

**UNIVERSIDAD GABRIELA MISTRAL
FACULTAD DE INGENIERIA**

**MODELO DE INTELIGENCIA DE NEGOCIOS
PARA LOS DISTRIBUIDORES ABARROTEROS
DE LA X REGION**

Memoria para optar al título de Ingeniero de Ejecución en Informática

Autor : César Mauricio Faúndez Faúndez
Profesor Guía : Marcelo Ramírez M.
Profesor Integrante : Roberto Carú C.

Puerto Varas – Chile
Noviembre, 2014

INDICE

Capítulo I - Introducción	01
Capítulo II - Marco Teórico	03
1. Hipótesis	03
2. Inteligencia de Negocios	03
3. DSS y Data Warehousing	06
3.1 Sistemas Fuente	08
3.2 Área de preparación de datos	09
3.3 Data Warehouse y Data Mart	12
3.4 OLAP, procesamiento analítico en línea	15
3.5 Servidor de presentación y Herramientas BI	18
4. Modelo Dimensional	20
4.1 Componentes del modelo dimensional	22
4.2 Esquema	26
4.3 Tipos de Esquema	30
5. Metodología Hefesto	34
5.1 Análisis de Requerimientos	35
5.2 Análisis de los OLTP	35
5.3 Modelo Lógico del DW	36
5.4 Procesos ETL	37
Capítulo III - Desarrollo del Trabajo	38
1. Adecuación de la metodología Hefesto	39
2. Análisis de requerimientos	41
2.1 Herramienta de recolección de información	41
2.2 Población de estudio	42
2.3 Nivel de inteligencia actual	42
2.4 Identificación de indicadores y dimensiones de análisis	44
3. Modelo Conceptual	50
4. Requerimientos para los OLTP	53
4.1 Existencia de procesos	53

4.2 Requerimiento de datos	54
4.3 Modelo conceptual ampliado y modelo dimensional	56
5. Modelo lógico del Data Warehouse	60
5.1 Tipo de Modelo Lógico del DW	61
5.2 Tablas de dimensiones	61
5.3 Tablas de hechos	63
5.4 Esquema completo.....	66
Capítulo IV - Conclusiones.....	67
Glosario.....	74
Bibliografía	78
Anexos	80

II. INTRODUCCIÓN

El retail en Chile es un mercado dedicado a la venta al por mayor y al detalle el cual ha ido evolucionando a través del tiempo. En un principio este se componía de pequeños y medianas empresas dedicadas a vender productos de primera necesidad las cuales en la década de los 80 y 90 empezaron a expandirse, con el principal objetivo de conseguir nuevas y mejores ubicaciones. Hoy en día debido a los acelerados cambios e innovaciones el panorama es otro, las preocupaciones se centran en la diferenciación y generación de nuevas oportunidades destinadas a satisfacer al cliente.

Esta competitividad ha vuelto a este sector muy agresivo en donde inevitablemente el más débil va desapareciendo sin saber que se puede hacer algo al respecto. En Estados Unidos la tendencia es que estos mercados pequeños se unan para poder competir, apoyados en tecnologías de información orientadas a la Inteligencia de Negocios, herramientas que las grandes empresas han explotado generando grandes beneficios.

En el sur de Chile este es un escenario muy lejano, las ventajas competitivas que podrían entregar estas tecnologías de información no son conocidas por todos segmentos del retail. Por ejemplo en el sector abarrotero las transnacionales avasallan la competencia y los distribuidores de mercados pequeños se han visto muy trastocados sobre todo con la entrada al mercado de filiales mayoristas. Ellos no conocen los beneficios de la Inteligencia de Negocios y viven analizando a pulso las posibles tendencias que los puedan posicionar un paso adelante, pero la realidad es otra, son organizaciones presas de la incertidumbre.

Esta tesis se enfoca justamente en este segmento en específico, empresas dedicadas al negocio de la distribución de abarrotes principalmente a negocios de barrio y pequeños mercados particulares en la región de los Lagos. Sus procesos de negocios más básicos están cubiertos por herramientas tecnológicas que no necesariamente son

sistemas y su principal nexo con el cliente son los vendedores o la fuerza de venta, la cual está capacitada para ser gestores comerciales, entregar asesoría, datos e información del negocio.

Investigaciones que se realizaron indican que las personas que se encargan de tomar las decisiones en estas empresas distribuidoras saben mucho sobre la teoría de Inteligencia de Negocios pero nunca se han atrevido a realizar una implementación de este tipo, excusándose en que ya tienen todo resuelto o de que el mercado es muy hermético como para crear una inteligencia más grande de la que manejan a través de planillas Excel, visitas a terreno o sus propios sistemas transaccionales. La idea es demostrar que se puede lograr con Inteligencia de Negocios, introducirlos a la tecnología y demostrarles que no hay que inventar la rueda porque muchas cosas ya se encuentran resueltas, que esta es la ventaja que necesitan para competir en el nuevo mercado de retail.

Realizar Inteligencia de Negocios sobre el cliente en este tipo de mercados va muy de la mano con tener a disposición los datos del mismo, los cuales generalmente se pueden encontrar en sus propios sistemas operacionales. Para las empresas distribuidoras no es así, ellos tienen los datos sobre cada uno de sus clientes pero estos, típicos negocios de barrio, también tienen sus propios datos los cuales conforman la dimensión a la cual no tienen acceso. La obtención de estos datos implica otro tipo de proyecto el cual la tesis no pretende abarcar, por lo que se omitirá y se dará énfasis en determinar cuáles son las necesidades de análisis.

II. MARCO TEÓRICO

1. Hipótesis

Generar un modelo de Inteligencia de Negocios para detectar las necesidades de análisis de los distribuidores abarroteros de la región de los Lagos apoyando la toma de decisiones.

Adicionalmente se desea indagar si este tipo de implementación puede representar un nuevo modelo de negocio que beneficie no solo al sector abarcado sino que transversalmente a todo el segmento retail pequeño que intenta competir contra las grandes empresas.

2. Inteligencia de Negocios.

La Inteligencia de Negocios es una rama de la informática que pretende aprovechar los datos generados por los sistemas transaccionales y en general cualquier tipo de dato que sea importante para la organización para tener el conocimiento necesario para tomar decisiones en tiempo real y transformar las empresas en entidades proactivas que puedan realizar acciones al identificar y corregir situaciones antes de que se conviertan en problemas, encontrar nuevas oportunidades o readaptarse frente la ocurrencia de sucesos inesperados, generalmente en el contexto del cliente pero también a todo nivel de la operación de la empresa. Es por esta razón que se nombra el término "Inteligencia", y mientras más relevante y útil sea esta inteligencia se podrán tomar mejores decisiones.

Existen muchas definiciones de Inteligencia de Negocios, las cuales se han ido depurando a través del tiempo debido a constante cambio de las necesidades de información y la evolución tecnológica, pero destaco la siguiente:

“La Inteligencia de Negocios son los procesos, tecnologías y herramientas necesarias para convertir los datos en la información y la información en conocimiento y el conocimiento en los planes que conducen a una acción de negocio rentable. BI abarca almacenamiento de datos, análisis de negocio y la gestión del conocimiento”. The Data Warehouse Institute, 2002

Desde aquí Inteligencia de Negocios, Business Intelligence o BI se refieren al mismo concepto.

Como se puede apreciar en la definición, la Inteligencia de Negocios tiene como objetivo que la entidad que la utiliza pueda tomar acciones que sean rentables para el negocio en base a los datos que esta genera, pero para ello se debe pasar por un proceso donde los datos se transformen en información y la información se transforme en conocimiento. En la siguiente figura se ilustra la definición:

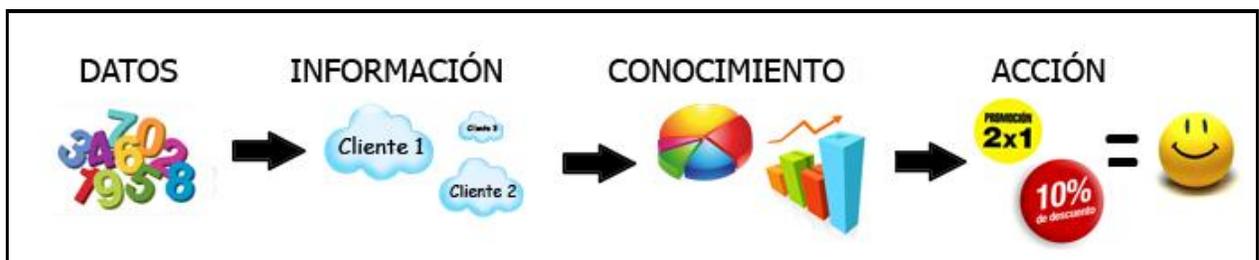


Figura 1. Concepto de Inteligencia de Negocios

Cada etapa del proceso implica un estado en el cual se encuentra el negocio en lo que respecta a inteligencia. Todas las empresas con adecuación tecnológica informática a nivel de sistemas transaccionales se encuentran en la etapa de “Datos”. Con los datos es posible mantener la parte operativa de la empresa pero de ninguna manera esta podrá predecir sucesos ni mucho menos reaccionar ante situaciones antes de que estas sucedan. En esta etapa la empresa es reactiva, por lo que reaccionará a medida que

vayan dando los acontecimientos. Una empresa a nivel de datos subsiste en un ambiente competitivo que no es agresivo, es decir, en donde no existe mucha competencia.

Cuando el segmento en donde se encuentra la empresa es más competitivo es necesario manejar “información” en vez de datos. En Chile es común ver este tipo de organizaciones las cuales realizan volcados de datos en planillas dándole un contexto, en donde el principal requisito es preguntarse lo que se necesita saber para atender ciertas necesidades. En Inteligencia de Negocios esto se logra a través de herramientas de almacenamiento que consolidan los diferentes datos de la empresa en una base de datos homogénea la cual es conocida como “Data Warehouse”, concepto que será explicado a fondo en capítulos posteriores.

Para lograr el “Conocimiento” se necesita analizar la información. La base de la inteligencia se encuentra en cómo utilizar la información obtenida, por lo que se necesita contar con un gran conocimiento sobre el negocio. Generalmente las empresas cuentan con analistas los cuales se especializan en crear conocimiento, tarea que se vuelve más sencilla con Inteligencia de Negocios ya que proporciona una serie de herramientas BI que contienen información estratégica ordenadas en reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, entre otros.

La “acción” la toma el usuario que tiene acceso a la información y se logra cuando se tiene el conocimiento suficiente para realizarla. No se debe considerar como una etapa, sino como un resultado de transformar los datos en información y la información en conocimiento. Como se puede observar en la figura 1 las acciones pueden repercutir en la sensación de satisfacción del cliente, si el contexto del conocimiento adquirido es en relación a él, pero como se decía anteriormente las acciones pueden realizarse a todo nivel, dependiendo de qué necesidad resuelva la solución de Inteligencia de Negocios implementada. Una empresa que haya alcanzado el nivel de conocimiento como para realizar acciones en base a él, ya no es una empresa reactiva sino que una empresa proactiva, por lo tanto puede sobrevivir en un ambiente competitivo más agresivo.

Como se puede deducir, llegar a ser una empresa que pueda tomar mejores decisiones en base al conocimiento adquirido se puede lograr lentamente de forma orgánica a través de planillas y otros métodos manuales, pero la diferencia con la Inteligencia de Negocios es que esta entrega una serie de herramientas y tecnologías que permiten acelerar todo este proceso. Esto es de vital importancia ya que no sirve tomar una acción tarde, a pesar de que haya sido la mejor alternativa. Para poder estar preparado ante la incertidumbre y aprovechar las oportunidades que entrega el entorno se debe contar con la información a tiempo, por lo que se debe generar el conocimiento cuanto antes.

3. *DSS y Data Warehousing*

DSS o Sistema de Soporte a las Decisiones por sus siglas en inglés, es un término que regularmente se encuentra en la documentación relacionada con Inteligencia de Negocios y su utilización depende mucho del punto de vista de la persona que lo utilice, algunos profesionales del área los consideran iguales, otros no se convencen y hablan de Inteligencia de Negocios como una evolución del DSS. Un DSS en definitiva es usado para referirse a un sistema informático que utiliza información y modelos matemáticos para ayudar a los usuarios a tomar decisiones empresariales adecuadas según las condiciones del mercado y la situación interna de la compañía.

Un sistema no es más que un conjunto de componentes relacionados entre sí que contribuyen a un determinado objetivo, caracterizado por tener entradas y salidas, siendo su resultado afectado por condiciones externas e internas al sistema. ¿Pero Inteligencia de Negocios será un sistema?

La Inteligencia de Negocios se compone de diferentes tecnologías y herramientas que necesariamente deben estar relacionadas entre sí para lograr un objetivo común que es el mismo al de un DSS, ser soporte para la toma de decisiones. La entrada característica para ambas definiciones son los datos los cuales se convierten en información, para entregar como salida el conocimiento necesario para realizar

acciones, de hecho algunas herramientas BI tienen la funcionalidad de entregar información estadística en base a diferentes algoritmos matemáticos. Además el contexto sobre el conocimiento que explotará la Inteligencia de Negocios en la empresa donde se implemente siempre es definido por las necesidades internas y las condiciones competitivas externas.

Por todo lo anterior podemos deducir que los términos Inteligencia de Negocios y DSS se refieren a lo mismo, aunque este último es una definición más global en donde podría estar clasificado cualquier sistema que apoye a la toma de decisiones en una empresa, como por ejemplo una o varias planillas Excel preparadas por un analista en base a diferentes datos obtenidos en los sistemas fuente de su organización y que aporte el conocimiento necesario para realizar acciones estratégicas también puede ser considerado un DSS.

Aun así el término de Data Warehousing también es confundido como un DSS y es importante destacar su diferencia. El Data Warehousing es el principal proceso de la Inteligencia de Negocios en el cual convergen una serie de tecnologías que posibilitan la extracción de datos desde diferentes sistemas fuente de la empresa, ya sea de sus sistemas transaccionales o de otras fuentes externas, los consolida en una base de datos diferente y finalmente provee la información transformada y resumida a los usuarios finales en un servidor de presentación a través de diferentes herramientas de BI. Este proceso es ilustrado en la figura 2.

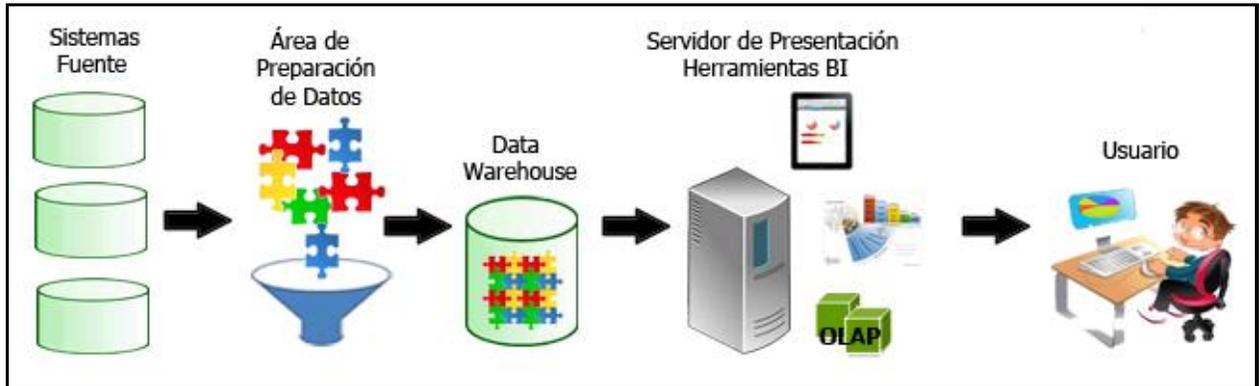


Figura 2. Proceso de Datawarehousing

Por lo tanto el Data Warehousing no es un DSS, pero si es válido decir que Inteligencia de Negocios es un DSS basado en el proceso de Data Warehousing.

3.1 *Sistemas Fuente*

Son los sistemas transaccionales que tradicionalmente poseen las empresas para soportar y capturar sus procesos operacionales, por lo que sus prioridades son el uptime (tiempo sin caída) y la disponibilidad.

Comúnmente son llamados sistemas heredados porque fueron desarrollados hace bastante tiempo pero aún siguen funcionando debido a que no se quieren reemplazar o no se pueden actualizar de forma sencilla. Generalmente fueron adquiridos para resolver un problema específico y se han ido actualizando de acuerdo a las necesidades de la empresa. Esta característica de software hecho a medida hace que los datos que contiene sean muy difíciles de integrar con otro sistema.

Cuando hablamos de sistemas fuentes es necesario mencionar el término OLTP o procesamiento de transacciones en línea por sus siglas en inglés. Un sistema fuente es OLTP y su principal funcionalidad es que las consultas realizadas a su base de datos son para responder las necesidades del negocio en tiempo real en base a CRUD (copia, lectura, actualización y borrado) y se encuentran fuertemente restringidas a la

estructura del sistema heredado. Un sistema OLTP no está hecho para guardar un histórico de datos muy amplio por lo que cada cierto tiempo sus bases de datos se deben respaldar en dispositivos externos.

La Inteligencia de Negocios genera información en base a los datos de los sistemas fuentes, pero también puede utilizar datos provenientes de fuentes externas o planillas, todo dependiendo del conocimiento que se quiera obtener. Para este marco teórico, se hablará de sistemas fuentes u OLTP en ambos casos.

3.2 Área de preparación de datos

Uno de los objetivos del Data Warehousing es tomar los datos de los sistemas OLTP y volcarlos dentro de una nueva base de datos, tarea que no es sencilla debido a la naturaleza de los sistemas fuente que los hace difícil de integrar. Ralph Kimball, uno de los investigadores más destacados en el área del Data Warehousing, define al área de preparación de datos como un área donde existen un conjunto de procesos que limpian, transforman, combinan, deduplican (eliminan datos redundantes), consolidan, guardan y preparan los datos extraídos de los diferentes sistemas fuente para luego ser usados en el Data Warehouse.

El proceso más destacado dentro de esta etapa es el llamado ETL o “Extracción, Transformación y Carga” por sus siglas en Ingles. Su diseño y desarrollo constituye entre un 60% a un 80% del esfuerzo de un proyecto de Inteligencia de Negocios, por lo que es el proceso más importante. En la figura 3 se puede apreciar el proceso ETL.

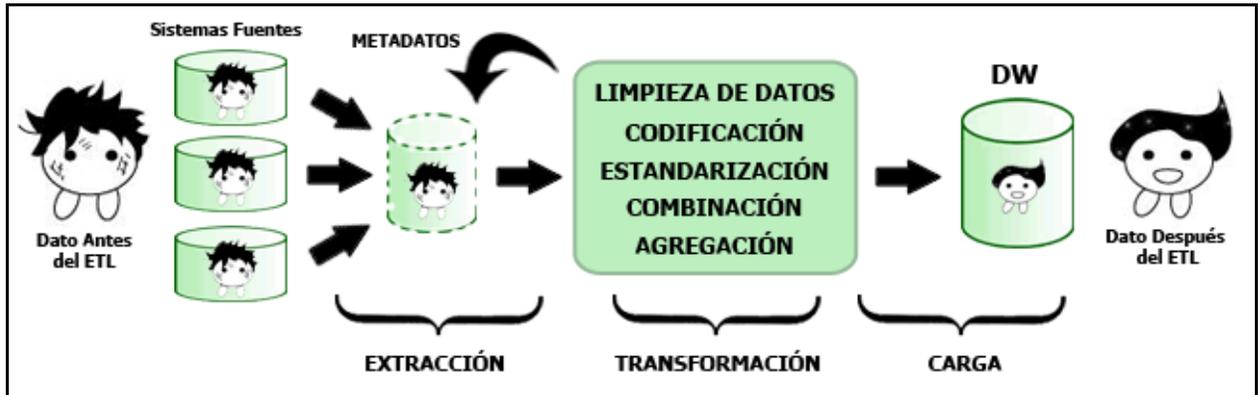


Figura 3. Proceso ETL

a) Extracción: en base a los requisitos y necesidades de los usuarios se exploran los diferentes sistemas fuentes que se tengan a disposición y se extrae la información relevante, dependiendo del contexto que se quiera analizar. La extracción puede ser sencilla si los datos se extraen de base de datos relacionales, a través de consultas SQL o rutinas programadas, pero si estos se encuentran en planillas o fuentes externas, como hojas de cálculo o archivos de texto, la extracción es más difícil y generalmente se usan herramientas prediseñadas específicas para cada caso.

Generalmente estos datos son extraídos en horarios donde los sistemas fuente no sean ocupados o se planifica un horario en caso de que las operaciones de la empresa sean 24/7 y son alojados en una base de datos intermedia de tal manera que no se entorpezcan los procedimientos diarios de la organización y del mismo Data Warehouse que es la base de datos donde finalmente serán alojados los datos luego de su transformación. Este procedimiento habitualmente es automático a través de un sistema de procesamiento por lotes (Batch).

En esta base de datos intermedia también son alojados los metadatos, los cuales incluyen información sobre los datos que serán volcados al Datawarehouse, como su ubicación, estructura y método de transformación en caso de que la haya tenido.

b) Transformación: luego de extraer los datos estos deben ser convertidos en un conjunto de datos congruentes y compatibles, sobre todo si son obtenidos desde diferentes sistemas fuente, para que los datos que sean volcados al Data Warehouse estén integrados, lo cual es un requisito fundamental en el proceso de Data Warehousing. Como se puede apreciar en la figura 3, nos encontraremos con diferentes niveles de transformación:

- *La limpieza de datos* realiza diferentes tipos de acciones contra los datos erróneos, inconsistentes o irrelevantes, como lo son la corrección de faltas de ortografía, arreglar los conflictos de dominio (por ejemplo que una ciudad no corresponda a un país), ignorarlos, eliminarlos o reemplazar el valor. También es válido en esta etapa crear valores nuevos si estos no existen.
- *La codificación* busca encontrar atributos en común que generalmente tienen diferente código y estandarizarlos, como por ejemplo si para un sistema fuente el atributo estado es 0 y 1 y para otro es “Activado” y “Desactivado” se elige una de las dos codificaciones o se crea una nueva. Existe un caso específico que se aplica a las claves primarias de los datos extraídos que para lograr una integración total se eliminan y se generan claves suplentes, las cuales serán los nuevos identificadores en el Data Warehouse.
- *La estandarización* busca que los datos signifiquen lo mismo, como son extraídos de diferentes sistemas fuente, los datos como las medidas o los pesos deben ser estandarizados, es decir que si nos encontramos con atributos medibles en kilos y toneladas, los datos sean transformados todos a kilos o si se miden en centímetros o metros todos sean medidos en metros. También considera la convención de nombres de los atributos, es decir si tenemos una columna que se llame “Nombre” y en otro sistema tenemos una columna que se llame “Nombre_Persona” y se refieren al mismo dato igualmente se tendrá que estandarizar.
- *La combinación* busca hacer coincidir varios datos en sus valores clave o en coincidencias parciales de los atributos no clave y combinarlos. Por ejemplo

desde un sistema contable y un sistema hotelero podríamos extraer más información de un cliente haciendo coincidir su atributo RUT.

- *La agregación* busca realizar cálculos de datos, como puede ser la sumarización de ventas en un mes. Este proceso puede ser realizado directamente desde los datos ya integrados en el Data Warehouse pero el realizarlos en esta etapa permite aumentar la performance de las consultas, debido a que no se ocupara procesamiento en realizar los cálculos más adelante.

c) Carga: al finalizar el proceso de transformación los datos son cargados al Data Warehouse generalmente a través de una carga masiva. Existen 2 tipos de carga.

- *La carga inicial* que se refiere a la primera carga que se realizará al Data Warehouse, caracterizada por contener registros históricos que idealmente deberían alcanzar los 5 años. Por esta razón realizar esta carga consume un tiempo bastante considerable.
- *La actualización o refresco periódico* como su nombre lo indica son cargas que se realizan cada cierto tiempo, con una demora de tiempo planificada. A diferencia de la carga inicial, la cantidad de datos volcados al Data Warehouse es menos ya que consideran solo datos nuevos que se fueron generando desde la última actualización. Como requisito fundamental los nuevos datos deben ser analizados antes de su carga para asegurar la consistencia del Data Warehouse y preferentemente deben ser indexados una vez cargados.
- *La carga total* se realiza cuando alguna actualización consume demasiado esfuerzo y tiempo, o por motivo particular no puede llevarse a cabo y consiste en volver a cargar todos los datos desde 0.

3.3 Data Warehouse y Data Mart

Hasta el momento se ha mencionado mucho el concepto de Data Warehouse y solo se ha dicho que es una base de datos en donde se alojan todos los datos que han pasado

por el área de preparación de datos de manera centralizada, sin entregar un mayor detalle, pero sin duda es la herramienta central de toda solución de Inteligencia de Negocios. Desde aquí en adelante Data Warehouse, almacén de datos o DW corresponden al mismo concepto.

La definición más popular de DW fue acuñada por William H. Inmon, considerado hasta los días de hoy el padre del Data Warehouse. Debido a su importancia se entregará parte de la definición sin traducir:

“Data Warehouse es una colección de datos subject oriented, nonvolatile, integrated, time variant para ayudar al proceso de toma de decisiones gerenciales”.

a) Subject Oriented: la traducción directa de esta definición sería “orientada al sujeto”, pero acá Inmon se refiere a sujetos de negocio. Un sujeto de negocio es una entidad dentro de todo el contexto de datos de particular interés analítico para la empresa. Esta propiedad caracteriza a un Data Warehouse ya que los datos que contiene son utilizados especialmente en el proceso de toma de decisiones, en cambio un sistema transaccional está pensado para soportar los procesos operacionales del negocio. Por ejemplo en un DW podremos encontrar entidades como clientes, productos, zonas o tiempo en contraste con los sistemas fuente que pueden contener código de producto, stock, compras o ventas. La ventaja de contar con entidades de alto nivel en un DW es asegurar una alta accesibilidad en los datos.

b) Nonvolatile: los datos en un Data Warehouse no son volátiles porque estos son estables, en el tiempo pueden ser agregados más datos pero los datos existentes no son removidos. La confianza para un buen análisis es que esta se base en información estable por lo que la característica de mantener datos históricos de la organización es fundamental. Esto es posible debido a que en un DW existen solamente 2 tipos de operaciones: la carga de datos y el acceso a ellos a través de consultas de solo lectura.

c) Integrated: que un Data Warehouse sea integrado significa que este contiene datos de diferentes sistemas fuentes que fueron previamente consolidados en el proceso ETL.

Como ya se mencionó, este proceso realiza una depuración en los datos lo cual permite que la utilización del almacén de datos por parte del usuario no se vea mermado por temas relacionados sobre la confiabilidad y validez de los mismos.

d) Time Variant: la principal entidad que se encontrara en un Data Warehouse es el tiempo y es la razón por la que es variante en el tiempo. Todos los datos del DW están asociados a un periodo de tiempo específico lo que significa que se podrán acceder a diferentes versiones de una misma información. El intervalo de tiempo y la periodicidad de los datos se definen de acuerdo a las necesidades y requisitos de los usuarios.

Años más tarde Ralph Kimball se dedicaría a estudiar a fondo los conceptos asociados al Data Warehousing, logrando ser uno de los principales exponentes del área. Para él un Data Warehouse es *“una copia de los datos transaccionales específicamente estructurados para consultas y análisis”*.

Adicionalmente propone el concepto Data Mart, el cual se refiere a un subconjunto del Data Warehouse, el cual puede ser implementado como un pequeño proyecto de DW generalmente orientado a un solo proceso de negocio o a un grupo de ellos relacionados en torno a un proceso de negocio en particular. Desde otra perspectiva un Data Warehouse está compuesto de un conjunto de Data Mart los cuales son fácilmente integrables debido a la particularidad de que la construcción de la base de datos se teniendo en cuenta su compatibilidad.

Kimball apunta a una perspectiva bottom-up en donde la construcción de un Data Warehouse completo, que puede resultar un proyecto que necesite de mucho esfuerzo, pueda realizarse a través de pequeños proyectos de Data Mart, realizándolos a medida que la empresa los vaya requiriendo, en contraste a la perspectiva clásica top-down de Inmon que apunta a la construcción completa de un Data Warehouse desde un inicio.

En resumen un Data Warehouse es una base de datos centralizada que se compone de datos extraídos desde diferentes fuentes, ya sean externas o internas los cuales han sido especialmente depurados y modificados para que estos puedan ser analizados

fácilmente por usuarios que tienen la tarea de tomar decisiones en la organización, sin la necesidad que estos tengan conocimientos informáticos avanzados.

3.4 OLAP, procesamiento analítico en línea.

En el ambiente de Inteligencia de Negocios nos encontraremos continuamente con la palabra OLAP la cual es acrónimo en inglés de procesamiento analítico en línea. El concepto OLAP está asociado a una solución que tiene como objetivo agilizar la consulta de grandes cantidades de datos a través de estructuras multidimensionales para ser utilizados en análisis. Por lo general los sistemas OLAP deben:

- Soportar requerimientos complejos de análisis.
- Analizar datos desde diferentes perspectivas.
- Soportar análisis complejos contra un gran volumen de datos.

Actualmente existen muchos tipos de herramientas OLAP ofrecidas en el mercado, y debido a esto se ha generado cierta confusión respecto al término, por eso es muy importante averiguar primero que está ofreciendo el proveedor antes de aceptar una implementación de este tipo en la empresa. Dentro de estas herramientas podremos encontrar:

- Programas para realizar ETL desde base de datos transaccionales.
- Motores para soportar bases de datos multidimensionales
- Diseñadores de cubos OLAP
- Generadores de reportes
- Graficadores
- Navegadores de datos

Un cubo OLAP es un subconjunto de información extraída del almacén de datos el cual contiene datos organizados y sumariados dentro de una estructura multidimensional, lo que permite a los usuarios acceder a dimensiones de información específicas sin

necesidad de consultarlos directamente en el DW. Esto implica una reducción considerable de tiempo al momento de consultar, ya que el almacén de datos generalmente tiene grandes cantidades de información, la cual corresponde a datos históricos de la organización. En la figura 4 se puede apreciar esta diferenciación:

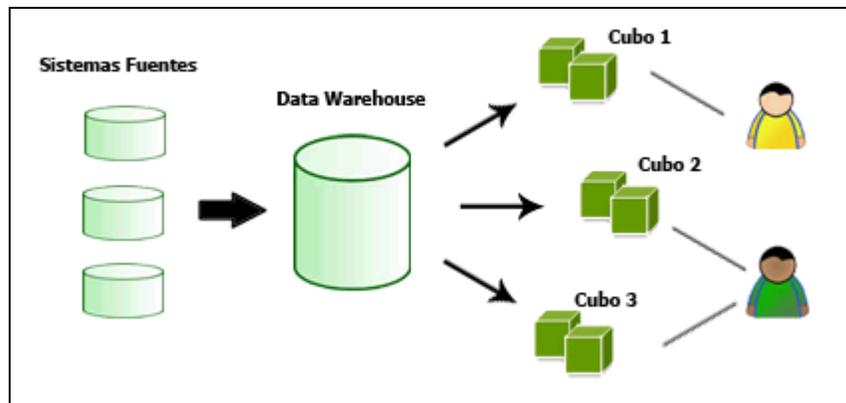


Figura 4. Cubos OLAP

Físicamente los cubos OLAP están compuestos por dimensiones y medidas.

Las dimensiones son el atributo estructural de un cubo el cual está compuesto por una serie de datos que pertenecen a una categoría similar según la percepción del usuario, como lo pueden ser clientes, productos o localidad.

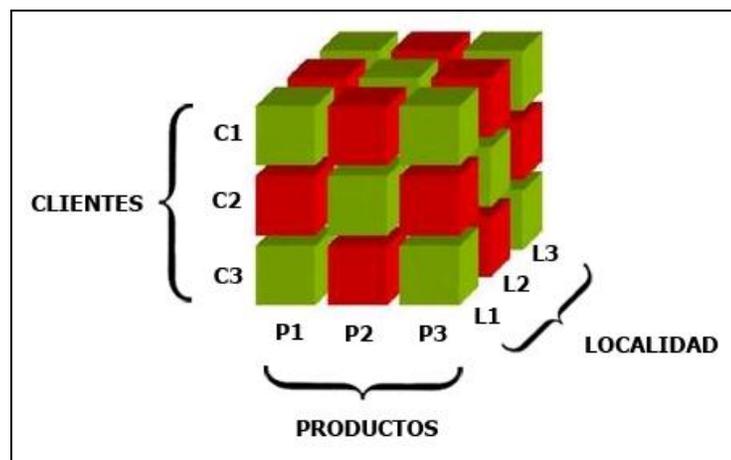


Figura 5. Dimensiones en cubo OLAP

Las medidas por su parte son datos numéricos de interés para los usuarios del cubo y que pueden ser analizadas desde diferentes perspectivas o dimensiones. En un cubo las medidas, también conocidas como indicadores, son precalculados para ayudar a disminuir los tiempos de respuestas en los procesos de búsquedas de información, dentro de estos cálculos se encuentran las agregaciones o sumas de datos. Algunas medidas comunes son ventas en unidades, ventas en pesos, gastos, conteo de la producción, presupuesto, entre otros. En la figura 5 por ejemplo podríamos consultar por las ventas en pesos a cada cliente por cada producto por cada localidad.

Además podremos encontrar diferentes categorías de OLAP diferenciados por su forma de funcionamiento y estructura:

a) ROLAP: procesamiento analítico relacional en línea, es un sistema que ha sido construido utilizando herramientas OLAP y una base de datos relacional. Los cubos son generados dinámicamente en el momento de generar las consultas, por lo que el manejo de estos es transparente para el usuario posibilitando consultas “ad hoc” o directas a la base de datos. Esta característica también se traduce directamente en una desventaja en cuanto al tiempo de respuesta de una consulta debido a que cada vez que se realizan, los cubos se deben volver a calcular.

En un sistema ROLAP realmente no existen los cubos pero si los simulan. Para el usuario analista es como si estuviera usando un cubo OLAP pero en realidad está trabajando directamente sobre una base de datos relacional.

b) MOLAP: procesamiento analítico multidimensional en línea, a diferencia del ROLAP los datos son almacenados en una base de datos multidimensional de tal manera que la representación interna y externa coincidan. Los cubos son generados de forma previa para que posteriormente sean consultados obteniendo así una performance alta. A pesar de ello cada vez que se requiera realizar cambios a un cubo este debe ser recalculado, por lo que tiene poca flexibilidad.

c) **HOLAP**: procesamiento analítico híbrido en línea, que es una combinación de ROLAP y MOLAP. Los datos precalculados y agregados se almacenan en base de datos multidimensionales y los de mayor detalle en estructuras relacionales.

3.5 Servidor de presentación y Herramientas BI

El servidor de presentación es la maquina física en el cual los datos del Data Warehouse se organizan y almacenan en un marco de trabajo dimensional, generalmente cubos, para que los usuarios finales lo consulten directamente.

Para realizar esta consulta los usuarios usan diferentes herramientas BI las cuales son consideradas también como un cliente del Data Warehouse ya que mantienen una sesión con el servidor de presentación, enviando flujos de peticiones SQL. Estas herramientas generalmente presentan en pantalla informes, gráficos o alguna otra forma superior de análisis, de tal manera que los usuarios no se preocupen por conocer cuáles son las características y funcionalidades de las estructuras de datos utilizadas y solo se enfoque en el análisis. Algunas herramientas BI son descritas a continuación:

a) **Herramientas de consulta Ad Hoc**: esta herramienta en especial necesita de ciertos conocimientos para poder ser utilizada, ya que el acceso a datos es la manipulación directa de tablas y relaciones en la base de datos, la cual presta una especial utilidad a la hora de analizar. Lamentablemente solo un 10% de los usuarios de la implementación de BI pueden utilizar esta herramienta mientras que el resto necesita de aplicaciones prediseñadas.

b) **Herramientas de reporte y consultas**: permite realizar consultas y crear reportes a través de pantallas graficas intuitivas. El usuario debe seguir una serie de pasos simples como seleccionar opciones de un menú o presionar botones para especificar los diferentes elementos que contenga el reporte, así como sus criterios de agregación.

c) Herramientas Dashboard: herramienta conformada por un conjunto de reportes altamente resumidos que son presentados generalmente a través de gráficos (torta, barras, semáforos, entre otros) de tal forma que represente un cuadro de mando en donde se pueda monitorear, controlar y gestionar los procesos de la empresa en una sola mirada. Uno de los elementos característicos son los KPI o indicadores de rendimiento, los cuales miden el desempeño de un proceso. En la figura 6 se puede observar el ejemplo de un dashboard.



Figura 6. Dashboard

d) Herramientas Data Mining: herramienta que cuenta con una poderosa tecnología en base a cálculos matemáticos para poder analizar y extraer conocimientos ocultos en los datos de un DW que permiten predecir o estimar comportamientos futuros y determinar eventos que provocan comportamientos específicos. De esta manera se

pueden responder preguntas como ¿Cuánto gasto aproximado tendremos el próximo mes? o ¿Cuánta ocupación de brasileños tendremos en nuestro hotel en el verano próximo?, las cuales permitirán tomar decisiones proactivas basadas en un conocimiento obtenido por la información acumulada en la organización.

Las herramientas de Data Mining utilizan métodos matemáticos como estadística, redes neuronales y sistemas expertos.

4. Modelo Dimensional

El modelo dimensional permite representar la estructura lógica de las bases de datos dimensionales de un Data Warehouse de manera que se puedan reunir las principales metas de diseño:

- Presentar la información a los usuarios de la forma más simple posible.
- Retornar los resultados a los usuarios lo más rápido posible.
- Proveer información relevante que guarde pistas de los procesos subyacentes.

A pesar de que se encuentra dentro de una metodología diferente a la que ocupará esta tesis, su entendimiento entregará los conceptos básicos que encontraremos en cualquier solución de Inteligencia de Negocios. La aplicación del modelo dimensional la podemos encontrar en la fase del diseño lógico descrita el año 1996 por Ralph Kimball en su metodología “Ciclo de vida Dimensional del Negocio” como propuesta para el diseño de Data Warehouse, partiendo de la visión multidimensional que los usuarios tienen de los datos de la empresa al momento de analizarlos.

Generalmente es comparado con el modelo entidad – relación, pero este último define la estructura física de una base de datos OLTP, en cambio el modelo dimensional define una visión lógica de los datos desde una perspectiva de negocio y no corresponde a la estructura física del DW.

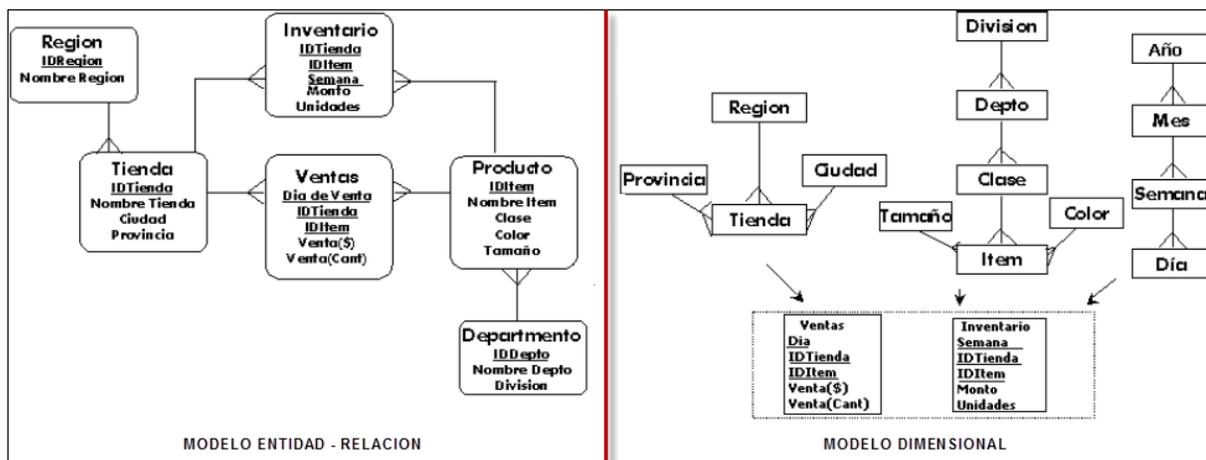


Figura 7. Modelo entidad relación y modelo dimensional

Como se puede apreciar en la figura 7, un modelo dimensional es agrupado en categorías o dimensiones de negocio que tienen sentido para el usuario analista, las cuales siempre estarán relacionadas con los indicadores del negocio contenidas en las tablas de hecho.

En un modelo dimensional las tablas de la base de datos pueden estar desnormalizadas a diferencia del modelo relacional en donde se deben aplicar las formas normales para normalizar los datos. Esto es posible debido a que en una base de datos dimensional encontraremos solamente un tipo de consulta SQL la cual es el SELECT, por lo que los problemas como la integridad referencial asociados a otras consultas como el UPDATE O DELETE no existen.

A pesar de lo anterior, la desnormalización genera redundancia de datos, lo cual hace crecer el tamaño de la base de datos. Si además consideramos que este tipo de base de datos tiene la tarea de almacenar datos históricos de la organización tenemos como resultado una gran cantidad de datos en comparación con una base de datos relacional. Como contraparte la desnormalización permite que a pesar del gran volumen de datos existente, las consultas se realicen más rápido debido a que se deberán recorrer menos

tablas, es decir que se tendrán que realizar menos joins en las consultas, aumentando el performance de lectura.

4.1 Componentes del modelo dimensional.

a) Dimensiones: como se ha mencionado anteriormente una dimensión es una perspectiva de negocio que tiene sentido para los usuarios, por lo cual su definición está dada de acuerdo a las necesidades de análisis de la empresa. Por ejemplo si se quieren medir las ventas de un negocio, para determinar las dimensiones se deben realizar preguntas de acuerdo a la perspectiva que queramos analizar del mismo.

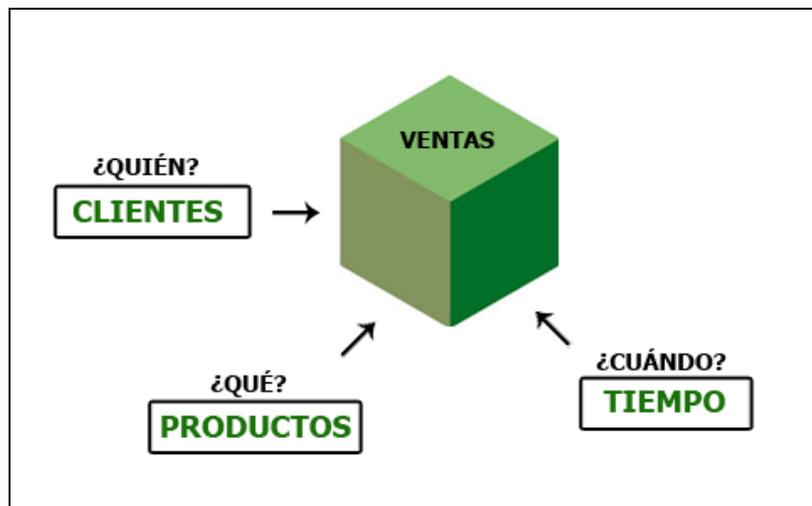


Figura 8. Dimensiones

En la figura 8 podemos ver que las necesidades de un negocio en particular son analizar los indicadores de ventas a través de las dimensiones cliente, producto y tiempo.

b) Atributos: los atributos permiten ver la información de cada dimensión a diferente nivel de detalle y agrupar los datos para ser analizados. Una dimensión está compuesta

por un conjunto de atributos relacionados entre ellos. Siguiendo el ejemplo anterior en la figura 9 se ilustra los posibles atributos que podrían tener las dimensiones cliente, producto y tiempo.

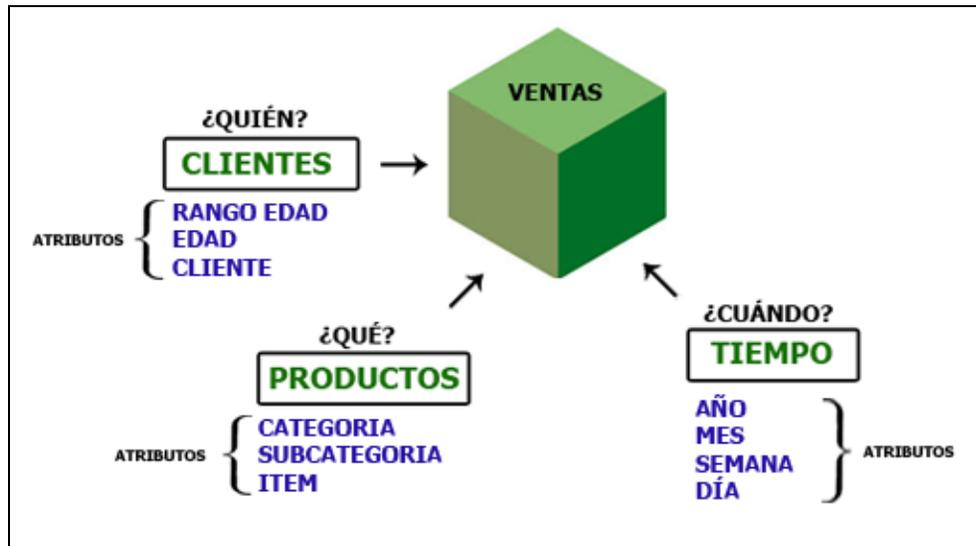


Figura 9. Dimensiones y atributos

c) Relaciones entre Atributos: al igual que en el modelo Entidad – Relación existen las relaciones 1: 1 (uno a uno), 1: M (uno a muchos) y N: M (muchos a muchos). En el modelo dimensional se utilizan para unir los diferentes atributos de una dimensión que corresponden a diferentes niveles, formando una estructura jerárquica. La relación entre atributos de diferentes dimensiones es representada por los indicadores definidos como intersección de las dimensiones.

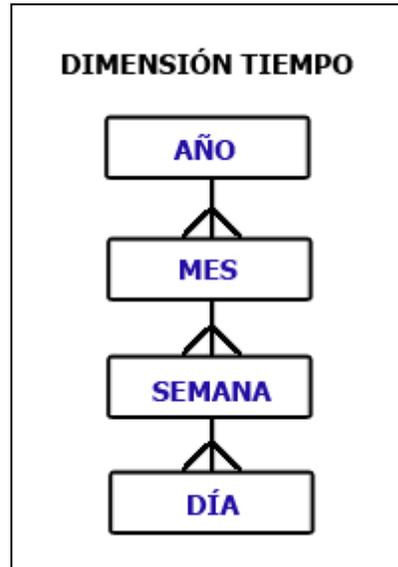


Figura 10. Jerarquía de la dimensión tiempo

d) Jerarquía: Como convención del modelado la jerarquía se dibuja verticalmente desde el atributo más agregado (arriba) hasta el más atómico (abajo). En una dimensión puede existir más de una jerarquía, siendo posible destacar la jerarquía principal de las jerarquías secundarias fácilmente.

Esta estructura de jerarquía permite la navegación en los cubos OLAP a través de las siguientes operaciones:

- Drill-Down, permite apreciar los datos en mayor detalle bajando por la jerarquía de una dimensión.
- Drill-Up, permite apreciar los datos en un menor nivel de detalle subiendo por la jerarquía de una dimensión.
- Drill- within, permite moverse entre atributos de una dimensión que se encuentran en jerarquías distintas.
- Drill-across, permite analizar información entre diferentes dimensiones.

e) **Hechos:** los hechos o también llamados indicadores representan la relación N: M que existe entre las dimensiones. En ellas se ubican todos los eventos que se necesitan medir para realizar un análisis.

Existen dos tipos de indicadores:

- Indicadores básicos, los cuales existen físicamente en el Data Warehouse, como por ejemplo las ventas.
- Indicadores derivados, los cuales deben ser calculados a partir de hechos básicos y pueden o no existir físicamente en el Data Warehouse, como por ejemplo el margen de ganancia, restando los hechos básicos de precio y el costo.

Cabe destacar que una dimensión puede ser común a varios hechos (dimensión global) o puede ser asociada a un hecho específico (dimensión local). El DW al tener la propiedad de “variante en el tiempo” convierte a la dimensión tiempo como la dimensión global más importante.

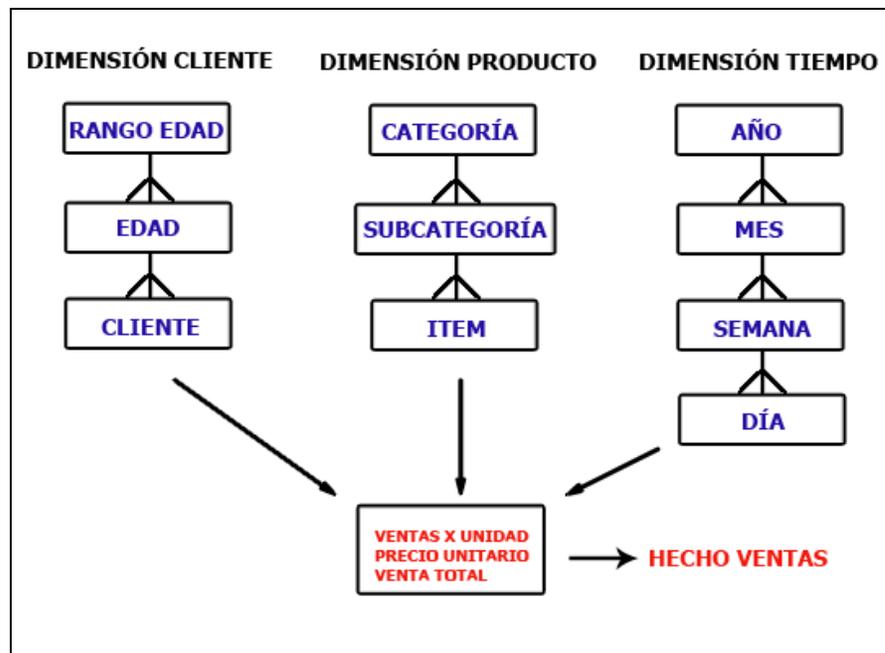


Figura 11. Ejemplo de modelo dimensional con todos los componentes.

4.2 Esquema

Como se había destacado anteriormente, un modelo dimensional es una visión lógica de un Data Warehouse que no representa su estructura física. Como contraparte un esquema si determina la representación física de un Data Warehouse el cual es mapeado desde el modelo dimensional creado. Por lo tanto la secuencia de actividades para el correcto diseño de las bases de datos dimensionales de un DW es la que se ilustra en la figura 12.

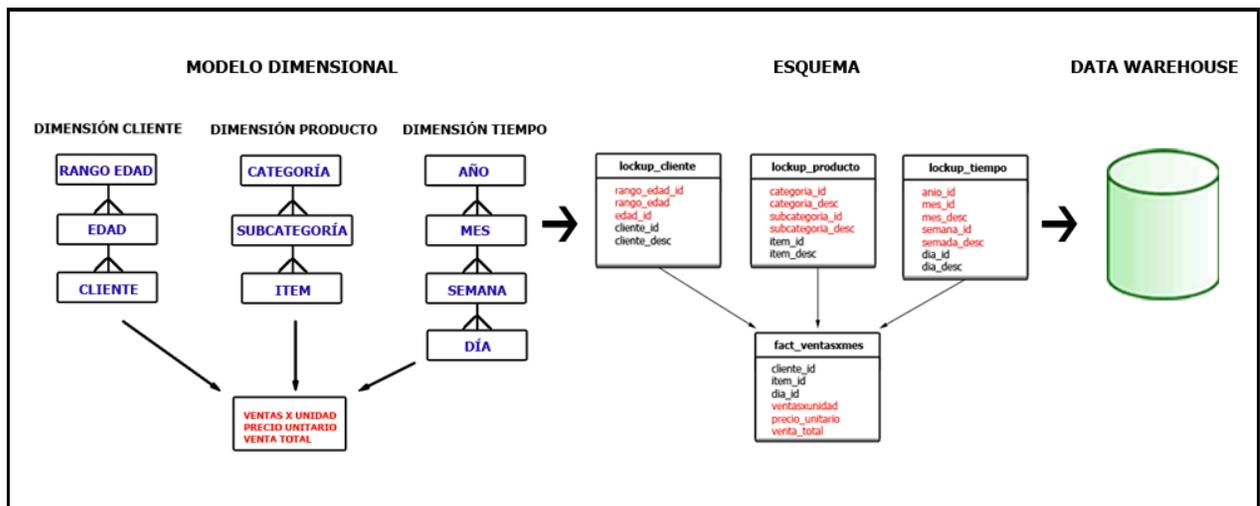


Figura 12. Diseño de una base de datos dimensional

La construcción de un esquema también involucra la definición de componentes propios, los cuales se presentan a continuación.

a) Tablas de Lookup: es la representación física de un atributo en el modelo dimensional. Para el mapeo se realizan los siguientes procedimientos:

- Crear una tabla Lookup por cada atributo que exista en una dimensión, por convención el nombre de la tabla debe empezar con “lookup” seguido del nombre del atributo. Se debe identificar obligatoriamente la clave principal del atributo y

su descripción es opcional, la cual puede ser evitada si la clave principal es usada con el mismo propósito de descripción.

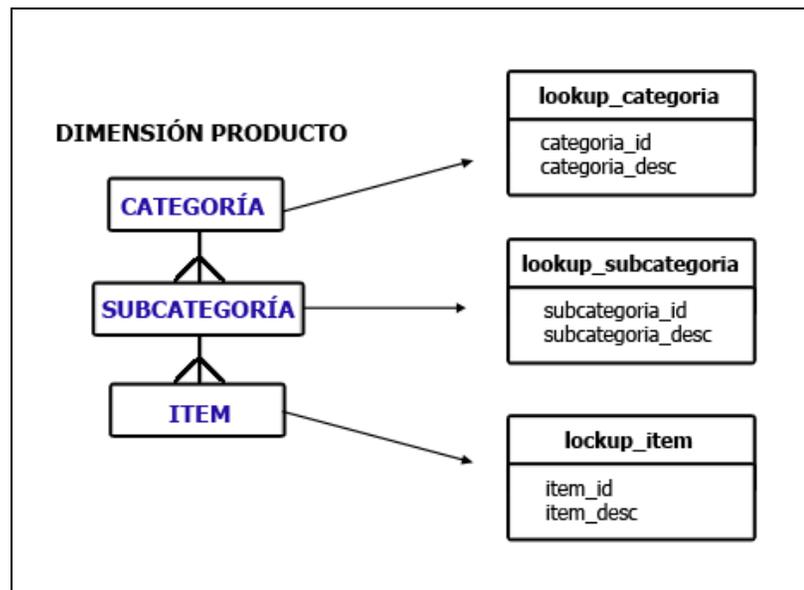


Figura 13. Construcción tablas de Lookup parte 1

- En cada tabla de Lookup se debe incluir una clave foránea que corresponda a la clave primaria de su padre dentro de la jerarquía del modelo dimensional, las cuales se utilizarán en las uniones o joins de las consultas. Para armar una estructura homogénea dentro de la base de datos es preferible utilizar los mismos nombres en las claves primarias y en las claves foráneas correspondientes.

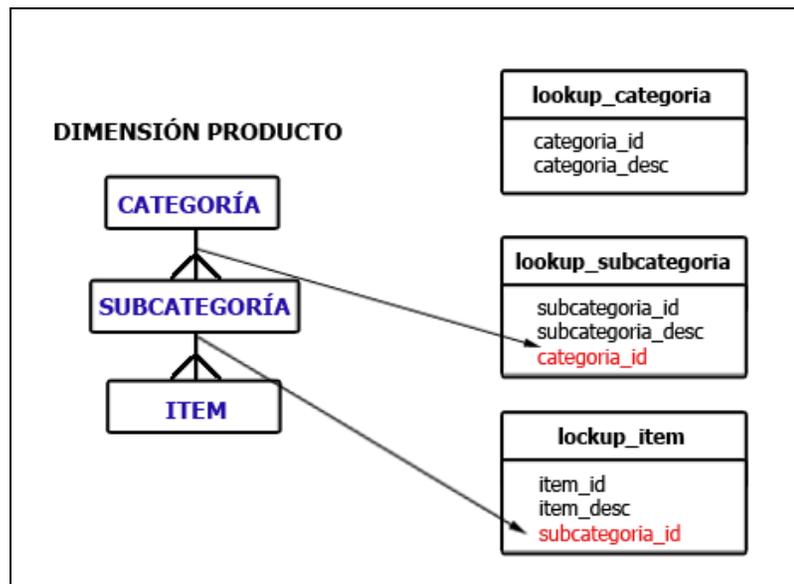


Figura 14. Construcción tablas de Lookup parte 2

b) Tabla Relacionales: es la representación física de una relación M: N entre dos atributos del modelo dimensional. En este caso además de crear las tablas de Lookup correspondientes, se crea la tabla relacional con un nombre que describa la relación de los atributos y debe incluir las claves primarias de las tablas de Lookup relacionadas.

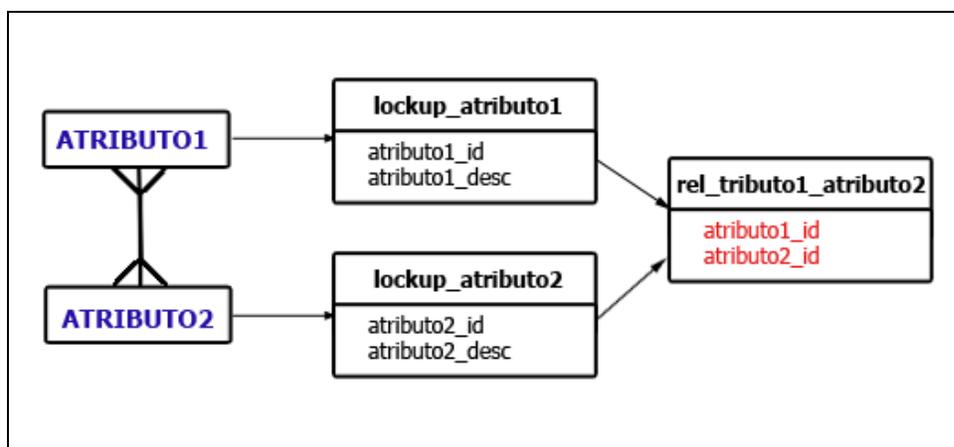


Figura 15. Ejemplo de tabla relacional

c) Tablas de Fact: es la representación física de un hecho en el modelo dimensional. Para el mapeo se siguen los siguientes procedimientos:

- Crear una tabla Fact por cada hecho que exista en el modelo dimensional, por convención la tabla debe empezar con el nombre “fact” seguido del nombre del hecho. Cada tabla Fact debe llevar una clave primaria de cada dimensión relacionada, la cual es escogida dependiendo del nivel de agregación que se necesite para esa tabla. Por ejemplo en la figura 16, el nivel de agregación tabla “fact_ventasxmes” en la dimensión tiempo es a nivel de la tabla “lookup_mes” debido a que la clave primaria relacionada es “mes_id”, por lo que el join de la consulta no pasara ni por la tabla “lookup_semana” ni “lookup_día”.

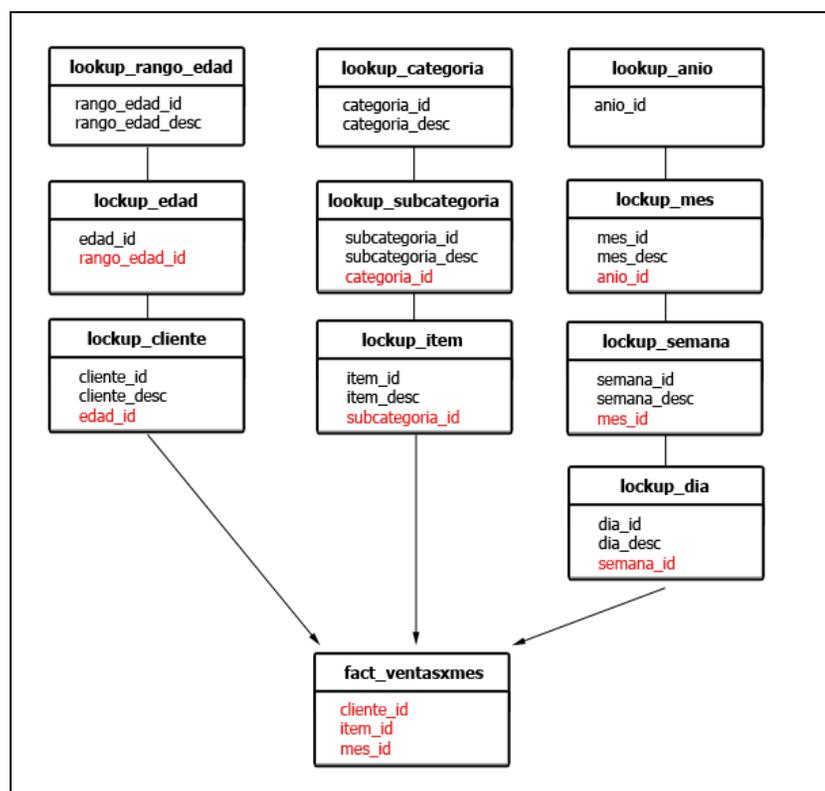


Figura 16. Construcción tabla Fact parte 1

- La tabla Fact debe considerar todos los indicadores básicos e indicadores derivados descritos en el modelo dimensional.

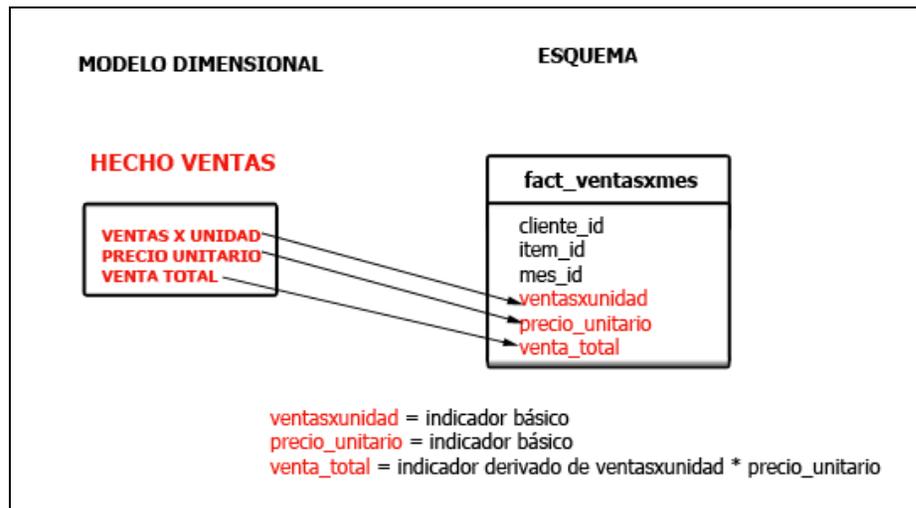


Figura 17. Construcción tabla Fact parte 2

4.3 Tipos de Esquema

a) Esquema copo de nieve

Seguir al pie de la letra procedimientos anteriores, entregará como resultado un esquema normalizado en 3 FN (Forma Normal), el cual es conocido como "Esquema copo de nieve". En resumidas cuentas una 3 FN se cumple cuando cada tabla tiene su respectiva clave primaria y el resto de elementos dependen directamente de ella.

En la figura 18 se ilustra un esquema copo de nieve correspondiente al modelo dimensional de la figura 11 antes presentada.

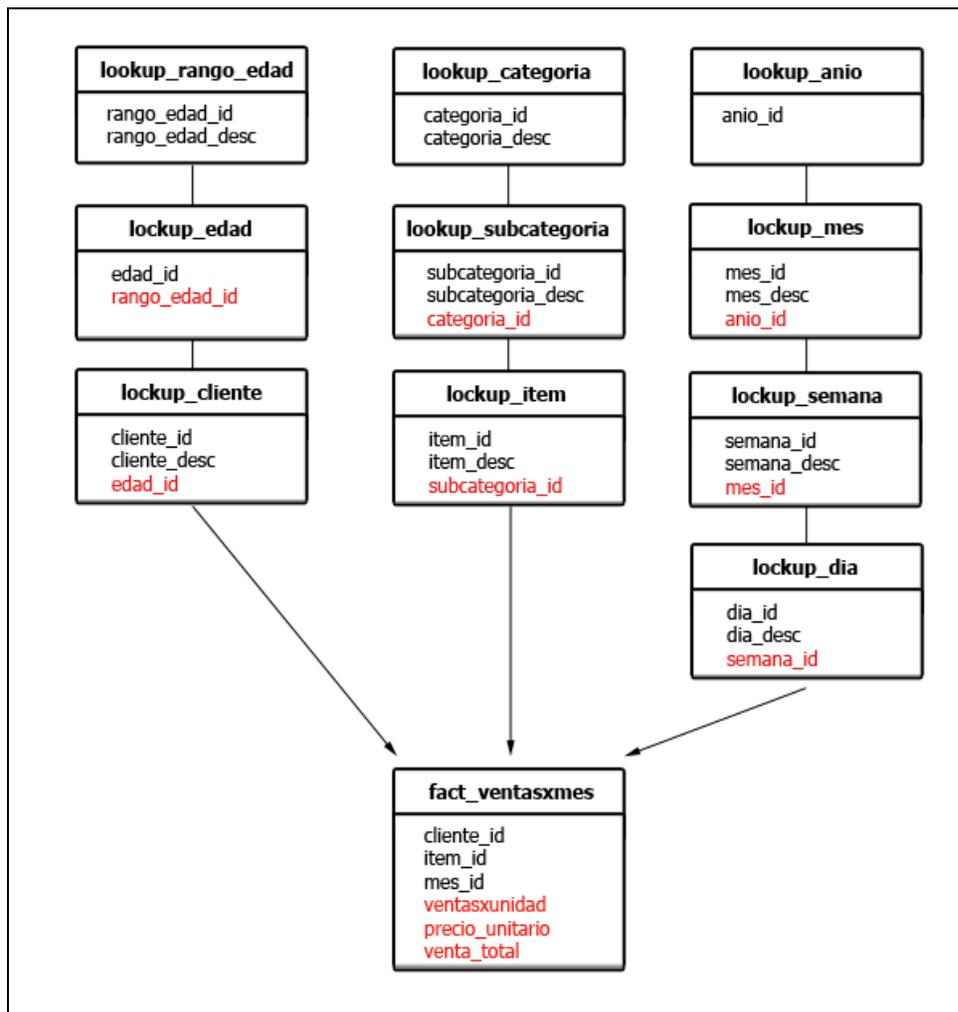


Figura 18. Esquema Copo de Nieve

b) Esquema Estrella

Por motivos que ya se conocen, un esquema puede pasar por un proceso de desnormalización, dependiendo de las necesidades y condiciones del Data Warehouse. Un DW puede llegar a almacenar una gran cantidad de información la cual puede afectar al tiempo de respuesta de las consultas. Esta situación puede mejorarse con la desnormalización.

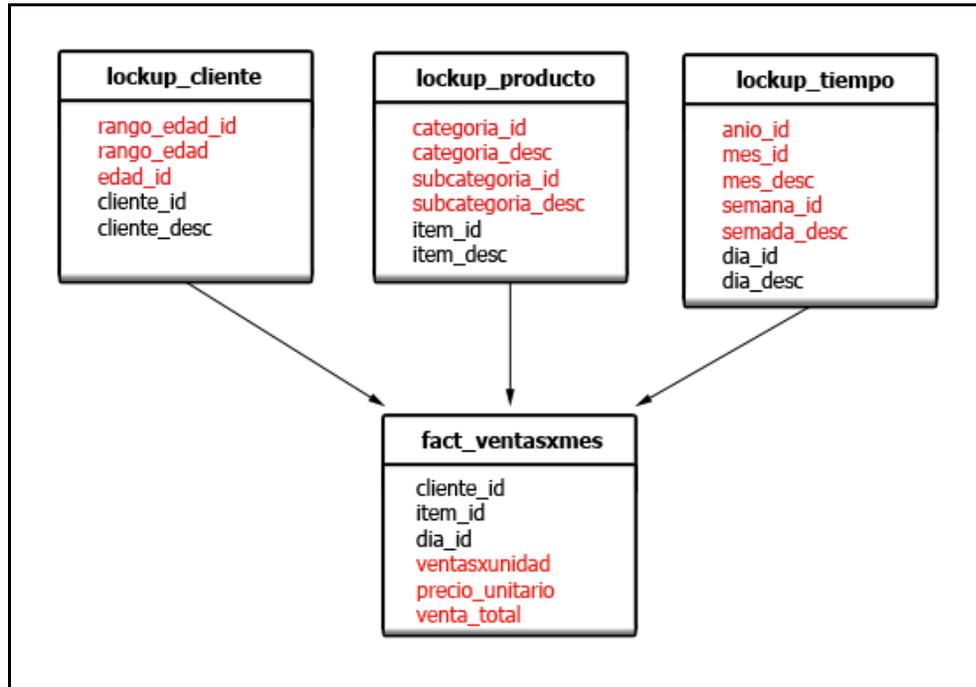


Figura 19. Esquema Estrella

Un esquema completamente desnormalizado es llamado “esquema estrella”. En ella existe solamente una tabla de Lookup por cada dimensión en donde se encuentran los elementos de todos los niveles de la jerarquía de atributos, por esta razón también es conocida como tabla de dimensión. Esta es la razón de por qué los tiempos en las consultas mejoran, mientras menos tablas se tengan que recorrer desde la tabla de Fact son menos joins que manejara la consulta. Un esquema estrella puede no estar completamente desnormalizada manteniendo algunos niveles de atributos, como se ha dicho anteriormente, su estructura depende completamente de las necesidades de análisis de los usuarios del Data Warehouse.

c) Esquema Constelación

Por su parte un esquema compuesto por varios esquemas estrella es llamado “esquema constelación” el cual tiene las siguientes características especiales:

- Tiene dos o más tablas de Fact. De todas ellas existe la tabla de Fact principal y el resto son tablas de Fact auxiliares.
- Una tabla de Fact auxiliar puede no estar relacionada con todas las dimensiones de la tabla de Fact principal.
- Una tabla de Fact auxiliar puede corresponder a sumalizaciones de la principal o corresponder a un proceso distinto de análisis.

Para propósitos de poder graficar un esquema constelación se ha agregado al ejemplo la tabla “fact_compra” y la tabla “lookup_geografía.

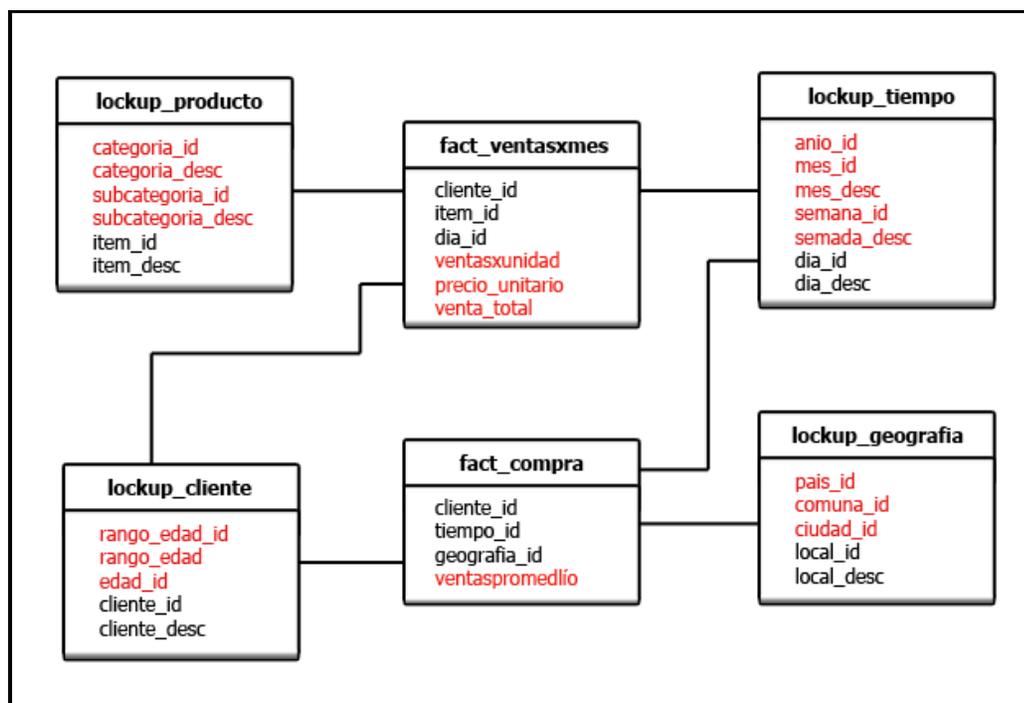


Figura 20. Esquema Constelación

5. Metodología Hefesto

Hefesto es una metodología formulada por el ingeniero argentino Ricardo Darío Bernabeu el cual entrega un marco de trabajo que guía la construcción e implementación de un Data Warehouse. El objetivo es entregar una primera implementación que satisfaga una parte de las necesidades de la organización, para mostrar las ventajas del DW y motivar a sus usuarios.

La metodología Hefesto puede resumirse en la figura 21.

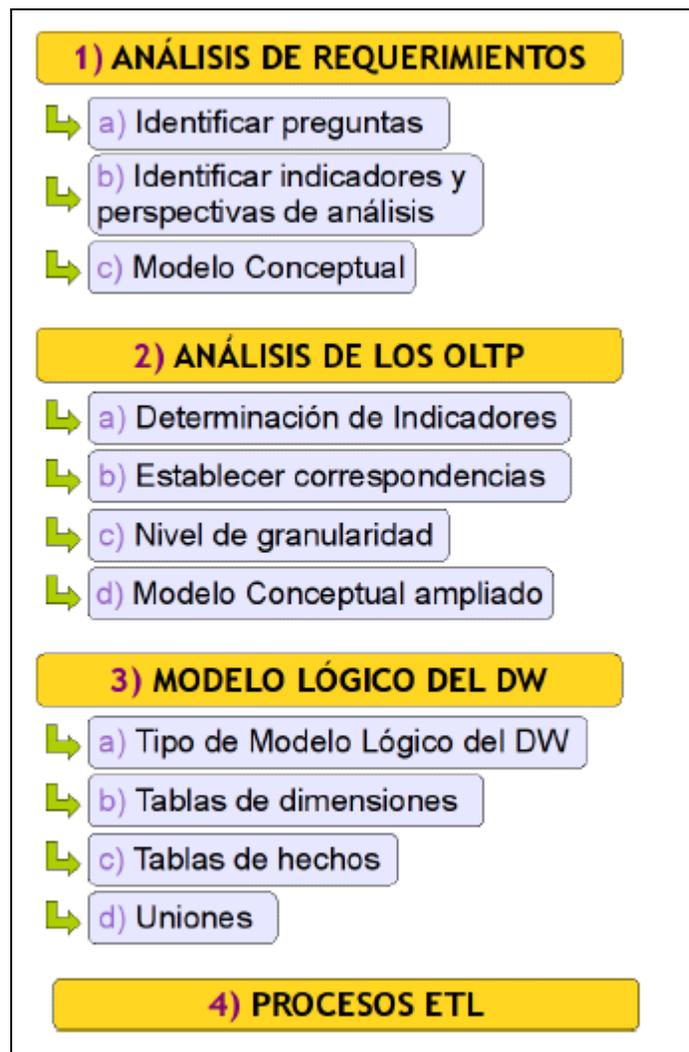


Figura 21. Metodología Hefesto

5.1 Análisis de Requerimientos

El primer paso de esta metodología es realizar una recolección de requerimientos de análisis directamente con los usuarios que ocuparan la solución de Data Warehouse, de tal manera de llegar a obtener un modelo conceptual de las dimensiones e indicadores de primera necesidad.

a) Identificar Preguntas: en esta etapa se recolectan las necesidades de análisis de información a través de técnicas y herramientas de recopilación de datos las cuales poseen diferentes características dependiendo de la estrategia a utilizar, como por ejemplo entrevistas, cuestionarios, observaciones, entre otros.

b) Identificar indicadores y perspectivas de análisis: luego de identificar las preguntas claves se debe proceder a su descomposición para descubrir los indicadores que se utilizaran y las perspectivas de análisis que intervendrán.

c) Modelo Conceptual: finalmente se construye un modelo conceptual a partir de los indicadores y perspectivas obtenidas anteriormente, de forma que se puedan observar claramente los alcances del proyecto para después trabajar sobre ellos.

5.2 Análisis de los OLTP

El segundo paso de esta metodología es analizar las fuentes transaccionales desde donde se obtendrá la información del Data Warehouse para poder determinar cómo se calcularán los indicadores, que la información requerida está disponible y poder establecer correspondencias entre el modelo conceptual creado anteriormente y lo que existe en los sistemas OLTP. Para ello se deben realizar cuatro etapas:

a) Determinación de Indicadores: en esta etapa se debe describir como se calcularan los indicadores.

b) Establecer Correspondencias: se debe 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.

c) Nivel de Granularidad: basándose en las correspondencias establecidas en el paso anterior se deben presentar al usuario los datos de análisis disponibles para cada perspectiva, de forma que se puedan seleccionar los campos que se consideren relevantes para el análisis. En este sentido, la metodología hace hincapié en la dimensión tiempo la cual debe estar presente y debido a sus propiedades se debe preguntar su nivel de granularidad, las cuales pueden ser días de la semana, quincena, mes, trimestres, semestres, año, etc.

d) Modelo conceptual ampliado: finalmente para graficar los resultados anteriores se ampliará el modelo conceptual, agregando bajo cada perspectiva los campos elegidos y bajo cada indicador su respectiva fórmula de cálculo.

5.3 Modelo Lógico del DW

La metodología Hefesto define una forma diferente de realizar el modelado dimensional. En base al modelo conceptual ampliado creado en el paso anterior se deben realizar tres etapas:

a) Tipo de Modelo Lógico del DW: se debe elegir cual será el tipo de esquema adecuado para soportar la estructura de base de datos prefiriendo la que se adapte mejor a las necesidades y requerimientos del usuario. Según lo visto anteriormente podemos elegir un esquema estrella, copo de nieve y constelación teniendo siempre presente que esta decisión puede afectar bastante al diseño del modelo lógico.

b) Tablas de dimensiones: se deberán diseñar las tablas de dimensiones que corresponden a cada perspectiva del modelo conceptual ampliado. Se deberá escoger entre sus campos una clave primaria y se elige un nombre que represente la tabla

dimensión. Se redefinen los nombres que no sean lo suficientemente intuitivos. Para los esquemas copos de nieve se debe definir la jerarquía normalizando la tabla.

c) Tablas de hechos: se definen las tablas de hechos que contendrán los hechos a partir de los cuales se calcularán los indicadores. Se deben seguir las siguientes reglas:

- Que la tabla tenga un nombre descriptivo que represente el área analizada.
- La clave primaria está compuesta por las id de cada dimensión relacionada.
- Se incluirán hechos que representen todos los indicadores asociados, los cuales pueden cambiar de nombre de tal manera que sean descriptivos.

d) Uniones: se realizan las uniones siguiendo el modelo conceptual ampliado, relacionando cada tabla de hecho con sus respectivas dimensiones.

5.4 Procesos ETL

Es la actividad relacionada con la extracción de datos de diferentes fuentes, para luego integrarlos, filtrarlos y depurarlos. Como existen varios software que facilitan esta tarea, Hefesto se centra solo en la generación de las sentencias SQL que contendrán los datos que sean de interés.

III. Desarrollo del trabajo

Las empresas distribuidoras de la región de los Lagos tienen un nivel tecnológico en donde sus procesos de operación son apoyados por sistemas transaccionales. A pesar de que los datos capturados por estos sistemas puedan entregar una idea de cómo realizar acciones estratégicas de marketing estos no son completamente certeros ya que parte de este análisis es deducido en base a tendencias o capturados por su misma fuerza de ventas en terreno.

Esto sucede porque no poseen los datos de sus clientes directos, los negocios de barrio a quienes abastecen productos. La falta de esta dimensión es porque los datos generados por sus procesos no son capturados por sistemas OLTP o por que no se tienen acceso a ellos fácil y completamente.

Para este estudio en particular se dará por hecho que los datos de los negocios de barrio pasaron por un proceso ETL y están listos para su uso, ya que el objetivo principal es identificar si este sector tiene una necesidad de Inteligencia de negocios.

Para identificar las necesidades de análisis de estas empresas distribuidoras se realizará un proceso de Data Warehousing que en primera instancia entregue un modelo de base de datos para un Data Warehouse.

Una vez teniendo la certeza que forma tiene este Data Warehouse se deducirá un nuevo modelo de negocios que beneficie al sector que se encuentre dentro de la solución de Inteligencia de Negocios entregada.

En primer lugar, para obtener las principales variables de análisis y construir un diseño físico del Data Warehouse se utilizará la metodología Hefesto, la cual tiene que ser modificada por las características explicadas anteriormente.

1. Adecuación de la metodología Hefesto

El ingeniero Ricardo Bernabeu proporciona una metodología que permite tener una visualización rápida de un modelo de Data Warehouse pero incluye las etapas del Data Warehousing asociadas al área de extracción de datos, tema que no es de interés para los propósitos de esta tesis. En la figura 22 se ilustra esta situación.

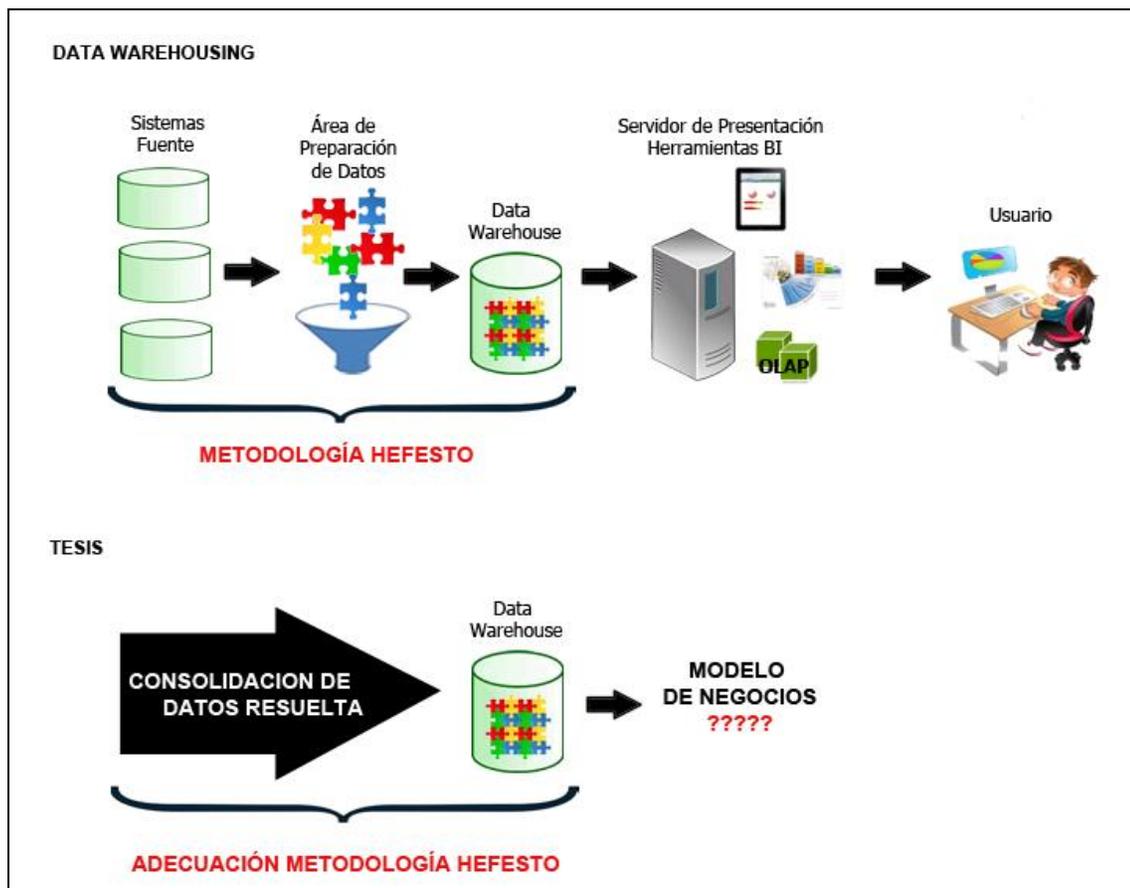


Figura 22. Adecuación de Metodología Hefesto

Bajo este punto de vista las actividades de la metodología que se verán afectadas son las siguientes:

a) Análisis de los OLTP

Esta etapa sufre modificaciones ya que no se tienen acceso a las bases de datos de los sistemas transaccionales para realizar las correspondencias a lo capturado en la etapa de análisis de requerimientos. Todo lo contrario, lo que se desea es que esta consolidación se realice en base a lo que requiere el Data Warehouse por lo que en esta etapa se realizará lo siguiente:

- Determinar si los procesos de los negocios de barrio generan los datos que el modelo conceptual del Data Warehouse necesita.
- Generar una lista de los datos que necesitan ser consolidados para el funcionamiento del futuro Data Warehouse y que deben contener los sistemas transaccionales que se unan al modelo.
- Generar un modelo lógico el cual pueda ser presentado a los clientes y que sirva como guía para la construcción de cubos.

El nombre de esta etapa cambiará a “Requerimientos para los OLTP”.

b) Procesos ETL.

Debido a que el proceso de homogenización de datos ya se encuentra realizado, esta actividad se elimina por completo de la metodología.

Por lo tanto, la adecuación de la metodología para este proyecto contempla las siguientes actividades:

- Análisis de requerimientos.
- Requerimientos para los OLTP
- Modelo Lógico del Data Warehouse

2. Análisis de requerimientos

2.1 Herramienta de recolección de información

Para realizar la tarea investigativa de esta etapa la herramienta de recolección de datos que se utiliza es la entrevista, pauta que puede ser revisada con mayor detalle en el anexo de este documento.

Debido a que no se sabe el nivel de conocimiento sobre el tema de Inteligencia de Negocios de los entrevistados se entregó una pequeña explicación del concepto y se ejemplifico como podría ser su uso una vez implementada.

Las preguntas realizadas son abiertas debido a que se necesita recabar detalladamente las perspectivas de análisis de los entrevistados las cuales no se conocen, determinando en el proceso las similitudes de necesidad.

Las preguntas apuntan a conocer el nivel de inteligencia en la empresa investigada y sus necesidades de Inteligencia de Negocio las cuales fueron conducidas de tal manera que se pudieran obtener indicadores y dimensiones de análisis deseadas en una supuesta consolidación de datos de los sistemas transaccionales de sus clientes. A continuación se enlistan las preguntas formuladas:

- ¿Conoce la Inteligencia de Negocios? ¿Cómo es aplicada en su empresa?
- ¿Si tuviera a su disposición los datos de venta, compra y stock de sus clientes, que información le gustaría obtener? ¿Cómo obtiene esta información actualmente?
- ¿Usted apoyaría un sistema de fidelización de clientes en los negocios de barrio a los cuales usted abastece para así verse beneficiado con los datos que se generarán? ¿De qué forma?
- ¿Usted cree que esta implementación le otorgará beneficios? ¿Cuáles?

2.2 Población de estudio

La población de estudio en estadística es un segmento el cual se tiene acceso para realizar una investigación. En este caso la población objetivo son las empresas de distribución abarrotera en la región de los Lagos convirtiéndose en la población de estudio tres de las empresas más representativas del sector, de tal manera de poder recabar información que pueda ser aplicada para realizar una implementación de Data Warehouse transversal en un corto plazo.

Adicionalmente, dada las características de este proyecto, las personas escogidas dentro de estas empresas para realizar esta investigación son la que están a cargo de tomar las decisiones estratégicas por lo que a su vez tienen un conocimiento amplio sobre el negocio y las necesidades de análisis de este.

Ya que se prefiere mantener anónimos los nombres de las empresas y de las personas entrevistadas se catalogarán de la siguiente manera:

Empresa	Persona Entrevistada
Empresa distribuidora 1	Gerente Zonal
Empresa distribuidora 2	Gerente Zonal
Empresa distribuidora 3	Dueño de la empresa

Tabla 1. Empresas y personas entrevistadas

2.3 Nivel de inteligencia actual.

Todas las empresas encuestadas tienen claro que se encuentran en una posición muy complicada en el mercado, en donde las empresas transnacionales las están avasallando, haciéndose cada vez más imprescindibles. La diferencia competitiva se

encuentra en el conocimiento del cliente, lo cual está bastante asumido: mientras más información tengan de su mercado objetivo mejores resultados van a obtener.

Este conocimiento actualmente se logra teniendo una certeza completa de lo que ellos compran y lo que venden a sus clientes lo cual es manejado dentro de sus datos transaccionales. Estos son volcados en planillas Excel para su análisis, pero desconocen completamente la dimensión de datos que generan sus clientes, por lo que esta parte la deducen a través de supuestos a base de algunas tendencias de mercado, teniendo márgenes de errores de por lo menos un 30%, lo cual hace que la toma de decisiones respecto a acciones de marketing no sean las ideales.

El nexo más directo con sus clientes es equipo de ventas los cuales informan diariamente lo que puedan percibir de su mercado objetivo a través de sus visitas a terreno. Existen varias técnicas para intentar conocer un poco más sobre sus clientes, de las cuales se pueden destacar:

- Concertación de acuerdos comerciales a cambio de información.
- Preguntas directas de determinada información.
- Examinación visual, por ejemplo el poder estar tres minutos en la bodega de un cliente es suficiente para saber más o menos que productos está comprando, el tamaño de la bodega, entre otras cosas.

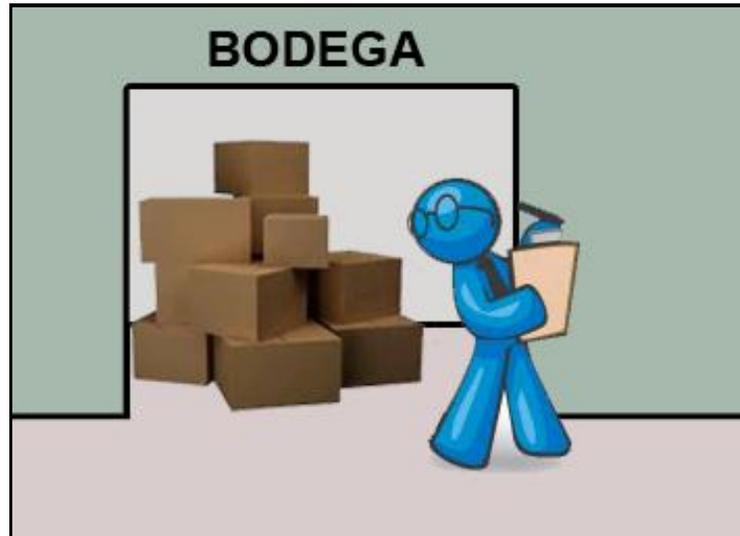


Figura 23. Examinación visual de un vendedor en terreno.

Lamentablemente en todos estos casos la veracidad de la información no es comprobable, sobre todo porque sus clientes son muy herméticos para entregar información y en caso de que pueda obtenerse, muchas veces la información no está completa, puede verse tergiversada o condicionada. Aun así la exploración visual puede ser más certera pero la posibilidad de poder realizarla es baja.

2.4 Identificación de indicadores y dimensiones de análisis

A pesar de lo anterior las personas encuestadas de la empresa 1 y 2 tienen algunos conocimientos de lo que significa la Inteligencia de Negocios, en contraste con la empresa 3 que solo la había escuchado nombrar.

Indistintamente todas las empresas tienen en mente proyectos que se relacionan con implementaciones informáticas que permitan tener un nexo más directo con sus clientes para saber en línea cuáles son sus necesidades y poder responder inmediatamente.

Esto es un buen comienzo, ya que indica que una implementación de Inteligencia de Negocios es posible.

A continuación se presentan las variables de análisis más importantes que han sido detectadas en esta investigación.

a) Control de stock del cliente.

Una de las razones por la que los distribuidores desean crear vínculos vía informática con sus clientes es para poder controlar su stock. Este requerimiento tiene tres necesidades claras:

- En primer lugar se necesita apoyar a sus clientes de tal manera de que no tengan ni un sobre stock ni un quiebre de stock, ya que ambas significan pérdida financiera, en un caso se tiene “dinero empozado” y en la otra no se tienen recursos que generen dinero. Ambas situaciones afectan a las ventas de los distribuidores.
- A pesar de que las empresas encuestadas tengan sistemas para controlar su stock y realizar pequeñas tendencias de ventas para controlar sus propios quiebres no se encuentran preparados ante un pedido inesperado. En todos los casos estas empresas también tienen sus distribuidores los cuales tienen sus centros de almacenaje muy alejados de la zona, es por eso que en base a tendencias se realizan los pedidos para mantener en sus propias bodegas un stock que les permitan reaccionar ante la demanda de sus clientes. Por ejemplo se tienen 100 cajas de productos y sus sistemas dicen que se venden 50 de estas cajas por semana, por lo que tendrían stock para 15 días, pero sucede una contingencia en donde un cliente compra estas 100 cajas en un solo pedido, en este caso quedarían con stock 0 de ese producto hasta que puedan volver a abastecerse.

- Finalmente el conocer el nivel de stock del cliente puede potenciar las acciones de marketing que se realizan a terreno por sus vendedores. Un vendedor podría ir directamente donde un cliente a ofrecer lo que le falte.

d) *Share of Wallet.*

La información relacionada con el potencial de venta que pueden tener estas empresas por cada uno de sus clientes parece ser de común interés. Al realizar las entrevistas todas las empresas coincidieron en aumentar su Share of Wallet, que en español equivale a la “participación en la billetera del consumidor”.

El Share of Wallet define la cantidad proporcional de dinero que un cliente invierte en los productos de estas empresas del 100% que este tiene destinado comprar. Por ejemplo si un negocio de barrio compra \$40.000 de fideos a una de estas empresas distribuidoras pero además compra \$60.000 de lo mismo a otras distribuidoras o mercados mayoristas el Share of Wallet del negocio para los fideos sería \$100.000, “billetera” que la empresa distribuidora siempre desea alcanzar.

El problema es que las empresas distribuidoras solamente manejan las ventas que ellos realizan a sus clientes, por lo que la torta completa no es conocida y como se dijo anteriormente, se deduce en base a tendencias. En la figura 24 se ilustra este ejemplo.

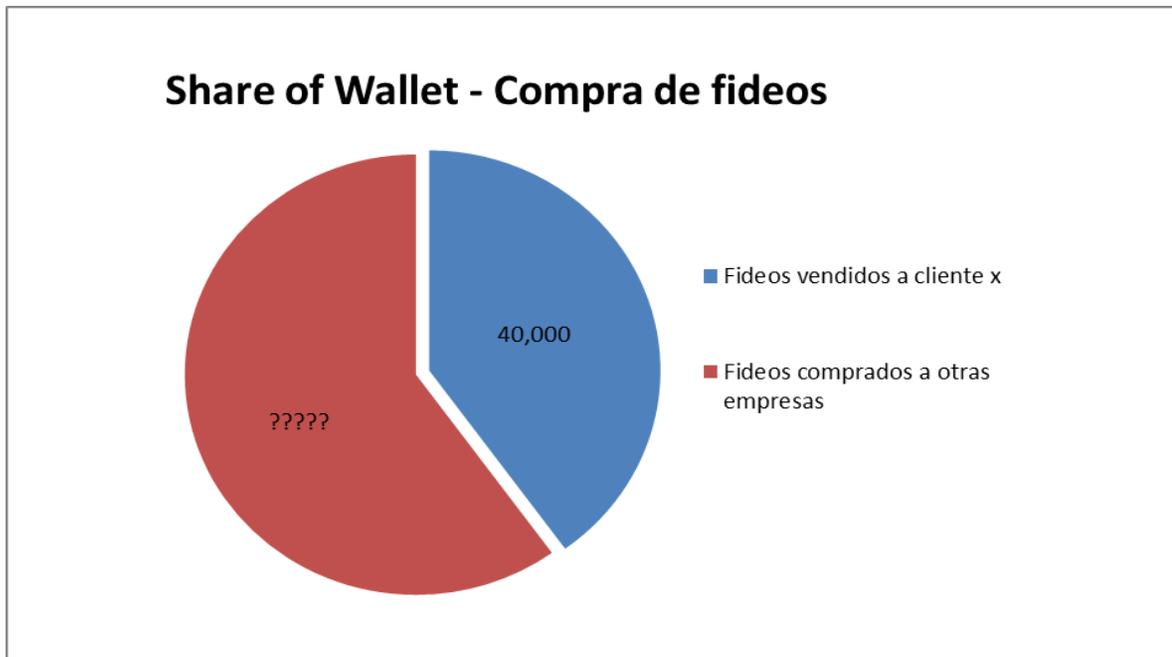


Figura 24. Share of Wallet – Compra de fideos

e) La fidelización del cliente final.

Una de las dimensiones más interesantes que se pueden explorar en una solución de Inteligencia de Negocios es la del cliente, pero en este caso es un poco diferente. Las necesidades de análisis de estas empresas son los datos de sus clientes inmediatos es decir los negocios de barrio o mercados particulares quienes compran sus productos directamente. Pero al final de la línea de este mercado encontramos al cliente final, quien es el que finalmente decide que comprar y es el que detona el movimiento de mercaderías entre las empresas.

La meta de análisis teniendo disponibles los datos del cliente final es la del Market Basket o “Canasta de Mercado” por su traducción al español, la cual está pensada en encontrar tendencias de compra de los clientes a los cuales se pueden asociar ciertas

promociones o descuentos. Por ejemplo se puede dar que un determinado segmento de cliente final vaya a su boliche más cercano a comprar siempre arroz y jurel todos los fines de mes. Con esta información el dueño del negocio podría realizar descuentos respecto a estos productos o venderlos en pack, ordenar sus estantes de productos de forma diferente y además se encontraría preparado para soportar este tipo de ventas anticipando compras para tener en el stock de su bodega.

A pesar de ello el conocimiento de este cliente final no tiene mucha relevancia para las empresas pero existe disposición para apoyar este tipo de iniciativas ya que cualquier situación que mejore el negocio de sus clientes será buena para ellos. El beneficio de uno directamente es el beneficio de otro creando un efecto domino, pero más aún, si se cuenta con este tipo de información de interés para el negocio de barrio se pueden realizar intercambios por otros tipos de información, como por ejemplo los potenciales de compra desconocidos.



Figura 25. Situación que puede ser evitada con Market Basket

f) Identificación de indicadores y dimensiones.

Utilizando la metodología Hefesto, se realiza la descomposición de las necesidades de análisis en indicadores y dimensiones para adecuarlos al modelo lógico del Data Warehouse.

- Se requiere conocer el potencial de ventas o el Share of Wallet del mercado objetivo de estas empresas. Para ello se necesita conocer la cantidad de productos comprados por un cliente como requisito fundamental y el valor de estos si se pudiera llegar a obtener en un tiempo determinado, distinguido por la categoría del producto, su marca y su descripción.
- Se requiere saber la cantidad de productos vendidos realizados por un cliente sobre productos de marcas propias en un tiempo determinado.
- Se necesita controlar el stock que los clientes tienen en sus bodegas semanalmente identificando categoría y marca de los productos.
- Se necesita saber el fraccionado de compras de los clientes, es decir determinar por cuanta cantidad compran, a quien les compran estos productos y cuál es el precio promedio mensual de estas compras, ordenados por categoría y marca.
- Se requiere conocer la cantidad de puntos de venta por sectores y cuál es su promedio de venta mensual ordenados por el mix de productos que manejan, es decir su categoría.
- Se desea conocer aspectos básicos sobre el cliente final, como su segmentación de edad y género para determinar que productos son los que compran en un determinado periodo de tiempo. Estos productos deben poder apreciarse por categoría y marca

A continuación se presenta una tabla con las descomposiciones de indicadores y dimensiones deducidos de los requerimientos de las empresas encuestadas.

Tipo de Análisis	Dimensiones	Indicadores
Share of Wallet	Productos, cliente, tiempo.	Cantidad de productos comprados, venta total en pesos.
Venta de Marcas Propias	Productos, cliente, tiempo.	Cantidad de productos vendidos
Control de Stock del cliente	Productos, cliente, tiempo.	Cantidad de productos en bodega.
Fraccionado de compras	Productos, cliente, proveedor, tiempo	Compra mensual, cantidad de productos promedio comprados.
Venta por sectores	Productos, cliente, geografía, tiempo	Cantidad de clientes, promedio de venta mensual
Market Basket	Productos, cliente, cliente final, tiempo	Cantidad de productos vendidos

Tabla 2. Dimensiones e indicadores descubiertos

3. Modelo conceptual

La metodología Hefesto define que al final de la etapa de análisis de requerimientos se construyan modelos conceptuales de acuerdo a las dimensiones e indicadores obtenidos para facilitar su visualización y poder trabajar con ellos posteriormente. En la figura 26 se ilustra cómo deben realizarse los modelos conceptuales.

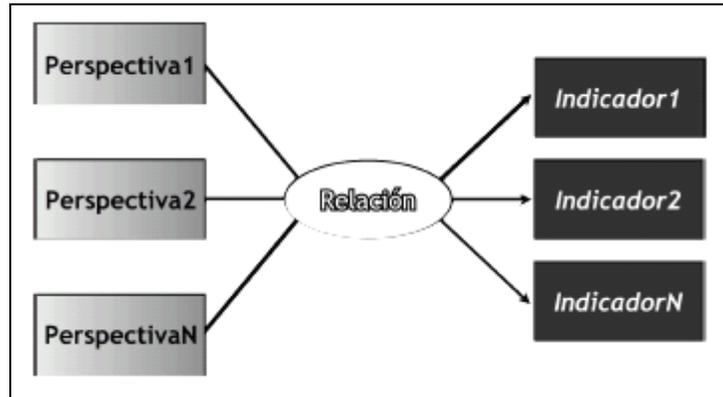


Figura 26. Modelo conceptual de la metodología Hefesto.

En la izquierda se colocan las perspectivas seleccionadas que serán unidas por un ovalo central que representa y lleva el nombre de la relación existente entre ellas, la cual es el proceso de negocio elegido para analizar. Finalmente desde esta relación se desprenden con flechas los diferentes indicadores ubicados a la derecha del modelo.

A continuación se presentan los modelos conceptuales realizados en base a la información recolectada en las encuestas y que se pueden apreciar resumidos en la tabla 2 de este capítulo. De acuerdo a la metodología Hefesto los modelos se han dividido de acuerdo al proceso de negocio que se analizará, de los cuales se pudieron identificar tres:

- Proceso de ventas.
- Proceso de compras.
- Proceso de inventario.

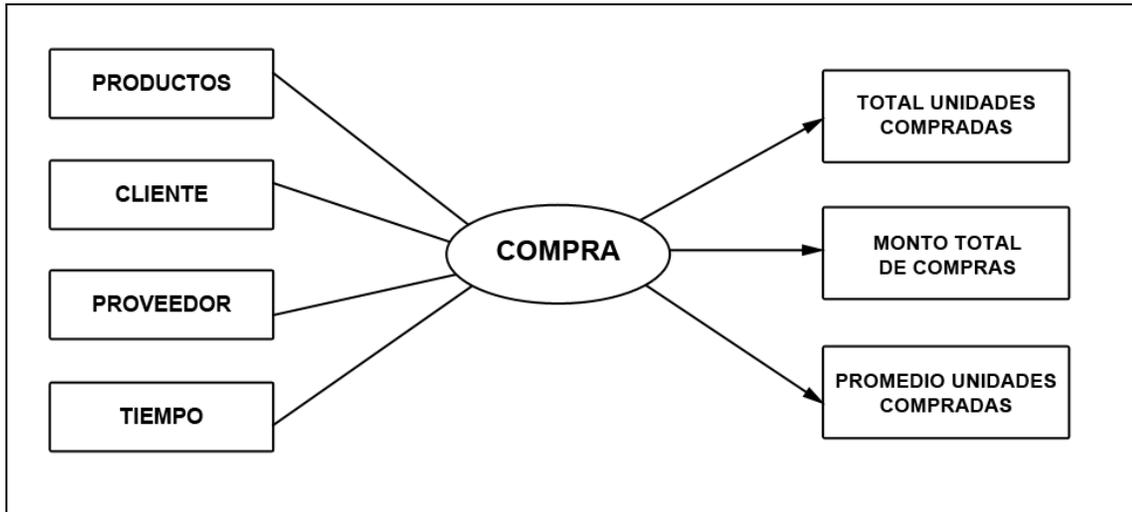


Figura 27. Modelo conceptual del proceso de compra.

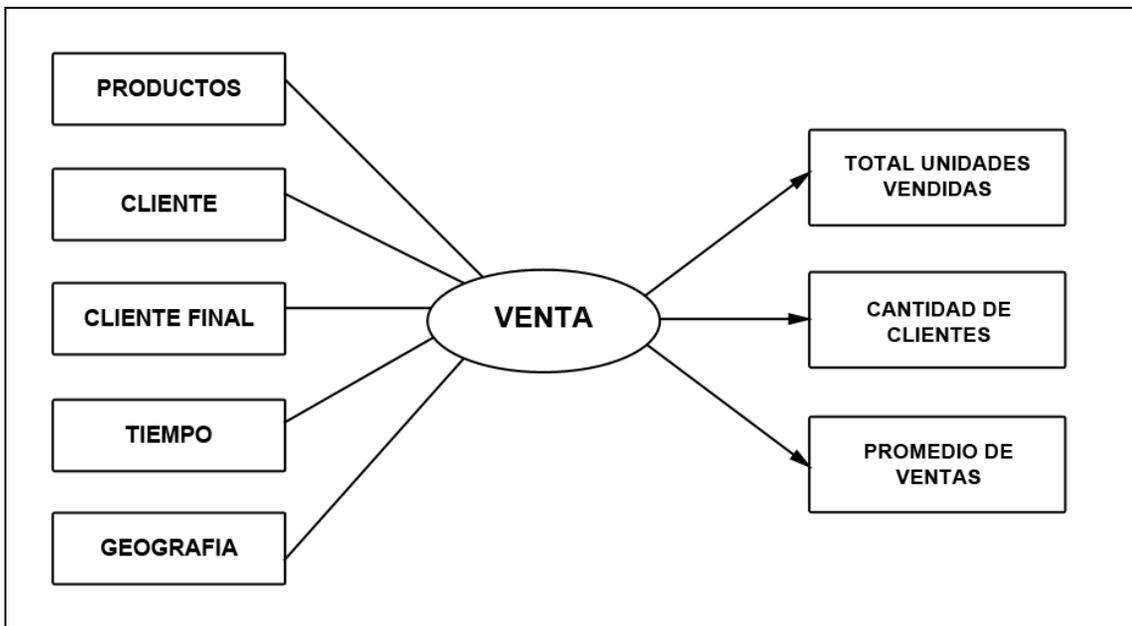


Figura 28. Modelo conceptual del proceso de venta

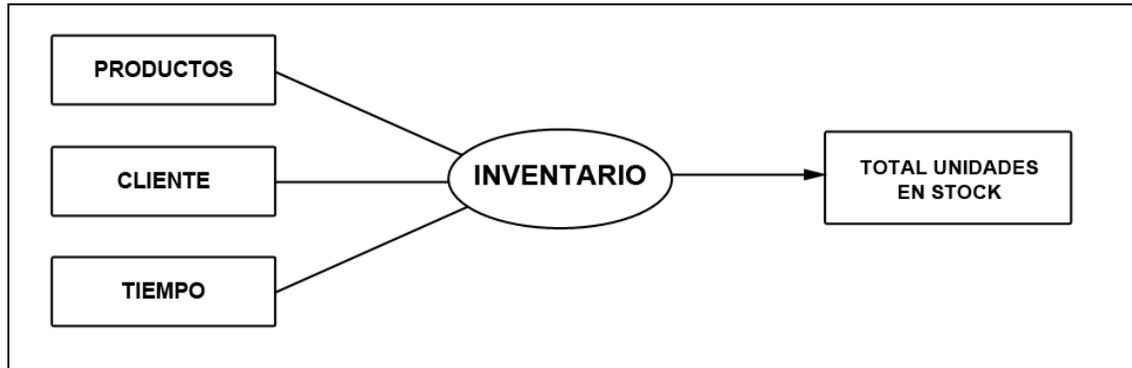


Figura 29. Modelo conceptual del proceso de inventario.

4. Requerimientos para los OLTP

4.1 Existencia de procesos

Como se ha dicho anteriormente, esta tesis no busca solventar la problemática del origen de los datos pero es necesario saber por lo menos si existen los procesos suficientes de donde se puedan obtener, ya que sin ellos una solución de Inteligencia de Negocios no se podría implementar.

Dada las características del retail al por menor, se deduce que todo el mercado objetivo de las empresas encuestadas por lo menos deben tener los procesos de compra y venta.

El problema radica en el proceso de inventario el cual no está presente en todos los negocios de barrio. Para comprobarlo se ha realizado una pequeña investigación en 50 mercados particulares de un sector de Puerto Montt, llegando a la conclusión que el 87% de ellos si tienen algún tipo de proceso de inventario manual. Los negocios que no tenían proceso de inventario eran muy pequeños y el método para determinar algún quiebre de stock es calculándolo visualmente.

Esta investigación determinó que el proceso de inventario existe en los mercados particulares más representativos del sector.

Teniendo la certeza de que los procesos están, se puede dar por hecho que la dinámica del negocio genera los datos suficientes para su posterior extracción.

4.2 Requerimiento de datos.

Como requerimiento los procesos deben estar soportados en un sistema OLTP, de esta manera su extracción será más sencilla. Los atributos de cada dimensión y los datos necesarios para cada indicador se consideran una pauta a seguir que indica los datos que debe capturar un sistema transaccional para adecuarse al modelo de Data Warehousing de esta tesis.

A continuación se listan los atributos de cada dimensión solicitados por las empresas distribuidoras encuestadas.

Dimensión	Atributo	Datos de Ejemplo
Producto	Categoría	Abarrotes, confites, helados
	Subcategoría	Fideos, legumbres, cereales
	Marca	Carozzi, Bonanza, San José
	Detalle	Arroz 1kilo, Aceite ½ litro, Ají 200 gramos
Cliente Final	Género	Masculino, Femenino
	Rango Edad	0 a 10, 11 a 20, 21 a 30, 31 a 40
	Cliente Final	Código de cliente interno

Tiempo	Año	2014, 2013, 2012
	Mes	2014/05, 2014/06, 2014/07
	Semana	2014/10/13 al 2014/10/19
	Día	2014/09/01, 2014/09/02, 2014/09/03
Proveedor	Proveedor	Razón social del proveedor
Geografía	Comuna	Frutillar, Llanquihue, FRESIA
	Ciudad	Puerto Montt, Puerto Varas, FRESIA
	Sector	Mirasol, La Paloma, Manuel Montt
	Cliente	Razón social del cliente

Tabla 3. Atributos correspondientes a cada dimensión.

Como se puede apreciar en la tabla 3, la dimensión “Cliente” ha desaparecido. Esto se debe a que se ha decidido que funcione como un atributo dentro de la dimensión “Geografía” como el atributo con menos granularidad.

Los datos necesarios y las funciones de cálculo para obtener los indicadores requeridos por las empresas encuestadas son los siguientes.

Indicador	Datos	Cálculo
Total unidades compradas	Cantidad de productos comprados	SUM(X)
Monto total de compras	Cantidad de productos comprados	SUM(X*Y)
	Precio producto comprado	
Promedio unidades compradas	Cantidad de productos comprados	AVG(X)
Total unidades vendidas	Cantidad de productos vendidos	SUM(X)
Promedio de Ventas	Cantidad de productos vendidos	AVG(X*Y)
	Precio producto vendido	
Cantidad de Clientes	Cliente	COUNT DISTINCT(X)
Total Unidades en Stock	Productos inventariados	SUM(X)

Tabla 4. Datos y funciones de cálculo para indicadores

4.3 Modelo conceptual ampliado y modelo dimensional

Se grafican los resultados obtenidos anteriormente para ampliar el modelo conceptual. En la figura 30 se ilustra este procedimiento.

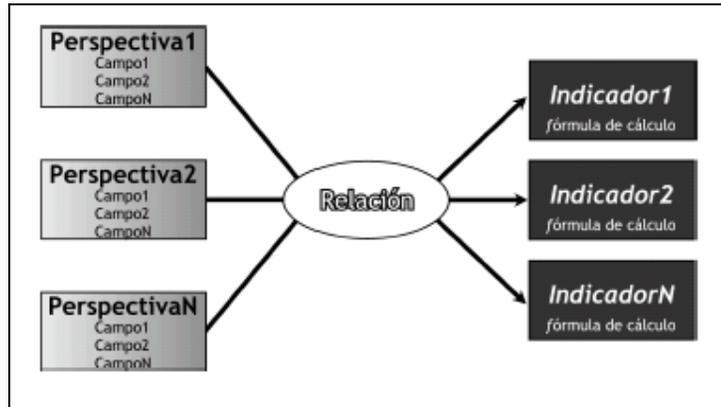


Figura 30. Modelo conceptual ampliado de la metodología Hefesto

Se debe colocar cada atributo debajo de su correspondiente perspectiva y bajo cada indicador su respectiva fórmula de cálculo.

A continuación se presentan los mapas conceptuales ampliados obtenidos en base a las tablas 3 y 4 de este capítulo.

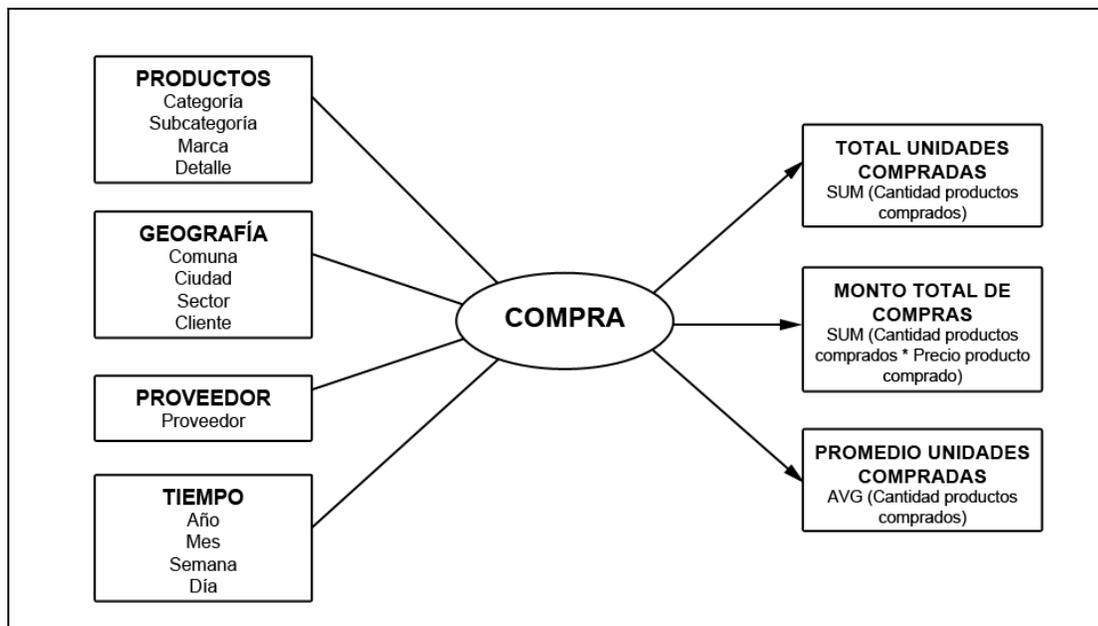


Figura 31. Modelo conceptual ampliado del proceso de compra

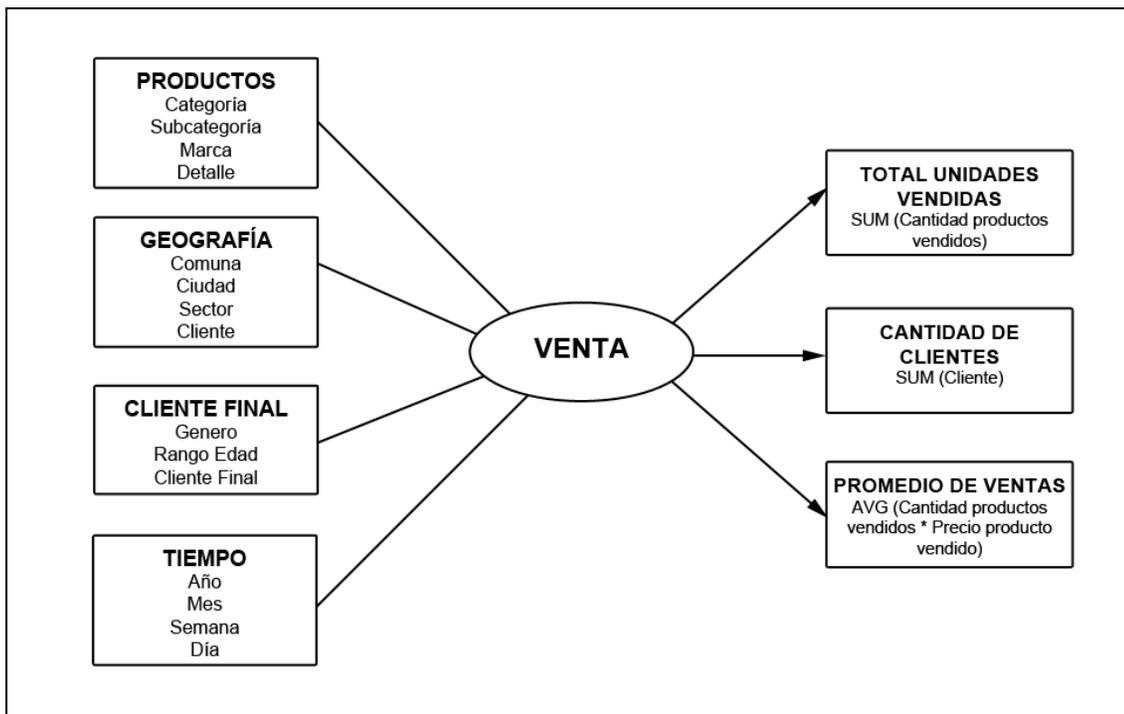


Figura 32. Modelo conceptual ampliado del proceso de venta

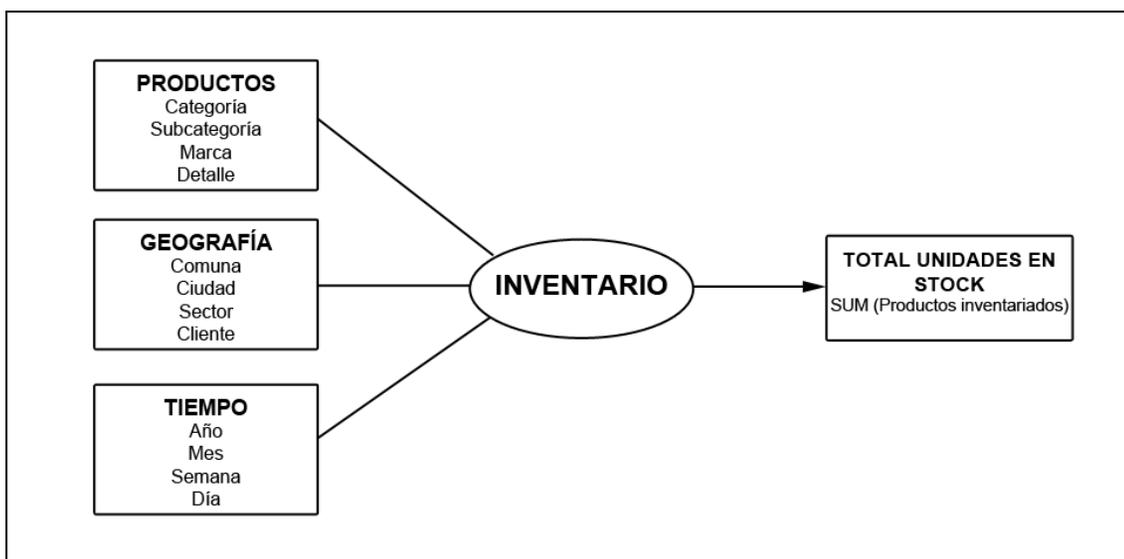


Figura 33. Modelo conceptual ampliado del proceso de inventario

La metodología Hefesto define que este tipo de modelos son suficientes para que el cliente pueda determinar cómo funcionará la solución de Inteligencia de Negocios, pero la realidad es otra. A través de estos modelos solo se da a conocer las diferentes dimensiones de análisis de un proceso de negocio con sus indicadores correspondientes, pero no demuestran la navegabilidad que este podría tener en un cubo OLAP. Para graficar esto es necesario construir un modelo dimensional que muestre las diferentes jerarquías entre los atributos de las dimensiones.

En la figura 34 se ilustra el modelo dimensional el cual es extraído de los diferentes modelos conceptuales ampliados construidos con la metodología Hefesto.

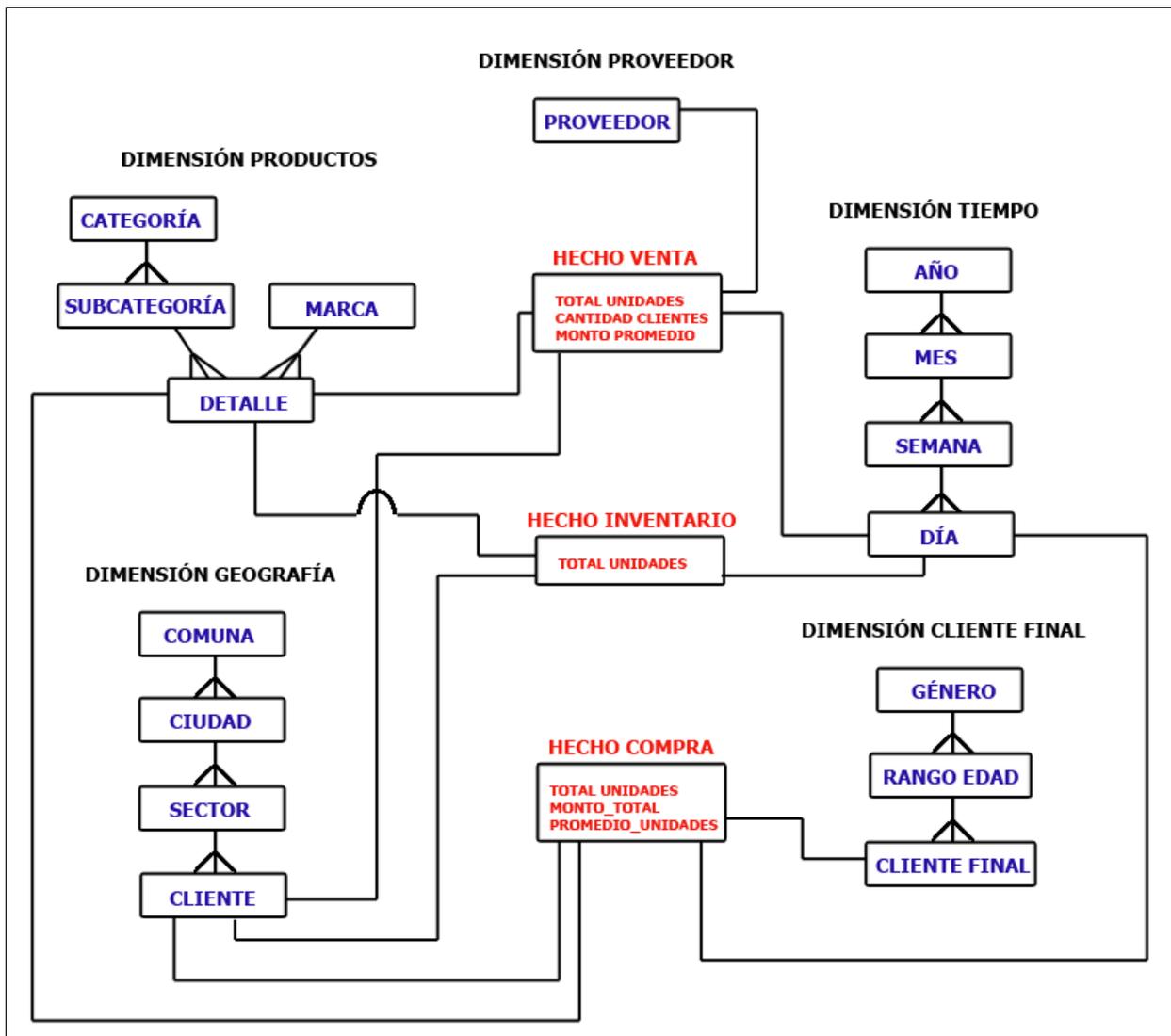


Figura 34. Modelo dimensional

5. Modelo lógico del Data Warehouse

A pesar de que Hefesto defina esta etapa como “Modelo lógico”, si se hiciera una comparación a la metodología de modelado dimensional de Kimball en esta parte se realizará el esquema físico a base de los modelos conceptuales ampliados.

5.1 Tipo de Modelo Lógico del DW

Debido a que existe más de un modelo conceptual, es decir existen diferentes procesos desde donde se quiere analizar la información se realizará una tabla de hecho por cada modelo conceptual. Al contar con más de una tabla de hecho necesariamente se tendrá que construir un esquema constelación.

5.2 Tablas de dimensiones

Para realizar tablas de dimensiones para un esquema constelación se debe seguir los mismos pasos que implicaría realizarlas en un esquema estrella. Para ello deben seguirse los siguientes pasos:

- Elegir un nombre que represente la tabla de dimensión. En esta oportunidad se utilizará las convenciones de nombres del modelado dimensional de Kimball, es decir se antepondrá la palabra “lookup”.
- Añadir solamente un campo que represente su clave principal, y no un campo clave compuesto por las id de cada atributo.
- Como los atributos son lo suficientemente intuitivos pasarán a ser campos en las tablas de dimensiones sin redefinir su nombre a excepción de las que contengan tildes y caracteres especiales.

En la figura 35 se ilustra cómo debe construirse una tabla de dimensión en la metodología Hefesto.

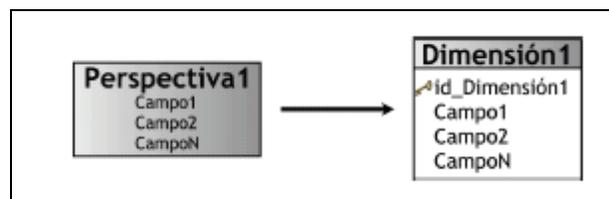


Figura 35. Diseño de tablas de dimensiones según la metodología Hefesto

A continuación se presentan las tablas de dimensión obtenidas a base de los modelos conceptuales ampliados.

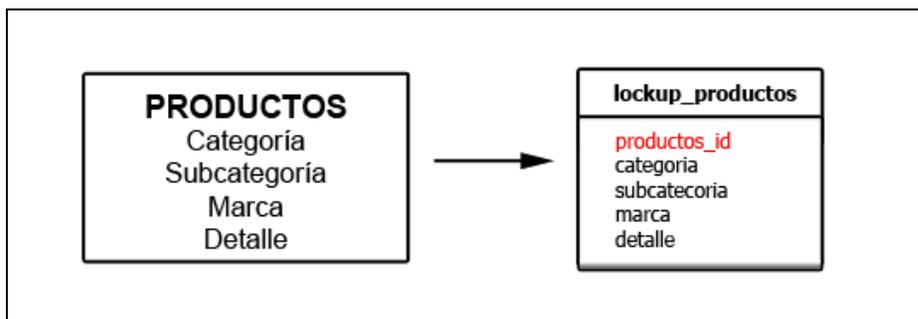


Figura 36. Tabla de dimensión global productos

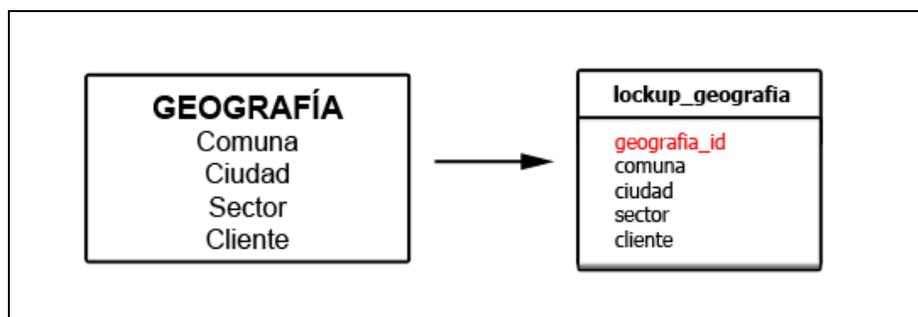


Figura 37. Tabla de dimensión global geografía

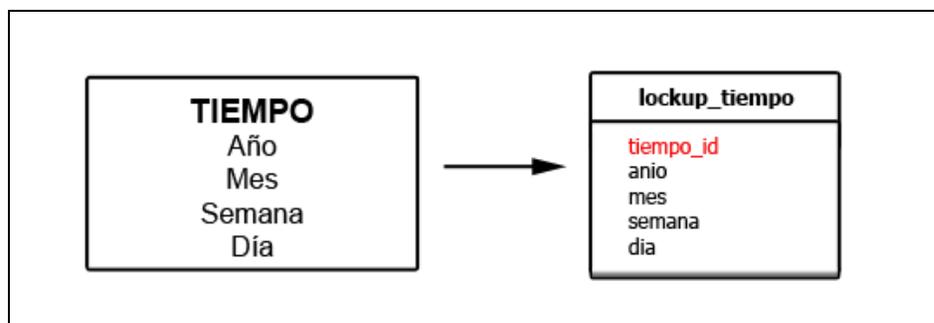


Figura 38. Tabla de dimensión global tiempo

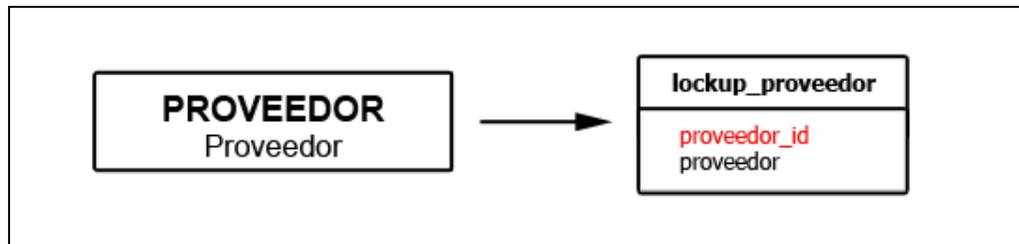


Figura 39. Tabla de dimensión local proveedor

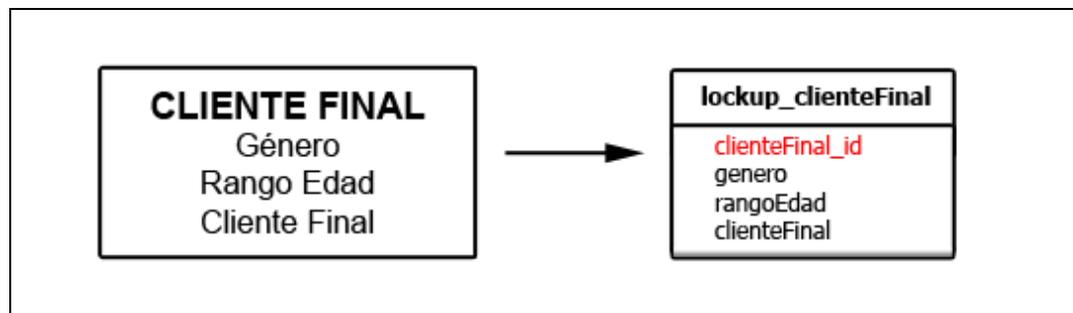


Figura 40. Tabla de dimensión local cliente final

5.3 Tablas de hechos

La construcción de tablas de hechos en un esquema constelación puede llegar a ser un poco más complejo que en otros esquemas. Para ello se deben realizar los siguientes pasos:

- Cada tabla de hecho tendrá el nombre del proceso que represente. Al igual que las tablas de dimensiones se utilizará la convención de nombre utilizada en la metodología Kimball anteponiendo la palabra “fact”.
- La clave primaria se compone de todas las id de las dimensiones asociadas.

- Incluir todos los hechos asociados eliminando la descripción del proceso ya que el nombre de la tabla lo describe. Además se eliminan todas las tildes y caracteres especiales.

En la figura 41 se ilustra cómo se construye una tabla de hechos en la metodología Hefesto.

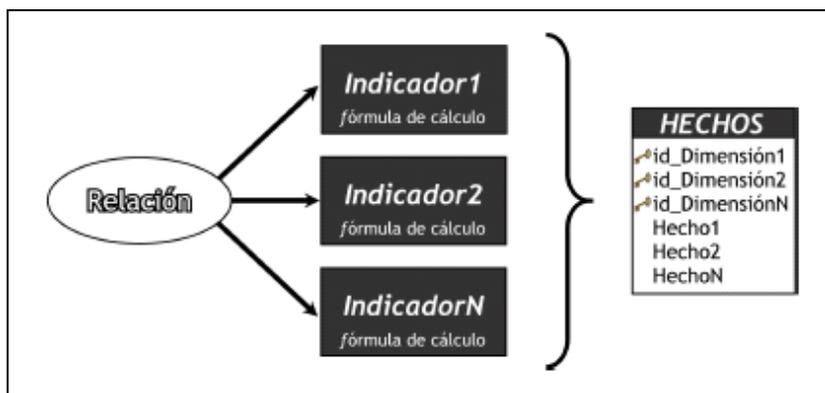


Figura 41. Diseño de tablas de hechos según la metodología Hefesto

A continuación se presentan las tablas de hechos obtenidas en base a los modelos conceptuales ampliados. Como en este caso se requiere saber cuáles son las dimensiones asociadas también se ilustraran en las figuras.

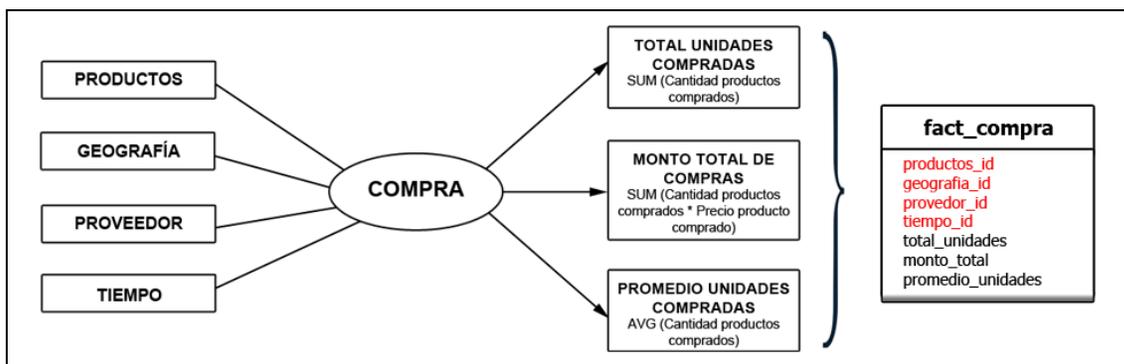


Figura 42. Tabla de hechos para el proceso de compra

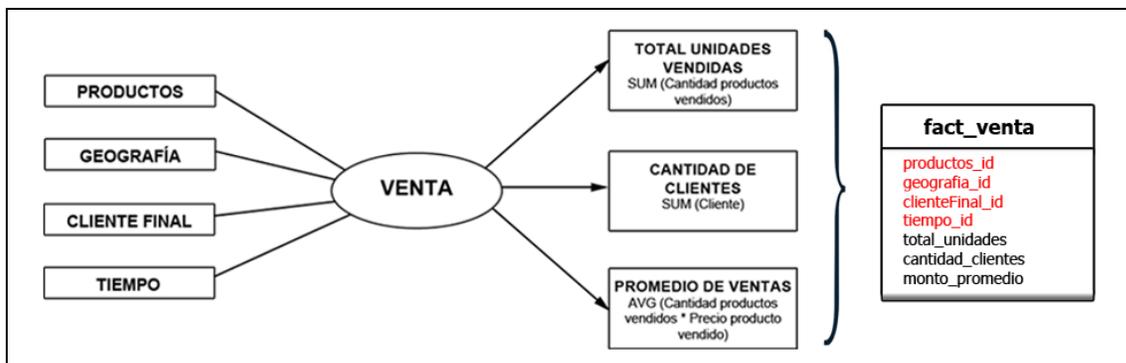


Figura 43. Tabla de hechos para el proceso de venta

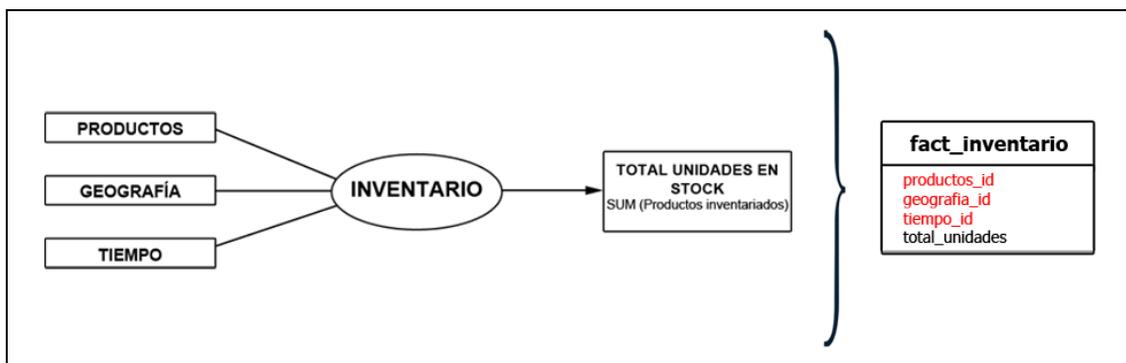


Figura 44. Tabla de hechos para el proceso de inventario

5.4 Esquema completo

En la figura 44 se presenta el esquema constelación completo resultante al aplicar la metodología Hefesto sobre los datos encuestados.

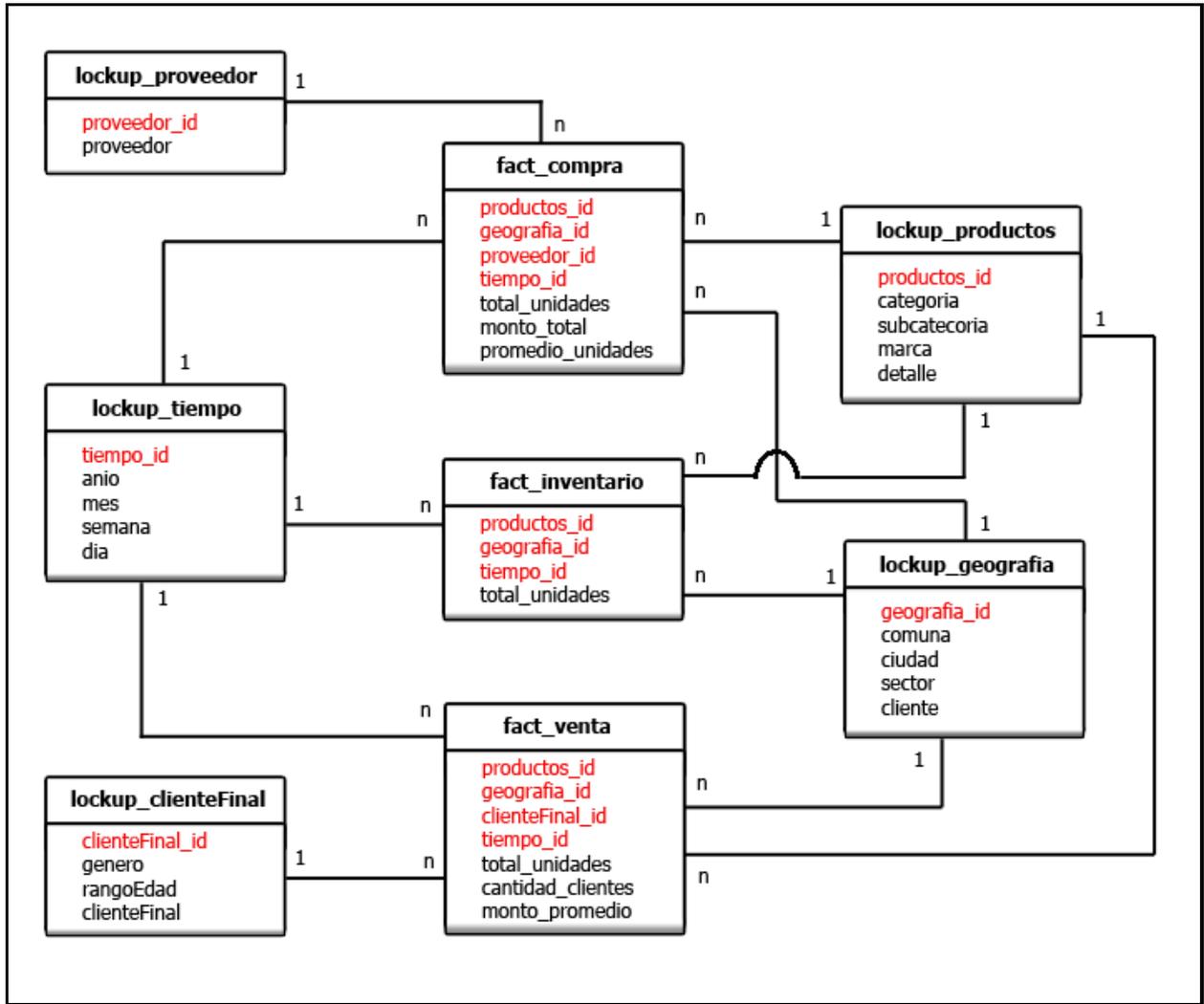


Figura 44. Esquema completo

IV. Conclusiones

El Data Warehouse creado se compone de datos especialmente extraídos para el análisis de dimensiones e indicadores clave para las empresas de distribución abarrotera. Estos datos se encuentran disponibles para crear visualizaciones basadas en cubos OLAP o en informes estáticos que permitan facilitar la toma de decisiones.

Independientemente de que el modelo de Inteligencia de Negocios haya sido definido para este sector, todas las empresas que tengan interés por la información contenida en él pueden verse beneficiadas:

- Los negocios de barrio obtienen un mayor conocimiento de su negocio y a través de las dinámicas de fidelización del cliente final se puede encontrar información que permitirá ordenar y redefinir sus procesos de venta para mejorarlas.
- Las empresas mayoristas que abastecen a las empresas distribuidoras también encontrarán información relevante de su dimensión de datos desconocida. Se recuerda que las compras de los negocios de barrio son las ventas de las empresas distribuidoras.

De acuerdo a esto el cuadro de beneficios en torno a la solución de Inteligencia de Negocios basada en el Data Warehouse entregado es el que se gráfica en la figura 45:

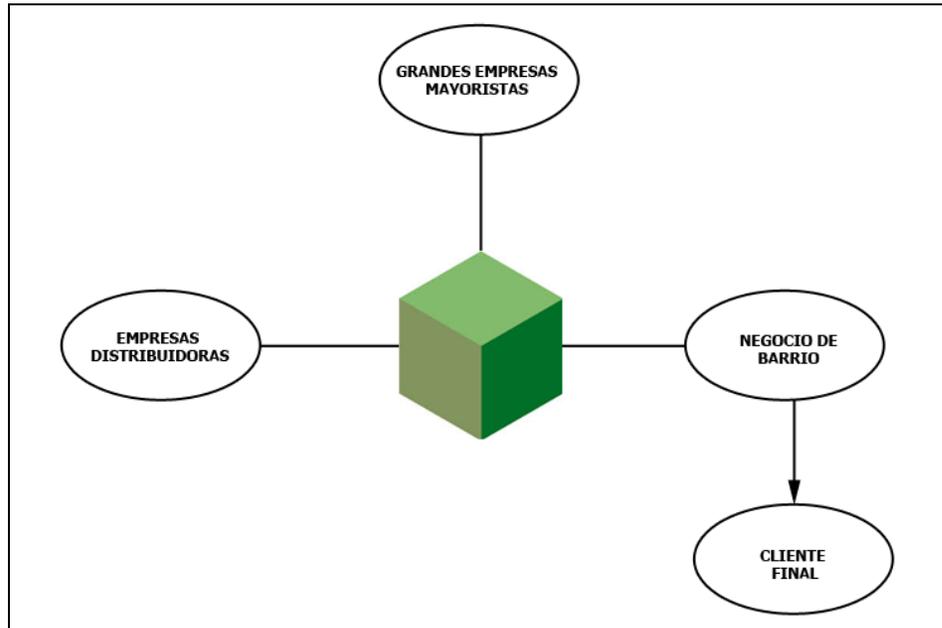


Figura 45. Empresas beneficiadas.

Se concluye además que la información solicitada por las empresas encuestadas puede generar otros tipos de análisis que pueden ayudar a potenciar el negocio.

a) Comprobación de compra o importación de productos

Es de interés común saber que productos son los que más se están vendiendo de manera que se pueda determinar si vale la pena comprarlos o importarlos (en caso de algunos de los grandes mayoristas), para su futura distribución.

Cubo 1: "Lista de los productos menos vendidos en un periodo de tiempo, a través de las dimensiones producto y tiempo".

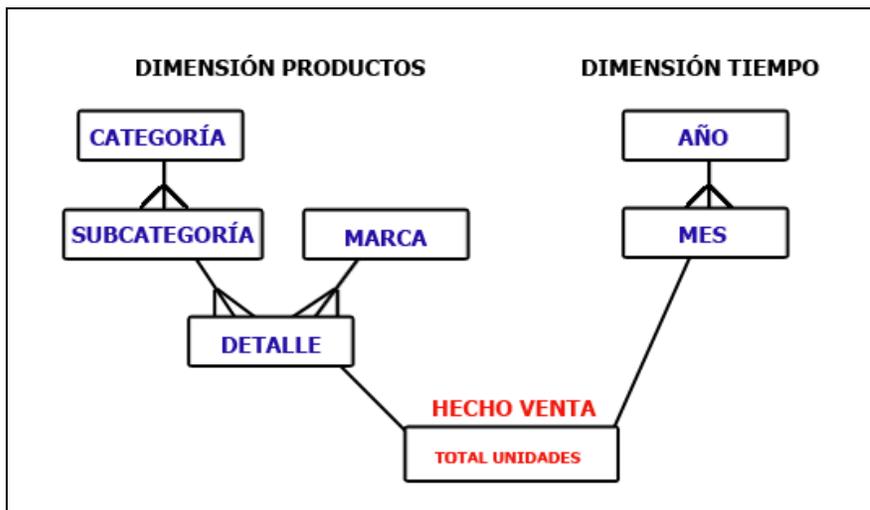


Figura 46. Cubo 1

Para el mismo propósito se puede determinar si algunos productos se estancan demasiado en bodega.

Cubo 2: “Lista de los productos que pasan más tiempo almacenados en bodega”.

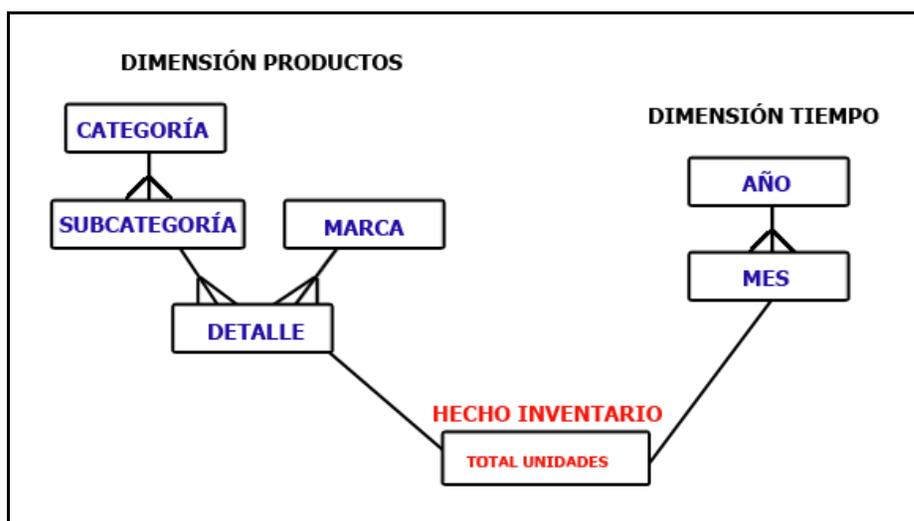


Figura 47. Cubo 2

b) Descubrir nuevos productos

Se puede verificar la existencia de productos que no estén siendo comprados al proveedor al cual se le realiza el análisis, de esta manera se podrán verificar nuevos mix de productos que han sido descubiertos por la competencia o productos que no son distribuidos por nadie y que los negocios pequeños han comprado en otros comercios, como en ferias o supermercados mayoristas.

Cubo 3: “Listado de productos vendidos en un tiempo determinado ordenados por proveedores, a excepción del proveedor a quien se le realice el análisis”

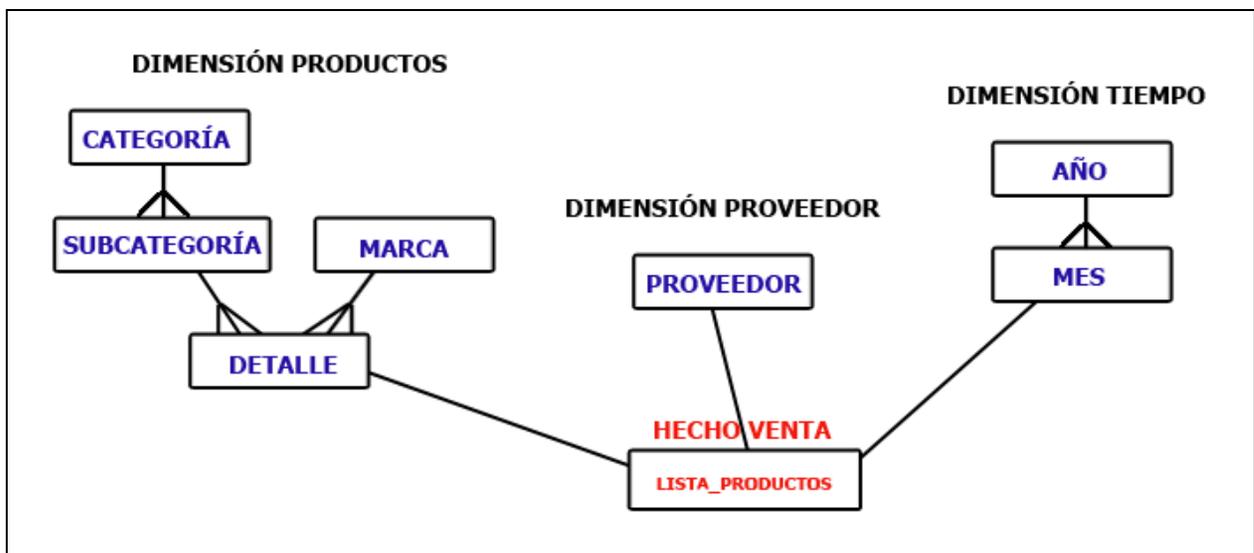


Figura 48. Cubo 3

Para este cubo en particular se debe crear un nuevo indicador que realice un SELECT DISTINCT al atributo “detalle” de la dimensión productos.

c) Mejoras en el margen gracias al comercio asociado

Generalmente las empresas de los sectores abarcados en esta tesis (ya sea mayorista, distribuidora o negocio de barrio) compiten entre ellas. Actualmente una fórmula diferente para competir frente al sucursalismo de las grandes cadenas es que cooperen entre ellas formando redes empresariales.

Un tipo de comercio asociado que puede adaptarse a las necesidades de los negocios de barrio de la región es la central de compra, la cual tiene como principal ventaja realizar compras en mayores cantidades y por lo tanto tener mayor poder negociador con los proveedores, lo que se traduce directamente en la obtención de menores precios de compra. Esto es posible realizarlo basado en la información generada desde el Data Warehouse creado.

Existen varias formas de aumentar las utilidades de una empresa, como puede ser la mejora de los procesos para aumentar la productividad o bajar costos como la mano de obra, pero el mayor beneficio para los negocios de barrio es aumentar los márgenes debido a que su negocio se centra principalmente a la compra y venta de productos, siendo las ganancias el margen que se pueda obtener.

El margen es el beneficio que se obtiene de la diferencia del costo de un producto y su valor de venta, por lo que si se consiguen los productos a un menor costo, mayores serán las ganancias de este tipo de negocios.

Esto se puede lograr con la central de compras, cuyas decisiones de negociación son basadas en el Data Warehouse creado generando el siguiente cubo:

Cubo 4: "Cantidad de productos vendidos mensualmente y su precio promedio ordenados por el detalle del producto y el cliente que los compra"

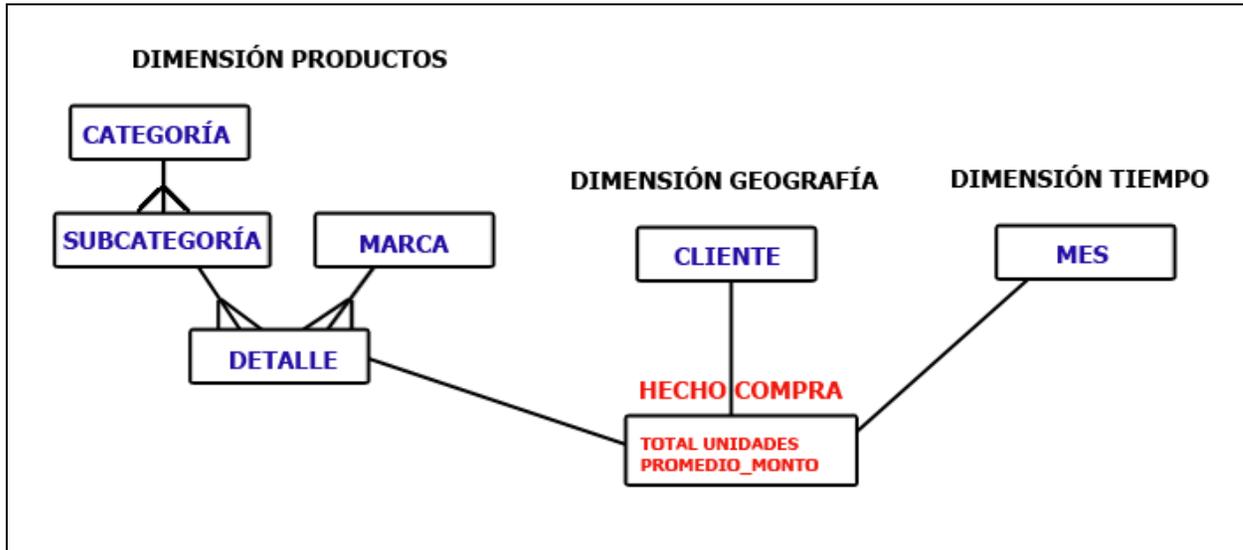


Figura 49. Cubo 4

Con la información extraída de este cubo se puede determinar a qué precio están comprando los pequeños negocios un determinado producto y consolidar un pedido a un conjunto de estos de manera que el volumen de un producto comprado aumente para poder negociar con las empresas distribuidoras un mejor costo. El modelo de negocio aplicado a una central de compras basada en Inteligencia de Negocios se ilustra en la figura 50.

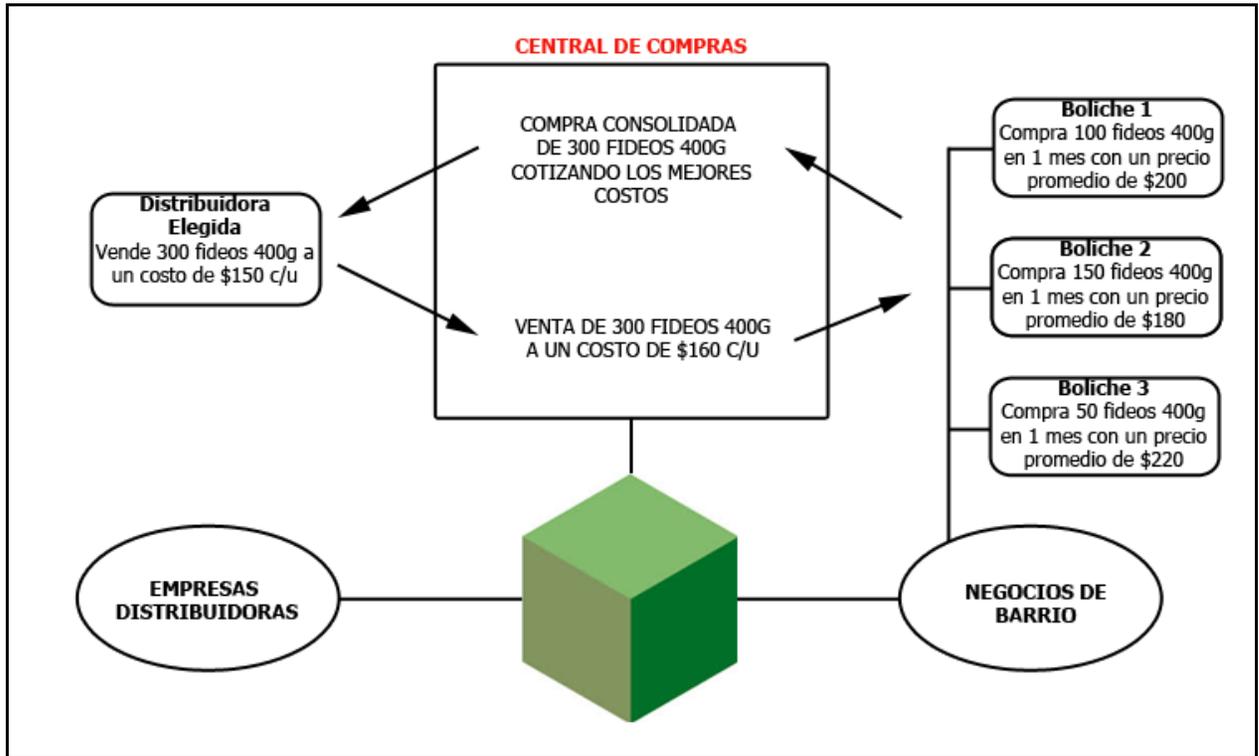


Figura 50. Central de Compras con Inteligencia de Negocios

Glosario

Agregación

Proceso de precalcular el cálculo de los datos a través de los niveles, para disminuir los tiempos de respuestas en los procesos de búsquedas de información.

Batch (Batch Processing)

Es la ejecución de un programa sin el control o supervisión directa del usuario. Este tipo de ejecución se utiliza en tareas repetitivas sobre grandes conjuntos de información, ya que sería tedioso y propenso a errores realizarlo manualmente.

BI (Business Intelligence)

Son los procesos, tecnologías y herramientas necesarias para convertir los datos en la información y la información en conocimiento y el conocimiento en los planes que conducen a una acción de negocio rentable.

CRUD (Create, Read, Update, Delete)

Acrónimo de crear, obtener, actualizar y borrar aludiendo a las cuatro operaciones básicas que se realizan sobre la información en una en base de datos.

Data Mart

Es un subconjunto del Data Warehouse, el cual puede ser implementado como un pequeño proyecto de DW generalmente orientado a un solo proceso de negocio o a un grupo de ellos relacionados en torno a un proceso de negocio en particular.

DSS (Decision Support System)

Es un sistema informático que utiliza información y modelos matemáticos para ayudar a los usuarios a tomar decisiones empresariales adecuadas según las condiciones del mercado y la situación interna de la compañía.

DW (Data Warehouse)

Es una colección de datos orientado al sujeto, no volátil, integrado y variante en el tiempo para ayudar al proceso de toma de decisiones gerenciales.

ETL (Extract, Transform and Load)

Es el proceso que permite mover datos desde múltiples fuentes, reformatearlos, limpiarlos y cargarlos en otra base de datos para propósitos de análisis

Granularidad

Representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando.

HOLAP (Hybrid Online Analytical Process)

Es una combinación de ROLAP y MOLAP. Los datos precalculados y agregados se almacenan en base de datos multidimensionales y los de mayor detalle en estructuras relacionales.

Integridad Referencial

Es un sistema de reglas que utilizan la mayoría de las bases de datos relacionales para asegurarse que los registros de tablas relacionadas son válidos y que no se borren o cambien datos relacionados de forma accidental produciendo errores de integridad.

Join

Es una sentencia SQL permite combinar registros de dos o más tablas en una base de datos relacional

Market Basket

La “Canasta de Mercado” por su traducción al español es un indicador que permite encontrar tendencias de compra de los clientes a los cuales se pueden asociar ciertas promociones o descuentos.

Marketing

Se trata de la disciplina dedicada al análisis del comportamiento de los mercados y de los consumidores. El marketing analiza la gestión comercial de las empresas con el objetivo de captar, retener y fidelizar a los clientes a través de la satisfacción de sus necesidades.

Mix de Productos

Son todos los bienes o servicios que ofrece al público una empresa, llámese tangibles o intangibles, los cuales si no se realizan buenas estrategias para promoverlos pueden decaer en cuanto en su demanda por lo que es muy importante tener claro cuáles son estos productos que ofrece la empresa y como estimular al consumidor para que los adquiera.

MOLAP (Multidimensional Online Analytical Processing)

Herramienta diseñada para el análisis de datos implementado en una base de datos dimensional. En un sistema MOLAP los cubos OLAP son generados con correspondencia a la estructura de datos.

OLAP (On-Line Analytical Processing)

Es una solución que tiene como objetivo agilizar la consulta de grandes cantidades de datos a través de estructuras multidimensionales para ser utilizados en análisis.

OLTP (On-Line Transaction Processing)

Representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer.

Proceso de Negocio

Es un conjunto de tareas lógicamente relacionadas que existen para conseguir un resultado bien definido dentro de un negocio; por lo tanto, toman una entrada y le agregan valor para producir una salida.

Retail

Es un sector económico que engloba a las empresas especializadas en la comercialización masiva de productos o servicios uniformes a grandes cantidades de clientes al por mayor o al detalle.

ROLAP

Herramienta diseñada para el análisis de datos implementado en una base de datos relacional. En un sistema ROLAP realmente no existen los cubos OLAP pero si los simulan, proceso que es transparente para el usuario

Share of Wallet

Es la “participación en la billetera del consumidor” que define la cantidad proporcional de dinero que un cliente invierte en los productos de una empresa del 100% que este tiene destinado comprar.

Stock

Conjunto de mercancías o productos que se tienen almacenados en espera de su venta o comercialización.

Bibliografía

Internet

- www-07.ibm.com/sg/events/blueprint/pdf/day1/Introduction_to_Business_Intelligence.pdf
- <http://www.microstrategy.com.ar/experienciaBI4/TeoriaDW.Argentina.pdf>
- <http://www.businessintelligence.info/docs/hefesto-v2.pdf>
- <http://cmapspublic2.ihmc.us/rid=1G50K43BC-7MFK48-C1Q/CARACTERISTICAS%20SISTEMAS%20TRANSACCIONALES.cmap>
- <http://www.businessintelligence.info/dss/dss-apoyo-decisiones.html>
- http://www.cs.uoi.gr/~pvassil/publications/2009_DB_encyclopedia/DW_metadata.pdf
- <http://www.informatica-hoy.com.ar/informatica-tecnologia-empresas/OLAP-Procesamiento-analitico-de-la-informacion-empresarial.php>
- <http://technet.microsoft.com/es-es/library/hh916543.aspx>
- <https://explorable.com/es/poblacion-de-la-investigacion>
- <http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>
- <http://olap.com/learn-bi-olap/olap-bi-definitions/dimension/>
- <http://es.slideshare.net/miguelorquera/modelo-dimensional-de-un-proceso-de-negocio-12802599>
- <https://es.scribd.com/doc/39775571/Share-of-Wallet-1>
- http://www.ehowenespanol.com/analisis-canasta-mercado-info_362288/
- <http://cubos-artus.blogspot.com/2011/10/definicion-de-agregaciones.html>
- http://www.aulaclic.es/sql/b_8_1_1.htm
- <http://definicion.de/marketing/>
- <http://www.todomktblog.com/2013/06/mix-prod.html>

Libros

- Inmon, W. H. Building the Data Warehouse Fourth Edition. Wiley Publishing, Inc. 2005.
- Kimball, R. The Data Warehouse Life Cycle Toolkit Second Edition. Wiley Publishing, Inc. 2008.

Anexos

Tesis:	Modelo de Inteligencia de Negocios para los distribuidores abarroteros de la X región.
Alumno:	Cesar Mauricio Faúndez F.
Universidad:	Gabriela Mistral

Entrevista a distribuidores mayoristas de la zona sur.

Nombre Entrevistado:

Cargo Entrevistado:

Fundamentación de la tesis

La tesis busca Generar un modelo de Inteligencia de Negocios para detectar las necesidades de análisis de los distribuidores abarroteros de la región de los Lagos apoyando la toma de decisiones.

La Inteligencia de Negocios son todos los procesos, tecnologías y herramientas necesarias para convertir los datos en la información y la información en conocimiento y el conocimiento en los planes que conducen a una acción de negocio rentable.

La información disponible para su análisis son todas las dinámicas de negocio de sus clientes directos a la que actualmente no tienen acceso, los cuales se consolidarán en una nueva base de datos que contendrá información histórica desde el momento en que se agreguen a la solución.

En esta fase de la tesis de carácter investigativa se busca obtener e identificar las necesidades de información clave de alto nivel que es necesario para llevar a cabo las metas y estrategias de su empresa y que facilitará una eficaz y eficiente toma de decisiones.

Preguntas

En base a lo anterior responda las siguientes preguntas

- ¿Conoce la Inteligencia de Negocios? ¿Cómo es aplicada en su empresa?
- ¿Si tuviera a su disposición los datos de venta, compra y stock de sus clientes, que información le gustaría obtener? ¿Cómo obtiene esta información actualmente?
- ¿Usted apoyaría un sistema de fidelización de clientes en los negocios de barrio a los cuales usted abastece para así verse beneficiado con los datos que se generarán? ¿De qué forma?
- ¿Usted cree que esta implementación le otorgará beneficios? ¿Cuáles?