

**UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR**

**FACULTAD DE INGENIERÍA Y GESTIÓN**

**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**



**“DESARROLLO DE UN DATAMART PARA MEJORAR LA TOMA  
DE DECISIONES EN ÁREA DE MARKETING EN UNA  
EMPRESA DE TRANSPORTE”**

**TRABAJO DE SUFICIENCIA PROFESIONAL**

Para optar el Título Profesional de

**INGENIERO DE SISTEMAS**

**PRESENTADO POR EL BACHILLER**

**ZAMUDIO MARTINEZ, LUIS ORLANDO**

**Villa El Salvador**

**2017**

## **DEDICATORIA**

Dedicado a mis padres y hermanos que son pilares en mi vida, a mi novia por el gran apoyo.

## **AGRADECIMIENTOS**

Primero y más valioso agradecimiento son para mis padres Héctor Zamudio y Ana Martínez quienes con sus ejemplos y enseñanzas me han permitido alcanzar lo poco o mucho que he logrado.

A mis hermanos y sobrinas pues siempre han tenido maneras de apoyarme en todo este largo camino.

A todos mis profesores que nos compartieron sus conocimientos y experiencias para poder crecer como profesionales y personas.

## INDICE

INTRODUCCIÓN .....	8
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA.....	10
1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA .....	10
1.2. JUSTIFICACIÓN DEL PROYECTO.....	12
1.3. DELIMITACIÓN DEL PROYECTO.....	13
1.3.1. DELIMITACIÓN ESPACIAL.....	13
1.3.2. DELIMITACIÓN TEMPORAL .....	13
1.4. FORMULACIÓN DEL PROBLEMA .....	13
1.4.1. PROBLEMA PRINCIPAL .....	13
1.4.2. PROBLEMAS ESPECÍFICOS .....	13
1.5. OBJETIVOS .....	14
1.5.1. OBJETIVO GENERAL .....	14
1.5.2. OBJETIVOS ESPECÍFICOS.....	14
CAPITULO II: MARCO TEÓRICO.....	15
2.1. ANTECEDENTES DE LA INVESTIGACIÓN .....	15
2.1.1. ANTECEDENTES INTERNACIONALES .....	15
2.1.2. ANTECEDENTES NACIONALES .....	16
2.2. BASES TEÓRICAS.....	20
2.2.1. DATAMART .....	20
2.2.2. DATAWAREHOUSE.....	21
2.2.3. MODELO MULTIDIMENSIONAL .....	22
2.2.4. ETL .....	25
2.2.5. INTELIGENCIA DE NEGOCIOS .....	26
2.2.6. METODOLOGÍA KIMBALL .....	28
2.3. MARCO CONCEPTUAL.....	32
2.3.1. ANÁLISIS DE INFORMACIÓN.....	32
2.3.2. ANALYSIS SERVICES.....	32
2.3.3. ARQUITECTURA.....	32
2.3.4. BASE DE DATOS .....	33
2.3.5. DATAMART .....	33
2.3.6. DATAWAREHOUSE.....	33
2.3.7. DIMENSIÓN .....	33
2.3.8. INDICADORES.....	34

2.3.9.	INTELIGENCIA DE NEGOCIOS .....	34
2.3.10.	TABLA HECHOS .....	34
	CAPÍTULO III: DESARROLLO DE LA METODOLOGÍA .....	35
3.1.	PLANIFICACIÓN DEL PROYECTO .....	35
3.2.	DEFINICIÓN DEL REQUERIMIENTO DEL NEGOCIO .....	36
3.3.	MODELO DIMENSIONAL.....	37
3.3.1.	ELEGIR EL PROCESO DEL NEGOCIO .....	37
3.3.2.	ESTABLECER EL NIVEL DE GRANULARIDAD .....	37
3.3.3.	ELEGIR LAS DIMENSIONES .....	38
3.3.4.	IDENTIFICAR LAS TABLAS DE HECHOS Y MEDIDAS.....	42
3.3.5.	MODELO GRÁFICO DE ALTO NIVEL .....	43
3.4.	DISEÑO FISICO.....	44
3.5.	DISEÑO E IMPLMENTACIÓN DEL SUBSISTEMA ETL.....	45
3.5.1.	FLUJO DE DATOS DIM_USUARIO.....	46
3.5.2.	FLUJO DE DATOS CATEGORIA.....	48
3.5.3.	FLUJO DE DATOS UBIGEO.....	50
3.5.4.	FLUJO DE DATOS DIM_METODO_PAGO.....	52
3.5.5.	FLUJO DE DATOS DESCUENTO .....	54
3.5.6.	FLUJO DE DATOS TIPO_MONEDA .....	56
3.5.7.	FLUJO DE DATOS TIEMPO .....	58
3.5.8.	FLUJO DE DATOS DIM_HORA .....	60
3.5.9.	FLUJO DE DATOS HECHOS_TRAYECTO .....	62
3.6.	SELECCIÓN DE PRODUCTOS.....	64
3.7.	DESARROLLO DE LA APLICACIÓN BI .....	65
3.7.1.	DESARROLLO DEL CUBO .....	65
3.7.2.	DESARROLLO POWER BI .....	68
	CONCLUSIONES .....	72
	RECOMENDACIONES.....	74
	BIBLIOGRAFIA .....	75

## Lista de figuras

Figura N° 1 - Componentes de un Datawarehouse.....	22
Figura N° 2 - Estructura de un cubo.....	24
Figura N° 3 - Proceso ETL.....	25
Figura N° 4 - Metodología de Ralph Kimball.....	29
Figura N° 5 - Modelo gráfico de alto nivel.....	43
Figura N° 6 - Diseño físico Datamart.....	44
Figura N° 7 - Diseño ETL.....	45
Figura N° 8 - Flujo de datos USUARIO.....	46
Figura N° 9 - OLDB Source USUARIO.....	47
Figura N° 10 - Flujo de datos CATEGORIA.....	48
Figura N° 11 - OLDB Source CATEGORIA.....	49
Figura N° 12 - Flujo de datos UBIGEO.....	50
Figura N° 13 - OLDB Source UBIGEO.....	51
Figura N° 14 - Flujo de datos METODO PAGO.....	52
Figura N° 15 - OLDB Source METODO PAGO.....	53
Figura N° 16 - Flujo de datos DESCUENTO.....	54
Figura N° 17 - OLDB Source DESCUENTO.....	55
Figura N° 18 - Flujo de datos TIPO MONEDA.....	56
Figura N° 19 - OLDB Source MONEDA.....	57
Figura N° 20 - Flujo de datos TIEMPO.....	58
Figura N° 21 - OLDB Source TIEMPO.....	59
Figura N° 22 - Flujo de datos HORA.....	60
Figura N° 23 - OLDB Source HORA.....	61
Figura N° 24 - Flujo datos HE_TRAYECTO.....	62
Figura N° 25 - OLDB Source HE_TRAYECTO.....	63
Figura N° 26 - Conexión vista.....	65
Figura N° 27 - Agregando Dimensiones.....	66
Figura N° 28 - Generación del cubo.....	67
Figura N° 29 - Reporte Trayectos por Perfil Usuario.....	68
Figura N° 30 - Reporte Metodos de Pago.....	69
Figura N° 31 - Reporte Códigos.....	70
Figura N° 32 - Reporte Tipo de Trayecto.....	71

## Listado de Tablas

Tabla N° 1 - Priorización de Procesos. ....	36
Tabla N° 2 - Dimensión Tiempo. ....	38
Tabla N° 3 - Dimensión Ubigeo ....	39
Tabla N° 4 - Dimensión Hora ....	39
Tabla N° 5 - Dimensión Método Pago.....	40
Tabla N° 6 - Dimensión Descuento. ....	40
Tabla N° 7 - Dimensión Usuario. ....	41
Tabla N° 8 - Dimensión Categoría. ....	41
Tabla N° 9 - Dimensión Tipo Moneda. ....	41
Tabla N° 10 - Hechos Trayecto.....	42

## INTRODUCCIÓN

Cabify es una empresa de redes de transporte a nivel internacional que conecta a usuarios con vehículos por medio de una app móvil o desde su página web.

Creada en el año 2011 en la actualidad opera en 12 países y 43 ciudades. Cada país cuenta con un Country Manager quién toma decisiones a nivel país y cada ciudad cuenta con un City Manager encargado de velar la funcionalidad de Cabify en esa ciudad. El crecimiento obtenido supero todas las expectativas ha generado que la empresa no establezca una uniformidad para el manejo de su información, lo cual es vital, fundamentalmente para el área de marketing pues son ellos quien lanzan las promociones y estudian a los usuarios, basados en esta información para poder seguir creciendo en el mercado que cada día se encuentra con mayor competencia por el incremento de creaciones de apps muy similares a Cabify.

El presente proyecto de ingeniería permitirá desarrollar un Datamart para mejorar la toma de decisiones, tendremos la posibilidad de establecer una uniformidad en el manejo de información, establecer perfiles de usuarios para promociones seleccionadas así como brindar unos informes con mayor impacto facilitando la toma de decisión y así poder comenzar nuevas estrategias en base a la información brindada.

Para la presentación de los informes y visualización de la información utilizaremos la herramienta de Inteligencia de negocios POWER BI pues es una herramienta de fácil manejo y gran impacto en las plantillas visuales para análisis el cual nos facilitara el desarrollo de las mismas y se tomara como país modelo a Perú.

El presente trabajo está organizado de la siguiente manera: el primer capítulo presenta el planteamiento del problema, el segundo capítulo presenta el marco teórico que considera una revisión de los antecedentes y las bases teóricas utilizadas en el desarrollo del proyecto de investigación; el tercer capítulo se desarrolla el Datamart basado en la metodología de Ralph Kimball, y por último se obtiene conclusiones y recomendaciones de la investigación.

## **CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA**

### **1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA**

Actualmente, el manejo de la información es vital para las decisiones de inversiones y estrategias en todas las empresas, tener conocimiento de los patrones y comportamientos de los clientes permite a la empresa tener decisiones acertadas. En la empresa Cabify se cuenta con 12 *Country Manager* y 42 *City Manager* los cuales a diario tiene que decidir qué tipo de promociones lanzar para capturar nuevos usuarios o mantener a los actuales por lo cual solicitan información al área de Data como base para sus siguientes decisiones en la creación de nuevas promociones o ampliar las promociones actuales.

En el área de Data la extracción e integración de los datos para la creación de reportes se realiza manualmente extrayendo la información, esto lleva a posibles errores en la información brindada por que se cuenta con una base de datos simple que no cuenta con una estructura orientada al análisis, además para generar la información la estructura de la base de datos no permite realizar consultas con respuestas más cortas.

Basados en los resultados obtenidos de los últimos códigos promocionales lanzados en Perú (durante el mes de Enero) se concluyó que existe un 20% que no retribuyen la inversión destinada para ese código promocional , lo cual genera perdida en la empresa al realizar una mala inversión, esto es por qué no se tiene identificado adecuadamente los indicadores apropiados para poder tomar con mejor criterio las decisiones, indicadores como el CAC (Costo de Adquisición por usuario), las cantidades y tipos de trayectos realizados, los diferentes métodos de pagos, todos estos datos arrojan diferentes perfiles de usuarios para diferentes tipos de promociones.

## 1.2. JUSTIFICACIÓN DEL PROYECTO

Toda empresa que crece siempre es por las decisiones que se toma, para invertir o para no invertir hay que tener en cuenta varios factores propios como externos del negocio, Cabify no es ajeno a esto y tanto los *City* y *Country Manager* buscan que estas decisiones que a diario toman sean las más acertadas para el beneficio de la empresa captando nuevos clientes y potenciando los actuales, creando estrategias en base a los datos brindados por el equipo de Data.

La creación de un Datamart en el área de Marketing de la empresa Cabify permitirá que la información brindada sea la más útil para la toma de decisiones teniendo una arquitectura amigable para las consultas, generando respuesta en menor tiempo, además de centralizar la información más relevante para la toma de decisiones con mayor grado de confianza en los datos pues serán generados desde la misma base de datos de producción de la empresa. Desde la cantidad de trayectos realizados por usuarios que activaron una promoción hasta el tiempo de retorno de inversión, son ejemplos de información que se brindara al equipo de Marketing para poder identificar los diferentes perfiles de los usuarios y así poder realizar lanzamientos de promoción segmentada. Para complementar y tener un mayor impacto visual se utilizará como herramienta de Power BI pues es muy fácil manejo para el usuario y tiene gran impacto visual.

### **1.3. DELIMITACIÓN DEL PROYECTO**

#### **1.3.1. DELIMITACIÓN ESPACIAL**

El área en estudio es Marketing de la empresa es Cabify sucursal Perú ubicada en el distrito de Miraflores, Lima.

#### **1.3.2. DELIMITACIÓN TEMPORAL**

El desarrollo del proyecto tendrá una duración de 4 meses empezando en el mes de Marzo hasta el mes de Junio del 2017.

### **1.4. FORMULACIÓN DEL PROBLEMA**

#### **1.4.1. PROBLEMA PRINCIPAL**

¿De qué manera el desarrollo de un datamart permitirá mejorar la toma de decisiones en el área de Marketing de la empresa Cabify?

#### **1.4.2. PROBLEMAS ESPECÍFICOS**

- ¿Cómo una base de datos multidimensional permitirá analizar la información para mejorar la toma de decisiones en el área de Marketing de la empresa Cabify?
  
- ¿Cómo los indicadores serán una buena fuente de información para mejorar la toma de decisiones en el área de Marketing de la empresa Cabify?

- ¿Cómo se podría representar los indicadores en una manera visual para mejorar la toma de decisiones en el área de Marketing de la empresa Cabify?

## **1.5. OBJETIVOS**

### **1.5.1. OBJETIVO GENERAL**

**1.5.1.1.** Desarrollar un datamart para la mejora en la toma de decisiones en el área de Marketing de la empresa Cabify.

### **1.5.2. OBJETIVOS ESPECÍFICOS**

**1.5.2.1.** Realizar una base de datos multidimensional para analizar la información que permita mejorar la toma de decisiones en el área de Marketing de la empresa Cabify.

**1.5.2.2.** Generar indicadores como salidas de información que permitan mejorar la toma de decisiones en el área de Marketing de la empresa Cabify.

**1.5.2.3.** Expresar de manera gráfica los indicadores para mejorar la toma de decisiones en el área de Marketing de la empresa Cabify con la herramienta de Inteligencia de negocios Power BI.

## **CAPITULO II: MARCO TEÓRICO**

### **2.1. ANTECEDENTES DE LA INVESTIGACIÓN**

#### **2.1.1. ANTECEDENTES INTERNACIONALES**

- *“DataMart de estadísticas medicas”* presentado por la ingeniera Ordoñez Leidi. La revolución de las tecnologías de información ha tenido un profundo efecto en la administración de las organizaciones, mejorando la habilidad de los administradores para coordinar y controlar las actividades de la organización y ayudándolos a tomar decisiones mucho más efectivas. Hoy en día, el uso de las TI, se ha convertido en un componente central de toda empresa o negocio que busque un crecimiento sostenido, el desafío constante por resolver las necesidades de información oportuna y confiable para los usuarios, finaliza con la implementación de un Datamart. La mayoría de los usuarios no necesita acceder más a los datos actuales operacionales, porque ellos tienen información más útil disponible desde el Datamart.

El aporte que encontramos en esta tesis es la importancia de un Datamart para poder brindar información exacta de lo que requieren los usuarios, además nos sirve como base para posterior desarrollo de minería de datos en la empresa. (Cabrera, 2007)

### **2.1.2. ANTECEDENTES NACIONALES**

- *“Solución integral para explotar eficientemente la información de los contactos con los clientes utilizando Datamart en telefónica del Perú”* presentado por el Ingeniero Alejandro Gamarra, Los sistemas de soporte a las decisiones son una necesidad cada vez más apremiante en las empresas ante un mercado más competitivo muestra el proceso seguido para el desarrollo de un Datamart cuyo objetivo será facilitar la información de los contactos con el cliente de Telefónica del Perú a todos los usuarios internos de la empresa y proporcionar las herramientas necesarias que ayuden a la correcta explotación y uso eficiente de la misma. La presencia de este Datamart contribuirá a comprender, aún mejor, cuáles son las necesidades reales del cliente y contribuir a la satisfacción de las mismas. El aporte que encontramos en esta tesis es el de saber que tan valiosa es la información para la captura y mantenimiento de los clientes. (Ramirez, 2011)

- *“Construcción de un Datamart que apoye en la toma de decisiones de la gestión de incidencias en una mesa de ayuda: caso Consorcio Peruano de empresas”*, presentado por los ingenieros Alfaro Mendoza, Luis Alfonso y Paucar Moreyra, Daphné Aurelia, Construye un Datamart basado en el funcionamiento de la gestión de incidencias de la mesa de ayuda del Consorcio Peruano de Empresas con el objetivo de obtener tendencias e indicadores de grandes cantidades de datos de los incidentes reportados por los usuarios a la mesa de ayuda. Para el análisis rápido y oportuno se generan reportes de manera automática con gráficos estadísticos y dashboards, con esto el usuario logra identificar los focos de incidencias de manera rápida y tomar las acciones correctivas que sean necesarias para mitigarlos ya que el impacto de las incidencias reportadas a la mesa de ayuda afecta la operatividad de los usuarios y esto impacta de manera directa a la operación del servicio de la organización. El aporte que encontramos en esta tesis es que automatizando los reportes empleando una herramienta a través del datamart diseñado se optimizará los tiempos en la toma de decisiones. (Mendoza & Moreyra, 2016)

- *“Desarrollo de un datamart para mejorar la toma de decisiones en el área de tesorería de la Municipalidad Provincial de Cajamarca”*, presentado por la ingeniera Guillén Rodríguez, Fiorelly Shirley, La Municipalidad Provincial de Cajamarca es una institución que maneja una gran cantidad de información día a día. Sin embargo, no saben cómo administrar

adecuadamente todas las cantidades recaudadas, debido a que su sistema actual no soporta el manejo adecuado de grandes volúmenes de información. Así la Municipalidad Provincial de Cajamarca tiene el problema de utilizar su información para emplearla en la toma de decisiones. El objetivo principal es proveer una solución de Inteligencia de Negocios que de soporte a las necesidades de información hacia los usuarios finales, que vienen a ser las demás áreas de la Municipalidad Provincial de Cajamarca, del área de Tesorería. La solución consistirá 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. El aporte que encontramos en esta tesis es que parte fundamental del desarrollo es el proceso de ETL donde los datos son procesados y depurados para ingresar al Datamart. (Rodríguez, 2012)

- *“Análisis, diseño e implementación de un datamart para el soporte de toma de decisiones y evaluación de las estrategias sanitarias en las direcciones de salud”* presentado por la ingeniera Rosales Sedano, Carmen Pamela. El presente proyecto de tesis tiene como objetivo la implementación de un datamart que permita apoyar la toma de decisiones necesarias para cumplir con los objetivos específicos de cada estrategia sanitaria nacional dentro de las direcciones de salud. Se tomó como

estrategia sanitaria piloto, la estrategia de Alimentación y Nutrición Saludable. Para ella, se realizó la captura de requerimientos, análisis, diseño y construcción del datamart. Los resultados obtenidos son los reportes analíticos útiles para tomar decisiones de forma rápida y confiable. El aporte que encontramos en esta tesis es que la generación de reportes por parte del datamart para cada estrategia permite el ahorro de tiempo, pues actualmente cada estrategia debe solicitar a un área de Data lo cual implica sobrecarga en dicha área. (Sedano, 2009)

- “Desarrollo de un datamart para optimizar la generación de información estratégica *de apoyo a la toma de decisiones en la Vicepresidencia de Banca Comercial de Interbank Perú*”, presentado por los ingenieros Espinoza Jenny y Palomino Carlos. Desarrollan un datamart que tiene por objetivo disminuir el alto consumo de recursos para optimizar la generación de información estratégica para el apoyo a la toma de decisiones en la Vicepresidencia de Banca Comercial Interbank Perú. La metodología que es la de Ralph Kimball. El datamart logró disminuir el tiempo de generación de información estratégica y reducir labores de carga operativa a recursos especializados. El aporte que encontramos en esta tesis son los beneficios económicos que nos brindara la creación de un correcto datamart, disminuyendo tiempo en generación de reportes y mejora en el análisis. (Vargas & Ruiz, 2016)

## **2.2. BASES TEÓRICAS**

### **2.2.1. DATAMART**

Se puede definir como un Datawarehouse sólo que en menor volumen; es decir, es un Datawarehouse orientada a algún tema. Los Datamart en la mayoría de ocasiones son usados por un departamento o grupo de usuarios en una empresa para un conjunto definido de tareas.

Se caracteriza por disponer de una estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicha área de negocio. Normalmente un datamart se alimenta a partir de los datos de un Datawarehouse. Existen 2 tipos de datamarts: OLAP y OLTP. (Buyto,2009)

#### **- OLTP - On-Line Transactional Processing**

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones que pueden involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales. El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura, como por ejemplo la enorme cantidad de transacciones que tienen que soportar diariamente las BD de bancos o hipermercados.

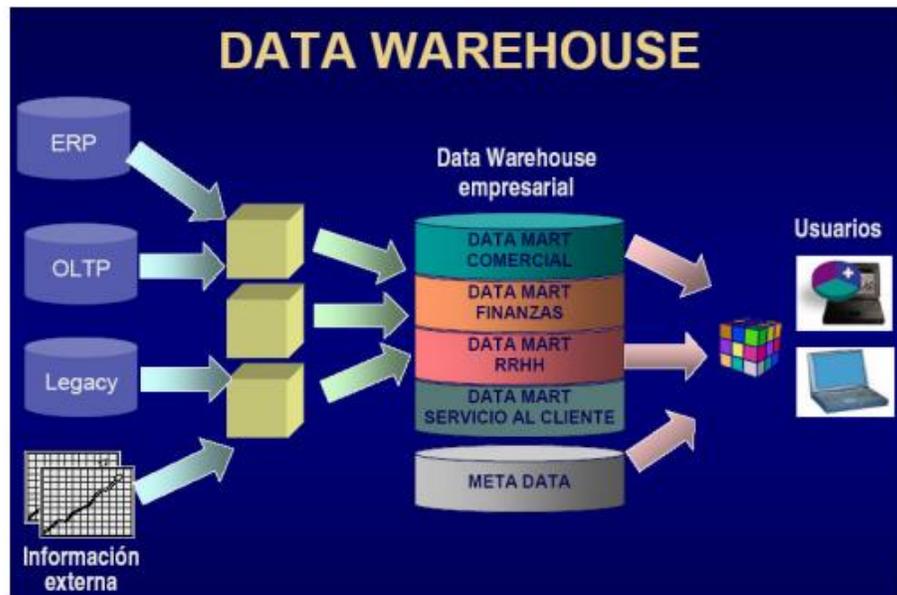
- **OLAP - On-Line Analytical Processing**

Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil como por ejemplo: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos, etc. El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones. Los datos se estructuran según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.

### **2.2.2. DATAWAREHOUSE**

Es un almacén de datos relacionados a las actividades de una organización y grabadas en una base de datos diseñada específicamente con el propósito de hacer informes para después analizar estos informes y conseguir información estratégica. El Datawarehouse tiene como objetivo agrupar los datos de toda la empresa con el fin de facilitar su análisis, de forma que sean útiles para acceder y analizar información sobre la propia empresa. Un Datawarehouse contiene la información de toda la empresa. Cualquier departamento puede acceder a la información de cualquier otro departamento mediante un único medio, así como obligar a que los mismos términos tengan el mismo significado para todos. (HUAMANTUMBA, 2007)

Figura Nº 1 - Componentes de un Datawarehouse.



Fuente: HUAMANTUMBA, 2007

### 2.2.3. MODELO MULTIDIMENSIONAL

Un modelo de datos es un sistema formal y abstracto que permite describir los datos de acuerdo con reglas y convenios predefinidos. Es formal, pues los objetos del sistema se manipulan siguiendo reglas definidas y utilizando los operadores definidos. (Alarcón, 2011)

La arquitectura de Datawarehouse se basa en un modelo de datos llamado modelo multidimensional. Este modelo permite modelar base de datos simples y entendibles al usuario final, debido que presenta la información en un marco estándar e intuitivo que permite un acceso de alto rendimiento. Además, permite resolver con problemas planteados en sistemas transaccionales.

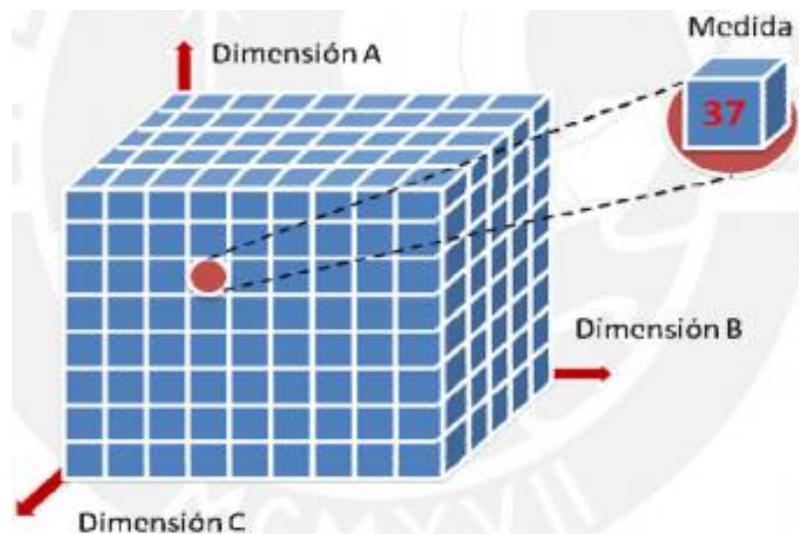
El modelo multidimensional está compuesto por dos componentes:

- El primer componente son las tablas que a su vez se dividen en dos tipos: Tablas de hechos y de dimensiones. Las tablas de hechos constituyen el objeto a analizar, poseen atributos de hechos que son del tipo cuantitativo cuyos valores se obtienen por aplicación de alguna función estadística que resumen un conjunto de valores en un único valor. Las tablas de dimensiones representan cada uno de los ejes en un espacio multidimensional. Sus atributos son del tipo cualitativo que proporcionan el contexto en el que se obtienen las medidas en un esquema de hecho. Las dimensiones poseen jerarquías, que son varios atributos unidos mediante una relación del tipo jerárquico.
  
- El segundo componente son los esquemas. Existen dos esquemas también: Esquema estrella y de copo de nieve o constelación de hechos. El esquema estrella forma un diagrama en forma de estrella teniendo en el centro de la estrella una o más tablas de hechos y las puntas de las estrellas a las tablas de dimensiones. En el caso del esquema de copo de nieve, las tablas de dimensiones se encuentran normalizadas, es decir, cada tabla dimensional sólo contiene el nivel que es la clave primaria en la tabla y la llave foránea de su parentesco del nivel más cercano.

La estructura básica del modelo multidimensional se conoce como Cubo o Hipercubo, ya que la información se representa como una matriz multidimensional, en los ejes de esta matriz se encuentran los criterios de análisis y en los cruces están los valores a analizar. Los Cubos están formados por:

- **Dimensiones:** Representan los criterios de análisis de los datos. Si una dimensión tiene más de un nivel entonces los miembros de la dimensión pueden ser organizados en una o más jerarquías.
- **Medida:** Dato numérico que representa una actividad específica de un negocio, mientras que una dimensión representa una perspectiva de los datos. Una medida contiene una propiedad numérica y fórmula.

Figura Nº 2 - Estructura de un cubo.

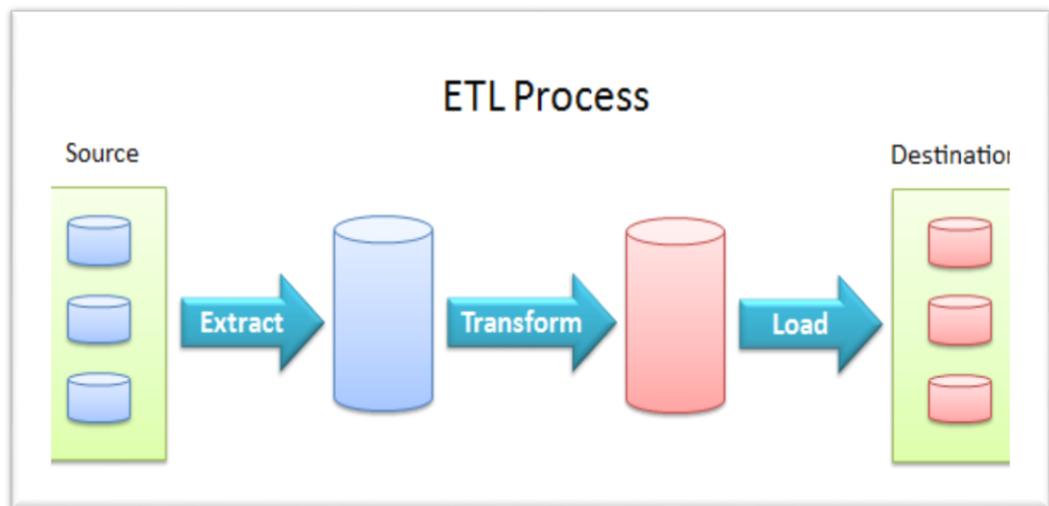


Fuente: Alarcón, 2011

#### 2.2.4. ETL

Los procesos Extracción, transformación y carga de datos o en sus siglas en ingles *Extract, Transform Load*. Son una parte de la integración de datos, pero es un elemento importante cuya función completa el resultado de todo el desarrollo de aplicaciones y sistemas.

Figura Nº 3 - Proceso ETL



Fuente: (INGENIMA, 2016)

- **Extracción:** En este proceso se extrae los datos desde el sistema de origen analizándolos y evaluándolos para verificar que estos datos cumplan con la estructura esperada de caso contrario los datos deberían ser rechazados y no tomados en cuenta para este proceso, finalmente convertir estos datos a un formato específico para poder iniciar el siguiente proceso.

- **Transformación:** En este proceso a los datos extraídos se les aplica una serie de reglas o funciones para tener el formato adecuado para la posterior carga.
- **Carga:** En esta etapa los datos extraídos del proceso de transformación son cargados al sistema 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 donde se realiza un resumen de todas las transacciones comprendidas en el periodo seleccionado y transporta 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 o Rolling dónde se busca tener varios niveles de granularidad.

### **2.2.5. INTELIGENCIA DE NEGOCIOS**

Es el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa. Este conjunto de herramientas y metodologías tienen en común las siguientes características:

- **Accesibilidad a la información:** Los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos con independencia de la procedencia de estos.
- **Apoyo en la toma de decisiones:** Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- **Orientación al usuario final:** Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

## 2.2.6. METODOLOGÍA KIMBALL

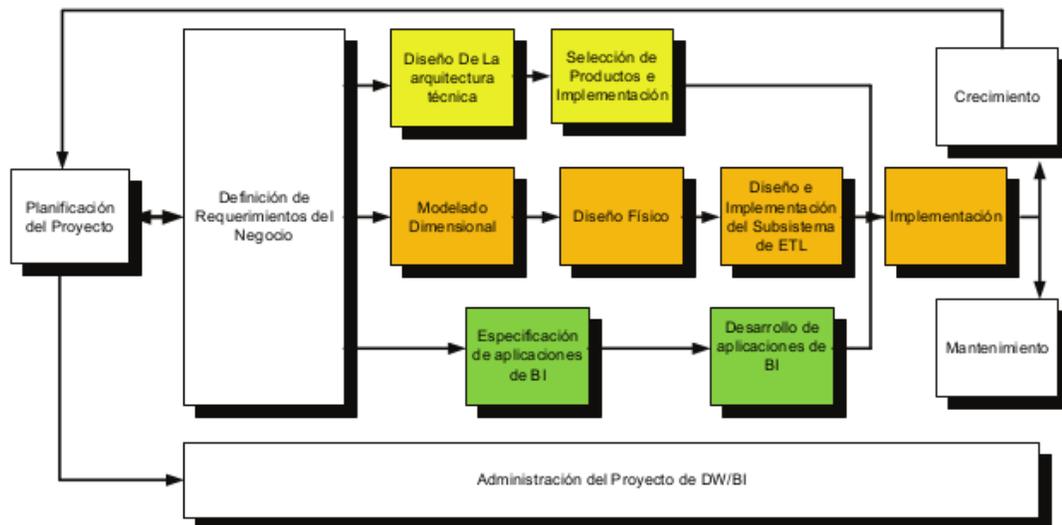
La Metodología Kimball, es una metodología empleada para la construcción de un almacén de datos (Datawarehouse, DW) que no es más que, una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. (Rivadera ,2014)

La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este ciclo de vida del proyecto de Datawarehouse, está basado en cuatro principios básicos:

- Centrarse en el negocio
- Construir una infraestructura de información adecuada
- Realizar entregas en incrementos significativos
- Ofrecer la solución completa.

La construcción de una solución de Datawarehouse/Business Intelligence es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. Las tareas de esta metodología (ciclo de vida) se describen a continuación:

Figura Nº 4 - Metodología de Ralph Kimball.



Fuente: (Rivadera ,2014)

- **Planificación del proyecto:** Busca identificar la definición y el alcance que tiene el proyecto de Datawarehouse. Esta etapa se concentra sobre la definición del proyecto, donde, a nivel de planificación, se establece la identidad del mismo, el personal, desarrollo del plan de proyecto, el seguimiento y la monitorización.
- **Definición de los requerimientos del negocio:** Es un factor determinante en el éxito de un proceso de DWH. Los diseñadores de los Datawarehouse deben tener en claro cuáles son los factores claves que guían el negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.

- **Modelado dimensional:** Se comienza con una matriz donde se determina la dimensión de cada indicador para luego especificar los diferentes grados de detalle dentro de cada concepto del negocio.
- **Diseño físico:** Se centra en 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 en esta etapa.
- **Diseño e implementación del subsistema ETL:** Tiene como principales actividades la extracción, transformación y carga (ETL). Estas actividades son altamente críticas ya que tienen que ver con la materia prima del Datawarehouse que son los datos.
- **Diseño de la arquitectura técnica:** En esta fase se deben tener en cuenta tres factores: los requerimientos de negocio, los actuales entornos técnicos, y las directrices técnicas y estratégicas futuras planificadas por la compañía, lo que permitirá establecer el diseño de la arquitectura técnica del entorno del Datawarehouse.
- **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.

- **Especificación de aplicaciones BI:** Se identifican los roles o perfiles de usuarios para los diferentes tipos de aplicaciones necesarias en base al alcance de los perfiles detectados.
- **Desarrollo de aplicaciones BI:** Involucra configuraciones de los metadatos y construcción de reportes específicos.
- **Implementación:** Representa el correcto funcionamiento de la tecnología, los datos y las aplicaciones de usuarios finales accesibles para el usuario del negocio.
- **Mantenimiento y crecimiento:** Se basa en la necesidad de continuar con las actualizaciones de forma constante para así lograr la evolución de las metas por conseguir.
- **Administración del proyecto de BI:** Asegura que todas las actividades del ciclo de vida se lleven a cabo de manera sincronizada.

## **2.3. MARCO CONCEPTUAL**

### **2.3.1. ANÁLISIS DE INFORMACIÓN**

El análisis de información parte desde la simple recopilación y lectura de textos hasta la interpretación. Es decir, el análisis es una actividad intelectual que logra el arte o la virtud de perfeccionar capacidades profesionales por parte del analista; todo esto gracias al empleo de métodos y procedimientos de investigación, ya sean cuantitativos o cualitativos que le permiten separar lo principal de lo accesorio y lo trascendental de lo pasajero o superfluo.

### **2.3.2. ANALYSIS SERVICES**

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

### **2.3.3. ARQUITECTURA**

Trata sobre 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.

#### **2.3.4. BASE DE DATOS**

Se le llama base de datos a los bancos de información que contienen datos relativos a diversas temáticas y categorizados de distinta manera, pero que comparten entre sí algún tipo de vínculo o relación que busca ordenarlos y clasificarlos en conjunto.

#### **2.3.5. 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.

#### **2.3.6. DATAWAREHOUSE**

Es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta.

#### **2.3.7. DIMENSIÓN**

Una dimensión de base de datos es una colección de objetos relacionados, denominados atributos, que se pueden usar para proporcionar información sobre los datos de hechos de uno o varios cubos. Por ejemplo, los atributos típicos de una dimensión de producto pueden ser el nombre, la categoría, la línea, el tamaño y el precio del

producto. Estos objetos están enlazados a una o varias columnas de una o varias tablas de una vista del origen de datos. De manera predeterminada, estos atributos están visibles como jerarquías de atributo y se pueden utilizar para comprender los datos de hechos en un cubo. Los atributos se pueden organizar en jerarquías definidas por el usuario que proporcionan rutas de navegación para ayudar a los usuarios al examinar los datos de un cubo.

### **2.3.8. INDICADORES**

Un indicador es una comparación entre dos o más tipos de datos que sirve para elaborar una medida cuantitativa o una observación cualitativa.

Esta comparación arroja un valor, una magnitud o un criterio, que tiene significado para quien lo analiza.

### **2.3.9. INTELIGENCIA DE NEGOCIOS**

Es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

### **2.3.10. TABLA HECHOS**

Los hechos son los indicadores de negocio que dan sentido al análisis de las dimensiones. Las tablas de hechos incluyen los indicadores asociados a un proceso de negocio en concreto y las claves de las dimensiones que intervienen en dicho proceso, en el mínimo nivel de granularidad o detalle.

## **CAPÍTULO III: DESARROLLO DE LA METODOLOGÍA**

### **3.1. PLANIFICACIÓN DEL PROYECTO**

En el proyecto nos enfocaremos en la información obtenida por las interacciones de los usuarios con la aplicación móvil Cabify, dicha información permite al personal de Marketing la toma de decisiones.

El único responsable del proyecto es quien redacta este trabajo escrito, actualmente me encuentro laborando en la empresa como Data Analyst.

### 3.2. DEFINICIÓN DEL REQUERIMIENTO DEL NEGOCIO

La metodología de Kimball nos indican que definir los requerimientos es un factor clave para el éxito del proyecto, es por esta razón que debemos analizar a quién entrevistar para que nos puedan brindar la información necesaria sobre los requerimientos que puedan causar mayor impacto en la empresa. En la empresa Cabify vengo trabajando más de 6 meses y en este tiempo pude determinar que la persona capaz de brindarme información precisa es Giovanni Bedoya quien trabaja como Head of Marketing Strategy.

A continuación, una matriz previa de los procesos del negocio versus las dimensiones.

**Tabla N° 1 - Priorización de Procesos.**

PROCESO DEL NEGOCIO	DIMENSIÓN							
	Tiempo	Ubigeo	Formas de pago	Descuentos	Usuario	Categorías	Hora	Tipo Moneda
Gestión de Marketing	X	X	X	X	X	X	X	X
Captación de Choferes	X	X			X	X		X
Gestión de Finanzas	X			X		X		X

Fuente: Elaboración propia

Como se comentó en la entrevista se pudo determinar los requerimientos más relevantes para Marketing los cuales especificamos a continuación:

- Cantidad de trayectos realizados en periodos de tiempo determinados por tipos.
- Cantidad de trayectos, usuarios, ingresos y descuentos por categorías en un tiempo determinado.
- Resultados de códigos, cantidad de personas que activaron la promoción, cantidad de usuarios nuevos, cantidad de usuarios nuevos con trayectos, cantidad de trayectos, ingresos y descuentos obtenidos por el código.
- Segmentación de usuarios por cantidad de trayectos en un tiempo determinado.

### **3.3. MODELO DIMENSIONAL**

#### **3.3.1. ELEGIR EL PROCESO DEL NEGOCIO**

Revisando la Tabla N° 1 Priorización de Procesos podemos determinar que el proceso prioritario es **Gestión de Marketing** por lo cual este proceso será el que se desarrollara.

#### **3.3.2. ESTABLECER EL NIVEL DE GRANULARIDAD**

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

### 3.3.3. ELEGIR LAS DIMENSIONES

- **Dimensión Tiempo:** Nos permitirá evaluar los indicadores en un tiempo determinado, también incluimos horas pues para el rubro de la empresa es importante tener en cuenta las horas donde se genera mayor demanda en la aplicación.

Tabla Nº 2 - Dimensión Tiempo.

DIM_TIEMPO			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_TIEMPO	INT IDENTITY (1,1)	ID Tiempo	1,2,..
FECHA	DATETIME	Fecha	DD/MM/YYYY
AÑO	INT	Número del año	2016,2017...
SEMESTE_NUM	INT	Número del Semestre	1 o 2
TRIMESTRE_NUM	INT	Número del trimestre	1, 2,3,4
BIMESTRE_NUM	INT	Número del Bimestre	1,2,3,4,5,6
MES_NUM	INT	Número del mes	1,2,3,...12
MES_NOM	VARCHAR(10)	Nombre del mes	Enero,Febrero...
DIA_NUM	INT	Número del día	1,2,...7
DIA_SEM	VARCHAR(10)	Nombre del día	Lunes,Martes...

Fuente: Elaboración propia

- **Dimensión Ubigeo:** Nos permitirá evaluar indicadores a nivel de ciudad o país según sea el criterio a tomar, permitiendo crear cuadros universales donde pueda filtrar cada *Marketing Manager* los resultados de su ciudad.

**Tabla Nº 3 - Dimensión Ubigeo.**

DIM_UBIGEO			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_UBIGEO	INT IDENTITY (1,1)	ID Tiempo	1,2,..
COD_UBIGEO	CHAR(4)	Código Ubigeo	1212,1220,...
COD_PAIS	CHAR(2)	Código país	12,13,...
NOM_PAIS	VARCHAR(100)	Nombre del país	Perú, México, ...
COD_CIUADAD	CHAR(2)	Código de la ciudad	12,20,...
NOM_CIUADAD	VARCHAR(100)	Nombre de la ciudad	Lima, Cali, ...

Fuente: Elaboración propia

- **Dimensión Hora:** Nos permitirá determinar la hora del trayecto para poder analizar la hora más puntas según filtros establecidos.

**Tabla Nº 4 - Dimensión Hora.**

DIM_HORA			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_HORA	INT IDENTITY (1,1)	ID Trayecto	1,2,..
HORADIA	INT	Hora del día	1,2,..

Fuente: Elaboración propia

- **Dimensión Método Pago:** Nos permitirá evaluar la forma de pago más usada por nuestros usuarios.

**Tabla Nº 5 - Dimensión Método Pago.**

DIM_METODO_PAGO			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_METODO_PAGO	INT IDENTITY (1,1)	ID Método pago	1,2,..
COD_METODO_PAGO	INT	Código de Método pago	1212,1220, ...
METODO_PAGO	VARCHAR(50)	Nombre Método Pago	Paypal, TC,..
COD_TIPO_PAGO	CHAR(2)	Código Tipo pago	12,13,...
TIPO PAGO	VARCHAR(100)	Nombre de Tipo de Pago	Paypal, TC,..
COD_TARJETA	CHAR(2)	Código Tipo tarjeta	12,13,...
TARJETA	VARCHAR(100)	Nombre de Tarjeta	MC, VISA, CASH...

Fuente: Elaboración propia

- **Dimensión descuento:** Nos permitirá evaluar el impacto y la efectividad obtenidos por los códigos promocionales lanzados por el área de Marketing.

**Tabla Nº 6 - Dimensión Descuento.**

DIM_DESCUENTO			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_DESCUENTO	INT IDENTITY (1,1)	ID Descuento	1,2,..
COD_DESCUENTO	CHAR(4)	Código Descuento	1212,1220,...
COD_TIPO_DESCUENTO	CHAR(4)	Código Tipo de Descuento	1212,1220,...
NOM_TIPO_DESCUENTO	VARCHAR(100)	Nombre del Tipo Descuento	Invitaciones, código crédito,...
NOM_DESCUENTO	VARCHAR(100)	Nombre descuento	INVITACABIFY, ACTIVAPE,...

Fuente: Elaboración propia

- **Dimensión Usuario:** Nos permitirá segmentar los usuarios por su tipo de role (privados o corporativos) debido a que las promociones a lanzar no siempre son para el mismo segmento.

**Tabla Nº 7 - Dimensión Usuario.**

DIM_USUARIO			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_USUARIO	INT IDENTITY (1,1)	ID Usuario	1,2,..
COD_USUARIO	CHAR(4)	Código Usuario	1,2,..
ROLE	VARCHAR(50)	Role del usuario	Private, Corp...

Fuente: Elaboración propia

- **Dimensión Categoría:** Nos permitirá segmentar los indicadores por la categoría del vehículo pues existen algunas categorías que se lanzan como medio promocional.

**Tabla Nº 8 - Dimensión Categoría.**

DIM_CATEGORIA			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_CATEGORIA	INT IDENTITY (1,1)	ID Categoría	1,2,..
COD_CATEGORIA	CHAR(4)	Código Categoría	1,2,..
CATEGORIA	VARCHAR(50)	Nombre de la Categoría	Lite, Executive...

Fuente: Elaboración propia

- **Dimensión Tipo Moneda:** Nos permitirá unificar todos los montos de moneda local poder convertirlo a una moneda universal (dólar).

**Tabla Nº 9 - Dimensión Tipo Moneda.**

DIM_TIPO_MONEDA			
CAMPO	TIPO DE DATO	DESCRIPCION	VALORES
ID_TIPO_MONEDA	INT IDENTITY (1,1)	ID Trayecto	1,2,..
COD_MONEDA	INT	Código de moneda	1,2,..
MONEDA	CHAR(3)	Tipo de Moneda	PEN, EUR, ...

Fuente: Elaboración propia

### 3.3.4. IDENTIFICAR LAS TABLAS DE HECHOS Y MEDIDAS

Tabla N° 10 - Hechos Trayecto

<b>HECHOS_TRAYECTO</b>			
<b>CAMPO</b>	<b>TIPO DE DATO</b>	<b>DESCRIPCION</b>	<b>VALORES</b>
<b>ID_HECHOS_TRAYECTO</b>	INT	ID Categoria	1,2,..
<b>ID_TIEMPO</b>	INT	ID Tiempo	1,2,..
<b>ID_UBIGEO</b>	INT	ID Ubigeo	1,2,..
<b>ID_METODO_PAGO</b>	INT	ID Metodo pago	1,2,..
<b>ID_USUARIO</b>	INT	ID Usuario	1,2,..
<b>ID_CATEGORIA</b>	INT	ID Categoria	1,2,..
<b>ID_DESCUENTO</b>	INT	ID Descuento	1,2,..
<b>ID_HORA</b>	INT	ID Hora	1,2,..
<b>ID_TIPO_MONEDA</b>	INT	ID Tipo Moneda	1,2,..
<b>FLAG_PURO</b>	INT	s/c descuento	1 0
<b>PRECIO</b>	DECIMAL(10,2)	Precio	\$x,xxx
<b>DESCUENTO</b>	DECIMAL(10,2)	Descuento	\$x,xxx
<b>TIPO_CAMBIO</b>	DECIMAL(10,2)	Tipo de cambio	\$x,xxx
<b>DURACION_SEG</b>	INT	Duracion segundos	ss
<b>RECORRIDO_MET</b>	INT	Recorrido metros	1500,...

Fuente: Elaboración propia

En esta imagen podemos observar la tabla hechos Trayecto la cual servirá como base del datamart y creación del cubo.

### 3.3.5. MODELO GRÁFICO DE ALTO NIVEL

Figura Nº 5 - Modelo gráfico de alto nivel.

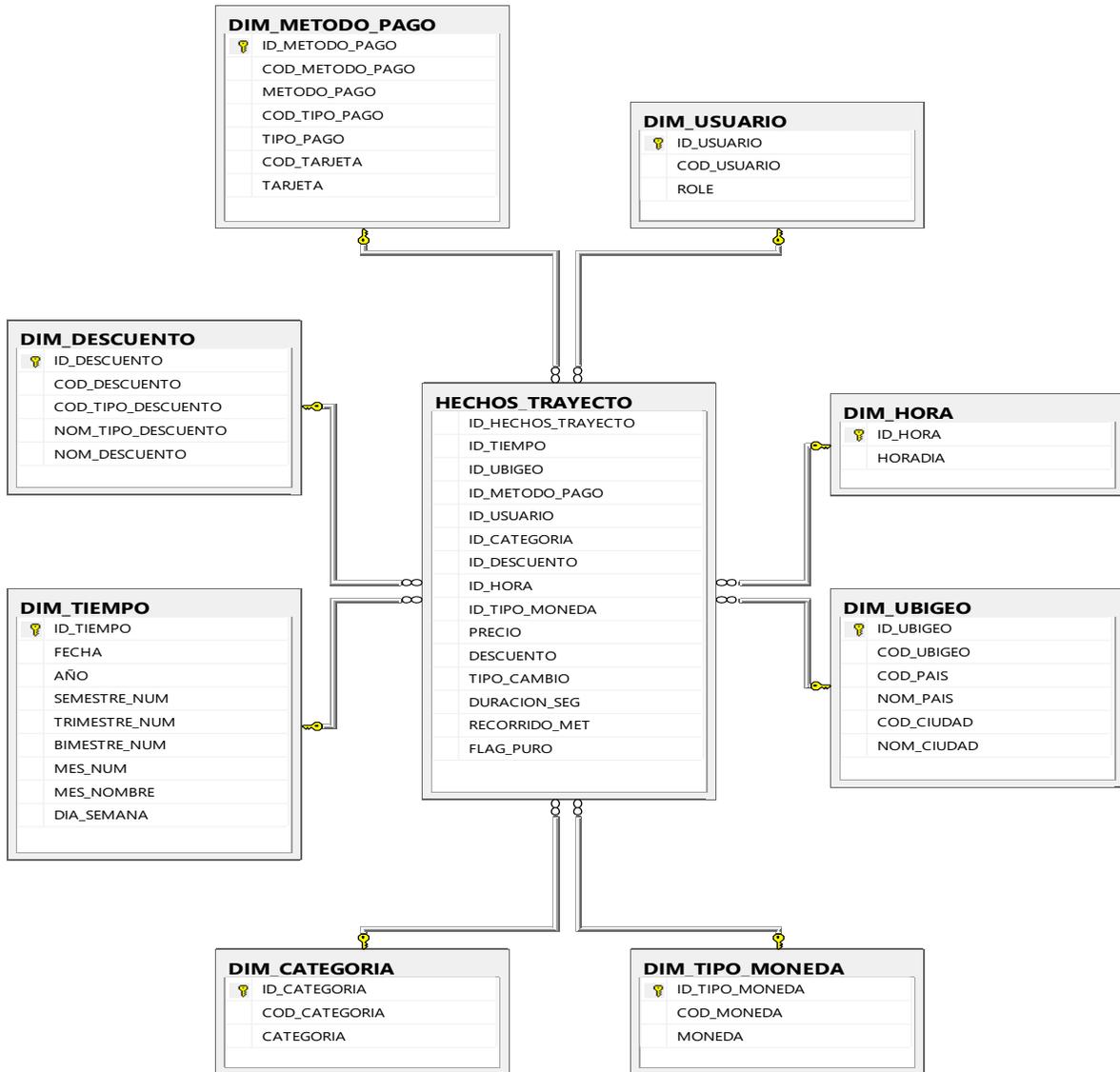


Fuente: Elaboración propia

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),

### 3.4. DISEÑO FISICO

Figura N° 6 - Diseño físico Datamart



Fuente: Elaboración propia

### 3.5. DISEÑO E IMPLMENTACIÓN DEL SUBSISTEMA ETL

Figura N° 7 - Diseño ETL

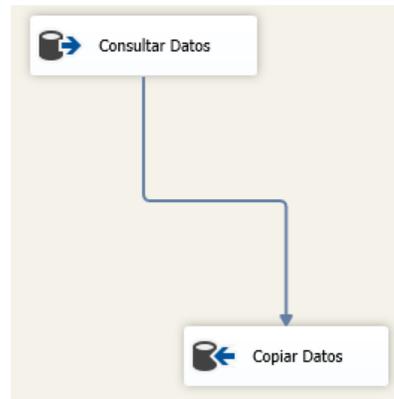


Fuente: Elaboración propia

Como se observa el diseño del ETL, cada Dimensión ubicada en la base de datos DM\_MKT se alimentará de la base de datos producción CABIFY\_DATA a través de consultas especificadas en los siguientes Data Flow.

### 3.5.1. FLUJO DE DATOS DIM\_USUARIO

Figura N° 8 - Flujo de datos USUARIO

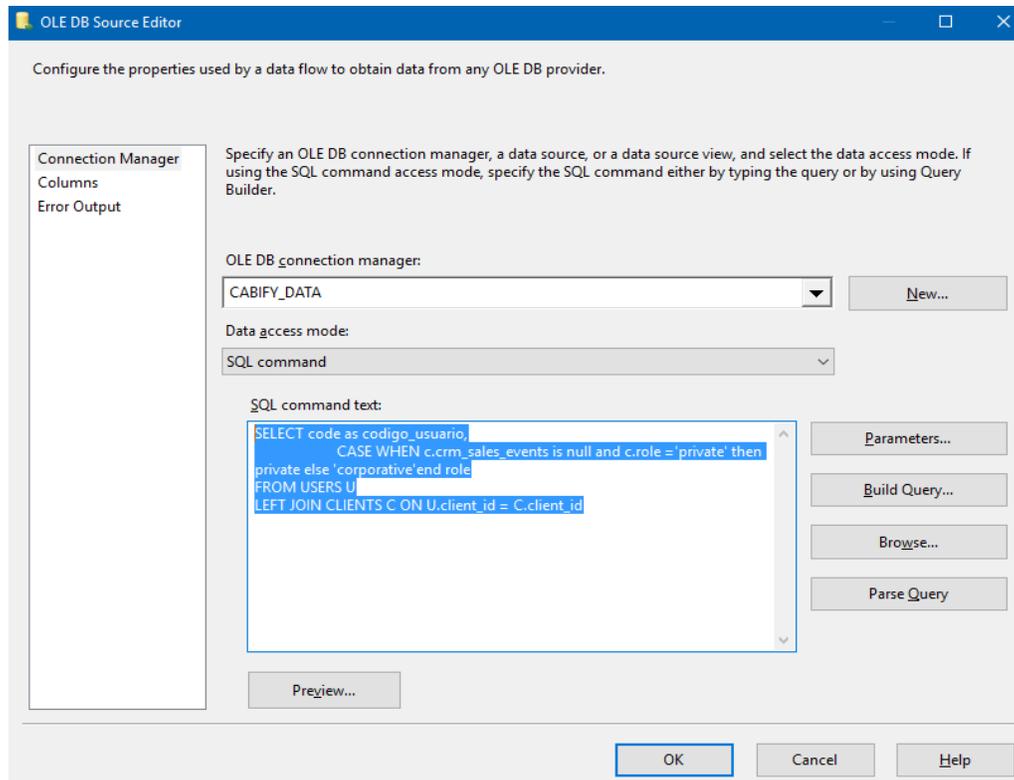


Fuente: Elaboración propia

#### Consulta:

```
SELECT code as codigo_usuario,  
       CASE WHEN c.crm_sales_events is null and c.role ='private' then  
         private else 'corporative'end role  
FROM USERS U  
LEFT JOIN CLIENTS C ON U.client_id = C.client_id
```

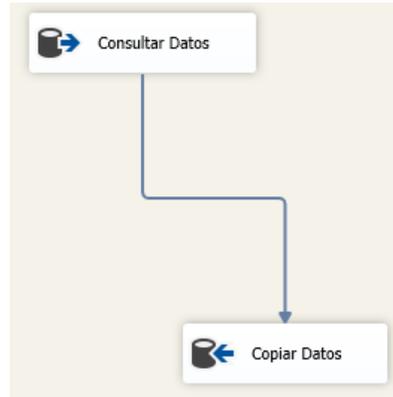
Figura N° 9 - OLDB Source USUARIO



Fuente: Elaboración propia

### 3.5.2. FLUJO DE DATOS CATEGORIA

Figura N° 10 - Flujo de datos CATEGORIA

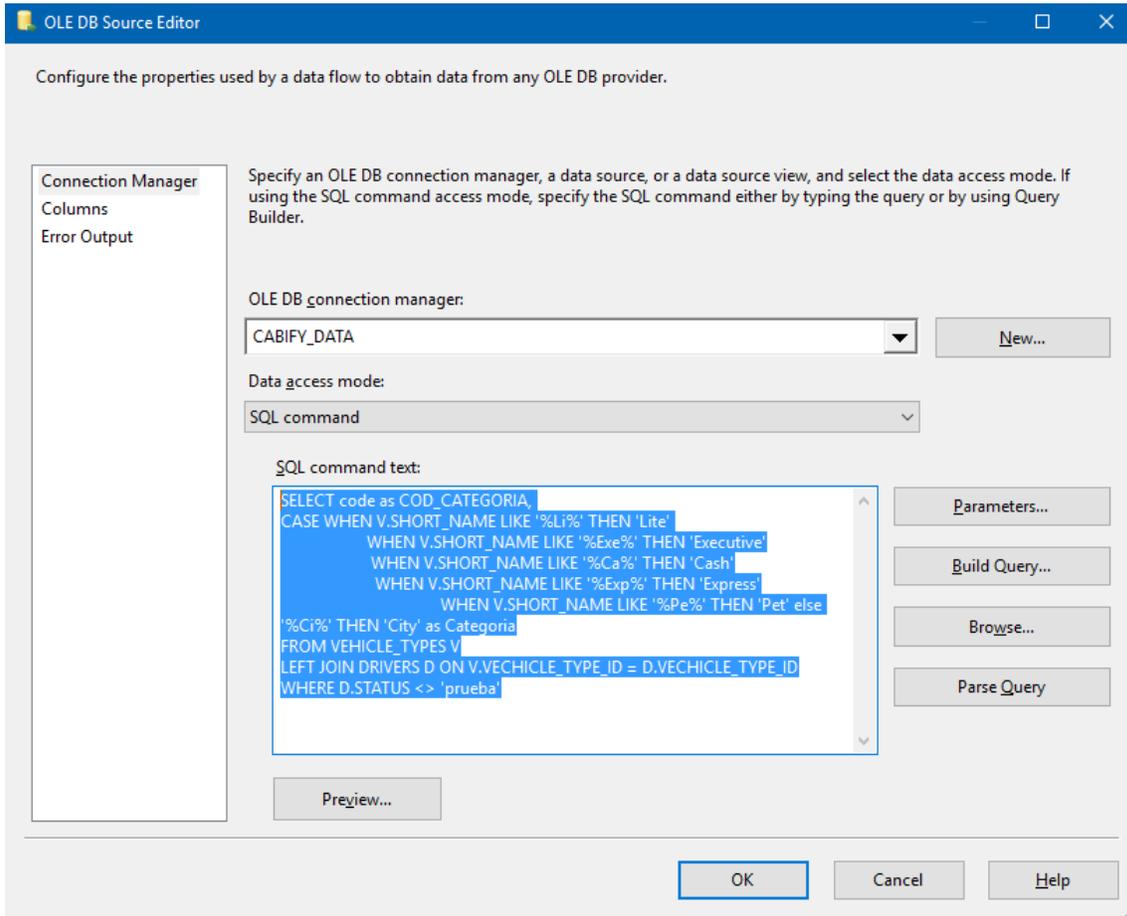


Fuente: Elaboración propia

#### Consulta:

```
SELECT code as COD_CATEGORIA,  
CASE WHEN V.SHORT_NAME LIKE '%Li%' THEN  
'Lite' WHEN V.SHORT_NAME LIKE '%Exe%' THEN  
'Executive' WHEN V.SHORT_NAME LIKE '%Ca%' THEN  
'Cash' WHEN V.SHORT_NAME LIKE '%Exp%' THEN  
'Express' WHEN V.SHORT_NAME LIKE '%Pe%' THEN  
'Pet' else '%Ci%' THEN 'City' as Categoria FROM  
VEHICLE_TYPES V LEFT JOIN DRIVERS D ON  
V.VECHICLE_TYPE_ID = D.VECHICLE_TYPE_ID  
WHERE D.STATUS <> 'prueba'
```

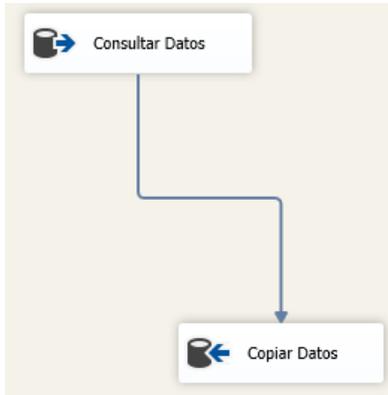
Figura Nº 11 - OLDB Source CATEGORIA



Fuente: Elaboración propia

### 3.5.3. FLUJO DE DATOS UBIGEO

Figura Nº 12 - Flujo de datos UBIGEO

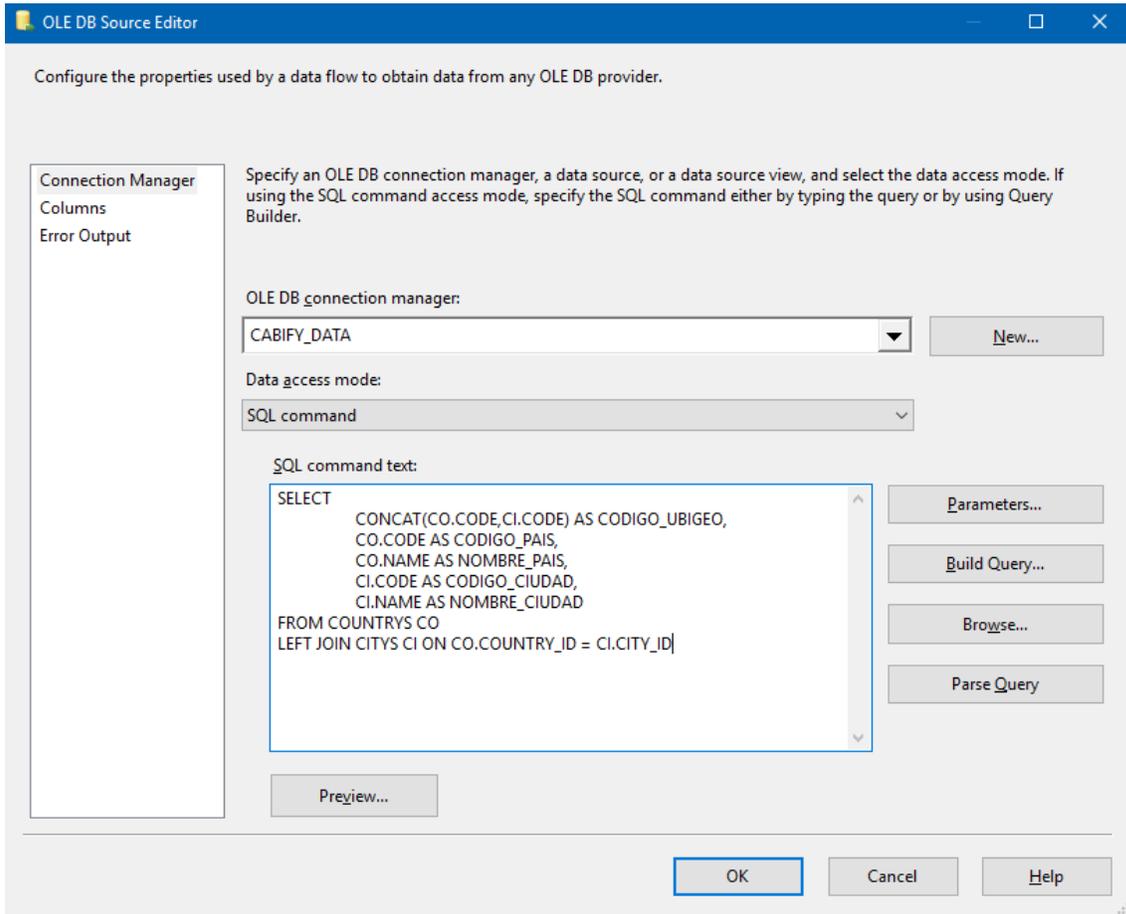


Fuente: Elaboración propia

#### Consulta:

```
SELECT
  CONCAT(CO.CODE,CI.CODE) AS CODIGO_UBIGEO,
  CO.CODE AS CODIGO_PAIS, CO.NAME AS
  NOMBRE_PAIS, CI.CODE AS CODIGO_CIUADAD,
  CI.NAME AS NOMBRE_CIUADAD FROM COUNTRYS
CO LEFT JOIN CITYS CI ON CO.COUNTRY_ID =
CI.CITY_ID
```

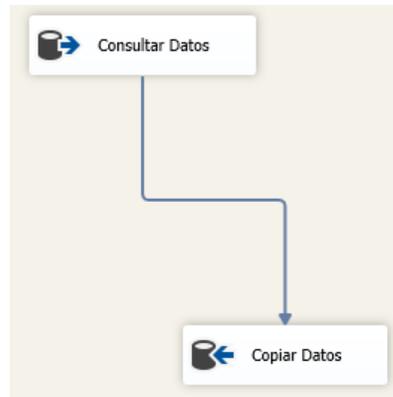
Figura Nº 13 - OLDB Source UBIGEO



Fuente: Elaboración propia

### 3.5.4. FLUJO DE DATOS DIM\_METODO\_PAGO

Figura N° 14 - Flujo de datos METODO PAGO

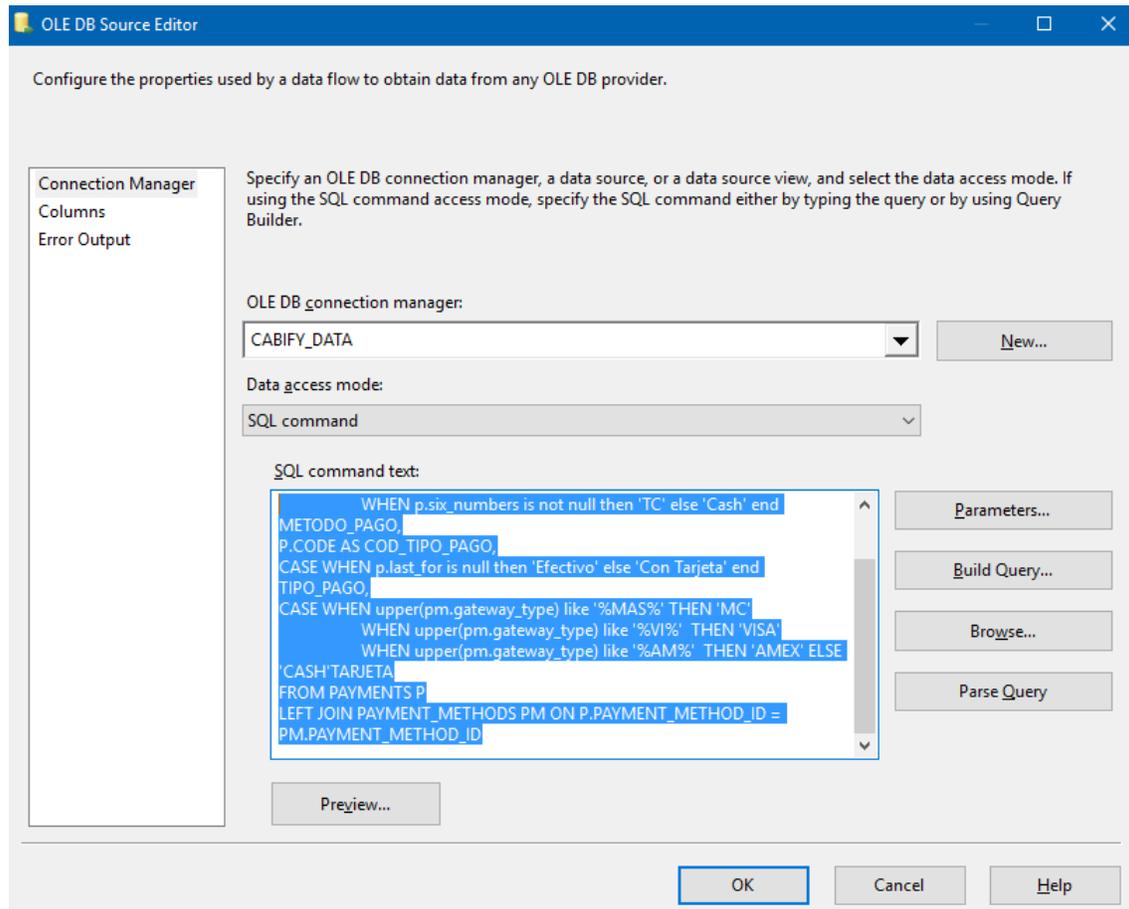


Fuente: Elaboración propia

#### Consulta:

```
SELECT      CONCAT(P.CODE,PM.CODE)      AS
CODIGO_UBIGEO AS COD_METODO_PAGO,
CASE WHEN gateway_type IS NULL 'Paypal'
      WHEN p.six_numbers is not null then 'TC' else
'Cash' end METODO_PAGO,
P.CODE AS COD_TIPO_PAGO,
CASE WHEN p.last_for is null then 'Efectivo' else 'Con
Tarjeta' end TIPO_PAGO,
CASE WHEN upper(pm.gateway_type) like '%MAS%'
THEN 'MC'  WHEN upper(pm.gateway_type) like
'%VI%' THEN 'VISA'  WHEN
upper(pm.gateway_type) like '%AM%' THEN 'AMEX'
ELSE 'CASH'TARJETA FROM PAYMENTS P LEFT
JOIN      PAYMENT_METHODS      PM      ON
P.PAYMENT_METHOD_ID= M.PAYMENT_METHOD_ID
```

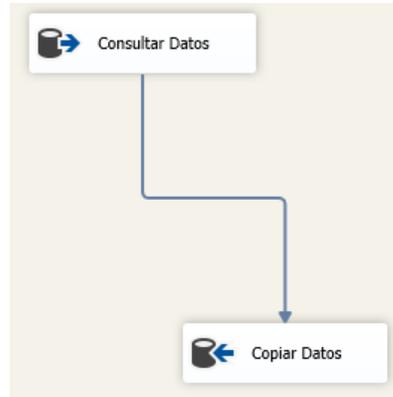
Figura Nº 15 - OLDB Source METODO PAGO



Fuente: Elaboración propia

### 3.5.5. FLUJO DE DATOS DESCUENTO

Figura N° 16 - Flujo de datos DESCUENTO

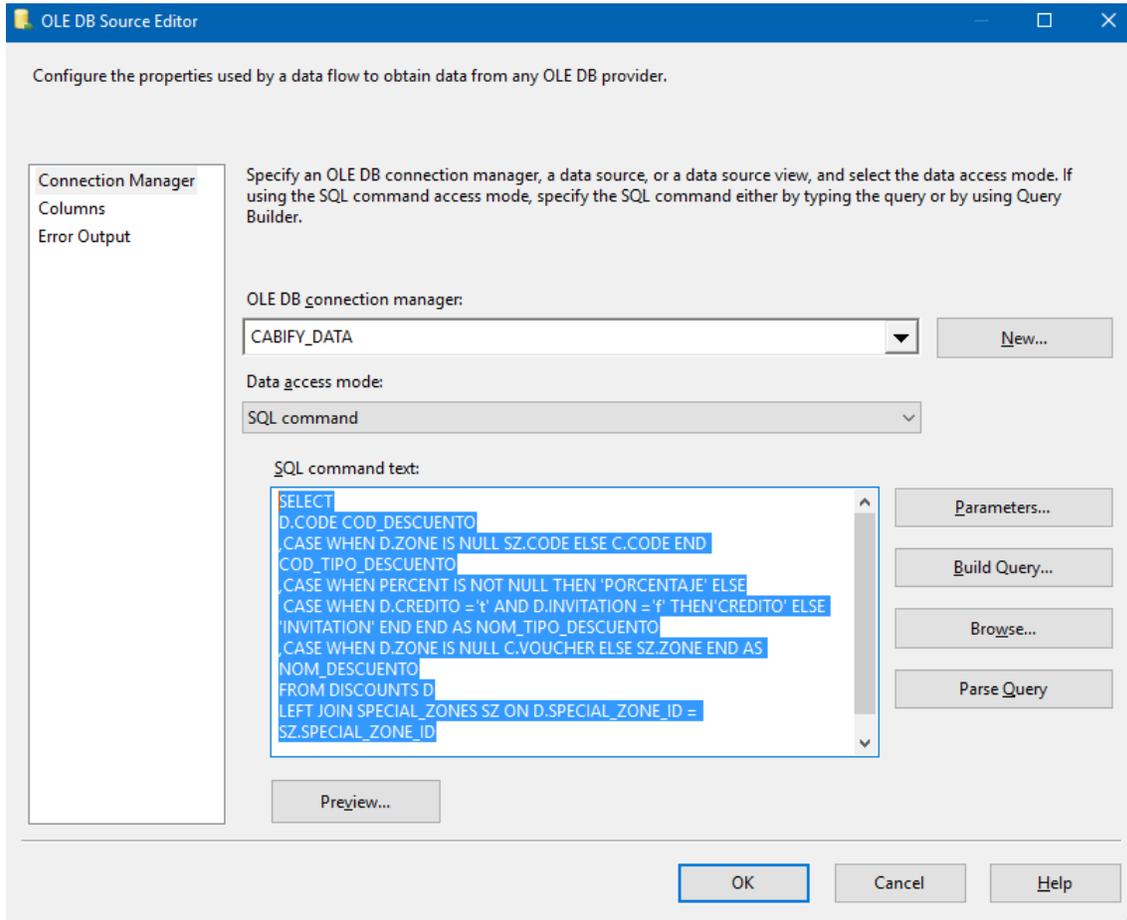


Fuente: Elaboración propia

#### Consulta:

```
SELECT
D.CODE COD_DESCUENTO
,CASE WHEN D.ZONE IS NULL SZ.CODE ELSE
C.CODE END COD_TIPO_DESCUENTO
,CASE WHEN PERCENT IS NOT NULL THEN
'PORCENTAJE' ELSE
CASE WHEN D.CREDITO ='t' AND D.INVITATION ='f'
THEN'CREDITO' ELSE 'INVITATION' END END AS
NOM_TIPO_DESCUENTO
,CASE WHEN D.ZONE IS NULL C.VOUCHER ELSE
SZ.ZONE END AS NOM_DESCUENTO
FROM DISCOUNTS DLEFT JOIN SPECIAL_ZONES
SZ ON D.SPECIAL_ZONE_ID =
SZ.SPECIAL_ZONE_IDLEFT JOIN CODES C ON
C.CODE = D.CODE
```

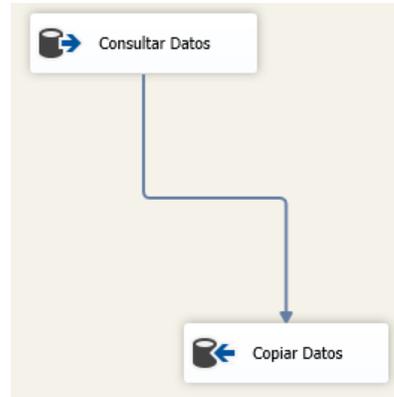
Figura Nº 17 - OLDB Source DESCUENTO



Fuente: Elaboración propia

### 3.5.6. FLUJO DE DATOS TIPO\_MONEDA

Figura N° 18 - Flujo de datos TIPO MONEDA

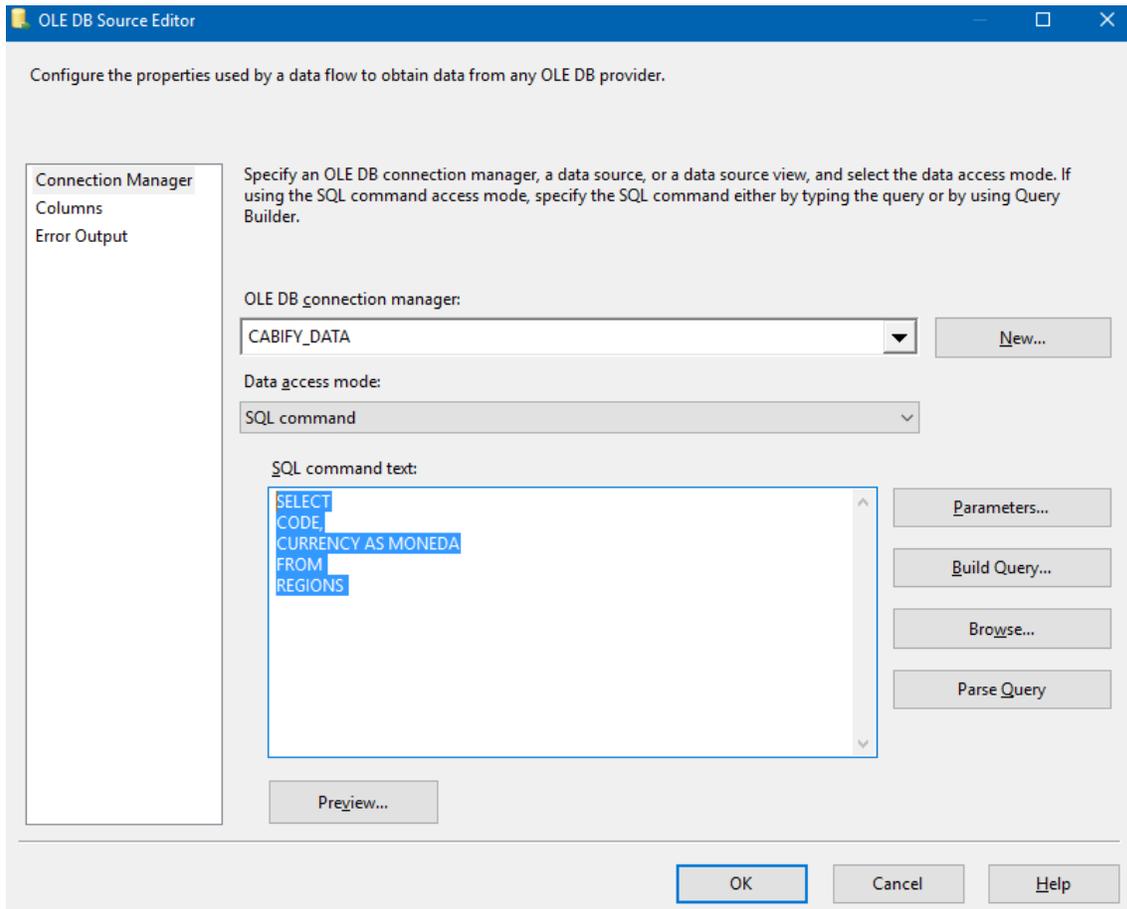


Fuente: Elaboración propia

#### Consulta:

```
SELECT CODE, CURRENCY AS MONEDA FROM  
REGIONS
```

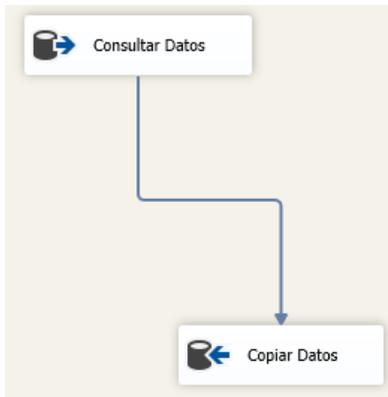
**Figura Nº 19 - OLDB Source MONEDA**



Fuente: Elaboración propia

### 3.5.7. FLUJO DE DATOS TIEMPO

Figura Nº 20 - Flujo de datos TIEMPO

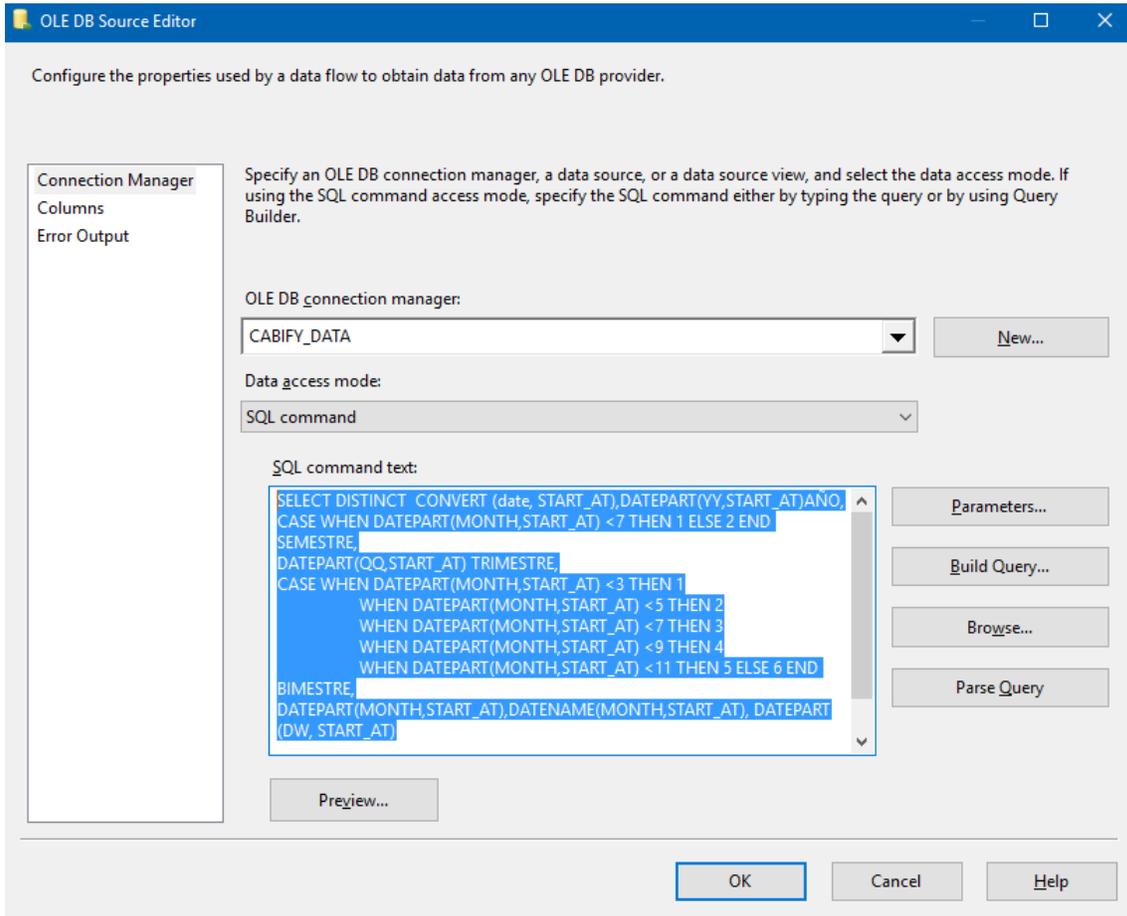


Fuente: Elaboración propia

#### Consulta:

```
SELECT DISTINCT CONVERT (date, START_AT),
DATEPART(YEAR,START_AT)AÑO,
CASE WHEN DATEPART(MONTH,START_AT) <7
THEN 1 ELSE 2 END SEMESTRE,
DATEPART(QUARTER,START_AT) TRIMESTRE,
CASE WHEN DATEPART(MONTH,START_AT) <3
THEN 1
      WHEN DATEPART(MONTH,START_AT) <5
THEN 2
      WHEN DATEPART(MONTH,START_AT) <7
THEN 3
      WHEN DATEPART(MONTH,START_AT) <9
THEN 4
      WHEN DATEPART(MONTH,START_AT) <11
THEN 5 ELSE 6 END BIMESTRE,
DATEPART(MONTH,START_AT),DATENAME(MONTH
,START_AT), DATEPART(DW, START_AT)
FROM JOURNEYS BY START_AT
```

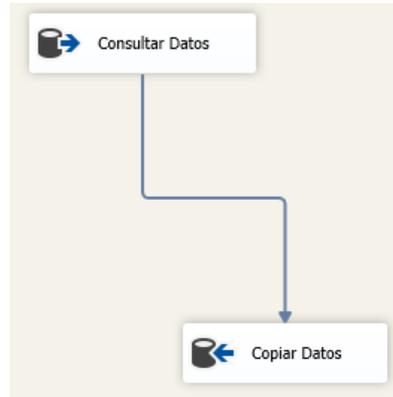
Figura Nº 21 - OLDB Source TIEMPO



Fuente: Elaboración propia

### 3.5.8. FLUJO DE DATOS DIM\_HORA

Figura N° 22 - Flujo de datos HORA

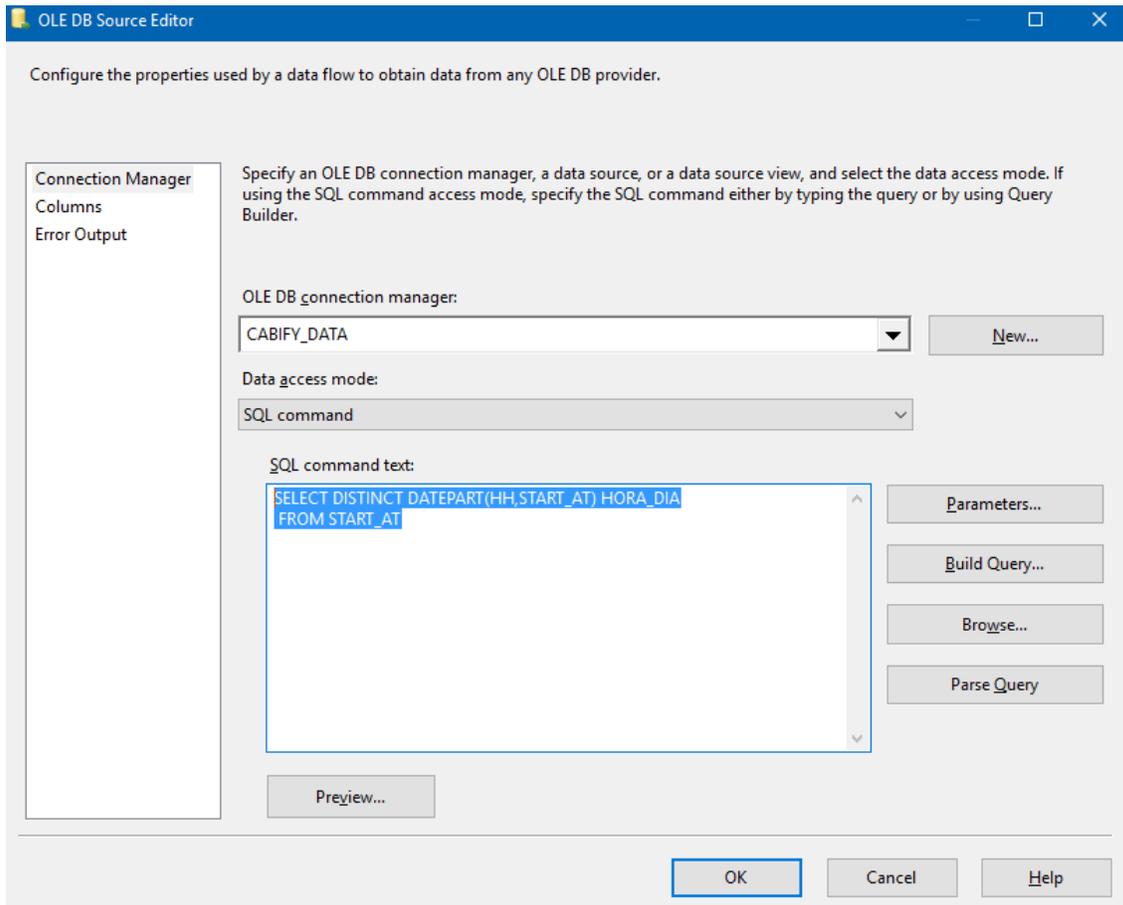


Fuente: Elaboración propia

#### Consulta:

```
SELECT DATEPART(HH,START_AT) HORA_DIA  
FROM START_AT
```

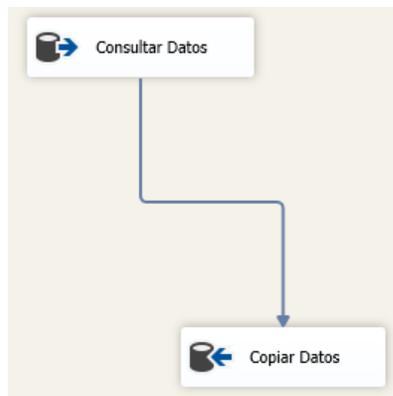
Figura Nº 23 - OLDB Source HORA



Fuente: Elaboración propia

### 3.5.9. FLUJO DE DATOS HECHOS\_TRAYECTO

Figura N° 24 - Flujo datos HE\_TRAYECTO

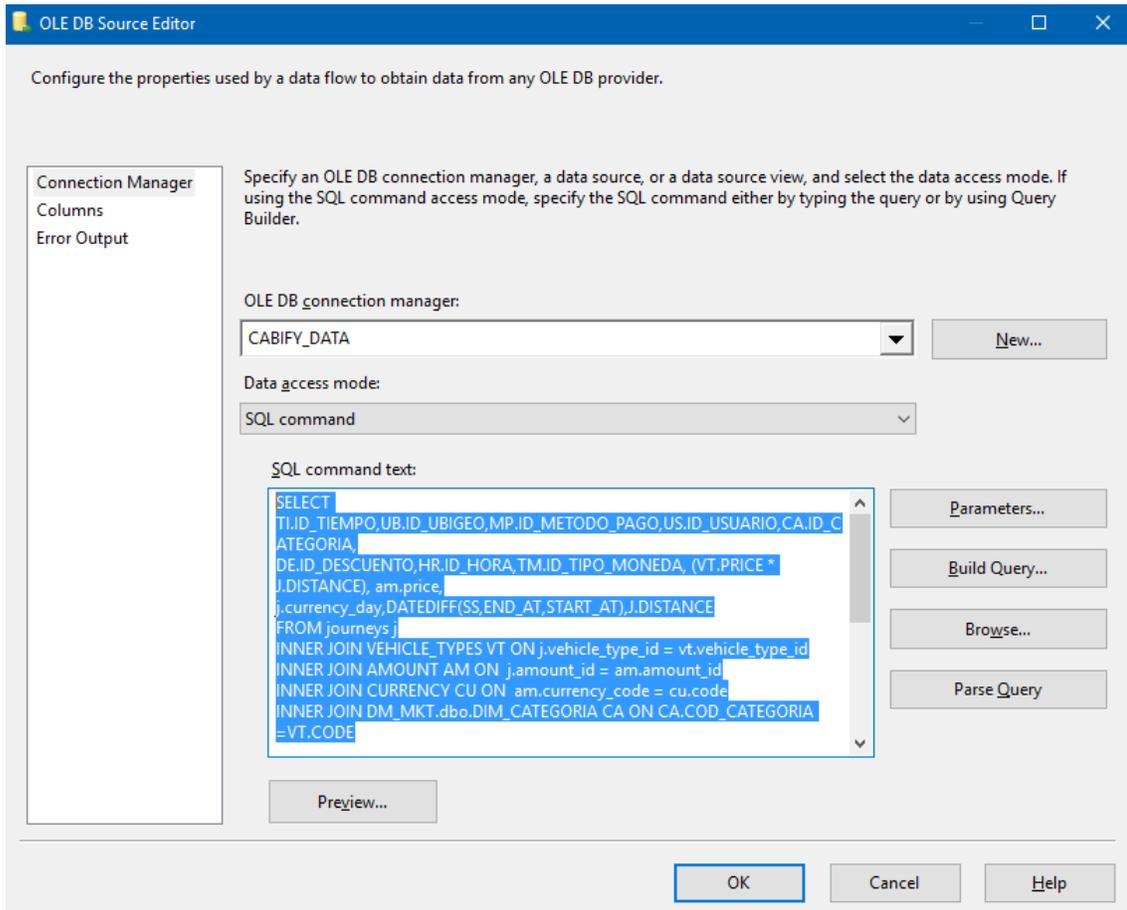


Fuente: Elaboración propia

#### Consulta:

```
SELECT
TI.ID_TIEMPO,UB.ID_UBIGEO,MP.ID_METODO_PAGO
,US.ID_USUARIO,CA.ID_CATEGORIA,
  DE.ID_DESCUENTO,HR.ID_HORA,TM.ID_TIPO_MON
EDA, (VT.PRICE * J.DISTANCE), am.price,
  j.currency_day,DATEDIFF(SS,END_AT,START_AT),J.
DISTANCE FROM journeys j INNER JOIN
VEHICLE_TYPES VT ON j.vehicle_type_id =
vt.vehicle_type_id INNER JOIN AMOUNT AM ON
j.amount_id = am.amount_id INNER JOIN CURRENCY
CU ON am.currency_code = cu.code INNER JOIN
DM_MKT.dbo.DIM_CATEGORIA CA ON
CA.COD_CATEGORIA =VT.CODE INNER JOIN
DM_MKT.dbo.DIM_DESCUENTO DE ON
DE.COD_DESCUENTO=AM.CODE INNER JOIN
DM_MKT.dbo.DIM_METODO_PAGO MP ON
am.payment_code = mp.COD_METODO_PAGO INNER
JOIN DM_MKT.dbo.DIM_TIEMPO TI ON j.start_at
=ti.FECHA INNER JOIN DM_MKT.dbo.DIM_UBIGEO UB
ON ub.NOM_CIUDAD = j.region INNER JOIN
DM_MKT.dbo.DIM_USUARIO US ON j.user_id =
us.COD_USUARIO INNER JOIN
DM_MKT.dbo.DIM_HORA HR ON
DATEPART(hh,start_at) = hr.horadia
```

Figura N° 25 - OLDB Source HE\_TRAYECTO



Fuente: Elaboración propia

### **3.6. SELECCIÓN DE PRODUCTOS**

Para el proceso ETL y la creación de cubo se estará usando la herramienta SQL Server Integration Services (SSIS) esto al gran soporte que se encuentra para esta herramienta, además de ser muy intuitivo y muy abierto a modificaciones futuras.

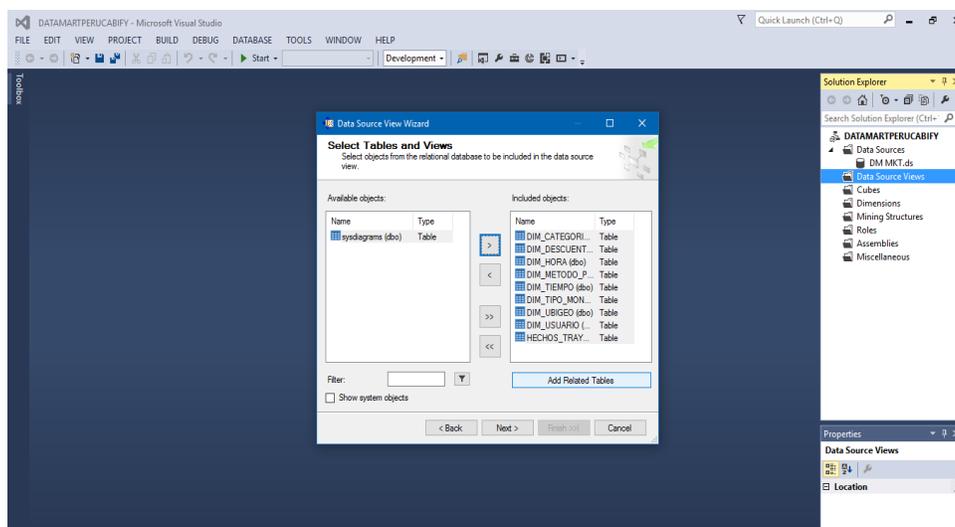
Para la visualización de los informes se está desarrollando POWER BI desktop, por su fácil uso y además de poder generar libros de trabajos los cuales servirán en un futuro a la empresa cuando empiece a desarrollar Datawarehouse.

### 3.7. DESARROLLO DE LA APLICACIÓN BI

#### 3.7.1. DESARROLLO DEL CUBO

Para comenzar el desarrollo del cubo, primero debemos tener la conexión de Analysis Services con la base de datos, luego tal como vemos en la Figura26 seleccionamos las dimensiones y la tabla de hechos para poder analizarlas.

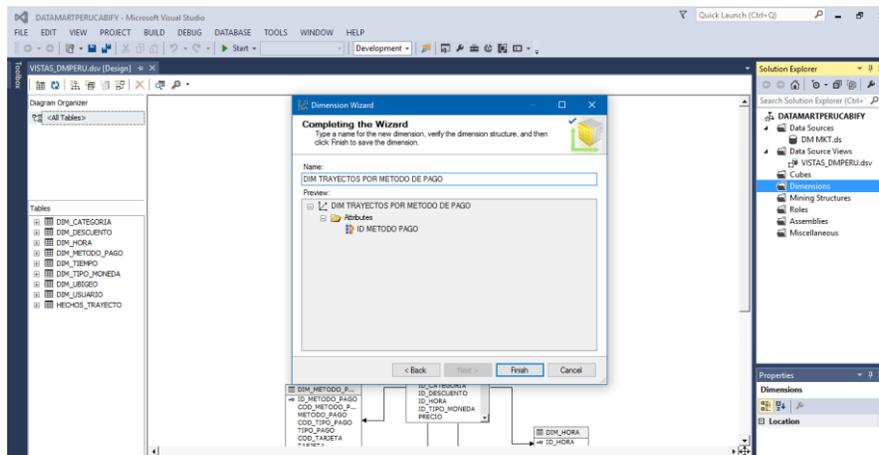
Figura Nº 26 - Conexión vista



Fuente: Elaboración propia

Luego vamos creando dimensiones dependiendo del cubo a diseñar, en esto caso se crearan 2 cubos (Códigos y Trayectos) y dependiendo se elegirán las dimensiones correspondientes para cada cubo.

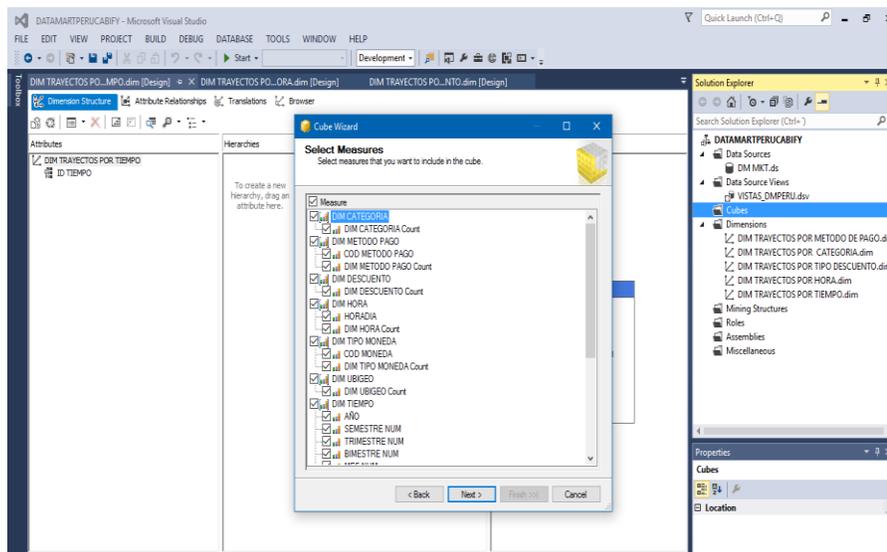
Figura Nº 27 - Agregando Dimensiones



Fuente: Elaboración propia

Finalmente ya cuando tenemos las dimensiones seleccionadas podemos crear el cubo escogiendo las métricas y los campos que deseamos según el análisis a realizar con el cubo para finalmente dar clic derecho y seleccionar process para que el cubo se realice satisfactoriamente y esté listo para usarse en la herramienta de BI.

**Figura N° 28 - Generación del cubo**



Fuente: Elaboración propia

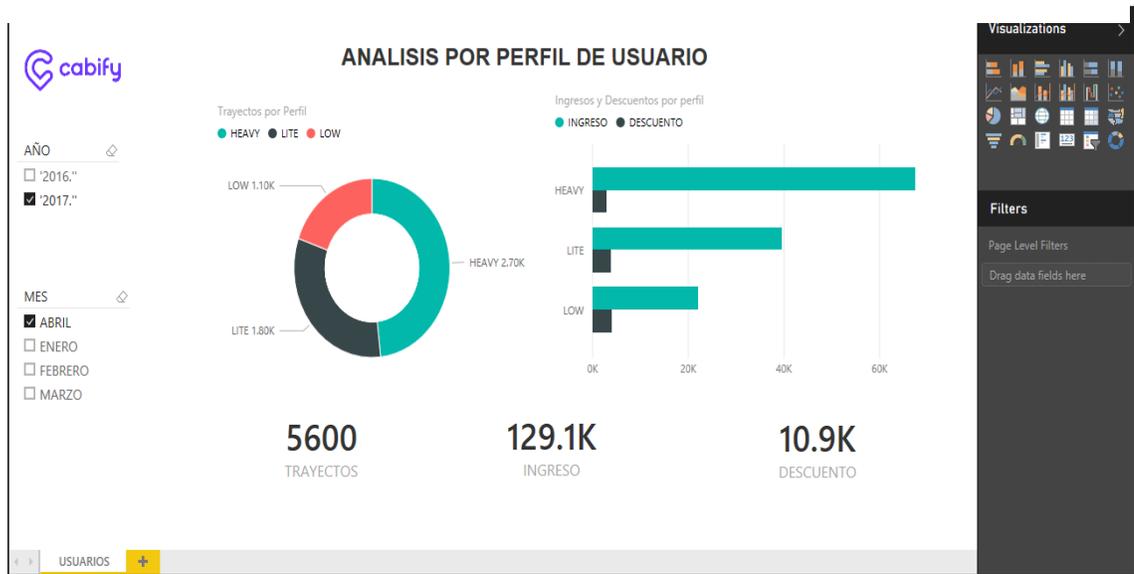
### 3.7.2. DESARROLLO POWER BI

Para la parte de visualización de los indicadores se empleó la herramienta de Inteligencia de Negocios POWER BI realizando la conexión con los cubos creados en el punto anterior.

Buscando satisfacer las necesidades de información del equipo de marketing se crearon 4 reportes iniciales.

-Reporte Análisis por perfil de usuario, acá se puede verificar la cantidad de trayectos realizados por el perfil de usuario, los ingresos y descuentos generados por estos en un rango de tiempo establecido lo cual permite al usuario revisar el comportamiento de los clientes según el perfil e identificar quienes generan mayor rentabilidad (menos uso de descuento).

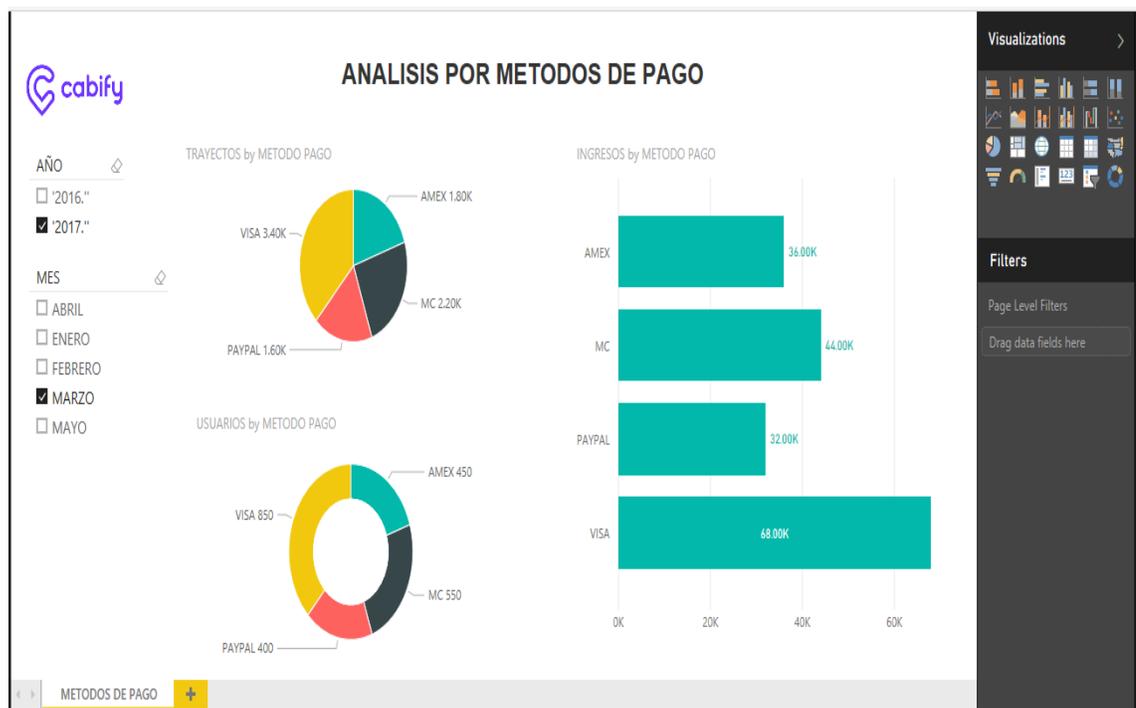
Figura N° 29 - Reporte Trayectos por Perfil Usuario



Fuente: Elaboración propia

-Reporte Análisis por método de pago, se observa la cantidad de trayectos, usuarios que realizaron y pagaron trayectos por distintos métodos de pago, gracias a esta información el usuario puede realizar distintas promociones por método de pago previniendo el retorno de estos usuarios basados en el uso de la tarjeta.

**Figura Nº 30 - Reporte Metodos de Pago**



Fuente: Elaboración propia

-Reporte Análisis de códigos, en esta tabla se observa los códigos las veces que activaron los trayectos ingresos y el CAC (costo adquisición por usuario) que sirve como indicador de rentabilidad del código pues mientras este sea menor se tiene un mejor retorno. Como se observa en este reporte con corte en el mes de Mayo se puede observar que sólo un código tiene un CAC que no genera retorno (mayor a 14).

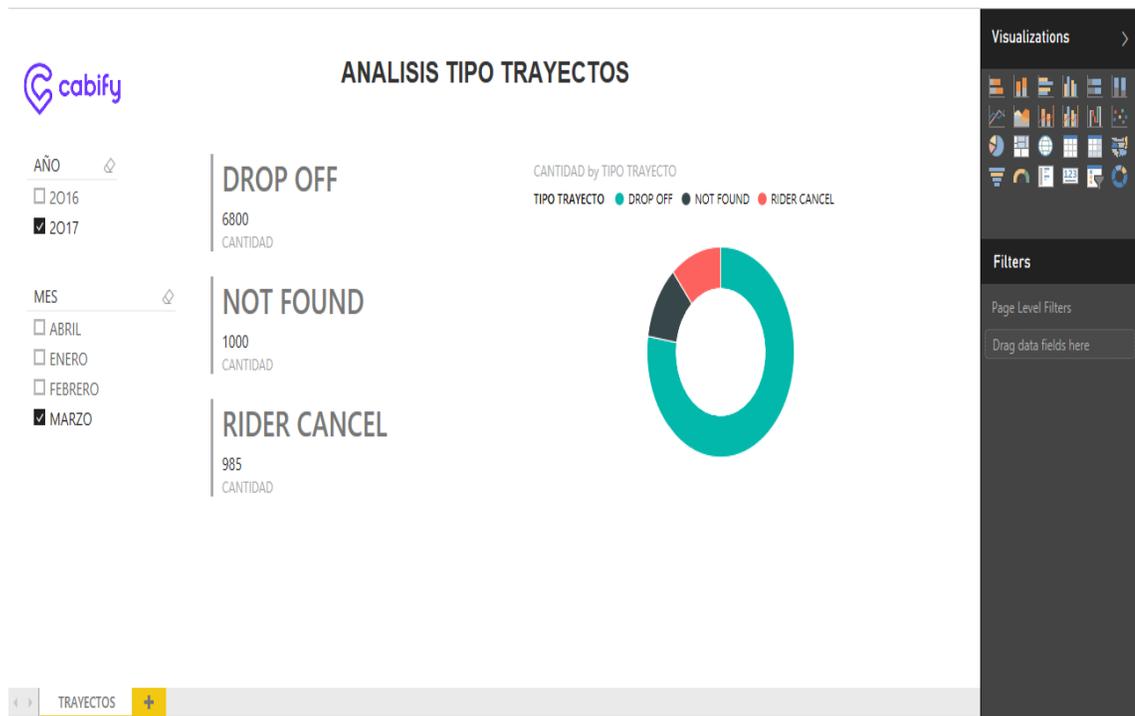
**Figura Nº 31 - Reporte Códigos**

CODIGO	CLAIMS	USU. NUEVOS	UN DO	DO	INGRESO	DESCUENTO	20% CABIFY	CAC
5ENVIOS	36	2	0	1	20	4	4	8.00
ACTIVAPE	240	48	40	81	1620	324	390	3.30
BREAKCABIFY	74	8	0	1	20	2	4	10.00
CABIFYPET	70	14	6	13	260	13	52	13.00
CABIVALENTIN	190	32	24	49	980	196	196	0.00
CLINICADELGADO	150	23	15	31	620	93	124	4.13
CONOCENOS2017	79	6	1	3	60	6	12	12.00
ESMIMAMITA	164	17	9	19	380	57	76	4.22
EXPODECO	4	0	0	1	20	4	4	15.00
EXPRESSPE	6	0	0	1	20	3	4	12.00
FRIDAYS	132	13	5	11	220	33	44	4.40
LIBERAPE	80	15	7	15	300	60	60	0.00
LIMA2017	450	62	54	109	2180	109	436	12.11
MAMA2017	362	54	46	93	1860	372	372	0.00
PORMASLIM	61	12	4	9	180	27	36	4.50
REACTIVAPE	120	24	16	33	660	132	132	0.00
ROSATEL	89	13	5	11	220	11	44	13.20
TRABAJADORPE	58	9	1	3	60	6	12	12.00
<b>Total</b>	<b>2933</b>	<b>442</b>	<b>296</b>	<b>614</b>	<b>12280</b>	<b>1727</b>	<b>2522</b>	<b>162.23</b>

Fuente: Elaboración propia

- En este reporte llamado análisis tipo trayectos permite determinar cuántos trayectos satisfactorios se realizaron (drop off), cuantos trayectos no se concretaron por que no se encontró conductor a asignar (not found) y cuantos fueron cancelados por los usuarios (rider cancel). Esta información permitirá al área de Marketing determinar si necesita generar una campaña.

**Figura N° 32 - Reporte Tipo de Trayecto**



Fuente: Elaboración propia

## CONCLUSIONES

1. Gracias al desarrollo del Datamart se mejoró la toma de decisiones con respecto a las inversiones, esto lo vemos reflejado pues se redujo a un 9% la cantidad de códigos sin retorno de inversión, esto debido a los indicadores que permitieron segmentar los usuarios permitiendo enviar códigos personalizados dependiendo del perfil del usuario lo cual podemos observarlo en el reporte Análisis de Códigos (Figura 31).
2. Se analizó y desarrollo una base de datos multidimensional la cual permite analizar y explotar los datos de una manera más eficaz haciendo más óptimo el tiempo de respuesta de las consultas.
3. Se cumplió en un 100% los requerimientos de indicadores analizados en conjunto con el área de marketing para poder realizar un análisis más profundo con la información brindada según lo coordinado en la entrevista con un representante del área de Marketing tal como se precisa en el punto 3.2 de este proyecto.

4. El uso de la herramienta Power BI para la visualización de los reportes ha permitido un manejo muy simple e intuitivo para el usuario final, reduciendo el tiempo de capacitación y ahorrando los costos de este.

## RECOMENDACIONES

1. Utilizar como base el Datamart desarrollado en este proyecto para poder replicar en los demás países y en un futuro poder desarrollar un Datawarehouse, pues al ser desarrollado bajo la metodología de Ralph Kimball permite estar abierto a nuevas incorporaciones de futuros Datamarts ampliando a las áreas implicadas de la empresa.
2. Generar alertas en la herramienta de BI el cual envíe correo al encargado de Marketing en caso de que los indicadores no estén llegando a las proyecciones pactadas.

## BIBLIOGRAFIA

- Alfaro Mendoza, L. A., & Paucar Moreyra, D. A. (2016). Construcción de un Datamart que apoye en la toma de decisiones de la gestión de incidencias en una mesa de ayuda: caso Consorcio Peruano de Empresas. Lima, Perú.
- Espinoza Vargas, J. M., & Palomino Ruiz, C. A. (2016). Desarrollo de un datamart para optimizar la generación de información estratégica de apoyo a la toma de decisiones en la Vicepresidencia de Banca Comercial de Interbank. Lima, Perú.
- Gamarra Ramirez, A. J. (2011). Solución integral para explotar eficientemente la información de los contactos con los clientes utilizando Datamart en Telefónica del Perú. Lima, Perú.
- Guillén Rodríguez, F. S. (2012). Desarrollo de un datamart para mejorar la toma de decisiones en el área de tesorería de la Municipalidad Provincial de Cajamarca. Cajamarca.
- Buyto (2009), Datamart. Obtenido de <http://www.buyto.es/general-business-intelligence/almacenamiento-de-datos-datawarehouse-datamart-en-business-intelligence.com>

- INGENIMA (2016). Obtenido de: <http://www.evaluandosoftware.com/etl-extraccion-transformacion-carga-datos/>
- HUAMANTUMBA, R. (2007). Obtenido de <http://www.raynerhd.com/wp-content/uploads/rayner-datamart.com>
- Rivera R, Gustavo (2014). La metodología de Kimball para el diseño de almacenes de datos. Argentina
- Ordoñez Cabrera, L. M. (2007). DataMart de estadísticas médicas. México.
- Rosales Sedano, C. P. (2009). Análisis, diseño e implementación de un datamart para el soporte de toma de decisiones y evaluación de las estrategias sanitarias en las direcciones de salud. Lima, Perú.
- Zambrano Alarcón, J. A. (2011). Análisis, diseño e implementación de un datamart para el área de mantenimiento y logística de una empresa de transporte público de pasajeros. Lima, Perú.