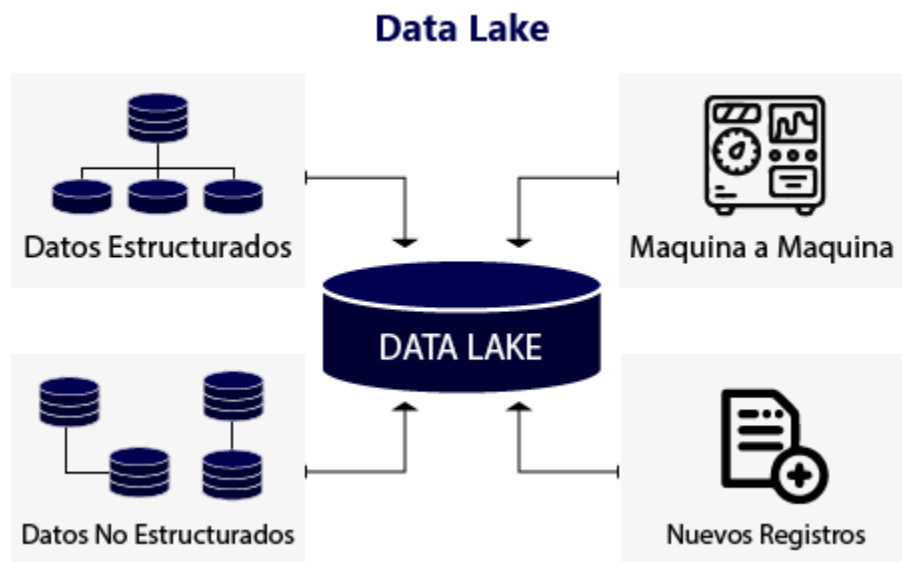



Ilustración 5 Fuentes de datos de un Data Lake

Data Warehouse Locales

Cuando una organización o empresa cuenta con datos ubicados en cierto lugar los cuales sean difíciles de trasladar debido a su tamaño, una gran opción será dotar a ese lugar con la capacidad de almacenar los datos de dichas fuentes de forma estructurada, de modo tal el acceso a los datos para su análisis se pueda realizar solo desde los dispositivos o maquinas con la potencia capaz de realizar las tareas de procesamiento (recordemos que un Data Warehouse no solo debe permitir almacenar los sino asegurar que el acceso a los datos se haga en el menor tiempo y adicionalmente contar con máquinas que permitir el procesamiento de los datos). Este tipo de Data Warehouse requiere de una gran inversión por parte de las empresas con relación a infraestructura, gasto energético, conectividad y capacidad de procesamiento para los datos.

Data Warehouse Virtuales o basados en la nube

En la actualidad sabemos que muchos servicios han migrado a la nube y se ha presentado como una gran oportunidad de negocios el presentarle a las empresas el poder usar infraestructura, plataformas y software que les permitan almacenar, gestionar y procesar los datos requeridos por la empresa, evitando problemas relacionados con la infraestructura, mantenimiento, seguridad y accesibilidad de los datos, ya que todo esto será suplido por el prestador de servicios cloud, dando ventajas como la velocidad en el acceso a los datos, permitir la integración de nuevas fuentes de datos fácil y eficientemente, además de ser mucho más rápido de implementar, aunque se debe contemplar que este tipo de servicio de

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	Conceptos sobre Data Warehouse	2021 - 2
--	---	-----------------------------------	----------

almacenamiento de datos aún esta en una etapa temprana, pero se asegura que será el futuro para las empresas y sus datos.

Arquitectura de Data Warehouse

Para comprender la arquitectura y construcción de un almacén de datos, debemos tener presentes los procesos y elementos que lo componen la forma en que la información se moverá como parte vital en la solución del Business Intelligence. En la siguiente imagen se pueden ver los cinco elementos y actividades requeridas para la construcción de un Data Warehouse. (Abella et al., 1999; Grupo Atico34, 2021; Gutiérrez, 2011; López, 2002; Sierra, 2020)




Ilustración 6 Arquitectura de un Data Warehouse

Fuentes de Datos

Un Data Warehouse requiere de componentes ya existentes en donde residan los datos que se requieran para la organización, desde bases de datos transaccionales, es decir, por medio de las cuales la empresa registra su actividad, de la cual se pueda obtener datos para su posterior análisis, también podemos tomar en cuenta repositorios de archivos o cualquier elemento que contenga datos requeridos para la toma de decisiones por parte de la empresa (Abella et al., 1999; Grupo Atico34, 2021; Gutiérrez, 2011; López, 2002; Sierra, 2020). Podemos definir dos tipos de fuentes de datos dependiendo de donde provengan:

- Fuentes externas: una organización puede requerir para la toma de decisiones datos que generen otras organizaciones, por ejemplo, datos generados por entidades gubernamentales o informes de otras empresas, aquí se pueden ubicar cualquier fuente que posea datos generados por de forma externa a la organización que posee el Data Warehouse.
- Fuentes internas: será cualquier fuente en donde se encuentren datos generados por la misma organización poseedora del Data Warehouse, aquí se pueden mencionar las bases de datos transaccionales, repositorios de archivos de la empresa, archivos o datos relacionadas con historiales de la organización y entre otros componentes que contengan datos relacionados

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

con la operación de la organización y que haya sido generado o creado por la misma.

Consolidación

Esta etapa es de vital importancia, ya que aquí nos encontramos con los procesos ETL (Extract – Transform – Load / Extracción – Transformación – Carga) en donde, ya con las fuentes previamente definidas, se realizarán ciertos procesos de manipulación de datos por medio de los cuales se extraiga la información de las fuentes, se modele o estructure acorde a las necesidades de la organización. Se filtren los datos y finalmente se carguen al Data Warehouse, Las actividades que se realizan en esta etapa siempre estarán enfocadas en generar datos de calidad los cuales aseguren un análisis óptimo para la toma de decisiones de la organización. (Abella et al., 1999; Grupo Atico34, 2021; Gutiérrez, 2011; López, 2002; Sierra, 2020)

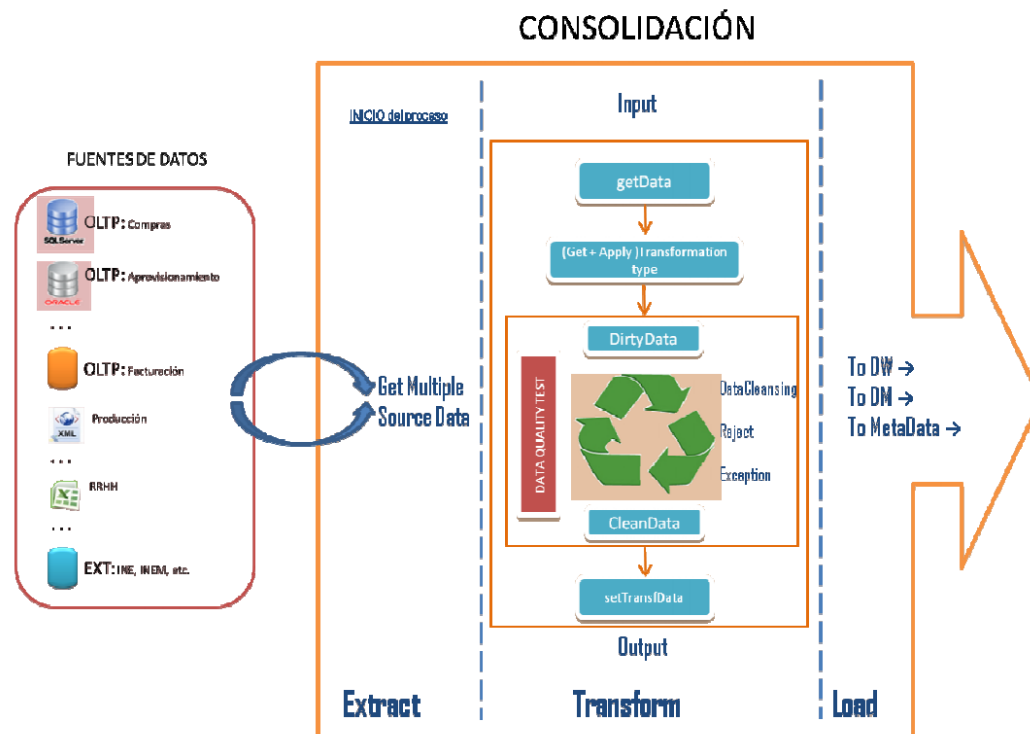



Ilustración 7 Proceso de Consolidación

Almacenamiento

Se puede considerar a este componente como el Data Warehouse en sí, esto es algo que queda a libre interpretación, ya que aquí nos encontraremos con los datos estructurados de modo tal las consultas que se realicen sobre el DW se realicen lo


 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

más rápido posible, esto quiere decir que en la etapa de consolidación los datos se deben cargar de modo tal se puedan generar conexiones entre si y formar de esta manera metadatos o diccionarios de datos, en los cuales se describen todos los datos almacenados, esto también comprenderá las relaciones que puedan tener los datos entre sí, las consultas que se deban realizar para rescatar cierta información, además que se vera la estructura u orden que tendrá la información al ser rescatada del DW. Toda esta explicación se puede hacer la analogía con una fotografía: esta fotografía puede estar almacenada con demás archivos multimedia, como lo son videos, ordenados de cierta forma dentro de una carpeta, la cual entre sí tendrá lazos indicadores relacionados con dichos archivos, como, por ejemplo, la fecha en que fueron tomados o almacenados, la duración de los videos, calidad, modificaciones realizadas y hasta la ubicación en donde fueron tomados los videos y fotos. Esta analogía es muy cercana a un servicio de alojamiento como lo es Google Fotos, el cual actualmente puede rescatar información de nuestros archivos multimedia tan precisa como las personas que aparecen en una foto. (Abella et al., 1999; Grupo Atico34, 2021; Gutiérrez, 2011; López, 2002; Sierra, 2020)

En el componente de almacenamiento se debe contemplar elementos relacionados con la accesibilidad, seguridad y disponibilidad de los datos, se deben definir políticas relacionadas con la carga, manipulación y accesibilidad de los datos, mantenimiento de la infraestructura, previamente a todo esto se debe contemplar la metodología por medio de la cual fluirán los datos (ver [Metodologías de Data Warehouse](#)). Como tareas relacionadas con el almacenamiento, se pueden contemplar los puntos de acceso que se tendrán para el DW, esto para cumplir con la siguiente etapa.

Acceso

Ya tenemos nuestra información organizada y almacenado tal cual lo requiere la empresa para sus necesidades, ahora debemos generar un componente el cual sirva de intermediario entre el usuario que va a explotar la información y el almacenamiento, y en el mundo de los datos ya existe un componente que nos permite realizar esta tarea: los middlewares. Recordemos que un middleware es cualquier software que permite la conectividad entre una fuente de datos y un cliente que consumirá los datos extraídos de dicha fuente, este componente se integra a entornos en donde se debe generar una transparencia sobre un sistema distribuido, en el caso de los DWs o Data Marts se debe asegurar por medio de los middlewares la conectividad entre todos los componentes, permitiendo que el DW sea explotable por una tecnología o sistema en concreto. (Abella et al., 1999; Grupo Atico34, 2021; Gutiérrez, 2011; López, 2002; Sierra, 2020)

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
--	---	---	-----------------


Explotación

La etapa o componente final será la encargada de generarle conocimiento a la empresa con el fin de mejorar la toma de decisiones, aprovechando todos los datos que se encuentren almacenados, la explotación de los datos se puede considerar como el cliente que se alimentara del servidor, que en este caso es el DW. También se puede comprender como un componente en el que encontraremos todas las interfaces por medio de las cuales los usuarios tendrán acceso a la información, dichas interfaces son quienes se encargaran del control de las herramientas encargadas de la presentación de la información, aquí podemos encontrarnos con productos como lo serán informes ejecutivos, modelos de predicción y análisis estadístico, además de ser el punto por medio del cual se podrá realizar minería de datos e implementar tecnologías tan potentes como lo son la inteligencia artificial herramientas OLAP y redes neuronales. (Abella et al., 1999; Grupo Atico34, 2021; Gutiérrez, 2011; López, 2002; Sierra, 2020)


Como se puede ver, la arquitectura o composición de un Data Warehouse conlleva una gran responsabilidad por parte de una o varias organizaciones, ya que es son fuentes de información vitales en procesos empresariales, no solo por permitir realizar consultas de gran nivel, sino al permitir obtener conocimiento de grandes volúmenes de datos, algo que hace algunos años era una tarea casi impensable y hoy en día las grandes empresas requieren para controlar las actividades que desempeñan, los servicios o productos que tengan, sus clientes y demás elementos que de por si son difíciles de analizar individualmente y que al estar enlazados entre si es una tarea obligatoria para la toma de decisiones el obtener conocimiento para la continuidad de la organización.

Modelos de Data Warehouse

En el mundo de los almacenes de datos los datos se estructuran de modo tal existan relaciones que permitan a su vez facilitar las búsquedas, en las bases de datos operativas nos encontrábamos con estructuras que se generaban en base a relaciones las cuales necesitaban estar normalizadas para poder acceder a los datos, en DW nos encontramos con la necesidad de tener un sistema multidimensional, el cual permita un análisis correcto de los datos y que estos puedan ser almacenados con el pasar del tiempo (Berzal, 2016; Díaz Villanueva, 2015; León, 2007; Valbuena, 2018). Antes de explicar los algunos modelos de Data Warehouse, debemos contemplar los elementos por medio de los cuales se componen estos:

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

- Dimensiones: una organización, al momento de implementar un Data Warehouse, debe contemplar las entidades o elementos de intereses sobre los cuales quiere dirigir sus análisis, regularmente se pueden ver modelos multidimensionales asociados al tiempo, la ubicación y un componente de la organización, como lo son los productos, servicios o proveedores.
- Miembros: dentro de cada una de las dimensiones que definamos podemos encontrarnos identificadores que nos ayuden a obtener mas detalles, por ejemplo, usando la dimensión se tiempo, esta se puede componer por días, meses, años, semestres, bimestres, todos estos son miembros o componentes del tiempo que le dan un mayor detalle a las demás dimensiones, por ejemplo, obtener las ventas de un producto en cierta ciudad en ciertas fechas.
- Jerarquías: con las dimensiones y sus miembros definidos debemos comprender que los miembros tienen una jerarquía, dicha jerarquía nos ayudara a comprender el tamaño de las consultas, por ejemplo, no es lo mismo consultar las ventas de una sola ciudad al consultar las ventas de una región, aquí es donde debemos tener en cuenta un orden para los miembros de cada una de las dimensiones.
- Hechos: los hechos se pueden definir como la colección de datos asociados a cada uno de los miembros de cada dimensión, de modo tal se obtenga en contexto de el hecho por medio de las dimensiones que lo componen, un ejemplo de un hecho seria una venta, la cual debe tener especificada como sus dimensiones los productos, el tiempo y la ubicación, pudiendo contener dicha venta otros valores.
- Medidas: como se menciona anteriormente, un hecho puede estar acompañado de otros valores, estos valores son precisamente las medidas, valores o atributos numéricos medibles, apegándonos al ejemplo de una venta, esta se realizo en cierto lugar, en la cual se compro cierto producto a cierta hora, las medidas que pueden acompañar este hecho puede ser la cantidad de unidades compradas o el valor total de la venta.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

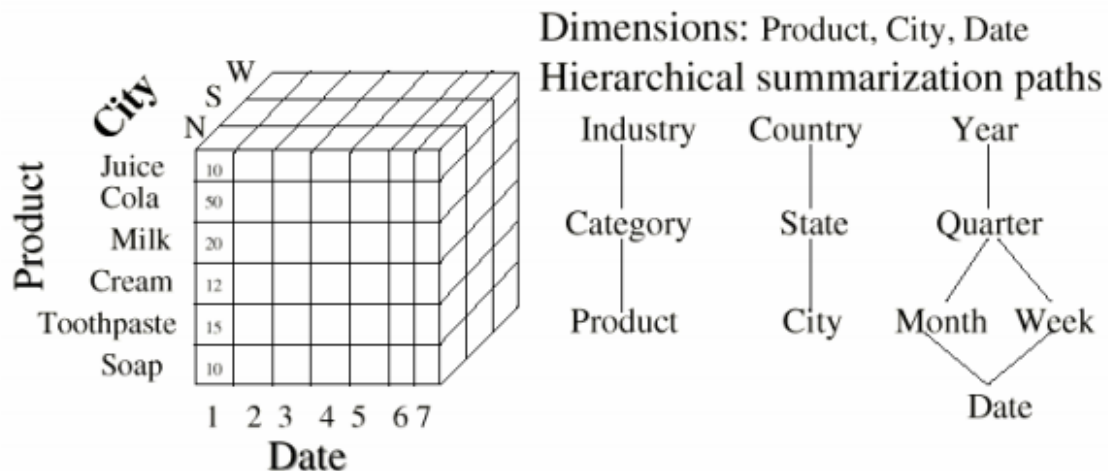


Ilustración 8 Ejemplo de un modelo multidimensional

Modelo de estrella

En este modelo nos encontraremos con una única tabla de hechos en donde se encontrarán los datos para los análisis obtenidos de las tablas de las dimensiones, se puede asimilar como si tuviéramos distintas tablas, de las cuales vamos a extraer su clave primaria para componer con esta la tabla de hechos. Este modelo es uno de los mas simples pero que a su vez esta optimizado para realizar consultas de grandes conjuntos de datos. (Berzal, 2016; Díaz Villanueva, 2015; León, 2007).

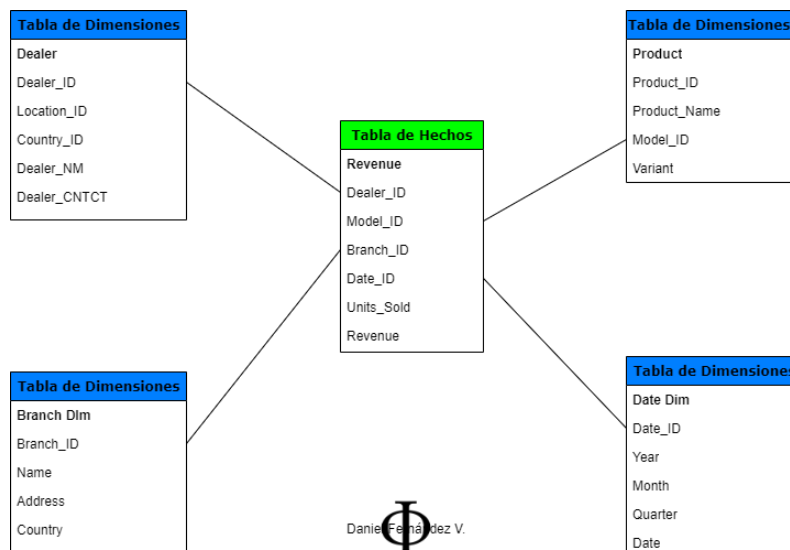



Ilustración 9 ejemplo de un Modelo de Estrella

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

Modelo de copo de nieve

Teniendo presente la estructura de un modelo de estrella, ahora pensemos si a las tablas de dimensiones existentes le agregamos dimensiones adicionales, esto con el fin de obtener consultas mucho más específicas y enriquecidas, sin cambiar nuestra tabla de hechos, así es como podremos obtener un modelo de copo de nieve, el cual se puede considerar como una evolución del modelo de estrella, manteniendo una única tabla de hechos. (Berzal, 2016; Díaz Villanueva, 2015; León, 2007).

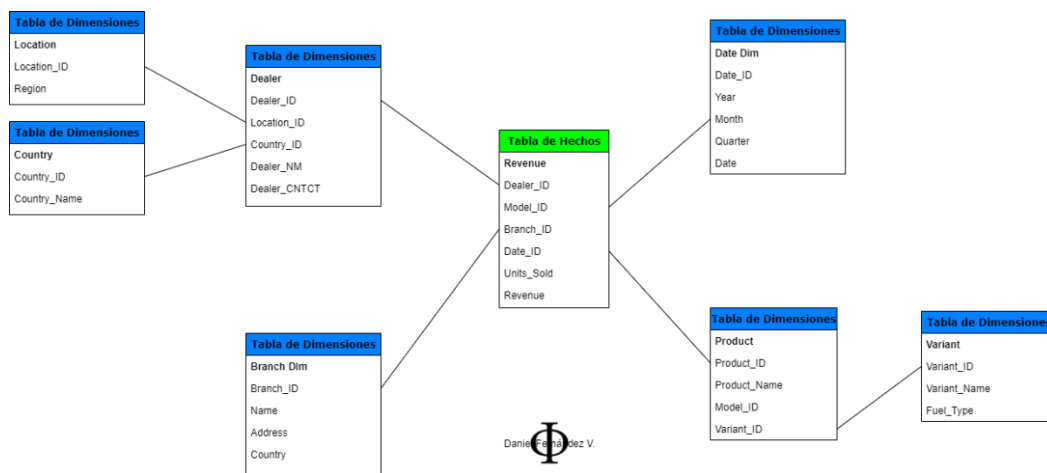

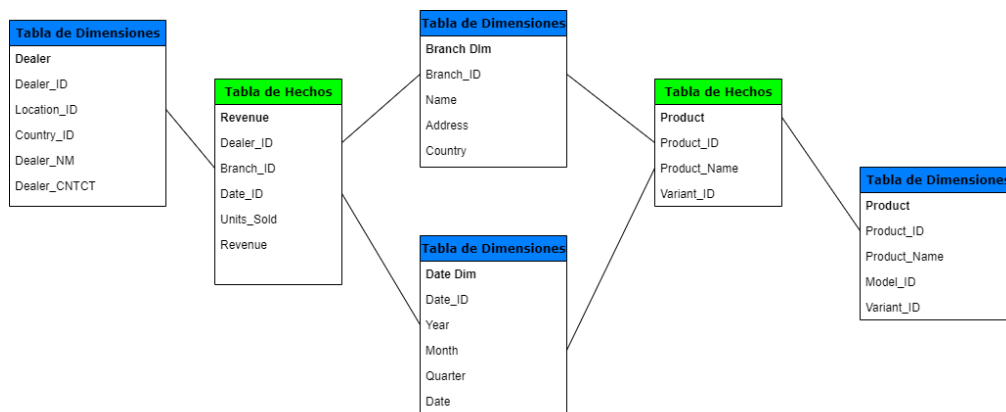


Ilustración 10 ejemplo de un Modelo de Copo de Nieve

Modelo de galaxia o estelar

El nombre de este modelo nos da una idea de cual será su estructura, en la cual tendremos dos o más modelos de estrella los cuales comparten entre si tablas de dimensiones, las dimensiones que comparten los hechos se conocen como dimensiones compartidas, además que este modelo presenta particularidades como lo es el poder agregar tablas de hechos para una mejor comprensión, la construcción de este modelo puede realizarse por medio del estudio de un modelo de estrella para dividirlo en otros modelos de estrella y finalmente, las dimensiones que componen este modelo se dividen en función a los distintos niveles de jerarquía, de modo tal las dimensiones de las cuales se alimenta un hecho pueden ser en realidad la descomposición de una dimensión por medio de sus miembros (Berzal, 2016; Díaz Villanueva, 2015; León, 2007).

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------



Daniel F. Rodríguez V.


Ilustración 11 Ejemplo de un Modelo de Galaxia

Metodologías de Data Warehouse

Al momento de diseñar un almacén de datos nos encontraremos con dos nombres muy importantes en este ámbito, los cuales nos dejaron dos metodologías por medio de las cuales podemos diseñar e implementar en una organización, todo esto dependiendo de la infraestructura y las fuentes de datos, pero antes de analizar cada metodología, debemos conocer a Kimball e Inmon. (Bustamante-Granda et al., 2018; Gravitar, 2020; Gutiérrez, 2011)

Kimball e Inmon

Ralph Kimball y Bill Inmon son dos estándares en el mundo del Data Warehousing, cada uno definiendo por medio de un enfoque el diseño de un DW, los cuales, desde la década de los 90s han sido de gran ayuda para poder comprender la importancia en la forma en que los desarrolladores y organizaciones abordan la implementación de DW en sus entidades. Dando un breve resumen de sus vidas, Kimball nació en 1944 en Minnesota, Estados Unidos, titulándose como ingeniero eléctrico de la Universidad de Stanford especializado en sistemas humano-maquina, se convirtió en uno de los primeros desarrolladores de los sistemas de almacenamiento de datos contando con una bibliografía en la cual nos aporta gran cantidad de información con relación al Data Warehousing, siendo su libro más conocido *The Data Warehouse Toolkit* (1996), por otro lado, Inmon nació en California, Estados Unidos en el año 1945, es un licenciado en Matemáticas de la universidad de Yale (1967) y el título más importante el cual es su maestría en ciencias de la computación de la universidad de Nuevo México, siendo uno de los padres o promotores del Data Warehousing y su mantenimiento, teniendo en su bibliografía libros como *Building the Data Warehouse* (1992) y en la actualidad es dueño de Forest Rim Technology,

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

empresa que implementa soluciones de Data Warehousing para las empresas. (Naeem, 2020)

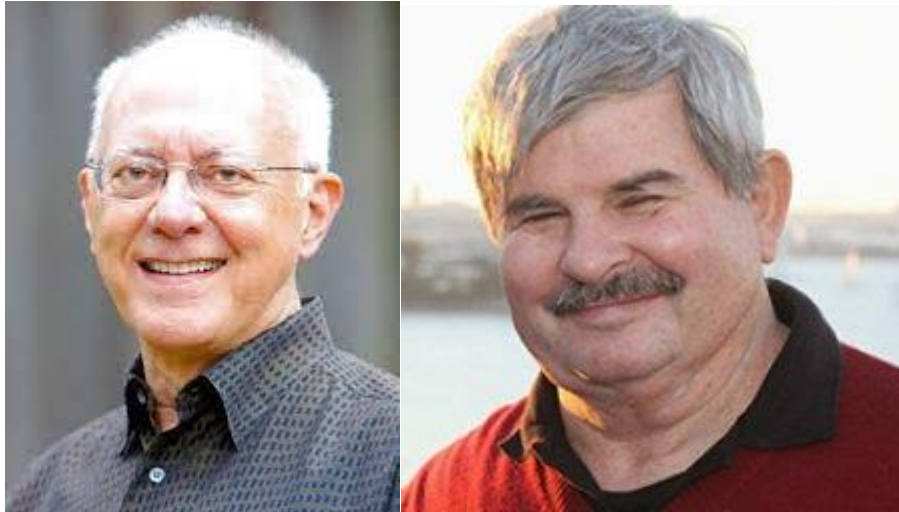



Ilustración 12 En orden, Ralph Kimball y Bill Inmon en la actualidad

Bottom-Up – Ascendente – Metodología Multidimensional

Esta metodología propuesta por Kimball e, la cual la creación del DW se realiza después de la creación de cada uno de los Data Marts para cada sector de la empresa, de modo tal el proceso de consolidación se realiza para la creación de cada Data Mart, obteniendo un DW que integrara todos los datos almacenados y organizados de cada Data Mart. Esta metodología garantiza que, al consolidar la información para cada Data Mart, se obtendrán consultas específicamente estructuradas para su posterior análisis (Bustamante-Granda et al., 2018; Gravitator, 2020; Gutiérrez, 2011; Naeem, 2020). Alguna de las ventajas que presenta esta metodología son las siguientes:

- Menor costo: esta metodología le permitirá a la organización el poder mantener el DW con grupos pequeños de profesionales en datos.
- Procesos focalizados: Al estar enfocado en los procesos, sectores o áreas de la empresa, esta metodología asegura una buena funcionalidad, generación de informes más específicos y buen seguimiento de métricas.
- Administración “distribuida”: se hace referencia a una administración distribuida en este caso debido a como vamos a gestionar los datos por medio de cada uno de los Data Marts, determinando que el DW solo requerirá una administración de la carga y consulta de los datos.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

Algunas desventajas de esta metodología podrían ser las siguientes:

- Al focalizarse en ciertos procesos o sectores de la organización, no se podrá garantizar una total cobertura en el manejo de los requisitos para los informes requeridos por la organización.
- La modificación de un DW que emplee esta metodología será poco flexible ya que esta se enfocara en base a los sectores de datos que posee cada Data Mart.

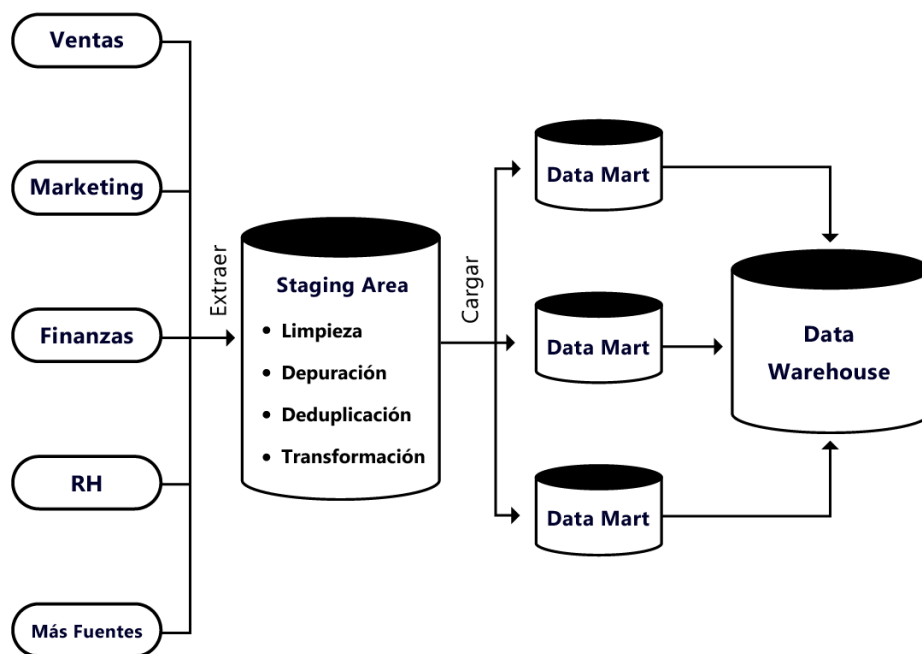



Ilustración 13 Diagrama de la metodología Bottom Up

Top-Down – Descendente – Metodología Relacional

En esta metodología Inmon nos presenta un diseño que toma como centro de información empresarial al DW tomando como fuente de la información para ciertos sectores o áreas al mismo DW y de ahí crear los Data Marts requeridos, generando de esta forma una única fuente confiable. Al momento de cargar los datos al DW se debe tener en cuenta que el volumen de los datos, todos los procesos de consolidación y las fuentes de datos requieren de un mayor control para obtener una estructura normalizada que satisfaga las necesidades de la empresa (Bustamante-Granda et al., 2018; Gravitator, 2020; Gutiérrez, 2011; Naeem, 2020). Algunos beneficios que presenta esta metodología son los siguientes:

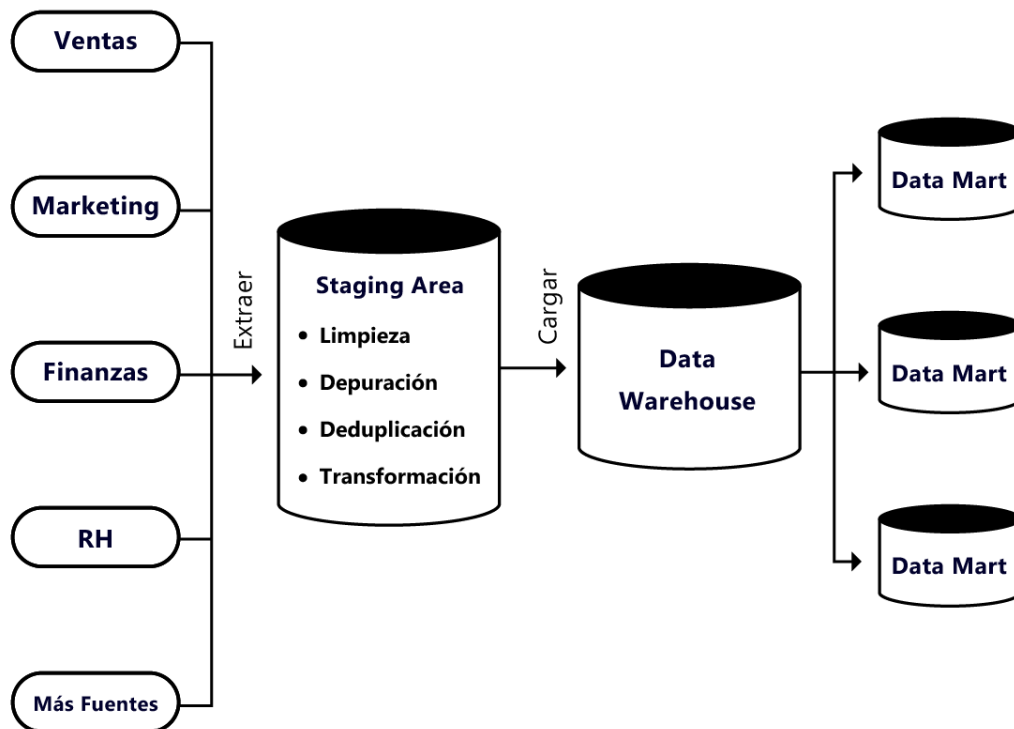
- Menor probabilidad de fracaso: los procesos de ETL serán más fáciles de realizar, al igual que los procesos de actualización y corrección de anomalías.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

- Mayor compresión de la actividad empresarial: al tener toda la información de la empresa concentrada y estructurada en un solo lugar esto garantizara que las consultas que se realicen sean más confiables que las que se podrían realizar de un Data Warehouse cargado con datos provenientes de Data Marts (se evita redundancia de los datos).
- Mayor flexibilidad: si se requieren realiza<r cambios, modificaciones o adiciones en base a las necesidades analíticas de la organización se podrá realizar sin ningún problema.


Algunas desventajas que presenta esta metodología son las siguientes:

- Mayores costos: debido a que estructurar y levantar un DW demanda mayores recursos, como lo son el almacenamiento, la infraestructura o los profesionales, se requiere que la organización invierta más dinero.
- Mayor tiempo de respuesta: al tener todos los datos concentrados, los tiempos de respuesta y de los procesos involucrados son mayores.



COMPARACIÓN DE DATA WAREHOUSE Y DATA LAKE


A continuación, se presenta un cuadro comparativo de las principales características de un Data Warehouse y un Data Lake (ver [definición](#)) en donde se

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	Conceptos sobre Data Warehouse	2021 - 2
--	---	-----------------------------------	----------

contemplan características relacionadas con el coste, usuarios, propósito, actividades y tipos de datos que se almacenan en cada una de estas estructuras:

Data Warehouse	Data Lake
Su propósito es el permitirle a la organización realizar análisis requeridos en la toma de decisiones de la organización.	Su propósito es almacenar Big-data a bajo costo y con la mejor efectividad posible
Se almacenan datos históricos los cuales requieren de procesos para poder almacenarse dentro del almacén de datos	Pueden almacenar datos estructurados o no estructurados provenientes de diferente fuente de datos de la organización
Las actividades que se realizan en esta estructura en la carga de datos históricos y la consulta de estos resumidos para la realización de informes.	Las actividades que se realizan en esta estructura es el almacenar datos de gran volumen y el análisis de estos en ámbitos como el Deep Learning o el análisis en tiempo real.
Quienes accederán a los datos almacenados serán Analistas de datos y Analistas empresariales.	Quienes tendrán acceso a los datos almacenados serán profesionales en ciencia de datos e ingenieros
Los datos almacenados serán de gran tamaño, pero solo serán aquellos de vital importancia para la organización y la toma de decisiones (pueden ser Gigas o Teras de datos los que se almacenen)	Los datos almacenados serán todos aquellos que puedan o no ser usados, lo cual derivara en que se tengan datos que consuman hasta Petabytes de almacenamiento.
El diseño de este sistema de almacenamiento estará orientado a los sucesos.	El diseño de ese sistema de almacenamiento estará orientado a las aplicaciones.

Condensando estos datos, si lo que requerimos es para un entorno empresarial, enfocado en el análisis de datos para la toma de decisiones dentro de la organización y requiriendo almacenar datos de vital importancia para la empresa, ahí sabremos que requerimos desarrollar o implementar un DW, por otro lado, si los datos a almacenar son muy variados, sin ninguna estructura fija, si el entorno en donde se va a aplicar no es una organización o empresa y su principal necesidad será tener un lugar en donde se almacenen datos que podrían o no ser usados, lo que requeriremos de implementar será un Data Lake. (Grupo Atico34, 2021; Mendez et al., 2004; Sulmont, 2020)


 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

BUSINESS INTELLIGENCE

Dentro del mundo de los negocios o de las organizaciones, la toma de decisiones es una actividad crítica y recurrente, la cual no se puede realizar sin contemplar la información que posee la empresa o que puede adquirir de otras fuentes, para esto requerimos del uso de herramientas tecnológicas, procesos y metodologías que nos ayuden a convertir los datos almacenados en información por medio de la cual la organización o empresa obtenga conocimiento, todo esto es una definición muy burda de lo que es el Business Intelligence, basado en (Cano, 2007), podemos obtener distintas definiciones de BI, como la siguiente que es muy apropiada *“Business Intelligence (BI) es un término paraguas que abarca los procesos, las herramientas, y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los negocios*“, en adición se nos da a conocer como el Data Warehousing va de la mano con el BI, ya que por medio de los almacenes de datos se puede obtener la información que requiera la empresa, pero teniendo siempre en cuenta que toda actividad o herramienta por medio de la cual se utilice los datos de interés de una empresa con el fin de obtener conocimiento.

Cabe resaltar que la inteligencia de negocios no es una actividad que solo se reserva para los directivos o dueños de una empresa, sino para cualquier integrante de la organización que tenga como responsabilidad tomar alguna decisión que afecte la actividad de la empresa. Y una particularidad que nos revela () es como las empresas u organizaciones consumen la mayoría de su tiempo obteniendo datos, estructurándolos y organizándolos, pero dejan como una tarea de segundo nivel el analizar, y estudiar los datos obtenidos, siendo esto una necesidad que debe aumentar a la par que una empresa genera una estructura mas compleja en su actividad comercial, ya que una empresa que tiene solo 5 clientes y que solo produce un único producto no requiere tomar decisiones tan complejas como una empresa como 20000 clientes alrededor del mundo y que genera infinidad de productos y servicios, cada una de estas empresas tiene contempla la toma de decisiones y el análisis de datos de una forma diferente, pero no desvalora el como una empresa desde su estado primitivo debe tomar decisiones que aseguren su supervivencia en el mercado. (Bustamante-Granda et al., 2018)

La implementación de BI dentro de una organización en la actualidad es más que necesario y más si se tiene acceso a datos para su análisis, no olvidemos que en la actualidad la información es oro puro, pero solo lo será si se es utilizada a favor de quien la posea, y dentro de una empresa lo que siempre se busca es obtener beneficios y BI presenta los siguientes tipos de beneficios:

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

- Beneficios tangibles: este tipo de beneficios tienen relación con ganancias, con beneficios monetarios, como lo son generar ingresos, la reducción de costes al detectar anomalías, riesgos u oportunidades.
- Beneficios intangibles: en el caso de la BI, lo intangible es la información y los beneficios que se obtienen con esta, como por ejemplo el conocimiento sobre alguna problemática en algún sector de la empresa o el conocimiento que le permita a la empresa aprovechar una oportunidad y así mejorar su posición.


Beneficios estratégicos: en este apartado nos referimos a todo aquellos que le permita a la empresa el poder formularse estrategias, como, por ejemplo, el saber a que publico o clientes hay que dirigirse o los productos que se deben ofertar en cierto momento del año. (Bustamante-Granda et al., 2018; Cano, 2007)

TIPOS DE DATOS

Al momento de elegir nuestras fuentes de datos nos encontraremos que estas no siempre albergan el mismo tipo de datos, un ejemplo muy simple puede ser el comparar una base de datos relacional en la cual se almacenan los datos bajo cierta estructura que permite que sean consultados los datos y visualizados fácilmente, por otro lado tenemos un archivo PDF el cual contienen ciertos datos dentro de él pero que el sistema de almacenamiento como lo puede ser un repositorio o servidor de archivos solo requiere almacenar más no puede comprender los datos que este archivo contiene, ahí es donde nos enfrentamos al reto de la gestión de los datos para su almacenamiento por parte del DW, a continuación se explicaran a más detalle los tipos de datos existentes: (AyudaLey, 2020; KYOCERA, 2021)

Datos Estructurados

Los datos estructurados son todos aquellos conjuntos de datos que se encuentran en archivos que pueden ser descifrados, ordenados y procesados por herramientas de software, esta definición se hace muy aproximada a la mayoría de las bases de datos, en las cuales los datos se ordenan por medio de atributos e índices. Dentro de una organización este tipo de datos es muy familiar, ya que lo encontramos en las bases de datos transaccionales, en las cuales la misma empresa consigna su actividad económica diariamente, utilizando los datos almacenados para extraer información por medio de consultas de bajo nivel y hasta llegar a ser explotados por medio de herramientas para la minería de datos. Otros tipos de datos estructurados son las hojas de Excel, archivos CSV, formularios web o fichas de clientes estandarizados, y en general cualquier conjunto de datos que tengan una estructura identificable.

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	Conceptos sobre Data Warehouse	2021 - 2
--	---	-----------------------------------	----------


ID Libro	Título	Autor	Editorial	Género	ISBN
001	<i>Justicia Auxiliar</i>	Ann Leckie	Nova	Ciencia Ficción	xxxxxxxxx
002	<i>La ciudad que nos unió</i>	N.K. Jemesin	Nova	Ciencia Ficción	xxxxxxxxx
003	<i>La Historia Interminable</i>	Michael Ende	Santillana	Fantasía	xxxxxxxxx
004	<i>Sakura</i>	Matilde Asensí	Esfera de los libros	Ficción/Suspense	xxxxxxxxx
005	<i>Neverwhere</i>	Neil Gaiman	Roca Libros	Fantasía Urbana	xxxxxxxxx

Ilustración 14 Ejemplo de datos estructurados

En la imagen anterior podemos encontrar con una tabla en la cual vemos un conjunto de datos estructurados y que no solo son comprensibles al ojo humano, sino que también un sistema de gestión de bases de datos podrá comprender dicha estructura y responder a consultas realizadas sobre esta tabla.

Datos No estructurados

Al comprender que la mayoría de las bases de datos que conocemos son las estructuras en donde encontraremos datos estructurados y procesables por herramientas de software, podremos dar el paso a los datos no estructurados, los cuales pueden parecer insignificantes, pero son, en promedio, el 80% de los datos que puede poseer una organización. Este tipo de datos se puede expresar como datos binarios sin ninguna estructura interna identificable, los elementos que conforman este tipo de datos pueden ser identificados y darle de cierta forma algunos atributos por medio de los cuales pueden ser clasificados en algún sistema. En este tipo de datos debemos comprender que todos los datos están en bruto y se requieren de actividades como la intervención humana o de herramientas de software especializadas que permitan explotar o extraer la mayor cantidad de datos posibles, algunos ejemplos de este tipo de datos son los archivos de algún procesador de texto como Word, PDFs, imágenes, videos, audios y publicaciones en medios sociales como Facebook. Todo este grupo de datos tienen en común el poder ser almacenados y administrados por algún sistema, pero los datos que estos archivos o elementos poseen requieren de procesos o tecnologías que permitan extraer los datos almacenados en estos.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------



NEVERWHERE

Prólogo

La noche antes de marcharse a Londres, Richard Mayhew no se estaba divirtiendo.

La velada había comenzado bien: había disfrutado leyendo las tarjetas de despedida y recibiendo los abrazos de varias chicas que conocía y que no estaban nada mal; le había gustado escuchar las advertencias sobre los vicios y los peligros que le acecharían en Londres, y le había encantado el paraguas blanco con el mapa del metro de Londres que los colegas le habían regalado entre todos. Las primeras pintas le habían sentado muy bien, pero después, con cada pinta de más le parecía que se divertía un poco menos. Ahora estaba sentado en la acera, tiritando, delante de un pub situado en una pequeña localidad escocesa, tratando de decidir si sería mejor vomitar o no, y no se divertía en absoluto.

Dentro del pub, los amigos de Richard seguían celebrando su inminente partida con un entusiasmo que, en opinión de Richard, comenzaba a rayar en lo siniestro. Sentado en la acera, agarrando con fuerza el paraguas cerrado, se preguntaba si realmente sería una buena idea trasladarse a Londres.

—Ándate con ojo —dijo una voz cascada y vieja—. Te echarán de aquí antes de que puedas decir esta boca es mía. O te detendrán, que tampoco me sorprendería.

Dos penetrantes ojos lo miraban fijamente desde un mugriento y aguileño rostro.

—¿Te encuentras bien?


—Sí, gracias —dijo Richard. Era un joven de rostro aniñado, con el cabello oscuro levemente ondulado y grandes ojos de color avellana; su aspecto descuidado, como si acabara de levantarse, lo hacía más atractivo para el sexo opuesto de lo que él mismo podía entender o creer.

Ilustración 15 Ejemplo de datos no estructurados

En la anterior imagen nos encontramos con un archivo de texto en formato PDF el cual puede ser visualizado por medio de software como Adobe Reader o los mismos navegadores, pero los datos que se encuentran dentro de este archivo no guardan ninguna estructura que permitan su procesamiento por software.

Datos Semiestructurados

Existen un tipo de datos los cuales están en un punto intermedio de las categorías anteriormente explicadas, en los cuales los datos tienen cierta estructura, organización o jerarquía, pero esta no se encuentra determinada por un esquema fijo, sino que varían con relación a las necesidades, enredando datos en estructuras de árbol. Los datos semiestructurados se componen por medio de metadatos, un

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

conjunto de etiquetas y elementos que permiten su clasificación y almacenamiento, sin tener la posibilidad de ser datos gestionables o automatizados fácilmente. Dentro de este tipo de datos nos podemos encontrar los correos electrónicos, lenguajes de etiquetado o marcado y archivos binarios, teniendo como peculiaridad que, aun siendo HTML y JSON datos de este tipo, estos datos han sido estructurados por Google permitiendo marcar datos estructurados dentro de las paginas web por medio de la implementación de etiquetas.



 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
--	---	---	-----------------


TABLA DE ILUSTRACIONES

Ilustración 1 Principales características de un almacén de datos	5
Ilustración 2 Flujo de los datos e información de un Data Warehouse empresarial	8
Ilustración 3 Flujo de datos para un Data Warehouse Operativo	9
Ilustración 4 Flujo de los datos para la creación de un Data Mart con una Metodología Top-Down	10
Ilustración 5 Fuentes de datos de un Data Lake	11
Ilustración 6 Arquitectura de un Data Warehouse	12
Ilustración 7 Proceso de Consolidación	13
Ilustración 8 Ejemplo de un modelo multidimensional	17
Ilustración 9 ejemplo de un Modelo de Estrella	17
Ilustración 10 ejemplo de un Modelo de Copo de Nieve	18
Ilustración 11 Ejemplo de un Modelo de Galaxia	19
Ilustración 12 En orden, Ralph Kimball y Bill Inmon en la actualidad	20
Ilustración 13 Diagrama de la metodología Botton Up	21
Ilustración 14 Ejemplo de datos estructurados	26
Ilustración 15 Ejemplo de datos no estructurados	27

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

BIBLIOGRAFIA

- Abella, R., Cópola, L., & Olave, D. (1999). *Sistema DataWarehousing*.
- AyudaLey. (2020, April 28). *Diferencias entre datos estructurados y no estructurados*. AyudaLey. <https://ayudaleyprotecciondatos.es/bases-de-datos/diferencias-datos-estructurados-no-estructurados/>
- Berzal, F. (2016). *Acceso a los datos Bases de datos relacionales: SQL O/R Mapping Bases de datos distribuidas Bases de datos NoSQL Bases de datos multidimensionales: Data Warehousing*.
- Bustamante-Granda, W. X., Estela, Macas-Ruiz, M., Fanny, & Cevallos-Macas, B. (2018). Data Warehouse: Análisis Multidimensional de BAFICI utilizando Power Pivot Data Warehouse: Multidimensional Analysis of BAFICI using Power Pivot Contenido. *Espacios*, 39.
- Cano, J. L. (2007). BUSINESS INTELLIGENCE: COMPETIR CON INFORMACIÓN. *BanESCO*, 01(01). https://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf
- De Souza, I. (2019, December 26). *Data Warehouse: ¿qué es un almacén de datos y para qué sirve?* Rock Content. <https://rockcontent.com/es/blog/data-warehouse/>
- Díaz Villanueva, W. (2015). *Almacenes de datos (Data Warehouses)*. <http://informatica.uv.es/iiguia/DBD/Teoria/data-warehouses.pdf>
- Gravitar. (2020, August 28). *Metodologías de Data Warehouse | Business Intelligence, Data Warehouse, Monterrey, México: Gravitar*. Gravitar. <https://gravitar.biz/datawarehouse/metodologias-data-warehouse/>
- Grupo Atico34. (2021, February 11). *Data warehouse. Definición y funciones*. Grupo Atico34. https://protecciondatos-lopd.com/empresas/data-warehouse/#Funciones_de_los_data_warehouses
- Gutiérrez, P. (2011). *DATA WAREHOUSE: MARCO DE CALIDAD*. UNIVERSIDAD CARLOS III.
- IBM. (2020, August 19). *¿Qué es un Data Warehouse? - Colombia | IBM*. IBM Db2 Warehouse Edición Empresarial. <https://www.ibm.com/co-es/analytics/data-warehouse>
- KYOCERA. (2021). *Diferencia entre datos estructurados y no estructurados*. KYOCERA. <https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/insights-hub/articles/diferencia-entre-datos-estructurados-y-no-estructurados.html>

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>Conceptos sobre Data Warehouse</p>	<p>2021 - 2</p>
---	---	---	-----------------

León, E. (2007). *Modelamiento Dimensional*.
<https://disi.unal.edu.co/~eleonguz/cursos/bda/presentaciones/S3-modelamiento.pdf>

López, B. (2002). *GUIAN PARA LA CONSTRUCCION DE UN DATA WAREHOUSE* (01 ed., Vol. 01). Universidad Autónoma de Nuevo León.
<http://eprints.uanl.mx/1131/1/1020147975.pdf>

Mendez, A., Britos, A., & Garcia-Martínez, P. Y. (2004). Fundamentos de Data Warehouse. *Reportes Técnicos En Ingeniería Del Software*, 5(1).
<http://www.itba.edu.ar/capis/rtis>

Mendoza, A. (2020, December 10). *Tipos de Data Warehouse | Business Intelligence, Data Warehouse*, Monterrey, México: Gravitar. Gravitar.
<https://gravitar.biz/datawarehouse/data-warehouse-tipos/>

Naeem, T. (2020, February 3). *Conceptos de Data Warehouse: enfoque de Kimball vs. Inmon | Astera*. Astera. <https://www.astera.com/es/type/blog/data-warehouse-concepts/>

Sierra, Y. (2020, July 20). *DWH: ejemplos, características y arquitectura del Data Warehouse - ADN Cloud*. Media Cloud. https://blog.mdcloud.es/dwh-ejemplos-arquitectura-data-warehouse/#Ejemplos_del_uso_de_DWH_en_distintos_sectores

Sulmont, L. (2020, January 6). *Data Lakes vs. Data Warehouses - DataCamp*. Data Lakes vs. Data Warehouses. <https://www.datacamp.com/community/blog/data-lakes-vs-data-warehouses>

Valbuena, D. (2018, September 18). *Esquemas en Data Warehousing*. Data Management. <https://datamanagement.es/2020/04/03/esquemas-data-warehousing/>

Zeus Data Talk. (2021, October 8). *¿Qué es y para qué sirve el Data Warehouse en tu empresa? | Big Data | Zeus*. Zeus Data Talk. <https://datablog.zeus.vision/2017/06/02/que-es-data-warehouse/>