

**CONCEPTOS SOBRE DATA WAREHOUSE - 2**

**OPTATIVA II: DATA WAREHOUSE**

**ANDRES SANCHEZ**

**LUIS FELIPE VELASCO TAO**


**UNIVERSIDAD DE SAN BUENAVENTURA**

**FACULTAD DE INGENIERIA**

**BOGOTÁ**


**18 DE AGOSTO**

**2021**

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

## Contenido


<b>TABLA DE ILUSTRACIONES .....</b>	<b>3</b>
<b>ARQUITECTURA DE DATA WAREHOUSE .....</b>	<b>4</b>
FUENTES DE DATOS .....	4
CONSOLIDACIÓN – STAGING AREA.....	5
Extracción (E).....	5
Transformación (T).....	6
Carga (L) .....	6
ALMACENAMIENTO.....	7
ACCESO .....	8
EXPLOTACIÓN .....	8
ARQUITECTURAS OLAP.....	9
ROLAP – Relational OLAP .....	9
MOLAP – Multidimensional OLAP .....	11
HOLAP – Hybrid OLAP .....	14
Operaciones Básicas OLAP .....	15
<b>MODELOS DE DATA WAREHOUSE.....</b>	<b>17</b>
MODELO DE ESTRELLA.....	18
MODELO DE COPO DE NIEVE .....	19
COMPARACIÓN MODELO DE ESTRELLA VS. MODELO DE COPO DE NIEVE .....	20
<b>METODOLOGÍAS DE DATA WAREHOUSE .....</b>	<b>21</b>
KIMBALL E INMON .....	21
BOTTOM-UP – ASCENDENTE – METODOLOGÍA MULTIDIMENSIONAL .....	22
Ciclo de vida metodología Kimball .....	23
TOP-DOWN – DESCENDENTE – METODOLOGÍA RELACIONAL.....	27
METODOLOGÍA RAPID WAREHOUSING .....	28

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

## **BIBLIOGRAFIA ..... 31**

### **TABLA DE ILUSTRACIONES**

Ilustración 1 Arquitectura de un Data Warehouse .....	4
Ilustración 2 Proceso de Consolidación .....	5
Ilustración 3 ETL y actividades a realizar en cada etapa .....	7
Ilustración 4 estructura ROLAP .....	9
Ilustración 5 Oracle presenta gran variedad de soluciones de almacenamiento...	10
Ilustración 6 SQL Server es una buena opción para Data Warehousing .....	11
Ilustración 7 DB2 ayuda de forma intuitiva a la construcción de sistemas relacionales .....	11
Ilustración 8 estructura de arquitectura MOLAP .....	12
Ilustración 9 Hyperion es una alternativa poco conocida para arquitecturas MOLAP .....	13
Ilustración 10 SQL server es un sistema versátil para Data Warehousing .....	13
Ilustración 11 Oracle como gestor de arquitecturas MOLAP .....	14
Ilustración 12 estructura HOLAP .....	15
Ilustración 13 demostración de las operaciones OLAP .....	16
Ilustración 14 Ejemplo de un modelo multidimensional .....	18
Ilustración 15 ejemplo de un Modelo de Estrella .....	19
Ilustración 16 ejemplo de un Modelo de Copo de Nieve .....	19
Ilustración 17 comparativas modelos de estrella y copo de nieve .....	20
Ilustración 18 En orden, Ralph Kimball y Bill Inmon en la actualidad .....	21
Ilustración 19 Diagrama de la metodología Botton Up .....	23
Ilustración 20 ciclo de vida de la metodología de Kimball .....	23
Ilustración 21 Estructura metodología top-down .....	28
Ilustración 22 metodología rápida para DW .....	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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

## 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 1 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 (OLTP), repositorios de archivos de la empresa, archivos o datos relacionadas con historiales de la organización y entre otros componentes que contengan datos relacionados con la operación de la organización y que haya sido generado o creado por la misma.

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

## Consolidación – Staging Area

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)

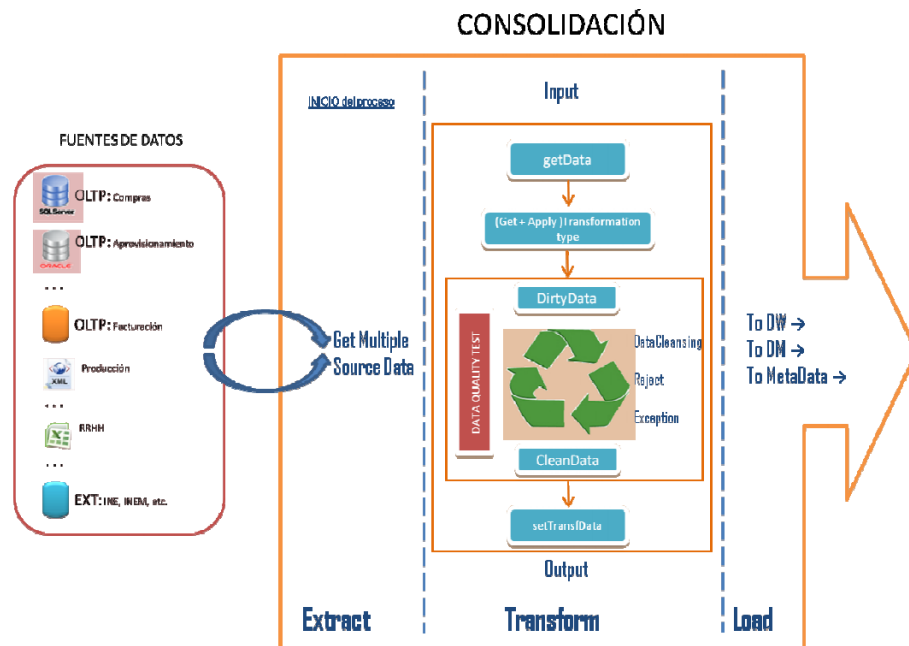



Ilustración 2 Proceso de Consolidación

## Extracción (E)

Previamente se realizó el proceso de selección de fuentes de datos, las cuales pueden proveernos de datos, sean estructurados o no estructurados, tales como las bases de datos transaccionales de la empresa, en donde los datos tienen cierta estructura y permiten la extracción de información, y por otro lado nos podemos encontrar con una variedad de archivos los cuales también pueden proveer ciertos datos pero conllevando procesos específicos para la obtención de los datos y su posterior estructuración. En esta etapa se deben tener en cuenta las necesidades o requerimientos expuestos previamente por la empresa y de esta forma se podrá

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------


identificar, seleccionar y unir todos aquellos datos que cumplan con dichas necesidades, se puede definir que en esta etapa se filtran las fuentes de datos seleccionadas con el fin de dejar en un área de almacenamiento temporal a convertir en datos históricos. (Gutiérrez, 2011; Vega Fajardo & Chiriguaya Freire, 2014)

### Transformación (T)

Al seleccionar los datos que cumplen con las necesidades y requerimientos debemos contemplar que estos datos se encuentran estructurados todos de una forma diferente entre sí, esto se puede expresar bajo un ejemplo muy simple, contemplando el resultado de una consulta de una base de datos como SQL, en donde los datos son de cierto tipo y por otro lado tenemos los datos que se encuentran que una hoja de calculo de Excel, ambos tipos de datos son necesarios y se van a almacenar dentro del almacén de datos pero para esto se deben aplicar actividades por medio de las cuales se realice una corrección de errores de tipado, eliminación de campos innecesarios y demás procesos que permitan la normalización y estructuración de los datos obtenidos. (Vega Fajardo & Chiriguaya Freire, 2014)

### Carga (L)

Finalmente, al realizar los procesos anteriores, nos encontraremos con datos de calidad, los cuales ya han sido tratados con el fin de ser almacenados en el DW, en esta etapa se implementaran métodos de carga, los cuales garanticen que los datos se almacenen dentro del almacén, bajo cierta estructura y contemplando que los datos que se cargan son de gran volumen, dichos métodos de carga deben garantizar los tiempos más óptimos, además de contemplar la carga de datos históricos o actualizados y el control de las transformaciones aplicadas en el proceso de transformación. (Gutiérrez, 2011; Vega Fajardo & Chiriguaya Freire, 2014)

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------




*Ilustración 3 ETL y actividades a realizar en cada etapa*

### 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 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](#)

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

[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)

### 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 alimentará 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



 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

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.

### Arquitecturas OLAP

El procesamiento de datos en línea o OLAP es una tecnología por medio de la cual se puede obtener de forma rápida y eficiente consultas de analíticas rápidas, de modo tal se puede obtener información de un DW y presentarla de forma entendible y comprensible por los usuarios, por lo que se deben tomar en cuenta las distintas arquitecturas existentes y de esta forma elegir aquella que se adecue al DW y a las necesidades de la organización.

### ROLAP – Relational OLAP


Esta forma de almacenar los datos nos presenta una arquitectura de tres capas indicados a continuación:

- Se tendrá una base de datos relacional en la cual se almacenarán los datos.
- Un motor OLAP por medio del cual se realizará el análisis de los datos almacenados en la base de datos.
- Una herramienta o software que sirva para la presentación de la información generada por el motor OLAP, se puede considerar como la capa de vista.



*Ilustración 4 estructura ROLAP*

Esta arquitectura nos presenta la capacidad de realizar consultas directamente sobre las bases de datos ya existentes, por lo que el motor OLAP es capaz de realizar la transformación de los datos extraídos de las bases de datos y estructurarlos de forma multidimensional y así ser almacenados en el DW y a su vez

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

luego permitirle al usuario realizar consultas sobre los datos para presentarlos a los usuarios mediante la ultima capa de vista al usuario.

Algunas de las ventajas de esa arquitectura son las siguientes:

- **Aprovechamiento de tecnologías relacionales:** las empresas podrán aprovechar el hardware y SGBD ya implementados, lo cual facilita la manipulación y no genera gastos adicionales.
- **Seguridad:** este punto se apega al anterior, ya que se la seguridad de los datos se dará por medio de los SGBD relacionales ya implementados y los protocolos de seguridad que estos tengan integrados.
- **Manipulación de grandes volúmenes de datos:** esta arquitectura soporta grandes volúmenes de datos estructurados en varias dimensiones.
- **Escalabilidad:** permite la adición de dimensiones al esquema ya existente.

Algunas desventajas de esta arquitectura son las siguientes:

- **Rendimiento:** debido a que se basa en SGBD relacionales, el rendimiento es mucho menos a lo que nos brindan modelos multidimensionales.
- **Almacenamiento:** requiere de tener mucho espacio de almacenamiento para soportar los datos requeridos.
- **Capacidad de procesamiento:** esta arquitectura presenta limitaciones en la capacidad de realizar consultas esto debido a que se usan sistemas relacionales los cuales se basan en procesos de unión para obtener ciertos datos.


### SGBD ROLAP

Dentro de los sistemas más populares para la gestión de esta arquitectura OLAP encontramos las siguientes:

- **Oracle:** presenta soporte desde bases de datos relacionales hasta DW con arquitectura de almacenamiento relacional, siendo muy versátil lo cual garantiza que las personas se capaciten en la manipulación de BD relaciones sean más fáciles de encontrar.



*Ilustración 5 Oracle presenta gran variedad de soluciones de almacenamiento*

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

- **SQL Server:** al igual que Oracle o cualquier SGBD relacional, se opera en base a la recolección de datos contenidos en los objetivos relacionales y se permite la extracción, consulta y análisis por parte del mismo sistema.



*Ilustración 6 SQL Server es una buena opción para Data Warehousing*

- **DB2:** es una herramienta bastante flexible para el análisis de datos de DW, permitiendo un manejo mucho más simple ya que presenta configuraciones base que permiten su rápida implementación.




*Ilustración 7 DB2 ayuda de forma intuitiva a la construcción de sistemas relacionales*

### MOLAP – Multidimensional OLAP

Este tipo de arquitectura solo emplea dos capas las cuales son descritas a continuación:

- Base de datos multidimensional: este tipo de base de datos estructura los datos por medio de dimensiones por medio de las cuales se podrá manejar, acceder y obtener los datos.
- Motor analítico: esta capa se encargará de la ejecución de las consultas sobre la base de datos, aplicando mecanismos de consulta OLAP.
  - La presentación de los datos estará a cargo del motor analítico, el cual procesará los datos rescatados de las consultas y estructurará la forma en que serán presentados a los usuarios.

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

El funcionamiento de esta arquitectura inicia en la carga de los datos provenientes de las bases de datos transacciones dentro de la base de datos multidimensional, de modo tal se generan las tablas de hechos y dimensiones, después de esto se realizarán ciertos cálculos por medio de los cuales se obtengan las medidas requeridas para todas las tablas involucradas, esto con el fin de enriquecer los datos almacenados y dejar la base de datos multidimensional lista para su uso por medio del motor analítico y sus interfaces de usuario.




*Ilustración 8 estructura de arquitectura MOLAP*

Algunas de las ventajas que esta arquitectura presenta son las siguientes:

- **Transparencia:** la forma en que la base de datos ha sido estructurada será imperceptible por parte de los usuarios de la base de datos multidimensional.
- **Análisis mejorado:** la forma en que los datos se encuentran ordenados facilita la consulta de los datos, permitiendo un grado de flexibilidad y rapidez frente a otras arquitecturas.
- **Rendimiento:** frente a las arquitecturas ROLAP se obtiene un mejor rendimiento ya que se trabaja con datos agregados, totales, subtotales, series y grados de detalle.

Algunas de las desventajas que se enfrentaría emplear esta arquitectura son las siguientes:

- **Construcción compleja:** se requiere de tener el conocimiento suficiente para construir un sistema multidimensional, además de que la estructura de los datos y la carga de esta demanda mucho tiempo.

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

- **Complejidad limitada:** los sistemas multidimensionales para almacenar los datos pueden estar limitados a diez dimensiones, ya que al tener muchas dimensiones la manipulación de los datos se hace mas compleja, además de que no se pueden manejar datos de tanto volumen.
- **Costos altos:** se requiere del uso de herramientas pagas para el manejo de MOLAP además de personal con mayores conocimientos con relación a sistemas multidimensionales.

### SGBD MOLAP

Algunos Sistemas de gestión para MOLAP que se pueden emplear son los siguientes:

- **Hyperion Essbase OLAP Server:** creado por Hyperion, empresa conocida por la creación de herramientas que permiten el análisis en entornos OLAP, y esta en especial permite la creación de informes, análisis, modelos y presupuestos de forma optima y acorde a las necesidades de la empresa, es un sistema multiusuarios, con capacidad de almacenar grandes volúmenes de datos y con un motor de gran potencia para consultas complejas.




*Ilustración 9 Hyperion es una alternativa poco conocida para arquitecturas MOLAP*

- **SQL Server - Microsoft Analysis Services:** este sistema es muy versátil, ya que, como antes mencionamos también permiten la manipulación de ROLAP, y en este caso, también soporta la creación y gestión de cubos multidimensionales, por lo que se puede considerar que este SGBD es una buena opción si se quiere aprender de DW y su gestión, tanto de forma relacional como multidimensional.



*Ilustración 10 SQL server es un sistema versátil para Data Warehousing*

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

- **Oracle Express:** este sistema de gestión presenta una gran cantidad de herramientas soportadas sobre Oracle Express Server, un motor de calculo potente y un gestor de memoria cache de datos. Las herramientas que componen este sistema se enfocan mucho en las necesidades del usuario y el análisis de datos por medio de consultas OLAP.




*Ilustración 11 Oracle como gestor de arquitecturas MOLAP*

### HOLAP – Hybrid OLAP

En base a las dos arquitecturas presentadas previamente se han desarrollado soluciones en las cuales se usan elementos tanto relacionales como multidimensionales, en donde podemos encontrar como los registros más detallados que sean muy grandes se seguirán almacenando de forma relacional, por otro lado, de forma multidimensional, se almacenaran los datos agregados como las medidas y los resultados de cálculos sobre los datos almacenados de forma relacional. La implementación de esta arquitectura demanda de personal muy capacitado con relación a las dos arquitecturas que componen a este sistema, de modo tal las ventajas y desventajas que las arquitecturas ROLAP y MOLAP se mantienen, añadiendo como desventaja el grado de complejidad para su implementación, incluyendo la gestión de la interconexión de los datos relacionales y multidimensionales, y la manipulación de los SGBD de forma óptima, acorde 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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

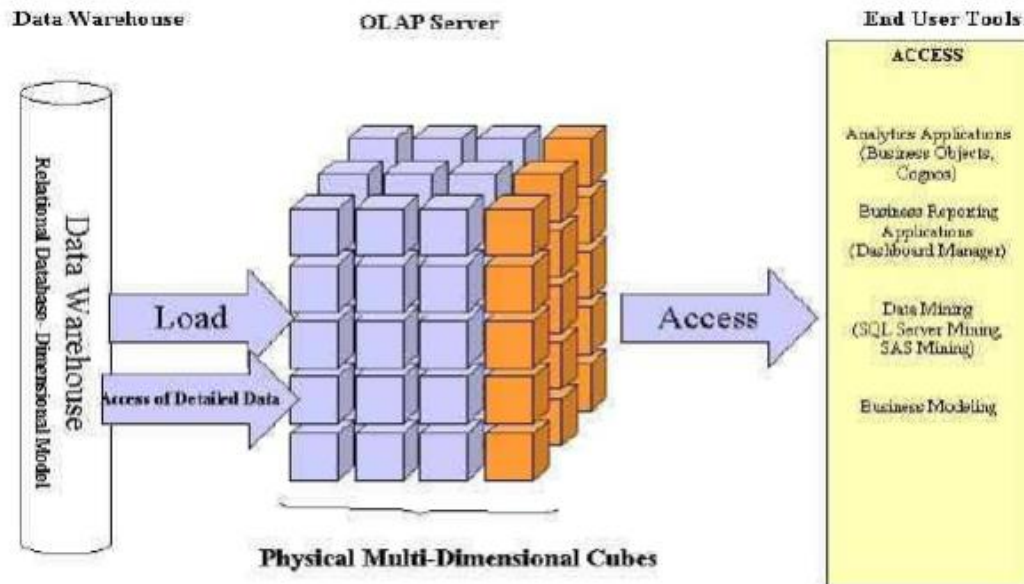



Ilustración 12 estructura HOLAP

### Operaciones Básicas OLAP

Dentro de los sistemas OLAP nos encontramos un conjunto de operaciones por medio de las cuales podremos manipular las consultas, esto con el fin de obtener el nivel de detalle que requiramos. Las operaciones básicas en arquitecturas OLAP son las siguientes:

- **Roll-up** (Enrollar - Compactar): por medio de esta operación se va a generar un resumen de los datos a lo largo de una dimensión en específico.
- **Drill-down** (profundizar): a lo contrario de la anterior operación, por medio de esta se podrá obtener una vista mucho más detallada entre las dimensiones, por ejemplo, se obtendrá información que parta desde las jerarquías más altas de una dimensión hasta las más bajas, aquí se pueden obtener, por ejemplo, datos sobre la dimensión de tiempo desde años, meses y hasta días.
- **Slice** (Corte - segmento): al realizar análisis, muchas veces queremos concentrarnos sobre un sector de una dimensión en específico, por lo que podremos obtener una parte específica de alguna dimensión, por ejemplo, obtener información asociada al año 2020.
- **Dice** (cubo): apoyado en la operación anterior, se aplican slice por medio de los cuales se obtenga un sector determinado de cada una de las dimensiones, por ejemplo, en un DW que tiene por dimensiones los productos, ciudades y fechas, podremos obtener las ventas de **arroz** en **Bogotá** en el mes de **abril**. Aquí puede variar el nivel de detalle de cada una

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

de las dimensiones, lo importante es definir un filtro o limitante para cada una de las dimensiones existentes.

- **Pivot** (Pivotar): se pueden obtener distintos puntos de vista “girando” o moviéndonos por cada una de las vistas o caras que nos presenta el cubo.

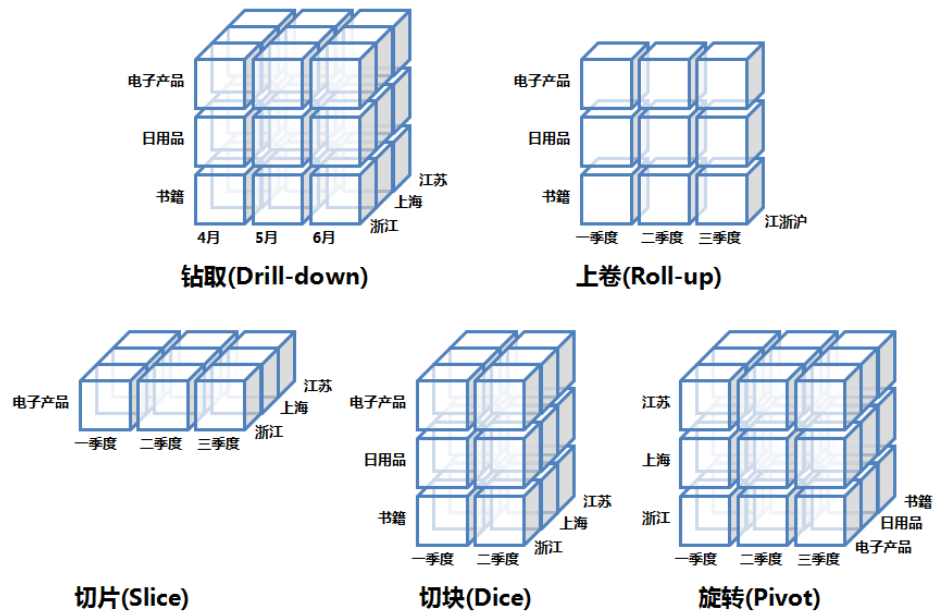



Ilustración 13 demostración de las operaciones OLAP




 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

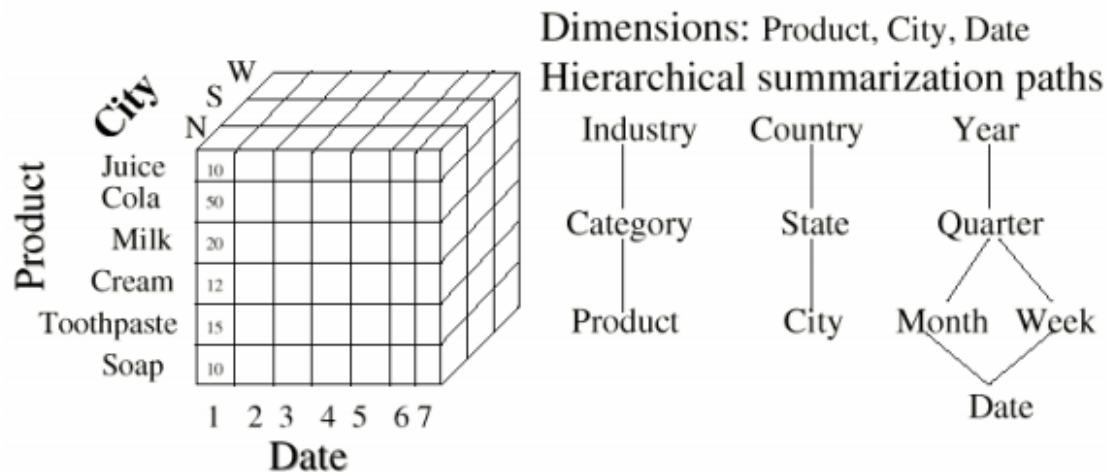
## MODELOS DE DATA WAREHOUSE

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:

- **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 o relaciones dimensionales 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

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

cierta hora, las medidas que pueden acompañar este hecho puede ser la cantidad de unidades compradas o el valor total de la venta.



*Ilustración 14 Ejemplo de un modelo multidimensional*

En la anterior imagen se expresa como las dimensiones se relacionan entre sí, por lo que el cubo representa la colección de hechos en los cuales se relacionan los productos, ciudades y fechas, además de que se expresa como cada una de las dimensiones hacen parte de una jerarquía por medio de la cual se pueden obtener agrupaciones de elementos que componen cada una de las dimensiones. (Aguilar Mallea & Pilco Romero, 2013)

### 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).

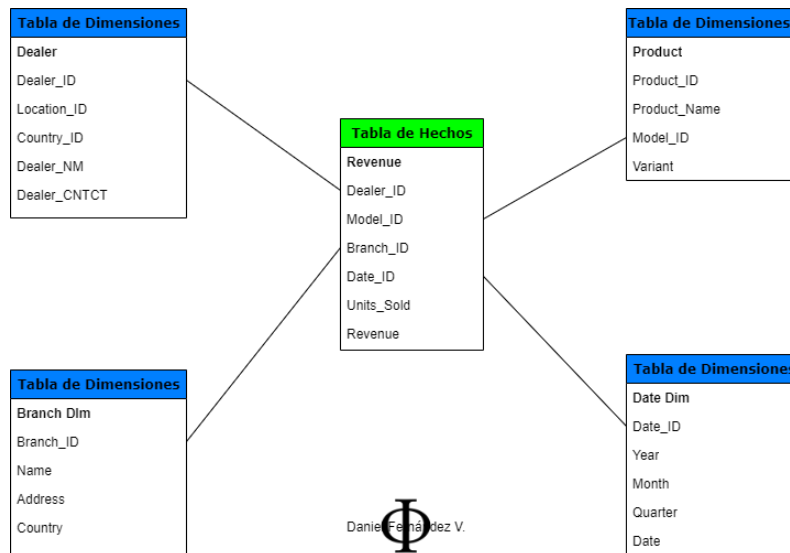


Ilustración 15 ejemplo de un Modelo de Estrella

### 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).

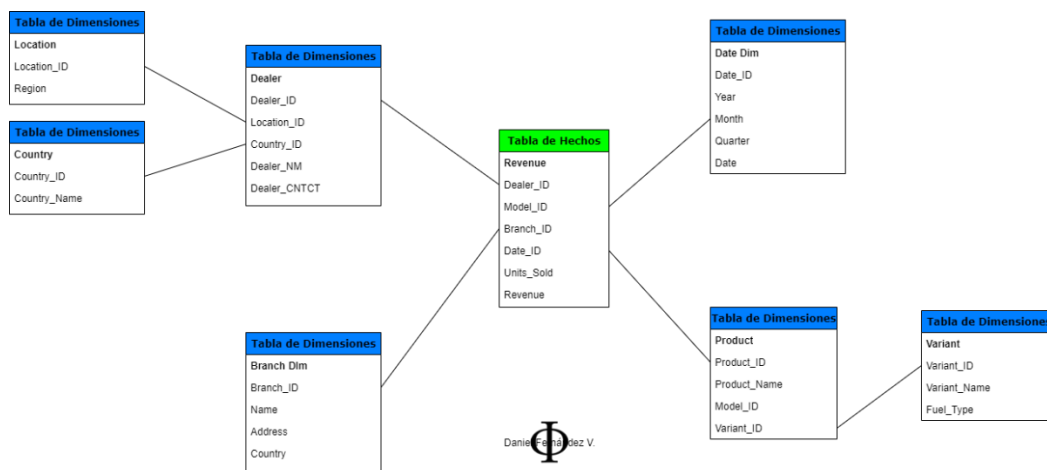



Ilustración 16 ejemplo de un Modelo de Copo de Nieve


 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

### Comparación modelo de estrella vs. Modelo de copo de nieve

Con base a la información obtenida en (Berzal, 2016; Ramos, 2016), se puede realizar una comparativa entre los dos modelos anteriormente expuestos:

Modelo estrella	Modelo copo de nieve
El acceso a los datos se realizará por medio de consultas de bajo nivel de complejidad debido a que se requerirán pocos joins para obtener los datos.	Las queries por medio de las cuales se accederá a los datos tienen un nivel de complejidad más alto, ya que deben contemplar las dimensiones agregadas por medio de una gran cantidad de joins.
Este modelo regularmente tiene una mayor redundancia con los datos almacenados, lo cual puede dificultar su mantenimiento.	Al tener tablas de dimensiones que ayudan a la especificación de las dimensiones conectadas a la tabla de hechos, se puede obtener una menor redundancia, lo cual facilitara su mantenimiento.
Debido a que las consultas tienen un bajo nivel de complejidad el rendimiento de este modelo será optimo, obteniendo consultas de forma rápida.	El acceso a los datos por medio de las consultas presenta una mayor lentitud ya que se deben contemplar una mayor cantidad de joins.
Se aconseja usar este modelo cuando las tablas de dimensiones tienen pocas filas. Por lo que se obtienen regularmente tablas de dimensiones y hechos desnormalizadas.	Si las tablas de dimensiones presentan una gran cantidad de filas, se aconseja emplear ese modelo y normalizar las tablas de dimensiones.
Se tendrán N tablas para cada una de las N dimensiones existentes.	Al normalizar las tablas de las dimensiones, se pueden tener más de una tabla para las N dimensiones existentes.
Este modelo se apoya en la metodología Top-Down.	Este modelo se apoya en la metodología Button-up.

*Ilustración 17 comparativas modelos de estrella y copo de nieve*

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

## 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; Gravitator, 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 del 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) (Kimball & Ross, 2013), 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, empresa que implementa soluciones de Data Warehousing para las empresas. (Naeem, 2020)

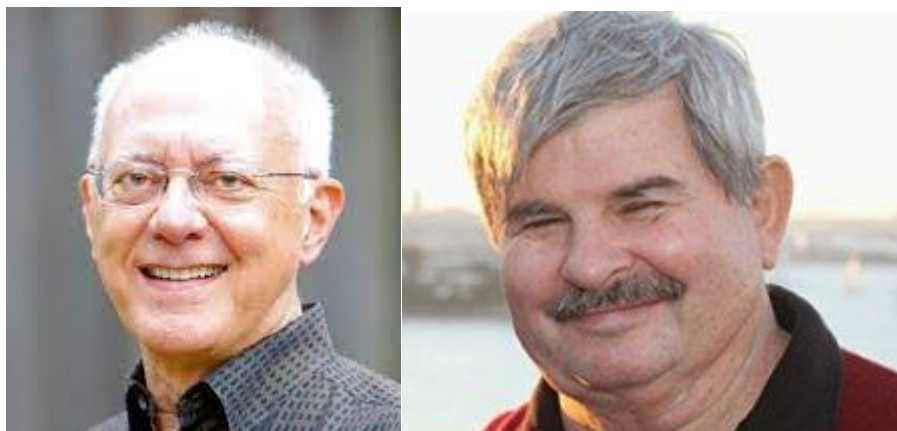



Ilustración 18 En orden, Ralph Kimball y Bill Inmon en la actualidad

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

### 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.

Algunas desventajas de esta metodología podrían ser las siguientes:

- **Sin cobertura total:** 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.
- **Poco flexible:** 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.



 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

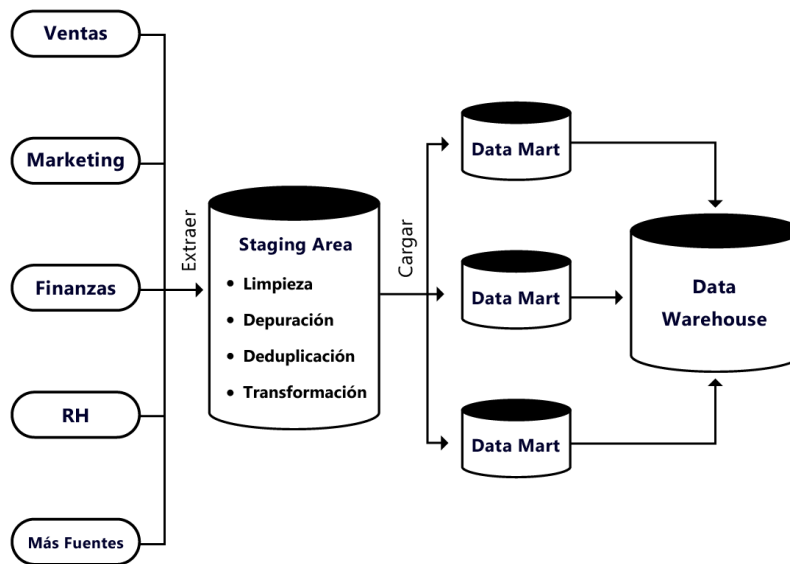


Ilustración 19 Diagrama de la metodología Bottom Up

### Ciclo de vida metodología Kimball

Kimball en su libro *The Data Warehouse Lifecycle Toolkit* (Kimball & Ross, 2013) nos presenta un ciclo de vida para el Data Warehousing el cual especifica los procesos que se deben seguir al momento de implementar o mejorar un DW. En la siguiente imagen se expresa la ruta que debemos seguir para estos procesos:

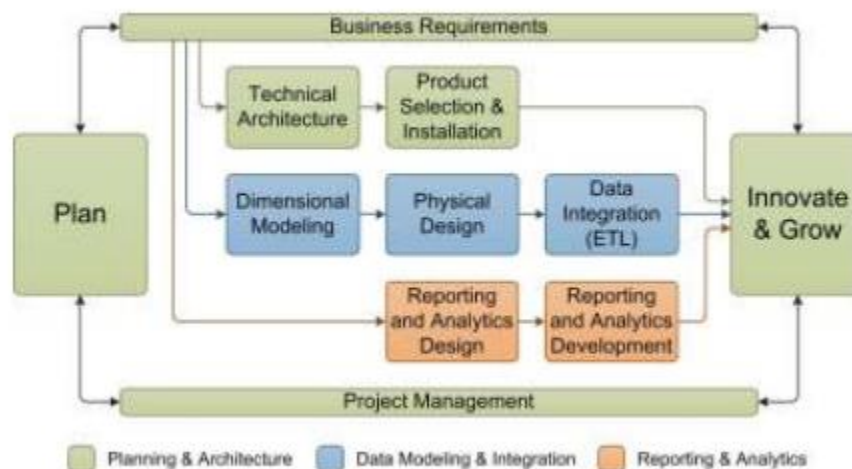




Ilustración 20 ciclo de vida de la metodología de Kimball

Las fases y actividades que componen el ciclo de vida son las siguientes (Gutiérrez, 2011; Kimball & Ross, 2013):

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

1. **Fase de planeación:** en esta etapa nos concentraremos en definir los objetivos que queremos cumplir con la implantación o mejora del DW, aquí debemos definir si es necesario o no su implementación, esto basándose en que se requieren definir los usuarios, que estos a su vez tengan motivos fuertes y solidos para el manejo de datos históricos. También se deben definir elementos como el alcance, los encargados en el proceso de implementación y aspectos relacionados con infraestructura, esto último de forma preliminar.
2. **Fase de definición de requerimientos:** esta fase se considera como la central, de la cual partirán tres ramificaciones o flujos, esto demuestra la importancia de definir correctamente, según Kimball (Kimball & Ross, 2013) “*los requerimientos del negocio se posicionan en el centro del universo del Data Warehouse*”, esto se puede interpretar a que, en base a estos se contemplan las necesidades de los analistas, el alcance que tendrá el DW, los datos a manipular, periodos para la actualización y carga de datos y los puntos de acceso al sistema.
3. **Fase de flujo tecnológico:** esta es una de las ramas en donde se debe trabajar con base a los requerimientos definidos, en esta se realizará un especial enfoque en la tecnología y la implementación de esta para el DW. Se realizan las siguientes actividades:
  - 3.1. Diseño de arquitectura técnica: en esta etapa se contemplan los tres pilares de una correcta arquitectura: requerimientos del negocio, entornos técnicos para el despliegue y ordenes técnicas y estratégicas proyectadas por la organización. Cada uno de estos elementos acondicionarán y limitarán que se desarrolle un DW acorde a las necesidades de la organización, encajando con la capacidad técnica, de infraestructura y monetaria y siempre teniendo en cuenta a los ejecutivos. Para que todo esto sea así, Kimball define 8 pasos por medio de los cuales se procura un correcto diseño arquitectónico y en poco tiempo. Los pasos son los siguientes:
    - 3.1.1. *Definir un grupo de trabajo de arquitectura:* se aconseja definir un grupo de tres personas, lideradas por un arquitecto técnico quien estará encargado de los datos de diseño y de definir las tareas de los otros dos integrantes del grupo.
    - 3.1.2. *Requisitos relacionados con la arquitectura:* en este paso se contemplará los requisitos por medio de los cuales se cree un sistema 100% acorde a las necesidades de la organización y al contexto de esta.
    - 3.1.3. *Documento de requisitos arquitectónicos:* este se basa en la documentación de los pasos previos.
    - 3.1.4. *Desarrollo de modelo arquitectónico de alto nivel:* este paso se comenzará a desarrollar los modelos bajo los requerimientos definidos, de modo tal el equipo se puede dividir acorde a los



 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

componentes principales del DW (Ir a [arquitectura de Data Warehouse](#)). Se pueden fijar reuniones por medio de las cuales se controle el avance de las actividades definidas para cada uno de los miembros del grupo y finalmente integrar todos los modelos.

3.1.5. *Diseño y especificaciones de subsistemas*: Este paso se enfoca en el diseño de cada uno de los subsistemas o componentes, acorde a los elementos de infraestructura, rendimiento, software y hardware requerido.

3.1.6. *Definición de fases de aplicación de la arquitectura*: al tener fijos los elementos para el despliegue del DW, se definirán fases para su implementación, acorde a como deben ir conectados o relacionados cada uno de los componentes.

3.1.7. *Documento de arquitectura técnica*: en esta etapa se documentarán todos los pasos previos, recordemos la importancia de las evidencias documentales en cualquier desarrollo.


3.1.8. *Revisión final de la arquitectura técnica*: finalmente se hace una revisión a toda la documentación generada, tanto por parte del grupo definido como por parte de los directivos de la organización y ahí dar el inicio a la siguiente actividad de esta fase.

3.2. Selección e instalación del Data Warehouse: en esta etapa vamos a elegir las tecnologías requeridas para todo el despliegue y posterior manipulación o gestión del DW, algunos de los componentes o herramientas de software que se contemplaran son las siguientes:

- Plataforma DW: esto en el caso que la organización en donde se vaya a desplegar el DW no cuente con dicha plataforma.
- SGBD: esto en el caso que la organización en donde se vaya a desplegar el DW no cuente con un SGBD para DW.
- Herramienta ETL: software que este encargado con los procesos de manipulación de los datos provenientes de las fuentes.
- Herramienta de consultas y/o reportes: en el caso que el SGBD no cuente con herramientas para las consultas, se debe buscar una que satisfaga las capacidades técnicas y económicas de la organización.

4. **Fase de flujo de datos**: en esta fase nos centraremos en como los datos se moverán a través de la arquitectura definida, contemplando la transformación de los datos provenientes de las fuentes, su carga al DW y demás acciones que se podrán ejercer sobre los datos. Podemos contemplar las siguientes etapas dentro de esta fase:


4.1. Modelo dimensional o lógico: apegándonos a los requerimientos del negocio, definiremos la forma en que se encontraran estructurados

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

enfocados en la forma en que los datos serán analizados, aquí veremos la definición de las dimensiones, los grados de detalle, jerarquías y demás elementos por medio de los cuales se defina la forma en que serán almacenados los datos.

- 4.2. Diseño físico: en esta etapa le daremos vida al modelo dimensiona, contemplando la definición de estándares del entorno de la base de datos, la indexación y estrategias del particionamiento. Aquí debemos contemplar que se debe trabajar en base a las jerarquías definidas para cada una de las dimensiones, llegando al manejo de los datos en su nivel atómico, otros elementos a considerar son las consultas más frecuentes que se pueden realizar, las medidas y valores pre-calculados requeridos para su posterior análisis.
- 4.3. Procesos ETL: en esta etapa se realizarán los procesos de extracción, transformación y carga de datos al DW. (ver [procesos ETL](#))
5. **Fase de integración de aplicaciones para Business Intelligence:** en esta fase nos concentraremos en los conceptos de análisis de datos y emisión de reportes, todo lo relacionado con las herramientas que les permitan a los usuarios la obtención de analítica y toma de decisiones. Podemos contemplar las siguientes etapas dentro de esta fase:
  - 5.1. Diseño de análisis y reportes: en esta etapa se identificarán y definirán las herramientas o software de BI las cuales consumirán los datos almacenados en el DW, teniendo presente el diseño de una interfaz o de varias interfaces que permitan la comunicación entre el software de BI y el DW, tener muy presente que será el software de BI el cual le de valor y le de mayores posibilidades a los usuarios de tomar las mejores decisiones sobre la organización.
  - 5.2. Desarrollo de análisis y reportes: con la etapa previa, solo restaría el despliegue del software de BI e interfaces requeridas para la comunicación con los usuarios y con el DW, validando el correcto acceso a los metadatos.
6. **Despliegue y crecimiento:** en esta última fase se comprobará que todos los procesos se ejecutaron correctamente, revisando el flujo de los datos, el correcto desempeño de las tecnologías empleadas y que el software de BI responda de la forma esperada por los usuarios. Teniendo en cuenta que, hasta que todas las etapas (incluida la documentación total) se cumplan cabalidad, se verifique su funcionamiento y se realicen las correcciones pertinentes, se podrá realizar el despliegue del DW.

Cuando se hace referencia al crecimiento se hace referencia a que el DW no será un elemento estático, el cual no requerirá cambios, por lo que siempre estará presto a volver a implementar el ciclo de vida anteriormente expuesto,

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

revisando cada uno de los componentes del DW y se deben contemplar las siguientes actividades para el mantenimiento y control del DW:

- Monitorización de uso del DW.
- Tuning del desempeño.
- Mantenimiento de la tabla de índices.
- Procedimiento de Back Up del DW.
- Arreglo de incidentes que afecten el DW.


### 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.
- **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 realizar 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.

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

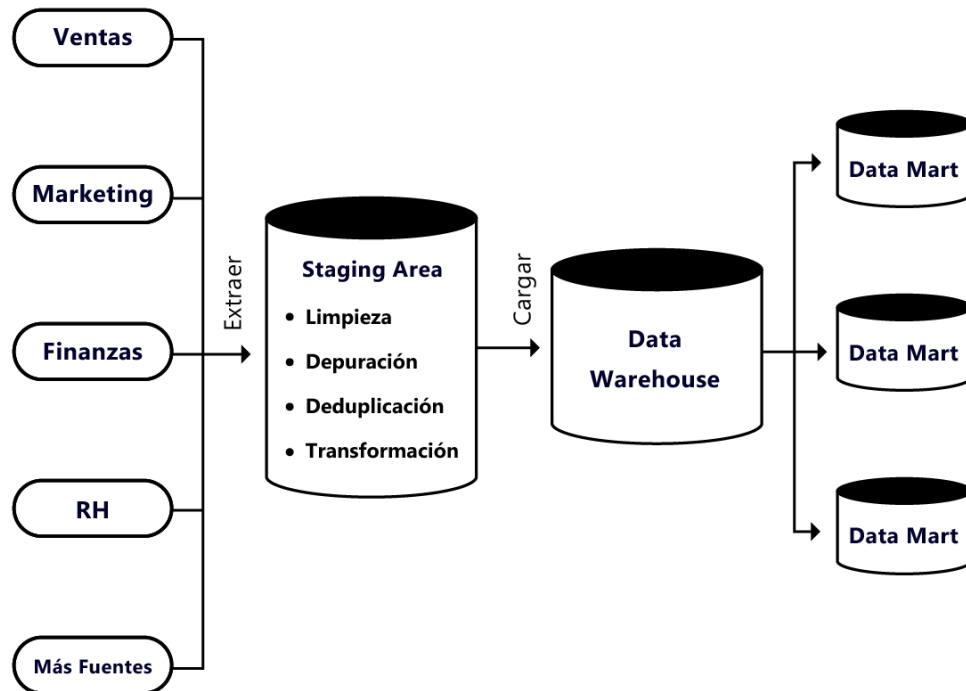



Ilustración 21 Estructura metodología top-down

### Metodología Rapid Warehousing

Esta es una metodología de tipo iterativo propuesta por el Instituto SAS (empresa conocida por sus productos de BI) la cual se basa en desarrollar de forma incremental el despliegue de un DW en una organización, todo esto por medio de cinco fases muy apegadas al ciclo de vida del software, en dichas fases se contempla la recolección de datos, definición de objetivos y requerimientos del DW, diseño, desarrollo, implementación y revisión, de modo tal se contempla que, por medio de la ultima fase el ciclo de vida de Data Warehouse se enriquezca o retroalimente (Brizuela & Blanco, 2013; Gutiérrez, 2011; Martha, 2011). Las fases que componen esta metodología son las siguientes:


1. **Definición de objetivos:** en esta fase se definen los alcances del DW, el grupo encargado de todo el proceso de despliegue del DW, las funciones que cumplirá el DW al ser instalado y puesto en operación, con todos estos elementos se podrá medir el nivel de éxito de la aplicación de la metodología.
2. **Definición de requerimientos de la información:** en la anterior fase se definen principalmente los objetivos con relación a aspectos técnicos, en esta podemos definir todas las necesidades de la organización que debemos satisfacer con relación a los datos, su almacenamiento y posterior análisis, para esto se realizaran reuniones con las partes interesadas para obtener la

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

mayor cantidad de información con relación a que datos se deben manipular, las fuentes que se requieren utilizar y demás elementos que hacen parte de la toma de decisiones de la organización.

3. **Diseño y modelado:** en esta fase realizaron el modelo dimensional o lógico a implementar del DW todo esto en base a todos los datos recopilados en la anterior fase, teniendo en cuenta la identificación de las fuentes de datos, los proceso ETL requeridos para almacenar los datos en el DW, siempre orientando a que el modelo físico final debe incluir todos los requerimientos definidos, especificar correctamente las dimensiones, hechos, medidas y documentar todo el proceso. No olvidar definir en esta fase las herramientas de acceso y analítica de los datos.
4. **Implementación:** la anterior fase nos dejó con las bases para realizar todas las actividades que sigan se especificaron en la arquitectura del DW, tales como la exploración de las fuentes de datos, todos los procesos de ETL definidos y la implementación de las herramientas que permitan el acceso y explotación de los datos.
5. **Revisión:** la revisión puede considerarse como la fase la cual se aplica no solo al final del despliegue del DW sino también una fase que se debe definir como una actividad recurrente para mantener optimo y en funcionamiento el DW. Al finalizar la implementación del DW se debe revisar la documentación generada en cada una de las fases previas, y por otro lado, como parte de las actividades de mantenimiento se debe tener un control critico con relaciones a inconvenientes y comentarios emitidos por parte de los usuarios del DW.


Después de definir las fases que generan una metodología iterativa y simple de aplicar, debemos fijarnos en el centro de esta metodología: la gestión del proyecto. Este componente es el eje central de toda la actividad ya que será una fase por medio de la cual se controlará el flujo en la ejecución de cada una de las fases, de modo tal la gestión del proyecto estará alineada a las directrices de los ejecutivos de la organización ya que son ellos quienes seguirán el cumplimiento de sus necesidades con relación al análisis de los datos y el progreso en la implementación del DW. (Martha, 2011)

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------




*Ilustración 22 metodología rápida para DW*



 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

## BIBLIOGRAFIA

- Abella, R., Cópola, L., & Olave, D. (1999). *Sistema DataWarehousing*.
- Aguilar Mallea, F. W., & Pilco Romero, W. F. (2013). Diseño de datos de un data warehouse. *Revista Ventana Científica*, 1(1), 24.  
[http://www.revistasbolivianas.org.bo/scielo.php?pid=S2305-60102013000100005&script=sci\\_arttext](http://www.revistasbolivianas.org.bo/scielo.php?pid=S2305-60102013000100005&script=sci_arttext)
- 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*.
- Brizuela, E. I. L., & Blanco, Y. C. (2013). Metodologías para desarrollar Almacén de Datos. *Revista de Arquitectura e Ingeniería*, 7(3), 1–12.  
<https://www.redalyc.org/pdf/1939/193930080003.pdf>
- 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.
- 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](https://protecciondatos-lopd.com/empresas/data-warehouse/#Funciones_de_los_data_warehouses)
- Gutiérrez, P. (2011). *DATA WAREHOUSE: MARCO DE CALIDAD*. UNIVERSIDAD CARLOS III.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley.  
<https://www.safaribooksonline.com/library/view/the-data-warehouse/9781118530801/>
- 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>

 <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 - 2</p>	<p>2021 - 2</p>
---	---	---	-----------------

Martha, H. (2011). Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular. *Revista Cubana de Información En Ciencias de La Salud*, 22(4). <http://www.rcics.sld.cu/index.php/acimed/rt/printerFriendly/208/168>

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/>

Ramos, S. (2016). Data Warehouse, data marts y modelos dimensionales. *Un Pilar Fundamental Para La Toma de Decisiones*. Albaterra: SolidQ, 1. [http://www.solidq.com/ebs/DataWarehouse\\_DataMarts\\_ModelosDimensionales\\_v2.pdf](http://www.solidq.com/ebs/DataWarehouse_DataMarts_ModelosDimensionales_v2.pdf)

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](https://blog.mdcloud.es/dwh-ejemplos-arquitectura-data-warehouse/#Ejemplos_del_uso_de_DWH_en_distintos_sectores)

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

Vega Fajardo, M., & Chiriguaya Freire, M. (2014). *Estudio y Diseño de un Data center alternativo para el Centro de Cómputo de la Universidad de Guayaquil* [Universidad de Guayaquil]. <http://repositorio.ug.edu.ec/bitstream/redug/17295/1/BFILO-PIN-0504.pdf>