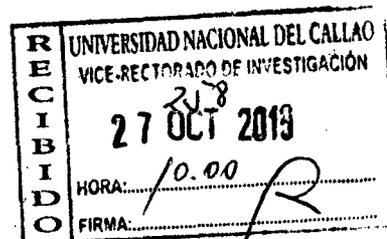


UNIVERSIDAD NACIONAL DEL CALLAO

**FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS
INSTITUTO DE INVESTIGACIÓN DE LA FACULTAD DE INGENIERÍA
INDUSTRIAL Y DE SISTEMAS**



INFORME FINAL DEL PROYECTO DE INVESTIGACIÓN

NOV 2015

**“METODOLOGIA PARA LA CREACION DE UNA
BASE DE DATOS MULTIDIMENSIONAL, USANDO
TECNOLOGÍA DE INTELIGENCIA DE NEGOCIOS Y
SQL SERVER EN UNA EMPRESA PERUANA.”**

AUTOR: BERTILA LIDUVINA, GARCÍA DÍAZ

**(PERIODO DE EJECUCIÓN: Del 01 de Noviembre del 2013 al 31 de
Octubre del 2015)**

**(Resolución de aprobación N° 1056-2013-R.- Callao, 25 de Noviembre del
2013)**

Callao, 2015

Handwritten initials

INDICE

I.- INDICE	1
II.- RESUMEN Y ABSTRACT	6
III.- INTRODUCCIÓN	
3.1 Planteamiento del problema	8
3.2 Importancia y justificación	9
IV.- MARCO TEORICO	
4.1 Inteligencia de negocios	10
4.1.1 Beneficios que aporta la inteligencia de negocios	12
4.1.2 Modelo de datos	13
4.1.3 Datawarehouse o almacén de datos	16
4.1.4 Cubos	22
4.1.5 Componentes de inteligencia de negocios	23
4.1.6 Proceso de extracción, transformación y carga (ETL)	24
4.1.7 Procesamiento analítico en línea (OLAP)	26
4.2 Lenguaje estructurado de consultas (SQL)	28
V.- MATERIALES Y METODOS	
5.1 Materiales	31
5.2 Población de la investigación y la muestra correspondiente	31
5.3 Técnicas, procedimientos e instrumentos de Recolección de datos	32
5.4 Técnicas de análisis o métodos estadísticos aplicados	32
5.5 Metodología	33
VI.- RESULTADOS	
6.1 Creación de la base de datos Ventas	34
6.2 Creación del datawarehouse	37
6.3 Proceso de extracción, transformación y carga (ETL)	39
6.4 Creación de Vistas	43
6.5 Creación de Cubos	45

6.3	Proceso de extracción, transformación y carga (ETL)	39
6.4	Creación de Vistas	43
6.5	Creación de Cubos	45
6.5.1	Creación del cubo detalle venta	45
6.5.2	Creación del cubo completo de ventas	47
6.5.3	Creación del cubo ventas	53
6.6	Creación de cubos en Visual Studio 2010 y sus respectivos gráficos en Excel	53
6.6.1	Por categoría e importe	54
6.6.2	Por categoría, año e importe	55
6.6.3	Por categoría, año, trimzestre e importe	56
6.6.4	Por compañía e importe	57
6.6.5	Por producto	58
6.7	Gráficos adicionales en Excel	59
VII.- DISCUSIÓN		60
VIII.- REFERENCIALES		62
IX.- APÉNDICE		64
1.-	Apéndice N° 1 Mapa mental del Datawarehouse	64
2.-	Apéndice N° 2 Separata de Datawarehouse	65
X.- ANEXOS		72
1.-	Anexo N° 1 Matriz de Consistencia	73
2.-	Anexo N° 2 Términos importantes	74

INDICE DE FIGURAS

Figura N° 4.1	Definición de inteligencia de negocio	12
Figura N° 4.2	Definición de modelización del negocio	13
Figura N° 4.3	Definición de modelo de datos	14
Figura N° 4.4	Esquema estrella	15
Figura N° 4.5	Esquema copo de nieve	16
Figura N° 4.6	Arquitectura del datawarehouse	18
Figura N° 4.7	Mapa mental de un datawarehouse	19
Figura N° 4.8	Ejemplo de un datamart	20
Figura N° 4.9	Ejemplo de cubo	22
Figura N° 4.10	Componentes de inteligencia de negocios	24
Figura N° 6.1	Creación de la base de datos ventas	34
Figura N° 6.2	Crea la estructura de bdventas	35
Figura N° 6.3	Crea las claves foráneas	37
Figura N° 6.4	Diagrama de ventas	37
Figura N° 6.5	Crea el datawarehouse dwhventas	38
Figura N° 6.6	Realiza el proceso ETL	40
Figura N° 6.7	Diagrama de dwhventas	43
Figura N° 6.8	Creación del proyecto multidimensional ventas	43
Figura N° 6.9	Creación de la vista ventas	44
Figura N° 6.10	Creación de la vista detalle ventas	44
Figura N° 6.11	Creación de la vista completa ventas	45
Figura N° 6.12	Creación del cubo detalle venta	45
Figura N° 6.13	Mostrar el acumulado de importe	46
Figura N° 6.14	Mostrar el importe por categorías	46
Figura N° 6.15	Mostrar el importe por categoria y producto	47
Figura N° 6.16	Creación del cubo completo de ventas	47
Figura N° 6.17	Mostrar el importe de la factura por producto y por año	48
Figura N° 6.18	Mostrar el importe de la facturacion por producto, por año y por nombre de mes	48
Figura N° 6.19	Mostrar el importe de la facturacion por	49

	categoría, por año y por nombre de mes	
Figura N° 6.20	Mostrar el importe por compañía y país	49
Figura N° 6.21	Mostrar el importe por país	50
Figura N° 6.22	Mostrar el importe por país y año	50
Figura N° 6.23	Mostrar el importe por transportista	51
Figura N° 6.24	Mostrar el importe por transportista y año	51
Figura N° 6.25	Mostrar el importe por empleado	52
Figura N° 6.26	Mostrar el importe por empleado y por año	52
Figura N° 6.27	Creación del cubo ventas	53
Figura N° 6.28	Creación del cubo por categoría e importe	54
Figura N° 6.29	Creación del cubo por categoría, año e importe	55
Figura N° 6.30	Creación del cubo por categoría, año, trimestre e importe	56
Figura N° 6.31	Creación del cubo por compañía e importe	57
Figura N° 6.32	Creación del cubo del importe de la factura por producto	58

INDICE DE GRAFICOS

Grafico N° 4.1	Ejemplo del gráfico de un cubo	23
Grafico N° 6.1	Creación del gráfico por categoria e importe	54
Grafico N° 6.2	Creación del gráfico por categoria, año e importe	55
Grafico N° 6.3	Creación del gráfico por categoria, año, trimestre e importe	56
Grafico N° 6.4	Creación del gráfico por compañía e importe	57
Grafico N° 6.5	Creación del gráfico del importe de la factura por producto	58
Grafico N° 6.6	Mostrar el importe de la facturación por año	59
Grafico N° 6.7	Mostrar el importe de la facturación por año y por trimestre	59

II. RESUMEN

“Metodología para la creación de una base de datos multidimensional, usando tecnología de inteligencia de negocios y SQL server en una empresa peruana.”.

El objetivo de esta investigación fue desarrollar una metodología que permita la creación de bases de datos multidimensionales.

Para recopilar los datos de esta investigación se creó una Base de Datos Ventas, que a su vez fue la entrada para la creación del Datawarehouse DwhVentas. El método usado fue el inductivo.

Se analizó el proceso de ventas, para lo cual se creó la tabla de hechos: factVentas, y se eligió como métrica: el importe de las ventas.

También se diseñó varias tablas de dimensiones, que permitieron evaluar el proceso de ventas por diferentes vistas. Ejemplo: por fecha, cliente, empleado. etc.

El Datawarehouse DwhVentas fue la entrada a su vez para crear cubos en Visual Basic 2010, que se logró con la ayuda del Analysis Services Multidimensional. Estos cubos dan informes estadísticos por diferentes variables definidas en las dimensiones.

Los cubos creados fueron exportados a Excel, donde se creó varios gráficos de estos cubos, por las diferentes dimensiones disponibles.

En conclusión esta metodología permitió la creación de diferentes vistas multidimensionales (Procesamiento analítico en línea), en cuestión de segundos, lo cual no sería posible con un enfoque tradicional (Procesamiento de transacciones en línea).

PALABRAS CLAVES:

Base de datos multidimensional, Inteligencia de negocios, SQL server, vistas multidimensionales, Procesamiento analítico en línea, Procesamiento de transacciones en línea.

ABSTRACT

"Methodology for the creation of a multidimensional database using business intelligence technology and SQL Server on a Peruvian company."

The objective of this research was to develop a methodology for creating multidimensional databases.

To collect research data to a database was created Sales, which in turn was the entrance to the creation of the data warehouse DwhVentas. The method used was the inductive.

The process of sales was analyzed, for which created the table of facts: factVentas, and it was chosen as metrics: the amount of the sales.

More dimension tables that allowed evaluate the sales process is also designed different views. Example: by date, customer, employee. etc.

The Datawarehouse DwhVentas was the entry in turn to create buckets in Visual Basic 2010, which was achieved by the help of the Multidimensional Analysis Services. These buckets give statistical reports for different variables defined in the dimensions.

The created buckets were exported to Excel, where there were created several graphs of these buckets, for the different available dimensions.

In conclusion this methodology allowed the creation of different views multidimensional (Online Analytical Processing), in seconds, which would not be possible with a traditional approach (online transaction processing).

KEY WORDS:

Multidimensional database, business Intelligence, SQL server, multidimensional conference, analytical Processing on line, Processing transaction on line.

III. INTRODUCCIÓN

3.1 Planteamiento del problema

“Metodología para la creación de una base de datos multidimensional, usando tecnología de inteligencia de negocios y SQL server en una empresa peruana”.

Esta investigación de tipo cualitativa, utilizó la información almacenada en las grandes Bases de Datos transaccionales con el objetivo de convertirla en información para la toma de decisiones, para este fin fue necesario crear una metodología para hacerla realidad.

La demanda de beneficios, la contracción de la economía, el crecimiento de la competencia y clientes más inteligentes, son factores que requieren que las Empresas tomen las mejores decisiones posibles en base a la información existente, generalmente almacenada en las Bases de Datos Operacionales.

A medida que las Empresas buscan nuevas formas de diferenciarse y crear ventajas competitivas, la Inteligencia de Negocios se ha convertido en prioridad estratégica. Ampliar y mejorar el uso de la Inteligencia de Negocios permite otorgar al personal de todas las organizaciones información más puntual, relevante y lista para usar que puede impulsar mejores decisiones, acciones y procesos comerciales.

Inteligencia de Negocios caracteriza una amplia variedad de tecnologías, plataformas de software, especificaciones de aplicaciones y procesos.

Sin embargo no existe una metodología para crear Bases Multidimensionales que permita realizar un análisis multidimensional.

La presente investigación creó una metodología para la creación de una base de datos multidimensional, usando inteligencia de negocios, con un software como SQL y utilizando la información almacenada en las Bases de datos operacionales para construir un

data warehouse con el fin de ayudar a la toma de decisiones en las Empresas.

En resumen el planteamiento del problema es el siguiente:

¿Cómo crear un procedimiento o Metodología para crear Bases de Datos multidimensionales que apoyen a la toma de decisiones, usando tecnología de Inteligencia de Negocios y SQL server?

El objetivo principal fue desarrollar una Metodología para crear una Base de Datos Multidimensional usando tecnología de inteligencia de Negocios y el Sistema de Gestión de Base de Datos SQL server, utilizando la información almacenada en las Bases de Datos Operacionales. Corresponde a un tipo de investigación básica.

3.2 Importancia y justificación

La ejecución del presente trabajo de investigación, es importante porque no existe un camino o método para diseñar una Base de Datos Multidimensional, de ahí la importancia de realizar un estudio al respecto, para que con elementos de juicio podamos crearlas a partir del uso de la información existente en las Bases de Datos Operacionales (OLTP), y llegar a una conclusión elaborada y discutida. Dicho aporte es tecnológico.

Los hechos observados que motivaron la realización de la presente investigación fueron la carencia de información de una metodología para incorporar la Inteligencia de Negocios en las Empresas y la utilidad que se puede obtener de las Bases de Datos Operacionales para ayudar a la toma de decisiones, y observar por ejemplo dónde no hay ventas o porqué sí hay ventas en determinados lugares o porqué hay más ventas en determinados periodos y en otros no. Es decir analizando los resultados positivos y negativos del análisis multidimensional.

El valor de la investigación radica en los resultados de la investigación que repercuten en el desarrollo tecnológico ya que no existe un trabajo similar que le preceda.

IV. MARCO TEORICO

En cuanto a las teorías y/o leyes y/o doctrinas estrechamente relacionadas con el problema y objetivos dados en el proyecto de investigación, de acuerdo a la operacionalización de la hipótesis del proyecto de investigación, existen las siguientes variables independientes: **Inteligencia de Negocios y SQL Server**, que son necesarios conocer para demostrar la variable dependiente que es una metodología para crear Bases de Datos Multidimensionales, las cuales expondré a continuación.

No Existen trabajos anteriores que tengan el mismo objetivo que el presente trabajo de investigación.

4.1 Inteligencia de negocios

La Inteligencia de Negocios pretende convertir datos en información y a partir de la información ser capaces de descubrir conocimiento. Definiremos Inteligencia de Negocios a partir de la definición de términos de Gartner:

*“Inteligencia de Negocios es un **proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones** .”* (Banesto, 2014, p.23).

La Inteligencia de Negocios ayuda a tomar mejores decisiones analizando si las acciones tomadas están dando resultados hacia los objetivos de la compañía.

La Inteligencia de Negocios es necesaria cuando:

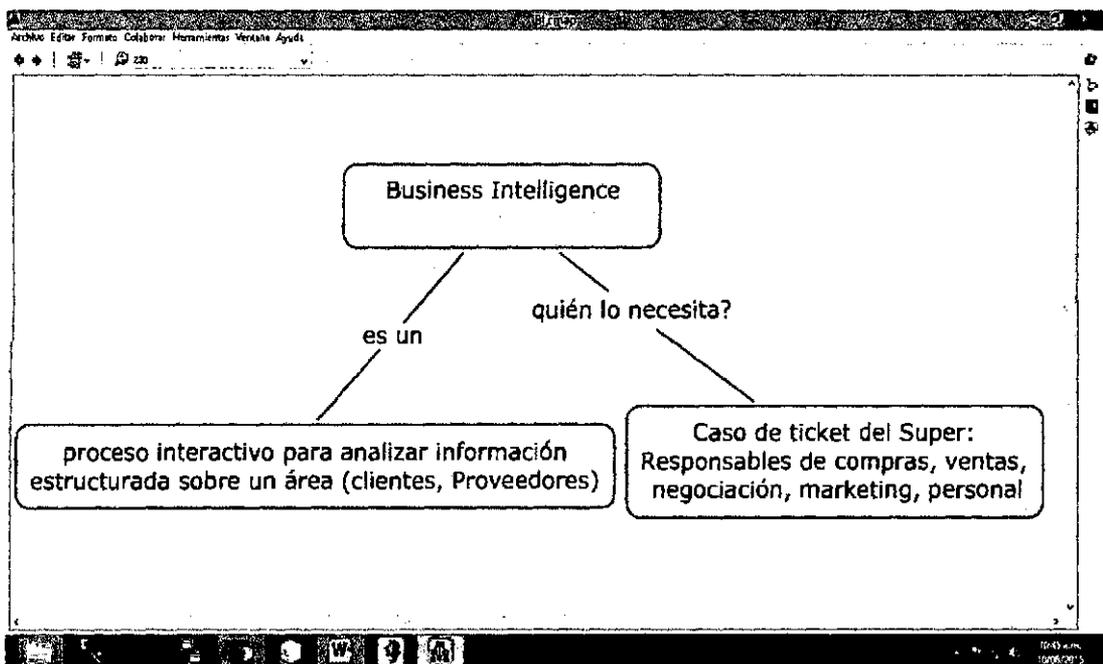
- La toma de decisiones se realiza de forma intuitiva en la organización.
- Problemas de calidad de información.
- Uso de excel.
- Cuando es necesario analizar un problema o problemas interrelacionados.

- Existe demasiada información en la organización.

En resumen la Inteligencia de Negocios buscan responder a las preguntas:

- ¿Qué pasó?
- ¿Qué pasa ahora?
- ¿por qué pasó?
- ¿Qué pasará?

FIGURA N° 4.1
DEFINICIÓN DE INTELIGENCIA DE NEGOCIOS



Fuente: Autor

4.1.1 Beneficios que aporta la inteligencia de negocios

Los beneficios que se pueden obtener a través del uso de la inteligencia de negocios pueden ser de distintos tipos:

- **Beneficios tangibles.**- por ejemplo: reducción de costes, generación de ingresos, reducción de tiempos para las distintas actividades del negocio.
- **Beneficios intangibles.**- el hecho de que tengamos disponible la información para la toma de decisiones hará que más usuarios

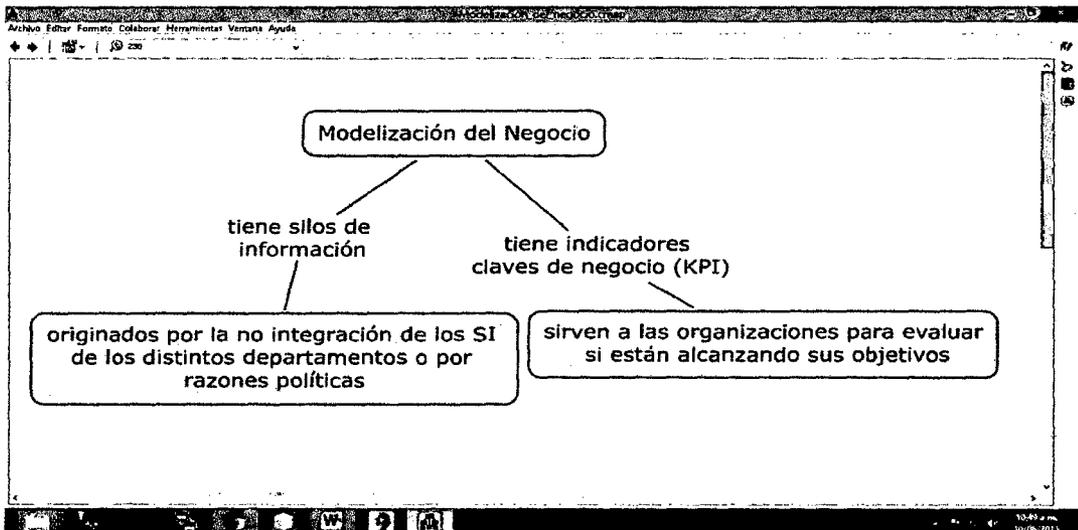
utilicen dicha información para tomar decisiones y mejorar nuestra posición competitiva. Ejemplo: optimizar la atención a los clientes, Aumentar la satisfacción de los clientes, Mejorar el acceso a los datos a través de consultas, análisis.

- **Beneficios estratégicos.-** Todos aquellos que nos facilitan la formulación de la estrategia, es decir, a qué clientes, mercados o con qué productos dirigirnos. Ejemplo: Mayor habilidad para analizar estrategias de precios, Mejorar la toma de decisiones, realizándola de forma más rápida, informada y basada en hechos, Mayor visibilidad de la gestión.

4.1.2 Modelo de datos

Utiliza el Modelo Relacional desarrollado por E.F.Codd, y está formado por tablas y relaciones entre las mismas. El modelo relacional utiliza el Standard Query Language (SQL).

FIGURA N° 4.2
DEFINICIÓN DE MODELIZACIÓN DEL NEGOCIO



Fuente: Autor

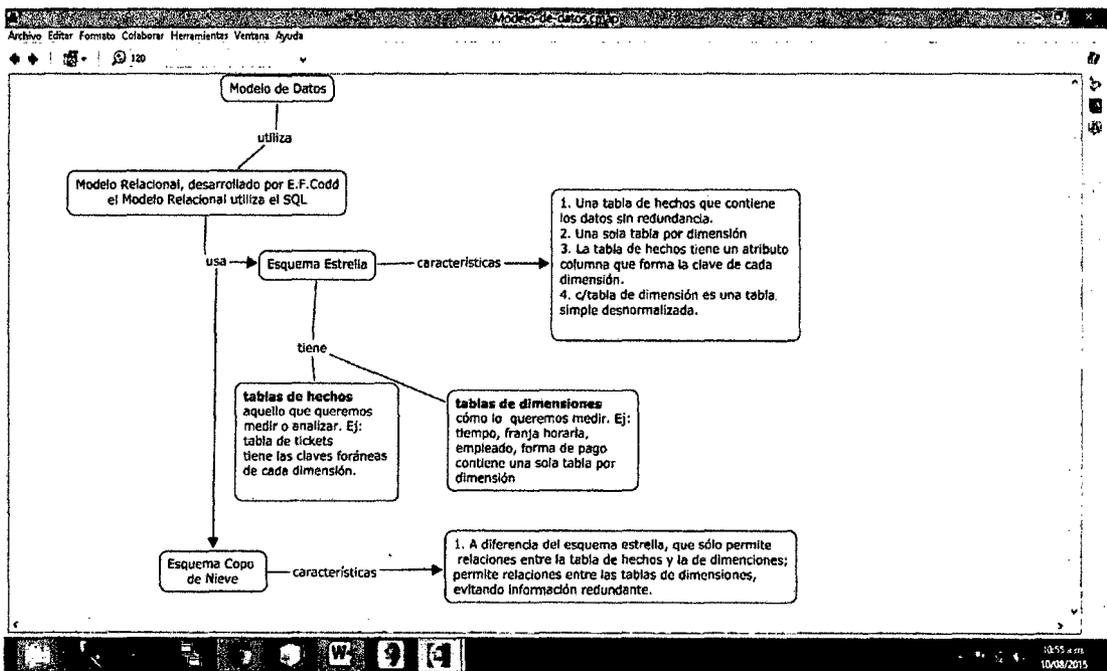
La idea principal es que la información sea presentada desnormalizada para optimizar las consultas. Para ello debemos identificar, en el seno de la empresa, los procesos de negocio, las

vistas para el proceso de negocio las medidas cuantificables asociadas a los mismos.

Los elementos de un Datawarehouse son:

- **Tabla de hecho:** Es la representación en el data warehouse de los procesos de negocio de la organización. Ejemplo: una venta.
- **Dimensión:** Es la representación en el Datawarehouse de una vista para un cierto proceso de negocio. Ejemplo: en un proceso de venta, sería el cliente que ha comprado, la fecha en que se ha realizado la venta.
- **Métrica:** son los indicadores de negocio. Son los conceptos cuantificables que permiten medir nuestro proceso de negocio. Ej: en una venta tenemos el importe de la misma.

FIGURA N° 4.3
DEFINICIÓN DE MODELO DE DATOS



Fuente: Autor

Utiliza los siguientes esquemas:

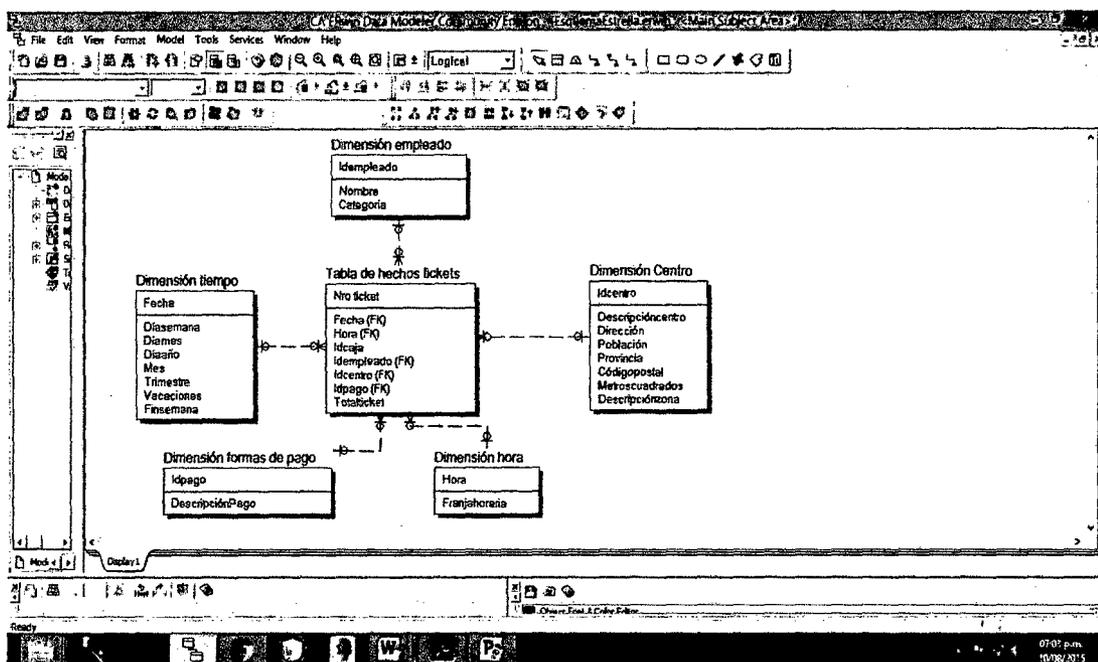
Esquema estrella

Esquema copo de Nieve

Esquema estrella.- Consiste en una tabla de hechos en el centro para el hecho objeto de análisis y una o varias tablas de dimensión por cada punto de vista de análisis que participa de la descripción de ese hecho. En la tabla de hecho encontramos los atributos destinados a medir sus métricas.

Para la construcción del esquema “estrella” debemos distinguir entre las tablas de hechos (aquello que queremos medir o analizar) y las tablas de dimensiones (cómo lo queremos medir), en un sistema de tickets de un supermercado, la tabla de hechos será la de los tickets y los queremos analizar por las dimensiones siguientes: tiempo, franja horaria, centro, empleado y forma de pago. El esquema “estrella” sería:

FIGURA N° 4.4
ESQUEMA ESTRELLA



Fuente: Banesto, Business Intelligence: Compartir con información,
2014

Las características del esquema estrella son:

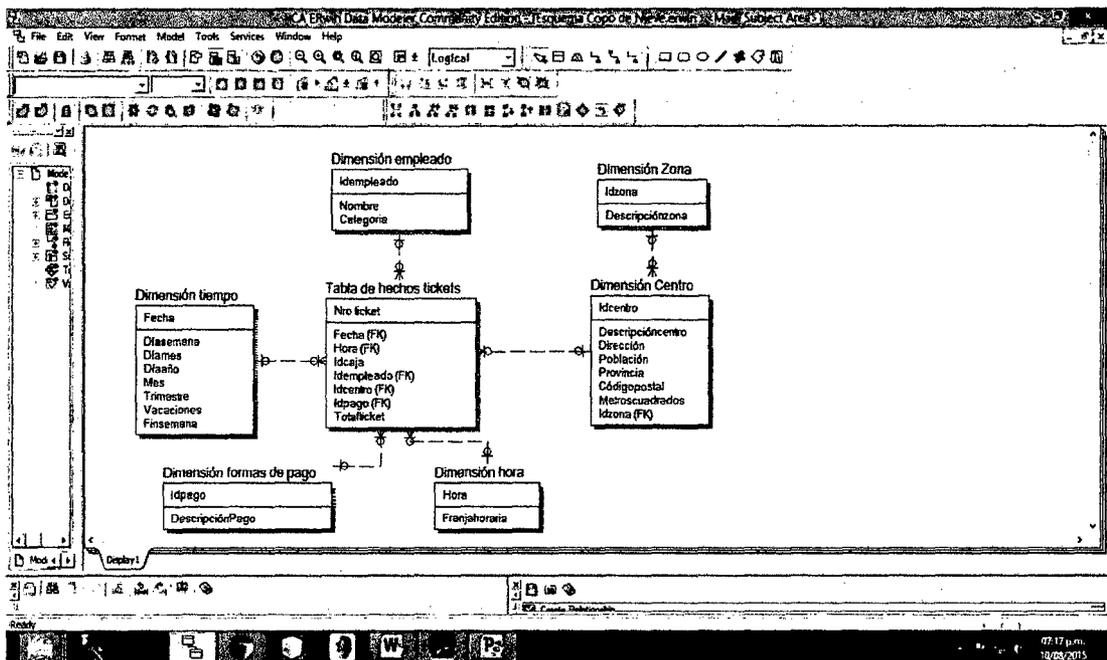
- Una tabla de hechos que contiene los datos sin redundancias.
- Una sola tabla por dimensión.

- La tabla de hechos tiene un atributo columna que forma la clave de cada dimensión.
- Cada tabla de dimensión es una tabla simple desnormalizada.

Esquema copo de nieve

Es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. En el esquema “copo de nieve” aparecen relaciones entre las tablas de dimensiones, mientras que en el esquema “estrella” sólo hay relaciones entre la tabla de hechos y las de dimensiones.

FIGURA N° 4.5
ESQUEMA COPO DE NIEVE



Fuente: Banesto, Business Intelligence: Compartir con información,
2014

4.1.3 Datawarehouse o almacén de datos

Un Data Warehouse (DW) es un repositorio central que contiene la información más valiosa de la empresa.

Los datos que aquí se almacenan han pasado por un proceso de calidad que asegura su consistencia. Además, el repositorio está construido de tal manera que el acceso sea lo más rápido posible.

Bill Inmon fue el que definió las características que debe cumplir un *datawarehouse*, que son: Orientado sobre un área, integrado, indexado al tiempo, es un conjunto no volátil de información que soporta la toma de decisiones.

Analicemos cada una de estas características detalladamente:

- “Orientado a un área”.- significa que cada parte del *datawarehouse* esta construida para resolver un problema de negocio. Por ejemplo: Entender los hábitos de compra de nuestros clientes, analizar la calidad de nuestros productos, etc.
- “Integrado”.- La información debe ser transformada en medidas comunes, códigos comunes y formatos comunes para que pueda ser útil.
- “Indexado en el tiempo “.- significa que se mantiene la información histórica y se almacena referida a determinadas unidades de tiempo, tales como horas, días, semanas, meses, trimestres o años. Ello nos permitirá analizar, por ejemplo, la evolución de las ventas en los periodos que queramos.
- “No volátil”.- significa que los usuarios no la mantienen, como lo harían en los entornos transaccionales. La información se almacena para la toma de decisiones. No se va actualizando continuamente, sino periódicamente, de forma preestablecida.

FIGURA N° 4.6 ARQUITECTURA DEL DATAWAREHOUSE

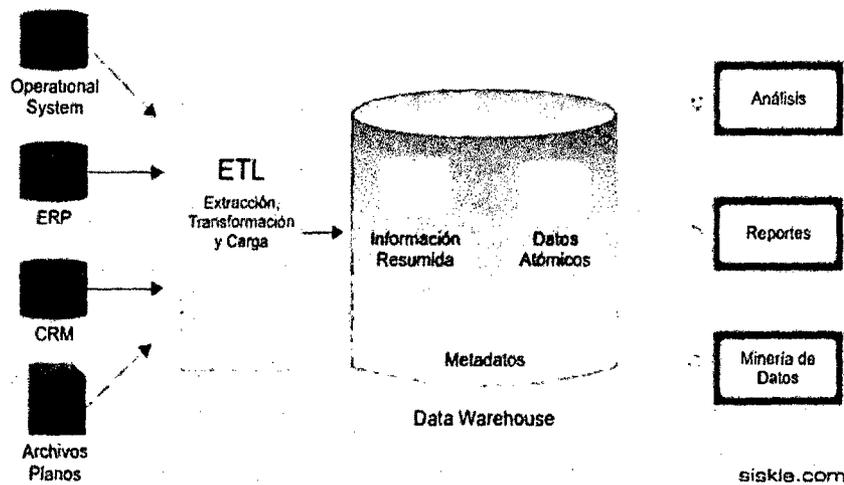


Figura 1. Arquitectura del DW

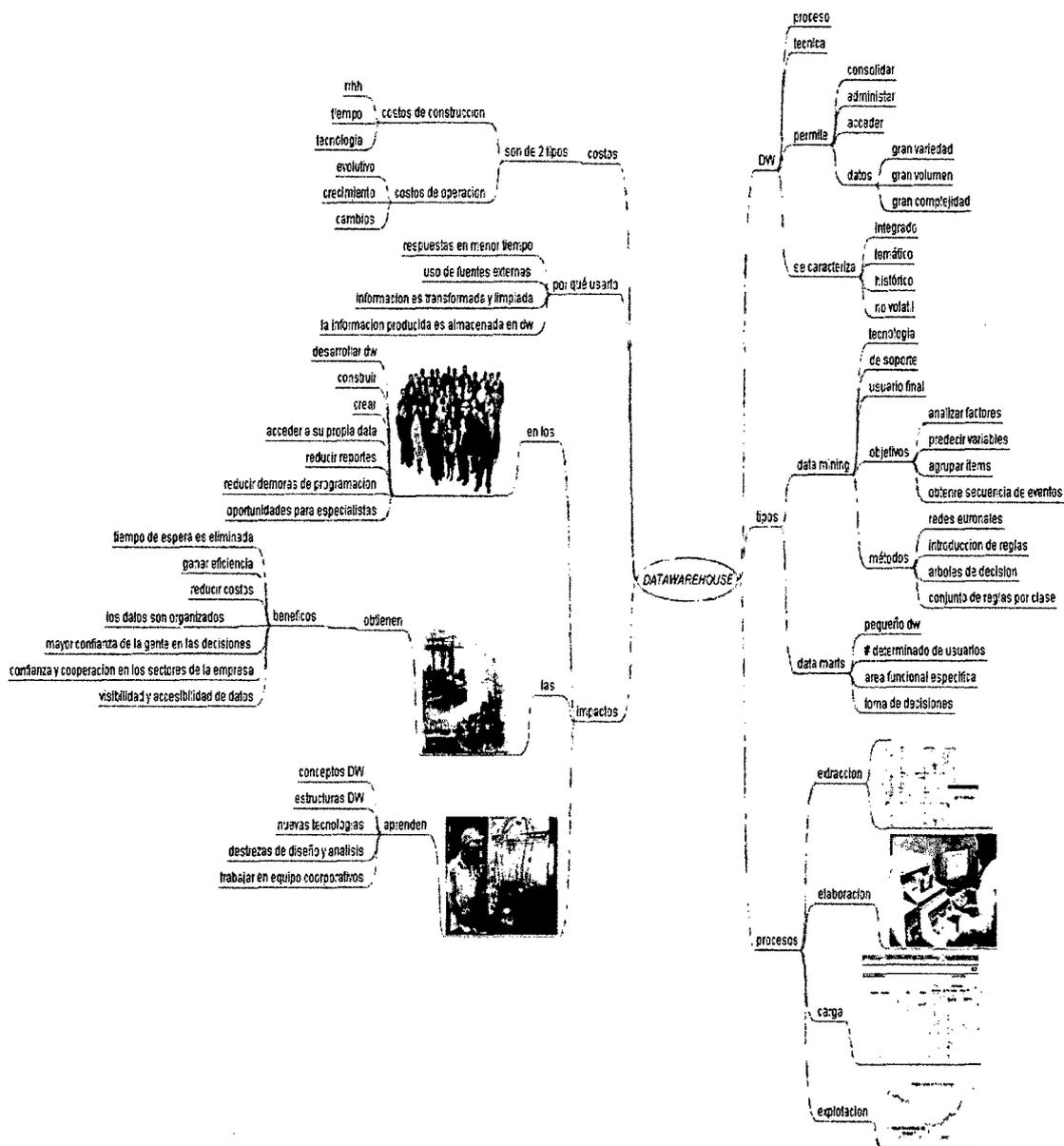
Fuente: División de alta tecnología: Cibertec

Principales aportaciones de un datawarehouse

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión espectaculares.

FIGURA N° 4.7

MAPA MENTAL DE UN DATAWAREHOUSE

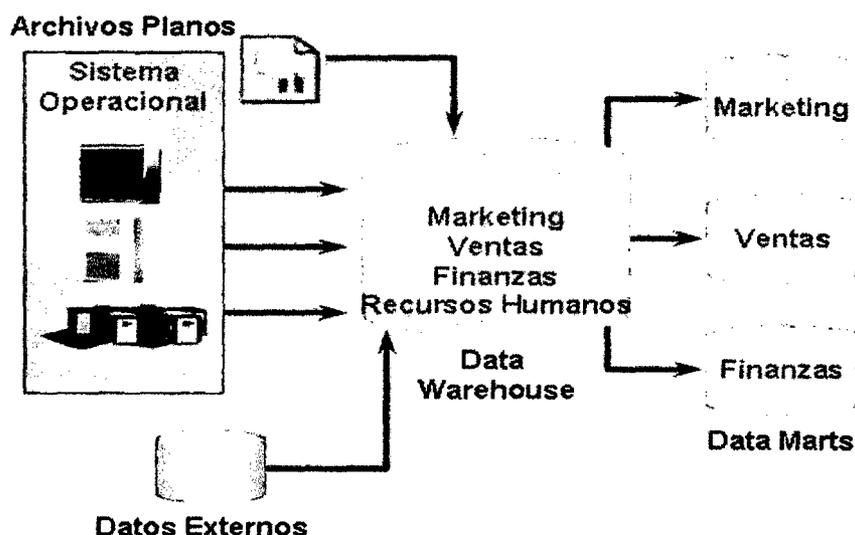


Fuente: Autor

Datamart

Un Datamart es un Data Warehouse mas pequeño. Normalmente la información contenida en un datamart es un subconjunto de un datawarehouse.

FIGURA N° 4.8
EJEMPLO DE UN DATAMART



Fuente: División de alta tecnología: Cibertec

(Revista de Investigación de Sistemas e Informática, 2013) se encontró lo siguiente:

Los pasos para la implementación del DataMart como solución de inteligencia de negocios según Business Intelligence Roadmap son:

1. Evaluación
 - Estado del acceso de la información en la empresa.
 - Reglas del negocio.
 - Estado del entorno que soporta la toma de decisiones.
 - Justificación de decisiones.
 - Retorno de inversión.
2. Definición de Requerimientos
 - Requerimientos funcionales.
 - Requerimientos de información.
 - Requerimientos de seguridad y performance.
3. Análisis de Datos
 - Fuente de datos.
 - Calidad de datos.

- Limpieza de datos.
4. Prototipo de Aplicación
 - Objetivos de prototipos.
 - Participación de stakeholders.
 - Herramientas y métodos para el prototipo.
 5. Diseño de la Base de Datos
 - Demanda de reportes y consultas.
 - Consideraciones para el diseño de la base de datos.
 - Selección de un DBMS.
 6. Diseño del ETL
 - Selección de herramientas ETL.
 - Fases de ETL.
 - Flujo de procesos ETL.
 - Métricas de calidad.
 7. Desarrollo del ETL
 - Extracción desde sistemas fuente.
 - Herramientas ETL.
 - Flujo del desarrollo ETL.
 - Pruebas.
 8. Desarrollo de la Aplicación
 - Resultados del prototipo.
 - Herramientas de acceso y análisis de la información.
 - Capacitación de usuarios.
 - Alcance de la aplicación.
 9. Certificación
 - Elaboración de casos de prueba.
 - Pruebas de carga.
 - Pruebas de funcionalidades.
 10. Implementación
 - Componentes para producción.
 - Mantenimiento de la base de datos.
 - Capacitación y soporte a la solución. (p. 52)

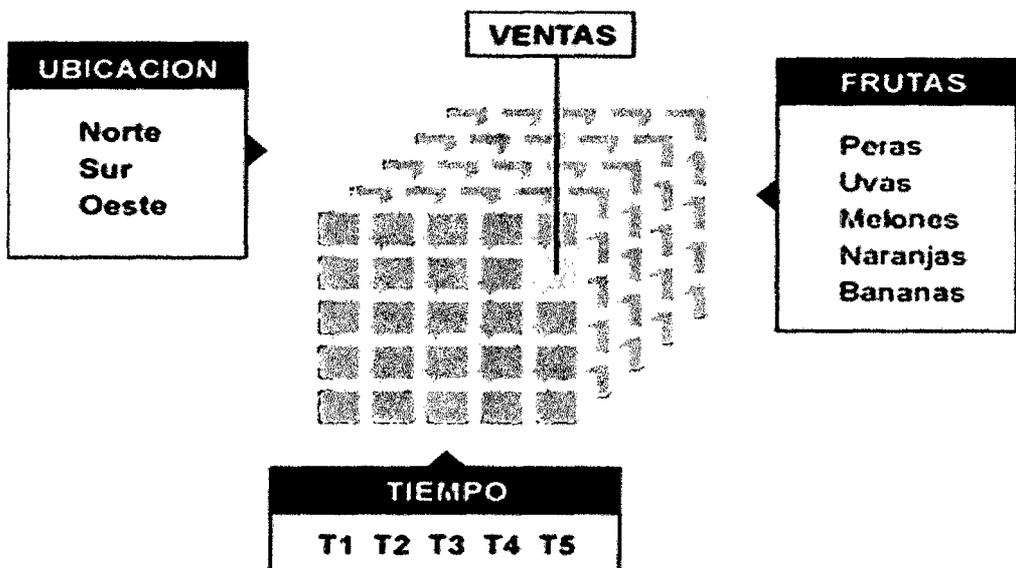
4.1.4 Cubos

Un cubo es una base de datos multidimensional, en la cual el almacenamiento físico de los datos se realiza en un vector multidimensional. Los cubos OLAP se pueden considerar como una ampliación de las dos dimensiones de una hoja de cálculo.

La propuesta de Codd consistía en realizar una disposición de los datos en vectores para permitir un análisis rápido. Estos vectores son llamados cubos. Disponer los datos en cubos evita una limitación de las bases de datos relacionales, que no son muy adecuadas para el análisis instantáneo de grandes cantidades de datos.

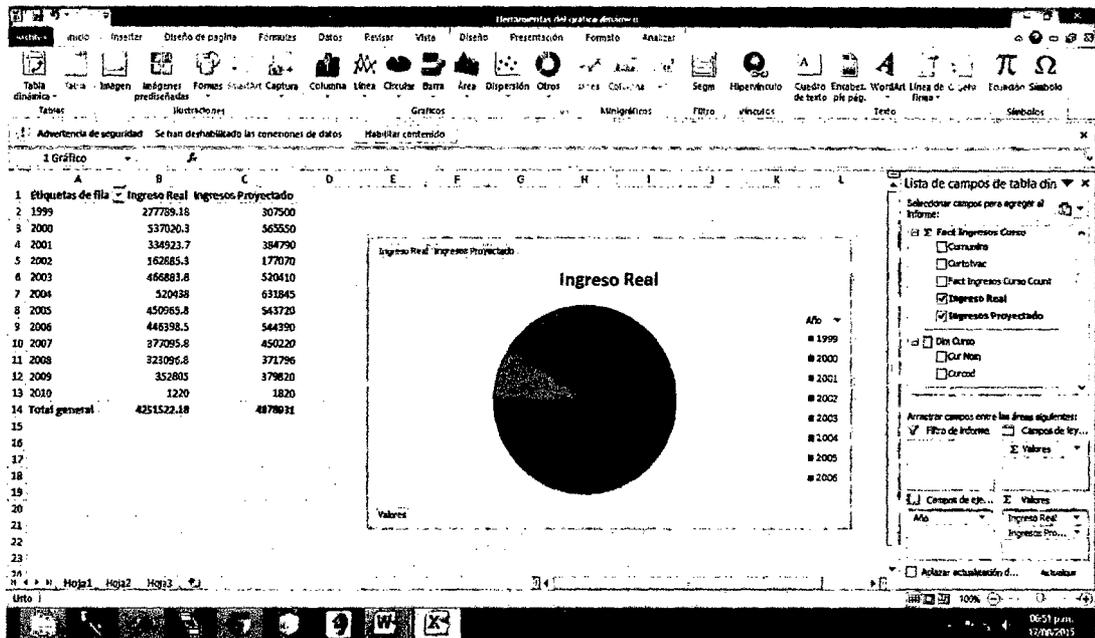
Los cubos son construidos con base en los hechos de negocio que resultan de interés para un cliente y que le permiten responder las preguntas de negocios.

FIGURA N° 4.9
EJEMPLO DE CUBO



Fuente: División de alta tecnología: Cibertec

GRAFICO N° 4.1 EJEMPLO DEL GRÁFICO DE UN CUBO



Fuente: Autor

4.1.5 Componentes de inteligencia de negocios

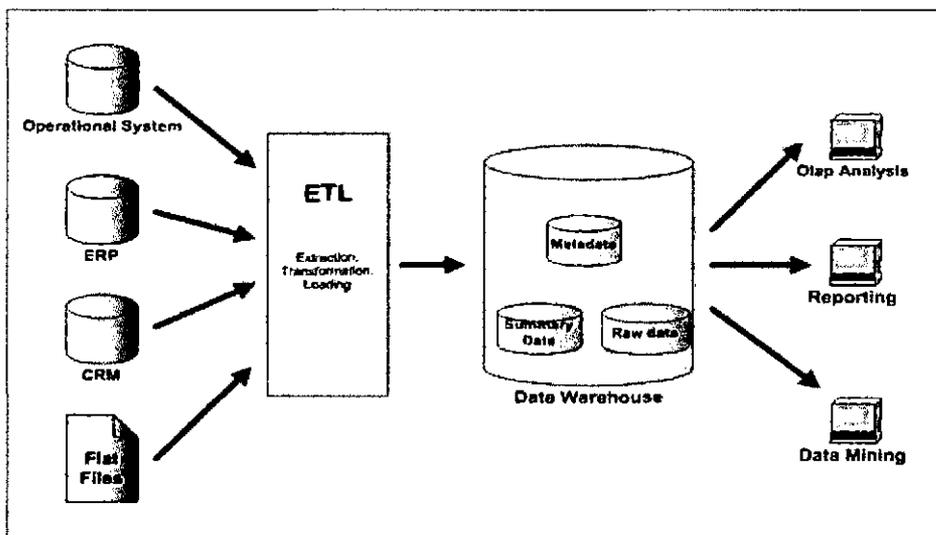
(BANESTO, 2014) se encontró lo siguiente:

Los componentes son:

- Fuentes de información, de las cuales partiremos para alimentar de información el *datawarehouse*.
- Proceso ETL de extracción, transformación y carga de los datos en el *datawarehouse*. Antes de almacenar los datos en un *datawarehouse*, éstos deben ser transformados, limpiados, filtrados y redefinidos. Normalmente, la información que tenemos en los sistemas transaccionales no está preparada para la toma de decisiones.
- El propio *datawarehouse* o almacén de datos, con el Metadata o Diccionario de datos. Se busca almacenar los datos de una forma que maximice su flexibilidad, facilidad de acceso y administración.

- El motor OLAP, que nos debe proveer capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos.
- Las herramientas de visualización, que nos permitirán el análisis y la navegación a través de los mismos. (p. 93)

FIGURA N° 4.10
COMPONENTES DE INTELIGENCIA DE NEGOCIOS



Fuente: Banesto, Business Intelligence: Compartir con información,
2014

La calidad de los datos en un *datawarehouse* es fundamental, como afirma Bill Inmon en su artículo 53 aparecido en "*Business Intelligence Network*" sobre calidad de Datos:

"Las organizaciones actúan bajo la suposición de que la información de la que disponen es precisa y válida. Si la información no es válida, entonces no pueden responder de las decisiones basadas en ella." (BANESTO, 2014, p. 98).

4.1.6 Proceso de extracción, transformación y carga (ETL)

El proceso trata de recuperar los datos de las fuentes de información y alimentar el *datawarehouse*. Dicho proceso consume entre el 60%

y el 80% del tiempo de un proyecto de Inteligencia de Negocios, por lo que es un proceso clave en la vida de todo proyecto.

El proceso ETL se divide en 5 subprocesos:

1. *Extracción*: Este proceso recupera los datos físicamente de las distintas fuentes de información.
2. *Limpieza*: Este proceso recupera los datos en bruto y comprueba su calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos y completa los valores vacíos, es decir se transforman los datos para reducir los errores de carga.
3. *Transformación*: Este proceso recupera los datos limpios y de alta calidad y los estructura y resume en los distintos modelos de análisis. El resultado de este proceso es la obtención de datos limpios y resumidos.
4. *Integración*: Este proceso valida que los datos que cargamos en el *datawarehouse* son consistentes con las definiciones y formatos del *datawarehouse*.
5. *Actualización*: Este proceso es el que nos permite añadir los nuevos datos al *datawarehouse*.

Integración de datos ETL

Las herramientas ETL, han sido la opción usual para alimentar el Datawarehouse.

- **Técnicas de integración de datos**

Entre las técnicas de integración de datos tenemos:

- ✓ Propagación de datos.- consiste en copiar datos de un lugar de origen a un entorno destino local. Los datos pueden extraerse del origen mediante programas que generen un fichero que debe ser transportado al destino. Una aproximación más eficiente es descargar sólo los datos que han cambiado en origen respecto a la última propagación.
- ✓ Consolidación de datos.- consiste en capturar los cambios realizados en múltiples entornos origen y propagarlos a un único entorno destino. Ejemplo: un datawarehouse.

- ✓ Federación de datos.- proporciona a las aplicaciones una visión lógica virtual común de una o más bases de datos.

- **Tecnologías de integración de datos**

Tenemos el ETL, que extrae datos del entorno origen, los transforma según nuestras necesidades de negocio para integración de datos y cargan estos datos en el entorno destino.

Entre los tipos de ETL, tenemos:

ETL de generación de código.- genera un programa de tercera generación. Ejemplo: cobol.

ETL basados en Motor.- crea flujos de trabajo en tiempo de ejecución definidos mediante herramientas gráficas.

ETL integrado en la Base de Datos (también incluye OLAP y Minería de Datos). Se clasifican a su vez en:

- ✓ ETL cooperativo.- Usan funciones avanzadas de la Base de Datos como: Procedimientos almacenados, SQL complejo.
- ✓ ETL complementarios.- Ofrecen funciones complementarias a los ETL comerciales.

- **Uso de la integración de datos**

Los procesos de integración de datos se usan en múltiples tipologías de proyectos. Podemos destacar los siguientes:

- ✓ Migración de datos
- ✓ Procesos de calidad de datos
- ✓ Corporate Performance Management (CPM)
- ✓ Master Data Management (MDM)
- ✓ Data Warehousing
- ✓ Inteligencia de negocios. Etc.

4.1.7 Procesamiento analítico en línea (OLAP)

Es una solución utilizada en el campo de la Inteligencia de negocios, cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

Para ello utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de datos (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

“Se entiende por OLAP, o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas o incluso informes” (CONESA/CURTO, 2011, p.94).

Codd lo define como el nombre dado al análisis dinámico de la Empresa, requerida para crear, manipular, animar y sintetizar información. Incluye la habilidad para discernir nuevas o no anticipadas relaciones entre las variables, la habilidad para identificar los parámetros necesarios para manejar grandes cantidades de datos, para crear un número ilimitado de dimensiones. Un análisis multidimensional representa datos como tablas, implica operaciones típicas: Mayor, menor, comparación entre periodos, porcentajes de diferencia, funciones estadísticas y financieras.

Comprende la consulta interactiva y el análisis de los datos.

Recupera y exhibe datos tabulares en 2 o 3 dimensiones, cuadros y gráficos.

Las estructuras OLAP permiten realiza preguntas que serían sumamente complejas mediante SQL.

Ejemplos de Aplicaciones OLAP:

- ✓ Número de unidades vendidas por línea y modelo, trimestres, años?.
- ✓ Porqué las Ventas de automóviles de determinado modelo disminuyen en el segundo trimestre?.
- ✓ Porqué la ganancia generada por los 10 mejores clientes decreció un 30% en relación al mismo periodo del año anterior?.

- **Tipos de OLAP**

Los tipos de OLAP difieren en cómo se guardan los datos:

- ✓ MOLAP.- son los cubos para sistemas pequeños de datos. Precalculan los datos.
- ✓ ROLAP.- Trabajan con las Bases de Datos relacionales. No se precaculan los datos.
- ✓ HOLAP.- los datos se dividen en almacenaje relacional y multidimensional.

- **Elementos OLAP**

OLAP permite el análisis multidimensional. Lo cual significa que la información está estructurada en ejes y celdas.

En el contexto OLAP existen diferentes elementos comunes a las diferentes tipologías OLAP:

- ✓ ESQUEMA.- colección de cubos, dimensiones, tablas de hecho y roles.
- ✓ CUBO.- es una colección de dimensiones asociadas a una tabla de hecho.
- ✓ TABLA DE HECHO, DIMENSION Y METRICA
- ✓ JERARQUIA.- conjunto de miembros organizados en niveles.
- ✓ NIVEL.- grupo de miembros en una jerarquía.
- ✓ MIEMBRO.- es un punto en la dimensión de un cubo que pertenece a un determinado nivel de una jerarquía.
- ✓ ROLES.- permisos asociados a un grupo de usuario.
- ✓ MDX.- es el lenguaje de consulta de estructuras OLAP.

4.2 Lenguaje estructurado de consultas (SQL)

Los orígenes del SQL están ligados a los de las bases de datos relacionales. En 1970 E. F. Codd propone el modelo relacional. Sin embargo, fue Oracle quien lo introdujo por primera vez en 1979 en un programa comercial. Hubo pues una demora entre la creación del modelo Relacional y su implementación computacional de aproximadamente 10 años, pues no había un software que soportará

el Modelo. Oracle SQL fue un gran éxito y dio lugar a toda una industria en torno a SQL. Sybase, Informix, Microsoft y otras compañías se presentaron con sus implementaciones de un sistema de gestión de base de datos relacional basado en SQL (RDBMS).

El SQL pasa a ser el lenguaje por excelencia de los diversos sistemas de gestión de bases de datos relacionales surgidos en los años siguientes y es por fin estandarizado en 1986 por el ANSI, dando lugar a la primera versión estándar de este lenguaje. Al año siguiente este estándar es también adoptado por la ISO.

Debido a que el objetivo principal de SQL es comunicar las acciones en el servidor de base de datos, no tiene la flexibilidad de un lenguaje de propósito general. La mayoría de la funcionalidad de SQL son preocupaciones entrada y salida de la base de datos: añadir, modificar, eliminar y leer datos.

En la actualidad el SQL es el estándar *de facto* de la inmensa mayoría de los SGBD comerciales.

El lenguaje SQL cumple con las reglas del Modelo Relacional, establecidas por el Dr. Codd, que son usados por el ORACLE y el SQL, que además tienen como antecedentes el Modelo Entidad Relación de Peter Chen.

El lenguaje SQL se puede considerar como una de las principales razones del éxito comercial de las bases de datos relacionales. Como se convirtió en un estándar para estas últimas, los usuarios perdieron el temor a migrar sus aplicaciones de base de datos desde otros tipos de sistemas de bases de datos a los sistemas relacionales.

El lenguaje SQL proporciona una interfaz de lenguaje declarativo del más alto nivel, por lo que el usuario sólo especifica lo que debe ser el resultado, dejando para el DBMS la optimización y las decisiones de cómo ejecutar la consulta.

El SQL tiene un rol muy importante en la comunicación con la base de datos, que también funciona embebido en otros lenguajes de uso

general, como los lenguajes procedimentales: Cobol, C, C++ y los lenguajes orientados a objetos como: Java, PHP y Python.

SQL es un lenguaje de consultas estructurado que opera sobre bases de datos relacionales y la normalización.

"Aunque el lenguaje SQL se considere un lenguaje de consultas, contiene muchas otras capacidades además de la consulta en bases de datos. Incluye características para definir la estructura de los datos, para la modificación de los datos en la base de datos y para la especificación de restricciones de seguridad." (SILBERSCHATZ, Abraham y KORTH, Henry, 2002, p.107).

SQL utiliza los términos *tabla*, *fila* y *columna* para los términos: *relación*, *tupla* y *atributo* del Modelo Relacional, respectivamente.

El principal comando de SQL para definir datos es la sentencia *create*, que se utiliza para crear esquemas, tablas y dominios, así como otras estructuras, como vistas, aserciones y triggers.

SQL es un lenguaje de bases de datos global: cuenta con sentencias para definir datos, consulta y actualizaciones. Se comporta como DDL (data definition language) y como DML (data manipulation language) y dispone de características para especificar temas de seguridad y autorización, definir restricciones de integridad y especificar controles de transacciones como el DCL (data control Language).

SQL, ha crecido a la par con los últimos adelantos tecnológicos como es inteligencia de negocios y a través de su interface con Visual Studio 2010, generan los cubos y los reportes que son el sustento de la Inteligencia de negocios.

En el SQL se puede generar el Datawarehouse que vendría a ser el input del Lenguaje de programación Visual Studio 2010, donde generaremos los cubos y reportes respectivos.

También es posible generar los reportes de los cubos en Excel, pero teniendo como entrada los cubos generados en Visual Studio 2010, por la gran facilidad que ofrece Excel para generar distintos tipos de gráficos como: diagramas de barras, diagramas circulares.

V. MATERIALES Y METODOS

5.1 Materiales

Los materiales usados los clasificamos en: software y hardware.

software

El Universo de esta investigación son las bases de datos en general, y hemos utilizado una base de datos "Ventas" creada para realizar este proyecto, y un Datawarehouse "dwhVentas" creado a partir de la Base de datos "Ventas". Estas 2 bases de datos creadas en SQL2012.

Usamos las siguientes "vistas" creadas en Visual Studio 2010:

Vista ventas.dsv

Vista DetalleVentas.dsv

Vista CompletoVentas.dsv

Asímismo creamos los siguientes "cubos" en Visual Studio 2010:

Cubo ventas

Cubo DetalleVentas

Cubo CompletoVentas

Las versiones de software que se utilizarán serán:

Para la creación de la Base de Datos y el Datawarehouse se utilizó el Microsoft SQL server 2012.

Para la creación de Vistas y Cubos se utilizó el Visual Studio 2010.

hardware

Se utilizó una PC con las siguientes características:

- ✓ Procesador Intel core i5
- ✓ Memoria RAM de 8 GB
- ✓ Sistema operativo de 64 bits
- ✓ Disco duro de 1 TB.

5.2 Población de la investigación y la muestra correspondiente

El universo de esta investigación son las Bases de Datos Multidimensionales en general.

Utilizaremos como muestra una base de datos multidimensional correspondiente a una empresa peruana. Creada a partir de Bases de Datos operacionales. Por ser una investigación cualitativa no hay tipo de muestreo. El tipo de muestra es no probabilística.

5.3 Técnicas, procedimientos e instrumentos de Recolección de datos

Las técnicas de recopilación de datos utilizada es la **Observación** del comportamiento de una base de datos multidimensional.

En la Base de Datos creada, se comprobó de manera práctica, las características de ésta y se convirtió en un Datawarehouse, con el fin de realizar un análisis multidimensional.

Para la demostración de la hipótesis se aplicó las herramientas de Inteligencia de Negocios y SQL server.

Adicionalmente, también se realizaron cuadros de contrastación de resultados.

5.4 Técnicas de análisis o métodos estadísticos aplicados

Para la conversión de la información colectada en datos elaborados, se han usado varios pasos:

- a) Conversión de la base de datos Ventas a un Datawarehouse, usando el SQL.
- b) Transformación del datawarehouse Ventas a diferentes vistas. A través del visual studio 2010.
- c) Transformación del datawarehouse Ventas a diferentes cubos. A través del visual studio 2010.
- d) Obtener reportes gráficos de los cubos obtenidos, en diferentes tipos de gráficos, usando el visual studio 2010 y el Excel (entorno gráfico).

La presente investigación no utiliza métodos estadísticos por ser una investigación cualitativa.

5.5 Metodología

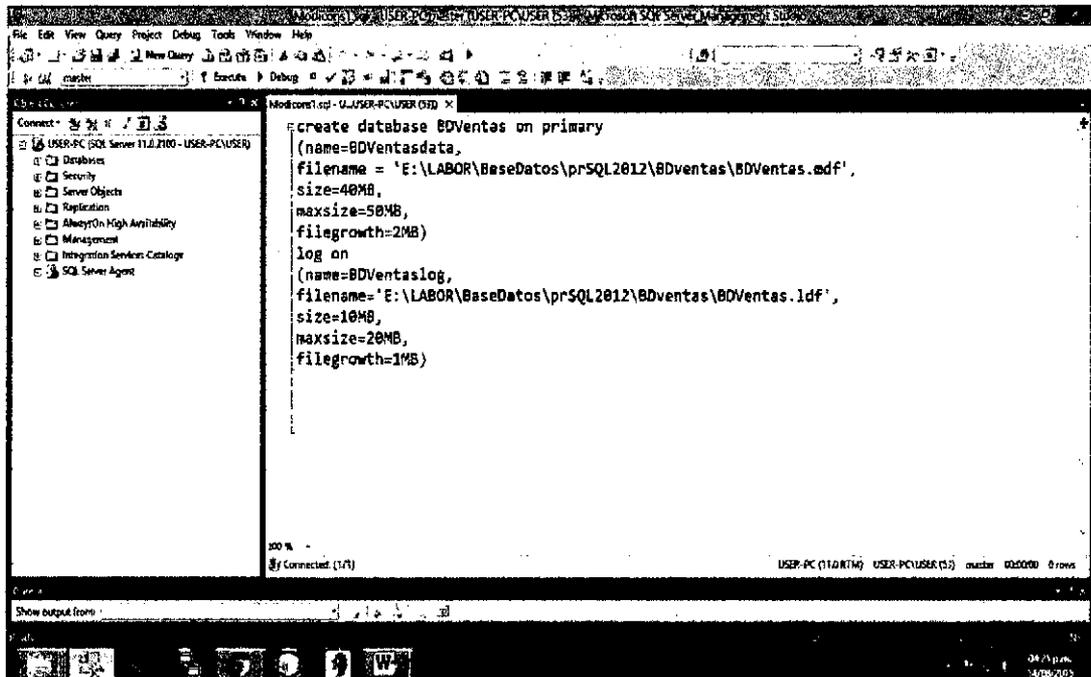
En cuanto al método utilizado es el Método inductivo, ya que hemos hecho una transición de lo particular a lo general, se ha usado la generalización inductiva, porque la investigación se basa en hechos reales, para lograr la evidencia.

VI. RESULTADOS

6.1 Creación de la base de datos Ventas

A continuación mostraremos el código para crear la Base de Datos y las tablas de la Base de Datos Ventas en SQL 2012.

FIGURA N° 6.1
CREACIÓN DE LA BASE DE DATOS VENTAS



Fuente: Autor

FIGURA N° 6.2
CREA LA ESTRUCTURA DE BDVENTAS

CREATE TABLE CLIENTE (
ClienteID	nchar(5) NOT NULL
PRIMARY KEY,	
NombreCia	nvarchar(40) NOT NULL,
NombreContacto	nvarchar(30),
TituloContacto	nvarchar(30),
Direccion	nvarchar(60),
Ciudad	nvarchar(15),
Pais	nvarchar(15),
Telefono	nvarchar(24),

```

CREATE TABLE EMPLEADO (
    EmpleadoID          int NOT NULL PRIMARY
    KEY,
    Apellido            nvarchar(20) NOT NULL,
    Nombre              nvarchar(10) NOT NULL,
    Titulo              nvarchar(30),
    Fechanacim         datetime,
    Direccion           nvarchar(60),
    Ciudad              nvarchar(15),
    Pais                nvarchar(15),
    Telefono            nvarchar(24),
)

```

```

CREATE TABLE CATEGORIA(
    Categoriad          int NOT NULL PRIMARY
    KEY,
    NombreCategoria    nvarchar(15) NOT NULL,
    Descripcion         ntext,
)

```

```

CREATE TABLE PRODUCTO(
    ProductID          int NOT NULL PRIMARY
    KEY,
    Nombreproducto     nvarchar(40) NOT NULL,
    ProveedorID        int,
    Categoriad         int,
    CantidadPorUnidad  nvarchar(20),
    PrecioUnitario     money,
    UnidadesStock      smallint,
    descontinuado      bit,
)

```

```

CREATE TABLE ORDEN(
    OrdenID            int NOT NULL PRIMARY
    KEY,
    ClienteID          nchar(5),
    EmpleadoID         int,
    FechaOrden         datetime,
    FechaEmbarque      datetime,
    ViaEmbarque        int,
    Flete              money,
    DirEmbarque        nvarchar(60),
    CiudadEmbarque     nvarchar(15),
    PaisEmbarque       nvarchar(15),
)

```

```

CREATE TABLE Detalle_Orden(
    OrdenID int NOT NULL,
    ProductID int NOT NULL,
    PrecioUnitario money,
    Cantidad smallint,
    Descuento real,
    PRIMARY KEY (OrdenID,ProductID)
)

```

```

CREATE TABLE PROVEEDOR(
    ProveedorID int NOT NULL
    PRIMARY KEY,
    NombreCia nvarchar(40),
    NombreContacto nvarchar(30),
    TituloContacto nvarchar(30),
    Direccion nvarchar(60),
    Ciudad nvarchar(15),
    Pais nvarchar(15),
    Telefono nvarchar(24),
    Fax nvarchar(24),
)

```

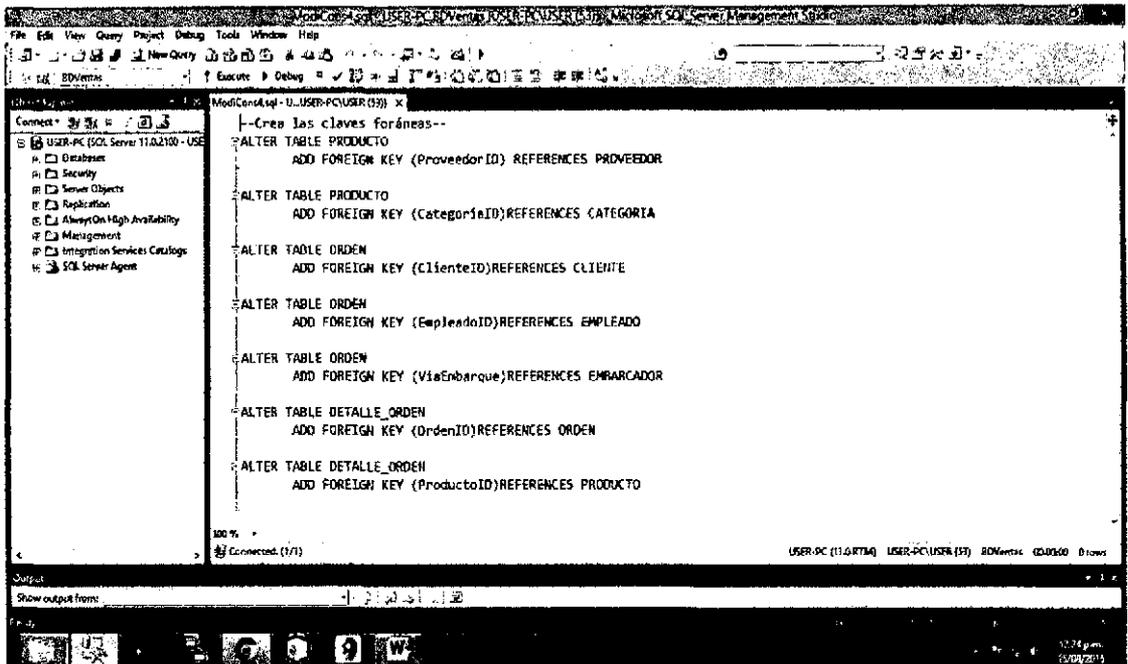
```

CREATE TABLE EMBARCADOR(
    EmbarcadorID int NOT NULL
    PRIMARY KEY,
    NombreCia nvarchar(40),
    Telefono nvarchar(24),
)

```

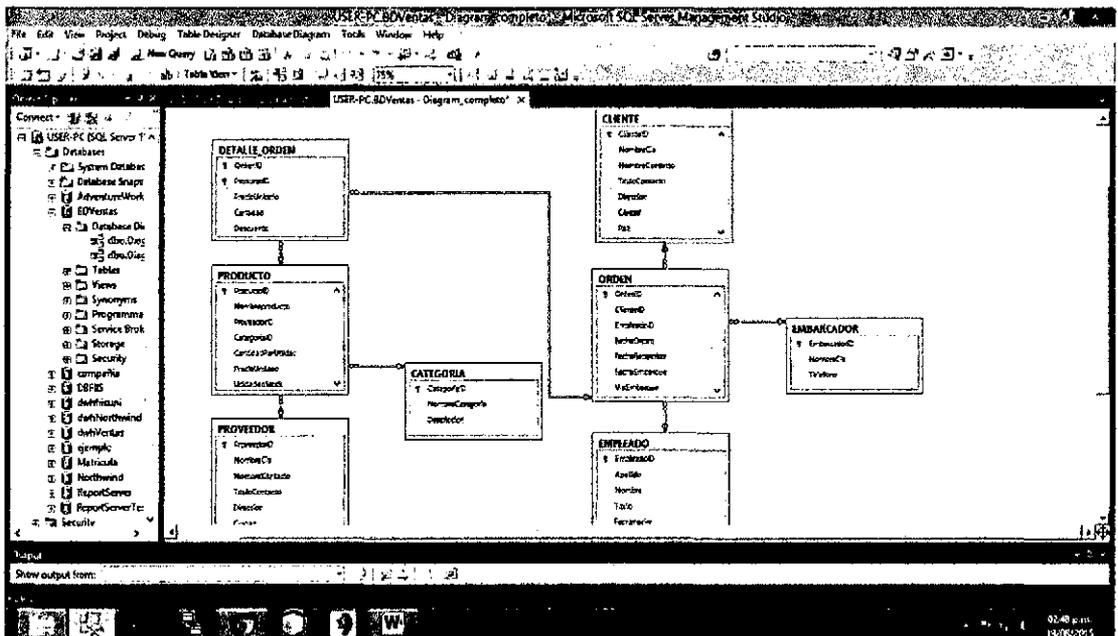
Fuente: Autor

FIGURA N° 6.3
CREA LAS CLAVES FORANEAS



Fuente: Autor

FIGURA N° 6.4
DIAGRAMA DE VENTAS



Fuente: Autor

6.2 Creación del datawarehouse

La creación del datawarehouse se realizó en el SQL 2012, teniendo como entrada la BDVentas. El datawarehouse dwhVentas tiene como tabla de hechos a la tabla: factVentas y varias tablas de dimensiones: Transportista, Cliente, Empleado, Fecha.

FIGURA N° 6.5
CREA EL DATAWAREHOUSE dwhVentas

```
/* CREANDO EL DATAWAREHOUSE */
create database dwhVentas

use dwhVentas

/*CREANDO LAS TABLAS DE DIMENSIONES*/

create table dimTransportista(
transpID int primary key,
nombreTransp nvarchar(80) )
go

create table dimCliente(
clienteID nchar(10) primary key,
companyname nvarchar(80),
city nvarchar(30),
country nvarchar(30) )
go

create table dimEmpleado(
empleadoID int primary key,
companyname nvarchar(80),
city nvarchar(30),
country nvarchar(30) )

create table dimFecha(
fechaID int primary key,
fecha datetime,
año smallint,
trimestre tinyint,
mes tinyint,
nomMes varchar(20),
dia tinyint,
diaSem tinyint,
nomDiaSem varchar(15) )
```

```
/*CREANDO LA TABLA DE HECHOS*/
```

```
create table factVentas(  
orderid int primary key,  
clienteID nchar(10) references dimCliente,  
empleadoID int references dimEmpleado,  
tranpID int references dimTransportista,  
fechID int references dimFecha,  
importe money,  
freight money )  
go
```

Fuente: Autor

6.3 Proceso de extracción, transformación y carga (ETL)

En el proceso ETL, se utilizó la sentencia INSERT para el cargado de datos a las diferentes tablas de dimensiones, provenientes de la BDVentas.

En algunos casos se usaron funciones para adecuarse a la estructura de las tablas de dimensiones. Ejemplo: la función fn_NomMes que convierte el número de mes al nombre del mes correspondiente.

También tenemos la función fn_NomDiaSem que convierte el número de día de la semana al nombre del correspondiente día.

FIGURA N° 6.6
REALIZA EL PROCESO ETL

```
use dwhVentas
go

/* INSERTANDO DATOS A LAS TABLAS DE DIMENSIONES
PROVENIENTES DE LA BDventas*/
insert dimTransportista
select EmbarcadorID,NombreCia from BDVentas.dbo.EMBARCADOR
go

insert dimEmpleado
select EmpleadoID,Apellido+', '+Nombre,Ciudad,Pais
from BDVentas.dbo.EMPLEADO
go

insert dimCliente
select ClienteID,NombreCia,Ciudad,Pais
from BDVentas.dbo.CLIENTE
go

/*CREANDO FUNCIONES PARA LA TRANSFORMACIÓN DE LOS DATOS*/

create function fn_NomMes(@mm tinyint)returns varchar(20)
begin
    declare @mes table(codigo int, nombre varchar(20))
    insert @mes
values(1,'Enero'),(2,'Febrero'),(3,'Marzo'),
(4,'Abril'),(5,'Mayo'),(6,'Junio'),(7,'Julio'),(8,'Agosto'
),(9,'Setiembre'),(10,'Octubre'),(11,'Noviembre'),(12,'Dic
iembre')
    return(select nombre from @mes where codigo=@mm)
    end
    go

create function fn_NomDiaSem(@dd tinyint)returns
varchar(15)
begin
    declare @semana table(codigo int, nombre varchar(15))
    insert @semana
values(1,'Domingo'),(2,'Lunes'),(3,'Martes'),
(4,'Miércoles'),(5,'Jueves'),(6,'Viernes'),(7,'Sábado')
    return(select nombre from @semana where codigo=@dd )
    end
```

```
/* INSERTANDO DATOS USANDO LAS FUNCIONES ANTERIORES*/
```

```
insert dimFecha
select distinct
year(FechaOrden)*10000+month(FechaOrden)*100+day(FechaOrden),
FechaOrden,
year(FechaOrden), datepart(QUARTER, FechaOrden),
month(FechaOrden), dbo.fn_NomMes(month(FechaOrden)), day(FechaOrden),
DATEPART(WEEKDAY, FechaOrden), dbo.fn_NomDiaSem(DATEPART(WEEKDAY, FechaOrden))
from BDVentas.dbo.ORDEN
go
```

```
insert factVentas
select o.OrdenID, o.ClienteID, o.EmpleadoID, o.ViaEmbarque,
year(FechaOrden)*10000+month(FechaOrden)*100+day(FechaOrden),
sum(od.PrecioUnitario*od.Cantidad*(1-od.Descuento)) as
Importe, o.Flete
from BDVentas.dbo.ORDEN as o
join BDVentas.dbo.[DETALLE_ORDEN] as od on
o.OrdenID=od.OrdenID
group by
o.OrdenID, o.ClienteID, o.EmpleadoID, o.FechaOrden, o.Flete,
o.ViaEmbarque
go
```

```
/* CREANDO LAS DIMENSIONES FALTANTES */
use dwhVentas
go
```

```
create table dimCategoria(
categoryID int primary key,
categoryname nvarchar(30) )
go
```

```
create table dimProducto(
productID int primary key,
categoryID int references dimCategoria,
productname nvarchar(80),
unitprice money )
go
```

```

create table factVtaProducto(
orderID int references factVentas,
productID int references dimProducto,
quantity smallint,
discount real,
importe money,
constraint pk_factVtaProd primary
key(orderID,productID) )
go

/* ETL detalle ventas */

insert dimCategoria
select CategoriaID,NombreCategoria from
BDVentas.dbo.CATEGORIA
go

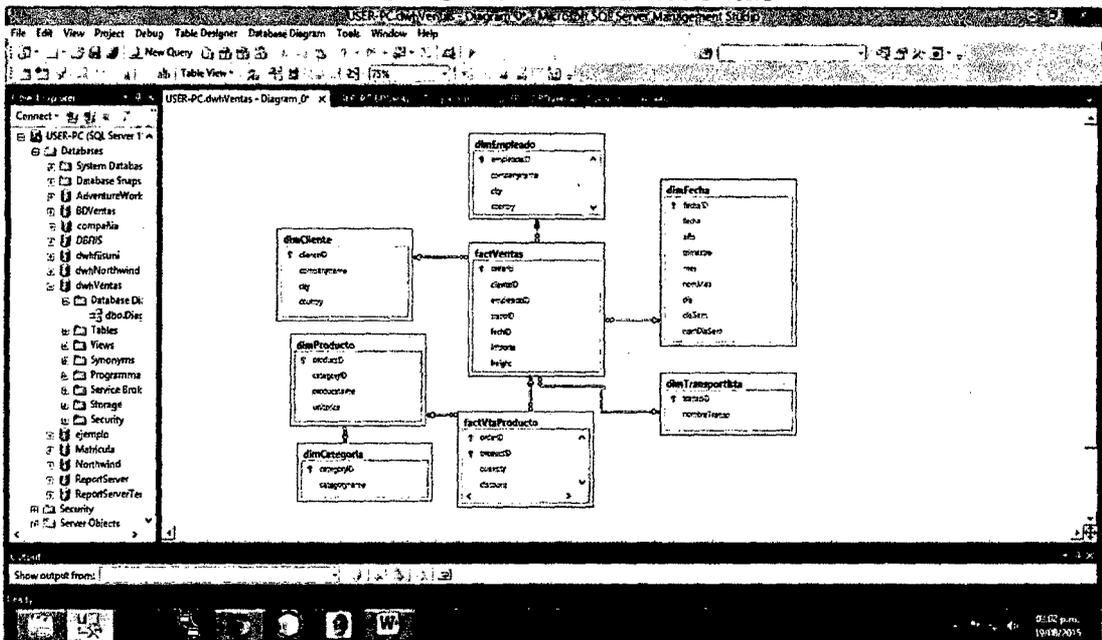
insert dimProducto
select
ProductoID,CategoriaID,Nombreproducto,PrecioUnitario
from BDVentas.dbo.PRODUCTO
go

insert factVtaProducto
select OrdenID,ProductoID,Cantidad,Descuento,
PrecioUnitario*Cantidad*(1-Descuento)as Importe
from BDVentas.dbo.[DETALLE_ORDEN]
go

```

Fuente: Autor

FIGURA N° 6.7
DIAGRAMA DE dwhVentas

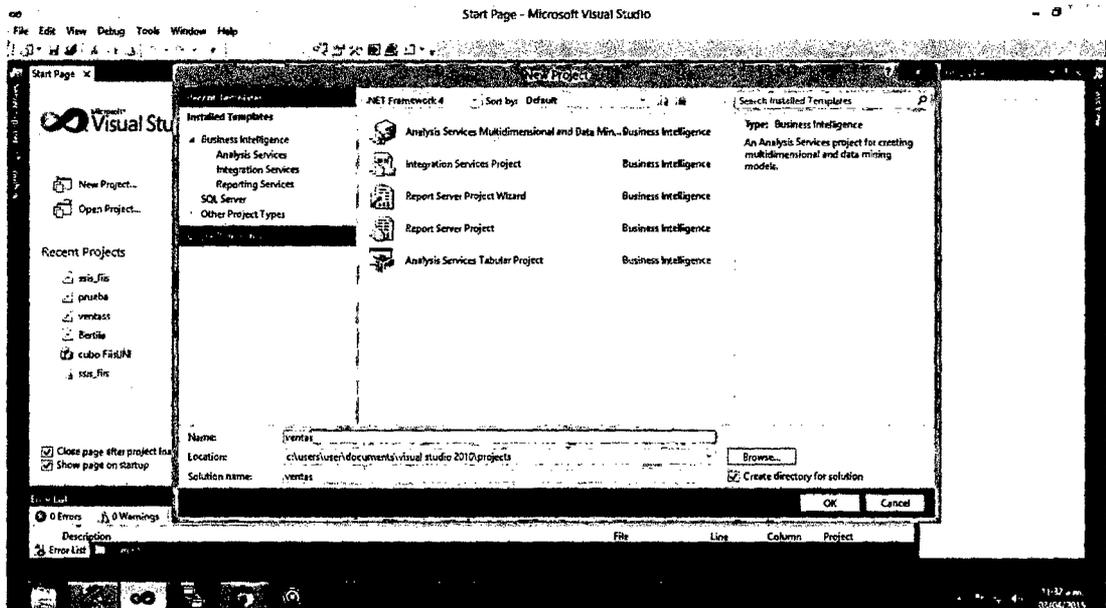


Fuente: Autor

6.4 Creación de Vistas

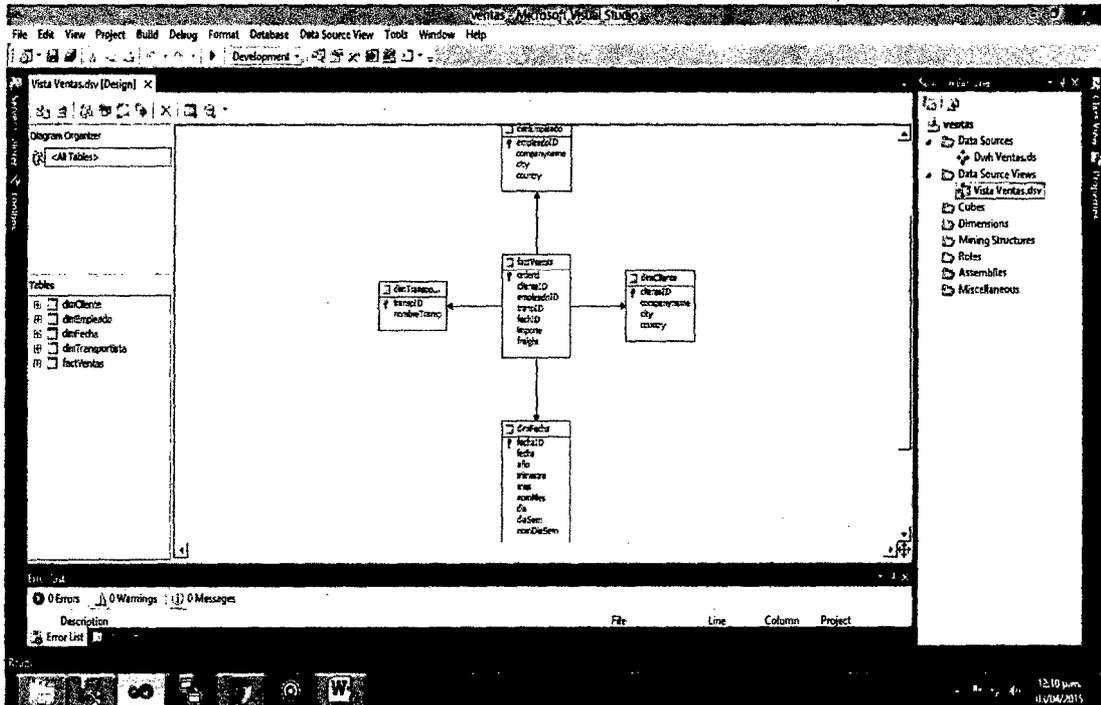
La generación de vistas es un paso previo a la creación de cubos.

FIGURA N° 6.8
CREACION DEL PROYECTO MULTIDIMENSIONAL VENTAS



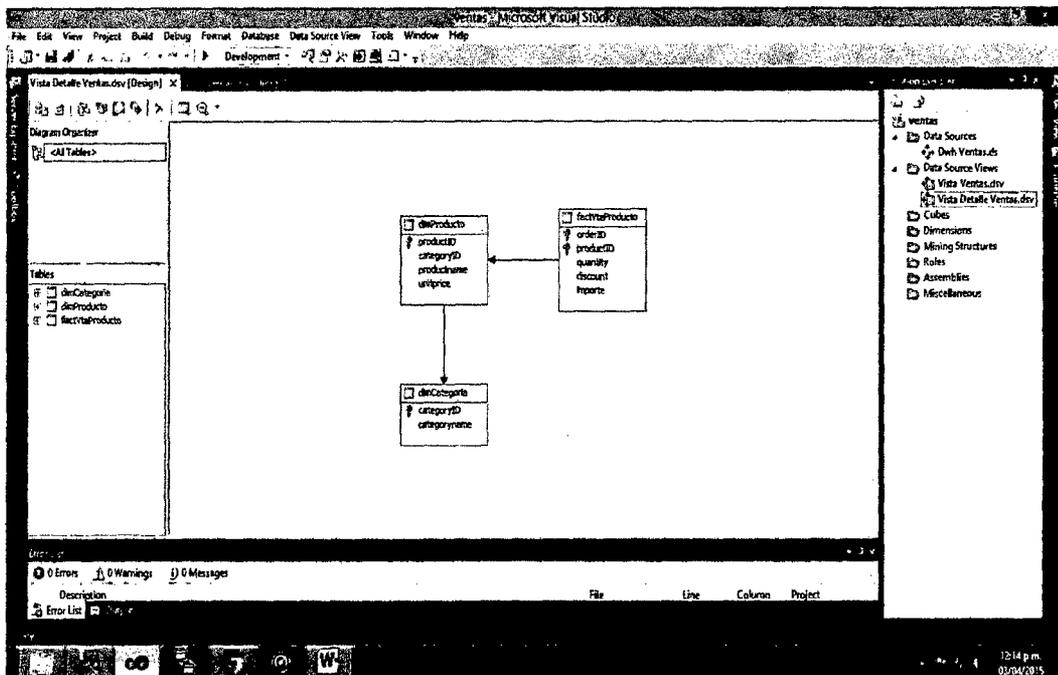
Fuente: Autor

FIGURA N° 6.9
 CREACIÓN DE LA VISTA VENTAS



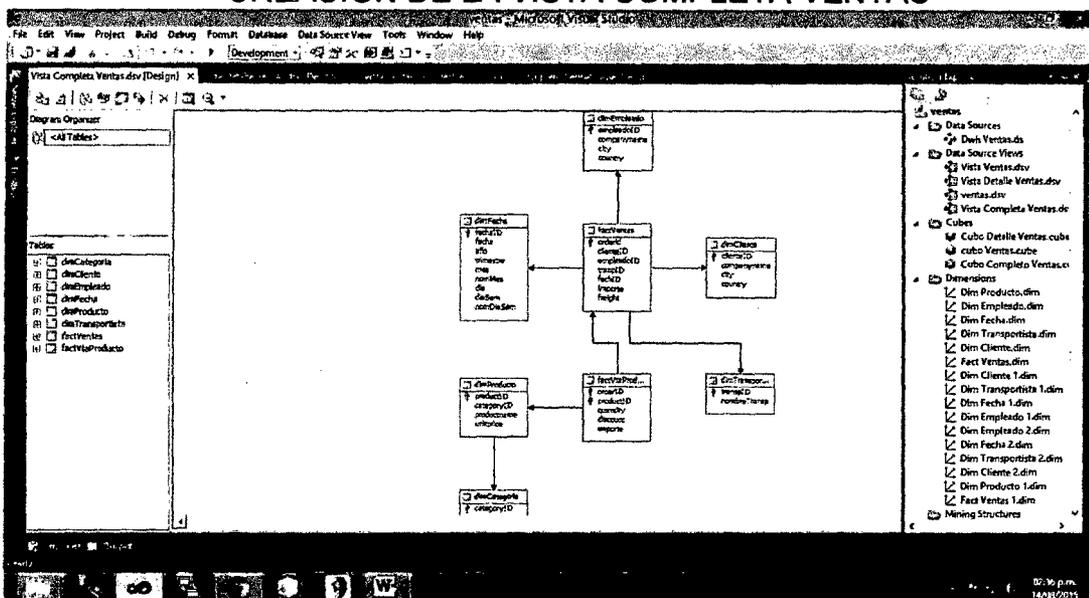
Fuente: Autor

FIGURA N° 6.10
 CREACION DE LA VISTA DETALLE VENTAS



Fuente: Autor

FIGURA N° 6.11
CREACION DE LA VISTA COMPLETA VENTAS



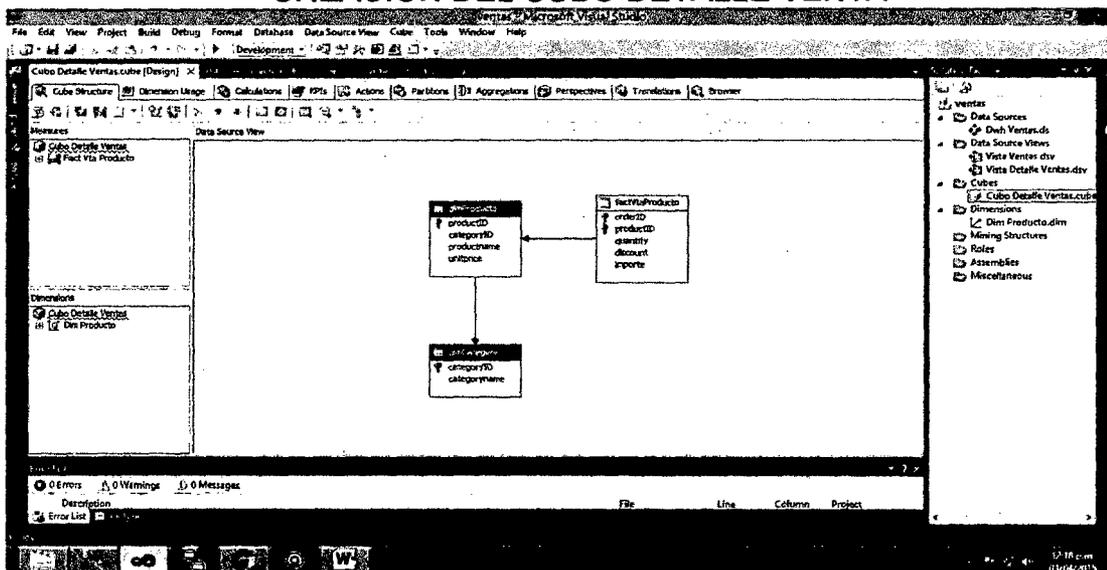
Fuente: Autor

6.5 Creación de Cubos

Los cubos permiten hacer resúmenes estadísticos de una, dos o más dimensiones.

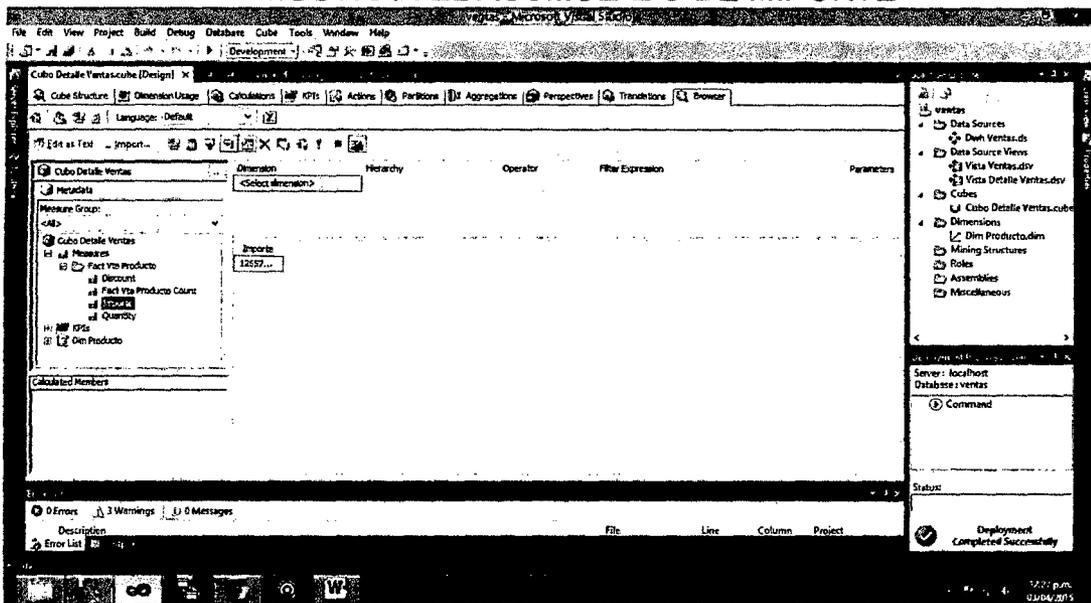
6.5.1 Creación del cubo detalle venta

FIGURA N° 6.12
CREACION DEL CUBO DETALLE VENTA



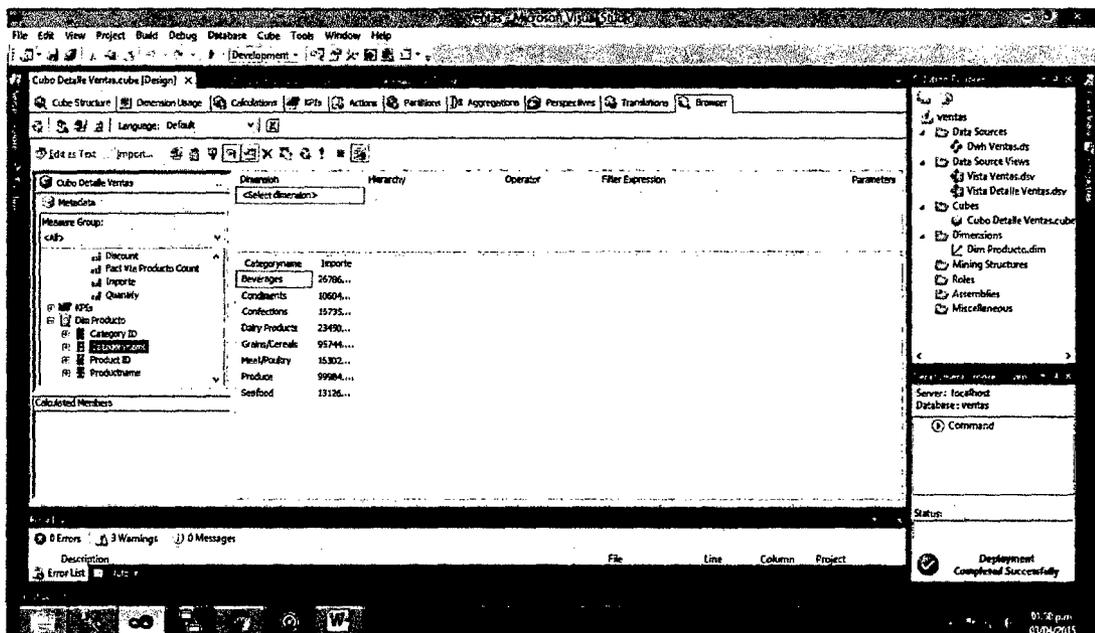
Fuente: Autor

FIGURA N° 6.13
 MOSTRAR EL ACUMULADO DE IMPORTE



Fuente: Autor

FIGURA N° 6.14
 MOSTRAR EL IMPORTE POR CATEGORIAS

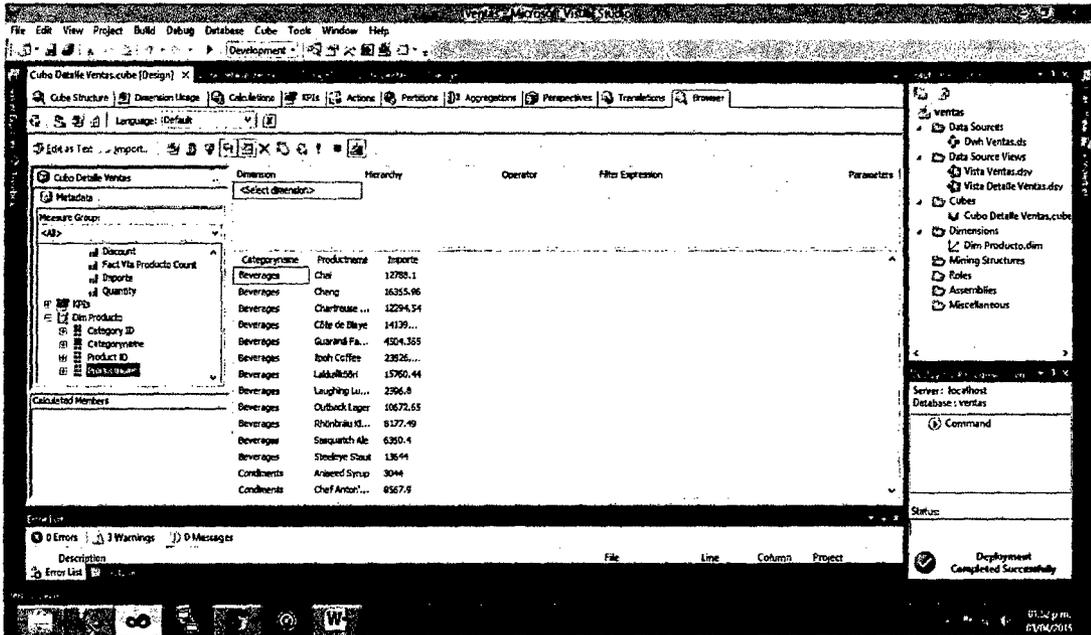


Fuente: Autor

Handwritten signature

FIGURA N° 6.15

MOSTRAR EL IMPORTE POR CATEGORIA Y PRODUCTO

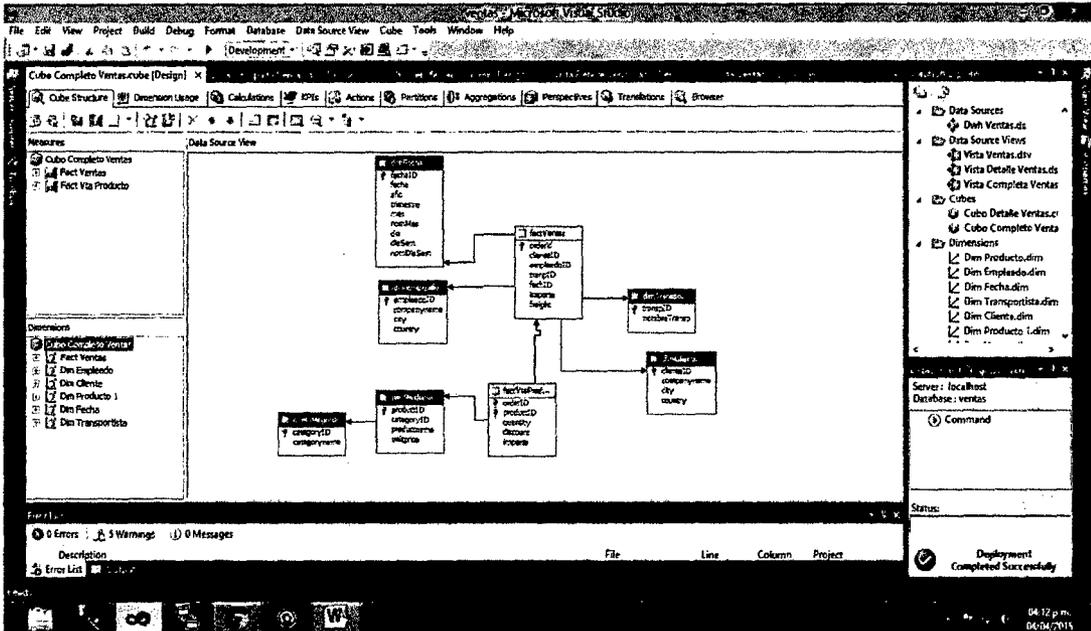


Fuente: Autor

6.5.2 Creación del cubo completo de ventas

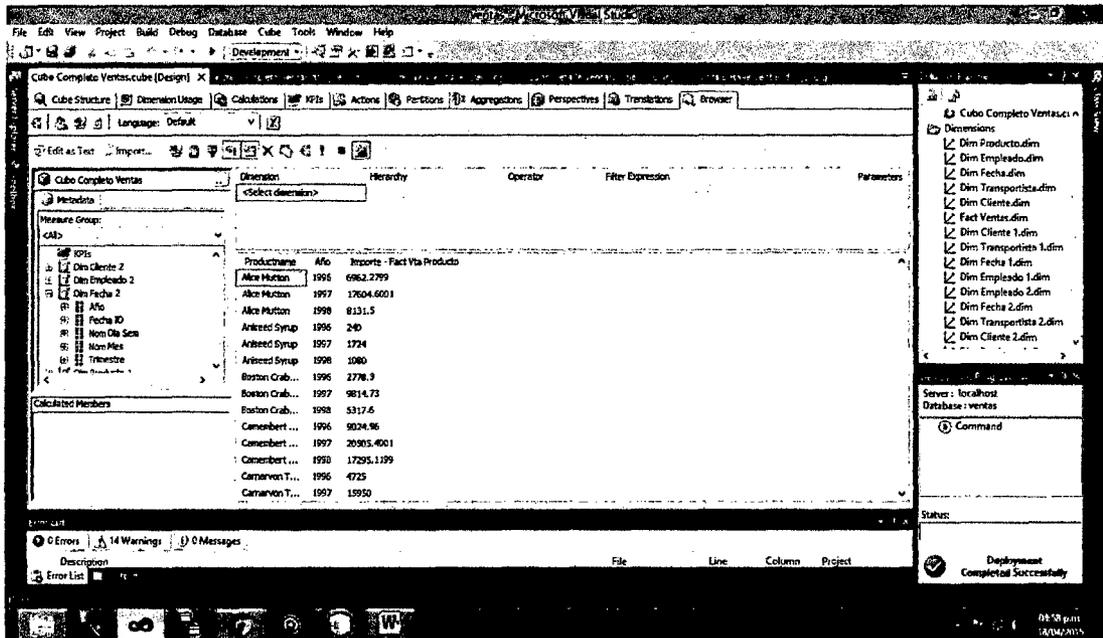
FIGURA N° 6.16

CREACION DEL CUBO COMPLETO DE VENTAS



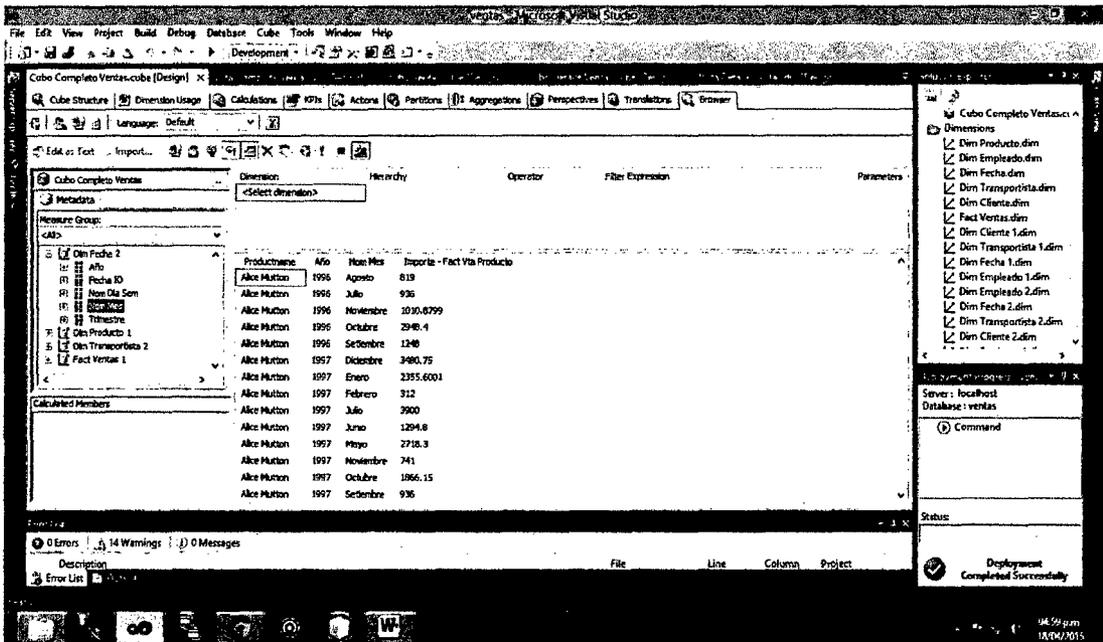
Fuente: Autor

FIGURA N° 6.17
 MOSTRAR EL IMPORTE DE LA FACTURA POR PRODUCTO Y
 POR AÑO



Fuente: Autor

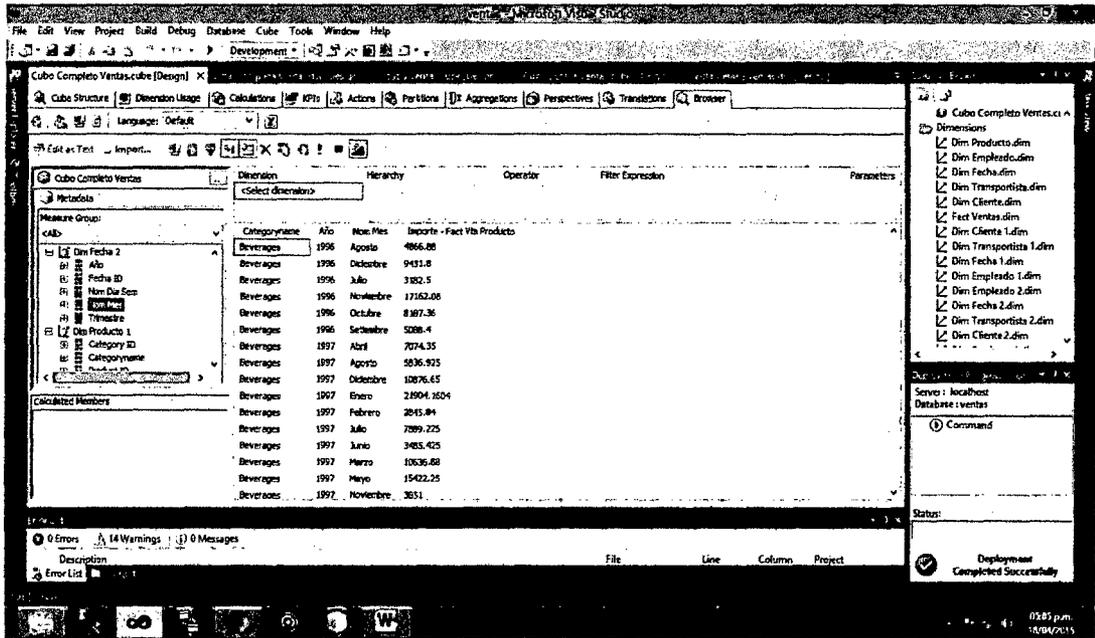
FIGURA N° 6.18
 MOSTRAR EL IMPORTE DE LA FACTURACION POR
 PRODUCTO, POR AÑO Y POR NOMBRE DE MES



Fuente: Autor

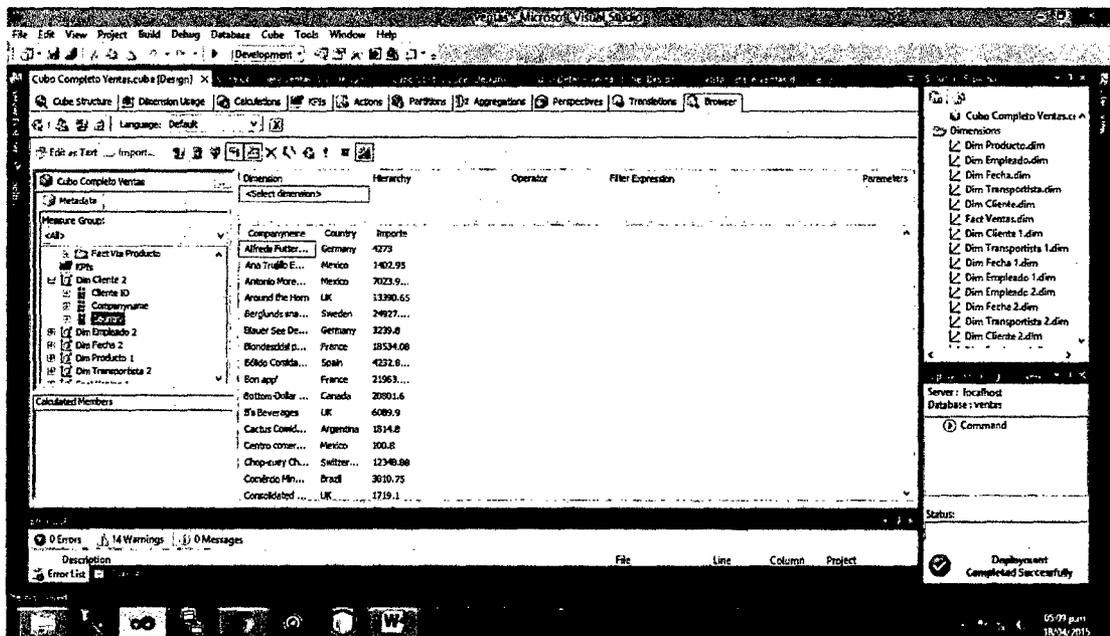
RUP

FIGURA N° 6.19
 MOSTRAR EL IMPORTE DE LA FACTURACION POR
 CATEGORIA, POR AÑO Y POR NOMBRE DE MES



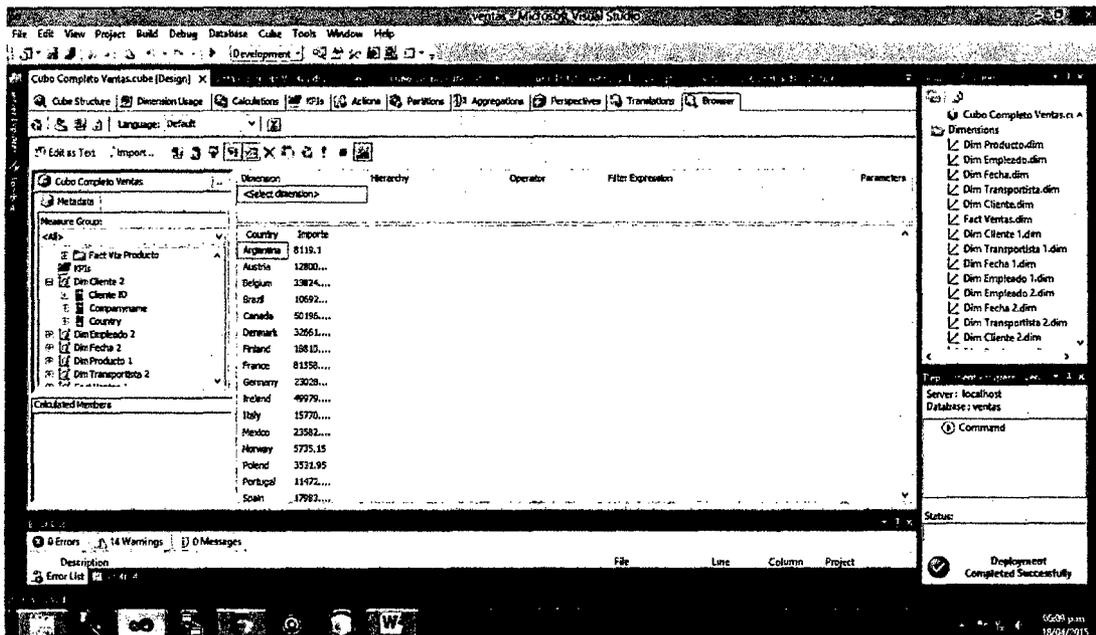
Fuente: Autor

FIGURA N° 6.20
 MOSTRAR EL IMPORTE POR COMPAÑÍA Y PAIS



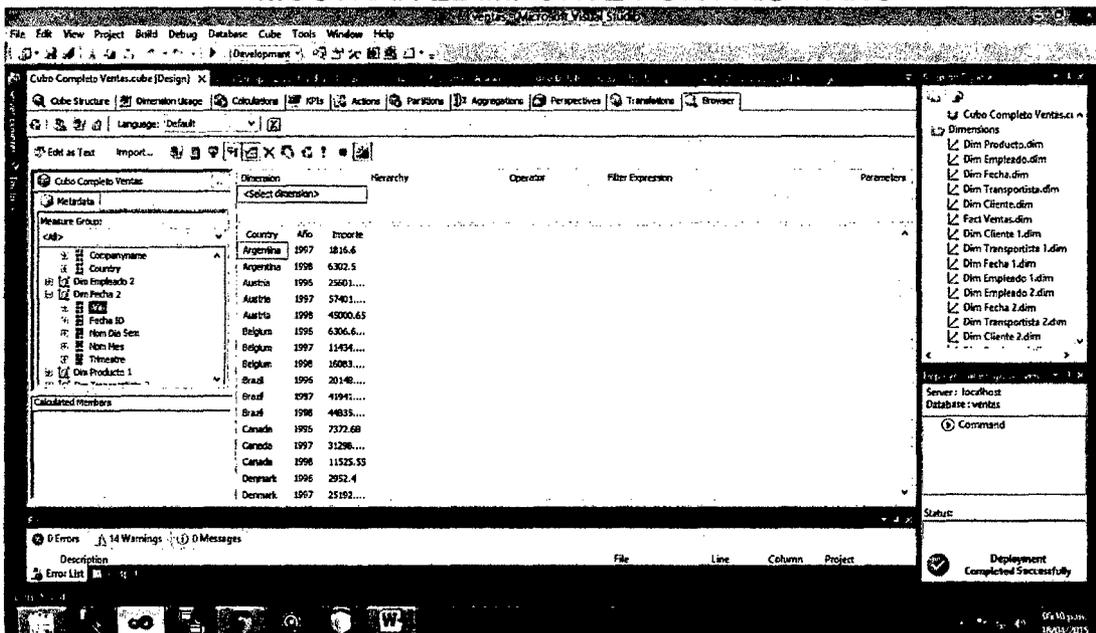
Fuente: Autor

FIGURA N° 6.21
 MOSTRAR EL IMPORTE POR PAIS



Fuente: Autor

FIGURA N° 6.22
 MOSTRAR EL IMPORTE POR PAIS Y AÑO

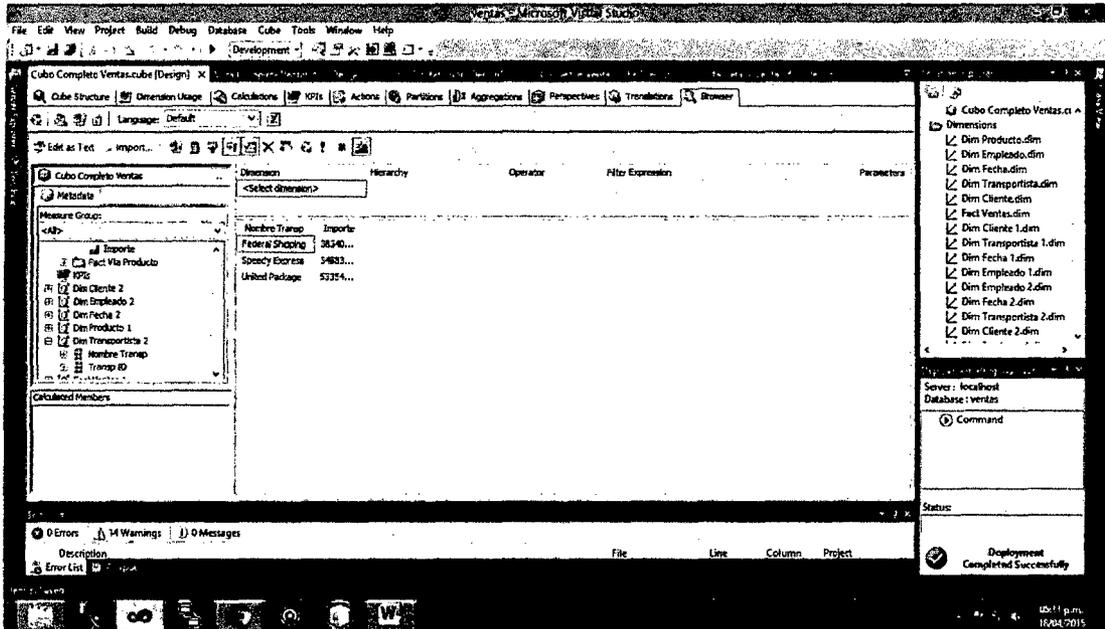


Fuente: Autor

Rese

FIGURA N° 6.23

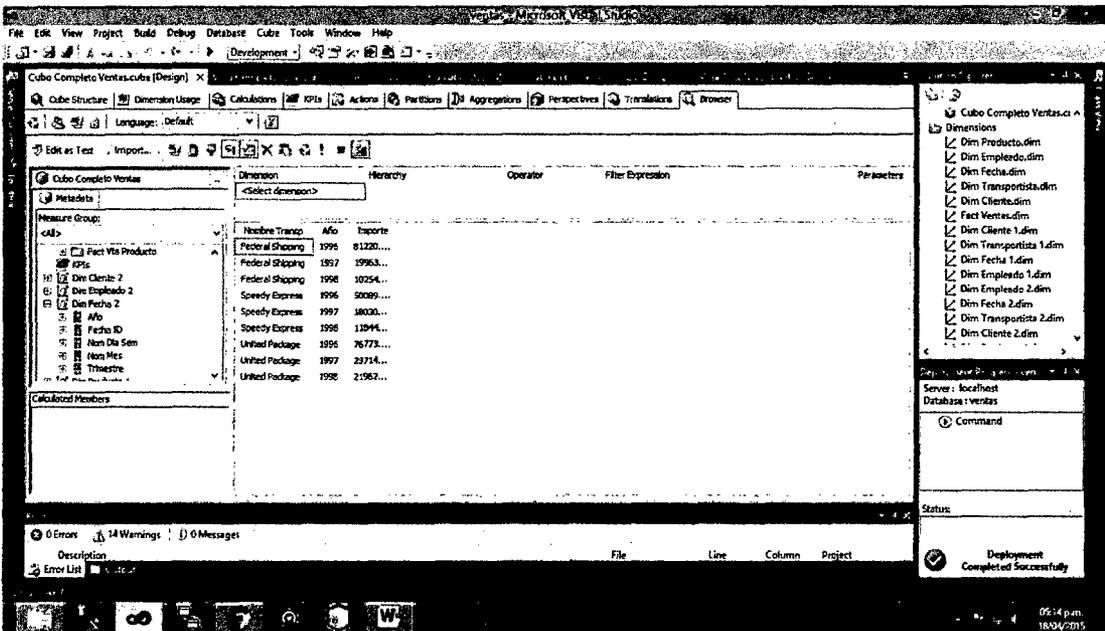
MOSTRAR EL IMPORTE POR TRANSPORTISTA



Fuente: Autor

FIGURA N° 6.24

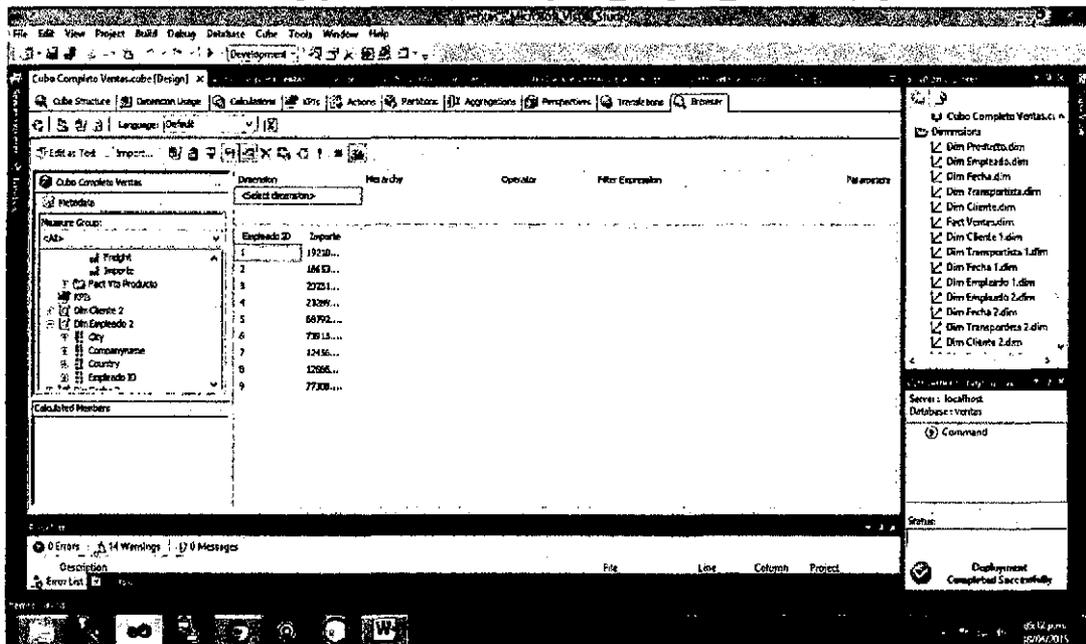
MOSTRAR EL IMPORTE POR TRANSPORTISTA Y AÑO



Fuente: Autor

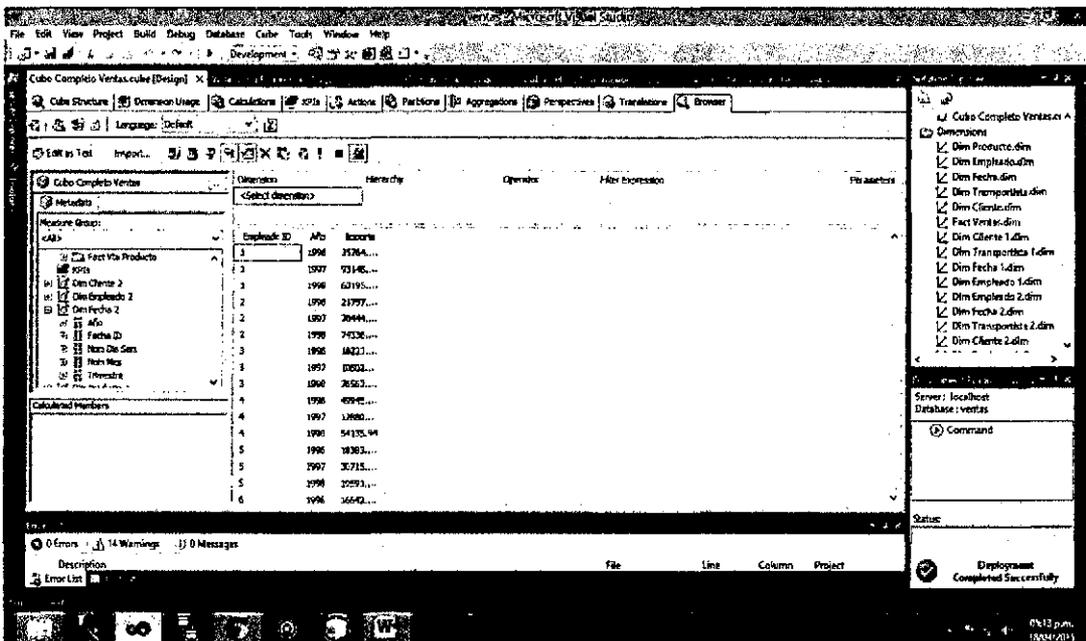
Handwritten signature

FIGURA N° 6.25
 MOSTRAR EL IMPORTE POR EMPLEADO



Fuente: Autor

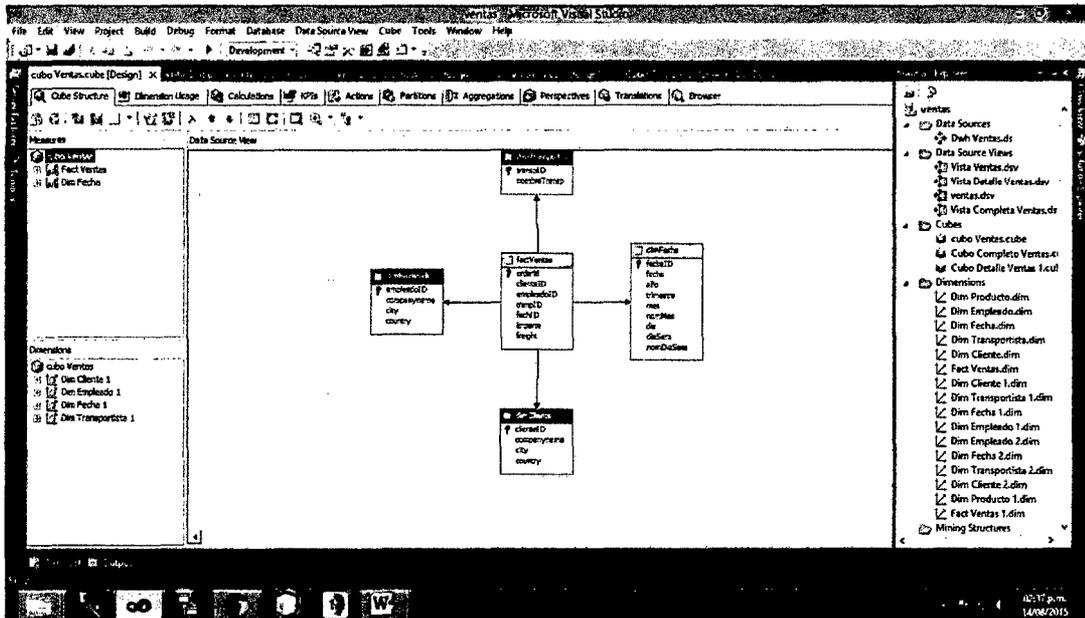
FIGURA N° 6.26
 MOSTRAR EL IMPORTE POR EMPLEADO Y POR AÑO



Fuente: Autor

6.5.3 Creacion del cubo ventas

FIGURA N° 6.27
CREACION DEL CUBO VENTAS



Fuente: Autor

6.6 Creación de cubos en Visual Studio 2010 y sus respectivos gráficos en Excel

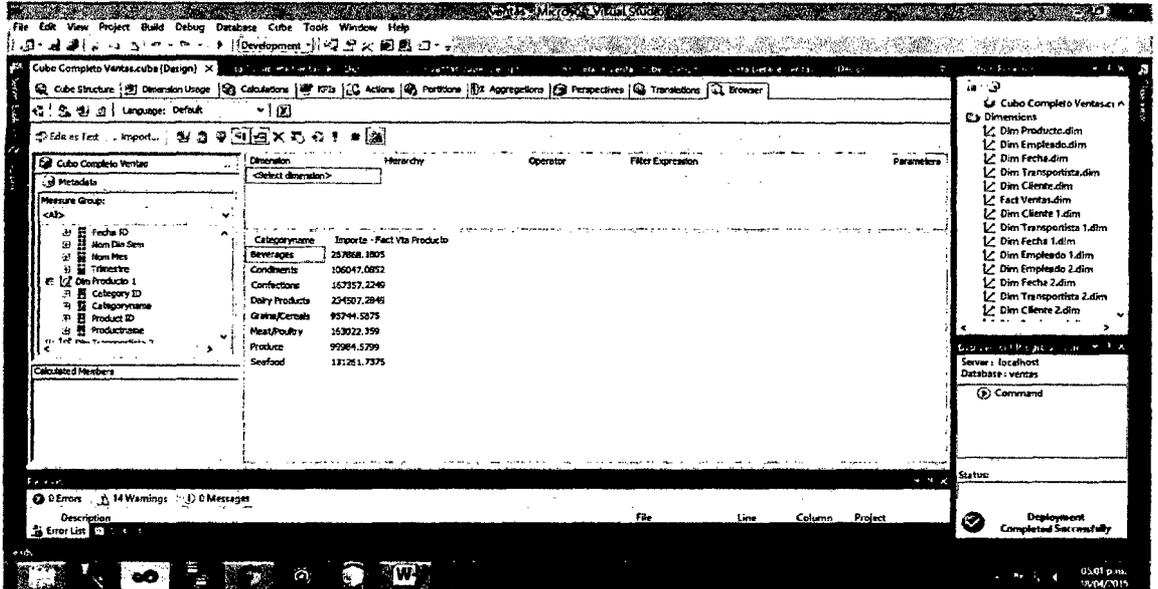
A continuación mostramos los cubos creados y su respectiva representación gráfica en Excel.

Handwritten signature

6.6.1 Por categoría e importe

FIGURA N° 6.28

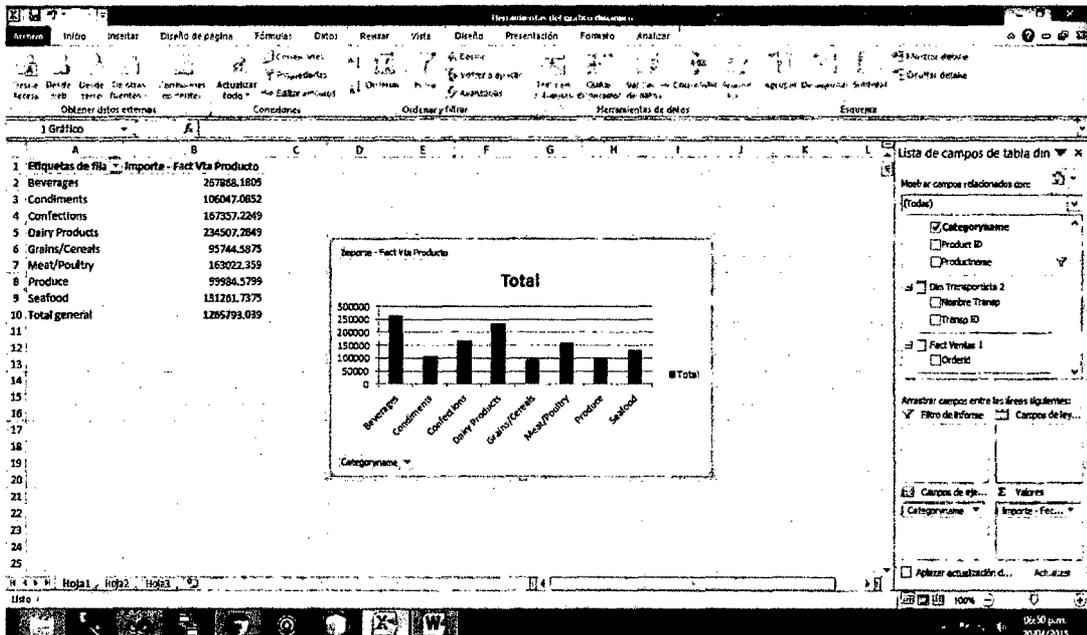
CREACION DEL CUBO POR CATEGORIA E IMPORTE



Fuente: Autor

GRAFICO N° 6.1

CREACION DEL GRAFICO POR CATEGORIA E IMPORTE

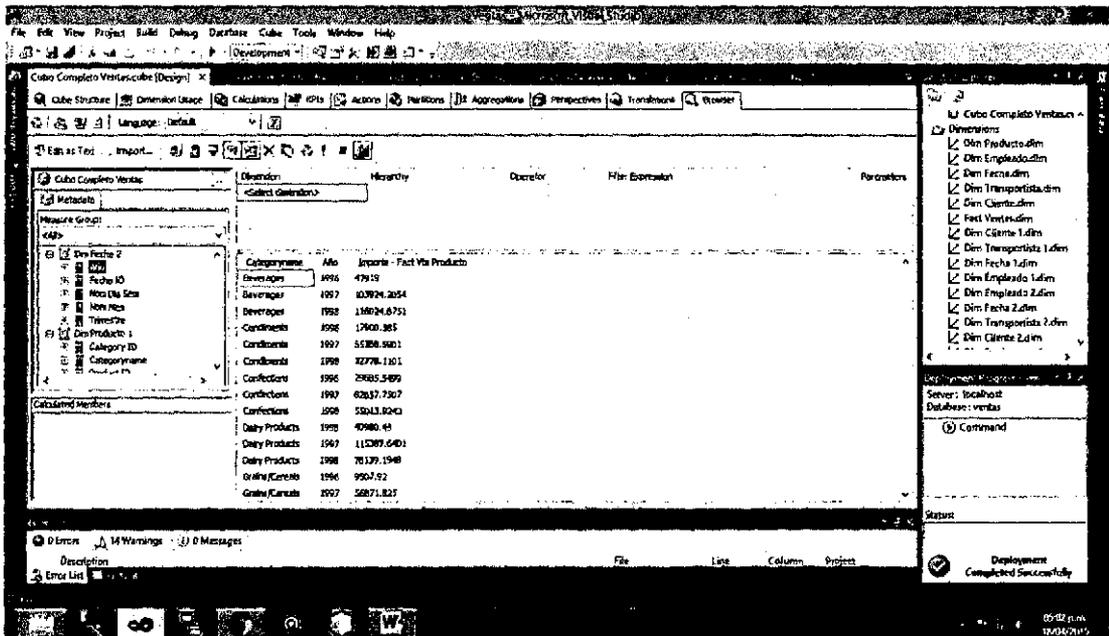


Fuente: Autor

6.6.2 Por categoría, año e importe

FIGURA N° 6.29

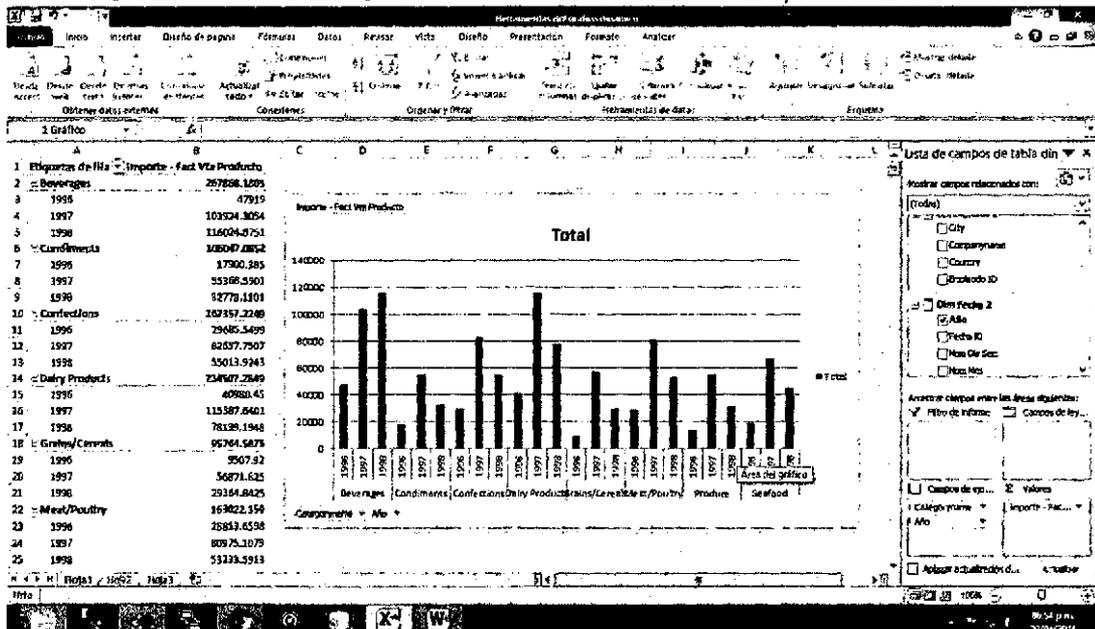
CREACION DEL CUBO POR CATEGORIA, AÑO E IMPORTE



Fuente: Autor

GRAFICO N° 6.2

CREACION DEL GRAFICO POR CATEGORIA, AÑO E IMPORTE

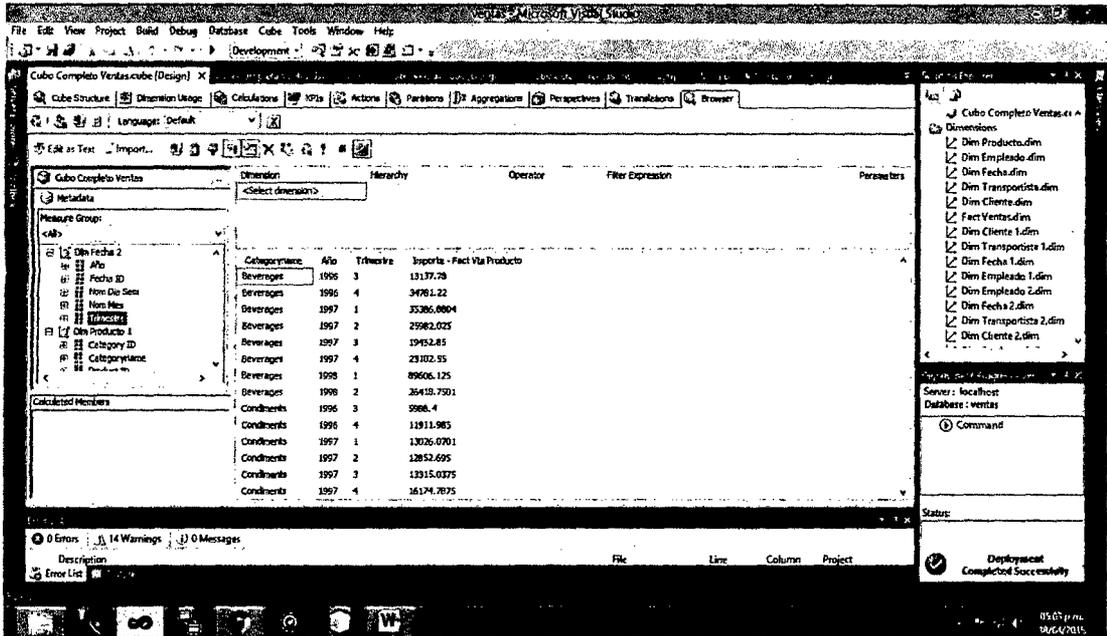


Fuente: Autor

6.6.3 Por categoría, año, trimestre e importe

FIGURA N° 6.30

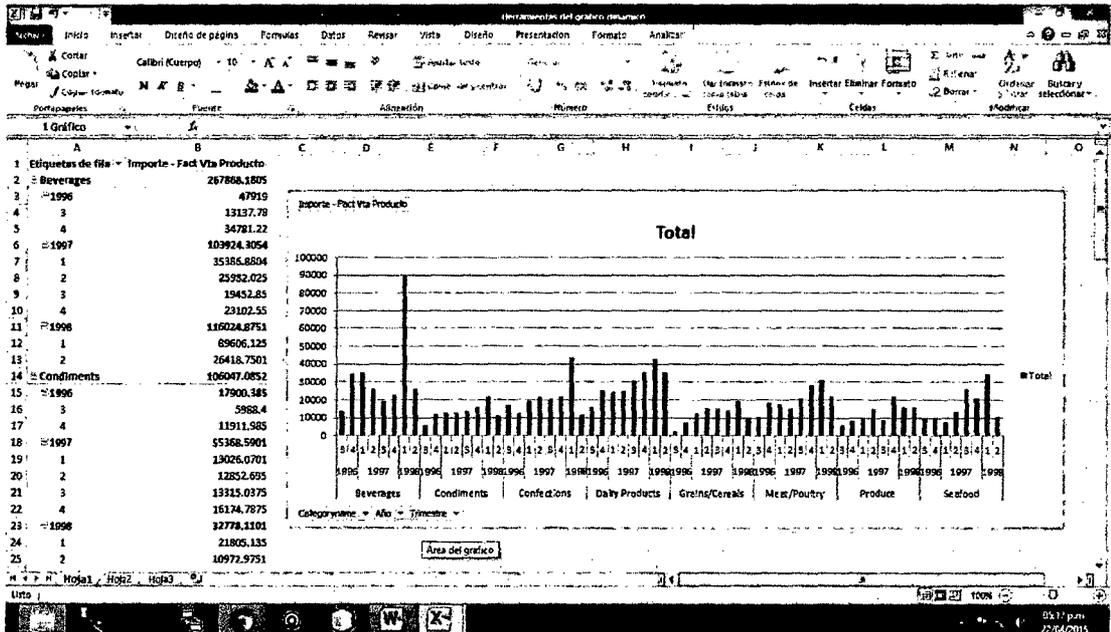
CREACION DEL CUBO POR CATEGORIA, AÑO, TRIMESTRE E IMPORTE



Fuente: Autor

GRAFICO N° 6.3

CREACION DEL GRAFICO POR CATEGORIA, AÑO, TRIMESTRE E IMPORTE

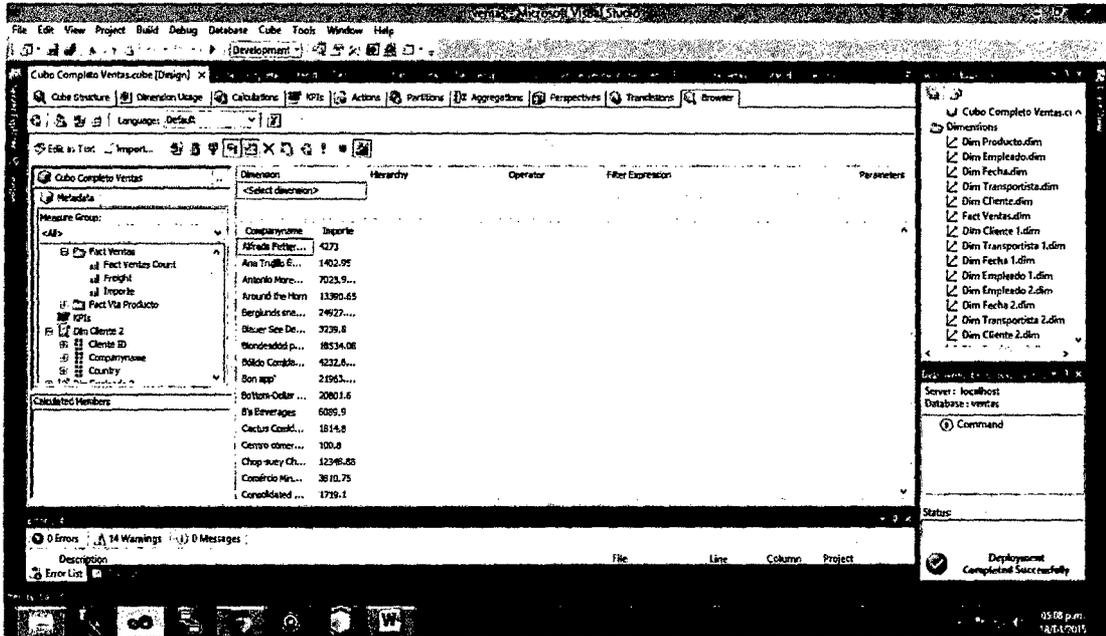


Fuente: Autor

6.6.4 Por compañía e importe

FIGURA N° 6.31

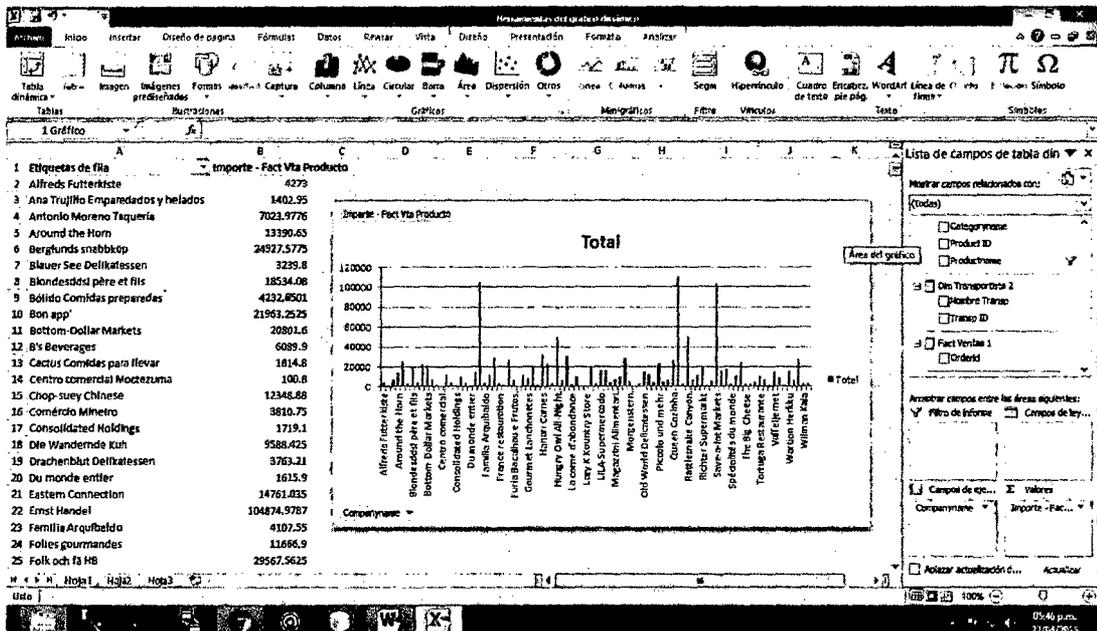
CREACION DEL CUBO POR COMPAÑÍA E IMPORTE



Fuente: Autor

GRAFICO N° 6.4

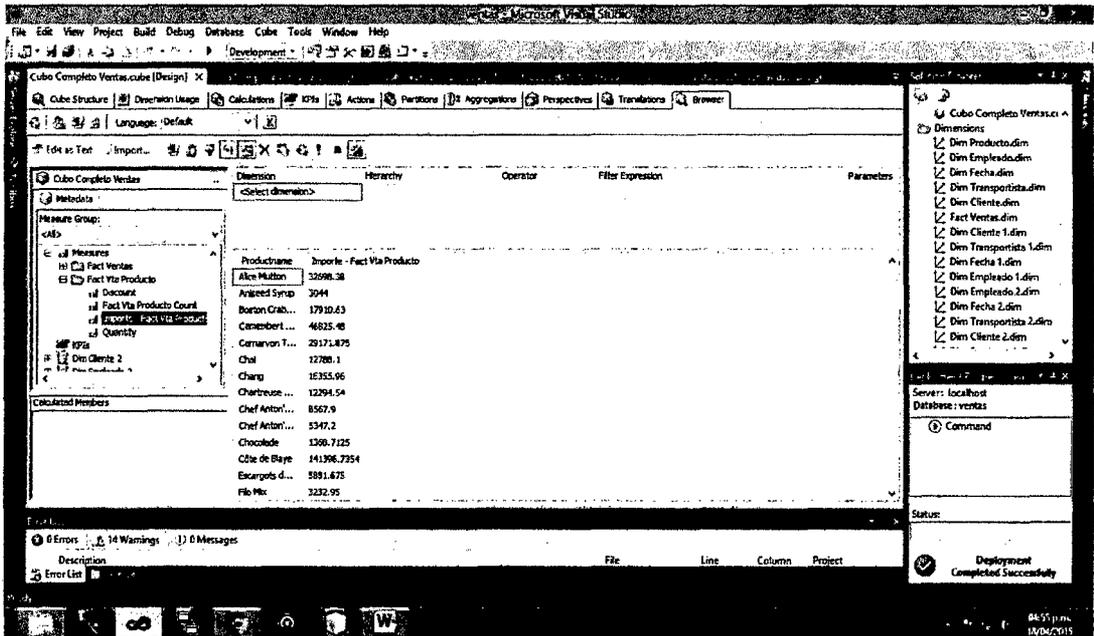
CREACION DEL GRAFICO POR COMPAÑÍA E IMPORTE



Fuente: Autor

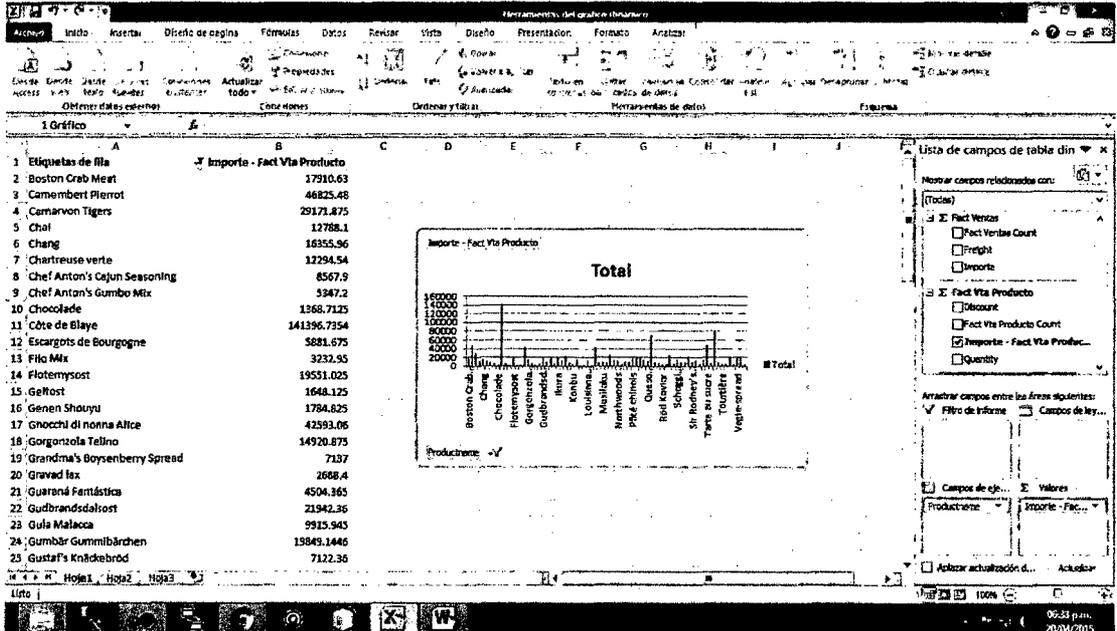
6.6.5 Por producto

FIGURA N° 6.32
CREACION DEL CUBO DEL IMPORTE DE LA FACTURA POR PRODUCTO



Fuente: Autor

GRAFICO N° 6.5
CREACION DEL GRAFICO DEL IMPORTE DE LA FACTURA POR PRODUCTO

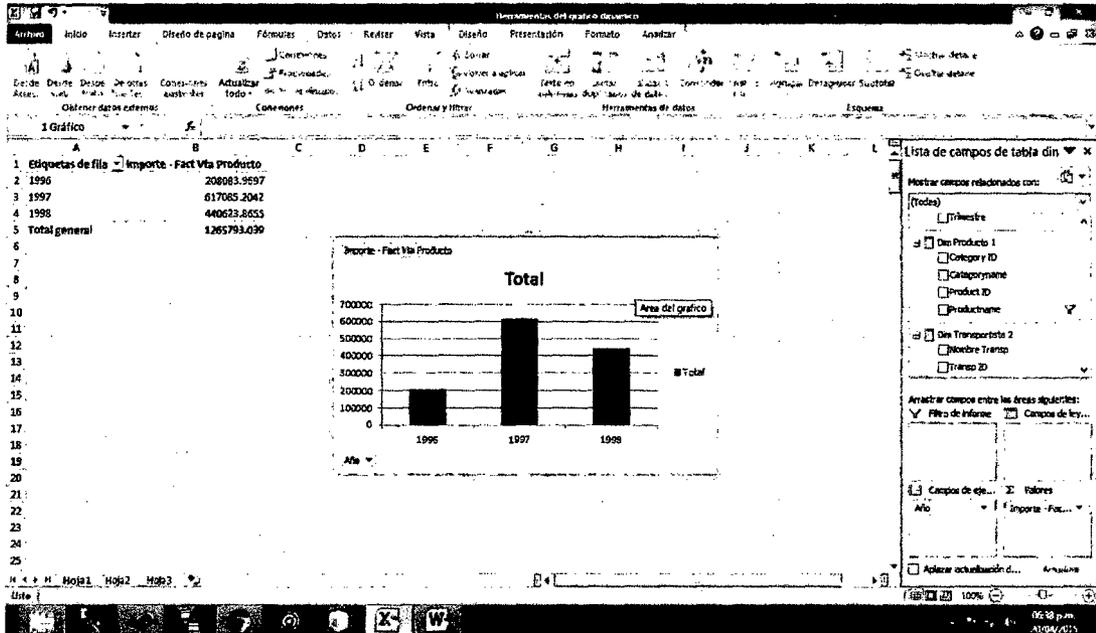


Fuente: Autor

6.7 Gráficos adicionales en excel

GRAFICO N° 6.6

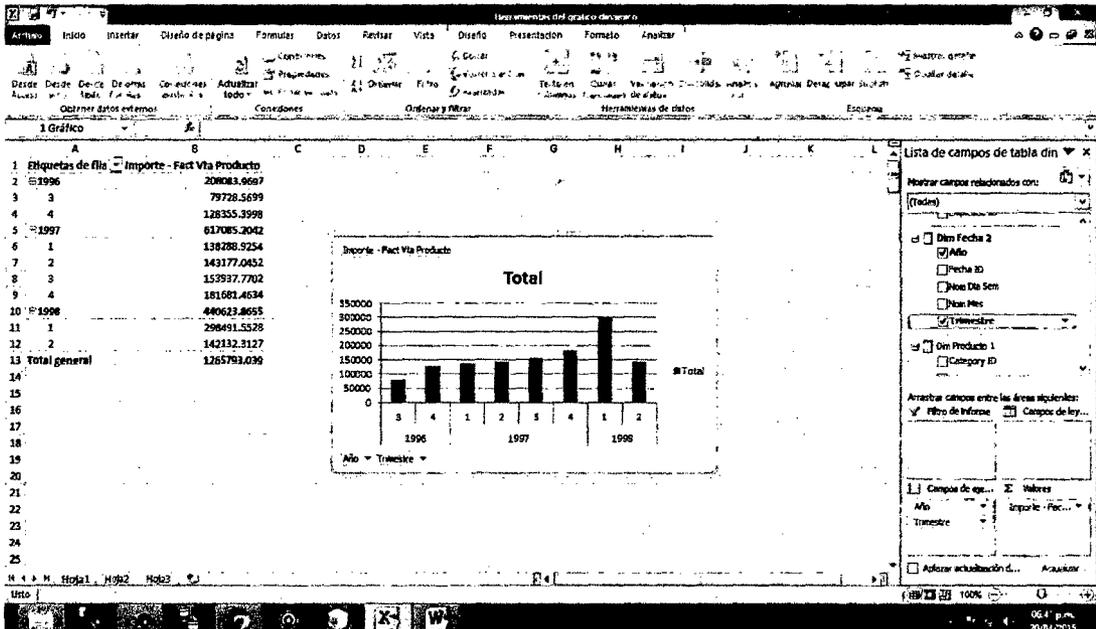
MOSTRAR EL IMPORTE DE LA FACTURACION POR AÑO



Fuente: Autor

GRAFICO N° 6.7

MOSTRAR EL IMPORTE DE LA FACTURACION POR AÑO Y POR TRIMESTRE



Fuente: Autor

VII. DISCUSION

1.- De los resultados obtenidos podemos concluir que los cubos y su expresión gráfica nos brindan mucha información en segundos, ejemplo: podemos seleccionar cubos por año; por trimestres; por año y por mes; por mes; por año y trimestre, Etc. lo cual facilita la toma de decisiones. Todo esto no sería posible con un enfoque tradicional.

2.- Es importante la limpieza de los datos de las Bases de Datos, ya que podrían generar Datawarehouses equivocados y concluir con análisis de inteligencia de negocios errados. Por lo que podemos concluir que es de mayor importancia la consistencia de los datos de las Bases de Datos tradicionales antes de que sean procesadas en un Datawarehouse.

3.- El Datawarehouse permitió apoyar al área de ventas en la toma de decisiones, mediante la información relevante y oportuna en un tiempo de segundos.

4.- El Datawarehouse se convirtió en una herramienta que mediante los cubos generados y exportándolos a Excel, permitió la generación de diferentes tipos de gráficos para su análisis, a la sola selección de los campos requeridos de las diferentes dimensiones existentes en un esquema de tipo estrella.

5.- Las Bases de Datos tradicionales no permiten un análisis de inteligencia de negocios, ya que tienen otro diseño y su objetivo es servir al día a día, pero no contienen datos para un cierto proceso del negocio. Y resulta muy difícil obtener cubos a través de los sistemas tradicionales o transaccionales.

6.- Si bien es posible obtener gráficos de los cubos a través de Visual Studio 2010, la versatilidad de Excel para la generación de diferentes tipos de gráficos es mejor, siempre y cuando estos cubos sean exportados del Visual Studio 2010 y no sean generados los cubos en Excel.

7.- Se podría enseñar Bases de Datos Tradicionales en un curso inicial de Base de Datos o Pre-Grado, por su mayor simplicidad y en un curso más avanzado Inteligencia de Negocios, por ser más complejo, como por ejemplo en Post-Grado.

8.- Inteligencia de Negocios es un tema de Administración con el apoyo de la Tecnología informática. La Administración juntamente con la Tecnología Informática se complementan para dar origen a la Inteligencia de negocios.

VIII. REFERENCIALES

- 1.- BANESTO. **Business Intelligence: Competir con Información.** España. Fundación Cultural Banesto. 2014.
- 2.- CONESA, JORDI y CURTO, JOSEP. **Introducción al Business Intelligence.** España. Editorial UOC. 1era edición. 2011.
- 3.- ELMASRI, RAMEZ y NAVATHE, SHAMKANT. **Fundamentos de Sistemas de Bases de Datos.** México. Editorial Prentice Hall. 5ta edición. 2011.
- 4.- MEDINA, EDISON. **Business Intelligence una guía Práctica.** Perú. Editorial Universidad Peruana de Ciencias. 2da edición. 2012.
- 5.- PASCO, IVAN. **Business Modeling: La generación de valor a través de las tecnologías de la información.** Perú. Ediciones PEISA S.A.C. 1era edición. 2006.
- 6.- PEREZ, CESAR y SANTIN, DANIEL. **Minería de Datos, técnicas y herramientas.** España. Thomson Editores Spain. 1era edición. 2007.
- 7.- REINOSA, ENRIQUE y MALDONADO, CALIXTO y OTROS. **Base de Datos.** Argentina. Editorial Alfaomega Grupo editor Argentino. 1era edición. 2012.
- 8.- SILBERSCHATZ, ABRAHAM y KORTH, HENR. **Fundamentos de Bases de Datos.** España. Editorial Mc Graw Hill. 4ta Edición. 2002.
- 9.- VITT, ELIZABETH y LUCKEVITH, MICHAEL y MISNER, STACIA. **Business Intelligence: Técnicas de análisis para la**

toma de decisiones estratégicas. España. Editorial Mc Graw-Hill/Interamericana. 1era edición, 2003.

10.- YALÁN, J. y PALOMINO, L., **Implementación de un Datamart como una solución de Inteligencia de Negocios para el área de Logística de T-Impulso.** *Revista de investigación de Sistemas e Informática.* Facultad de ingeniería de Sistemas e Informática. Universidad Nacional Mayor de San Marcos. Perú. Vol. (10): 52 a 63. Enero – Junio 2013.

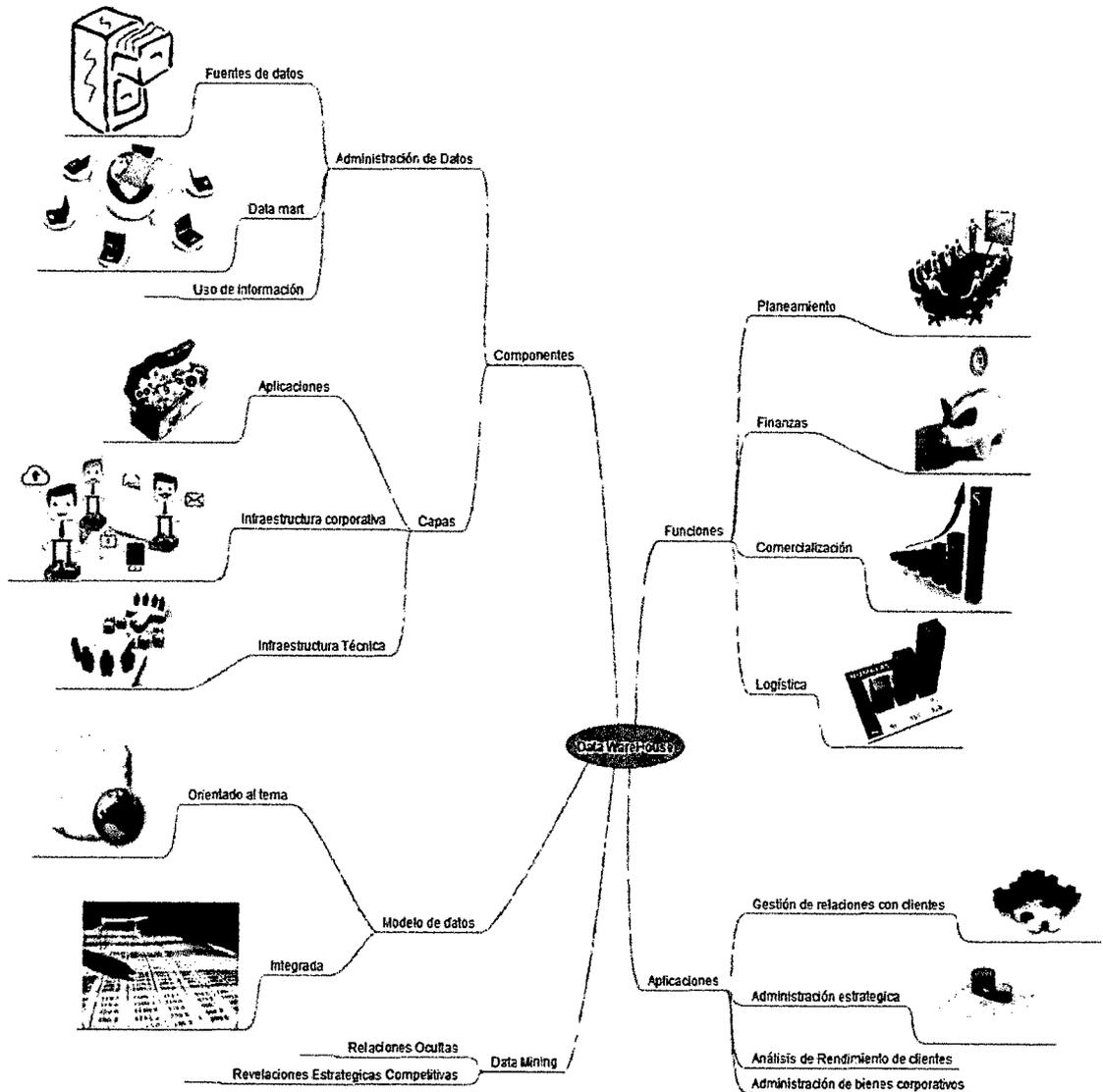
11.- ZORRILLA, MARTA. **Analysis Server 2008, Diseño multidimensional Tecnología OLAP tutorial.** España. Universidad de Cantabria. 2010.

12.- **Fundamentos de aplicaciones Business Intelligence.** División de Alta Tecnología. Cibertec, 2012.

IX. APÉNDICE

APENDICE N° 1

MAPA MENTAL DEL DATAWAREHOUSE



Fuente: Autor

Handwritten signature

DATAWAREHOUSING

Propedeutico 2014
Curso: Base de Datos
Prof. Mg. Bertila García Díaz
Sesión 9

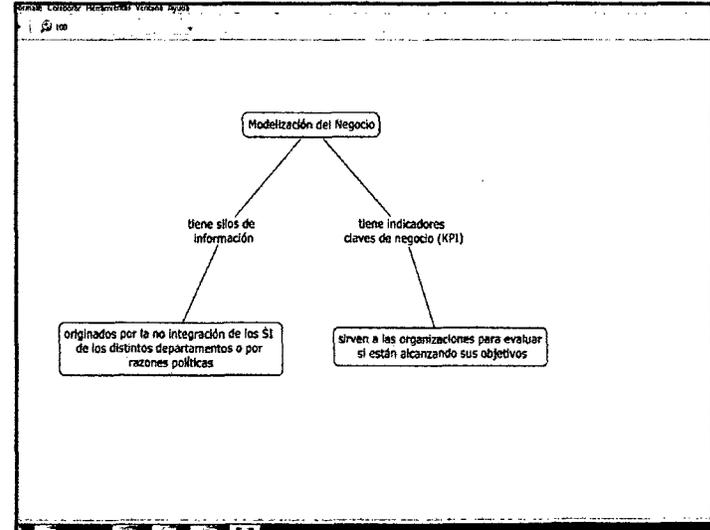
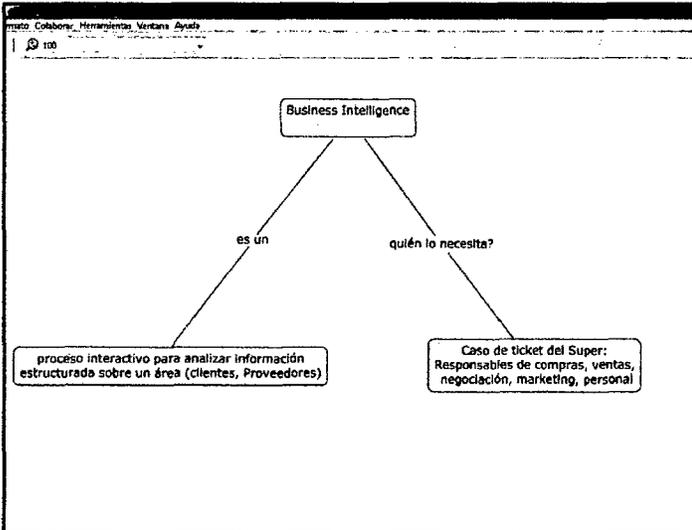
1

Términos importantes

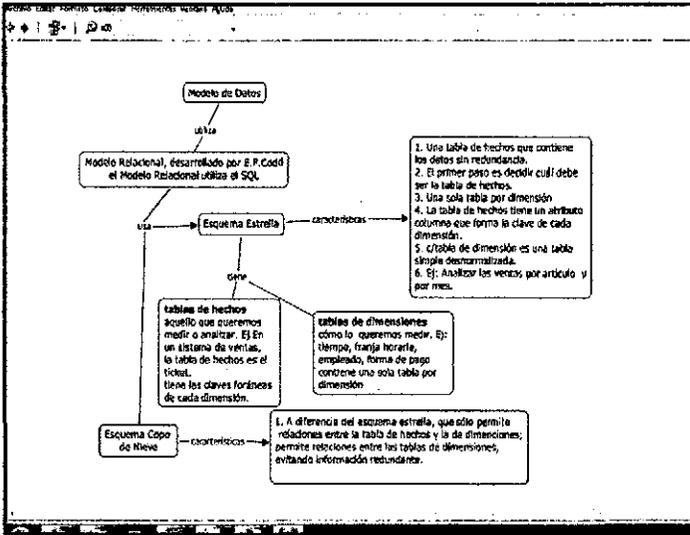
- *OLAP*.- On line Analytical Processing
-Procesamiento analítico en línea (Datawarehouse)
- *OLTP*.-Bases de datos operacionales
-Bases de Datos Operacionales

2

APENDICE N° 2
SEPARATA DE DATAWAREHOUSE



ppp



BENEFICIOS DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS

- Crear un círculo virtuoso de la información
- Permitir una visión única, conformada, histórica y de calidad.
- Crear, manejar y mantener métricas, indicadores claves de rendimiento (KPI) e indicadores claves de metas (KGI).
- Aportar información actualizada tanto a nivel agregado como en detalle.

6

¿Cuándo es necesaria la inteligencia de negocios?

Entre las situaciones, tenemos:

- La toma de decisiones se realiza de forma intuitiva en la organización.
- Problemas de calidad de información.
- Uso de excel.
- Existe demasiada información en la organización.

7

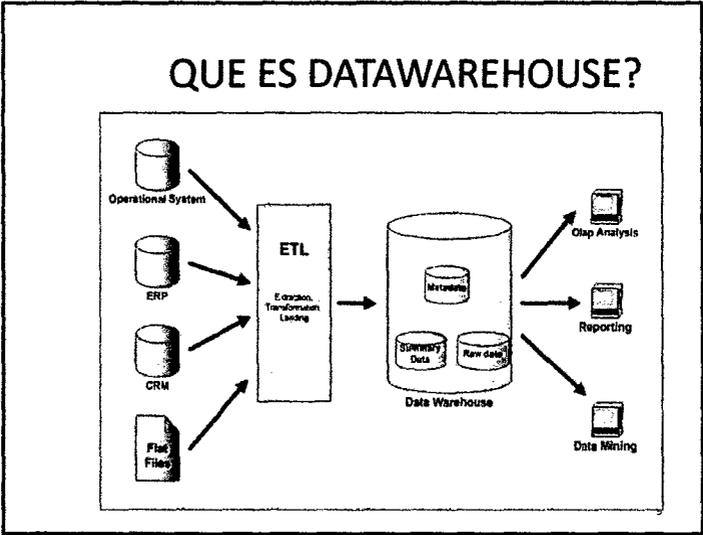
En Resumen

Los BI buscan responder a las preguntas:

- ¿Qué pasó?
- ¿Qué pasa ahora?
- ¿por qué pasó?
- ¿Qué pasará?

8

99



Data warehouse

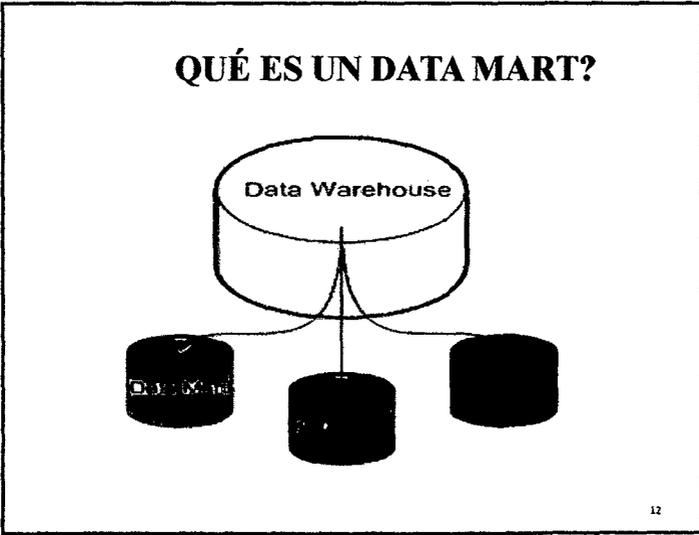
- Es el núcleo de un sistema de inteligencia de negocio.
- Es un repositorio de datos que proporciona una visión global, común e integrada de los datos .
- Las Bases de datos relacionales son el soporte técnico más usado para almacenar las estructuras de estos datos

10

QUE ES DATAWAREHOUSE?

- Variante en el tiempo.- Se realizan fotos de los datos basados en fechas o hechos.
- Orientada al tema.- organiza una colección de información alrededor de un tema central.
- Integrada.- incluye datos de múltiples orígenes.
- No volátil.- sólo de lectura para los usuarios finales.

11



ppp 67

OTROS ELEMENTOS

- **DATA WAREHOUSING** es el proceso de extraer y filtrar datos de las operaciones comunes de la organización, procedentes de los distintos sistemas de información operacionales y/o sistemas externos, para transformarlos, integrarlos y almacenarlos en un almacén de datos con el fin de acceder a ellos, para dar soporte en el proceso de toma de decisiones.
- **DATA MART**: está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización.

13

OTROS ELEMENTOS

- **Procesos ETL**: tecnología de integración de datos basada en la consolidación de datos que se usa tradicionalmente para alimentar data warehouse, data mart

14

ELEMENTOS DE UN DATA WAREHOUSE

- La idea principal es que la información sea presentada desnormalizada para optimizar las consultas. Para ello debemos identificar, en el seno de la empresa, los procesos de negocio, las vistas para el proceso de negocio, las medidas cuantificables asociadas a los mismos.

15

ELEMENTOS DE UN DATA WAREHOUSE

- **Tabla de hecho**: es la representación en el data warehouse de los procesos de negocio de la organización. Ejemplo: una venta.
- **Dimensión**: es la representación en el DW de una vista para un cierto proceso de negocio. Ejemplo: en un proceso de venta, sería el cliente que ha comprado, la fecha en que se ha realizado.
- **Métrica**: son los indicadores de negocio. Son los conceptos cuantificables que permiten medir nuestro proceso de negocio. Ej: en una venta tenemos el importe de la misma

16

TIPOS DE ESQUEMA

- **Esquema estrella** : consiste en una tabla de hechos en el centro para el hecho objeto de análisis y una o varias tablas de dimensión por cada punto de vista de análisis que participa de la descripción de ese hecho. En la tabla de hecho encontramos los atributos destinados a medir sus métricas.
- **Esquema en copo de nieve**: es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas.

17

DISEÑO DE ANALISIS OLAP

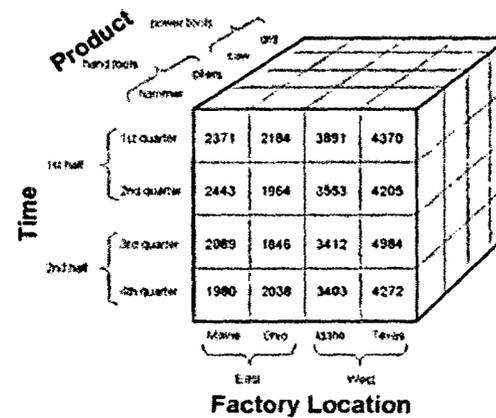
18

DISEÑO DE ANALISIS OLAP

- Es una tecnología que permite un análisis multidimensional a través de tablas matriciales o pivotantes.
- Se entiende por OLAP al método ágil y flexible para organizar datos sobre un objeto, y cuyo objetivo es recuperar y manipular datos a través de consultas.
- Una herramienta OLAP está formada por un motor y un visor

19

CUBOS



20

69

CUBOS

- es una base de datos multidimensional, en la cual el almacenamiento físico de los datos se realiza en un vector multidimensional. Los **cubos OLAP** se pueden considerar como una ampliación de las dos dimensiones de una hoja de cálculo.

21

CUBOS

- La propuesta de **Codd** consistía en realizar una disposición de los datos en vectores para permitir un análisis rápido. Estos vectores son llamados **cubos**. Disponer los datos en cubos evita una limitación de las bases de datos relacionales, que no son muy adecuadas para el análisis instantáneo de grandes cantidades de datos

22

OLAP

- Es una solución utilizada en el campo del *Business Intelligence* cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o **Cubos OLAP**) que contienen datos resumidos de grandes Bases de datos (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

23

OLAP

- Codd lo define como el nombre dado al análisis dinámico de la Empresa, requerida para crear, manipular, animar y sintetizar información. Incluye la habilidad para discernir nuevas o no anticipadas relaciones entre las variables, la habilidad para identificar los parámetros necesarios para manejar grandes cantidades de datos, para crear un número ilimitado de dimensiones

24

Hand
70

OLAP

- Un análisis multidimensional representa datos como tablas, implica operaciones típicas: Mayor, menor, comparación entre periodos, porcentajes de diferencia, funciones estadísticas y financieras.
- Comprende la consulta interactiva y el análisis de los datos.
- Recupera y exhibe datos tabulares en 2 o 3 dimensiones, cuadros y gráficos.

25

Ejemplos de Aplicaciones OLAP

- Número de unidades vendidas por línea y modelo, trimestres, años?
- Porqué las Ventas de automóviles de determinado modelo disminuyen en el segundo trimestre?
- Porqué la ganancia generada por los 10 mejores clientes decreció un 30% en relación al mismo periodo del año anterior?

25

X. ANEXOS

ANEXO N° 1

MATRIZ DE CONSISTENCIA

“METODOLOGIA PARA LA CREACION DE UNA BASE DE DATOS MULTIDIMENSIONAL, USANDO TECNOLOGÍA DE INTELIGENCIA DE NEGOCIOS Y SQL SERVER EN UNA EMPRESA PERUANA.”

Formulación del problema	Objetivos	Hipótesis	Variables	Metodología
<p>¿Cómo crear un procedimiento o Metodología para crear Bases de Datos multidimensionales que apoyen a la toma de decisiones, teniendo en cuenta que esta tecnología es reciente, usando tecnología de Inteligencia de Negocios y SQL Server?.</p>	<p>Determinar una Metodología para crear una Base de Datos Multidimensional usando tecnología de inteligencia de Negocios y el Sistema de Gestión de Base de Datos SQL server, utilizando la información almacenada en los sistemas OLTP, como una etapa posterior a las Bases de Datos transaccionales.</p>	<p>“Si se toma en cuenta la Inteligencia de Negocios y el SQL server para la creación de una Base de Datos multidimensional en las empresas, estas podrán mejorar su capacidad de tomar decisiones acertadas y rápidas.”</p>	<p>Variable X: (independiente) Inteligencia de Negocios</p> <p>Variable Y: (independiente) SQL server</p> <p>Variable Z: (dependiente) Metodología para crear Bases de Datos multidimensionales</p>	<p>Nivel de la investigación: Es de tipo Aplicada.</p> <p>Tipo de la investigación: cualitativa y básica.</p> <p>Método: inductivo.</p> <p>Diseño de la investigación: Experimental, para propiciar la creación de la Metodología que se midió con la generación de los cubos en Excel, que son los resultados obtenidos, luego de crear un largo proceso de creación de software.</p> <p>Universo: BD Multidimensionales.</p> <p>Muestra: BD ventas.</p>

ANEXO N° 2

TERMINOS IMPORTANTES

- BI.- (business intelligence) inteligencia de negocios; conjunto de estrategias y aspectos relevantes enfocados a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o empresa.
- CRM (Customer relationship management).- Gestión de relaciones con clientes; Sistemas informáticos de apoyo a la gestión de las relaciones con los clientes, a la venta y al marketing. Dicho software puede comprender varias funcionalidades para gestionar las ventas y los clientes de la empresa: automatización y promoción de ventas, tecnologías data warehouse (almacén de datos) para agregar la información transaccional.
- DBMS.- (Database management system) sistemas de administración de bases de datos; permiten almacenar y posteriormente acceder a los datos de forma rápida y estructurada.
- ETL.- (Extract, Transform and Load) extraer, transformar y cargar; es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar.
- ERP.- (Enterprise Resource Planning), Planificación de Recursos Empresariales; son sistemas informáticos destinados a la administración de recursos en una organización.
- OLAP.- (On-Line Analytical Processing) procesamiento analítico en línea; Es una solución utilizada en el campo de la llamada Inteligencia de negocios cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales.
- OLTP.- (OnLine Transaction Processing) procesamiento de transacciones en línea; es un tipo de procesamiento que facilita y

administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones. Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor.

- SQL.- (Structured Query Language) lenguaje de consulta estructurado; es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en ellas.