



UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR

(UNTELS)

**“ANÁLISIS Y DISEÑO DE UN DATAMART COMO HERRAMIENTA DE
APOYO PARA LA CREACIÓN DE ESTRATEGIAS EN EL ÁREA DE VENTAS
DE UNA EMPRESA DEL RUBRO DE ASESORÍAS Y SEMINARIOS”**

TRABAJO DE INVESTIGACIÓN PARA OPTAR EL TÍTULO DE

INGENIERO DE SISTEMAS

PRESENTADO POR LA BACHILLER

SHEILA MIRIAM LIVANO CASTILLO

LIMA-PERÚ

2014

Dedicatoria

A mis padres, que son el motor de cada esfuerzo y cada paso que doy, así como también, al propósito por el cual Dios me puso en este camino de vida, y estoy convencida que esta experiencia es parte de ése propósito.

Bachiller: Sheila Livano Castillo

Agradecimientos

Principalmente a Dios, que conoce el camino que he recorrido para llegar aquí.

También quiero agradecer por su contribución en el presente proyecto a Jefes y Compañeros del trabajo, porque permitieron y entendieron en muchas ocasiones mi ausencia por motivos del proyecto.

Amigos que nunca se negaron a orientarme, y aquí los quiero mencionar: El Lic. Luis Hondermann, El Ing. Ángel Gonzales H., Hans Moreno S. y Luis Ordoñez M.

A mi asesor, el Ing. Frank Escobedo, que ha sido un maestro y amigo en todo este tiempo que ha durado el presente proyecto, así como en toda mi etapa Universitaria. Y a la Universidad que me formó académicamente y de donde tengo lindos recuerdos.

Bachiller: Sheila Livano Castillo

ÍNDICE GENERAL

INTRODUCCIÓN.....	10
CAPÍTULO I.....	11
PLANTEAMIENTO DEL PROBLEMA	11
1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA	11
1.1. JUSTIFICACIÓN	13
1.2. DELIMITACIÓN DE LA INVESTIGACIÓN	14
1.2.1. Espacial:	14
1.2.2. Temporal:	14
1.3. FORMULACIÓN DEL PROBLEMA.....	14
1.3.1. Problema principal.....	14
1.4. OBJETIVO	14
1.4.1. Objetivo principal.....	14
CAPÍTULO II.....	15
MARCO TEÓRICO	15
2.1. ANTECEDENTES	15
2.2. BASES TEÓRICAS	19
2.2.1. Necesidad de un Datawarehouse	19
2.2.2. Conceptos de Datawarehouse.....	19
2.2.2.1. Beneficios de un Datawarehouse	20
2.2.3. Data Mart	21
a. Dependientes	22
b. Independientes	22
2.2.4. Paradigmas de Datawarehouse y Datamart según Inmon y Kimball	22
2.2.4.1. Bill Inmon	22
2.2.4.2. Ralph Kimball	23
2.2.5. ETL (Extract, Transform and Load).....	24
2.2.5.1. Fases del ETL.....	24
a. Extracción	24
b. Transformación	25
c. Carga	25

2.2.6.	Modelado Multidimensional	26
2.2.6.1.	Hechos	27
2.2.6.2.	Dimensiones	27
2.2.6.3.	Medidas	28
2.2.6.4.	Esquemas	28
a.	Esquema Estrella	28
b.	Esquema en copo de nieve	29
2.2.6.5.	OLAP (On-Line Analytical Processing)	30
2.2.6.6.	MOLAP (Multidimensional OLAP)	30
2.2.6.7.	ROLAP (Relational OLAP)	31
2.2.7.	Metodología Kimball	31
2.2.7.1.	Planificación del Proyecto	32
2.2.7.2.	Definición de los Requerimientos del Negocio	32
2.2.7.3.	Modelado Dimensional	33
2.2.7.4.	Diseño Físico	33
2.2.7.5.	Diseño y Desarrollo de la Presentación de Datos	34
2.2.7.6.	Diseño de la Arquitectura Técnica	34
2.2.7.7.	Selección de Productos e Instalación	35
2.2.7.8.	Especificación de Aplicaciones para Usuarios Finales	35
2.2.7.9.	Desarrollo de Aplicaciones para Usuarios Finales	35
2.2.7.10.	Implementación	36
2.2.7.11.	Mantenimiento y crecimiento	36
2.2.7.12.	Gestión del Proyecto	36
2.3.	MARCO CONCEPTUAL	37
	CAPITULO III	51
	DESCRIPCIÓN, DISEÑO Y CONSTRUCCIÓN DEL MODELO	51
3.1.	ANÁLISIS DEL MODELO	51
3.1.1.	La Empresa	51
a.	Misión	51
b.	Visión	52
c.	Servicios que ofrece INPECHI	52
d.	Datos	53
3.1.2.	Aplicación de la Metodología	53

3.1.2.1.	Planificación del Proyecto	53
3.1.2.2.	Definición de los Requerimientos del Negocio	55
a.	<i>Algunos Requerimientos en base al área definida:</i>	55
b.	<i>Otros Requerimientos en base al modelo utilizado:</i>	56
3.1.2.3.	Modelado Dimensional	56
a.	Definición del Proceso de Negocio:	56
b.	Nivel de Granularidad:	56
c.	Elección de las Dimensiones	57
d.	Identificación de los Hechos	57
e.	Modelo Gráfico de Alto Nivel	57
f.	Detalle de las Tablas de Dimensión	58
3.1.2.4.	Diseño Físico del Data Mart	60
3.1.2.5.	Diseño y Desarrollo de la Presentación de Datos	61
3.1.2.6.	Diseño de la Arquitectura Técnica	62
3.1.2.7.	Selección de Producto e Instalación	62
3.2.	CONSTRUCCIÓN DEL MODELO	63
3.2.1.	Construcción de las Dimensiones	63
3.2.2.	Construcción de la Fact Table DM_Venta	64
3.2.3.	Construcción de los ETLs	65
3.2.4.	Construcción de Dimensiones y Cubo	69
3.3.	REVISIÓN Y CONSOLIDACIÓN DE RESULTADOS	83
	CONCLUSIONES	88
	RECOMENDACIONES	89
	BIBLIOGRAFÍA	90
	ANEXOS	91

ÍNDICE DE FIGURAS

Figura 1: Esquema de un Datawarehouse.....	20
Figura 2: Esquema de un Datawarehouse según Bill Inmon	23
Figura 3: Esquema de un Datawarehouse según Ralph Kimball	23
Figura 4: Fases del Sistema ETL.....	24
Figura 5: Ejemplo de un Proceso de ETL	26
Figura 6: Esquema modelado de Datos Tipo Estrella	28
Figura 7: Esquema modelado de Datos Tipo Copo de Nieve	29
Figura 8: Metodología de Ralph Kimball.....	32
Figura 9: Logo de la empresa Inpechi Consultores Internacionales S.A.C.	51
Figura 10: Elección de las Dimensiones y la Tabla Hechos o Fact Table	57
Figura 11: Modelo Gráfico de Alto Nivel	58
Figura 12: Atributos de las Dimensiones.....	58
Figura 13: Atributos de las Dimensiones, Persona, Servicio,	59
Figura 14: Atributos de la Fact Table	59
Figura 15: Diseño Físico del Data Mart y sus Dimensiones.....	60
Figura 16: Diseño del ETL	61
Figura 17: Arquitectura para el Proyecto de DM en INPECHI	62
Figura 18: Configuración para la conexión de Origen de Datos.....	69
Figura 19: Elección de la Base de datos en SQL Server a la cual se accederá desde el Analysis Services.	70
Figura 20: Lectura de todas las tablas en la Base de Datos OLAP	71
Figura 21: Esquema OLAP	72
Figura 22: Creación de las Dimensiones.....	72
Figura 23: Primeros pasos para crear las dimensiones.....	73
Figura 24: Selección de la dimensión CLIENTE	74
Figura 25: Selección de los atributos por Dimensión.	75
Figura 26: Generación de un Cubo	76
Figura 27: Creación del cubo	77
Figura 28: Visualización de las dimensiones y el cubo que se creará	78
Figura 29: Creación del segundo Cubo	79
Figura 30: Últimos pasos a considerar para el Cubo.....	80
Figura 31: Conexión establecida para el Cubo	81
Figura 32: Visualización de los cubos generados.....	82

ÍNDICE DE CUADROS

Cuadro 1: Cuadro de Comparación entre BD_Transaccional y un DW	21
Cuadro 2: Cronograma de Actividades para la realización del Proyecto	54
Cuadro 3: Matriz de procesos -Dimensiones.....	55
Cuadro 4: Principales clientes de la categoría A (Expresado en cantidad de personas)	87

ÍNDICE DE GRÁFICOS

Gráfico 1: Asistencia por tipo de Servicio a nivel nacional - Año 2013	83
Gráfico 2: Cantidad de Personas que asistieron al Seminario de Importación a nivel nacional – Vista a detalle por mes.....	84
Gráfico 3: Medios Publicitarios en Seminario de Importación a nivel nacional	85
Gráfico 4: Medios Publicitarios en Seminario de Importación a nivel nacional por meses.....	86
Gráfico 5: Publicidad por diarios en Seminario de Importación a nivel nacional	87

INTRODUCCIÓN

El presente trabajo de investigación lleva por título **“ANÁLISIS Y DISEÑO DE UN DATAMART COMO HERRAMIENTA DE APOYO PARA LA CREACIÓN DE ESTRATEGIAS EN EL ÁREA DE VENTAS DE UNA EMPRESA DEL RUBRO DE ASESORÍAS Y SEMINARIOS”**, para optar por el título de Ingeniero de Sistemas, presentado por la bachiller SHEILA MIRIAM LIVANO CASTILLO.

Gestionar la información en las empresas es, hoy en día, una herramienta clave para poder sobrevivir en un mercado cambiante, dinámico y global. Aprender a competir con esta información es fundamental para la toma de decisiones, el crecimiento y la gestión de las empresas.

Todo proceso o persona involucrada en una organización va dejando información con el paso del tiempo, que tal vez, a simple vista, podrían ser meras transacciones o consumo del bien o servicio que una empresa brinda. Pero lo que muchas empresas no prestan atención es que esta información puede ser de gran utilidad para la toma de decisiones permitiendo así ventajas competitivas en el mercado en que se desarrolla.

Por este motivo se requieren herramientas que permitan minimizar el tiempo de análisis de dicha información a una velocidad y precisión deseable. El poseer conocimientos correctos significa tener respuestas correctas y realizar decisiones estratégicas para su ejecución en beneficio de la empresa.

El objetivo de este trabajo de investigación es diseñar un Datamart para apoyar al área de ventas en una Mediana empresa que dicta asesorías y seminarios para Importar desde China y lograr con esta nueva herramienta que sus estrategias sean apoyadas y fundamentadas con la explotación de la información almacenada en el Datamart.

La estructura que hemos seguido en este proyecto se compone de 3 capítulos. El primer capítulo comprende el planteamiento del problema, el segundo capítulo el desarrollo del marco teórico y el tercer capítulo corresponde al desarrollo del proyecto.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA

Inpechi Consultores Internacionales S.A.C. es una organización privada, orientada a realizar capacitaciones, asesorías y consultorías de tal manera que las personas puedan realizar su negociación de Importación con China con la mayor seguridad.

Su principal servicio es dictar seminarios para aprender a Importar con China, estos se realizan una vez por semana, dictándose así 4 seminarios por mes, pudiendo ser en Lima o en otras Provincias. También cuenta con el servicio de asesorías personalizadas y 3 seminarios más: Negociación con el proveedor Chino, Cómo contactar proveedores Chinos seguros, y un seminario de Inteligencia Comercial. Estos tres seminarios fueron incorporados en el año 2012.

Para Inpechi el trabajo diario se apoya en un sistema informático que le permite almacenar el registro de sus clientes, datos generales de éste, inscribir a sus servicios de consultoría (asesorías o seminarios), organizar e ingresar al sistema resultados de encuestas que realizan al culminar cada evento, entre otros. Pero a pesar de contar con buena información que sus clientes dejan al pasar por sus consultorías

y/o seminarios, Inpechi nunca se ha dedicado a explotar dicha información.

Es por ello que jamás se ha visto en la necesidad de mantener su información integrada de tal manera que pueda obtener de ella respuestas estratégicas, es más, las campañas de publicidad o de seguimiento a clientes han sido creadas intuitivamente por la gerencia, que si bien no ha dado malos resultados, se podría optar por herramientas que simplifiquen el trabajo de la intuición, y que brinden tendencias exactas, verdaderas, que apoyen a la toma de decisiones; con pruebas que el camino que se está siguiendo es el que se debe tomar. Pero para implementar soluciones como estas en una empresa donde solo se cuenta con bases de datos simples, es necesario, optar por un nuevo repositorio de datos; uno pensado, y estructurado estratégicamente en base a lo que queremos llegar o a lo que queramos analizar. Para ello se propuso la construcción de un Datamart donde se guardará información estratégica, previamente analizada y habiéndole dado mantenimiento de depuración, limpieza, y reestructuración si lo necesitase.

La propuesta surgió después de laborar un año en Inpechi Consultores Internacionales S.A.C., en ese tiempo se pudo observar que clientes de algunas empresas volvían a inscribirse en los nuevos seminarios que Inpechi apertura en el 2012, y con ellos también venían nuevos participantes. Tal es así que se vio la oportunidad de categorizar a mucho de los clientes frecuentes en un rango [A-D] y tomarlos en cuenta para los nuevos seminarios que Inpechi podría abrir. El deseo de análisis de esta fuente de información respecto a los clientes, los lugares donde dictar nuevos seminarios, con que ponente contar, que promociones crear en base a los clientes, entre otros, generó una fuerte necesidad de explotar la información que Inpechi guarda en sus diferentes bases de datos, e interesarse en esta propuesta de solución con el fin de comenzar nuevas estrategias en base a su información almacenada y utilizarla de forma correcta en la toma de decisiones en el área de Ventas.

1.1. JUSTIFICACIÓN

Existe una gran necesidad de explotación y análisis de la información que se encuentra almacenada en las Bases de Datos de la empresa en estudio.

Si el principal objetivo es contar con mayor audiencia en los seminarios futuros, y generar ventaja competitiva frente a sus competidores, no hay más que buscar estrategias para conseguirlas, pero estas estrategias deben estar fundamentadas en un perfil, una tendencia o quizá un elemento que haga única a la empresa de captar nuevos clientes y/o fidelizar con los que ya cuentan. Así como también adelantarse a los acontecimientos mediante análisis predictivos y de esta forma aumentar la probabilidad de éxito de la toma de decisiones.

La creación de un Data Mart permitirá a Inpechi recopilar data de sus diferentes bases de datos con una dirección definida. No es el hecho de implementar un nuevo repositorio y tener almacenada la información un tiempo más; es por el contrario el primer paso para filtrar información útil que servirá de apoyo a la gerencia; almacenando en el Datamart información relevante que será el corazón de cualquier solución de Inteligencia de Negocios que se quiera implementar.

Además la tecnología de un Datamart resulta sumamente importante para las compañías medianas, las cuales típicamente no tienen tan vastos recursos como los líderes de la industria pero tienen la agilidad para implementar decisiones de negocio significativas de una manera rápida y relativamente sencilla.

Para complementar la implementación del Datamart se utilizará la herramienta Microsoft Excel que nos permitirá visualizar en gráficos lo que la gerencia desea saber, sin embargo cabe resaltar que Excel funcionará como un añadido al trabajo que se está desarrollando en la presente, ya que se puede prescindir de esta herramienta y utilizar otros productos para la generación de reportes, esto hace del trabajo aún más económico y adaptable pues se podrá trabajar con software que la empresa ya cuenta.

1.2. DELIMITACIÓN DE LA INVESTIGACIÓN

1.2.1. Espacial:

Este trabajo de investigación se desarrollará en Lima en el distrito de Miraflores, en el C. C. CompuPalace Av. Arequipa 4947 – Piso 3. Stand 3031; ubicación de la Oficina principal de Inpechi Consultores Internacionales.

1.2.2. Temporal:

Comprende el periodo de

INICIO: 2 de Setiembre de 2013

TÉRMINO: 17 de Enero de 2014

1.3. FORMULACIÓN DEL PROBLEMA

1.3.1. Problema principal.

¿De qué manera el análisis y diseño de un Datamart como herramienta de apoyo influye en la creación de estrategias en el área de ventas de una empresa del rubro de asesorías y seminarios?

1.4. OBJETIVO

1.4.1. Objetivo principal.

Determinar la influencia de analizar y diseñar un Datamart, como herramienta de apoyo, en la creación de estrategias del área de ventas en una empresa del rubro de asesorías y seminarios.

CAPÍTULO II

MARCO TEÓRICO

2.1. ANTECEDENTES

Cada Empresa tiene su razón de ser, su propia visión y misión e incluso pueden contar con estrategias propias del negocio o de quien las dirige; pero sea cual fuere el rubro o sector a quien van dirigidos existe un común denominador de todas las empresas y es que siempre estarán enfocadas en alcanzar ventajas que superen a sus competidores y ser líder en su sector.

Las investigaciones y aplicaciones de Datamart van más allá de aplicar soluciones nuevas e innovadoras en una organización. Tienen objetivos a nivel gerencial y buscan fortalecer el valor de la empresa convirtiéndola en un ente capaz de construir un conjunto de estrategias que lo lleven al éxito.

Aquí mencionaremos algunos trabajos Aplicativos de implementación de Datamart desarrollados a **nivel Internacional**:

Abril del 2007- Ecuador, López Beltrán, Carlos Patricio, presenta su tesis para optar por el título de Ingeniero en Sistemas e Informática en la Escuela Politécnica el Ejército, con título "ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA MART PARA LA DIRECCIÓN FINANCIERA Y RECURSOS HUMANOS DE LA ESCUELA

POLITÉCNICA DEL EJÉRCITO PARA UNA TOMA DE DECISIÓN EFECTIVA” se busca el desarrollo de un Datamart, el cual permitiría mejorar el manejo de la información de la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército, plasmándose en decisiones efectivas y oportunas; y además integra dos áreas estrechamente relacionadas como son las áreas financieras y recursos humanos para así mejorar la gestión administrativa financiera.

Junio 2012-Chile, Arias Cuevas, Jaime Mijail presenta su tesis para optar por el título de Ingeniero Civil Industrial en la Universidad de Chile, con título "DISEÑO Y CONSTRUCCIÓN DE UN DATAMART PARA EL ANÁLISIS DE PREFERENCIAS DEL USUARIO EN LA WEB A PARTIR DE DATOS ORIGINADOS EN EL PORTAL EDUCARCHILE" propone un modelo capaz de satisfacer las necesidades de información de la organización, entregando conocimiento útil a la hora de evaluar y definir nuevas políticas de publicación de contenidos que se ajustan a las reales preferencias de los usuarios del portal. Esto, gracias a que a través de la creación de indicadores limpios y consolidados respecto del uso de las secciones y viñetas que componen el portal, y que se almacenan en un Data Mart, el equipo de administración del sitio puede acceder a una información detallada acerca del comportamiento de sus visitantes, la que hasta ese momento no había sido obtenida.

Año 2013 – Ecuador, Aimacaña Quilumba Doris Eulalia presenta su trabajo de graduación para optar por el Título de Ingeniero Informático en la Universidad Central del Ecuador, con título “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART ACADÉMICO USANDO TECNOLOGÍA DE BI PARA LA FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA”, desarrolla un sistema de BI para colocar la información al alcance de los responsables de tomas de decisiones, de modo que permita agilizar el proceso de análisis de datos, formulación de estrategias de prevención y planificación de actividades de una forma más rápida y eficaz; puesto a que se contaba con la información académica de estudiantes y docentes almacenada en una

base de datos, pero no existía un sistema que permita la toma de decisiones de manera proactiva.

Otras tesis desarrolladas a **nivel nacional** respecto al Análisis y Desarrollo de Datamart.

En el 2012 - Cajamarca, Guillén Rodríguez, Fiorelly Shirley en su Tesis para optar por el Título de Ingeniero de Sistemas en la Universidad Privada del Norte, con título "DESARROLLO DE UN DATA MART PARA MEJORAR LA TOMA DE DECISIONES EN EL ÁREA DE TESORERÍA DE LA MUNICIPALIDAD PROVINCIAL DE CAJAMARCA" desarrolla un Datamart para el apoyo en la toma de decisiones del área de Tesorería de la Municipalidad. La solución consiste en brindar una gama de reportes de análisis gerenciales que permitan visualizar las recaudaciones que obtienen para las demás áreas existentes dentro de la Municipalidad Provincial de Cajamarca. Los resultados obtenidos son los reportes para la toma de decisiones de forma rápida y confiable.

En el 2008 – Trujillo, un grupo de estudiantes de la Universidad Nacional de Trujillo presentaron la tesis titulada “SISTEMA DE SOPORTE PARA LA TOMA DE DECISIONES Y LA MEJORA DE LA GESTIÓN ADMINISTRATIVA DE LA ESCUELA DE INGENIERÍA DE SISTEMAS DE LA FACULTAD DE INGENIERÍA DE LA U.N.T.” La investigación es descriptiva, aplicada a la elaboración de un modelo de Datawarehouse (Bases de datos multidimensionales) para la Escuela de Ingeniería de Sistemas de la Universidad Nacional de Trujillo, la población de estudio son los datos de los 11 docentes y 250 alumnos con los que cuenta, utilizándose un muestreo simple para el tamaño de muestra. El objetivo es elaborar un sistema de Soporte que sirva de base para agilizar la toma de decisiones de los ejecutivos de la Escuela de Ingeniería de Sistemas: Director de la Escuela y Jefe del Departamento de Ingeniería de Sistemas.

Otras tesis desarrolladas a **nivel local** respecto al Análisis y Desarrollo de Datamart.

En enero del 2009, Rosales Sedano, Carmen Pamela en su Tesis para optar por el Título de Ingeniero Informático en la Pontificia Universidad Católica del Perú con título, "ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA MART PARA EL SOPORTE DE TOMA DE DECISIONES Y EVALUACIÓN DE LAS ESTRATEGIAS SANITARIAS EN LAS DIRECCIONES DE SALUD" Implementa un sistema de información que permite el monitoreo y la evaluación de las estrategias dentro de cada dirección regional de salud siguiendo siempre las políticas fijadas en el Plan Concertado, y así poder controlar los avances que se realizan y el cumplimiento de las políticas. La inteligencia de negocios se utiliza para la evaluación de los resultados de cada estrategia sanitaria de las Direcciones de salud.

Diciembre del 2009, Fernández Ochoa, Eddy Armando en su Tesis para optar por el título de Ingeniero Informático en la Pontificia Universidad Católica del Perú, con título "ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART DE CLIENTES PARA EL AREA DE MARKETING DE UNA ENTIDAD ASEGURADORA", gestiona la información orientando el manejo de datos hacia una arquitectura propia de Datawarehouse, con un diseño flexible para afrontar el problema de duplicación de clientes existente en la organización.

Noviembre del 2011, Zambrano Alarcón, Jaime Alexander en su Tesis para optar por el Título de Ingeniero Informático en la Pontificia Universidad Católica del Perú, con título "ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL ÁREA DE MANTENIMIENTO Y LOGÍSTICA DE UNA EMPRESA DE TRANSPORTE PÚBLICO DE PASAJEROS" en esta tesis el objetivo es, proveer una solución de BI que da soporte a las necesidades de información a los usuarios finales del área de mantenimiento y logística.

2.2. BASES TEÓRICAS

2.2.1. Necesidad de un Datawarehouse

Una de las claves de éxito de las organizaciones modernas es el acceso a la información correcta, en el tiempo adecuado, en el lugar correcto y en la forma adecuada. Un Datawarehouse permite a los usuarios acceder a grandes volúmenes de datos operacionales, integrados para analizar y responder a necesidades de planeación, predicción y toma de decisiones.

Generalmente, los sistemas transaccionales usan estructuras normalizadas, en las cuales se optimizan las inserciones y actualizaciones, pero es menos probable que el sistema se organice de forma tal que produzca reportes eficientes para datos resumidos con cierta jerarquía. Y es aquí donde debería usarse el DWH, que usa los datos relevantes de fuentes existentes y los combina en una estructura que ha sido optimizada para las selecciones.

2.2.2. Conceptos de Datawarehouse

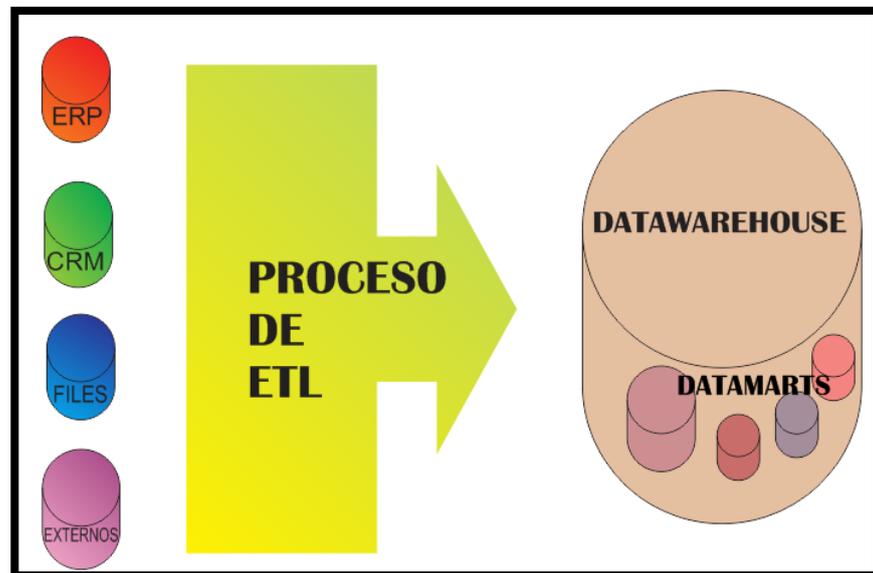
“Es un conjunto de base de datos relacionados, que se han reunido y almacenado en conjunto para sacar de ellos el máximo provecho. La idea fundamental del almacenamiento de datos es reunir tanta información como sea posible con la esperanza que pueda surgir un panorama más significativo”¹

“Es el centro de la arquitectura para los sistemas de información en la década de los '90. Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se

¹ Diccionario de términos de Computación, Bryan Pfaffenberger. Pág. 136

necesitan para el procesamiento analítico, informático sobre una amplia perspectiva de tiempo.”²

Figura 1: Esquema de un Datawarehouse



Fuente: Elaboración propia.

2.2.2.1. Beneficios de un Datawarehouse

- Integra y globaliza la información del negocio para proporcionar una herramienta de soporte a la toma de decisiones en cualquier área funcional.
- 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 relación con el cliente.

² <http://www.programacion.com/bbdd/tutorial/warehouse/> 2 agosto 2004

Cuadro 1: Cuadro de Comparación entre BD_Transaccional y un DW

Aspectos	BD operacional	Data Warehouse
Objetivo	De tipo operativo (operaciones del día a día)	Análisis y toma de decisiones
Proceso	De transacciones. Repetitivo y conocido.	De consultas masivas. Puntual y no conocido.
Actividad	Predomina la actualización	Predomina la consulta
Rendimiento	Importancia del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva.
Explotación	Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa relacionada con el negocio
Volatilidad	Actualizable	Carga, pero no actualización
Usuarios	Usuarios de perfiles medios o bajos	Usuarios de perfiles altos
Organización	Estructura normalmente relacional	Visión multidimensional
Granularidad	Datos generales desagregados, al detalle	Datos en distintos niveles de detalle y agregación
Horizonte histórico	30 a 90 días	5 a 10 años
Perspectiva	Importancia del dato actual	Importancia del dato histórico
Volumen de datos	Pequeño/medio. Del orden del Mb a Gb.	Medio/grande. Del orden del Gb a Tb.

Fuente: <http://anabuigues.com/2010/01/14/data-warehouse-y-las-bases-de-datos-operacionales/>

2.2.3. Data Mart

“Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un

Datamart puede ser alimentado desde los datos de un Datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información”.³

Existen dos clases de Datamart:

- a. Dependientes.** Los datos utilizados para alimentar el Datamart provienen del Datawarehouse. Esta estrategia es particularmente apropiada cuando el Datawarehouse crece muy rápidamente y los distintos departamentos requieren sólo una pequeña porción de los datos contenidos en él.
- b. Independientes.** Integra por sí mismo un compendio de distintas fuentes de información, sistemas operacionales y/o fuentes externas.

Los Datamarts que mantienen estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información
- Facilidad para la historización de los datos

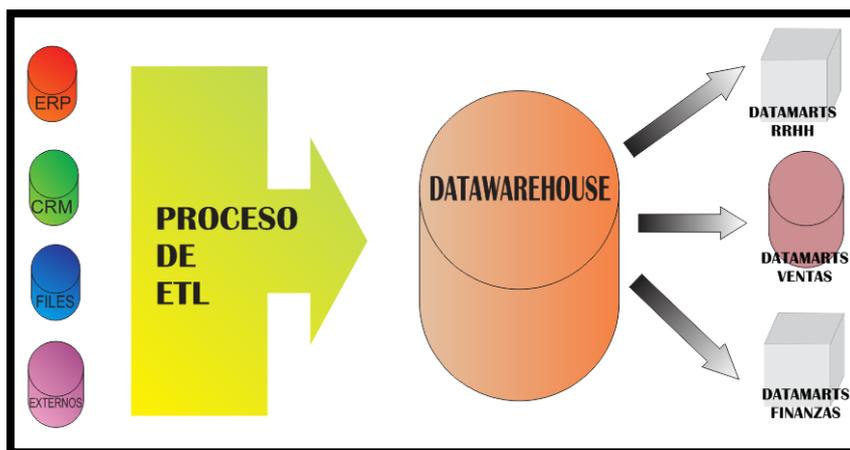
2.2.4. Paradigmas de Datawarehouse y Datamart según Inmon y Kimball

2.2.4.1. Bill Inmon

Define un Datawarehouse como *“un conjunto de datos integrados, históricos, variantes en el tiempo y unidos alrededor de un tema específico, que es usado por la gerencia para la toma de decisiones”*, defiende una metodología descendente (**top-down**) a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos corporativos, es decir, los Datamarts se crearán después de haber terminado el Datawarehouse completo de la organización.

³ http://www.sinnexus.com/business_intelligence/datamart.aspx Junio 2012

Figura 2: Esquema de un Datawarehouse según Bill Inmon

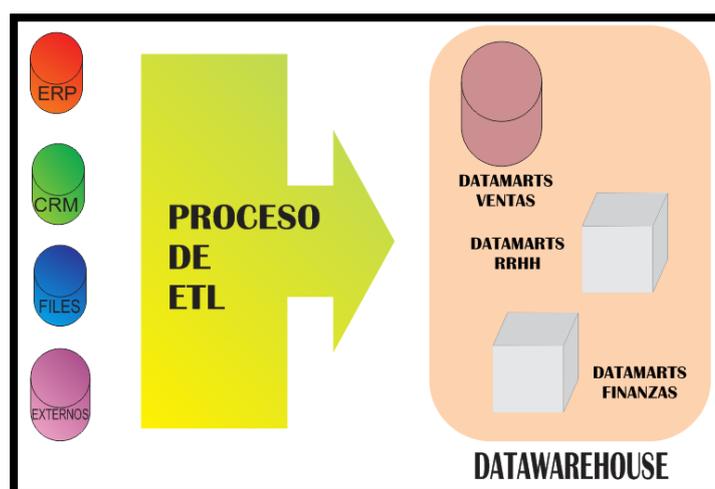


Fuente: Elaboración propia

2.2.4.2. Ralph Kimball

Define un Datawarehouse como "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". Kimball determinó que un Data Warehouse no era más que: "la unión de todos los Data Marts de una entidad". Defiende por tanto una metodología ascendente (**bottom-up**) a la hora de diseñar un almacén de datos.

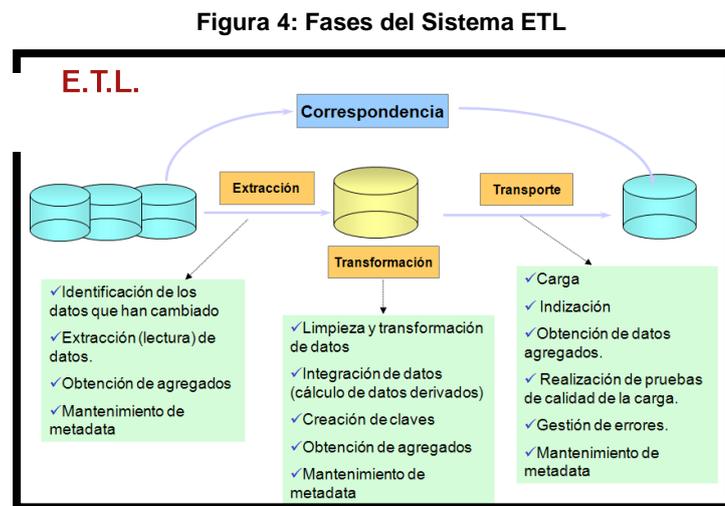
Figura 3: Esquema de un Datawarehouse según Ralph Kimball



Fuente: Elaboración propia

2.2.5. ETL (Extract, Transform and Load)

“Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, Datamart, o Datawarehouse, para analizar, o en otro sistema operacional para apoyar un proceso de negocio. Son todas las actividades necesarias relacionadas a la administración de datos y metadatos para satisfacer las necesidades de información.”⁴



Fuente: <http://www.carlosproal.com/dw/dw05.html>

2.2.5.1. Fases del ETL

a. Extracción

Para llevar a cabo correctamente el proceso de extracción del ETL, se debe considerar lo siguiente:

- Extraer los datos desde los sistemas de origen.
- Analizar los datos extraídos obteniendo una revisión.
- Interpretar esta revisión para verificar que los datos extraídos cumplen la pauta o estructura que se esperaba. Si no fuese así, los datos deberían ser rechazados.
- Convertir los datos a un formato preparado para iniciar el proceso de transformación

⁴ http://prezi.com/ri9m9dmz_q8b/etl/ Alfredo Ramírez. 2013

Se debe exigir siempre que esta tarea cause un impacto mínimo en el sistema de origen ya que, si los datos a extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que no pudiera volver a ser utilizado con normalidad.

b. Transformación

Esta fase aplica una serie de reglas de negocio o funciones, sobre los datos extraídos para convertirlos en datos que serán cargados. Estas directrices pueden ser declarativas, pueden basarse en excepciones o restricciones pero, para potenciar su eficacia, hay que asegurarse de que sean:

- Declarativas.
- Independientes.
- Claras.
- Inteligibles.
- Con una finalidad útil para el negocio.

c. Carga

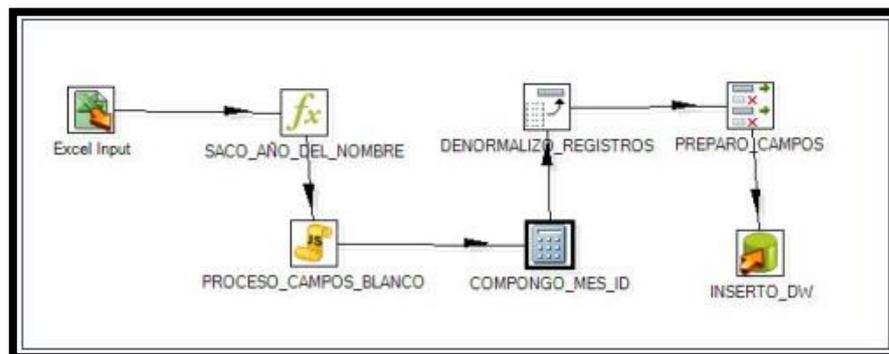
En esta fase, los datos procedentes de la fase anterior son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes.

Existen dos formas básicas de desarrollar el proceso de carga:

Acumulación simple: esta manera de cargar los datos consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el Datawarehouse, almacenando un valor calculado que consistirá típicamente en un sumatorio o un promedio de la magnitud considerada. Es la forma más sencilla y común de llevar a cabo el proceso de carga.

Rolling: este proceso sería el más recomendable en los casos en que se busque mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada.

Figura 5: Ejemplo de un Proceso de ETL



Fuente: <http://churriwifi.wordpress.com/2010/05/30/16-3-5-etl-carga-de-hechos-de-presupuestos-con-pdi/>

2.2.6. Modelado Multidimensional

“El modelado multidimensional provee un camino viable para agregar hechos a lo largo de múltiples atributos, llamados dimensiones. Los datos son almacenados como hechos y dimensiones en un modelo de datos relacional”.⁵

Otra definición nos explica: “Es un sistema de software diseñado para permitir el almacenamiento y la recuperación eficiente de grandes volúmenes de datos íntimamente relacionados y almacenados, vistos y analizados desde diferentes perspectivas (llamadas dimensiones). Un Modelo Multidimensional es un componente fundamental de una base de datos multidimensional es un arreglo o matriz de n dimensiones. Busca presentar la información de una manera estándar, sencilla y sobre todo intuitiva para los

⁵ Bellatreche, Karlapalem, Mohania. 2001

usuarios, además de que permite accesos a la información mucho más rápida por parte de los manejadores de bases de datos.”⁶

Se debe tener en cuenta al llenar un modelo dimensional:

- No albergar en el modelo dimensional datos nulos.
- Cada dimensión tiene un código auto correlativo independiente que es su llave principal (IDENTITY).
- Los datos deben estar estandarizados. Este tipo de conversión se hacen en el proceso de ETL Extracción, transformación y carga. Generalmente es un DTS el encargado de hacer ese trabajo.
- La forma de organizar los datos como hechos y dimensiones es mediante tablas.

2.2.6.1. Hechos

Contiene información de datos reales, de medidas, de hechos en bruto. Estas tienden a contener gran volumen de información ya que contienen la información histórica de todos los hechos de un determinado tipo que se han producido en una empresa. Son las que contiene la información a consultar y formarían el núcleo del Datawarehouse construido.

2.2.6.2. Dimensiones

Contienen información dimensional, que permite filtrar, organizar, agregar, etc. la información contenida en las tablas de hechos. Son típicamente tablas que contienen información des-normalizada de un concepto respecto con el cual se relacionan los hechos individuales. Las dimensiones podríamos definir las como diferentes puntos de vista por los que podemos analizar la información.

⁶ http://prezi.com/q0yca7u0v0_3/modelo-multidimensional-2/ Francisco Quirós Fallas. Enero 2013

2.2.6.3. Medidas

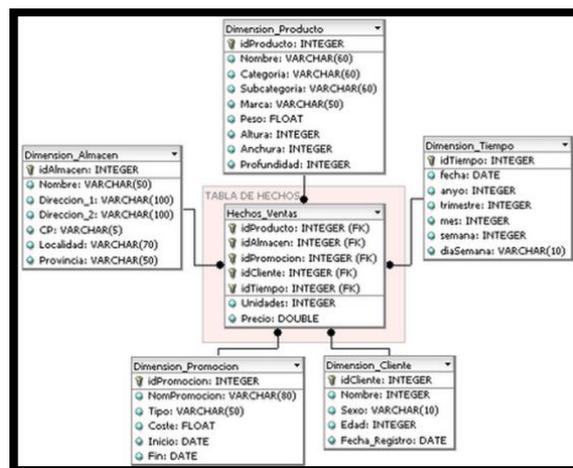
Una medida es un tipo de dato cuya información es usada por los analistas (usuarios) en sus query's para medir la performance del comportamiento de un proceso o un objeto del negocio. Por ejemplo: Cantidades, tamaños, montos, duraciones, etc.

2.2.6.4. Esquemas

a. Esquema Estrella

“El modelo estrella presenta una tabla de hechos rodeado de una serie de dimensiones en profundidad simple, es decir, sin relaciones con otras tablas más allá de la misma dimensión y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP”⁷. Tiene como ventaja mayor velocidad de procesamiento, al prescindir de operaciones “join” para relacionar a las demás tablas. Su desventaja es el almacenamiento por la des-normalización, pero esto es poco crítico.

Figura 6: Esquema modelado de Datos Tipo Estrella



Fuente: <http://jcesar-27.blogspot.com/2011/09/datawarehouse-esquemas-de-estrella-y.html>

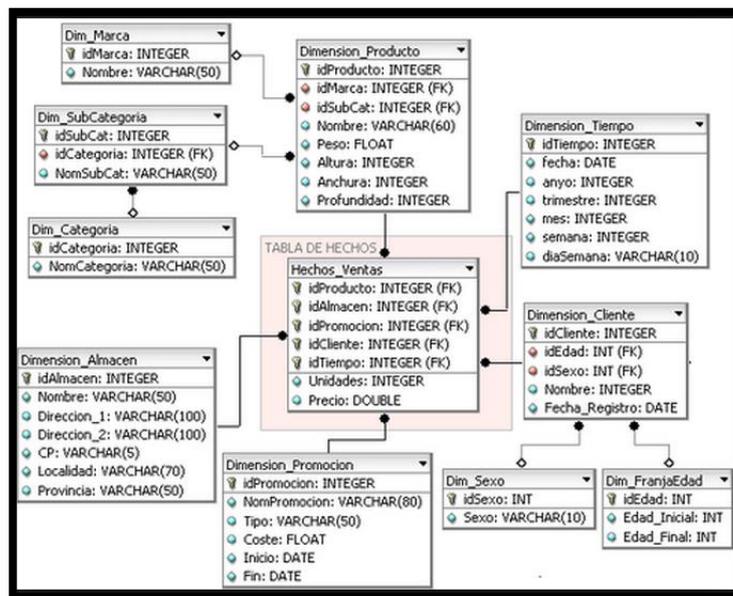
⁷ http://etl-tools.info/es/bi/almacenedatos_esquema-estrella.htm, 2006

b. Esquema en copo de nieve

“El 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. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas joins gracias a que las dimensiones de análisis se representan ahora en tablas de dimensión normalizadas.”⁸

El problema es que para extraer datos de las tablas en esquema de copo de nieve, a veces hay que vincular muchas tablas en las sentencias SQL que puede llegar a ser muy complejo y difícil para mantener.

Figura 7: Esquema modelado de Datos Tipo Copo de Nieve



Fuente: <http://jcesar-27.blogspot.com/2011/09/datawarehouse>

-esquemas-de-estrella-y.html

⁸ Blog: <http://informationmanagement.wordpress.com/category/data-warehouse/esquema-en-copo-de-nieve/> por José Curto 2007

2.2.6.5. OLAP (On-Line Analytical Processing)

“OLAP es una categoría de tecnología de software. Se basa en el uso de bases de datos multidimensionales para proveer a analistas, administradores y ejecutivos una visión rápida, consistente e interactiva de una amplia variedad de información. Se diferencia al uso exclusivo de un Datawarehouse en que provee acceso rápido a información estratégica para su análisis posterior”.⁹

- OLAP es un modelo multidimensional y permite ver los datos a través de dimensiones o múltiples filtros.
- Proporciona un modelo de datos intuitivo y conceptual.
- Es rápido; Permitiendo así a los gerentes y responsables preguntarse y resolver situaciones en un corto período de tiempo.
- Los sistemas OLAP tienen un motor de cálculo robusto para manejar las necesidades de cálculo especializado que una estructura multidimensional impone. Este motor de cálculo de OLAP organiza los datos en una forma que permite a los analistas escribir sencillas y directas fórmulas que se ejecutan a través de múltiples dimensiones con sólo unas pocas líneas de código.

2.2.6.6. MOLAP (Multidimensional OLAP)

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis. Su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente.

⁹ Forsman, 1997.

2.2.6.7. ROLAP (Relational OLAP)

La arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales. La arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del Datawarehouse, y soporta técnicas de optimización de accesos para acelerar las consultas.

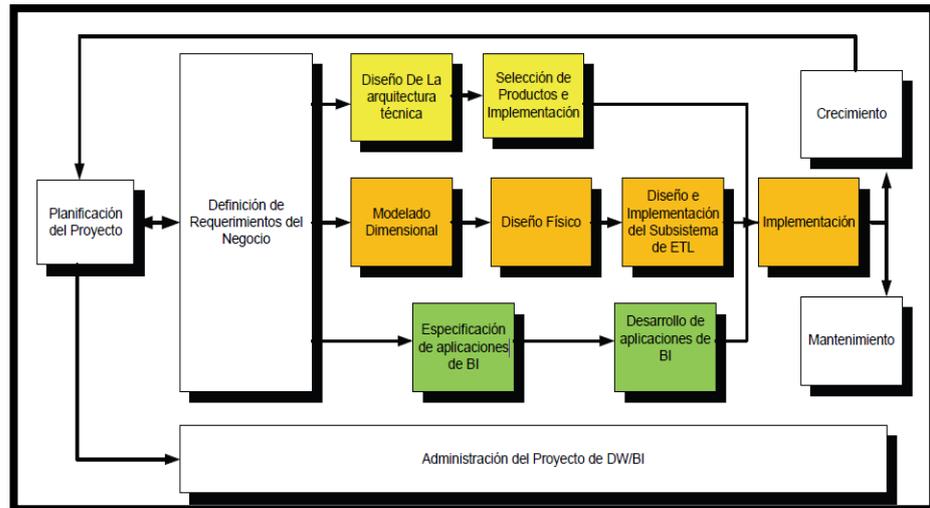
2.2.7. Metodología Kimball

Esta Metodología comprende 4 principios básicos:

- **Centrarse en el negocio:** Esto es identificar los requerimientos del Negocio y desarrollar una relación sólida con este.
- **Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará los requerimientos del negocio.
- **Realizar entregas en incrementos significativos:** crear el almacén de datos (DW) en incrementos entregables. En esto la metodología se parece a las metodologías ágiles de construcción de software.
- **Ofrecer la solución completa:** proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible.

La construcción de una solución de DW / BI (Datawarehouse / Business Intelligence) es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. La siguiente imagen detalla los pasos a seguir:

Figura 8: Metodología de Ralph Kimball



Fuente : Tareas de la Metodología de Kimball, denominada Business Dimensional Lifecycle (Kimball et al 98,08,Mundy & Thornthwaite 06)

2.2.7.1. Planificación del Proyecto

Como metodología, en esta etapa propone identificar el alcance preliminar basándose en los requerimientos del negocio y no en fechas límites, construyendo la justificación del proyecto en términos del negocio.

A nivel de planificación del proyecto se establece la identidad del mismo, el personal (los usuarios, gerentes del proyecto, equipo del proyecto), desarrollo del plan del proyecto, el seguimiento y la monitorización.

2.2.7.2. Definición de los Requerimientos del Negocio

Un factor determinante en el éxito de un proyecto de DWH es la interpretación correcta de los diferentes niveles de requerimientos expresados por los distintos grupos de usuarios.

Los diseñadores de los DWH deben entender los factores claves que guían el negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.

2.2.7.3. Modelado Dimensional

La definición de los requerimientos del negocio determina los datos necesarios para cumplir los requerimientos analíticos de los usuarios. Diseñar los modelos de datos para soportar estos análisis requiere un enfoque diferente al usado en los sistemas operacionales. Básicamente, se comienza con una matriz donde se determina la dimensionalidad de cada indicador y luego se especifican los diferentes grados de detalle dentro de cada concepto del negocio, así como la granularidad de cada indicador y las diferentes jerarquías que dan forma al modelo dimensional del negocio.

2.2.7.4. Diseño Físico

El diseño físico de la base de datos se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Un elemento principal de este proceso es la definición de estándares del entorno de la base de datos. La indexación y las estrategias de particionamiento se determinan también en esta etapa.

En la estrategia de particionamiento o agregación, el DWH tiene, y debe tener, todo el detalle de información en su nivel atómico. Las tablas agregadas deben ser versiones reducidas de las dimensiones asociadas con la granularidad de la tabla de hechos y añadir los indicadores de las tablas de detalle a un nivel superior.

La existencia de estas tablas agregadas debe ser completamente transparente para el usuario de negocio. Es decir, tanto el jefe de área como el jefe de producto trabajarán con él. Por otro lado, en la estrategia de indexación los índices son estructuras opcionales optimizadas y orientadas a conjuntos de operaciones. Según Ralph Kimball, las tablas de dimensión deben tener un único índice sobre las claves primarias y sería

recomendable que el índice estuviera compuesto de un único atributo.

2.2.7.5. Diseño y Desarrollo de la Presentación de Datos

Esta etapa es típicamente la más subestimada de las tareas en un proyecto de DWH. Las principales actividades de esta fase del ciclo de vida son: la extracción, la transformación y la carga (ETL process). Se definen como procesos de extracción aquellos requeridos para obtener los datos que permitirán efectuar la carga del Modelo Físico diseñado. Así mismo, se definen como procesos de transformación los procesos para convertir o recodificar los datos fuente a fin de poder efectuar la carga efectiva del Modelo Físico. Por otra parte, los procesos de carga de datos son los procesos requeridos para poblar el DWH.

Todas estas tareas son altamente críticas pues tienen que ver con la materia prima del DWH: los datos. La desconfianza y pérdida de credibilidad del DWH provocará efectos inmediatos e inevitables si el usuario se encuentra con información inconsistente. Es por ello que la calidad de los datos es un factor determinante en el éxito de un proyecto de DWH. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente. Para cumplir con estas premisas es necesario tener en cuenta ciertos parámetros a la hora de desarrollar las tablas de dimensión y la tabla de hechos.

2.2.7.6. Diseño de la Arquitectura Técnica

Los entornos de DWH requieren la integración de numerosas tecnologías. Se deben tener en cuenta tres factores: los requerimientos del negocio, los actuales entornos técnicos y las directrices técnicas y estratégicas futuras planificadas por la compañía para poder establecer el diseño de la arquitectura técnica del entorno de algunos equipos de trabajo.

2.2.7.7. Selección de Productos e Instalación

Utilizando el diseño de arquitectura técnica como marco es necesario evaluar y seleccionar los componentes específicos de la arquitectura, como la plataforma de hardware, el motor de base de datos, la herramienta de ETL, las herramientas de acceso, etc.

Una vez evaluados y seleccionados los componentes determinados se procede con la instalación y prueba de los mismos en un ambiente integrado de DWH.

2.2.7.8. Especificación de Aplicaciones para Usuarios Finales

No todos los usuarios del DWH necesitan el mismo nivel de análisis. Es por ello que en esta etapa se identifican los roles o perfiles de usuarios para los diferentes tipos de aplicaciones necesarias en base al alcance de los perfiles detectados (gerencial, analista del negocio, vendedor, etc.)

2.2.7.9. Desarrollo de Aplicaciones para Usuarios Finales

A continuación de la especificación de las aplicaciones para usuarios finales, el desarrollo de las aplicaciones de los usuarios finales involucra configuraciones de los metadatos y construcción de reportes específicos.

Los usuarios acceden al DWH por medio de herramientas de productividad basadas en GUI (Graphical User Interface). De hecho existen multitud de estas herramientas con las que proveer a los usuarios. Las herramientas pueden incluir software de consultas, generadores de reportes, procesamiento analítico en línea o herramientas de Datamining dependiendo de los tipos de usuarios y sus requerimientos particulares. Sin embargo, una sola herramienta puede no satisfacer todos los requerimientos, por lo que quizás sea necesaria la integración de herramientas hechas bajo petición expresa de los usuarios para satisfacer sus necesidades de consulta sobre el DWH.

2.2.7.10. Implementación

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesibles para el usuario del negocio. Hay varios factores extras que aseguran el correcto funcionamiento de todos estos elementos, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación y las estrategias de feedback. Todas estas tareas deben tenerse en cuenta antes de que cualquier usuario pueda tener acceso al DWH.

2.2.7.11. Mantenimiento y crecimiento

Como se remarca siempre, la creación de un DWH es un proceso (de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral) que acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con las actualizaciones de forma constante para poder seguir la evolución de las metas por conseguir.

Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

2.2.7.12. Gestión del Proyecto

La gestión del proyecto asegura que las actividades del ciclo de vida se lleven a cabo de manera sincronizada. La gestión del proyecto acompaña todo el ciclo de vida. Entre sus actividades principales se encuentra la monitorización del estado del proyecto y el acoplamiento entre los requerimientos del negocio y las restricciones de los sistemas de información para poder manejar correctamente las expectativas en ambos sentidos.

2.3. MARCO CONCEPTUAL

- **Análisis de Datos:**

Es el proceso de inspeccionar, limpiar y transformar datos con el objetivo de resaltar información útil, lo que sugiere conclusiones, y apoyo a la toma de decisiones.

- **Análisis Gerencial:**

Es de naturaleza estratégica y no operativa, pues con este se examina la agregación de valor de cada una de las funciones mayores de la empresa. A partir del conocimiento de la estrategia empresarial se determina la estructura organizacional, con la cual se pretende llevar a cabo la estrategia, para dar cumplimiento a la misión y objetivos mayores de la compañía.

- **Análisis Predictivo:**

Se basan en el análisis de los datos actuales e históricos para hacer predicciones sobre futuros eventos. Dichas predicciones raramente suelen ser afirmaciones absolutas, pareciéndose más a eventos y su probabilidad de que suceda en el futuro.

En el mundo de los negocios los modelos predictivos explotan los patrones de comportamiento encontrados en el pasado para poder identificar riesgos y oportunidades. Los modelos capturan las relaciones entre muchos factores permitiendo capturar riesgos potenciales asociados a un conjunto de condiciones, guiando así a la toma de decisiones.

- **Analysis Services:**

Es una herramienta de Microsoft Visual Studio que permite la elaboración de objetos de análisis multidimensional llamado cubos.

- **Área funcional:**

Conjunto de actividades realizadas por las personas que tienen asignado un objetivo de la producción y explicitan una función del sistema organizativo

- **Arquitectura:**

Tiene que ver con el diseño y la implementación de estructuras de software de alto nivel. Es el resultado de ensamblar un cierto número de elementos arquitectónicos de forma adecuada para satisfacer la mayor funcionalidad y requerimientos de desempeño de un sistema, así como requerimientos no funcionales, como la confiabilidad, escalabilidad, portabilidad, y disponibilidad.

- **Atributos:**

Son las características por medio de las cuales se puede describir una entidad. Por ejemplo: Cliente tiene como atributo: nombre, dirección, edad, etc.

- **Audiencia:**

Número de individuos destinatarios que reciben mensajes a través de un medio de comunicación y que, en ocasiones, interactúa con ellos. Las audiencias suelen dividirse según diversas variables, como la edad o el sexo, para determinar los contenidos que se les ofrecen.

- **Base de Datos (BD):**

Es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

- **Competidores:**

Aquellos negocios que ofrecen productos o servicios que satisfacen la misma necesidad en el cliente y aun cuando lo hagan de diferente manera o estén localizados dentro o no de una misma localidad.

- **Convergencia:**

Unión de dos o más objetos que confluyen en un mismo punto.

- **Cubo:**

Las bases de datos multidimensionales dependen de estructuras llamadas cubos. Un cubo es una colección de medidas y dimensiones. Este puede tener “n” dimensiones, las medidas dentro de un cubo son evaluadas en la intersección de todas las “n” dimensiones. Los cubos permiten la agregación a través de jerarquías dimensionales y la navegación hacia arriba /abajo rápida, siendo así mucho más flexible que una construcción basada en tablas.

- **Datamart:**

Es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento.

- **Data Mining (Minería de datos):**

Es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.

- **Datawarehouse (DWH):**

Es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se

utiliza. Se trata, sobre todo, de un expediente completo de una organización, más allá de la información transaccional y operacional, almacenada en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos

- **Datos Operacionales:**

Son datos que tienen que ver con un área temática específica. Un almacén de datos operacionales puede ser utilizado como un área transitoria para la reorganización física de los datos operacionales extraídos, para proporcionar informes operacionales y apoyar la toma de decisiones operacionales. También pueden servir como un punto de consolidación si es que los datos operacionales proceden de varias fuentes.

- **Depuración de datos:**

Es el descubrimiento, corrección o eliminación de datos erróneos de una base de datos. Este proceso permite identificar datos incompletos, incorrectos, inexactos, no pertinentes, etc. y luego substituir, modificar o eliminar estos datos sucios. Después de la limpieza, la BD podrá ser compatible con otras bases de datos similares en el sistema.

- **Dimensión:**

Llamada también lookup table (tabla de búsqueda). Es un elemento conformado por componentes que en su conjunto caracterizan a dicho elemento. Una dimensión puede ser el tiempo, el cual está compuesto por día, mes, año, semestre, etc. Una dimensión se representa por una tabla donde sus columnas indican cada uno de sus componentes.

- **DTS:**

Muchas organizaciones tienen que centralizar sus datos para mejorar la toma de decisiones corporativas. Sin embargo, los datos podrían estar almacenados en muchos formatos diferentes y en distintas

ubicaciones. Los Servicios de transformación de datos (DTS) permiten satisfacer esta necesidad crucial de la empresa. Para ello, proporcionan un conjunto de herramientas que permiten extraer, transformar y consolidar datos de distintos orígenes en uno o varios destinos compatibles con la conectividad DTS.

- **Empresa líder:**

Aquella que marca la pauta en cuanto a precio, innovaciones, publicidad, etc., siendo normalmente imitada por el resto de los actuantes en el mercado.

- **Estrategia:**

Es el desarrollo de una amplia fórmula de cómo la empresa va a competir, cuáles deben ser sus objetivos y qué políticas serán necesarias para alcanzar tales objetivos.

- **ETL:**

Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, Datamart, o Datawarehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

- **Explotación de Información:**

Es una sub-disciplina de la Informática que aporta herramientas de Minería de Datos para extraer el conocimiento no trivial que se encuentra (implícitamente) en los datos disponibles de diferentes fuentes de información dentro del marco de un proyecto ingenieril.

- **Fidelizar:**

Se refiere a la fidelización de los clientes. La fidelización es el fenómeno por el que un público determinado permanece fiel a la compra de un producto concreto de una marca concreta, de una forma continua o periódica.

- **Fuente de Información:**

Cualquier origen de información susceptible de ser representado mediante una señal analógica y/o digital. El objetivo es poder procesar, almacenar o transmitir la información que supone las alteraciones del medio. Así como diversos tipos de documentos que contienen datos útiles.

- **GUI (Graphical User Interface):**

Es un programa informático que actúa de interfaz de usuario, utilizando un conjunto de imágenes y objetos gráficos para representar la información y acciones disponibles en la interfaz. Su principal uso, consiste en proporcionar un entorno visual sencillo para permitir la comunicación con el sistema operativo de una máquina o computador.

- **Granularidad:**

Se refiere a la especificidad a la que se define un nivel de detalle en una tabla, es decir, si hablamos de una jerarquía la granularidad empieza por la parte más alta de la jerarquía, siendo la granularidad mínima, el nivel más bajo. En Datawarehouse, no sólo existe granularidad para las tablas de hechos (Fact's), también existe granularidad en las dimensiones.

- **Historización:**

Acción de almacenar data histórica en una BD.

- **Indexación:**

Tiene como propósito ejecutar la elaboración de un índice que contenga de forma ordenada la información, esto con la finalidad de obtener resultados de forma sustancialmente más rápida y relevante al momento de realizar una búsqueda. Es por ello que la indexación

es un elemento fundamental de elementos como los motores de búsqueda y las bases de datos.

- **Indicador:**

Son claves que constituyen un tipo de medida cuyo objetivo es proporcionar información sobre el rendimiento de una actividad o sobre la consecución de una meta.

- **Integrado:**

Hace referencia a que la BD debe contener los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

- **Integration Services:**

Tecnología Microsoft, introducida con SQL Server 2005, que permite implementar procesos ETL, entre otros.

- **Inteligencia de Negocios (IN o BI):**

Se denomina inteligencia de negocios o BI (del inglés Business Intelligence) al conjunto de estrategias y llaves enfocadas a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o empresa.

- **Jerarquía:**

Las jerarquías pueden existir en una dimensión en la cual sirven como navegaciones predefinidas y están compuestas de uno o más niveles. Una dimensión puede tener una o más jerarquías.

- **Join:**

Permite combinar registros de dos o más tablas en una base de datos relacional. En el Lenguaje de Consultas Estructurado (SQL) hay tres tipos de JOIN: interno, externo y cruzado.

- **Metadatos:**

Información no relevante para el usuario final pero sí de suma importancia para el sistema que maneja la data. Los metadatos son enviados junto a la información cuando se realiza alguna petición o actualización de la misma.

- **Microsoft Excel:**

Es una aplicación distribuida por Microsoft Office para hojas de cálculo. Este programa es desarrollado y distribuido por Microsoft, y es utilizado normalmente en tareas financieras y contables.

- **Metodología de Bill Inmon:**

Define una metodología descendente top-down (hacia abajo) donde los Datamarts se crearán después de haber terminado el Datawarehouse completo de la organización.

- **Metodología de Ralph Kimball:**

Determina que un Datawarehouse es la unión de todos los Datamarts de una organización. Define la metodología bottom-up (hacia arriba).

- **Microsoft Visual Studio:**

Visual Studio permite a los desarrolladores crear aplicaciones, sitios y aplicaciones web, así como servicios web en cualquier entorno que soporte la plataforma .NET (a partir de la versión .NET 2002). Así se pueden crear aplicaciones que se comuniquen entre estaciones de trabajo, páginas web, dispositivos móviles, dispositivos embebidos, consolas, etc.

- **Modelización:**

Es una técnica orientada por los datos que representa un sistema en función de sus datos, independientemente de cómo se procesen dichos datos para producir información. Si se capturan los datos y se

almacenan en estructuras flexibles de archivos y bases de datos, todas las necesidades actuales y futuras de información pueden satisfacerse por medio del uso de dichos datos.

- **Modelo copo de nieve:**

En las bases de datos usadas para Data Warehousing, un esquema en copo de nieve es una estructura más compleja que el esquema en estrella. Se da cuando existen un gran número de tablas de hechos sin que sea factible reducir su número. Aunque puede reducir espacio, tiene la contrapartida de tener menor rendimiento al tener que crear más tablas de dimensiones y más relaciones entre las tablas.

- **MOLAP (Multidimensional Online Analytical Processing):**

Se trata de una alternativa a la tecnología ROLAP (OLAP-Relacional). Aunque ambos tipos de herramientas están diseñados para realizar análisis de datos a través de un modelo de datos multidimensional, MOLAP se diferencia significativamente en que requiere un pre-procesamiento y almacenamiento de la información contenida en el cubo OLAP.

MOLAP almacena estos datos en una matriz de almacenamiento multidimensional optimizado, más que en una base de datos relacional (o en un ROLAP).

- **Modelo entidad relación:**

Constituye una forma de representar conceptualmente la realidad mediante su abstracción en entidades y relaciones. De esta forma intentamos representar el mundo que nos rodea, los datos de nuestro problema mediante una serie de entidades que representan objetos o conceptos así como las relaciones que se dan entre ellos tales como su uso y composición.

- **Modelo estrella:**

Es una arquitectura de almacén de datos simple. En este diseño del almacén de datos la tabla de variables (hechos) está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP.

- **Multidimensional:**

Son bases de datos multidimensionales que se utilizan principalmente para crear aplicaciones OLAP y pueden verse como bases de datos de una sola tabla, su peculiaridad es que por cada dimensión tienen un campo (o columna), y otro campo por cada métrica o hecho.

- **Normalización:**

Es el proceso de organizar los datos de una base de datos. Incluye la creación de tablas y el establecimiento de relaciones entre ellas según reglas diseñadas tanto para proteger los datos como para hacer que la base de datos sea más flexible al eliminar la redundancia y las dependencias incoherentes.

- **OLAP (On-Line Analytical Processing):**

Es una solución utilizada en el campo de la llamada Inteligencia empresarial (o 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 o Sistemas Transaccionales (OLTP).

- **Organización Privada:**

Empresa comercial que es propiedad de inversores privados, no gubernamentales, accionistas o propietarios (generalmente en conjunto, pero puede ser propiedad de una sola persona).

- **Proactivo:**

Sistema de apoyo a la toma de decisiones en el desarrollo de acciones creativas y audaces para generar mejoras.

- **Procesamiento informático:**

Ejecución de diversas instrucciones por parte del microprocesador, de acuerdo a lo que indica un programa.

- **Recopilación de datos:**

Se refiere al uso de una diversidad de técnicas y herramientas que pueden ser utilizadas por el analista para desarrollar los sistemas de información, los cuales pueden ser la entrevistas, la encuesta, el cuestionario, la observación, el diagrama de flujo y el diccionario de datos.

- **Registro:**

En el contexto de una base de datos relacional, un registro (también llamado fila o tupla) representa un objeto único de datos implícitamente estructurados en una tabla.

- **Reportes:**

Un reporte es un Documento, generado por el Sistema, que nos presenta de manera Estructurada y/o Resumida, datos relevantes guardados o generados por la misma aplicación de tal manera que se vuelvan útiles para los fines de la empresa.

- **Repositorio:**

Es un sitio centralizado donde se almacena y mantiene información digital, habitualmente bases de datos o archivos informáticos.

- **ROLAP (Procesamiento Analítico Online Relacional):**

Se trata de sistemas y herramientas OLAP (Procesamiento Analítico Online) construidos sobre una base de datos relacional.

- **Sector:**

Es la división de la actividad económica de un Estado o territorio, atendiendo al tipo de proceso productivo que tenga lugar.

- **Sincronización:**

Describe el ajuste temporal de eventos. Se habla de sincronización cuando determinados fenómenos ocurran en un orden predefinido o a la vez

- **Sistema de Gestión Integral:**

Conjunto formado por la estructura de la organización, las responsabilidades, los procedimientos, los procesos y los recursos que se establecen para llevar a cabo la gestión integral de los sistemas.

- **Sistema Gestor de Base de Datos (SGBD o DBMS)**

Permiten almacenar y posteriormente acceder a los datos de forma rápida y estructurada.

- **Sistema Informático:**

Es el conjunto de partes interrelacionadas, hardware, software y de recurso humano que permite almacenar y procesar información.

- **Sistemas Transaccionales:**

Es un tipo de sistema de información diseñado para recolectar, almacenar, modificar y recuperar todo tipo de información que es generada por las transacciones en una organización. Una transacción es un evento o proceso que genera o modifica la información que se

encuentran eventualmente almacenados en un sistema de información.

- **Software:**

Conjunto de los componentes lógicos necesarios que hacen posible la realización de tareas específicas, en contraposición a los componentes físicos que son llamados hardware.

- **SQL:**

Es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en ellas. Una de sus características es el manejo del álgebra y el cálculo relacional que permiten efectuar consultas con el fin de recuperar de forma sencilla información de interés de bases de datos, así como hacer cambios en ella.

- **Tabla de Hechos:**

Llamada también Fact Table. Es la tabla central que representa datos numéricos en el contexto de las entidades del negocio.

- **Tecnología de Información:**

Utilización de tecnología, específicamente computadoras y ordenadores electrónicos, para el manejo y procesamiento de información, específicamente la captura, transformación, almacenamiento, protección, y recuperación de datos e información.

- **Tendencia:**

En un sentido general, es un patrón de comportamiento de los elementos de un entorno particular durante un período. En términos del análisis técnico, la tendencia es simplemente la dirección o rumbo del mercado.

- **Toma de decisiones:**

Es el proceso mediante el cual se realiza una elección entre las opciones o formas para resolver diferentes situaciones de la empresa (utilizando metodologías cuantitativas que brinda la administración). La toma de decisiones consiste, básicamente, en elegir una opción entre las disponibles, a los efectos de resolver un problema actual o potencial (aun cuando no se evidencie un conflicto latente).

- **Ventaja Competitiva:**

Se denomina ventaja competitiva a una ventaja que una compañía tiene respecto a otras compañías competidoras.

CAPITULO III

DESCRIPCIÓN, DISEÑO Y CONSTRUCCIÓN DEL MODELO

3.1. ANÁLISIS DEL MODELO

3.1.1. La Empresa

INPECHI, Importaciones y Negocios Perú – China se constituye como una Organización Privada, cuyo objetivo está orientado a realizar capacitaciones, asesorías y consultorías de tal manera que sus clientes puedan realizar su negociación de Importación con China con la mayor seguridad.

Figura 9: Logo de la empresa Inpechi Consultores Internacionales S.A.C.



Fuente: <http://importardechina.inpechi.org/>

a. Misión

Brindar soluciones en gestión, negociación y comercio exterior a micro, pequeñas y medianas empresas, que las impulse a ser más productivas, rentables y competitivas a través de consultoría, capacitación y servicios externos (Outsourcing) en materia de Importaciones y Negocios con calidad, seguridad y rentabilidad.

b. Visión

Ser una empresa con una continua rentabilidad y reconocimiento de los clientes, por su personal competitivo y por sus servicios de capacitaciones, consultoría y gestión de altísima seriedad, honestidad y calidad.

c. Servicios que ofrece INPECHI

Los servicios que ofrecen están enfocados a la Promoción de Negocios entre Perú y China.

La Institución brinda los siguientes servicios:

- Asesorías para Importar de China:
 - Creación y constitución de empresas importadoras.
 - Asesoría contable y tributaria para la importación.
 - Selección y búsqueda de proveedores Chinos confiables
 - Inteligencia comercial. Análisis e investigación de la competencia.
 - Contacto con agentes de carga, agentes de aduanas, agentes de seguros y operadores logísticos.
 - Creación de marcas y patentes.
 - Planeamiento estratégico para la comercialización de los productos importados.
- Seminarios de Importación de China, Negociación con el proveedor Chino, Cómo contactar proveedores Chinos seguros, y un seminario de Inteligencia comercial.
- Visitas y contactos con proveedores chinos en diversas Ferias de Negocios
- Tour a la Feria de Cantón. Viaje a China para conocer proveedores y fábricas.

Los servicios abarcan el ámbito de personas naturales interesadas en hacer negocios con China y también a Pequeñas y Medianas empresas que requieran importar sus propios productos.

d. Datos

- Dirección: Av. Arequipa 5095 Of. 3031 Miraflores Lima
- Teléfonos: 01-4472164
- RPM: #999966374
- RPC: 962385273
- Nextel: 814*6800
- E-Mail: informes@inpechi.org
- Web: www.ImportardeChina.com.pe

Teniendo en cuenta la información de los apartados anteriores y el interés de este proyecto en centrar el estudio en ciertos aspectos dentro de la Empresa de Consultoría y Seminarios, la metodología de Ralph Kimball se ajusta más a lo que se quiere desarrollar al permitir la creación del DWH partiendo de los DM.

La metodología de Kimball nos permite, por tanto, ofrecer soluciones en un plazo inferior al que resultaría de abordar un proyecto global destinado a toda la institución.

Por otro lado, la metodología de Kimball ofrece una clara exposición de las fases y sus respectivas actividades, así como un buen número de ejemplos documentados en los cuales apoyarse cuando no se dispone de gran experiencia en el desarrollo de DM y DWH.

3.1.2. Aplicación de la Metodología

3.1.2.1. Planificación del Proyecto

El presente proyecto se enfocará en el área de Ventas, sin embargo se utilizará algunas tablas del área de Marketing como complemento para que los reportes arrojen información más completa y útil a la Gerencia de la Empresa.

La información que se utilizará a fin de analizar el entorno de la empresa, serán de las tablas más completas e importantes de su Base de Datos y entre otras fuentes con las que cuentan.

Respecto a los periodos de Historicidad se trabajará con los años 2011, 2012 y parte del 2013.

En cuanto al personal que desarrollará el proyecto, será únicamente quien desarrolla este trabajo escrito, con ayuda del personal de la empresa quien colaborará con las dudas que se tenga frente a los sistemas que manejan y/o recopilación de información necesaria para la elaboración de este documento.

A continuación se presenta la planificación prevista para el proyecto de DM que se realizó:

Cuadro 2: Cronograma de Actividades para la realización del Proyecto

PROCESOS	INICIO	FIN
Definición de los Requerimientos del Negocio	lun 02/09/2013	jue 05/09/2013
Modelo Dimensional	vie 06/09/2013	lun 25/11/2013
Definir Modelo de Negocio	vie 06/09/2013	lun 16/09/2013
Definir el Grano	mar 17/09/2013	lun 23/09/2013
Elegir las Dimensiones	mar 24/09/2013	jue 10/10/2013
Identificar los Hechos	vie 11/10/2013	jue 31/10/2013
Detallar las Dimensiones	lun 04/11/2013	lun 25/11/2013
Diseño Físico del Data Warehouse	mar 26/11/2013	jue 12/12/2013
Diseño y Desarrollo de la presentación de datos	vie 13/12/2013	jue 19/12/2013
Diseño de la Arquitectura técnica	vie 20/12/2013	vie 10/01/2014
Selección de Productos e Instalación	lun 13/01/2014	vie 17/01/2014

Fuente: Elaboración Propia

3.1.2.2. Definición de los Requerimientos del Negocio

Con ayuda del Gerente de la Empresa se obtuvieron los dos procesos más relevantes e importantes que maneja la empresa y sobre todo con las que se podrá tener acceso a la explotación de datos.

Para ello se construye la matriz de procesos/ dimensiones, de tal forma que se logre identificar el requerimiento o proceso del negocio más crítico.

Cuadro 3: Matriz de procesos -Dimensiones

Proceso de Negocio	DIMENSIONES									
	Clientes del servicio	Asistentes al servicio	Servicio	Tiempo	Categoría	Descuento	Publicidad	Ponente	Lugar	Importes
Análisis de Ventas	X		X	X	X			X	X	X
Gestión de Marketing		X				X	X			X

Fuente: Elaboración Propia

En este caso podemos observar que el requerimiento más crítico está en el Análisis de Ventas, y será el punto de partida para comenzar a diseñar nuestra solución Datamart.

a. Algunos Requerimientos en base al área definida:

- Porcentaje de Personas que asistieron por Tipo de Seminarios.
- Mostrar la Evolución de cantidad de personas del Tipo de Seminario más asistido en los tres últimos años.
- Cantidad de personas que asistieron en total a los seminarios de Lima en base al medio de publicidad por el cual se enteraron del evento, todo esto del último año y definido por mes.
- Porcentaje de Ingreso Obtenido vs el ingreso total de los últimos seminarios asistidos por los 5 primeros clientes de categorización A.

- Entre otros requerimientos que el dueño y/o administrador deseen consultar en base a la información que se traerá a la Fact Table.

b. Otros Requerimientos en base al modelo utilizado

- Migrar la información guardada en archivos Excel hacia la base de datos para una mejor manipulación de la data todo desde una sola fuente de origen.
- Determinar las tablas y atributos más importantes de la base de datos de la empresa para llevar a cabo la implementación del Datamart.
- Contar con un repositorio de datos con información útil y bien estructurada para futuros usos de soluciones BI.

3.1.2.3. Modelado Dimensional

a. Definición del Proceso de Negocio:

Ya que el proceso más crítico es el de Ventas, se utilizará como complemento algunas tablas del área de Marketing como complemento y se trabajará con ambos para medir el impacto y la influencia de uno en base al otro.

b. Nivel de Granularidad:

Se llegará a un nivel de Granularidad medio, sin un detalle a profundidad pero tampoco muy general.

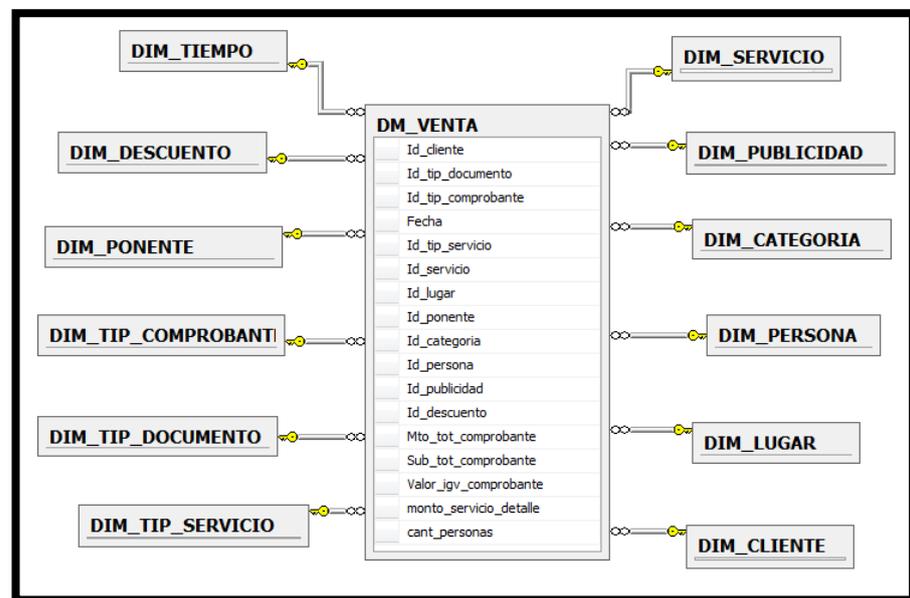
Por detallar algunos tenemos:

- Fecha de inicio del servicio.
- Categorización de los Clientes.
- Tipos de Seminarios en la Empresa
- Lugares donde se encuentran los auditorios, por servicio.
- Medios de Comunicación con más captación de personas registradas a los eventos, etc.

c. Elección de las Dimensiones

Para el área de Ventas y Marketing se identificó 12 tablas dimensiones, 9 de ellas son dimensiones directamente del negocio y 3 son atributos que fueron considerados como dimensiones por dos motivos; primero para no entrar a una tipología copo de nieve, y segundo porque pueden ser utilizados en algún momento como filtro de personas Naturales o Jurídicas.

Figura 10: Elección de las Dimensiones y la Tabla Hechos o Fact Table



Fuente: Elaboración Propia

d. Identificación de los Hechos

Para la tabla hechos iniciamos identificando las medidas; y como el área crítica es Ventas, sus medidas estarán determinadas por los montos y cantidades que aparecen en sus facturas y/o boletas, los demás campos de la Fact Table serán las llaves primarias de las dimensiones.

e. Modelo Gráfico de Alto Nivel

Para concluir con el proceso dimensional inicial se realiza un gráfico denominado Modelo Dimensional de alto nivel (o

gráfico de burbujas, Bubble chart, en el léxico de Kimball), como ilustra la siguiente figura:

Figura 11: Modelo Gráfico de Alto Nivel



Fuente: Elaboración Propia

f. Detalle de las Tablas de Dimensión

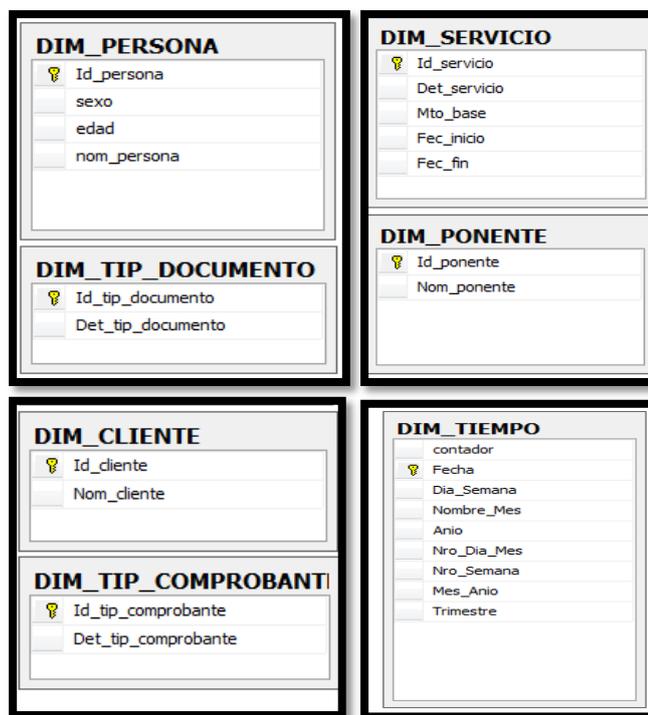
Figura 12: Atributos de las Dimensiones

Publicidad, Lugar, Descuento y Tipo_Servicio



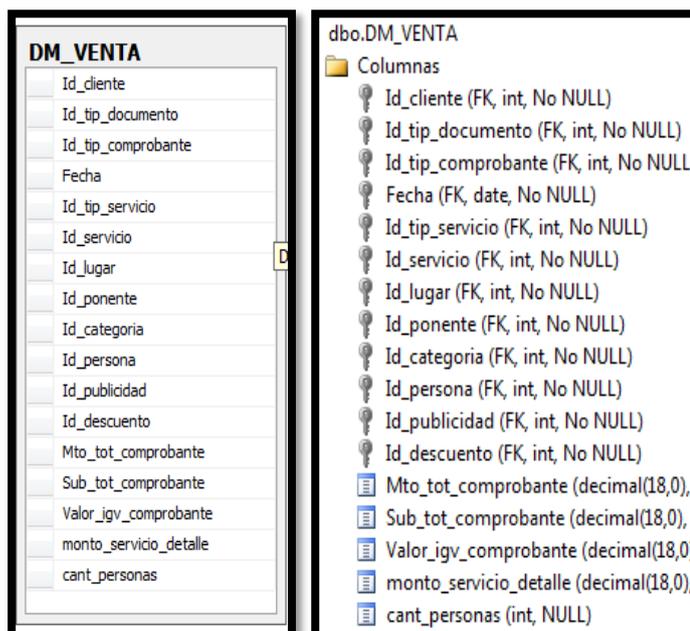
Fuente: Elaboración Propia

Figura 13: Atributos de las Dimensiones, Persona, Servicio, Tipo_Documento, Ponente, Cliente, Tipo_Cprobante, y Tiempo.



Fuente: Elaboración Propia

Figura 14: Atributos de la Fact Table



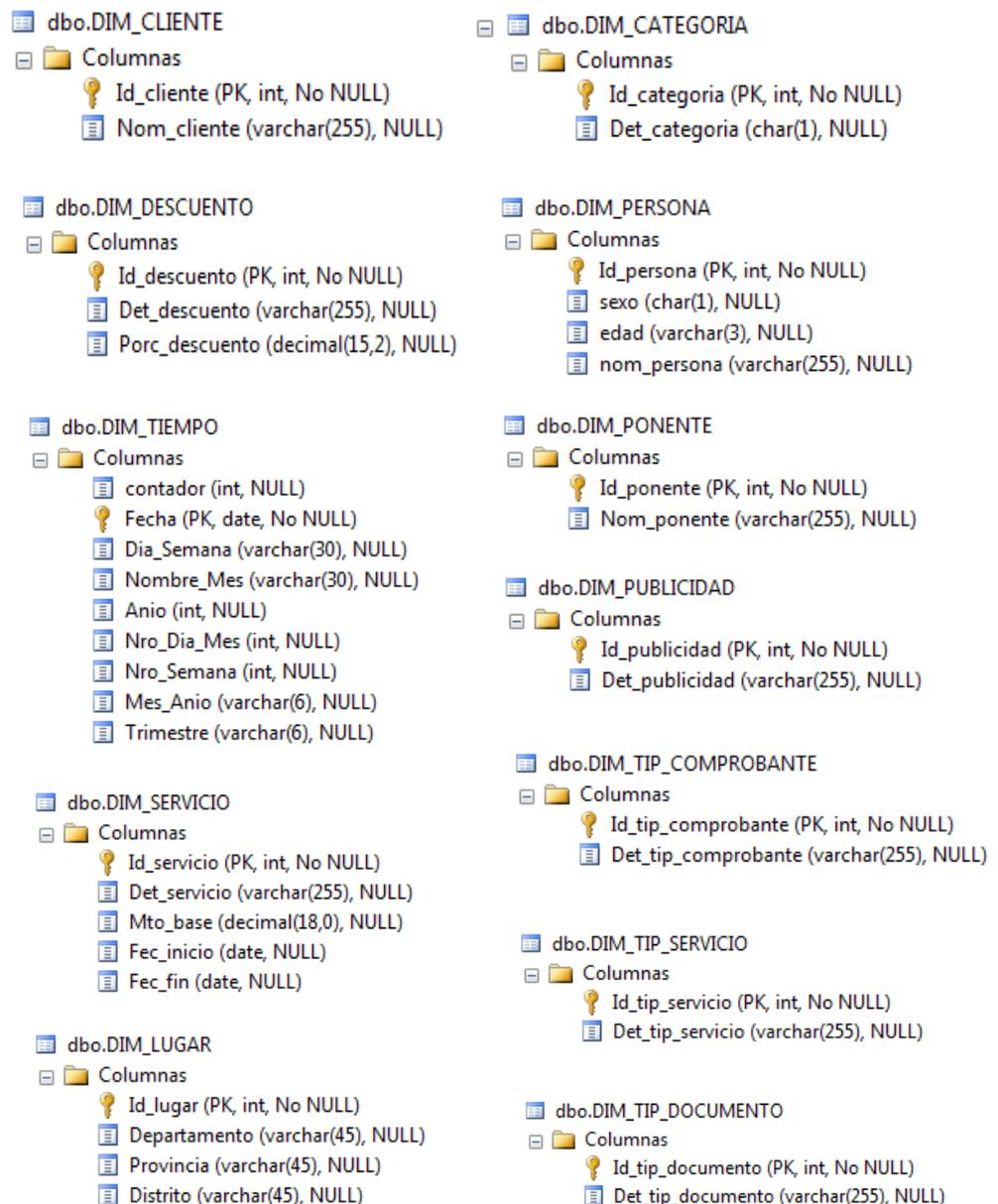
Fuente: Elaboración Propia

3.1.2.4. Diseño Físico del Data Mart

Siguiendo con la metodología de Ralph Kimball, el siguiente paso en el desarrollo del DM es el diseño físico del mismo. El modelo dimensional desarrollado en el capítulo anterior debe convertirse ahora en un diseño físico.

En el diseño físico se deben incluir los nombres de columna físicos, los tipos de los datos, las declaraciones de clave (si procede) y la posibilidad de incluir valores nulos.

Figura 15: Diseño Físico del Data Mart y sus Dimensiones



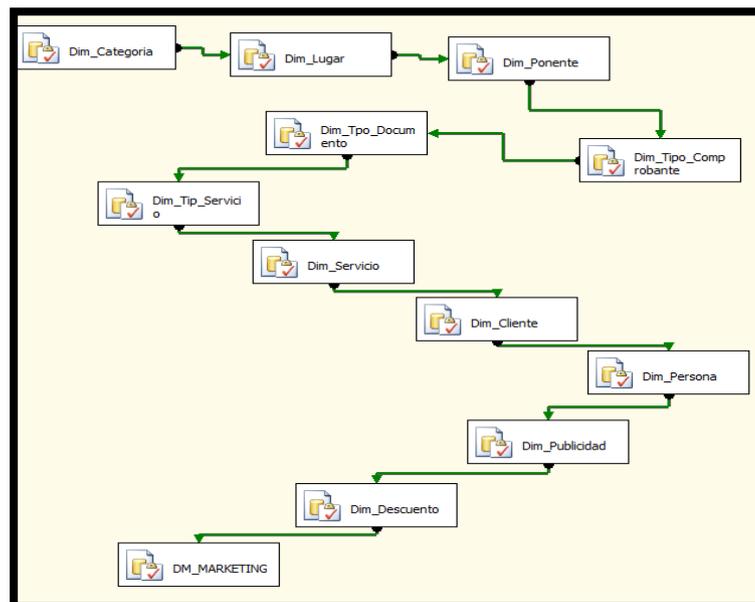
- dbo.DM_VENTA
 - Columnas
 - Id_cliente (FK, int, No NULL)
 - Id_tip_documento (FK, int, No NULL)
 - Id_tip_comprobante (FK, int, No NULL)
 - Fecha (FK, date, No NULL)
 - Id_tip_servicio (FK, int, No NULL)
 - Id_servicio (FK, int, No NULL)
 - Id_lugar (FK, int, No NULL)
 - Id_ponente (FK, int, No NULL)
 - Id_categoria (FK, int, No NULL)
 - Id_persona (FK, int, No NULL)
 - Id_publicidad (FK, int, No NULL)
 - Id_descuento (FK, int, No NULL)
 - Mto_tot_comprobante (decimal(18,0), No NULL)
 - Sub_tot_comprobante (decimal(18,0), NULL)
 - Valor_igv_comprobante (decimal(18,0), NULL)
 - monto_servicio_detalle (decimal(18,0), NULL)
 - cant_personas (int, NULL)

Fuente: Elaboración Propia

3.1.2.5. Diseño y Desarrollo de la Presentación de Datos

En esta etapa se lleva a cabo la extracción, la transformación y la carga (ETL process) de los datos dentro del diseño del DM. Además, se definen como Procesos de Transformación los procesos para convertir o recodificar los datos fuente a fin de poder efectuar la carga efectiva del Modelo Físico.

Figura 16: Diseño del ETL

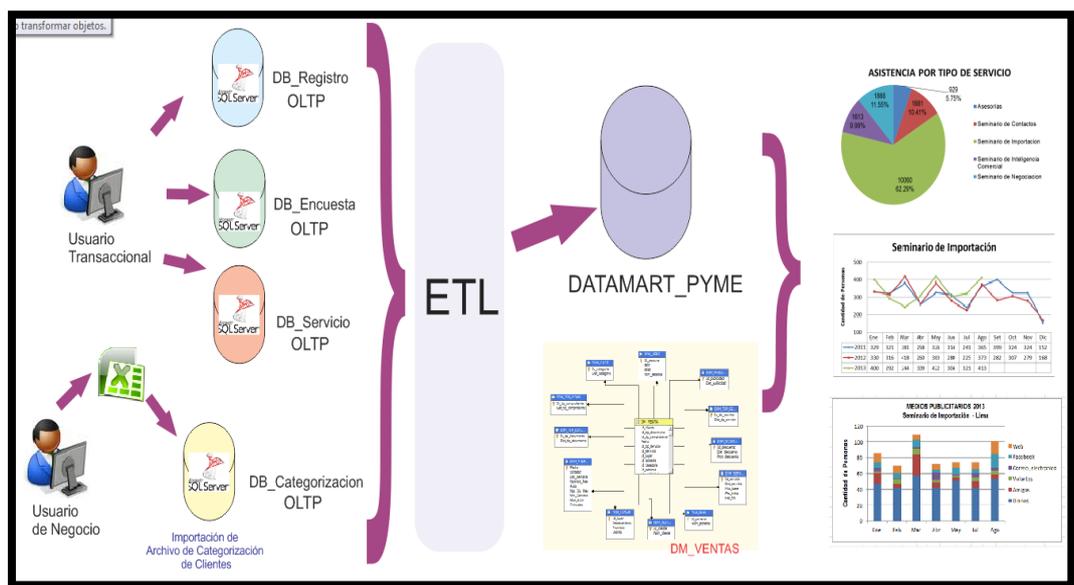


Fuente: Elaboración Propia

3.1.2.6. Diseño de la Arquitectura Técnica

Al igual que un plan para un nuevo hogar, la arquitectura técnica es el modelo de los servicios técnicos del DM y de sus elementos. El plan de la arquitectura técnica sirve como marco de organización para apoyar la integración de las tecnologías. La arquitectura permite detectar los problemas a priori y trata de minimizar al comienzo del proyecto las sorpresas que pudieran surgir.

Figura 17: Arquitectura para el Proyecto de DM en INPECHI



Fuente: Elaboración Propia

3.1.2.7. Selección de Producto e Instalación

Para la creación del Datamart se utilizará SQL Server 2008 ya que es la misma tecnología de BD que utiliza la empresa. Para generar cubos se dispondrá de Microsoft Visual Studio Analysis Services, así como también Integration Services para el uso de los ETLs y como muestra del correcto funcionamiento de la implementación del Datamart se mostrará algunos gráficos realizados en Excel, previa conexión desde el cubo de Ventas o el de Marketing generado.

3.2. CONSTRUCCIÓN DEL MODELO

3.2.1. Construcción de las Dimensiones

```
CREATE TABLE DIM_SERVICIO (  
  Id_servicio INT NOT NULL,  
  Det_servicio VARCHAR(255) NULL,  
  Mto_base DECIMAL NULL,  
  Fec_inicio DATE NULL,  
  Fec_fin DATE NULL,  
  PRIMARY KEY(Id_servicio));
```

```
CREATE TABLE DIM_TIP_SERVICIO (  
  Id_tip_servicio INT NOT NULL,  
  Det_tip_servicio VARCHAR(255) NULL,  
  PRIMARY KEY(Id_tip_servicio));
```

```
CREATE TABLE DIM_TIP_DOCUMENTO (  
  Id_tip_documento INT NOT NULL,  
  Det_tip_documento VARCHAR(255) NULL,  
  PRIMARY KEY(Id_tip_documento));
```

```
CREATE TABLE DIM_TIP_COMPROBANTE (  
  Id_tip_comprobante INT NOT NULL,  
  Det_tip_comprobante VARCHAR(255) NULL,  
  PRIMARY KEY(Id_tip_comprobante));
```

```
CREATE TABLE DIM_TIEMPO (  
  contador INT NULL,  
  Fecha DATE,  
  Dia_Semana VARCHAR(30) NULL,  
  Nombre_Mes VARCHAR(30) NULL,  
  Anio INT NULL,  
  Nro_Dia_Mes INT NULL,  
  Nro_Semana INT NULL,  
  Mes_Anio VARCHAR(6) NULL,  
  Trimestre VARCHAR(6) NULL,  
  PRIMARY KEY(Fecha));
```

```
CREATE TABLE DIM_PONENTE(  
  Id_ponente INT NOT NULL,  
  Nom_ponente VARCHAR(255) NULL,  
  PRIMARY KEY(Id_ponente));
```

```
CREATE TABLE DIM_CLIENTE (  
  Id_cliente INT NOT NULL,  
  Nom_cliente VARCHAR(255) NULL,  
  PRIMARY KEY(Id_cliente));
```

```
CREATE TABLE DIM_CATEGORIA (  
  Id_categoria INT NOT NULL,  
  Det_categoria CHAR(1) NULL,  
  PRIMARY KEY(Id_categoria));
```

```
CREATE TABLE DIM_LUGAR (  
  Id_lugar INT NOT NULL,  
  Departamento VARCHAR(45) NULL,  
  Provincia VARCHAR(45) NULL,  
  Distrito VARCHAR(45) NULL,  
  PRIMARY KEY(Id_lugar));
```

```
CREATE TABLE DIM_PERSONA (  
  Id_persona INT NOT NULL,  
  sexo char(1) NULL,  
  edad varchar(3) NULL,  
  nom_persona varchar(255) NULL,  
  PRIMARY KEY(Id_persona));
```

```
CREATE TABLE DIM_DESCUENTO (  
  Id_descuento INT NOT NULL,  
  Det_descuento VARCHAR(255) NULL,  
  Porc_descuento decimal(15,2) NULL,  
  PRIMARY KEY(Id_descuento));
```

```
CREATE TABLE DIM_PUBLICIDAD (  
  Id_publicidad INT NOT NULL,  
  Det_publicidad VARCHAR(255) NULL,  
  PRIMARY KEY(Id_publicidad));
```

3.2.2. Construcción de la Fact Table DM_Venta

```
CREATE TABLE DM_VENTA (  
  Id_cliente INTEGER NOT NULL,  
  Id_tip_documento INTEGER NOT NULL,  
  Id_tip_comprobante INTEGER NOT NULL,  
  Fecha DATE NOT NULL,  
  Id_tip_servicio INTEGER NOT NULL,  
  Id_servicio INTEGER NOT NULL,  
  Id_lugar INTEGER NOT NULL,  
  Id_ponente INTEGER NOT NULL,  
  Id_categoria INTEGER NOT NULL,  
  Id_persona INTEGER NOT NULL,  
  Id_publicidad INTEGER NOT NULL,  
  Id_descuento INTEGER NOT NULL,  
  Mto_tot_comprobante DECIMAL NOT NULL ,  
  Sub_tot_comprobante DECIMAL NULL,
```

Valor_igv_comprobante **DECIMAL** NULL,
 monto_servicio_detalle **DECIMAL** NULL,
 cant_personas **INTEGER** NULL,
FOREIGN KEY(Id_cliente)
REFERENCES dbo.DIM_CLIENTE(Id_cliente),
FOREIGN KEY(Id_tip_documento)
REFERENCES dbo.DIM_TIP_DOCUMENTO(Id_tip_documento),
FOREIGN KEY(Id_tip_comprobante)
REFERENCES
 dbo.DIM_TIP_COMPROBANTE(Id_tip_comprobante),
FOREIGN KEY(fecha)
REFERENCES dbo.DIM_tiempo(fecha),
FOREIGN KEY(Id_tip_servicio)
REFERENCES DIM_TIP_SERVICIO(Id_tip_servicio),
FOREIGN KEY(Id_servicio)
REFERENCES dbo.DIM_SERVICIO(Id_servicio),
FOREIGN KEY(Id_lugar)
REFERENCES dbo.DIM_LUGAR(Id_lugar),
FOREIGN KEY(Id_ponente)
REFERENCES dbo.DIM_PONENTE(Id_ponente),
FOREIGN KEY(Id_categoria)
REFERENCES dbo.DIM_CATEGORIA(Id_categoria) ,
FOREIGN KEY(Id_persona)
REFERENCES dbo.DIM_PERSONA(Id_persona),
FOREIGN KEY(Id_publicidad)
REFERENCES dbo.DIM_PUBLICIDAD(Id_publicidad),
FOREIGN KEY(Id_descuento)
REFERENCES dbo.DIM_DESCUENTO(Id_descuento),);

3.2.3. Construcción de los ETLs

Se trabajó a través de pequeños procedimientos almacenados.

Dimensiones

```

Create proc PA_DIM_Tipo_Documento
AS
delete from dbo.DIM_TIP_DOCUMENTO
INSERT dbo.DIM_TIP_DOCUMENTO
(Id_tip_documento,Det_tip_documento)
SELECT [Cod_Tipo_doc],[Tpo_Documento]
FROM [DB_Servicio].[dbo].[Tipo_Documento]
GO
  
```

```

Create proc PA_DIM_Tipo_Comprobante
as
delete from dbo.DIM_TIP_COMPROBANTE
  
```

```

insert dbo.DIM_TIP_COMPROBANTE
(Id_tip_comprobante,Det_tip_comprobante)
select Cod_Tipo_Comprobante,Detalle_Comprobante
from DB_Servicio.dbo.Tipo_Comprobante
go

```

```

create proc PA_DIM_Lugar
as
delete from dbo.DIM_LUGAR
insert dbo.DIM_LUGAR (Id_Lugar,Departamento,Provincia,Distrito)
select a.Cod_Lugar,a.Departamento,a.Provincia,a.Distrito
from DB_registro.dbo.lugar a
go

```

```

create proc PA_DIM_Ponente
as
delete from dbo.DIM_PONENTE
insert dbo.DIM_PONENTE (Id_ponente,Nom_ponente)
select a.cod_ponente, a.nom_ponente
from DB_Servicio.dbo.Ponente a
go

```

```

create proc PA_DIM_Categoria
as
delete from dbo.DIM_CATEGORIA
insert dbo.DIM_CATEGORIA (Id_categoria,Det_categoria)
select a.COD_CATEGORIA, A.DESC_CATEGORIA
from DB_categorizacion.dbo.DET_CATEGORIA a
go

```

```

create proc PA_DIM_Tip_Servicio
as
delete from dbo.DIM_TIP_SERVICIO
insert dbo.DIM_TIP_SERVICIO(Id_tip_servicio,Det_tip_servicio)
select Cod_Tipo_Servicio,Tipo_Servicio
from DB_Servicio.dbo.Tipo_Servicio a
go

```

```

create proc PA_DIM_Servicio
as
delete from dbo.DIM_SERVICIO
insert
dbo.DIM_SERVICIO(Id_servicio,Det_servicio,Mto_base,Fec_inicio,Fe
c_fin)
select a.cod_servicio, a.desc_servicio, a.monto_base, a.fecha_inicio,
a.fecha_fin
from DB_Servicio.dbo.Servicio a
go

```

```

create proc PA_DIM_Cliente
as
delete from dbo.DIM_Cliente
insert dbo.DIM_Cliente(Id_Cliente,Nom_cliente)
select a.Cod_Cliente,a.Nomb_Cliente
from DB_Servicio.dbo.Cliente a
go

```

```

create PROC PA_DIM_Persona
AS
delete from dbo.DIM_PERSONA
INSERT dbo.DIM_PERSONA (Id_persona,sexo,edad,nom_persona)
SELECT [a].cod_persona ,[a].Sexo_Persona,[a].edad
,[a].Nom_Persona
FROM [DB_registro].[dbo].[Persona] a
GO

```

```

create PROC PA_DIM_Descuento
as
delete from dbo.DIM_DESCUENTO
INSERT dbo.DIM_DESCUENTO (Id_descuento, Det_descuento,
Porc_descuento)
SELECT a.cod_descuento, a.detalle_descuento, a.porcent_descuento
FROM DB_Servicio.dbo.Descuento a
Go

```

```

create PROC dbo.PA_DIM_Publicidad
as
delete from dbo.DIM_PUBLICIDAD
INSERT dbo.DIM_PUBLICIDAD (Id_publicidad,Det_publicidad)
SELECT a.cod_publicidad,a.nom_publicidad
FROM DB_Encuesta.dbo.Publicidad a
Go

```

```

create PROCEDURE dbo.LLENAR_DIM_TIEMPO AS
BEGIN
SET LANGUAGE spanish
SET DATEFIRST 1
declare @fi datetime, @ff datetime
declare @Contador int
set @fi = '2010-01-01'
Set @ff = '2013-31-12'
Set @Contador = 367

```

```

WHILE @fi <= @ff
BEGIN

```

```

INSERT INTO DATAMART_PYME.dbo.DIM_TIEMPO

```

```

([contador]
,[Fecha]
,[Dia_Semana]
,[Nombre_Mes]
,[Anio]
,[Nro_Dia_Mes]
,[Nro_Semana]
,[Mes_Anio]
,[Trimestre])

```

```

SELECT @contador,CONVERT(DATE,@fi,112),
datename(weekday,@fi),
datename(Month,@fi),
datepart(YY,@fi),
datepart(Day,@fi),
datepart(WW,@fi),
datepart(Month,@fi),
datepart(quarter,@fi)

```

```

set @fi = @fi + 1
set @contador = @contador+1

```

END

END

Fact Table DM_Venta

```

select A.COD_CLIENTE,c.Cod_Tipo_doc, a.Cod_Tipo_Comprobante,
d.fecha_inicio, E.Cod_Tipo_Servicio, B.cod_servicio,
F.Cod_Lugar, d.Cod_Ponente, J.CATEGORIA,
z.cod_persona,h.cod_publicidad ,m.cod_descuento
,a.monto_total,a.sub_total,a.igv,b.monto_servicio,b.cantidad INTO
NUEVA_TABLA
from db_servicio.dbo.cabecera_comprobante a
INNER JOIN db_servicio.dbo.det_comprobante b
ON a.Cod_Comprobante=b.cod_comprobante
INNER JOIN DB_Servicio.dbo.Cliente C
ON a.Cod_Cliente=C.Cod_Cliente
INNER JOIN db_servicio.dbo.Servicio d
ON d.cod_servicio=b.cod_servicio
INNER JOIN DB_Servicio.dbo.Tipo_Servicio E
ON D.Cod_Tipo_Servicio=E.Cod_Tipo_Servicio
INNER JOIN DB_Servicio.dbo.AUDITORIO F
ON D.cod_auditorio =F.Cod_Auditorio
INNER JOIN db_categorizacion.dbo.CATEGORIA J
ON J.COD_CLIENTE=C.COD_CLIENTE
inner join DB_Servicio.dbo.Persona_Registrada z
ON z.Cod_Cliente=c.Cod_Cliente and z.cod_servicio =d.cod_servicio
left join DB_Encuesta.dbo.registro_encuesta h

```

```
ON h.cod_persona = z.cod_persona and h.cod_servicio
=d.cod_servicio
inner join DB_Servicio.dbo.Descuento m
ON m.cod_descuento=z.cod_descuento
```

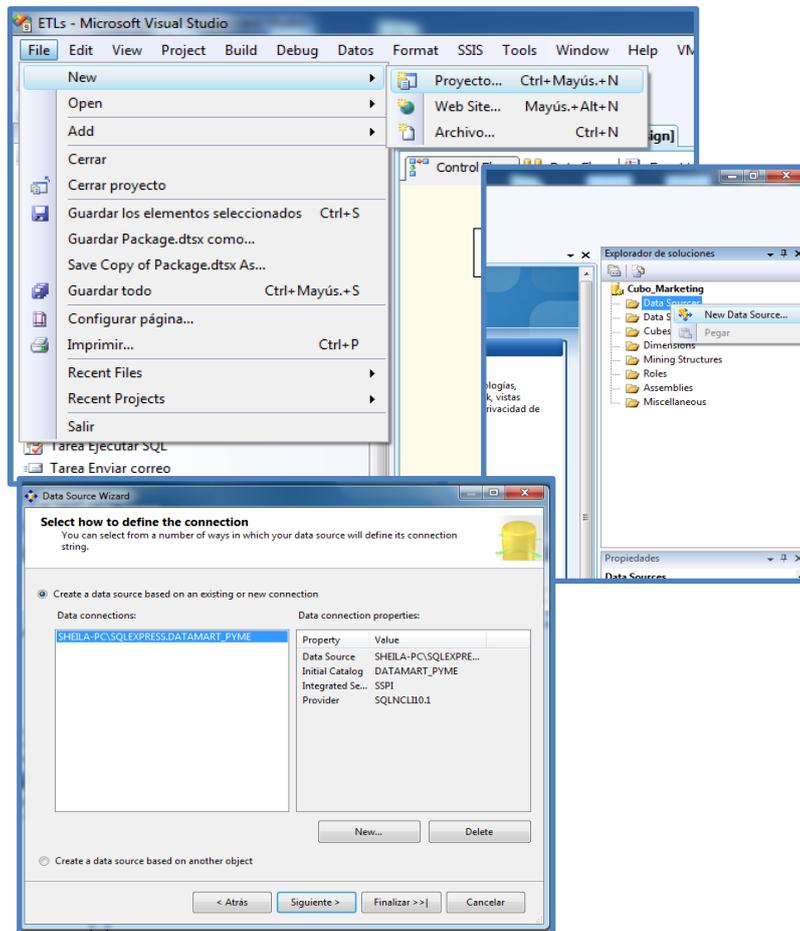
```
UPDATE NUEVA_TABLA set cod_publicidad= 10 where
cod_publicidad is null
```

```
INSERT INTO dbo.DM_VENTA SELECT * FROM
dbo.NUEVA_TABLA
```

3.2.4. Construcción de Dimensiones y Cubo

Antes de comenzar a crear las dimensiones y el Cubo de interés debemos configurar en el Analysis Services el origen de datos que vamos a utilizar, para este trabajo será una conexión a Sql Server 2008.

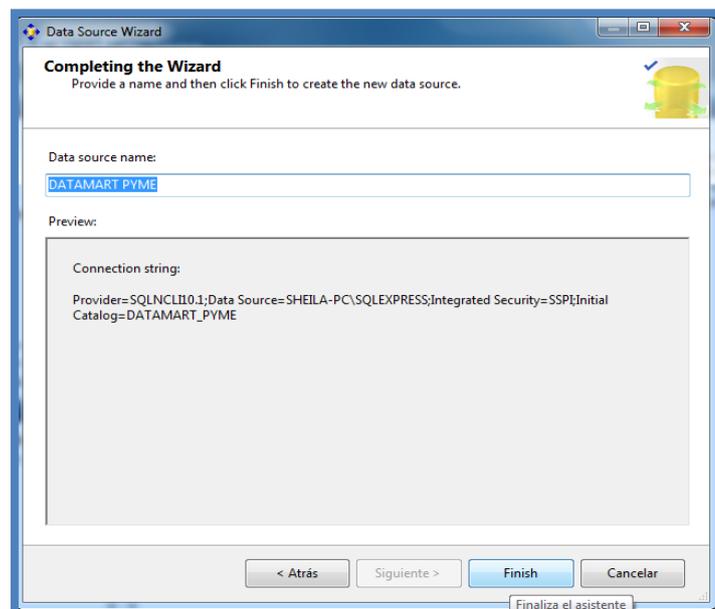
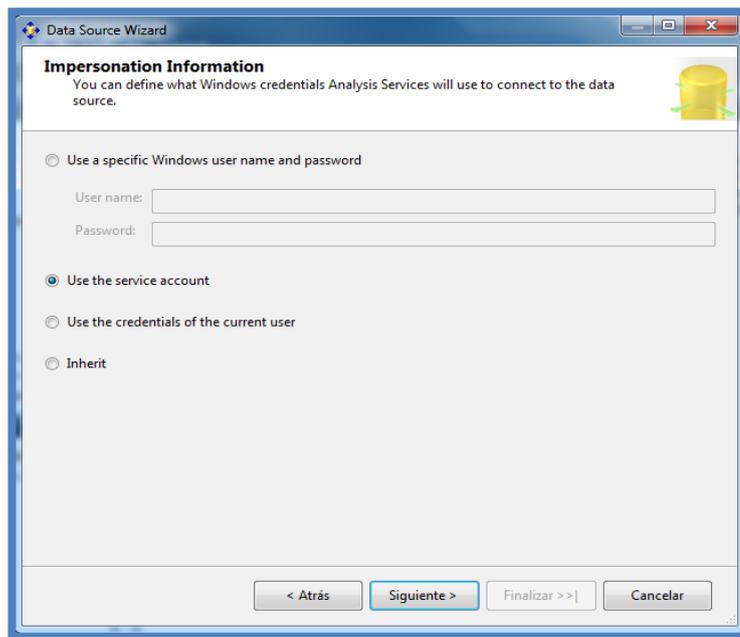
Figura 18: Configuración para la conexión de Origen de Datos



Fuente: Elaboración Propia

Seleccionamos la Base de Datos en SQL Server 2008 donde se encuentran las dimensiones así como la Fact Table.

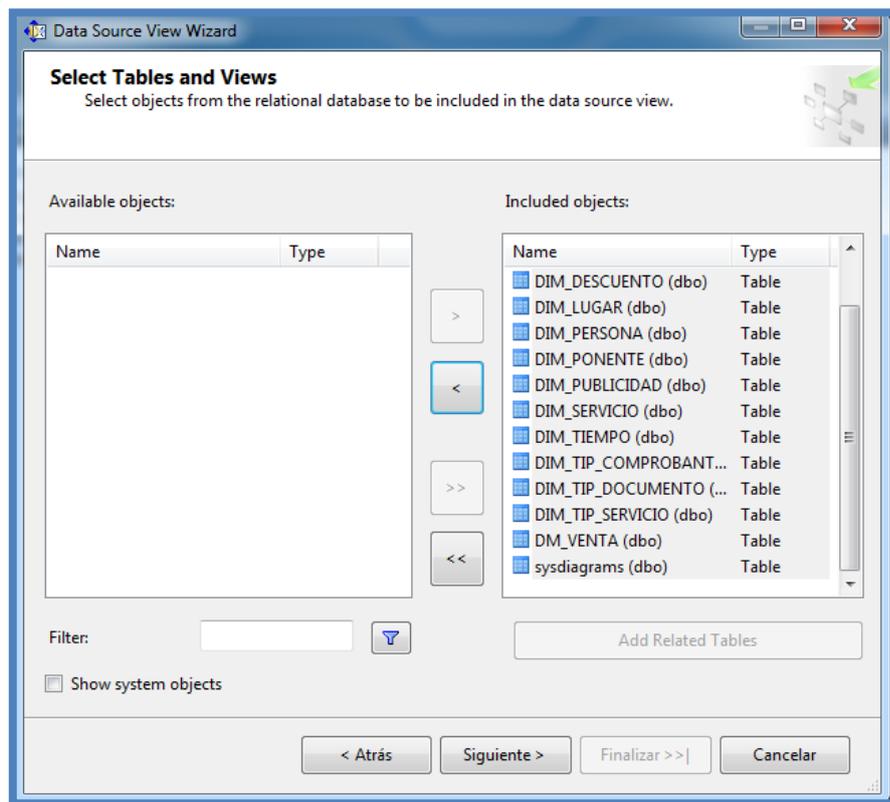
Figura 19: Elección de la Base de datos en SQL Server a la cual se accederá desde el Analysis Services.



Fuente: Elaboración Propia

Una vez terminada la configuración del Origen de datos de donde obtendremos toda la información, se deben escoger las dimensiones incluyendo a la Fact Table que utilizaremos y que ya han sido antes cargada con la información de interés en el Origen de Datos.

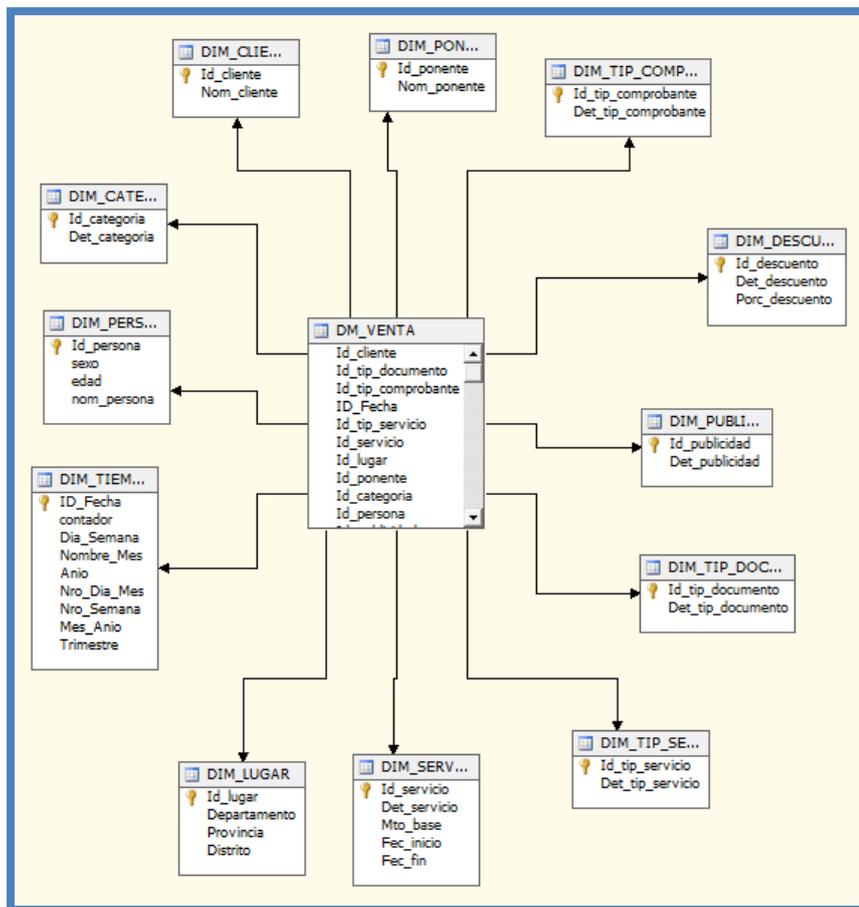
Figura 20: Lectura de todas las tablas en la Base de Datos OLAP



Fuente: Elaboración Propia

Al terminar con la lectura de las tablas de la OLAP se generará un dibujo del esquema OLAP .

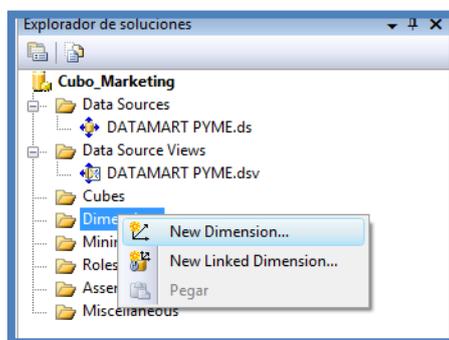
Figura 21: Esquema OLAP



Fuente: Elaboración Propia

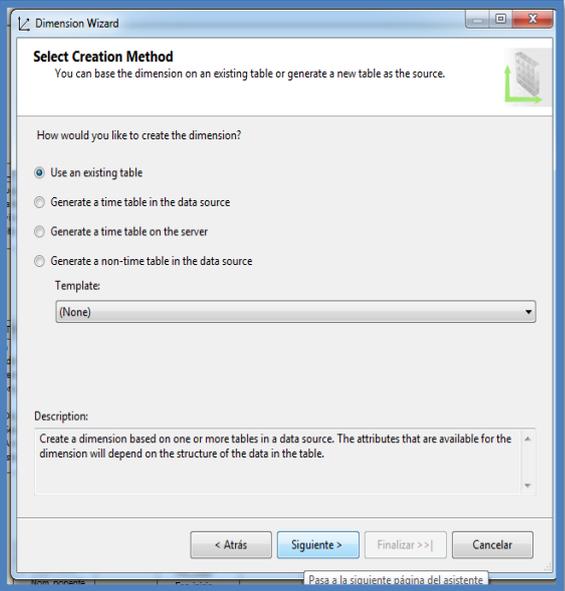
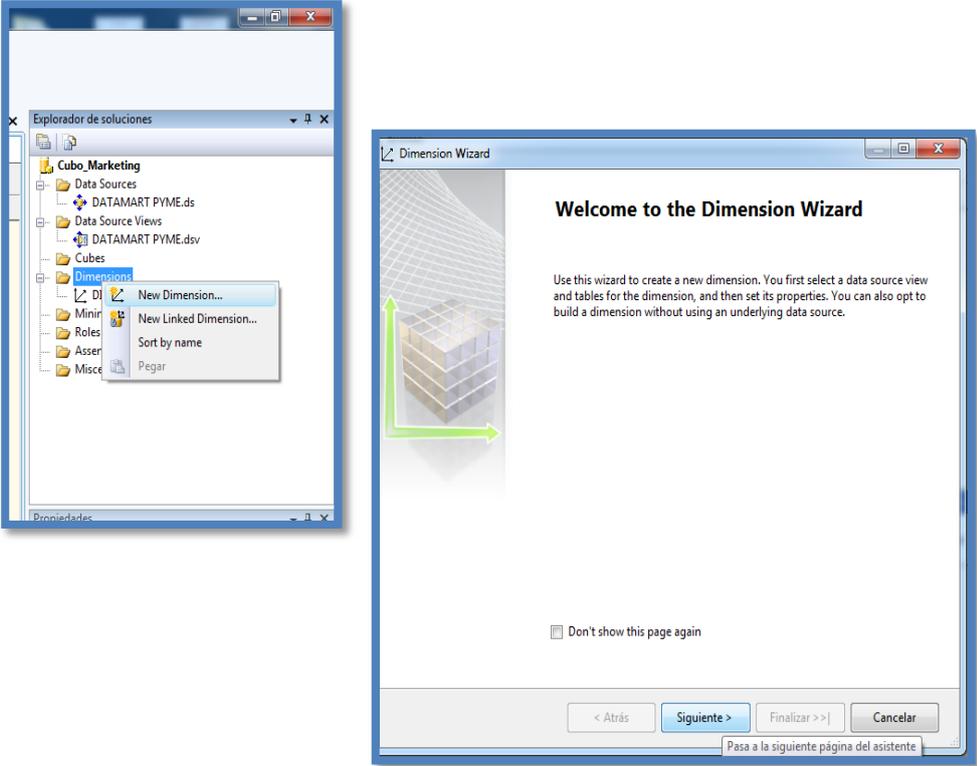
Luego de esta primera fase, ahora sí procedemos a crear las dimensiones.

Figura 22: Creación de las Dimensiones



Fuente: Elaboración Propia

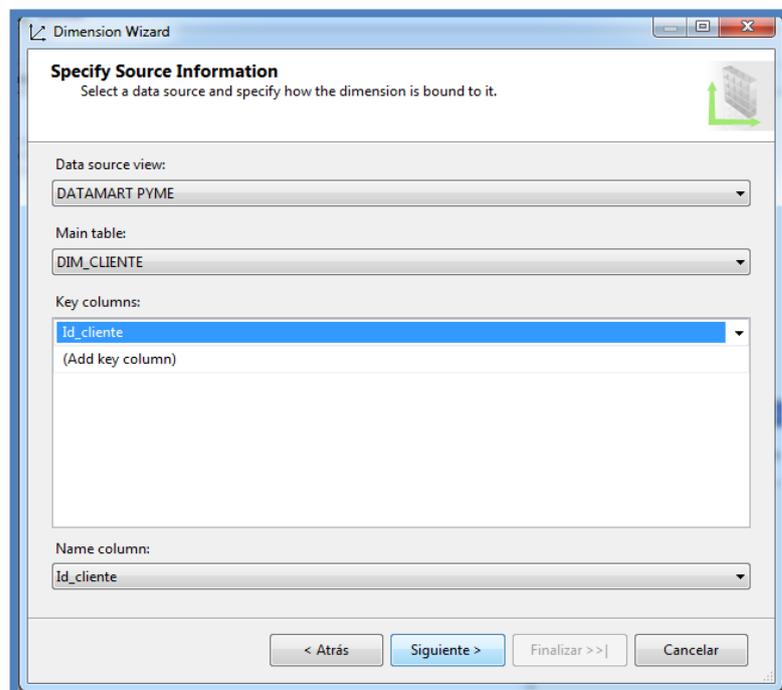
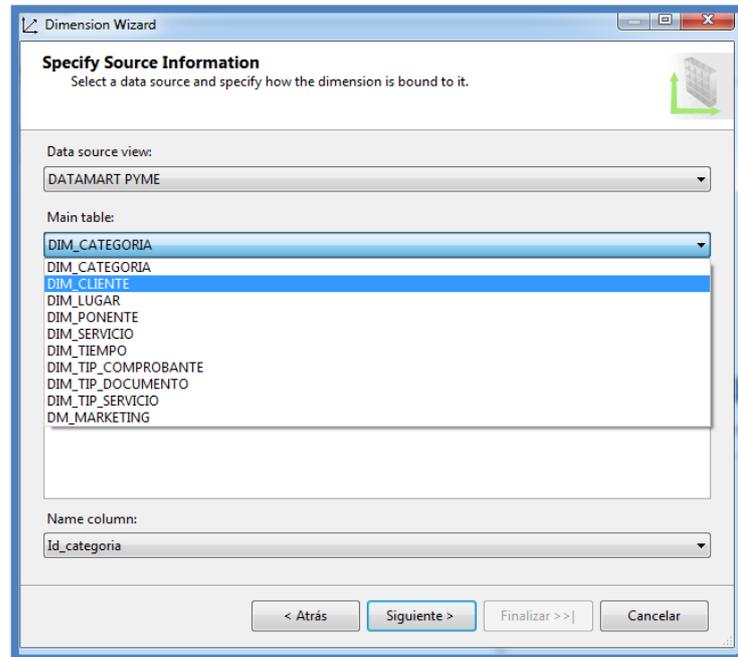
Figura 23: Primeros pasos para crear las dimensiones.



Fuente: Elaboración Propia

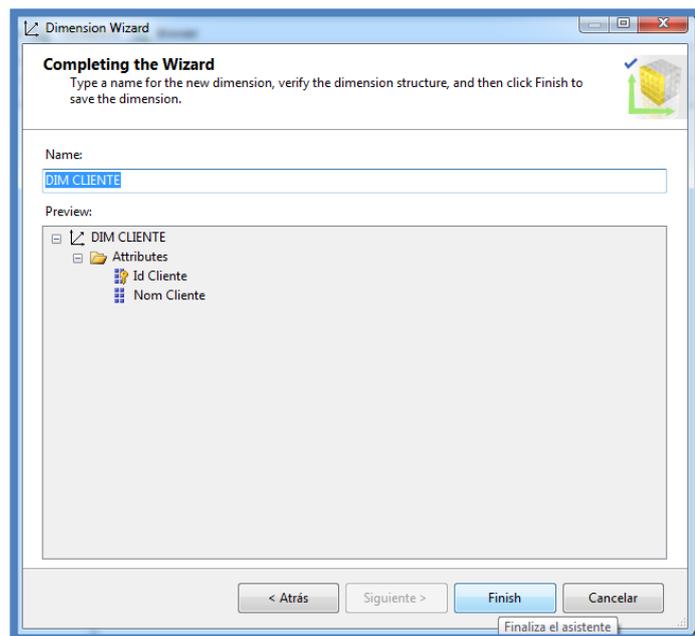
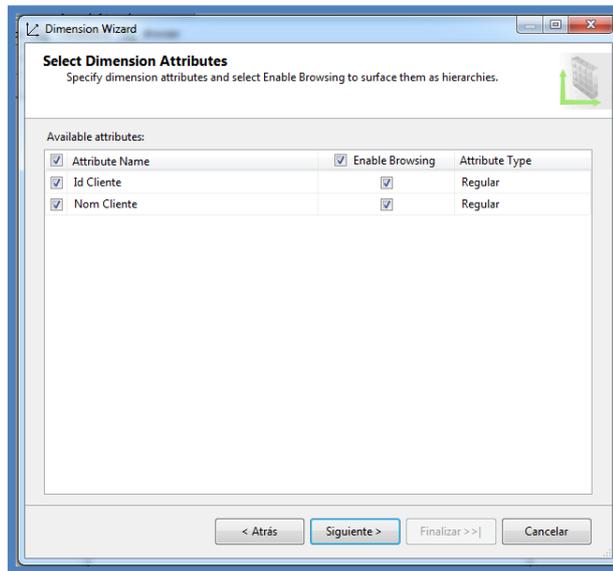
Debemos escoger una dimensión. Este paso se repetirá por todas las dimensiones que necesitemos traer al Analysis Services.

Figura 24: Selección de la dimensión CLIENTE



Fuente: Elaboración Propia

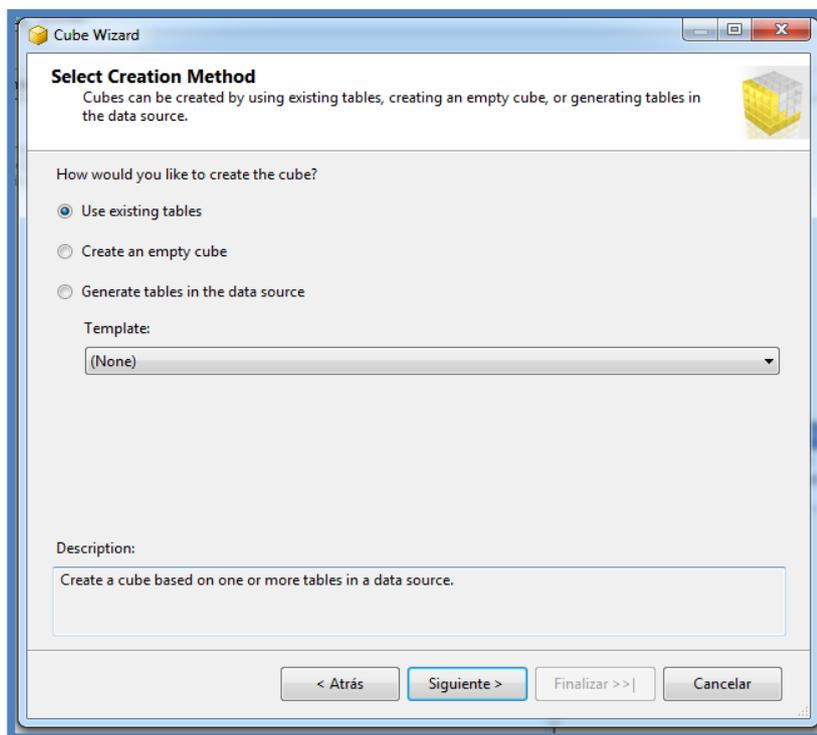
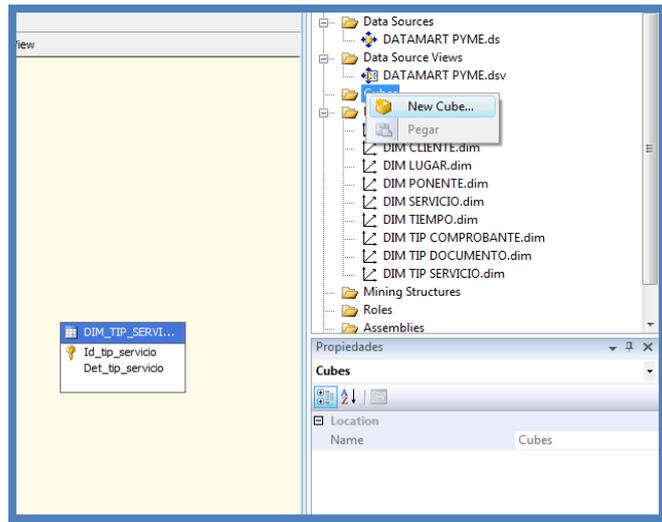
Figura 25: Selección de los atributos por Dimensión.



Fuente: Elaboración Propia

Después de crear todas las dimensiones, ya podemos generar el Cubo respectivo.

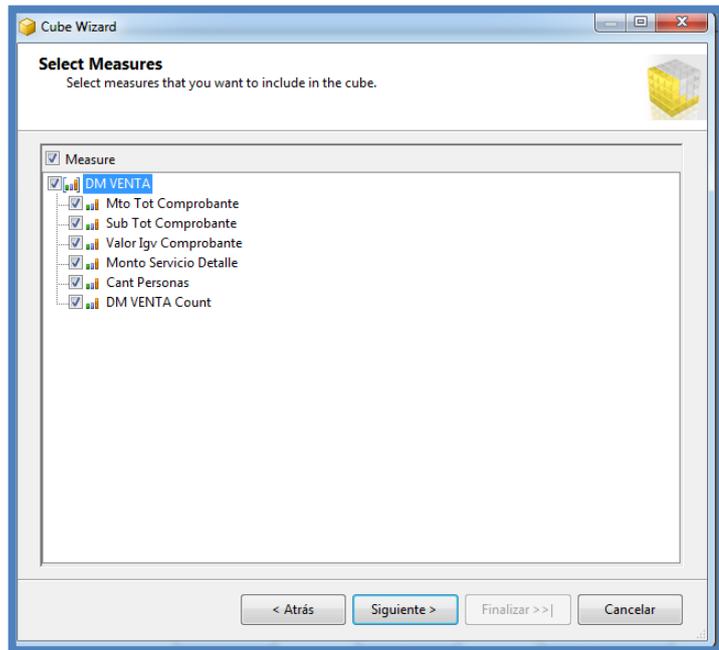
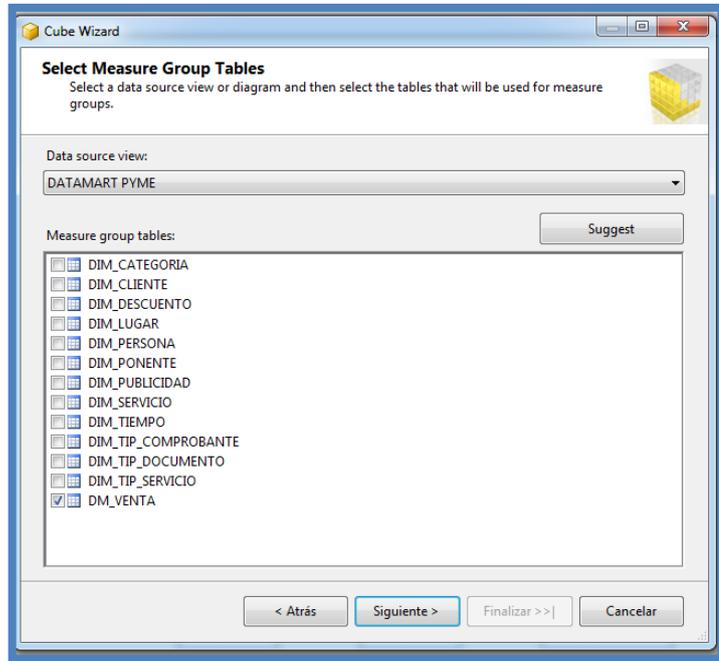
Figura 26: Generación de un Cubo



Fuente: Elaboración Propia

Se debe seleccionar la tabla que funciona como Fact Table en el OLAP.

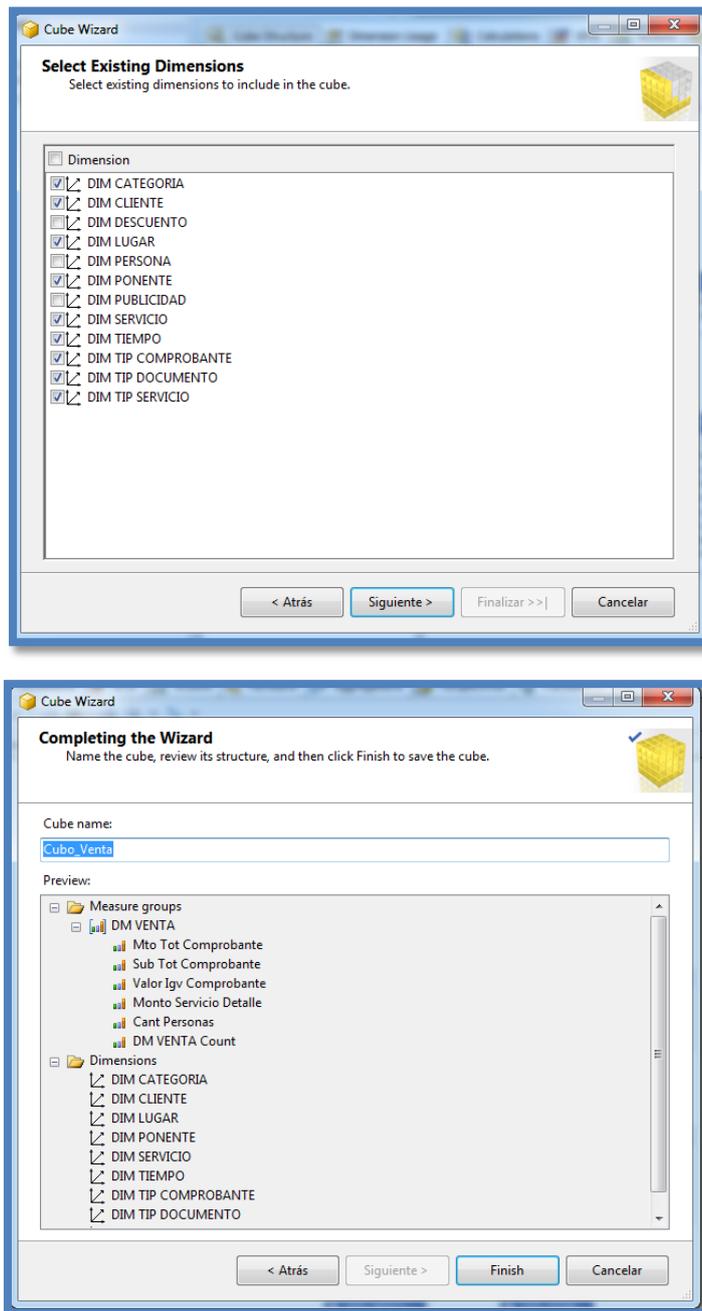
Figura 27: Creación del cubo



Fuente: Elaboración Propia

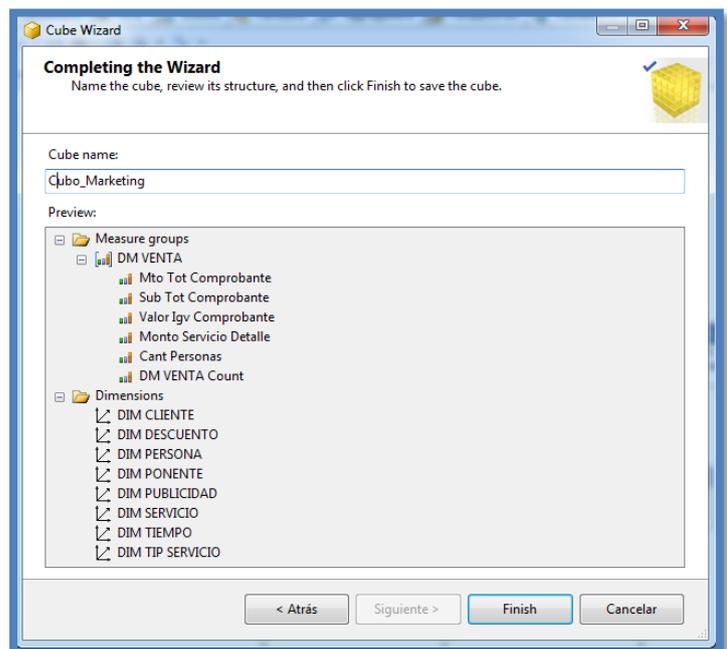
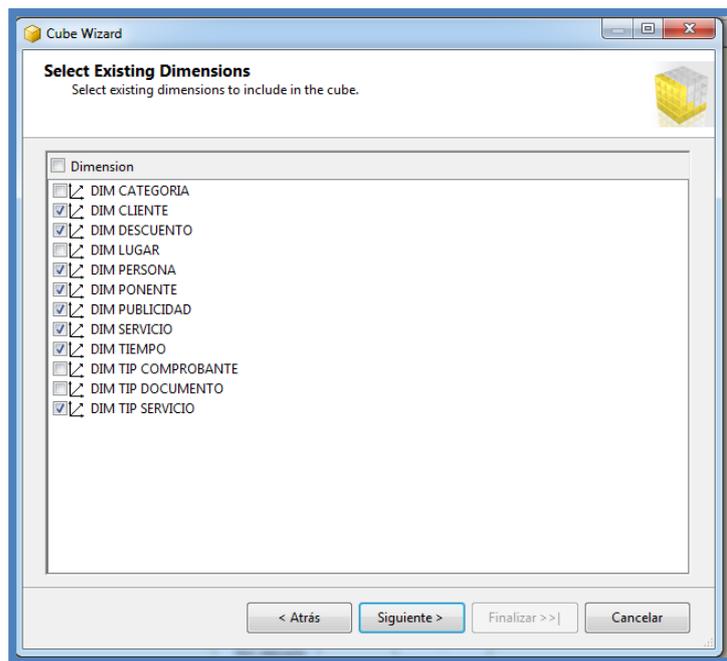
En los siguientes pasos al dar “siguiente” solo nos estará mostrando las dimensiones a escoger para el cubo, en este caso se creará dos cubos (Ventas y Marketing) por lo que no se escogerán todas las dimensiones.

Figura 28: Visualización de las dimensiones y el cubo que se creará



Fuente: Elaboración Propia

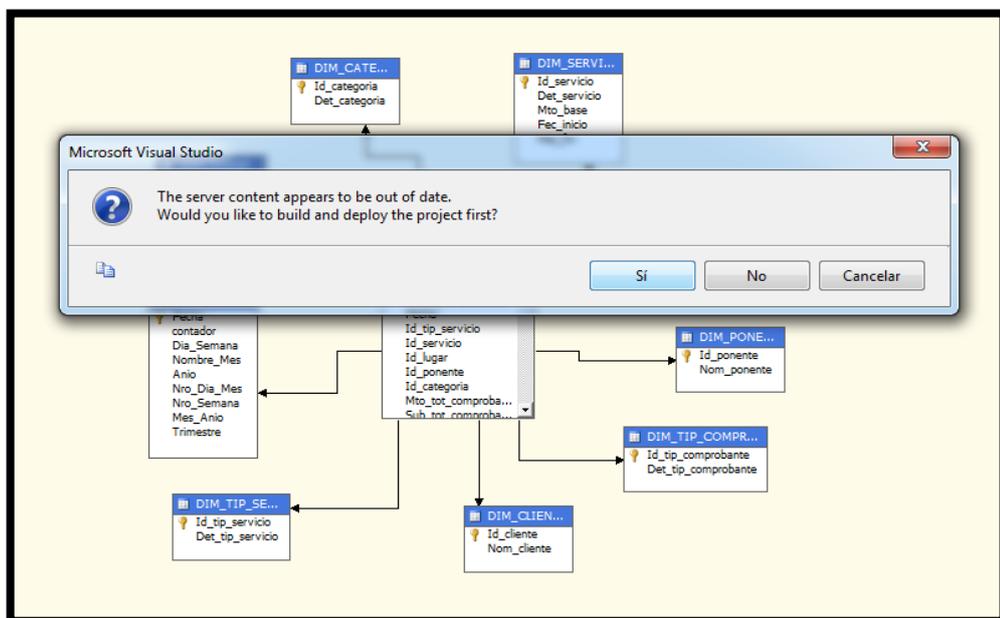
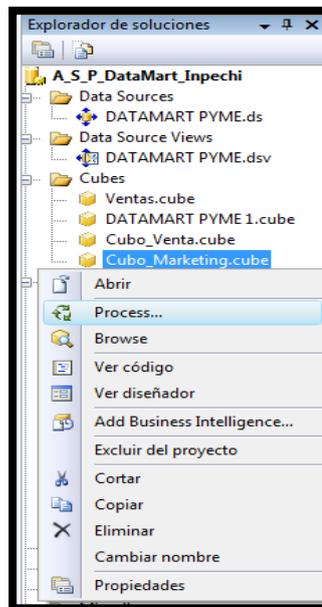
Figura 29: Creación del segundo Cubo



Fuente: Elaboración Propia

Ahora solo debemos dar un clic derecho al cubo y seleccionar: Process, de tal manera que se logre la conexión y la migración de los datos desde SQL Server2008 a los dos Cubos generados en Analysis Services.

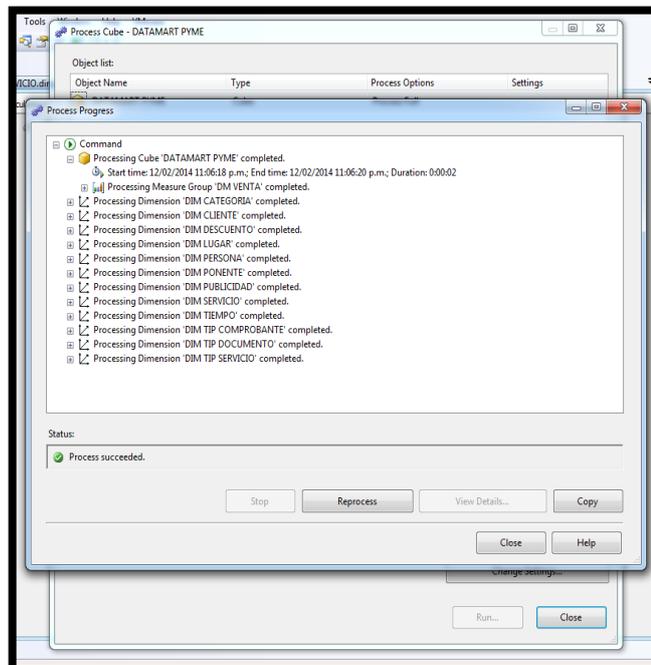
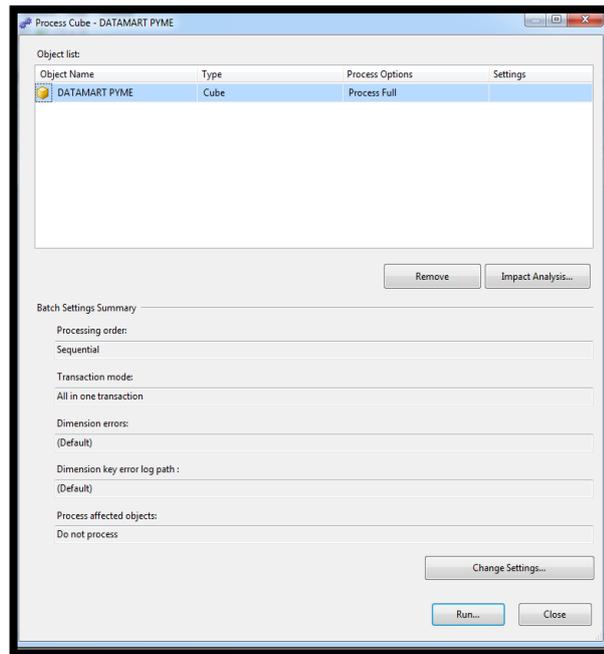
Figura 30: Últimos pasos a considerar para el Cubo



Fuente: Elaboración Propia

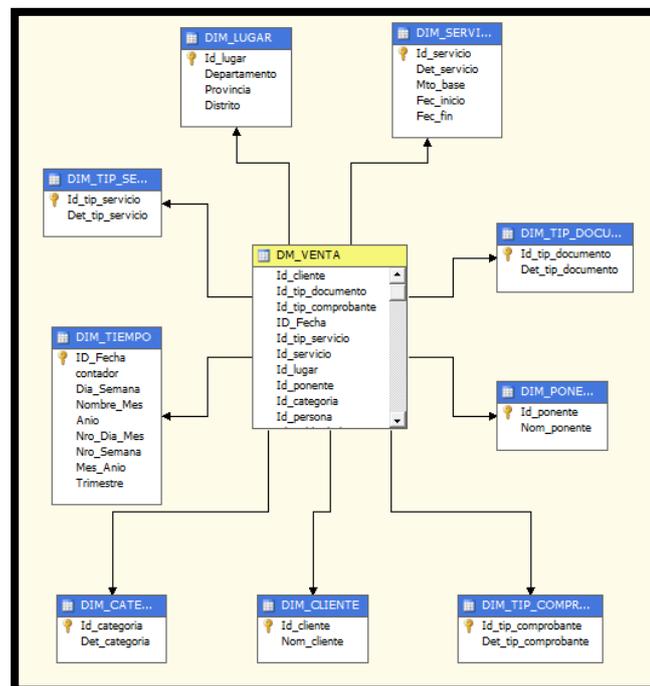
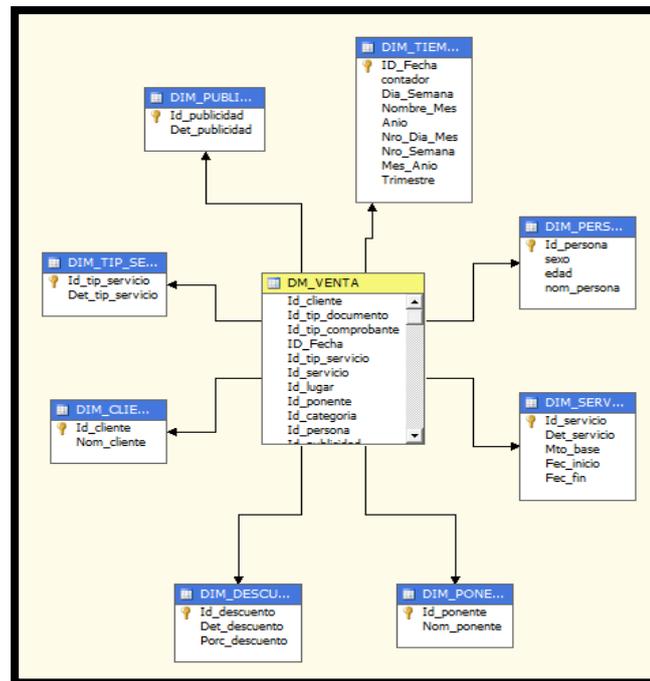
Una vez que nos muestra el mensaje de “*Deployment Completed Successfully*” sabremos que se estableció la conexión y el cubo ya puede ser utilizado por alguna herramienta de BI para el análisis de datos.

Figura 31: Conexión establecida para el Cubo



Fuente: Elaboración Propia

Figura 32: Visualización de los cubos generados

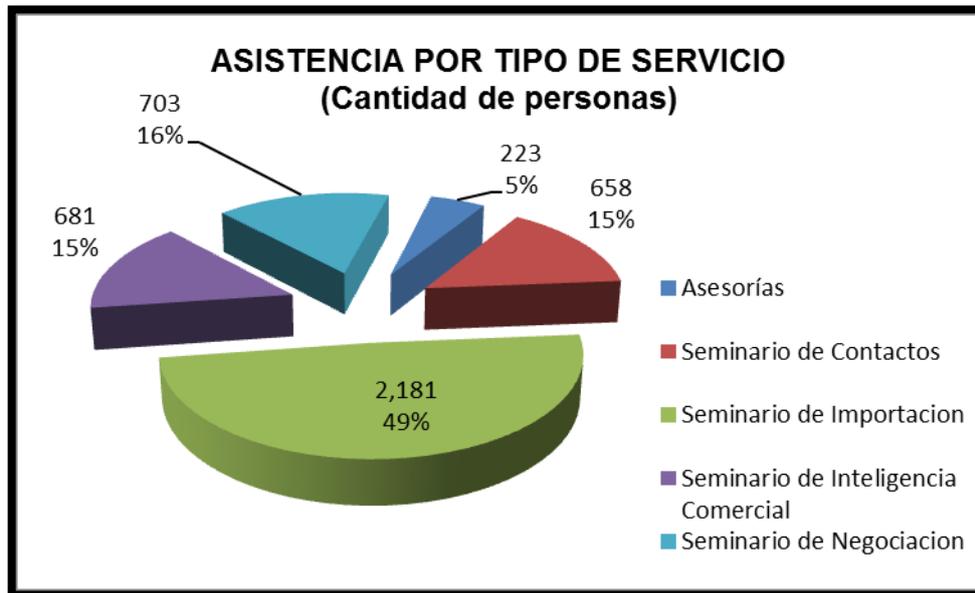


Fuente: Elaboración Propia

3.3. REVISIÓN Y CONSOLIDACIÓN DE RESULTADOS

- Se han diseñado y llevado a la práctica los modelos conceptual, lógico y físico que definen el cubo multidimensional que ha dado soporte al almacén de datos.
- Se han generado procesos automatizados de creación de tablas, transformación y carga de datos.
- Se han definido índices que aceleran las consultas sobre la base de datos.
- Como un agregado a la Implementación del Data Mart en base a los requerimientos de la Empresa se generaron algunos gráficos en Excel que son muestra del correcto funcionamiento de nuestro Data Mart.

Gráfico 1: Asistencia por tipo de Servicio a nivel nacional - Año 2013

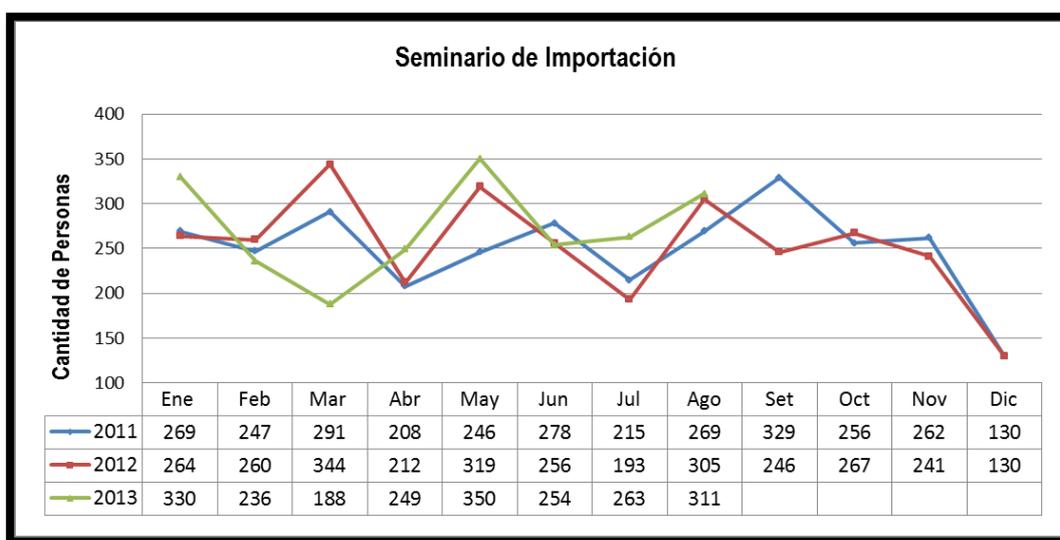


Fuente: Elaboración Propia

Esta información será útil para las dos áreas estudiadas: Ventas y Marketing, ya que al analizar el tipo de Servicio más asistido se podrían tomar decisiones como:

- Cuando realizar más eventos del Seminario más asistido.
- Metas para el año en curso respecto al Seminario más asistido.
- Añadir más publicidad a los Seminarios menos asistidos.
- Qué medidas estratégicas tomar para aumentar la asistencia de los Seminarios menos asistidos y más asistidos.
- Tener presente el Tipo de Servicio que le genera mayores ingresos y conservar o mejorar su tendencia.

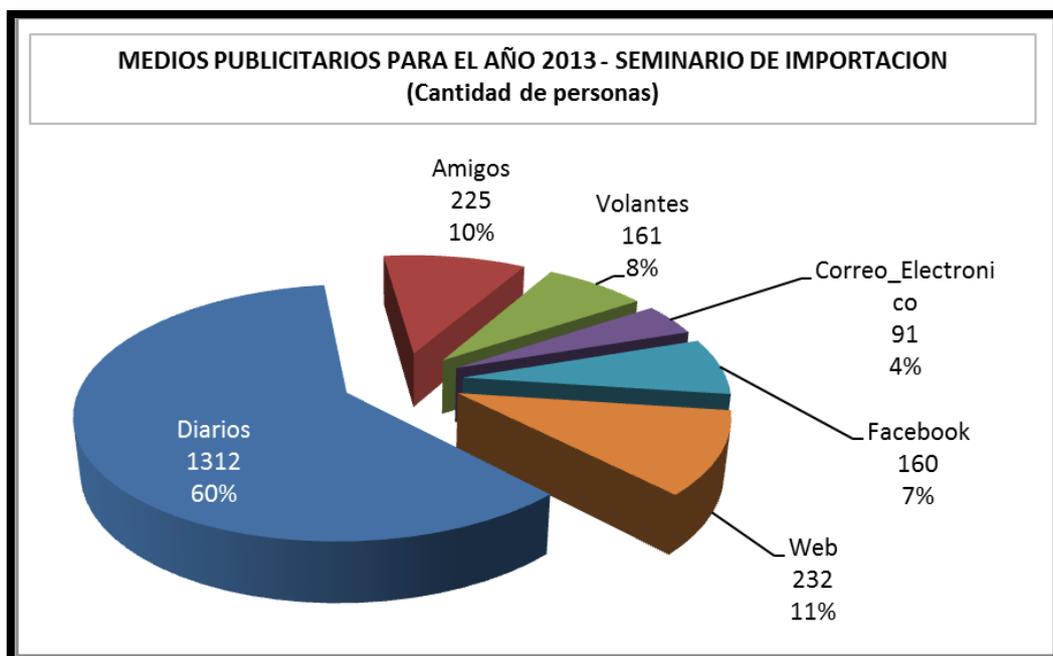
Gráfico 2: Cantidad de Personas que asistieron al Seminario de Importación a nivel nacional – Vista a detalle por mes



Fuente: Elaboración Propia

Este gráfico muestra la tendencia del Seminario más asistido: Seminario de Importación, en base a la cantidad de personas que asisten al evento. La visualización está determinada por mes y cada línea de color representa el año estudiado. Como se puede observar el año 2013 solo cuenta con información hasta el mes de Agosto, porque este es el último periodo del que se obtuvo la data. También se puede observar la caída de cantidad de asistentes al seminario en los meses de Julio y Diciembre, fechas importantes para el Perú, donde generalmente las personas invierten más en su hogar que en pagar cursos o capacitaciones.

Gráfico 3: Medios Publicitarios en Seminario de Importación a nivel nacional



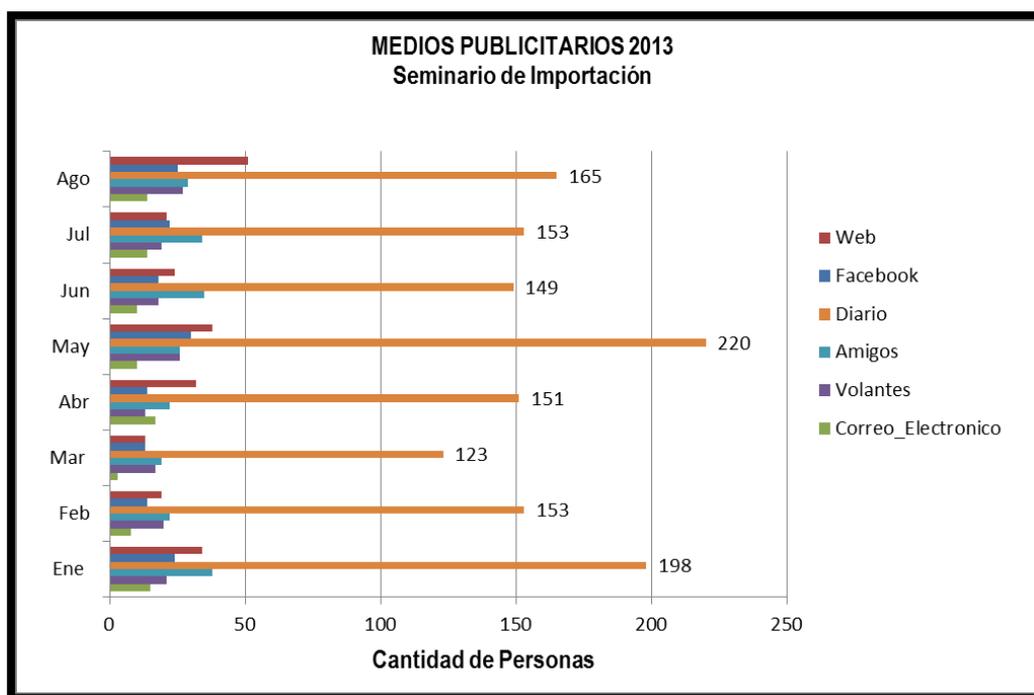
Fuente: Elaboración Propia

Este gráfico contiene información de gran importancia para el área de Marketing se puede observar aquí la publicidad que más se ha destacado en el 2013. El medio Publicitario por el cual las personas generalmente se entera de los Seminarios de Importación.

En el gráfico siguiente se puede observar la acogida de los medios publicitarios para el año 2013 en cada mes. Se destaca claramente que la publicidad por Diario es la que funciona mejor.

Gracias a esta información la Gerencia, ya podría comenzar a enfocarse en un medio publicitario como estrategia de captación de asistentes, sin dejar de lado los demás medios de publicidad, por el contrario, añadir nuevos temas para mejorar la efectividad de estos.

Gráfico 4: Medios Publicitarios en Seminario de Importación a nivel nacional por meses



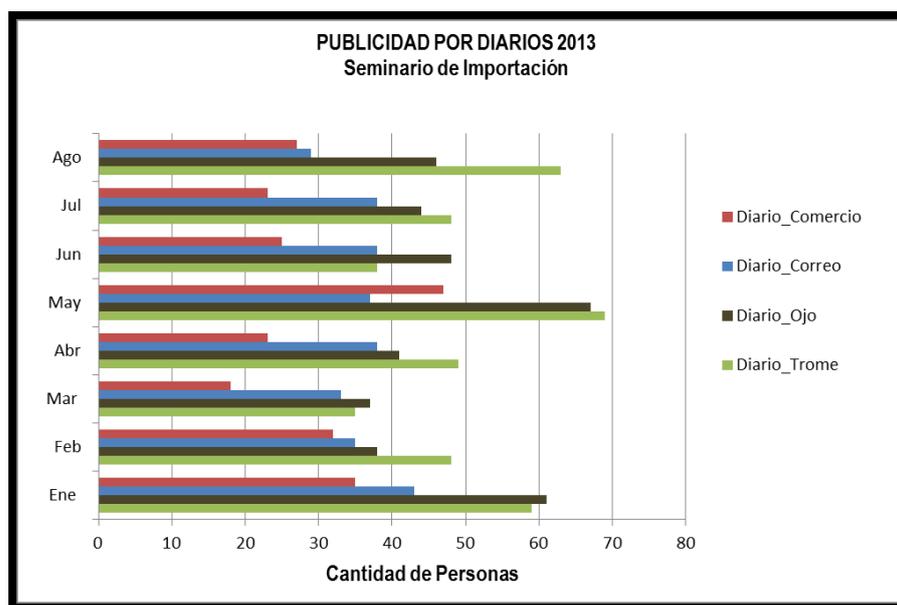
Fuente: Elaboración Propia

También se podría comenzar a analizar la cantidad de publicaciones o campañas que se hizo por Diarios en cada mes y contrastar con este gráfico, así se podría saber si tales campañas tuvieron éxito con respecto a las tendencias que se muestran.

El gráfico que viene muestra únicamente los diarios, utilizados como medios de publicidad y su acogida por mes en el año 2013 para el Seminario más importante con el que cuenta Inpechi, Aprender a Importar.

Se puede observar que publicar en Ojo y Trome tiene mejores resultados que en los demás, lo más seguro porque son diarios más baratos y cuentan con una gran cantidad de lectores. Analizar las secciones donde salieron publicadas también ayudaría mucho contrastando con esta información para las próximas publicaciones.

Gráfico 5: Publicidad por diarios en Seminario de Importación a nivel nacional



Fuente: Elaboración Propia

A continuación una pequeña tabla creada en Excel con las principales empresas de Categoría A y su participación en los diversos Seminarios, esto permitirá observar que clientes son los más concurrentes a los Seminarios, y más adelante poder considerarlos para invitarlos a eventos nuevos, o para lo que la empresa consigne necesario la ayuda de esta tabla.

Cuadro 4: Principales clientes de la categoría A (Expresado en cantidad de personas)

Empresa	Tipos de Seminario					Total
	Asesorías	Seminario de Contactos	Seminario de Importación	Seminario de Inteligencia Comercial	Seminario de Negociación	
empresa 1243			8		1	9
empresa 1547			7			7
empresa 1128		1	5	1		7
empresa 106	1	1			4	6
empresa 1225			6			6
empresa 972			6			6
empresa 1082			5			5
empresa 1397		2	2	1		5
empresa 866			5			5
empresa 1282			5			5
empresa 1463		1	2	1	1	5
empresa 1301			2	1	2	5
empresa 1582			1	1	3	5
empresa 1303			5			5
empresa 1353			5			5

Fuente: Elaboración Propia

CONCLUSIONES

1. Con esta aplicación del Datamart y los cubos generados se podrá analizar el comportamiento de la cartera de clientes frecuentes, el crecimiento de asistentes por año, los ingresos anuales, mensuales o por periodos; la influencia de sus medios de comunicación utilizados como marketing para difundir sus seminarios y si han tenido o no llegada en sus participantes.
2. Se podrá tomar mejores decisiones en cuanto al uso de los recursos, promociones y personal apoyados por una herramienta estratégica de gran alcance como lo es un Datamart.
3. La creación de este Data Mart permitirá la integración de casi toda la información que almacena Inpechi en sus bases de datos y contribuir de esta forma a optimizar sus operaciones, reduciendo el tiempo en el proceso y por lo tanto disminución de ciertos costos.
4. Se generó dos cubos: Ventas y Marketing para que ambos se apoyen al momento de generar los reportes, es decir, que la información que sea consultada sea más completa y variada, de este modo será de gran apoyo al análisis previo para la toma de decisiones.
5. Los usuarios contarán con información real y oportuna, sin que exista la dependencia de otra aplicación.
6. Los reportes fueron generados con Microsoft Excel 2010, sin embargo la utilización de esta herramienta es opcional, pues existen diversas herramientas para Inteligencia de Negocios que son aún más útiles. En este proyecto se utilizó Excel ya que es un programa con el que cuenta la empresa y esto lo hace un gasto menos en la compra de nuevas licencias.
7. Es importante la correcta definición de las jerarquías de las dimensiones del modelo de la base de datos, ya que de lo contrario se podría obtener data incompleta y/o errónea al momento de implementar una herramienta de Inteligencia de Negocios sobre todo en la explotación de los cubos, lo cual se solucionaría mediante el re-modelamiento de la base de datos y los cubos.

RECOMENDACIONES

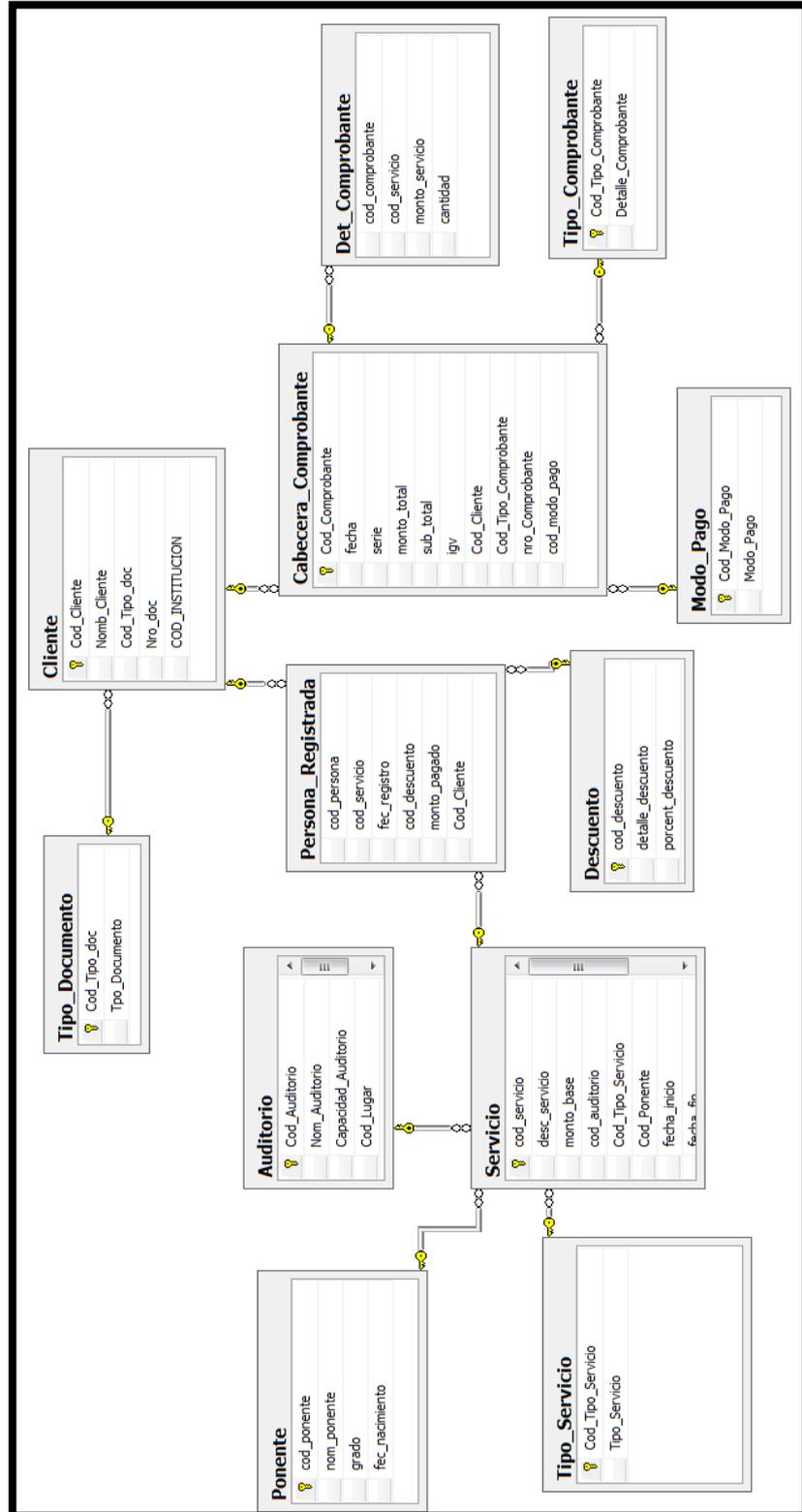
1. Se recomienda que la solución siga creciendo con la adición de información de todas las demás fuentes de origen con las que cuenta Inpechi, al modelo ya construido.
2. Se debe mantener la administración y continuidad del seguimiento a la carga de información hacia el Datamart, a fin de que se mantenga actualizado y realmente sea de beneficio para los usuarios finales y tomadores de decisiones.
3. Es importante que la información ingresada a la base de datos y con la cual es llenada la Fact Table sea la original, es decir datos verdaderos y no manipulados o cambiados, pues los reportes que luego se generarán no serán fiables y éste será un motivo principal de malas decisiones que se puedan tomar.
4. Se puede optar por utilizar una herramienta de BI para el análisis y presentación de datos más complejos, conectándolo al Datamart creado.
5. Como parte del ciclo de vida de la inteligencia de negocio, es recomendable crear un grupo encargado para que vele por el seguimiento de mejoras y adiciones al Datamart. También es importante que este grupo vele por el control de calidad de la información que se tendrá disponible dentro de las estructuras del Datamart.

BIBLIOGRAFÍA

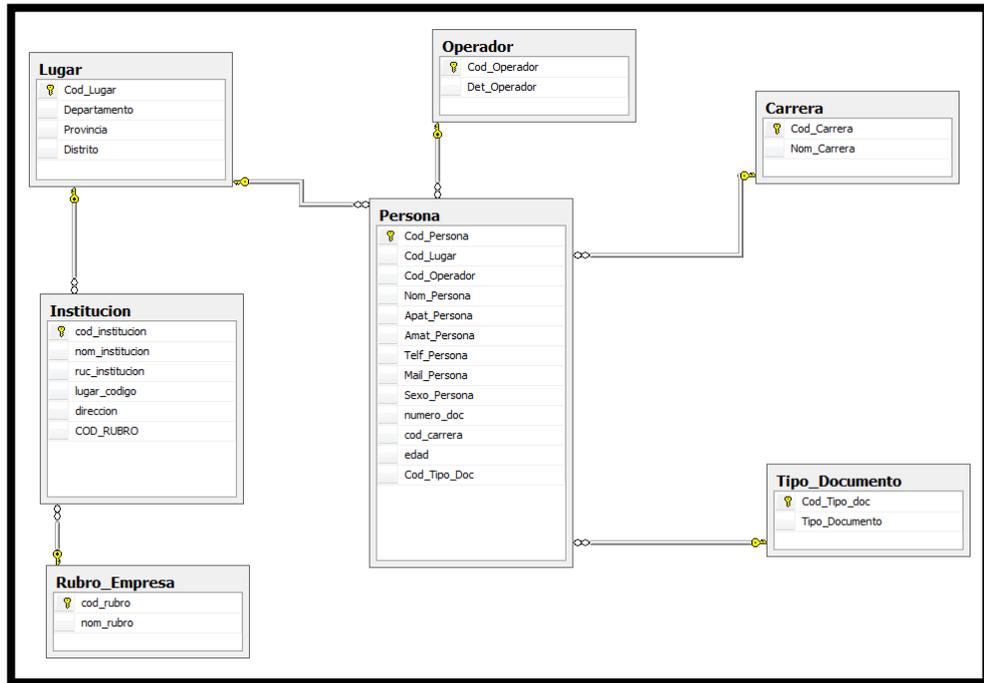
1. Rosales Sedano, Carmen. "Análisis, diseño e implementación de un Data Mart para el soporte de toma de decisiones y evaluación de las estrategias sanitarias en las direcciones de salud". Pontificia Universidad Católica del Perú. Perú. 2009
2. Fernández Ochoa, Eddy. "Análisis, diseño e implementación de un Data Mart de clientes para el área de marketing de una entidad aseguradora." Pontificia Universidad Católica del Perú. Perú. 2009
3. Zambrano Alarcón, Jaime. "Análisis, diseño e implementación de un Data Mart para el área de mantenimiento y logística de una empresa de transporte público de pasajeros". Pontificia Universidad Católica del Perú. Perú. 2011
4. Guillén Rodríguez, Fiorelly. "Desarrollo de un Data Mart para mejorar la toma de decisiones en el área de tesorería de la Municipalidad Provincial de Cajamarca". Universidad Privada del Norte. Perú 2012
5. López Beltrán, Carlos. "Análisis, diseño e implementación de un data MART para la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército para una toma de decisión efectiva". Escuela Politécnica del Ejército. Ecuador.2007
6. Arias Cuevas, Jaime. "Diseño y construcción de un Data Mart para el análisis de preferencias del usuario en la web a partir de datos originados en el portal Educarchile". Universidad de Chile. 2012
7. Aimacaña Quilumba. "Diseño e Implementación de un Data Dart académico usando tecnología de BI para la facultad de ingeniería, ciencias físicas y matemática". Universidad Central del Ecuador. Ecuador 2013.

ANEXOS

MODELO ENTIDAD-RELACION DE LA BD_SERVICIO TRANSACCIONAL DE INPECHI

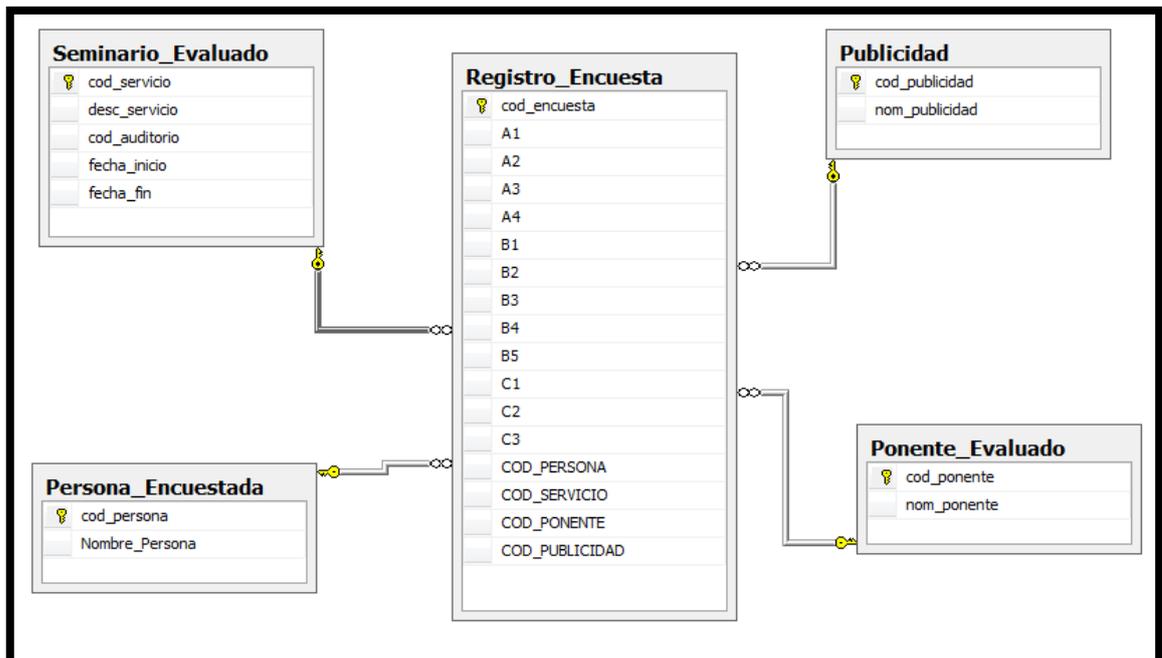


MODELO ENTIDAD-RELACION DE LA BD_REGISTRO DE INPECHI



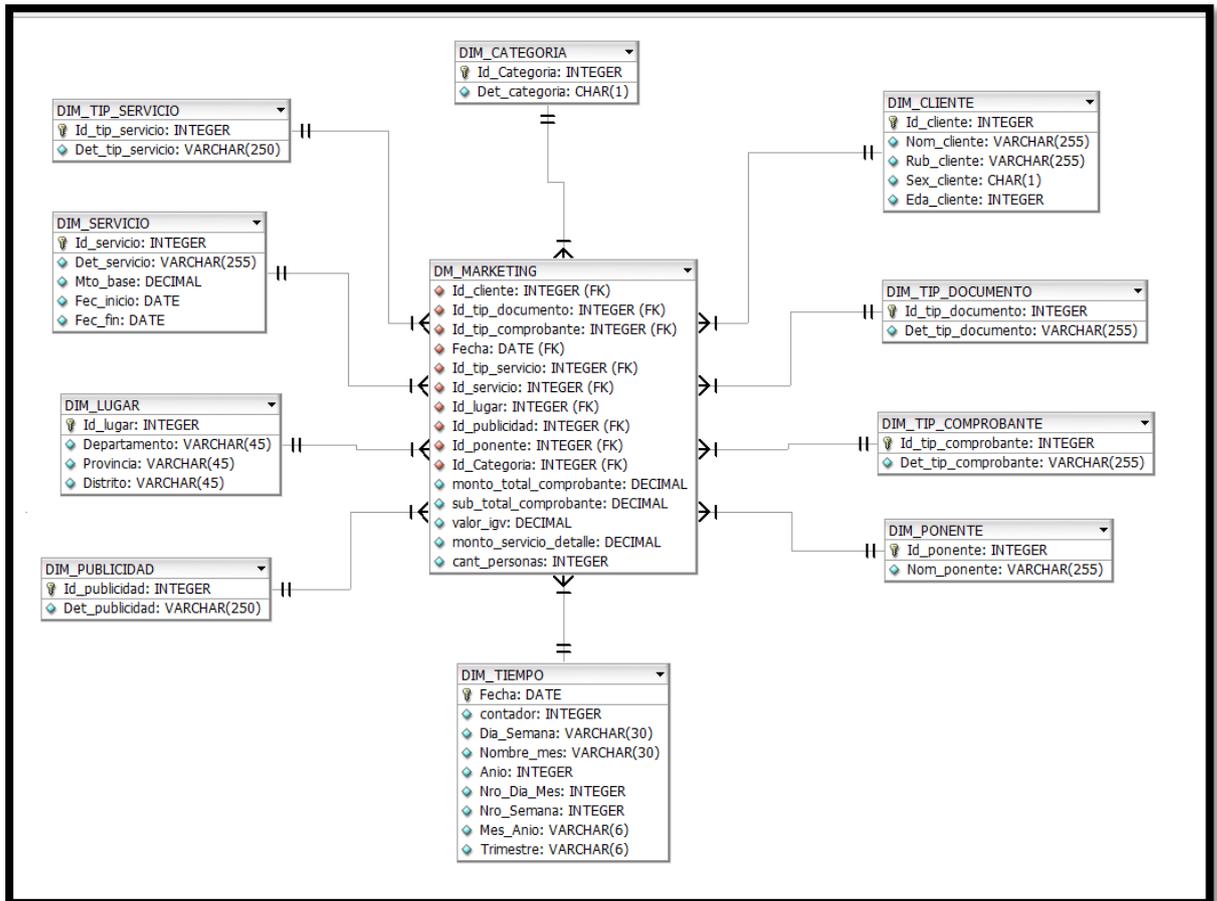
Fuente: Elaboración Propia

MODELO ENTIDAD-RELACION DE LA BD_ENCUESTA DE INPECHI



Fuente: Elaboración Propia

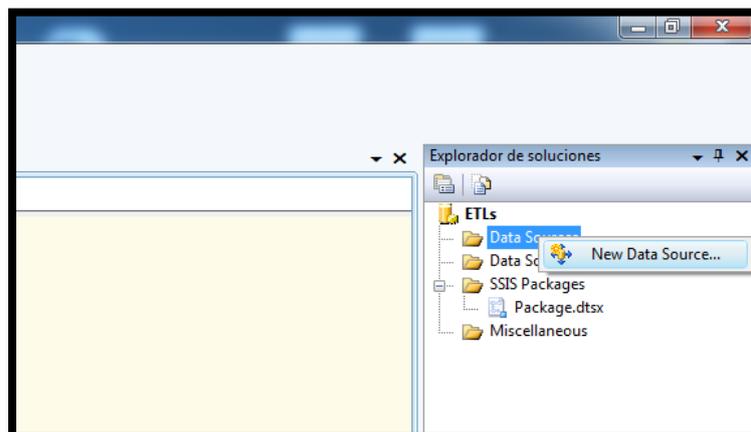
MODELO DE CONEXIÓN ENTRE LAS DIMENSIONES Y LA FACT_TABLE

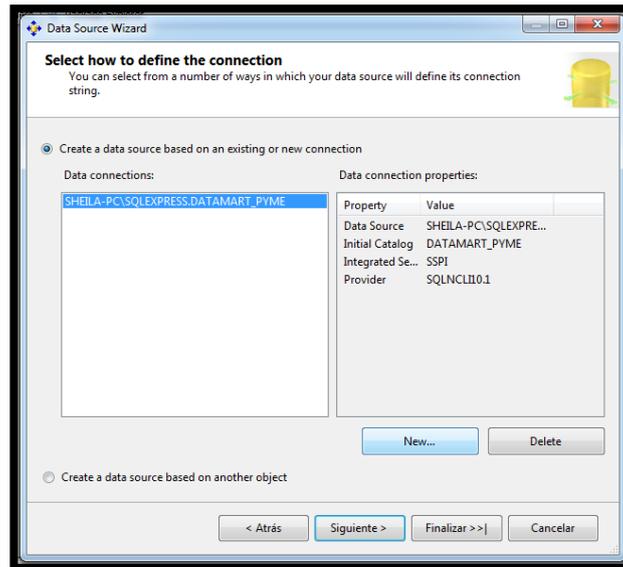


Fuente: Elaboración Propia

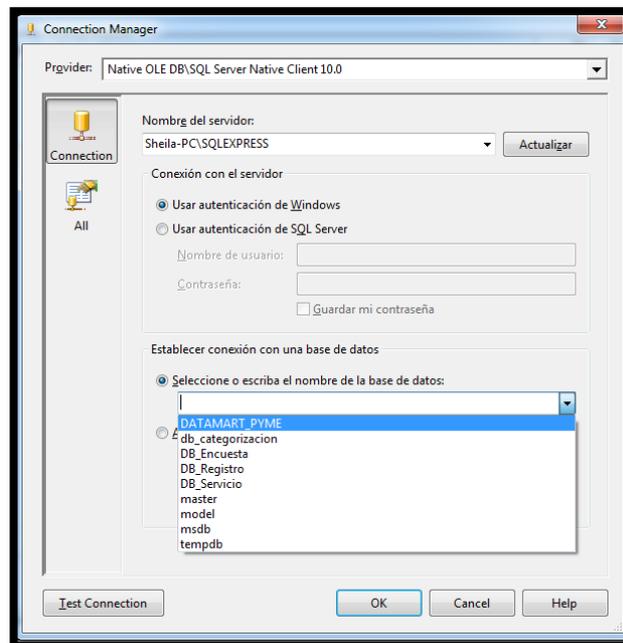
Para comenzar a ejecutar los ETLs a través del Microsoft Visual Studio, con Integration Services creamos los ETLs con los siguientes pasos.

Configuración para acceder a la base de datos de interés

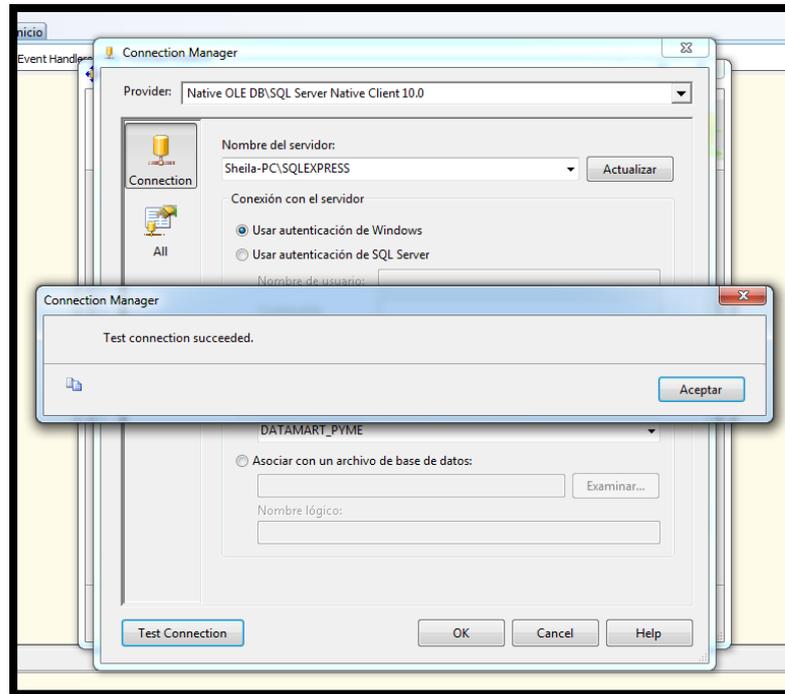




Escribimos el nombre del servidor y escogemos la base de datos donde se encuentran las tablas a cargar con data.



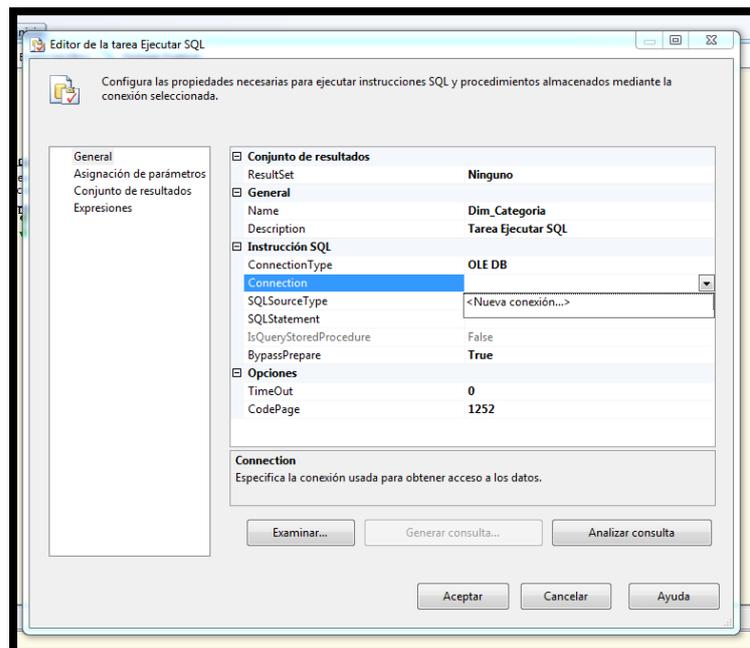
Damos en test connection



Para confirmar que todo está bien y le damos en aceptar

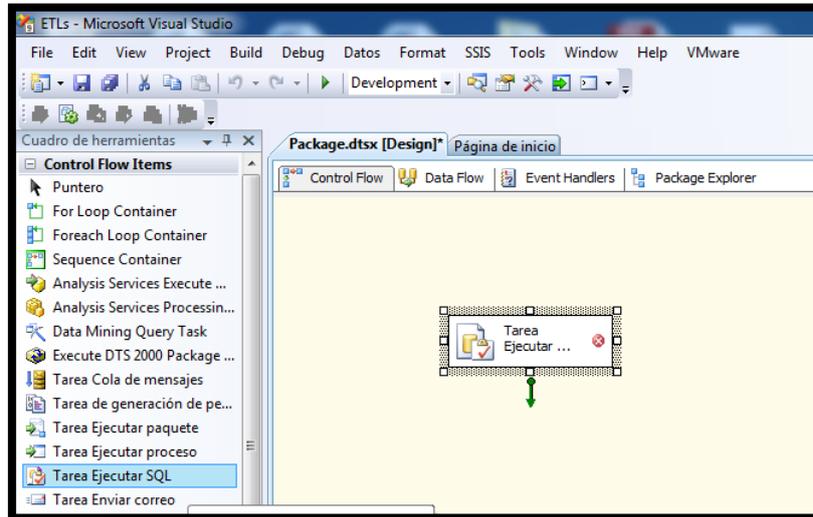
LUEGO → OK → FINALIZAR → FINISH

Luego volver a dar doble clic en la tarea EJECUTAR SQL y Configurar en Connection → nueva conexión...

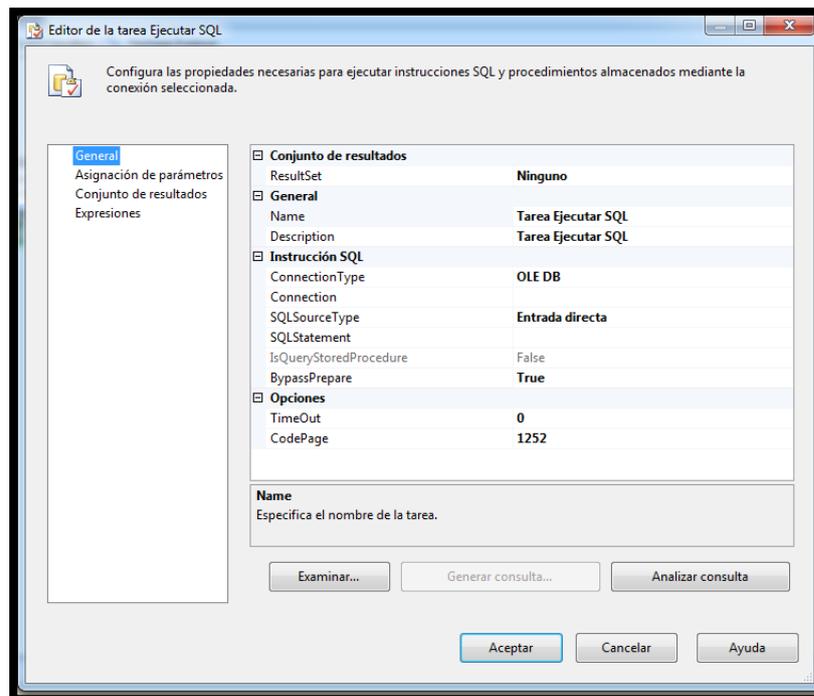


Seleccionar el servidor y dar ok. Y por último terminar de aceptar.

Ahora podemos comenzar a generar los ETL en el Visual Studio, doble click a la tarea creada.

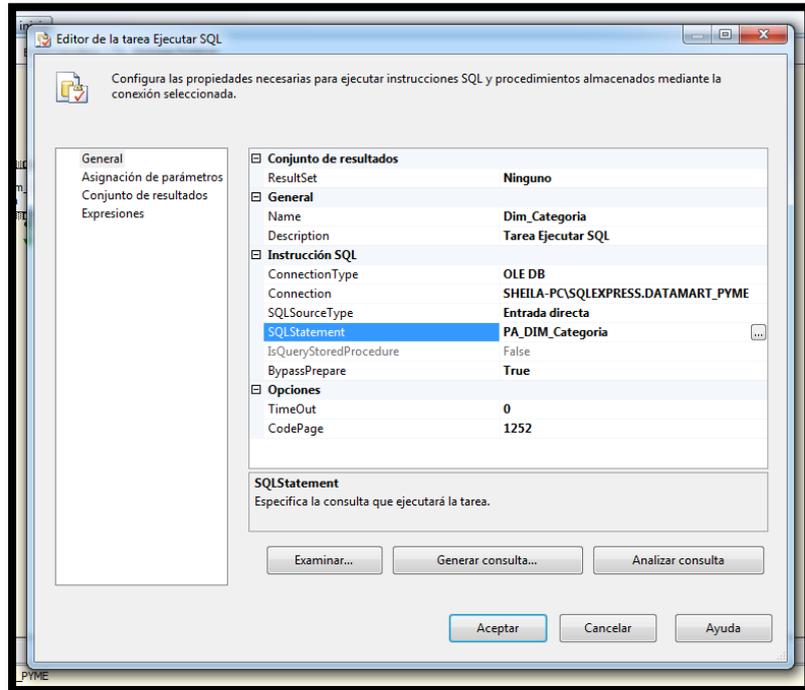


Y configuramos la ventana nueva que sale

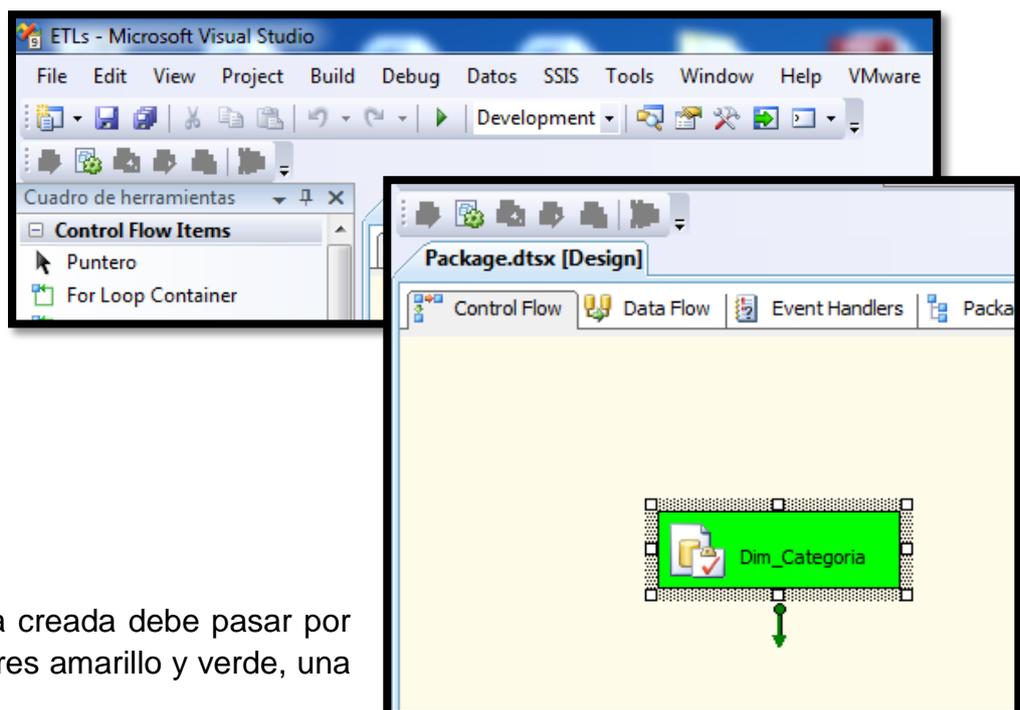


Configuración en:

Y en SQL STATEMENT: va el nombre del procedimiento almacenado anteriormente creado para cada dimensión. Y le damos en Aceptar.



Por ultimo damos en ejecutar en la parte superior



La Tarea creada debe pasar por dos colores amarillo y verde, una

vez que llego a verde quiere decir que el proceso se ejecutó correctamente, de lo contrario se pondrá en rojo y se tendrá que verificar el error para corregirlo.

Y así generamos todas las demás tareas para ejecutar los ETLs de las diferentes DIMENSIONES guardados en procedimientos almacenados dentro de nuestra base de datos en DM_MARKETING.

Finalmente cuando ya se tienen toda las dimensiones se procede a ejecutarlo masivamente o como el usuario prefiera (tal vez uno por uno)

