**Huertos Acmes Chile S.A**

Creación de un Datawarehouse. Informe 2

|  |  |  |  |
| --- | --- | --- | --- |
| **UNIVERSIDAD TECNOLOGICA DE CHILE** | | | |
| **Programa** | INGENIERIA EN INFORMATICA | **Número Documento** |  |
| **Materia** | TI1232-Tecnologías de Información y Comunicación II | **Versión** | V1.0 09-Junio/2016 |
| **Profesores** | Manuel Rojas | **Autores** | Ian Torres  Matias Silva  Luis Pinto |

Contenido

[**INTRODUCCIÓN** 3](#_Toc529982752)

[**1.** **ANÁLISIS DE REQUERIMIENTOS** 4](#_Toc529982753)

[**1.1.** **Identificar preguntas** 4](#_Toc529982754)

[**1.2.** **Identificar indicadores y perspectivas** 6](#_Toc529982755)

[**1.3.** **Modelo Conceptual** 7](#_Toc529982756)

[**2.** **ANÁLISIS DE LOS OLTP** 9](#_Toc529982757)

[**2.1.** **Conformar Indicadores** 9](#_Toc529982758)

[**2.2.** **Establecer correspondencias** 9](#_Toc529982759)

[**2.3.** **Nivel de granularidad** 11](#_Toc529982760)

[**2.4.** **Modelo Conceptual ampliado** 14](#_Toc529982761)

[**3.** **MODELO LÓGICO DEL DW** 16](#_Toc529982762)

[**3.1.** **Tipo de Modelo Lógico del DW** 16](#_Toc529982763)

[**3.2.** **Tablas de dimensiones** 16](#_Toc529982764)

[**3.3.** **Tablas de hechos** 19](#_Toc529982765)

[**3.4.** **Uniones** 23](#_Toc529982766)

[**4.** **INTEGRACIÓN DE DATOS** 25](#_Toc529982767)

[**4.1.** **Carga Inicial** 25](#_Toc529982768)

[**4.2.** **Actualización** 33](#_Toc529982769)

[BIBLIOGRAFIA 35](#_Toc529982770)

[Indicaciones 36](#_Toc529982771)

**INTRODUCCIÓN**

A continuación, se presenta el desarrollo de la actividad propuesta por el docente a cargo de la asignatura el cual consta de la ejecución de procesos ETL (Extracción, Transformación y Carga).

Comprende el análisis y desarrollo de un modelo relacional de datos con una breve descripción del modelo de negocio que permitirá elaborar un modelo preliminar de las bases de datos para finalmente obtener el modelo de datos, adjuntado el script para la migración de las estructuras.

También se llevará a cabo una clasificación de los requerimientos explicando los requerimientos de información del modelo de negocio.

Posteriormente se realizará la creación de un modelo dimensional de datos realizando una justificación, explicación de la granularidad, las dimensiones incluyendo un análisis de las dimensiones para finalmente confeccionar el modelo multidimensional.

Finalmente incluye una conclusión del desarrollo.

1. **ANÁLISIS DE REQUERIMIENTOS**

Lo primero que se hará será identificar los requerimientos de l@s usuari@s a través de preguntas que expliciten los objetivos de su organización. Luego, se analizarán estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. Finalmente se confeccionará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso.

Es muy importante tener en cuenta que Acme se puede utilizar para construir un Data Warehouse o un Data Mart a la vez, es decir, si se requiere construir por ejemplo dos Data Marts, se deberá aplicar la metodología dos veces, una por cada Data Mart. Del mismo modo, si se analizan dos áreas de interés de negocio, como el área de “Ventas” y “Compras”, se deberá aplicar la metodología dos veces.

* 1. **Identificar preguntas**

El primer paso comienza con el acopio de las necesidades de información, el cual puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc.

El análisis de los requerimientos de l@s diferentes usuari@s, es el punto de partida de esta metodología, ya que ell@s son l@s que deben, en cierto modo, guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación con sus funciones y cualidades.

El objetivo principal de esta fase es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la empresa, y que facilitará una eficaz y eficiente toma de decisiones.

Debe tenerse en cuenta que dicha información, es la que proveerá el soporte para desarrollar los pasos sucesivos, por lo cual, es muy importante que se preste especial atención al relevar los datos.

Una forma de asegurarse de que se ha realizado un buen análisis es corroborar que el resultado del mismo haga explícitos los objetivos estratégicos planteados por la empresa que se está estudiando.

Otra forma de encaminar el relevamiento es enfocar las necesidades de información en los procesos principales que desarrolle la empresa en cuestión.

La idea central es, que se formulen preguntas complejas sobre el negocio, que incluyan variables de análisis que se consideren relevantes, ya que son estas las que permitirán estudiar la información desde diferentes perspectivas.

Un punto importante que debe tenerse muy en cuenta es que la información debe estar soportada de alguna manera por algún OLTP, ya que, de otra forma, no se podrá elaborar el DW.

*Caso práctico:*

Se indagó a l@s usuari@s en busca de sus necesidades de información, pero las mismas abarcaban casi todas las actividades de la empresa, por lo cual se les pidió que escogieran el proceso que considerasen más importante en las actividades diarias de la misma y que estuviese soportado de alguna manera por algún OLTP. El proceso elegido fue el de Ventas.

A continuación, se procedió a identificar qué era lo que les interesaba conocer acerca de este proceso y cuáles eran las variables o perspectivas que debían tenerse en cuenta para poder tomar decisiones basadas en ello.

Se les preguntó cuáles eran según ell@s, los indicadores que representan de mejor modo el proceso de Ventas y qué sería exactamente lo que se desea analizar del mismo. La respuesta obtenida, fue que se deben tener en cuenta y consultar datos sobre la cantidad de unidades vendidas y el monto total de ventas.

Luego se les preguntó cuáles serían las variables o perspectivas desde las cuales se consultarán dichos indicadores. Para simplificar esta tarea se les presentó una serie de ejemplos concretos de otros casos similares.

Las preguntas de negocio obtenidas fueron las siguientes:

Se desea conocer cuántas unidades de cada producto fueron vendidas a sus clientes en un periodo determinado. O en otras palabras: “Unidades vendidas de cada producto a cada cliente en un tiempo determinado”.

Se desea conocer cuál fue el monto total de ventas de productos a cada cliente en un periodo determinado. O, en otras palabras: “Monto total de ventas de cada producto a cada cliente en un tiempo determinado”.

Debido a que la dimensión Tiempo es un elemento fundamental en el DW, se hizo hincapié en él. Además, se puso mucho énfasis en dejar en claro a l@s usuari@s, a través de ejemplos prácticos, que es este componente el que permitirá tener varias versiones de los datos a fin de realizar un correcto análisis posterior.

*Como se puede apreciar, las necesidades de información expuestas están acorde a los objetivos y estrategias de la empresa, ya que es precisamente esta información requerida la que proveerá un ámbito para la toma de decisiones, que en este caso permitirá analizar el comportamiento de l@s client@s a l@s que se pretende satisfacer ampliamente, para así lograr obtener una ventaja competitiva y maximizar las ganancias.*

* 1. **Identificar indicadores y perspectivas**

Una vez que se han establecido las preguntas de negocio, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

Para ello, se debe tener en cuenta que los indicadores, para que sean realmente efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc.

En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar, que el Tiempo es muy comúnmente una perspectiva.

**Caso práctico:**

A continuación, se analizarán las preguntas obtenidas en el paso anterior y se detallarán cuáles son sus respectivos indicadores y perspectivas.

|  |
| --- |
| PIC    **Figura** **5.3:   Caso práctico, indicadores y perspectivas.** |

 En síntesis, los indicadores son:

* Unidades vendidas.
* Monto total de ventas.

Y las perspectivas de análisis son:

* Clientes.
* Productos.
* Tiempo.
  1. **Modelo Conceptual**

En esta etapa, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior. *Modelo Conceptual: descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.*

A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante l@s usuari@s y explicado con facilidad.

La representación gráfica del modelo conceptual es la siguiente:

|  |
| --- |
| PIC  **Figura** **5.4:   Modelo Conceptual.** |

A la izquierda se colocan las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación, constituye el proceso o área de estudio elegida. De dicha relación y entrelazadas con flechas, se desprenden los indicadores, estos se ubican a la derecha del esquema.

Como puede apreciarse en la figura anterior, el modelo conceptual permite de un solo vistazo y sin poseer demasiados conocimientos previos, comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos.

**Caso práctico:**

El modelo conceptual resultante de los datos que se han recolectado es el siguiente:

|  |
| --- |
| PIC  **Figura** **5.5:   Caso práctico, Modelo Conceptual.** |

1. **ANÁLISIS DE LOS OLTP**

Seguidamente, se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definirán qué campos se incluirán en cada perspectiva. Finalmente, se ampliará el modelo conceptual con la información obtenida en este paso.

* 1. **Conformar Indicadores**

En este paso se deberán explicitar como se calcularán los indicadores, definiendo los siguientes conceptos para cada uno de ellos:

* Hecho/s que lo componen, con su respectiva fórmula de cálculo. Por ejemplo: Hecho1 + Hecho2.
* Función de sumarización que se utilizará para su agregación. Por ejemplo: SUM, AVG, COUNT, etc.

***Caso práctico:***

Los indicadores se calcularán de la siguiente manera:

* ” Unidades Vendidas”:
  + Hechos: Unidades Vendidas.
  + Función de sumarización: SUM.

Aclaración: el indicador” Unidades Vendidas” representa la sumatoria de las unidades que se han vendido de un producto en particular.

* ”Monto Total de Ventas”:
  + Hechos: (Unidades Vendidas) \* (Precio de Venta).
  + Función de sumarización: SUM.

Aclaración: el indicador” Monto Total de Ventas” representa la sumatoria del monto total que se ha vendido de cada producto, y se obtiene al multiplicar las unidades vendidas, por su respectivo precio.

* 1. **Establecer correspondencias**

El objetivo de este paso es el de examinar los OLTP disponibles que contengan la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

La idea es, que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

**Caso práctico:**

En el OLTP de la empresa analizada, el proceso de venta está representado por el diagrama relacional de la siguiente figura.

Diagrama Relacional: representa la información a través de entidades, relaciones, cardinalidades, claves, atributos y jerarquías de generalización.

|  |
| --- |
| Diagrama Entidad Relación del caso práctico  **Figura** **5.6:   Caso práctico, Diagrama Relacional.** |

A continuación, se expondrá la correspondencia entre los dos modelos:

|  |
| --- |
| http://www.dataprix.com/files/uploads/250image/HEFESTO%20v2_0/xhefesto,P20-,P20modelo,P20conceptual,P20-,P20correspondencia.png.pagespeed.ic.Z3gXp9apPV.png  **Figura** **5.7:   Caso práctico, correspondencia.** |

Las relaciones identificadas fueron las siguientes:

* La tabla ”Productos” se relaciona con la perspectiva ”Productos”.
* La tabla ”Clientes” con la perspectiva ”Clientes”.
* El campo ”fecha” de la tabla ”Facturas\_Venta” con la perspectiva ”Tiempo” (debido a que es la fecha principal en el proceso de venta).
* El campo ”cantidad” de la tabla ”Detalles\_Venta” con el indicador ”Unidades Vendidas”.
* El campo ”cantidad” de la tabla ”Detalles\_Venta” multiplicado por el campo ”precio\_Fact” de la misma tabla, con el indicador ”Monto Total de Ventas”.
  1. **Nivel de granularidad**

Una vez que se han establecido las relaciones con los OLTP, se deben seleccionar los campos que contendrá cada perspectiva, ya que será a través de estos por los que se examinarán y filtrarán los indicadores.

Para ello, basándose en las correspondencias establecidas en el paso anterior, se debe presentar a l@s usuari@s los datos de análisis disponibles para cada perspectiva. Es muy importante conocer en detalle que significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, es conveniente investigar su sentido, ya sea a través de diccionarios de datos, reuniones con l@s encargad@s del sistema, análisis de los datos propiamente dichos, etc.

Luego de exponer frente a l@s usuari@s los datos existentes, explicando su significado, valores posibles y características, est@s deben decidir cuáles son los que consideran relevantes para consultar los indicadores y cuáles no.

Con respecto a la perspectiva “Tiempo”, es muy importante definir el ámbito mediante el cual se agruparán o sumarizarán los datos. Sus campos posibles pueden ser: día de la semana, quincena, mes, trimestres, semestre, año, etc.

Al momento de seleccionar los campos que integrarán cada perspectiva, debe prestarse mucha atención, ya que esta acción determinará la granularidad de la información encontrada en el DW.

**Caso práctico:**

De acuerdo a las correspondencias establecidas, se analizaron los campos residentes en cada tabla a la que se hacía referencia, a través de dos métodos diferentes. Primero se examinó la base de datos para intuir los significados de cada campo, y luego se consultó con el encargado del sistema sobre algunos aspectos de los cuales no se comprendía su sentido.

De todas formas, y como puede apreciarse en el diagrama de relacional antes expuesto, los nombres de los campos son bastante explícitos y se deducen con facilidad, pero aun así fue necesario investigarlos para evitar cualquier tipo de inconvenientes.

* Con respecto a la perspectiva” Clientes”, los datos disponibles son los siguientes:
  + id\_Cliente: es la clave primaria de la tabla” Clientes”, y representa unívocamente a un cliente en particular.
  + Codigo: representa el código del cliente, este campo es calculado de acuerdo a una combinación de las iniciales del nombre del cliente, el grupo al que pertenece y un número incremental.
  + Razon\_Soc: nombre o razón social del cliente.
  + Telefono1: número de teléfono del cliente.
  + Telefono2: segundo número telefónico del cliente.
  + Fax1: número de fax del cliente.
  + Fax2: segundo número de fax del cliente.
  + Mail1: dirección de correo electrónico del cliente.
  + Mail2: segunda dirección de correo del cliente.
  + id\_Sit\_Fiscal: representa a través de una clave foránea el tipo de situación fiscal que posee el cliente. Por ejemplo: Consumidor Final, Exento, Responsable No Inscripto, Responsable Inscripto.
  + CUIT: número de C.U.I.T. del cliente.
  + ConvenioMultilateral: indica si el cliente posee o no convenio multilateral.
  + DGR: número de D.G.R. del cliente.
  + id\_Clasificación: representa a través de una clave foránea la clasificación del cliente. Por ejemplo: Muy Bueno, Bueno, Regular, Malo, Muy Malo.
  + id\_Nota: representa a través de una clave foránea una observación realizada acerca del cliente.
  + Cta\_Habilitada: indica si el cliente posee su cuenta habilitada.
  + id\_Rubro: representa a través de una clave foránea el grupo al que pertenece el cliente. Por ejemplo: Bancos, Construcción, Educación Privada, Educación Pública, Particulares.
  + idCuentaContable: representa la cuenta contable asociada al cliente, la cual se utilizará para imputar los movimientos contables que este genere.
  + Eliminado: indica si el cliente fue eliminado o no. Si fue eliminado, no figura en las listas de clientes actuales.
* En la perspectiva ”Productos”, los datos que se pueden utilizar son los siguientes:
  + id\_prod: es la clave primaria de la tabla ”Productos”, y representa unívocamente a un producto en particular.
  + stock: stock actual del producto.
  + stock\_min: stock mínimo del producto, se utiliza para dar alerta si el stock actual está cerca del mismo, al ras o si ya lo superó.
  + Precio: precio de venta del producto.
  + Detalle: nombre o descripción del producto.
  + id\_Rubro: representa a través de una clave foránea el rubro al que pertenece el producto.
  + id\_Marca: representa a través de una clave foránea la marca a la que pertenece el producto.
  + stock\_MAX: stock máximo del producto. Al igual que ”stock\_min”, se utiliza para dar alertas del nivel de stock actual.
  + tipo: clasificación del producto. Por ejemplo: Producto, Servicio, Compuesto.
  + Costo: precio de costo del producto.
  + codigo: representa el código del producto, este campo es calculado de acuerdo a una combinación de las iniciales del nombre del producto, el rubro al que pertenece y un número incremental.
  + Imagen: ruta de acceso a una imagen o dibujo mediante la cual se quiera representar al producto. Este campo no es utilizado actualmente.
  + Generico: indica si el producto es genérico o no.
  + Eliminado: indica si el producto fue eliminado o no. Si fue eliminado, no figura en las listas de productos actuales.
  + PrecioR: precio de lista del producto.
* Con respecto a la perspectiva “Tiempo”, que es la que determinará la granularidad del depósito de datos, los datos más típicos que pueden emplearse son los siguientes:
  + Año.
  + Semestre.
  + Cuatrimestre.
  + Trimestre.
  + Número de mes.
  + Nombre del mes.
  + Quincena.
  + Semana.
  + Número de día.
  + Nombre del día.

Una vez que se recolectó toda la información pertinente y se consultó con l@s usuari@s cuales eran los datos que consideraban de interés para analizar los indicadores ya expuestos, los resultados obtenidos fueron los siguientes:

* Perspectiva ”Clientes”:
  + ”Razon\_Soc” de la tabla ”Clientes”. Ya que este hace referencia al nombre del cliente.
* Perspectiva ”Productos”:
  + ”detalle” de la tabla ”Productos”. Ya que este hace referencia al nombre del producto.
  + ”Nombre” de la tabla ”Marcas”. Ya que esta hace referencia a la marca a la que pertenece el producto. Este campo es obtenido a través de la unión con la tabla ”Productos”
* Perspectiva ”Tiempo”:
  + ”Mes”. Referido al nombre del mes.
  + ”Trimestre”.
  + ”Año”.
  1. **Modelo Conceptual ampliado**

En este paso, y con el fin de graficar los resultados obtenidos en los pasos anteriores, se ampliará el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo. Gráficamente:

|  |
| --- |
| PIC    **Figura** **5.8:   Modelo Conceptual ampliado.** |

**Caso práctico:**

Teniendo esto en cuenta, se completará el diseño del diagrama conceptual:

|  |
| --- |
| PIC  **Figura** **5.9:   Caso práctico, Modelo Conceptual ampliado.** |

1. **MODELO LÓGICO DEL DW**

A continuación, se confeccionará el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se definirá el tipo de modelo que se utilizará y luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizarán las uniones pertinentes entre estas tablas. *Modelo Lógico: representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD.*

* 1. **Tipo de Modelo Lógico del DW**

Se debe seleccionar cuál será el tipo de esquema que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades de l@s usuari@s. Es muy importante definir objetivamente si se empleará un esquema en estrella, constelación o copo de nieve, ya que esta decisión afectará considerablemente la elaboración del modelo lógico.

**Caso práctico:**

El esquema que se utilizará será en estrella, debido a sus características, ventajas y diferencias con los otros esquemas.

* 1. **Tablas de dimensiones**

En este paso se deben diseñar las tablas de dimensiones que formaran parte del DW.

Para los tres tipos de esquemas, cada perspectiva definida en modelo conceptual constituirá una tabla de dimensión. Para ello deberá tomarse cada perspectiva con sus campos relacionados y realizarse el siguiente proceso:

* Se elegirá un nombre que identifique la tabla de dimensión.
* Se añadirá un campo que represente su clave principal.
* Se redefinirán los nombres de los campos si es que no son lo suficientemente intuitivos.

Gráficamente:

|  |
| --- |
| PIC    **Figura** **5.10:   Diseño de tablas de dimensiones.** |

Para los esquemas copo de nieve, cuando existan jerarquías dentro de una tabla de dimensión, esta tabla deberá ser normalizada. Por ejemplo, se tomará como referencia la siguiente tabla de dimensión y sus respectivas relaciones padre-hĳo entre sus campos:

|  |
| --- |
| PIC    **Figura** **5.11:   Jerarquía de ”GEOGRAFIA”.** |

Entonces, al normalizar esta tabla se obtendrá:

|  |
| --- |
| PIC    **Figura** **5.12:   Normalización de ”GEOGRAFIA”.** |

**Caso práctico:**

A continuación, se diseñarán las tablas de dimensiones.

* Perspectiva “Clientes”:
  + La nueva tabla de dimensión tendrá el nombre “CLIENTE”.
  + Se le agregará una clave principal con el nombre “idCliente”.
  + Se modificará el nombre del campo “Razon\_Soc” por “Cliente”.

Se puede apreciar el resultado de estas operaciones en la siguiente gráfica:

|  |
| --- |
| PIC    **Figura** **5.13:   Caso práctico, tabla de dimensión ”CLIENTE”.** |

* Perspectiva “Productos”:
  + La nueva tabla de dimensión tendrá el nombre “PRODUCTO”.
  + Se le agregará una clave principal con el nombre “idProducto”.
  + El nombre del campo “Marca” no será cambiado.
  + Se modificará el nombre del campo “Detalle” por “Producto”.

Se puede apreciar el resultado de estas operaciones en la siguiente gráfica:

|  |
| --- |
| PIC    **Figura** **5.14:   Caso práctico, tabla de dimensión ”PRODUCTO”.** |

* Perspectiva “Tiempo”:
  + La nueva tabla de dimensión tendrá el nombre “FECHA”.
  + Se le agregará una clave principal con el nombre “idFecha”.
  + El nombre los campos no serán modificados.

Se puede apreciar el resultado de estas operaciones en la siguiente gráfica:

|  |
| --- |
| PIC    **Figura** **5.15:   Caso práctico, tabla de dimensión ”FECHA”.** |

##### 

* 1. **Tablas de hechos**

En este paso, se definirán las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

* Para los esquemas en estrella y copo de nieve, se realizará lo siguiente:
  + Se le deberá asignar un nombre a la tabla de hechos que represente la información analizada, área de investigación, negocio enfocado, etc.
  + Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
  + Se crearán tantos campos de hechos como indicadores se hayan definido en el modelo conceptual y se les asignará los mismos nombres que estos. En caso de que se prefiera, podrán ser nombrados de cualquier otro modo.

Gráficamente:

|  |
| --- |
| PIC    **Figura** **5.16:   Tabla de hechos.** |

* Para los esquemas constelación se realizará lo siguiente:
  + Las tablas de hechos se deben confeccionar teniendo en cuenta el análisis de las preguntas realizadas por l@s usuari@s en pasos anteriores y sus respectivos indicadores y perspectivas.
  + Cada tabla de hechos debe poseer un nombre que la identifique, contener sus hechos correspondientes y su clave debe estar formada por la combinación de las claves de las tablas de dimensiones relacionadas.
  + Al diseñar las tablas de hechos, se deberá tener en cuenta:
    - Caso 1: Si en dos o más preguntas de negocio figuran los mismos indicadores, pero con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición. Por ejemplo:

|  |
| --- |
| PIC  **Figura** **5.17:   Caso 1, preguntas.** |

Entonces se obtendrá:

|  |
| --- |
| PIC    **Figura** **5.18:   Caso 1, diseño de tablas de hechos.** |

* + Caso 2: Si en dos o más preguntas de negocio figuran diferentes indicadores con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición. Por ejemplo:

|  |
| --- |
| PIC    **Figura** **5.19:   Caso 2, preguntas.** |

* + Entonces se obtendrá:

|  |
| --- |
| PIC    **Figura** **5.20:   Caso 2, diseño de tablas de hechos.** |

* + Caso 3: Si el conjunto de preguntas de negocio cumple con las condiciones de los dos puntos anteriores se deberán unificar aquellos interrogantes que posean diferentes indicadores pero iguales perspectivas de análisis, para luego reanudar el estudio de las preguntas. Por ejemplo:

|  |
| --- |
| PIC    **Figura** **5.21:   Caso 3, preguntas.** |

* + Se unificarán en:

|  |
| --- |
| PIC    **Figura** **5.22:   Caso 3, unificación.** |



***Caso práctico:***

A continuación, se confeccionará la tabla de hechos:

* La tabla de hechos tendrá el nombre “VENTAS”.
* Su clave principal será la combinación de las claves principales de las tablas de dimensiones antes definidas: “idCliente”, “idProducto” e “idFecha”.
* Se crearán dos hechos, que se corresponden con los dos indicadores y serán renombrados, “Unidades Vendidas” por “Cantidad” y “Monto Total de Ventas” por “MontoTotal”.

En el gráfico siguiente se puede apreciar mejor este paso:

|  |
| --- |
| PIC    **Figura** **5.23:   Caso práctico, diseño de la tabla de hechos.** |

##### 

* 1. **Uniones**

Para los tres tipos de esquemas, se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

**Caso práctico:**

Se realizarán las uniones pertinentes, de acuerdo corresponda:

|  |
| --- |
| PIC    **Figura** **5.24:   Caso práctico, uniones.** |

1. **INTEGRACIÓN DE DATOS**

Una vez construido el modelo lógico, se deberá proceder a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc.; luego se definirán las reglas y políticas para su respectiva actualización, así como también los procesos que la llevarán a cabo.

* 1. **Carga Inicial**

Debemos en este paso realizar la Carga Inicial al DW, poblando el modelo de datos que hemos construido anteriormente. Para lo cual debemos llevar adelante una serie de tareas básicas, tales como limpieza de datos, calidad de datos, procesos ETL, etc.

La realización de estas tareas puede contener una lógica realmente compleja en algunos casos. Afortunadamente, en la actualidad existen muchos softwares que se pueden emplear a tal fin, y que nos facilitarán el trabajo.

Se debe evitar que el DW sea cargado con valores faltantes o anómalos, así como también se deben establecer condiciones y restricciones para asegurar que solo se utilicen los datos de interés.

Cuando se trabaja con un esquema constelación, hay que tener presente que varias tablas de dimensiones serán compartidas con diferentes tablas de hechos, ya que puede darse el caso de que algunas restricciones aplicadas sobre una tabla de dimensión en particular para analizar una tabla de hechos se puedan contraponer con otras restricciones o condiciones de análisis de otras tablas de hechos.

Primero se cargarán los datos de las dimensiones y luego los de las tablas de hechos, teniendo en cuenta siempre, la correcta correspondencia entre cada elemento. En el caso en que se esté utilizando un esquema copo de nieve, cada vez que existan jerarquías de dimensiones, se comenzarán cargando las tablas de dimensiones del nivel más general al más detallado.

Concretamente, en este paso se deberá registrar en detalle las acciones llevadas a cabo con los diferentes softwares. Por ejemplo, es muy común que sistemas ETL trabajen con "pasos" y "relaciones", en donde cada "paso" realiza una tarea en particular del proceso ETL y cada "relación" indica hacia donde debe dirigirse el flujo de datos. En este caso lo que se debe hacer es explicar que hace el proceso en general y luego que hace cada "paso" y/o "relación". Es decir, se partirá de lo más general y se irá a lo más específico, para obtener de esta manera una visión general y detallada de todo el proceso.

Es importante tener presente, que al cargar los datos en las tablas de hechos pueden utilizarse preagregaciones, ya sea al nivel de granularidad de esta o a otros niveles diferentes.

**Caso práctico:**

Para simplificar la aplicación del ejemplo, el caso práctico solo se centrará en los aspectos más importantes del proceso ETL, obviando entrar en detalle de cómo se realizan algunas funciones y/o pasos.

El proceso ETL planteado para la ***Carga Inicial*** es el siguiente:

|  |
| --- |
| http://www.dataprix.com/files/uploads/250image/HEFESTO%20v2_0/xhefesto,P20-,P20carga,P20inicial.png.pagespeed.ic.dxLXu4L86Z.png  **Figura** **5.26:   Caso práctico, Carga Inicial.** |

Las tareas que lleva a cabo este proceso son:

* Inicio: inicia la ejecución de los pasos en el momento en que se le indique.
* Establecer variables Fecha\_Desde y Fecha\_Hasta: establece dos variables globales que serán utilizadas posteriormente por algunos pasos.
  + Para la variable "Fecha\_Desde" se obtiene el valor de la fecha en que se realizó la primera venta.
  + Para la variable "Fecha\_Hasta" se obtiene el valor de la fecha actual.
* Carga de Dimensión CLIENTE: ejecuta el contenedor de pasos que cargará la dimensión CLIENTE, más adelante se detallará el mismo.
* Carga de Dimensión PRODUCTO: ejecuta el contenedor de pasos que cargará la dimensión PRODUCTO, más adelante se detallará el mismo.
* Carga de Dimensión FECHA: ejecuta el contenedor de pasos que cargará la dimensión FECHA, más adelante se detallará el mismo.
* Carga de Tabla de Hechos VENTAS: ejecuta el contenedor de pasos que cargará la tabla de hechos VENTAS, más adelante se detallará el mismo.

A continuación, se especificarán las tareas llevadas a cabo por "***Carga de Dimensión CLIENTE***". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

|  |
| --- |
| http://www.dataprix.com/files/uploads/250image/HEFESTO%20v2_0/xhefesto,P20-,P20Carga,P20de,P20Dimension,P20CLIENTE.png.pagespeed.ic.oTC5dyr_y3.png  **Figura** **5.27:   Caso práctico, Carga de Dimensión CLIENTE.** |

* Obtener datos de OLTP: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión CLIENTE.

Se tomará como fuente de entrada la tabla “Clientes” del OLTP mencionado anteriormente.

Se consultó con l@s usuari@s y se averiguó que deseaban tener en cuenta solo aquellos clientes que no estén eliminados y que tengan su cuenta habilitada.

Es importante destacar que, aunque existían numerosos movimientos de clientes que en la actualidad no poseen su cuenta habilitada o que figuran como eliminados, se decidió no incluirlos debido a que el énfasis está puesto en analizar los datos a través de aquellos clientes que no cuentan con estas condiciones.

Los clientes eliminados son referenciados mediante el campo “Eliminado”, en el cual un valor “1” indica que este fue eliminado, y un valor “0” que aún permanece vigente. Cuando se examinaron los registros de la tabla, para muchos clientes no había ningún valor asignado para este campo, lo cual, según comunicó el encargado del sistema, se debía a que este se agregó poco después de haberse creado la base de datos inicial, razón por la cual existían valores faltantes. Además, comentó que en el sistema, si un cliente posee en el campo “Eliminado” un valor “0” o un valor faltante, es considerado como vigente.

Con respecto a la cuenta habilitada, el campo del OLTP que le hace mención es “Cta\_Habilitada”, y un valor “0” indica que no está habilitada y un valor “1” que sí.

Seguidamente, se expondrá la sentencia SQL que contiene este paso:

|  |
| --- |
| PIC    **Figura 5.28:   Caso práctico, CLIENTE - Obtener datos de OLTP.** |

* Cargar CLIENTE: almacena en la tabla de dimensión CLIENTE los datos obtenidos en el paso anterior.

A continuación, se especificará las tareas llevadas a cabo por "***Carga de Dimensión PRODUCTO***". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

|  |
| --- |
| http://www.dataprix.com/files/uploads/250image/HEFESTO%20v2_0/xhefesto,P20-,P20Carga,P20de,P20Dimension,P20PRODUCTO.png.pagespeed.ic.3lMA5LDF0H.png  **Figura** **5.29:   Caso práctico, Carga de Dimensión PRODUCTO.** |

* Obtener datos de OLTP: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la dimensión PRODUCTO.

Las fuentes que se utilizarán son las tablas “Productos” y “Marcas”.

En este caso, aunque existían productos eliminados, l@s usuari@s decidieron que esta condición no fuese tomada en cuenta, ya que había movimientos que hacían referencia a productos con este estado.

Es necesario realizar una unión entre la tabla “Productos” y “Marcas”, por lo cual se debió asegurar que ningún producto hiciera mención a alguna marca que no existiese, y se tomaron medidas contra su futura aparición.

El SQL que contiene este paso es el siguiente:

|  |
| --- |
| PIC    **Figura 5.30** **:   Caso práctico, PRODUCTO - Obtener datos de OLTP.** |

* *Cargar PRODUCTO: almacena en la tabla de dimensión PRODUCTO los datos obtenidos en el paso anterior.*

*A continuación, se especificarán las tareas llevadas a cabo por "****Carga de Dimensión FECHA****". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:*



|  |
| --- |
| **Figura 5.31** **:   Caso práctico, Carga de Dimensión FECHA.** |

Para generar esta tabla de dimensión, infaltable en todo DW, existen varias herramientas y utilidades de software que proporcionan diversas opciones para su confección. Pero, si no se cuenta con ninguna, se puede realizar manualmente o mediante algún programa, llenando los datos en un archivo, tabla, hoja de cálculo, etc, y luego exportándolos a donde se requiera.

Lo que se hizo, fue realizar un procedimiento que hace lo siguiente:

* Recibe como parámetros los valores de "Fecha\_Desde" y "Fecha\_Hasta".
* Recorre una a una las fechas que se encuentran dentro de este intervalo.
* Analiza cada fecha y realiza una serie de operaciones para crear los valores de los campos de la tabla de la dimensión FECHA:

|  |
| --- |
| PIC  **Figura 5.32:   Caso práctico, datos de FECHA.** |

* + idFecha = YEAR(fecha)\*10000 + MONTH(fecha)\*100 + DAY(fecha).
  + Año = YEAR(fecha).
  + Trimestre = CASE WHEN QUARTER(fecha) = 1 then '1er Tri' ... END.
  + – Mes = CASE WHEN MONTH(fecha) = 1 then 'Enero' ... END.
  + Inserta los valores obtenidos en la tabla de dimensión FECHA.

Como puede observarse, la clave principal "idFecha" es un campo numérico representado por el formato "yyyymmdd".

A continuación, se especificará las tareas llevadas a cabo por "***Carga de Tabla de Hechos VENTAS***". Este paso es un contenedor de pasos, así que incluye las siguientes tareas:

|  |
| --- |
| http://www.dataprix.com/files/uploads/250image/HEFESTO%20v2_0/xhefesto,P20-,P20Carga,P20de,P20Tabla,P20de,P20Hechos,P20VENTAS.png.pagespeed.ic.Q6qj69GMfK.png  **Figura** **5.33:   Caso práctico, Carga de Tabla de Hechos VENTAS.** |

* Obtener datos de OLTP: obtiene a través de una consulta SQL los datos del OLTP necesarios para cargar la tabla de hechos VENTAS.

Para la confección de la tabla de hechos, se tomaron como fuente las tablas “Facturas\_Ventas” y “Detalles\_Venta”. Al igual que en las tablas de dimensiones, se recolectaron las condiciones que deben cumplir los datos para considerarse de interés, y en este caso, se trabajará solamente con aquellas facturas que no hayan sido anuladas.

Se investigó al respecto, y se llegó a la conclusión de que el campo que da dicha información en “Anulada” de la tabla “Facturas\_Ventas” y si el mismo posee el valor “1” significa que efectivamente fue anulada.

Otro punto importante a tener en cuenta es que la fecha se debe convertir al formato numérico “yyyymmdd”.

Se decidió aplicar una preagregación a los hechos que formarán parte de la tabla de hechos, es por esta razón que se utilizará la cláusula GROUP BY para agrupar todos los registros a través de las claves primarias de esta tabla.

La sentencia SQL que contiene este paso fue la siguiente:

|  |
| --- |
| PIC    **Figura 5.34:   Caso práctico, VENTAS - Obtener datos de OLTP..** |

* Cargar VENTAS: almacena en la tabla de hechos VENTAS los datos obtenidos en el paso anterior.

* 1. **Actualización**

Cuando se haya cargado en su totalidad el DW, se deben establecer sus políticas y estrategias de actualización o refresco de datos.

Una vez realizado esto, se tendrán que llevar a cabo las siguientes acciones:

* Especificar las tareas de limpieza de datos, calidad de datos, procesos ETL, etc., que deberán realizarse para actualizar los datos del DW.
* Especificar de forma general y detallada las acciones que deberá realizar cada software.

**Caso práctico:**

Las políticas de Actualización que se han convenido con l@s usuari@s son las siguientes:

* La información se refrescará todos los días a las doce de la noche.
* Los datos de las tablas de dimensiones “PRODUCTO” y “CLIENTE” serán cargados totalmente cada vez.
* Los datos de la tabla de dimensión “FECHA” se cargarán de manera incremental teniendo en cuenta la fecha de la última actualización.
* Los datos de la tabla de hechos que corresponden al último mes (30 días) a partir de la fecha actual, serán reemplazados cada vez.
* Estas acciones se realizarán durante un periodo de prueba, para analizar cuál es la manera más eficiente de generar las actualizaciones, basadas en el estudio de los cambios que se producen en los OLTP y que afectan al contenido del DW.

Para evitar que se extienda demasiado la aplicación del ejemplo, el caso práctico solo incluirá lo que debería realizar el proceso ETL para actualizar el DW.

El proceso ETL para la actualización del DW es muy similar al de Carga Inicial, pero cuenta con las siguientes diferencias:

* Inicio: iniciará la ejecución de los pasos todos los días a las doce de la noche.
* Establecer variables Fecha\_Desde y Fecha\_Hasta:
  + La variable "Fecha\_Desde" obtendrá el valor resultante de restarle a la fecha actual treinta días.
  + La variable "Fecha\_Hasta" obtendrá el valor de la fecha actual.
* Carga de Dimensión CLIENTE: a la serie de tareas que realiza este paso, se le antecederá un nuevo paso que borrará los datos de la dimensión CLIENTE.
* Carga de Dimensión PRODUCTO: a la serie de tareas que realiza este paso, se le antecederá un nuevo paso que borrará los datos de la dimensión PRODUCTO.
* Carga de Dimensión FECHA: en este paso, en vez de recibir el valor de la variable "Fecha\_Desde", se tomará la fecha del último registro cargado en la dimensión FECHA.
* Carga de Tabla de Hechos VENTAS:
  + a la serie de tareas que realiza este paso, se le antecederá un nuevo paso que borrará los datos de la tabla de HECHOS correspondientes al intervalo entre "Fecha\_Desde" y "Fecha\_Hasta".
  + en el paso "Obtener datos de OLTP" se le agregará a la sentencia SQL la siguiente condición:
    - WHERE Facturas\_Venta.Fecha >= {Fecha\_Desde} AND Facturas\_Venta.Fecha <= {Fecha\_Hasta}

# BIBLIOGRAFIA

# Indicaciones

A continuación, se detallan aspectos de formato y redacción para la elaboración del informe.

Aspectos de formato y redacción

1. Considere las siguientes instrucciones de redacción y formato:

* Redactar en tercera persona singular tiempo presente. En lo posible, se debe cuidar que los párrafos completos estén en una misma hoja, evitando las líneas viudas o huérfanas. Esto permite captar mejor la idea que encierra el párrafo completo.
* Formato de entrega: documento Word y PDF.
* Papel tamaño carta. Márgenes estándar (superior e inferior de 2,5 cm. izquierdo y derecho de 2,5 cm).
* Párrafos: alineación justificada, interlineado sencillo.
* Tipo de letra o fuente: Arial o Calibri (Cuerpo).
* Tamaño títulos: 14 y en negrita
* Tamaño subtítulos: 12 y en negrita.
* Tamaño textos: 11 normal.

**No olvides respetar las reglas ortográficas y de redacción**





1. Incluir un listado de las fuentes bibliográficas utilizadas en la recopilación de información, con el título “Referencias bibliográficas”, según Norma APA 6° Edición. Se sugiere descargar la Guía para citas y referencias bibliográficas según Norma APA, en el sitio web de INACAP, en la Red de Bibliotecas:

<http://www.inacap.cl/tportalvp/red-de-bibliotecas-inacap>

Recopilar también distintos insumos gráficos (imágenes, fotos, diagramas, entre otros) que permitan complementar la información a presentar.

Una vez finalizado el informe, elimina las instrucciones e indicaciones y ejemplos.

Recuerda completar el pie de página y los datos de la portada con el nombre del Área académica y nombre de tu carrera.