

**UNIVERSIDAD GABRIELA MISTRAL
FACULTAD DE INGENIERIA**

**IMPLEMENTACION DE UN DATA MART DE
TARJETAS DE CREDITO EN UNA ENTIDAD
BANCARIA**

Memoria para optar al título de Ingeniero de Ejecución en Informática

Autor: Andrés Alé Araneda

Santiago – Chile
Diciembre, 2014

**UNIVERSIDAD GABRIELA MISTRAL
FACULTAD DE INGENIERIA**

**IMPLEMENTACION DE UN DATA MART DE
TARJETAS DE CREDITO EN UNA ENTIDAD
BANCARIA**

Memoria para optar al título de Ingeniero de Ejecución en Informática.

Autor : Andrés Mohamed Alé Araneda
Profesor Guía : Marcelo Ramírez Moraga
Profesor Integrante : Jorge Tapia Castillo

Santiago – Chile
Diciembre, 2014

INDICE

1	INTRODUCCION	6
1.1	Descripción del problema	7
1.2	Hipótesis.....	8
1.3	Objetivo General.....	8
1.4	Objetivos Específicos	8
1.5	Alcances	9
2	MARCO TEORICO	10
2.1	Sistemas OLTP o transaccionales.....	10
2.2	Ejemplos de Sistemas OLTP o transaccionales	10
2.3	Bases de datos multidimensionales: OLAP vs OLTP	11
2.4	Propiedades de las bases de datos OLAP	14
2.5	Data Warehouse.....	16
2.6	Principales aportes de un Data Warehouse	20
2.7	Data Mart.....	20
2.8	Modelo Relacional	33
2.9	Modelo Dimensional	35
3	CONTEXTO	43
3.1	La entidad bancaria	43
3.2	Tarjeta de crédito.....	43
3.2.1	Estado de cuenta Tarjeta de crédito.....	44
3.2.2	Número de cuenta (Tarjeta)	44
3.2.3	Cupos de la Tarjeta de Crédito.....	44
3.2.4	Tasa de Interés Vigente	45
3.2.5	CAE	45
3.2.6	Periodo Facturado	46
3.2.7	Fecha Tope de Pago	46
3.2.8	Promociones y Beneficios Tarjetas de Crédito.....	46
3.2.9	Cargo de Interés por Mora Tarjetas de Crédito	46
3.2.10	Tarjetas de Crédito en cartera vencida.....	47

3.2.11 Cierre de Tarjeta de Crédito Voluntario	47
3.3 Descripción del problema	47
3.4 Solución Propuesta.....	47
3.5 Justificación de la solución	48
4 DESARROLLO.....	50
4.1 Marco Metodológico	50
4.1.1 Planificación del Proyecto	50
4.1.2 Tareas por etapa	52
4.1.3 Flujo de Tareas.....	53
4.1.4 Análisis de requerimientos	54
4.1.5 Modelo Conceptual.....	58
4.1.6 Análisis de los datos existentes.....	60
4.1.7 Calculo de Indicadores	60
4.1.8 Establecer correspondencias	62
4.1.9 Nivel de Granularidad.....	72
4.1.10 Modelo Conceptual Ampliado.....	78
4.1.11 Modelo Lógico del Data mart.....	80
4.1.12 Tipo de Modelo Lógico del Data mart.....	80
4.1.13 Tablas de Dimensiones	81
4.1.14 Tablas de Hechos.....	85
4.1.15 Construcción de Tablas de Hechos.....	88
4.1.16 Uniones	91
4.2 ETL y Carga de Datos	93
4.3 Creación de Cubos Dimensionales.....	102
4.3.1 Creación de Indicadores.....	102
4.3.2 Creación de Atributos	103
4.3.3 Creación de Jerarquías	104
4.3.4 Otros Cubos Multidimensionales del Modelo	107
4.4 Prototipos de Informe	109
4.4.1 Informe de Mora	109

4.4.2	Informe de Ventas y Consumos	110
4.4.3	Informe de Cierres por Sucursal.....	111
5	Conclusión	112
6	Glosario de Terminos	114
7	Indice de Figuras.....	120
8	Bibliografía	122

Dedicatoria

A mi amada esposa Elizabeth, quien con su empuje, apoyo incondicional y paciencia logró impregnar la energía necesaria para conseguir esta importante meta en mi vida.

A Lorenzo, Matías y Vicente por comprender mis días y tardes de ausencia en virtud de conseguir este logro.

1 INTRODUCCION

Poder gestionar la información de las empresas, en la actualidad, es una herramienta clave para competir en un mercado cambiante, dinámico y global. El aprendizaje en base a esta información es fundamental para la toma de decisiones, el crecimiento y la gestión de cualquier empresa. La disciplina denominada como Business Intelligence nos acerca a los sistemas de información que nos ayudan en la toma de decisiones para la organización.

Tras las dificultades de los sistemas tradicionales en satisfacer las necesidades de información, surge el concepto de Data Warehouse, como solución a las necesidades de información globales de la empresa. Este término se traduce literalmente como almacén de datos.

La ventaja principal de este tipo de sistemas se basan en un concepto fundamental: la estructura de la información. Este concepto significa el almacenamiento de información homogénea y fiable en una estructura basada en la consulta y el tratamiento jerarquizado de la misma y, lo más importante, en un entorno diferenciado de los sistemas operacionales.

Un Data Warehouse se caracteriza por ser integrado, temático, histórico y no volátil. Una de las claves del éxito en la construcción de un Data Warehouse es el desarrollo de forma gradual, seleccionando un área funcional y expandiendo progresivamente el almacén de datos a otras áreas de la organización.

Un Data Mart es una versión especial de un almacén de datos (Data Warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro de la organización pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los datos según sus necesidades.

El Data Mart es un sistema orientado a la consulta, en el que se producen procesos batch de carga de datos con una frecuencia baja y conocida. Es consultado mediante herramientas OLAP, que ofrecen una visión multidimensional de la información.

Además sobre estas bases de datos se pueden construir Sistemas de Información Ejecutivos (EIS por sus siglas en Ingles) y Sistemas de ayuda a la toma de decisiones (DSS por sus siglas en Ingles).

En síntesis podríamos decir que los Data Mart son pequeños Data Warehouse centrados en un tema o un área de negocio específico dentro de la organización.

La presente tesis es sobre la implementación de tecnologías de Data Warehouse y Bussines Intelligence para el desarrollo de un Data Mart de tarjetas de crédito de una importante institución bancaria nacional.

Un Sistema de Información de gestión de Tarjetas de Crédito basado en un Data Mart para la explotación de la información, permite consultar grandes volúmenes de información sin afectar el rendimiento operacional de la entidad bancaria. El uso de Inteligencia de Negocios como herramienta tecnológica, permitirá a la alta gerencia, jefes de área, analistas comerciales y ejecutivos de marketing tener un acceso rápido y un análisis oportuno de la información de tarjetas de crédito, así como también, servir de apoyo en la toma de decisiones en la gestión de tarjetas de crédito.

1.1 Descripción del problema

El mercado chileno de tarjetas de crédito posee más de 25 millones de tarjetas en una población con casi 15 millones de tarjetas activas y un promedio de 3,2 tarjetas por persona, lo que moviliza unos US 1.200 millones mensuales.

Hoy existe en el mercado una alta incertidumbre sobre el comportamiento de los consumidores a corto plazo, adicionalmente se ha registrado un aumento de las provisiones crediticias en caso de no pago.

En el caso específico de estudio la entidad bancaria ha registrado un importante aumento en la cartera de clientes inactivos, un porcentaje importante se concentra en las tarjetas con menos de 1 año de antigüedad, producto de lo anterior también ha aumentado el número de clientes que solicitan el cierre de su tarjeta, adicionalmente ha visto afectados en forma exponencial los índices de mora y castigo de este producto crediticio. La institución bancaria no cuenta con un sistema de información que los apoye en este ámbito.

1.2 Hipótesis

Implementado tecnologías de Data Warehouse y soluciones de Business Intelligence, es posible realizar un análisis exhaustivo de la cartera de clientes de la entidad bancaria que permitan la toma de decisiones de la alta gerencia y la implementación y seguimiento de acciones de marketing y fidelización.

Una solución de Business Intelligence utiliza todas las facilidades que proporciona la tecnología informática actual, para ofrecer información y herramientas para soporte a la toma de decisiones en la banca. Se ha enfocado a proporcionar respuestas efectivas a preguntas de negocio de alta relevancia para la exitosa gestión en el sector bancario y financiero.

1.3 Objetivo General

Mejora en la gestión y el control de la entidad bancaria sobre su cartera de tarjetas de crédito que permita aumentar la rentabilidad y disminuir el riesgo.

1.4 Objetivos Específicos

- Analizar las características de los clientes y tipos de tarjetas que componen la cartera y así poder determinar acciones para alinear estrategias hacia el público objetivo de la Institución.
- Analizar el consumo y comportamiento de los clientes para ver distribuciones en el tiempo de dichos consumos, comercios más frecuentes de compra, productos comprados y características de dichos clientes.
- Controlar y gestionar la cartera morosa de tal forma de disminuir dicha cartera y lograr el recupero en tramos de temprana morosidad y prevenir futuros castigos.
- Generar reportes analíticos con distribución de montos según rango de días de atraso para todas las líneas de producto en los últimos meses, perfil de los clientes/productos con mayor problema de morosidad y predicciones para la morosidad futura.
- Evaluar las bajas (cierres de tarjetas) y las altas (apertura de tarjetas).

1.5 Alcances

- Desarrollar un modelo analítico y de control que permita analizar la cartera de tarjetas de crédito de la entidad bancaria.
- Desarrollo de un modelo lógico analítico que pueda dar soporte a los requerimientos de información de la entidad bancaria.
- Implementar un sistema de información de gestión de tarjetas de crédito que en base a un Data Mart, brinde un acceso rápido y análisis oportuno de la información de la cartera de tarjetas de crédito.
- Implementar un prototipo de reportes e indicadores en base a herramientas de Business Intelligence de forma de apoyar de forma sostenible y continua a la organización para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones.

2 MARCO TEORICO

2.1 *Sistemas OLTP o transaccionales*

OLTP es la sigla en inglés de Procesamiento de transacciones en línea (Online Transaction Processing), es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida.

La tecnología OLTP se utiliza en innumerables aplicaciones, como banca electrónica, procesamiento de pedidos, comercio electrónico, supermercados o industria.

El procesamiento de transacciones en línea cada vez necesita más recursos para las transacciones que se propagan por una red y que puedan integrar más de una empresa. Por esta razón el software actual para los sistemas OLTP utiliza procesamiento cliente-servidor y software de intermediación (middleware) que permite a las transacciones correr en diferentes plataformas en una red.

En grandes aplicaciones la eficiencia del OLTP puede depender de lo sofisticado que sea el software de gestión de transacciones o de que tácticas de optimización se utilizan para facilitar la gran cantidad de actualizaciones concurrentes que se pueden producir en una base de datos orientada a OLTP.

En los sistemas de bases de datos descentralizados, más exigentes, los programas de intermediación OLTP distribuyen el procesamiento de transacciones entre varios ordenadores de una red. A menudo OLTP se integra en una arquitectura orientada a servicios o en un servicio web.

2.2 *Ejemplos de Sistemas OLTP o transaccionales*

Las operaciones en un cajero automático de un banco son operaciones de tipo OLTP. Para cada solicitud: consulta de saldo, retiro de efectivo, etc. el sistema responde inmediatamente. El procesamiento de los pagos de una nómina de una

empresa no es OLTP, por que el resultado de este proceso no se obtiene de forma inmediata para cada empleado si no por el contrario, la respuesta de este procesamiento es el conjunto completo de todos los cálculos para el pago de la nómina y tampoco es inmediato si no que se obtiene de procesar a todos los empleados.

2.3 Bases de datos multidimensionales: OLAP vs OLTP

Las bases de datos multidimensionales son una variación del modelo relacional que utiliza cubos OLAP¹ para organizar los datos y expresar las relaciones entre ellos. Las principales ventajas de este tipo de bases de datos son la versatilidad para cruzar información y la alta velocidad de respuesta. Esto las convierte en herramientas básicas para soluciones de Business Intelligence o de Big Data, donde el análisis de los datos resulta crucial.

En general, estamos acostumbrados a trabajar con bases de datos orientadas a transacciones (conocidas como bases de datos OLTP).

Los sistemas OLTP son bases de datos relacionales (RDBMS) orientadas a transacciones. Una transacción es una secuencia de operaciones llevada a cabo por una base de datos de manera atómica. Las operaciones pueden ser de cuatro tipos diferentes: SELECT, INSERT, DELETE y UPDATE. Al tratarse de un proceso atómico, cada transacción solo tiene dos posibles finales: commit (si se han llevado a cabo correctamente todas las operaciones) o rollback (cuando una operación de la secuencia ha fallado, en cuyo caso hay que deshacer los cambios producidos por el resto de las operaciones de la transacción y alertar del error). Las transacciones son el pilar de prácticamente cualquier programa de gestión o página web del mundo. Su necesidad se ve muy clara, por ejemplo, en el sector bancario. Para que una base de datos OLTP pueda asegurar las transacciones es necesario que sea ACID compliant. ACID es un acrónimo de Atomicity, Consistency, Isolation and Durability:

¹ OLAP acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Solución utilizada en el campo de la llamada Inteligencia empresarial 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)

Atomicidad: asegura que la operación se ha realizado correcta y completamente, de forma que no pueda quedar a medias en caso de que surja cualquier error.

Consistencia: asegura la integridad de la base de datos, es decir, que solo se ejecutan aquellas operaciones que no van a romper las reglas, restricciones y claves foráneas. La propiedad de consistencia sostiene que cualquier transacción llevará a la base de datos desde un estado válido a otro también válido.

Aislamiento: asegura que dos transacciones que afectan a la misma información (tabla, fila o celda) son independientes y no generan errores ni deadlocks. Existen cuatro posibles niveles de aislamiento: READ UNCOMMITTED, READ COMMITTED, REPEATABLE READS y SERIALIZABLE, en función del tipo de bloqueos que se crean para proteger la información.

Durabilidad: asegura que una vez realizada la operación, ésta persistirá y no se podrá deshacer, aunque falle el sistema. Será almacenada en disco, no solo en memoria.

Los sistemas OLTP son la versión tradicional de una base de datos: se diseñan utilizando un modelo entidad-relación, se implementan en los motores típicos de base de datos (Oracle, SQLServer, MySQL, etc.) y dan soporte a la mayor parte del software del mercado:

Optimizadas para lecturas y escrituras concurrentes: gracias a las propiedades ACID, el acceso a los datos está adaptado para tareas frecuentes de lectura y escritura.

Organizadas según la capa de aplicación: las tablas y los datos se estructuran según el software que los maneja: programa de gestión a medida, ERP, CRM, BPM... etc.

Adaptadas a cada empresa o departamento: dado que en muchas ocasiones se utiliza software no integrado, los formatos de los datos no suelen ser uniformes en los diferentes departamentos. Es común la falta de compatibilidad y la existencia de islas de datos.

Consultas realizadas en SQL, modificaciones en DML: el SQL (Standard Query Language) es el lenguaje de consulta universal para leer bases de datos relacionales

(cláusula SELECT), mientras que DML (Data Manipulation Language) es el estándar para realizar modificaciones (cláusulas INSERT, UPDATE y DELETE).

Gestión de datos históricos inexistente: el historial de cambios suele limitarse a datos actuales o recientes. Salvo sistemas de backup, o que el software tenga una funcionalidad específica para ello, no se suelen manejar valores históricos para cada campo.

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: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos, etc.

Un cubo OLAP no es más que un vector de varias dimensiones. Desde un punto de vista relacional, puede verse como una tabla de hechos (fact table) que tiene dos tipos de columnas:

Indicadores: también denominados métricas o ratios, son los valores numéricos con los que se opera. Por ejemplo: nº de clientes, nº de proveedores, importe de las ventas, nº de ventas, importe de las compras, nº de compras... etc.

Dimensiones: son las características por las que se pueden filtrar y cruzar los indicadores. Por ejemplo: tiempo (fijando un determinado día, mes o año), geografía (fijando un determinado país, región o ciudad), proveedor, cliente, modo de pago... etc.

Las columnas correspondientes a las dimensiones tienen claves foráneas a tablas de dimensión, que generalmente son tablas de maestros con clave-valor (esquema en estrella) o tablas organizadas en jerarquías (esquema en copo de nieve) como: ciudad – provincia – país.

En general suele resultar necesario dimensionar la volumetría de los cubos para conseguir que generen los informes deseados, y monitorizar su tamaño para garantizar que los resultados se obtienen en el tiempo esperado. En este sentido suele resultar fundamental tener en cuenta dos características básicas: la cardinalidad del cubo, posibles combinaciones de todos los valores de todas las dimensiones, y la granularidad del cubo, nivel de detalle máximo de los datos, o lo que es lo mismo, nivel de agregación mínimo de la información.

2.4 Propiedades de las bases de datos OLAP

Así como los sistemas OLTP son típicos para bases de datos convencionales y Data Warehouse, los sistemas OLAP son propios de los Data Mart.

Optimizadas para operaciones de lectura: dado que la acción más común es la consulta, estas bases de datos disponen de valores agregados y resultados pre calculado que les permiten responder en tiempo récord. Evitar las restricciones ACID les da agilidad.

Organizadas según las necesidades analíticas: los datos están estructurados según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización. Se busca evitar islas de datos.

Asíncronas: no siempre se actualizan en tiempo real, sino que se suelen alimentar con información procedente de las bases de datos relacionales mediante un proceso de extracción, transformación y carga (ETL).

Consultas realizadas en MDX: este lenguaje, MDX (Multidimensional expressions) fue desarrollado inicialmente por Microsoft y adoptado posteriormente como estándar para leer cubos OLAP. Un cliente puede manipular el cubo de distintas formas: rotarlo, rebanarlo, cortarlo en dado, etc.

Gestión de datos históricos a largo plazo: una de las exigencias analíticas consiste en realizar estudios de evolución a lo largo del tiempo, esto requiere que estas bases de datos mantengan un histórico a largo plazo, normalmente no inferior a cinco años.

En los orígenes de la tecnología OLAP la mayor parte de las compañías asumió que la única solución para una aplicación OLAP era un modelo de almacenamiento no relacional. Sin embargo, otras compañías no tardaron en descubrir que a través del uso de estructuras de base de datos (esquemas de estrella y de copo de nieve), índices y el almacenamiento de agregados, se podrían utilizar sistemas de administración de bases de datos relacionales (RDBMS) para el OLAP.

Estos vendedores llamaron a esta tecnología OLAP relacional (ROLAP). En consecuencia, las primeras compañías adoptaron el término OLAP multidimensional (MOLAP). Ambos conceptos, MOLAP y ROLAP, se explican con más detalle en los siguientes párrafos. Las implementaciones MOLAP normalmente tienen mejor

rendimiento y velocidad que la tecnología ROLAP, pero tienen problemas de escalabilidad. Por otro lado, las implementaciones ROLAP son más escalables y son frecuentemente atractivas a los clientes debido a que aprovechan las inversiones en tecnologías de bases de datos relacionales preexistentes.

Sistemas MOLAP: la información procedente de los sistemas operacionales se carga en el sistema MOLAP mediante una serie de rutinas por lotes. Una vez cargado el dato elemental en la Base de Datos multidimensional (MDDDB), se realizan una serie de cálculos para obtener los datos agregados, a través de las dimensiones de negocio, rellenando la estructura MDDDB.

Tras rellenar esta estructura, se generan unos índices y algoritmos de tablas hash² para mejorar los tiempos de accesos a las consultas. Una vez que el proceso de compilación se ha acabado, la MDDDB está lista para su uso. Los usuarios solicitan informes a través de la interfaz, y la lógica de aplicación de la MDDDB obtiene el dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos pre compilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan pre calculados y almacenados previamente.

Sistemas ROLAP: los usuarios ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP es capaz de usar datos pre calculado 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 Data Warehouse, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones

² *Una tabla hash o tabla fragmentada es una estructura de datos que asocia llaves o claves con valores.*

son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples.

2.5 Data Warehouse

Un Data Warehouse 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. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales, etc.). Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales)

El término Data Warehouse fue acuñado por primera vez por Bill Inmon³, y se traduce literalmente como almacén de datos. No obstante, y como cabe suponer, es mucho más que eso. Según definió el propio Bill Inmon, un Data Warehouse se caracteriza por ser:

Integrado: los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Temático: sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo,

³ *William H. Inmon (nacido en 1945) es un científico informático estadounidense, reconocido por muchos como el padre del Data Warehouse.*

todos los datos sobre clientes pueden ser consolidados en una única tabla del Data Warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

Histórico: el tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

No volátil: el almacén de información de un Data Warehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Otra característica del Data Warehouse es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de refresco, su fiabilidad, forma de cálculo, etc.

Los metadatos serán los que permiten simplificar y automatizar la obtención de la información desde los sistemas operacionales a los sistemas informacionales.

Los objetivos que deben cumplir los metadatos, según el colectivo al que va dirigido, son:

Dar soporte al usuario final, ayudándole a acceder al Data Warehouse con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de Business Intelligence como DSS, EIS o CMI.

Dar soporte a los responsables técnicos del Data Warehouse en aspectos de auditoría, gestión de la información histórica, administración del Data Warehouse, elaboración de programas de extracción de la información, especificación de las

interfaces para la realimentación a los sistemas operacionales de los resultados obtenidos... etc.

Por último, destacar que para comprender íntegramente el concepto de Data Warehouse, es importante entender cuál es el proceso de construcción del mismo, denominado ETL (Extracción, Transformación y Carga), a partir de los sistemas operaciones de una compañía:

Extracción: obtención de información de las distintas fuentes tanto internas como externas.

Transformación: filtrado, limpieza, depuración, homogeneización y agrupación de la información.

Carga: organización y actualización de los datos y los metadatos en la base de datos.

Una de las claves del éxito en la construcción de un Data Warehouse es el desarrollo de forma gradual, seleccionando a un departamento usuario como piloto y expandiendo progresivamente el almacén de datos a los demás usuarios. Por ello es importante elegir este usuario inicial o piloto, siendo importante que sea un departamento con pocos usuarios, en el que la necesidad de este tipo de sistemas es muy alta y se puede obtener y medir resultados a corto plazo.

En la figura 01 se muestra un ejemplo de la arquitectura básica y los componentes que posee un Data Warehouse.

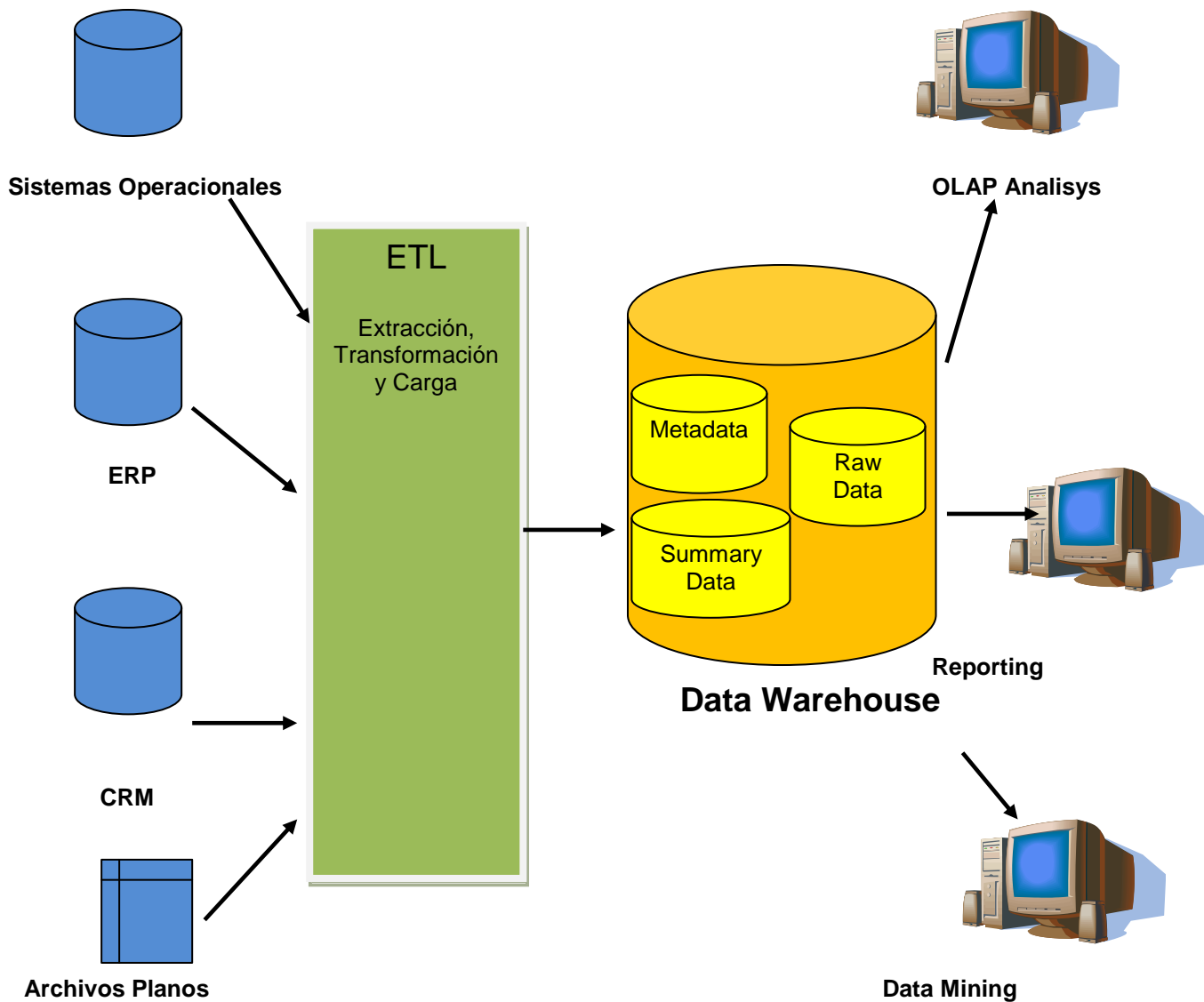


Figura 01, Ejemplo de componentes de un Data Warehouse.

2.6 Principales aportes de un Data Warehouse

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

2.7 Data Mart

Un Data Mart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer de la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un Data Mart puede ser alimentado desde los datos de un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

Cubo de Data Mart

Para crear el Data Mart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio Data Warehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. De esta forma se pueden plantear dos tipos de Data Marts:

Data Mart OLAP

Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.

Data Mart OLTP

Pueden basarse en un simple extracto del Data Warehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas report, que vienen a ser fact-tables reducidas (que agregan las dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de consultas (aunque sólo es posible en algunos SGBD avanzados, como Oracle).

Los Data Mart que están dotados con estas 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.

Diferencias entre OLTP y OLAP

OLTP	OLAP
On-Line Transaction Processing	On-Line Analytical Processing
Datos organizados por aplicación	Datos organizados por tema
Focalizado en aplicaciones específicas	Focalizado en requerimientos empresariales
No integradas	Integradas
Distintos tipos de datos	Mismo tipo de datos
Diferente formatos de archivos	Formatos de archivos estándar
Diferentes plataformas hardware	Un sólo servidor (lógico)
Realizan periódicamente altas, bajas y modificaciones	Solamente altas de datos
Se realizan acciones repetidas	Continuamente cambia el tipo de pregunta
Manipulación de datos registro a registro	Carga y acceso de datos en forma masiva
Transacciones y/o validación a nivel de registro	Validación antes o después de la carga (nunca a nivel de registro o transacción)

Normalmente se encuentra separado el Data Warehouse del OLTP, debido a:

- El Data Warehouse tiene alta demanda de recursos, puede entorpecer el desempeño del OLTP.
- Los datos del Data Warehouse normalmente son integrados de múltiples sistemas OLTP remotos.

La figura 02 representa las diferencias que existen entre un sistema transaccional (OLTP) y un Data Warehouse.

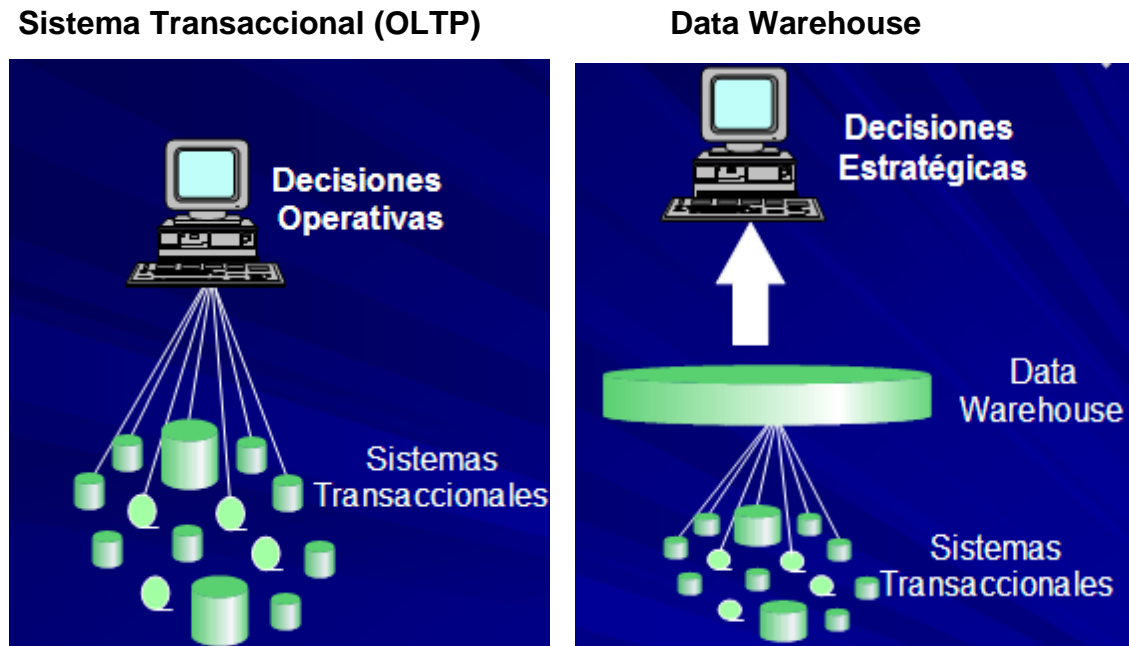


Figura 02, Representaciones OLTP y Data Warehouse.

La figura 03 describe la arquitectura general utilizada en un Data Warehouse.

Monitoreo y Administración

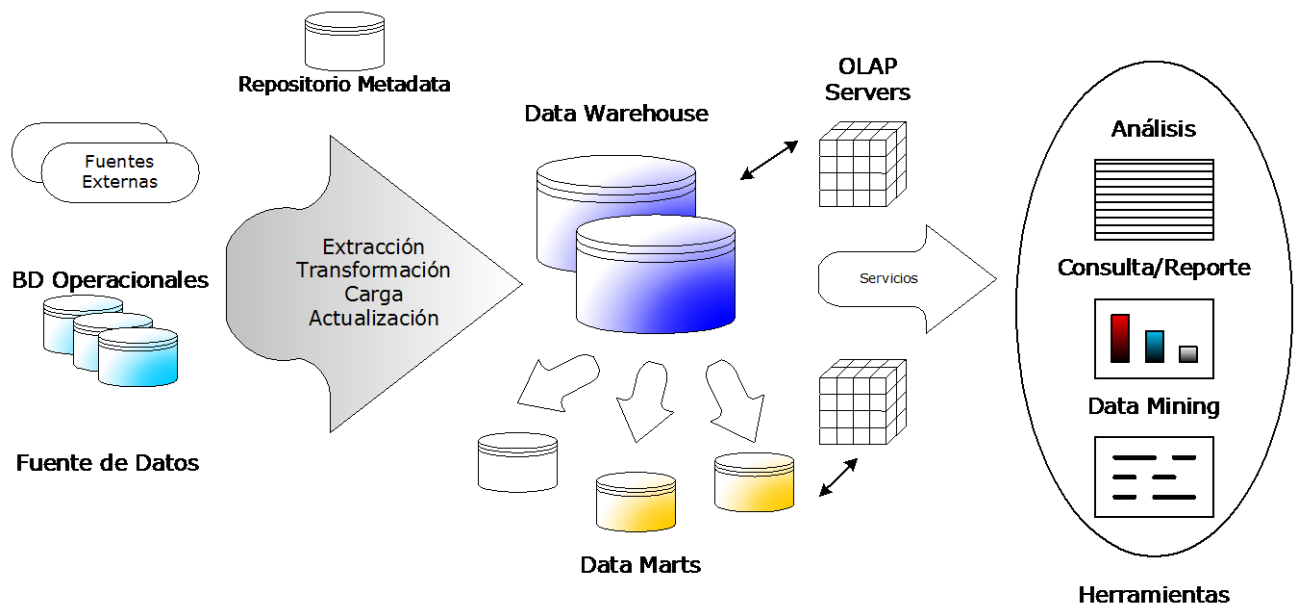


Figura 03, Ejemplo de arquitectura de un Data Warehouse.

- Extracción de datos de múltiples fuentes
- Transformación de datos
- Carga de datos
- Acceso de datos

Transformación de datos

Proceso de sumarización y cambios en los datos operacionales para reunir los objetivos de orientación a temas.

Extracción de datos

Se extraen datos de las distintas fuentes operativas a un espacio temporal para posterior limpieza y transformación.

Carga de datos

Inserción sistemática de datos en el componente de almacenamiento físico del Data Warehouse.

Acceso de datos

Los usuarios acceden al Data Warehouse mediante herramientas basados en GUI:

- Software de consultas
- Generadores de reportes
- Data Mining

Metadatos

Representan toda la información de administración y seguimiento necesarios para:

- Acceso a datos
- Compresión y utilización
- Semántica
- Origen
- Formato
- Reglas de agregación

Data Marts

Subconjuntos departamentales que focalizan objetos seleccionados.

Se caracteriza por una definición de requerimientos más rápida y fácil.

Pueden integrarse en un futuro en un Data Warehouse.

Data Mining

Extracción de información oculta y predecible de grandes bases de datos.

Predicción automatizada de tendencias y comportamientos.

Descubrimiento automatizado de modelos previamente desconocidos.

Data Warehouse

Un Data Warehouse es un conjunto de datos integrados, orientados a una materia que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones.

- No son transitorios
- No son volátiles
- No se llevan a cabo modificaciones o eliminaciones, solo inserciones
- Guarda datos sumariados

Orientados a una materia

Organiza y orienta los datos en función del usuario final y sus temas de interés

Ejemplo

Ventas, Competencias, Importaciones.

Datos integrados

Los datos provienen de diferentes fuentes, la integración de datos se logra mediante la consistencia en la convenciones de nombres, unidades de medida y codificación.

Unidades de medida

Las distintas fuentes de datos pueden tener un mismo elemento medido en:

- Centímetros
- Metros
- pulgadas

Codificación

Las distintas fuentes de datos pueden tener distintas codificaciones

Ejemplo: genero

- M, F
- 0, 1
- x, y

Convenciones de nombres

El mismo elemento puede estar referido con nombres diferentes en distintas aplicaciones

Varían con el tiempo

Mantiene tanto datos históricos como datos actuales

La información histórica es de gran importancia, permite analizar tendencias

Toma de decisión

Sistemas orientados para dar soporte a la toma de decisión dirigida a los trabajadores del conocimiento:

- Ejecutivos
- Administradores
- Analistas

Hipercubo

La figura 04 nos muestra un ejemplo de la construcción de un hipercubo.

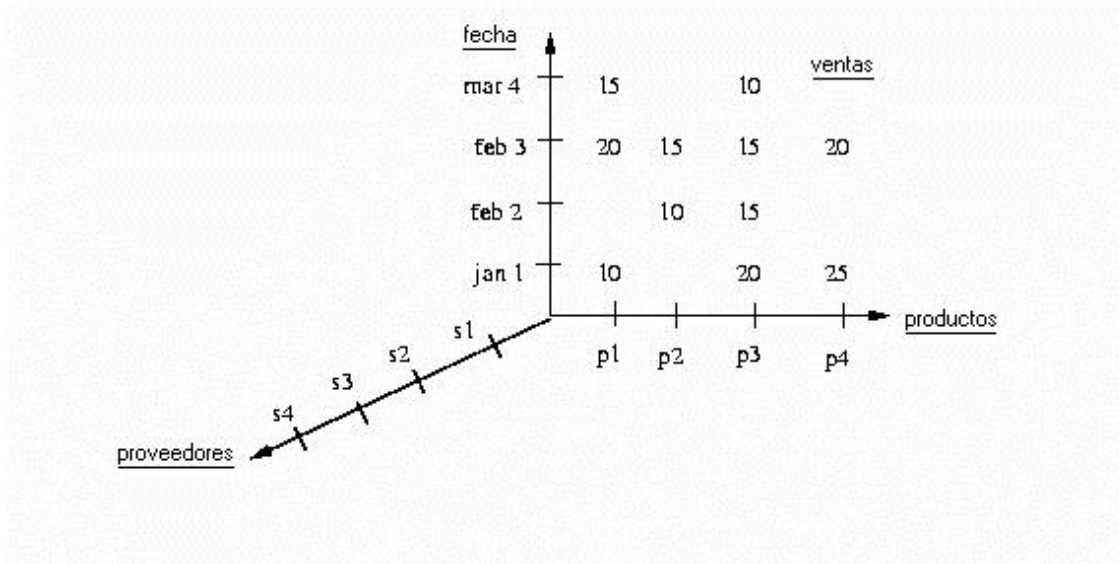


Figura 04, Representación gráfica de un hipercubo.

Operaciones

- Pivoting
- Slicing dicing
- Roll up
- Drill down

Pivoting

Rotar el cubo para ver una cara en particular.

Ejemplo:

- Analizar información referida a un fabricante

Slicing dicing

Seleccionar algún subconjunto de ese cubo.

Ejemplo:

- Analizar el cubo de datos restringiéndolo para algunos fabricantes, productos y fechas.

Roll up

Agrupamiento por alguna dimensión determinada.

Ejemplo:

- Analizar las ventas de producto a las ventas por tipo de producto.

Drill down

Operación inversa: muestra información detallada de cada agrupamiento.

Ejemplo:

- Analizar las ventas de tipo de producto a las ventas por producto.

Implementaciones relacionales

- Esquema estrella
- Copo de nieve

Esquema estrella

Un esquema estrella está compuesto por una tabla central (tabla de hechos) y un conjunto de tablas mostradas en forma radial alrededor de ésta (tablas dimensión).

La figura 05 se muestra un ejemplo de un esquema estrella.

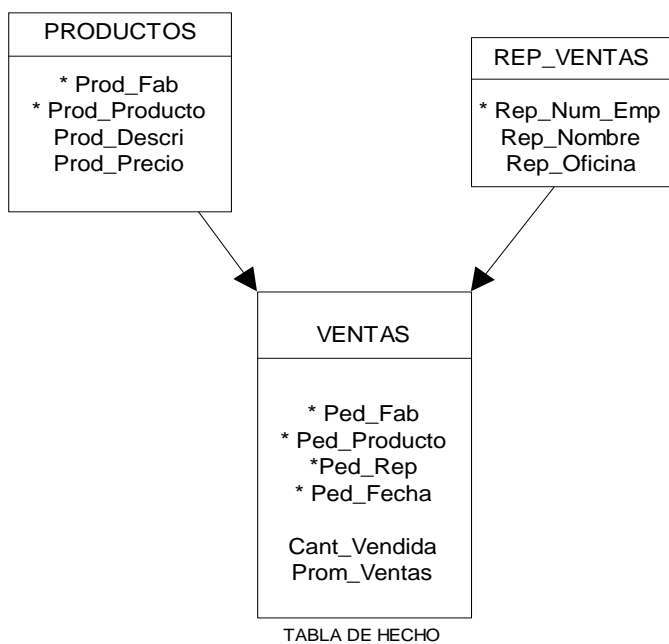


Figura 05, Ejemplo de esquema en estrella.

Copo de nieve

Un esquema copo de nieve es una extensión del esquema estrella, donde cada una de las tablas del esquema se divide en más tablas (tablas más normalizadas):

En la figura 06 se representa gráficamente un esquema copo de nieve.

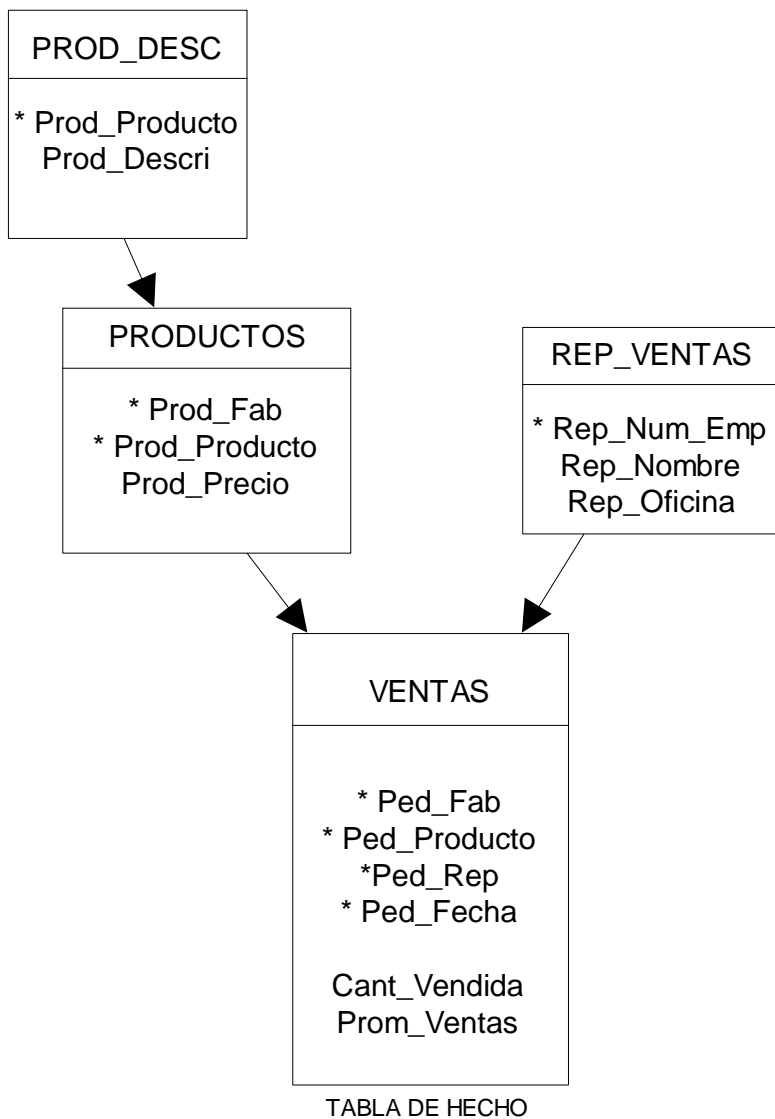


Figura 06, Ejemplo de esquema copo de nieve

Tabla de relaciones

La figura 07, representa un ejemplo de una tabla de relaciones.

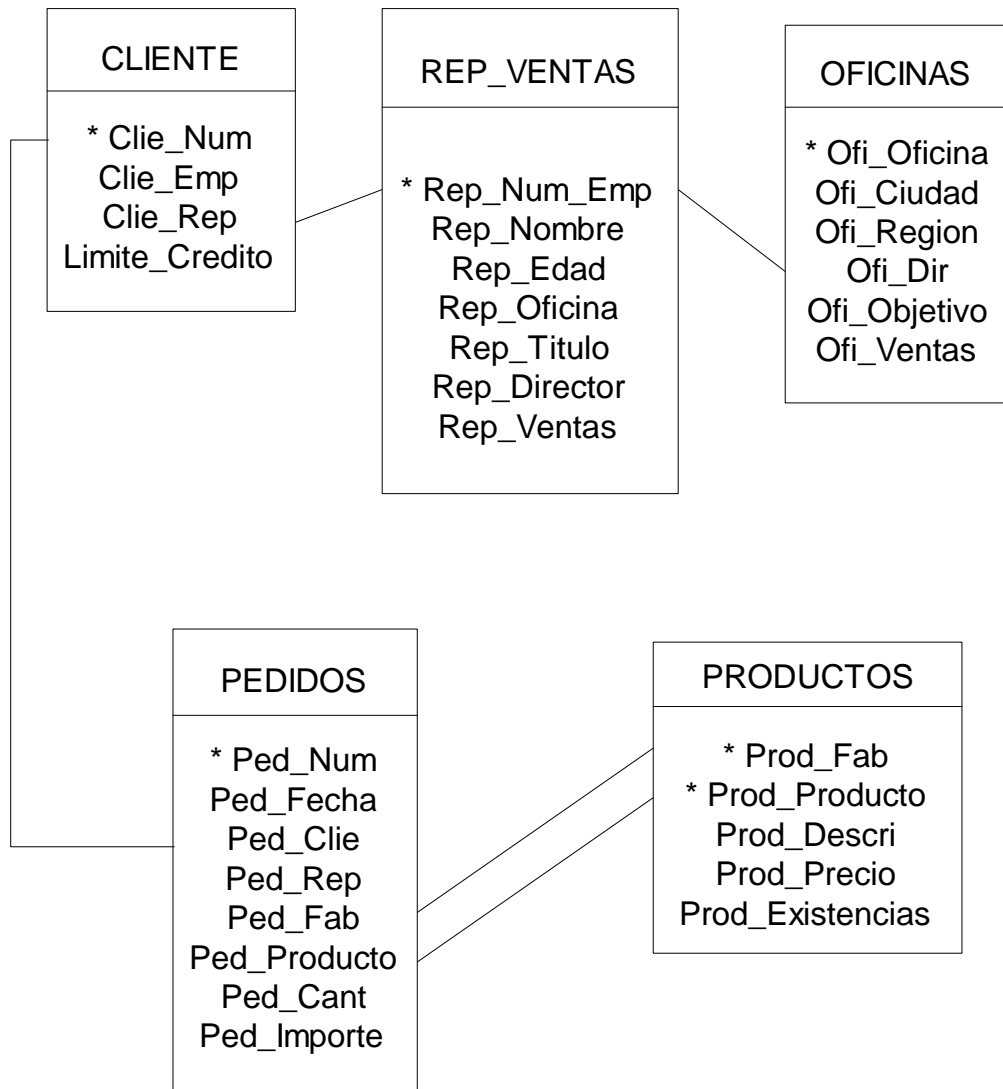


Figura 07, Ejemplo de Tabla de relaciones.

2.8 Modelo Relacional

El modelo relacional, para la gestión de una base de datos, es un modelo de datos basado en la lógica de predicados y en la teoría de conjuntos. Es el modelo más utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente. Sus bases fueron postuladas en 1970 por Edgar Frank Codd ⁴, de los laboratorios IBM, y no tardó en consolidarse como un nuevo paradigma en los modelos de base de datos.

Su idea fundamental es el uso de relaciones. Estas relaciones podrían considerarse en forma lógica como conjuntos de datos llamados tuplas. Pese a que ésta es la teoría de las bases de datos relacionales creadas por Edgar Frank Codd, la mayoría de las veces se conceptualiza de una manera más fácil de imaginar, esto es, pensando en cada relación como si fuese una tabla que está compuesta por registros (cada fila de la tabla sería un registro o tupla), y columnas (también llamadas campos).

En este modelo todos los datos son almacenados en relaciones, y como cada relación es un conjunto de datos, el orden en el que éstos se almacenen no tiene relevancia (a diferencia de otros modelos como el jerárquico y el de red). Esto tiene la considerable ventaja de que es más fácil de entender y de utilizar por un usuario no experto. La información puede ser recuperada o almacenada por medio de consultas que ofrecen una amplia flexibilidad y poder para administrar la información.

Este modelo considera la base de datos como una colección de relaciones. De manera simple, una relación representa una tabla que no es más que un conjunto de filas, cada fila es un conjunto de campos y cada campo representa un valor que interpretado describe el mundo real. Cada fila también se puede denominar tupla o registro y a cada columna también se le puede llamar campo o atributo.

Para manipular la información utilizamos un lenguaje relacional, actualmente se cuenta con dos lenguajes formales el álgebra relacional y el cálculo relacional. El álgebra relacional permite describir la forma de realizar una consulta, en cambio, el cálculo relacional sólo indica lo que se desea devolver.

⁴ *Edgar Frank Codd (Ted Codd), fue un científico informático inglés nacido en 1923, conocido por sus aportes a la teoría de bases de datos relacionales.*

Esquema

Un esquema contiene la definición de una estructura (generalmente relaciones o tablas de una base de datos), es decir, determina la identidad de la relación y qué tipo de información podrá ser almacenada dentro de ella; en otras palabras, el esquema contiene los metadatos de la relación. Todo esquema consta de un nombre de la relación (su identificador) y nombre de los atributos (o campos) de la relación y sus dominios; el dominio de un atributo o campo define los valores permitidos para el mismo, equivalente al tipo de dato por ejemplo carácter, numérico, fecha, string

Instancias

Una instancia de manera formal es la aplicación de un esquema a un conjunto finito de datos. Se puede definir como el contenido de una tabla en un momento dado, pero también es válido referirnos a una instancia cuando trabajamos o mostramos únicamente un subconjunto de la información contenida en una relación o tabla, como por ejemplo: ciertos caracteres y números (una sola columna de una sola fila), algunas o todas las filas con todas o algunas columnas. Cada fila es una tupla, el número de filas es llamado cardinalidad, el número de columnas es llamado aridad o grado.

Base de datos relacional

Una base de datos relacional es un conjunto de una o más tablas estructuradas en registros (líneas) y campos (columnas), que se vinculan entre sí por un campo en común, en ambos casos posee las mismas características como por ejemplo el nombre de campo, tipo y longitud; a este campo generalmente se le denomina ID, identificador o clave. A esta manera de construir bases de datos se le denomina modelo relacional. Estrictamente hablando el término se refiere a una colección específica de datos pero a menudo se le usa, en forma errónea como sinónimo del software usado para gestionar esa colección de datos. Ese software se conoce como SGBD (sistema gestor de base de datos) relacional o RDBMS (del inglés relational database management system).

Las bases de datos relacionales pasan por un proceso al que se le conoce como normalización de una base de datos, el cual es entendido como el proceso necesario para que una base de datos sea utilizada de manera óptima.

Entre las ventajas de este modelo están:

Garantiza herramientas para evitar la duplicidad de registros, a través de campos claves o llaves.

Garantiza la integridad referencial: Así al eliminar un registro elimina todos los registros relacionados dependientes.

Favorece la normalización por ser más comprensible y aplicable.

2.9 Modelo Dimensional

El modelado dimensional es una forma de acercar los datos a la manera en que estos serán convertidos en información útil para los usuarios del negocio. El objetivo final es que estos puedan encontrar de manera intuitiva y rápida la información que necesitan.

La aplicación del modelo dimensional tiene lugar en la fase de diseño lógico, lo que permite la traducción del esquema resultante del diseño conceptual al plano lógico. El modelo dimensional se describe en el año 1996 por Ralph Kimball⁵, como propuesta para el diseño de almacenes de datos (Data Warehouses), partiendo de la visión multidimensional que los usuarios tienen de los datos empresariales cuando se enfrentan a ellos con propósito de análisis (de análisis multidimensional, OLAP en concreto).

El análisis multidimensional consiste en analizar los datos que hacen referencia a hechos, sean económicos o de otros tipos, desde la perspectiva de sus componentes o dimensiones (utilizando para ello algún tipo de métrica o medida de negocio).

Este modelo tiene en cuenta que para el análisis multidimensional los datos se representan como si estuvieran en un espacio dimensional (cubo de datos), permitiendo

⁵ *Ralph Kimball, es reconocido como uno de los padres del concepto de Data Warehouse, se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que éste concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.*

su estudio en términos de hechos sujetos al análisis (facts, en inglés) y dimensiones que permiten diferentes puntos de vista por los que analizar esos hechos.

La analogía del cubo (recuerde el cubo de Rubik) con la visión multidimensional es válida para comprender el concepto desde un punto de vista gráfico, pero sólo es válido para un modelo de tres dimensiones. Un modelo de más de tres dimensiones suele denominarse hipercubo, y ya resulta más difícil su representación gráfica.

El modelo dimensional distingue tres elementos básicos:

Hechos: es la representación en el Data Warehouse de los procesos de negocio de la organización, por ejemplo: una venta puede identificarse como un proceso de negocio. Los hechos se podrán reconocer además porque siempre tienen asociada una fecha, y una vez registrados no se modifican ni se eliminan (para no perder la historia).

Métrica: son los indicadores de negocio de un proceso de negocio. Aquellos conceptos cuantificables que permiten medir nuestro proceso de negocio. Por ejemplo, en una venta tenemos el importe de la misma y la cantidad vendida. Existen métricas derivadas, como el precio unitario, que se obtiene al dividir el importe total por las unidades vendidas.

Dimensión: es la representación en el Data Warehouse de un punto de vista para los hechos de cierto proceso de negocio. Si regresamos al ejemplo de una venta, para la misma tenemos el cliente que ha comprado, la fecha en la que se ha realizado, el producto vendido, estos conceptos pueden ser considerados como vistas para este proceso de negocio. Puede ser interesante recuperar todas las compras realizadas por un cliente, o para un producto o familia de productos, o para un lapso determinado.

Pero ¿qué hay de su representación? Igual que sucede en el modelo relacional, el modelo dimensional adopta el concepto de relación (tabla) como estructura básica del modelo. Pero a diferencia del modelo relacional, que no hace distinción entre

relaciones, el modelo dimensional distingue entre relaciones de hecho (tablas de hecho) y relaciones de dimensión (tablas de dimensión).

Los conceptos básicos: hechos, métricas y dimensiones (facts, measures and dimensions) se representan en el modelo como relaciones (tablas) dentro de un esquema dimensional. Según las técnicas de modelado utilizadas, ese esquema dimensional puede adoptar forma de estrella o de copo de nieve.

En la figura 08 se muestra un ejemplo de un esquema en estrella para el caso de las ventas diarias en un conjunto de tiendas:

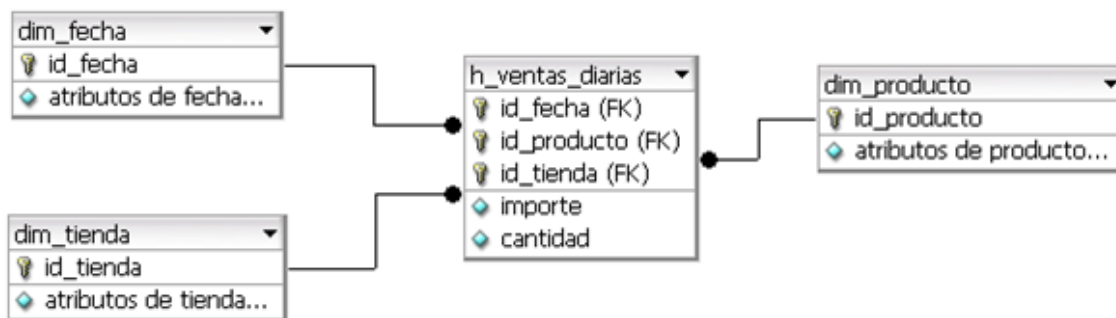


Figura 08, Representación esquema estrella para ventas.

Vemos que existe una tabla de hechos en el centro (h_ventas_diarias), y una tabla de dimensión para cada dimensión de análisis que participa de la descripción de ese hecho (dim_xxx).

Las columnas de la tabla de hechos incluyen las dimensiones que identifican el hecho, y los hechos o medidas del negocio. Las columnas en las tablas de dimensión incluyen atributos asociados a cada posible valor de la dimensión.

Los atributos de una dimensión permitirán agrupar los hechos jerárquicamente, para poder consolidar y desagregar las mediciones según se desee. Como ejemplos tenemos:

Para la dimensión fecha: año, mes, día, nombre del mes, día de la semana, trimestre.

Para la dimensión producto: tipo de producto, familia, unidad de medida.

Cuando a una dimensión no se le pueden asociar múltiples atributos, se dice que tenemos una dimensión degenerada, y solo aparecerá como una columna en la tabla de hechos. Por ejemplo, si tuviéramos en nuestro ejemplo una dimensión “Estado de la Venta”, con los posibles valores: “Confirmada”, “Pendiente”, “Cancelada”; en este caso la dimensión serviría solo para separar las ventas según su estado, y no se requieren atributos adicionales.

Eventualmente, pudiera surgir un esquema en estrella integrado solo por una tabla de hechos con dimensiones degeneradas.

Por otro lado, cuando se decide normalizar las tablas de dimensiones para eliminar campos redundantes, tenemos lo que se conoce como esquema en copo de nieve, tal como lo señala la figura 09.

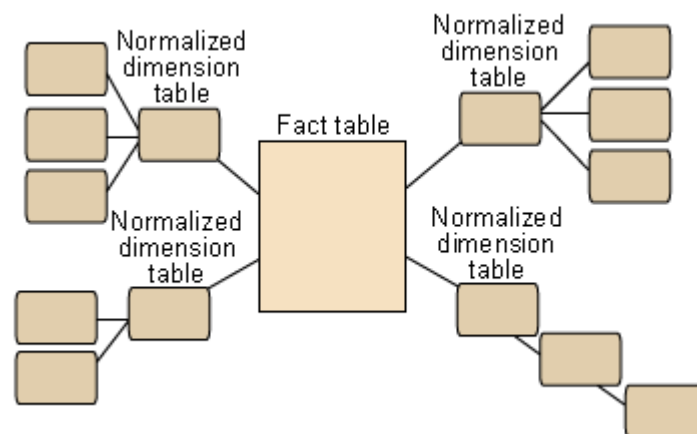


Figura 09, Ejemplo esquema copo de nieve.

Esquema Snow flake (copo de nieve)

La diferencia del esquema snowflake con el esquema estrella, está en la estructura de las tablas de lookup: las tablas de lookup en el esquema snowflake están normalizadas. Cada tabla lookup contiene solo el nivel que es clave primaria en la tabla y la foreign key de su parentesco del nivel más cercano del diagrama.

Tabla Fact o de Hechos

Es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. Estas medidas se hacen sobre el grano, o unidad básica de la tabla. El grano o la granularidad de la tabla queda determinada por el nivel de detalle que se almacenará en la tabla. Por ejemplo, para el caso de producto, mercado y tiempo antes visto, el grano puede ser la cantidad de madera vendida mensualmente. El grano revierte las unidades atómicas en el esquema dimensional. Cada medida es tomada de la intersección de las dimensiones que la definen. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. La razón de estas características es que así se facilita que los miles de registros que involucran una consulta sean comprimidos en unas pocas líneas en un set de respuesta. La clave de la tabla fact recibe el nombre de clave compuesta o concatenada debido a que se forma de la composición (o concatenación) de las llaves primarias de las tablas dimensionales a las que está unida.

Así entonces, se distinguen dos tipos de columnas en una tabla fact: columnas fact y columnas key. Donde la columna fact es la que almacena alguna medida de negocio y una columna key forma parte de la clave compuesta de la tabla.

Tablas Lock-up o Dimensionales

Estas tablas son las que se conectan a la tabla fact, son las que alimentan a la tabla fact. Una tabla lockup almacena un conjunto de valores que están relacionados a una dimensión particular. Tablas lock_up no contienen hechos, en su lugar los valores en las tablas lock_up son los elementos que determinan la estructura de las dimensiones. Así entonces, en ellas existe el detalle de los valores de la dimensión respectiva. Una tabla lock_up está compuesta de una primary key que identifica unívocamente una fila en la tabla junto con un conjunto de atributos, y dependiendo del diseño del modelo multidimensional puede existir una foreign key que determina su relación con otra tabla lock_up. Para decidir si un campo de datos es un atributo o un

hecho se analiza la variación de la medida a través del tiempo. Si varía continuamente implicaría tomarlo como un hecho, caso contrario será un atributo.

Los atributos dimensionales son un rol determinante en un DDW. Ellos son la fuente de todas las necesidades que debieran cubrirse. Esto significa que la base de datos será tan buena como lo sean los atributos dimensionales, mientras más descriptivos, manejables y de buena calidad, mejor será el Data Warehouse.

Pasos básicos del Modelamiento Multidimensional

1. Decidir cuáles serán los procesos de negocios a modelar, basándose en el conocimiento de éstos y de los datos disponibles. Ejemplo: Gastos realizados por cada mercado para cada ítem a nivel mensual. Productos vendidos por cada mercado según el precio en cada mes.

2. Decidir el Grano de la tabla Fact de cada proceso de negocio.

Ejemplo: Producto x mercado x tiempo. En este punto se debe tener especial cuidado con la magnitud de la base de datos, con la información que se tiene y con las preguntas que se quiere responder. El grano decidirá las dimensiones del DDW. Cada dimensión debe tener el grano más pequeño que se pueda puesto que las preguntas que se realicen necesitan cortar la base en caminos precisos (aunque las preguntas no lo pidan explícitamente).

3. Decidir las dimensiones a través del grano. Las dimensiones presentes en la mayoría de los DDW son: tiempo, mercado, producto, cliente. Un grano bien elegido determina la dimensionalidad primaria de la tabla fact. Es posible usualmente agregar dimensiones adicionales al grano básico de la tabla fact, donde estas dimensiones adicionales toman un solo valor para cada combinación de las dimensiones primarias. Si se reconoce que una dimensión adicional deseada viola el grano por causar registros adicionales a los generados, entonces el grano debe ser revisado para acomodar esta dimensión adicional.

4. Elegir las mediciones del negocio para la tabla fact. Se deben establecer los ítems que quedarán determinados por la clave compuesta de la tabla fact.

La Dimensión Tiempo

Virtualmente se garantiza que cada DW tendrá una tabla dimensional de tiempo, debido a la perspectiva de almacenamiento histórica de la información. Usualmente es la primera dimensión en definirse, con el objeto de establecer un orden, ya que la inserción de datos en la base de datos multidimensional se hace por intervalos de tiempo, lo cual asegura un orden implícito.

Dimensiones que varían lentamente en el tiempo

Son aquellas dimensiones que se mantienen “casi” constantes en el tiempo y que pueden preservar la estructura dimensional independiente del tiempo, con sólo agregados menores relativos para capturar la naturaleza cambiante del tiempo. Cuando se encuentra una de estas dimensiones se está haciendo una de las siguientes fundamentales tres elecciones. Cada elección resulta en un diferente grado de seguimiento sobre el tiempo:

Tipo 1: Sobrescribir el viejo valor en el registro dimensional y por lo tanto perder la capacidad de seguir la vieja historia.

Tipo 2: Crear un registro dimensional adicional (con una nueva llave) que permita registrar el cambio presentado por el valor del atributo. De esta forma permanecerían en la base tanto el antiguo como el nuevo valor del registro con lo cual es posible segmentar la historia de la ocurrencia.

Tipo 3: Crear un campo “actual” nuevo en el registro dimensional original el cual almacene el valor del nuevo atributo, manteniendo el atributo original también. Cada vez que haya un nuevo cambio en el atributo, se modifica el campo “actual” solamente. No se mantiene un registro histórico de los cambios intermedios.

Niveles

Un nivel representa un nivel particular de agregación dentro de una dimensión; cada nivel sobre el nivel base representa la sumarización total de los datos desde el nivel inferior. Para un mejor entendimiento, veamos el siguiente ejemplo: consideremos una dimensión tiempo con tres niveles: mes, semestre y año. El nivel mes representa el nivel base, el nivel semestre representa la sumarización de los totales por mes y el nivel año representa la sumarización de los totales para los semestres.

Agregar niveles de sumarización otorga flexibilidad adicional a usuarios finales de aplicaciones EIS/ DSS para analizar los datos.

Sobre Jerarquías

A nivel de dimensiones es posible definir jerarquías, las cuales son grupos de atributos que siguen un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, con cada nivel representando el total agregado de los datos del nivel inferior. Las jerarquías definen cómo los datos son sumarizados desde los niveles más bajos hacia los más altos. Una dimensión típica soporta una o más jerarquías naturales. Una jerarquía puede pero no exige contener todos los valores existentes en la dimensión. Se debe evitar caer en la tentación de convertir en tablas dimensionales separadas cada una de las relaciones muchos-a-uno presentes en las jerarquías. Esta descomposición es irrelevante en el Modelamiento Multidimensional y Planeamiento del espacio ocupado en disco y sólo dificulta el entendimiento de la estructura para el usuario final, además de destruir el desempeño del browsing.

3 CONTEXTO

3.1 La entidad bancaria

Los continuos cambios en la economía y en el entorno político mundial, han dado como resultado modificaciones estructurales en el sector financiero, siendo cada vez más importante cumplir con los retos que impone la globalización de los mercados y la correspondiente necesidad de autorregulación. Las instituciones y los gobiernos han visto la necesidad de llegar a consensos, como los impartidos por el Comité de Basilea, los cuales definen los principios para el monitoreo y control de las instituciones financieras.

Las instituciones financieras invierten importantes recursos en la aplicación de conceptos y metodologías de administración financiera y de control, buscando la información que sustente la toma de decisiones y permitiendo a los administradores evaluar el cumplimiento de sus planes estratégicos de forma permanente y oportuna. La entidad bancaria existe desde el año 1998 entregando a sus clientes servicios desde el lugar en que se encuentren enfocados en la atención a sus requerimientos financieros, esta entidad no cuenta con sucursales.

El modelo de negocios está enfocado en brindar una atención las 24 horas del día, los 7 días de la semana a través de distintos canales (internet, teléfono) por medio de ejecutivos integrales capacitados bajo altos estándares de calidad y seguridad en el servicio.

3.2 Tarjeta de crédito

Se entiende por "tarjeta de crédito", al documento que permite a su titular o usuario, disponer de un crédito del emisor, adquirir bienes o servicios en establecimientos afiliados al correspondiente sistema, además de otro tipo de beneficios que se le puedan otorgar.

Se le llama "tarjeta" debido a que consiste en un dispositivo de plástico del tamaño de una tarjeta de visita, que cuenta con una banda magnética en la que se almacena información con la identificación de su dueño y otros datos relacionados.

Contiene además el logotipo y nombre del banco que es la entidad que ha emitido la tarjeta, un número de identificación, el nombre del titular, la fecha de vencimiento de la misma (expresadas con mes y año). Las tarjetas de crédito son intransferibles.

En general en el mercado nacional operan varios sistemas de tarjetas de crédito emitidas por bancos, incluyendo marcas internacionales (Visa, Mastercard, Diners, American Express, entre otras).

Las tarjetas de crédito permiten hacer compras en los comercios que estén afiliados a la red de pagos de la tarjeta (en forma física o virtual, a través del teléfono o Internet) y también, es posible sacar dinero desde cajeros automáticos.

3.2.1 Estado de cuenta Tarjeta de crédito

El Estado de Cuenta (EECC) corresponde al documento mensual que contiene el resumen de toda la información relacionada con el uso de la Tarjeta de Crédito en el periodo de facturación, incluyendo las adquisiciones de bienes, contratación de servicios y Avances en Efectivo. En resumen, informa el detalle de todos los movimientos registrados en la Tarjeta de Crédito en un mes. Con el Estado de Cuenta se puede obtener información detallada de las tasas y cobros asociados al uso de la tarjeta, información detallada de compras, además, de conocer el CAE (Carga Anual Equivalente) asociado a la Tarjeta de Crédito.

3.2.2 Número de cuenta (Tarjeta)

Este es el número de la cuenta que coincide con el número de la primera tarjeta titular entregada.

3.2.3 Cupos de la Tarjeta de Crédito

Cupo Total : Es el monto total de crédito que tiene disponible para la utilización de la tarjeta de crédito, sea por uso como medio de pago o para realizar avances en efectivo, aunque este último podrá limitarse a una parte del cupo total o establecerse como un cupo independiente.

Cupo Utilizado: El cupo utilizado corresponde al monto total utilizado ya sea por el titular o los adicionales en compras y/o avances en efectivo.

Cupo Disponible: Corresponde al monto total de crédito disponible en la tarjeta de crédito con cargo al cual se pueden realizar adquisición de bienes o contratación de servicios, avances en efectivo o devengar cualquiera de los conceptos incluidos en la carga anual equivalente, los que pueden pagarse total o parcialmente en el plazo establecido para ello o en los siguientes periodos de facturación, devengando el saldo insoluto la tasa mensual de crédito rotativo, según sea el caso. Las compras en cuotas rebajan el cupo disponible en el monto total de la compra. Los avances en efectivo también rebajan el cupo disponible al solicitarlos.

3.2.4 Tasa de Interés Vigente

Corresponden a las tasas vigentes en caso de pagar un monto menor a la deuda facturada, para compras en cuotas y avances en efectivo. Estas tasas van variando de forma mensual.

3.2.5 CAE

Carga anual equivalente que corresponde a un indicador, expresado en forma de porcentaje, revela el costo del crédito disponible en la tarjeta de crédito en un periodo anual, cualquiera sea el plazo pactado para el pago de la obligación. En el estado de cuenta se informa la CAE de crédito rotativo, la CAE de compra en cuotas y la CAE de avance en efectivo, indicador que será de referencia para el próximo periodo. Toda CAE se calcula sobre un supuesto de gasto mensual de 20 UF y pagadero en 12 cuotas.

3.2.6 Periodo Facturado

Es el espacio de tiempo considerado para registrar los movimientos de la(s) Tarjeta(s) de Crédito que se incluyen en el Estado de Cuenta para informar al consumidor.

3.2.7 Fecha Tope de Pago

Corresponde al día del mes estipulado en el contrato de Tarjeta de Crédito, en el que se deben pagar las obligaciones registradas en el Estado de Cuenta. En resumen es la fecha límite estipulada en el contrato para pagar.

3.2.8 Promociones y Beneficios Tarjetas de Crédito

Las Tarjetas de Crédito cuentan con promociones las cuales se pueden confirmar por medio de la web de la entidad ya que varían constantemente y otras que son permanentes como por ejemplo el programa de puntos: con el cual los clientes acceden a una variedad de productos y servicios. Los clientes pueden utilizar los puntos en viajes, tiendas, etc. y para canjear productos. Tarjetas Adicionales Sin Costo y con Límite de Gasto Mensual: comparten los beneficios de las tarjetas titulares, pero con el control sobre el gasto, ya que ofrece la posibilidad de asignar cupos diferenciados entre la Tarjeta Titular y cada Tarjeta Adicional de acuerdo a sus necesidades.

3.2.9 Cargo de Interés por Mora Tarjetas de Crédito

Este cargo se aplica cuando cliente tiene retraso en el pago de su facturación. El cálculo se realiza en relación al pago que cliente efectúe y los días transcurridos de retraso.

3.2.10 Tarjetas de Crédito en cartera vencida

Cartera vencida corresponde a bloqueo de plástico cuando la Tarjeta de Crédito lleva tres estados de cuenta impagos, en este caso la tarjeta se elimina de sistema y si la deuda es igual o superior a UF 3 se genera una colocación de crédito.

Para realizar pago cliente tiene la alternativa de realizar cancelación total de deuda vía telefónica y la alternativa de repactar deuda. Se debe consultar a cliente cual la forma de pago que utilizará.

3.2.11 Cierre de Tarjeta de Crédito Voluntario

Este procedimiento se da cuando cliente solicita voluntariamente el cierre de su Tarjeta de Crédito (Visa y/o Mastercard), sin necesidad del cierre de su cuenta corriente. Si el cliente llama se procede a revisar la tarjeta de crédito para confirmar si posee deuda, los valores, motivos del cierre y se intenta la retención del cliente.

3.3 Descripción del problema

En este entorno, la institución financiera en estudio mantiene un sistema de información gerencial basado en la generación de reportes estáticos producidos por demanda. Estos reportes a su vez deben ser transformados con herramientas de análisis como hojas de cálculo, para producir la información mínima necesaria para la toma de decisiones gerenciales, demandando un gran esfuerzo técnico y analítico de diversas áreas de la institución.

3.4 Solución Propuesta

Se desarrollará una solución basada en un Data Mart que, aplicando inteligencia de negocios, pueda satisfacer las necesidades del área comercial de la entidad bancaria de contar con un sistema de información de gestión de tarjetas de crédito y que sirva de apoyo en la toma de decisiones con respecto a la captación de clientes,

reducción de cierres, control de la cartera morosa y para poder determinar acciones para alinear estrategias hacia el público objetivo de la Institución.

3.5 Justificación de la solución

La implementación del Data Mart permitirá a la organización mediante una base homologada, consolidar la información proveniente de las fuentes de datos internas y externas del Banco proporcionando un canal único de información y memoria corporativa.

La construcción del Data Mart ayudará a identificar debilidades en las fuentes de información, las cuales mediante un proceso correctivo permitirán contar con un sistema de información gerencial óptimo y una fuente de datos depurada.

La organización podrá contar con un análisis de la información con enfoques diversos del negocio: la información de operaciones, saldos y datos de los clientes del Banco contenida en el Data Mart, permitirá al usuario final obtener información gerencial desde diversos enfoques del giro del negocio.

La rentabilidad es el objetivo final de todo negocio, y la alta competitividad en el sector financiero hace crítico el adecuado control de este vital parámetro. Esta herramienta permitirá al Banco obtener una medición y un control efectivo de la rentabilidad, con el objetivo final de aplicar una fórmula detallada de cálculo de la rentabilidad proporcionando una visión completa y multidimensional de la rentabilidad por tipo de producto, por oficina y por cliente.

Este proyecto proporcionará al Banco las herramientas necesarias para el procesamiento de los datos, para la evaluación de la situación comercial, y la ejecución de la estrategia de marketing de relaciones, permitiendo la identificación, segmentación y selección de los clientes más valiosos para la institución, la determinación de los productos, servicios y puntos de atención que el cliente utiliza, así como los productos y servicios que de acuerdo a su perfil y características debería utilizar, para maximizar su rentabilidad y explotar al máximo el potencial de negocios de cada cliente.

Como solución a la difusión masiva de la información de negocio a las áreas estratégicas del Banco permitirá administrar y visualizar los reportes a medida del

usuario final, sin necesidad de estar conectado directamente a la base de datos transaccional.

4 DESARROLLO

4.1 *Marco Metodológico*

Para el desarrollo del Data Mart se utilizará la metodología de Kimball. La metodología indicada por Ralph Kimball en su libro *The Data Warehouse lifecycle toolkit* conduce a una solución completa en una cantidad de tiempo relativamente pequeña.

Además, debido a la gran cantidad de documentación que se puede encontrar y a los numerosos ejemplos aportados en diferentes entornos, permite encontrar una respuesta a casi todas las preguntas que puedan surgir, sobre todo cuando no se dispone de la experiencia previa necesaria.

Por otro lado, este tipo de metodología bottom-up permite que, partiendo de cero, podamos empezar a obtener información útil en cuestión de días y después de los prototipos iniciales, comenzar el ciclo de vida normal que nos ofrezca una solución completa de Business Intelligence.

Los Data Marts resultantes son fácilmente consultables tanto para los desarrolladores como para los usuarios finales. La relación directa entre los hechos y dimensiones conceden a cualquier usuario la posibilidad de construir consultas muy sencillas, la mayoría de las veces sin tener a mano la documentación de los metadatos.

La metodología indicada es ideal para los primeros pasos de implantación de Business Intelligence a un cliente, cuando la complejidad de almacenamiento de datos no es demasiado grande y donde la infraestructura del BI se encarga de los datos procedentes de un número limitado de fuentes.

4.1.1 **Planificación del Proyecto**

La planificación del proyecto comprende varias etapas considerando la organización, la dirección y el control de los recursos de la empresa/departamento/unidad, para alcanzar un objetivo en un plazo, con un costo y calidad preestablecidos.

No se debe olvidar que el proyecto se desarrolla para algunos usuarios de la organización, a los que además se les pide que colaboren con él. Se debe conseguir que las relaciones con estos usuarios sean excelentes, ya que ello asegurará su participación y mejorará su evaluación del proyecto, por lo que será necesario comunicarles tanto el inicio del proyecto como su posterior evolución.

La figura 10 nos muestra las etapas y actividades a desarrollar:

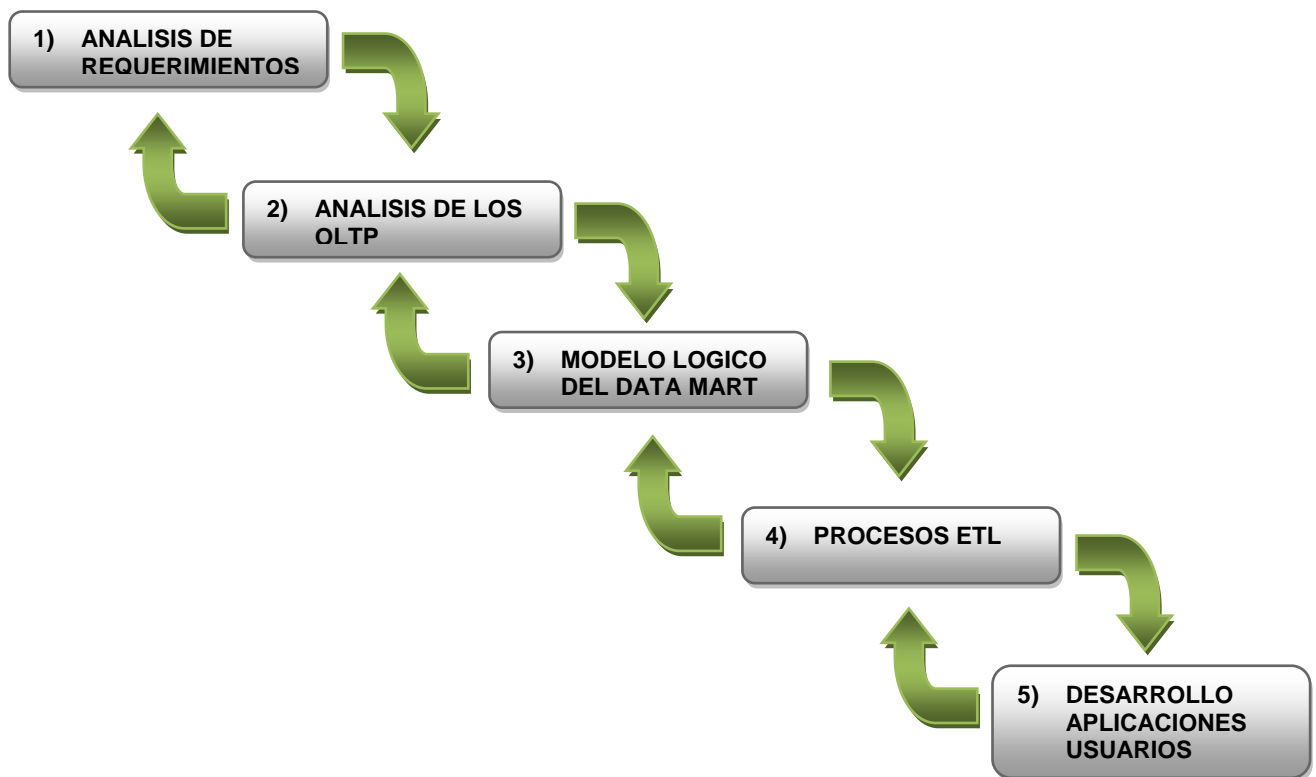


Figura 10, Representación gráfica etapas del proyecto.

4.1.2 Tareas por etapa

A continuación la figura 11 indica el detalle de las tareas que se realizarán en cada etapa:

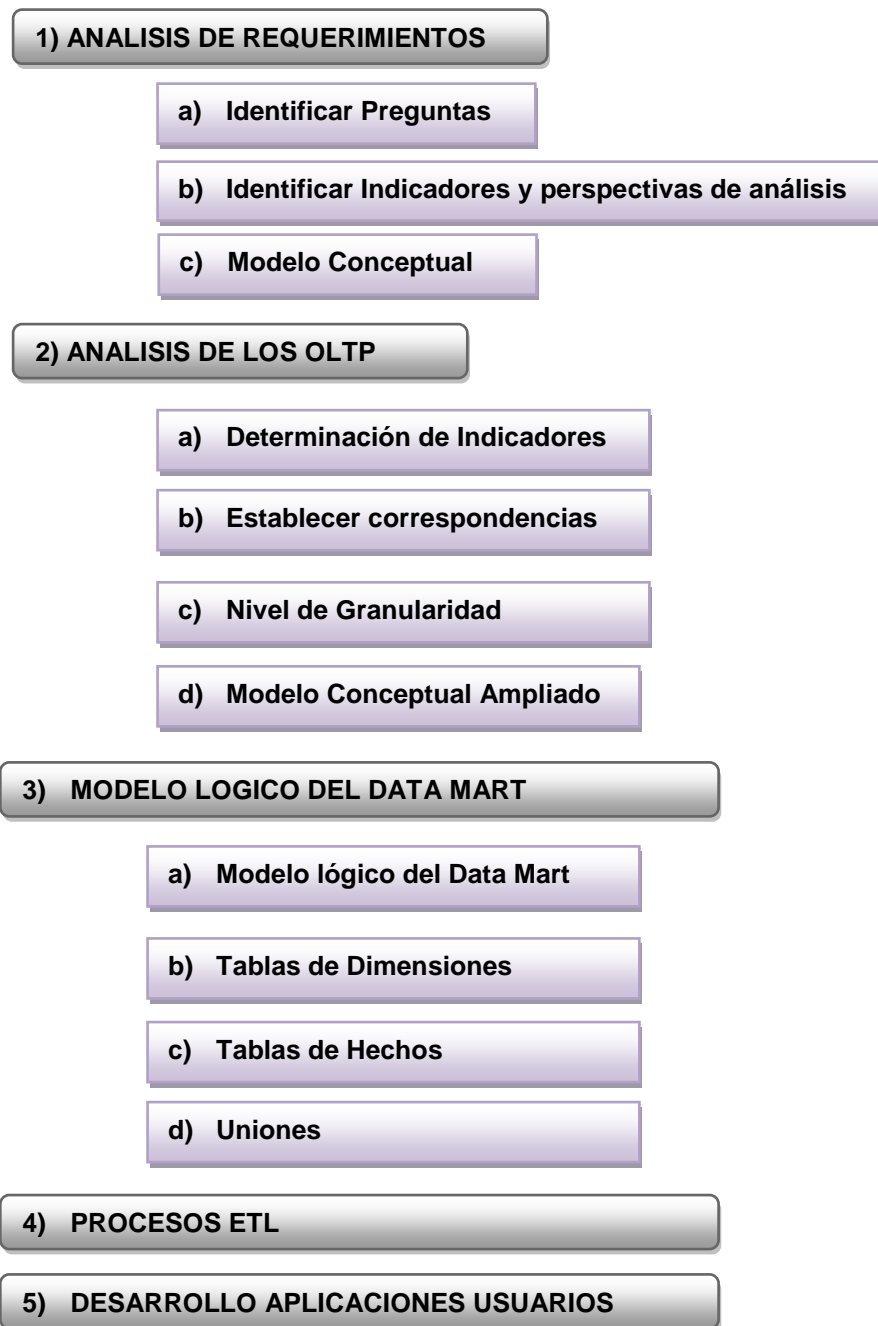


Figura 11, Descripción de tareas para cada etapa del proyecto.

4.1.3 Flujo de Tareas

En las figura 12 y 13 se muestra una representación gráfica con el flujo de las tareas que se desarrollaran en cada etapa y su respectiva Carta Gantt:

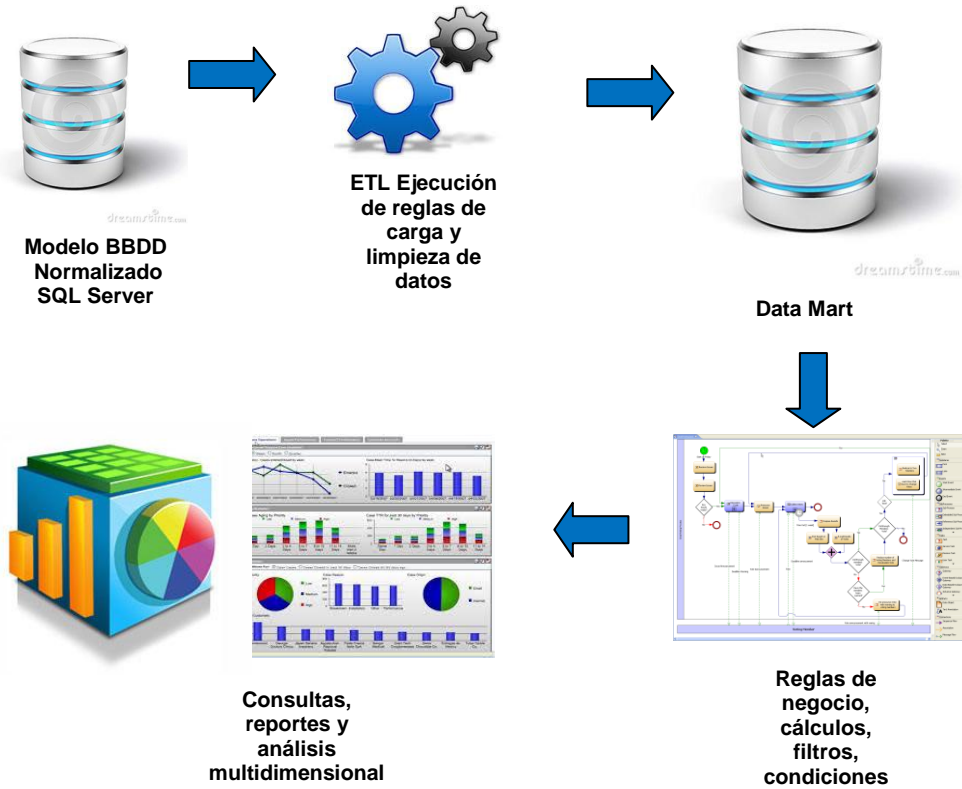


Figura 12, Descripción del flujo de tareas para el proyecto.

Etapa	Actividad	Inicio	Duración(Días)	Fin
Planificación del proyecto	Definición del proyecto	16-jun	2	17-jun
	Definición de la planificación y la gestión del proyecto	18-jun	3	20-jun
Análisis de Requerimientos	Identificar Preguntas	21-jun	10	30-jun
	Identificar Indicadores y Perspectivas de Análisis	01-jul	10	10-jul
Análisis de los OLTP	Desarrollo Modelo Conceptual	11-jul	10	20-jul
	Determinación de Indicadores	21-jul	5	25-jul
	Establecimiento de Correspondencias	26-jul	5	30-jul
	Determinación Nivel de Granularidad	31-jul	5	04-ago
Modelo Lógico del Data Mart	Desarrollo Modelo Conceptual Ampliado	05-ago	5	09-ago
	Desarrollo Modelo Lógico Data Mart	10-ago	10	19-ago
	Creación de Tablas de Dimensiones	20-ago	10	29-ago
	Creación de Tablas de Hechos	30-ago	5	03-sep
Procesos ETL	Creación de Uniones	04-sep	5	08-sep
	Construcción de Procesos	09-sep	5	13-sep
Desarrollo Aplicaciones Usuarios	Definición y Desarrollo de Reportes para usuarios	14-sep	5	18-sep
				95

Figura 13, Carta Gantt del proyecto Data Mart Tarjetas de Crédito.

4.1.4 Análisis de requerimientos

El primer paso en esta etapa será identificar los requerimientos de información de los principales usuarios a través de preguntas que sean claves para el logro de los objetivos estratégicos de la institución. Posteriormente se analizarán estas preguntas a fin de identificar cuáles serán los indicadores y dimensiones que serán consideradas para la construcción del Data Mart. Para finalizar se confeccionará un modelo conceptual, cuyo objetivo es visualizar el resultado obtenido en esta primera etapa.

4.1.4.1 Identificar preguntas

Comenzaremos con la recopilación de las necesidades de información, el cual se llevará a cabo a través de diferentes técnicas como por ejemplo entrevistas, cuestionarios, observaciones, etc.

El análisis de los requerimientos de los diferentes usuarios, es el punto inicial, ya que ellos son los que deben indicar claramente lo que se espera de la construcción de este Data Mart en relación a su funcionalidad y características.

El principal objetivo en esta fase, es obtener e identificar las necesidades de información y que serán indispensables para cumplir las metas y estrategias definidas por la organización, y que ayudará a la correcta toma de decisiones.

Esta información nos entregará el principal soporte para desarrollar los siguientes pasos en la construcción de este proyecto por lo cual se prestará especial atención al relevar los datos.

Se comprobará que el resultado del mismo haga explícitos los objetivos estratégicos planteados por la empresa en estudio, se enfocarán las necesidades de información en los procesos principales relativos al producto tarjeta de crédito. Se formularán preguntas complejas sobre el negocio, que incluyan variables de análisis que se consideren relevantes, ya que son estas las que permitirán estudiar la información desde diferentes perspectivas.

Toda la información que se requiere está contenida de una forma u otra en la base de datos relacional de SQL Server existente en la entidad ya que de otra forma no se podría elaborar el Data Mart.

Para este trabajo de investigación se realizó una encuesta entre los usuarios en busca de necesidades de información abarcando los procesos más importantes insertos en las actividades diarias de la empresa y que de alguna manera estuvieran soportados por la base de datos existente. Los 4 procesos elegidos fueron: Facturación, Morosidad, Ventas y Cierres, todos los procesos asociados al producto Tarjeta de Crédito. Posteriormente se procedió a identificar cual era la información de interés acerca de estos procesos y cuáles eran las variables o perspectivas importantes para la toma de decisiones. Se consultó sobre cuáles eran los indicadores que representan cada uno de los procesos y que es lo que se desea analizar, a continuación se indica para cada proceso las respuestas obtenidas:

Ventas: Cantidad de clientes adquiridos, cupo de crédito entregado, marca de la tarjeta.

Facturación: monto de consumos, cantidad de clientes y nombre de comercio.

Morosidad: Cantidad de clientes morosos y monto en mora para cada tramo de morosidad.

Cierres: Cantidad de tarjetas cerradas, cupo entregado y características de estos clientes (segmento al que pertenecen).

Luego se consultó cuáles serían las variables o perspectivas desde las cuales se consultarán los indicadores anteriormente descritos.

Los resultados obtenidos fueron:

Ventas: Se desea conocer cuántos clientes de cada marca fueron adquiridos en un periodo determinado y el monto de crédito entregado (cupos).

Facturación: Se desea conocer la cantidad de clientes que compraron en un determinado comercio y el monto del consumo realizado esto para un periodo determinado.

Morosidad: Se desea conocer para cada tramo de morosidad la cantidad de clientes que pasan a mora en un periodo.

Se desea conocer para cada tramo de morosidad los montos en mora en un periodo determinado.

Tramo Mora

Segmento

4.1.5 Modelo Conceptual

En esta etapa se construirá un modelo conceptual tomando como base los indicadores y dimensiones obtenidas a partir del estudio.

El objetivo de este modelo es visualizar con claridad cuáles serán los alcances de este trabajo en los cuales profundizaremos, además de obtener un modelo que sea fácil de entender por los usuarios finales.

La representación gráfica que se utilizará para el modelo conceptual es la que se muestra en la figura 14:

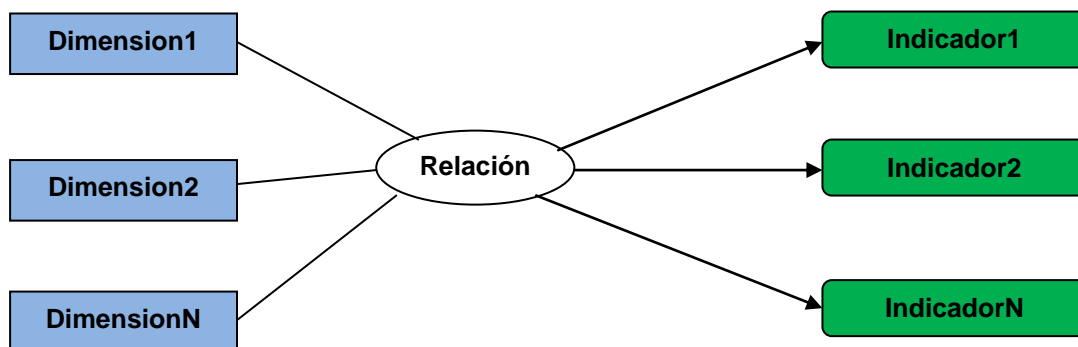


Figura 14, Representación gráfica de un modelo conceptual.

A la izquierda de la figura 14 se especifican las dimensiones que serán unidas a un punto central que representa la relación existente entre ellas, de dicha relación se desprenden los indicadores desplegados a la derecha de la figura.

En las figuras 15, 16, 17 y 18 se muestran los modelos conceptuales resultantes en base a los datos recopilados:

Ventas

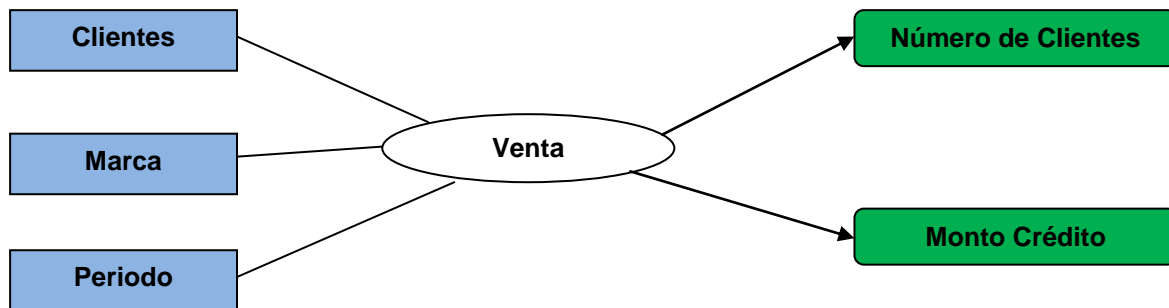


Figura 15, Representación gráfica del modelo conceptual Ventas.

Facturación

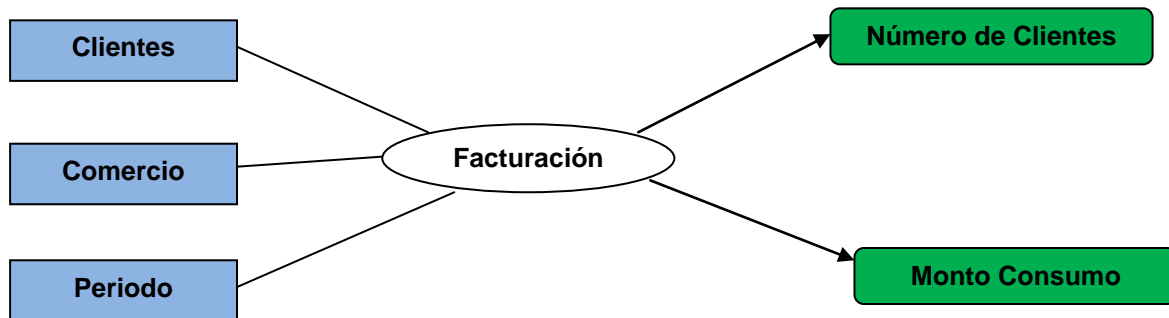


Figura 16, Representación gráfica del modelo conceptual Facturación.

Morosidad

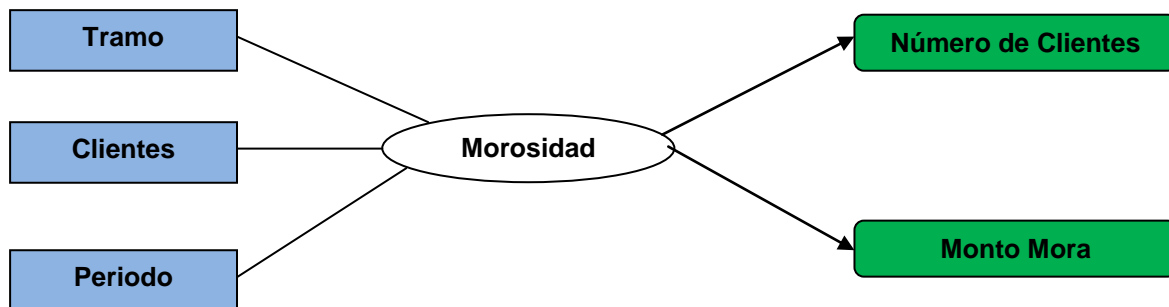


Figura 17, Representación gráfica del modelo conceptual Morosidad.

Cierres

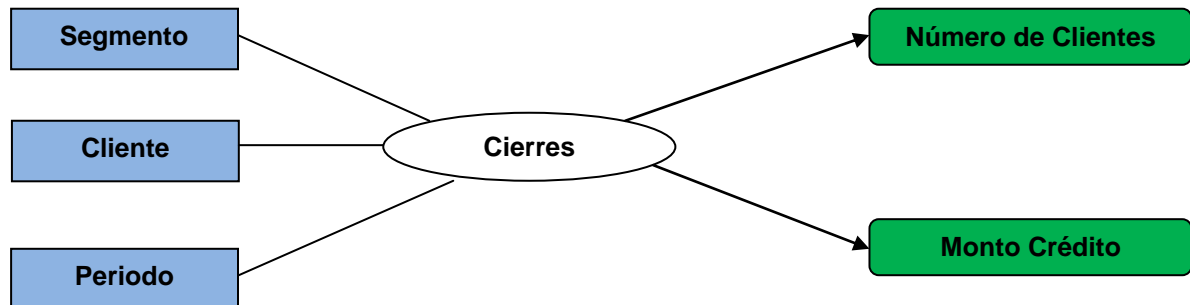


Figura 18, Representación gráfica del modelo conceptual Cierres.

4.1.6 Análisis de los datos existentes

El siguiente paso será analizar las fuentes de datos existentes, en este caso una base de datos relacional montada en un servidor SQL Server, esto nos permitirá determinar cómo serán calculados los indicadores y establecer las relaciones entre el modelo conceptual anteriormente descrito y las fuentes de datos, posteriormente definiremos que campos incluiremos en cada dimensión para finalmente ampliar un modelo conceptual con toda la información obtenida.

4.1.7 Calculo de Indicadores

En este punto explicaremos como se realizará el cálculo de los indicadores, definiremos los conceptos para cada uno de ellos:

Composición de hechos y su forma de cálculo: Hecho1 + Hecho2

Función de agregación a utilizar: suma, contar, promedio, etc.

Para este caso los indicadores se calcularan de la siguiente forma:

Ventas

Numero de Adquisiciones:

Hechos: Número de Clientes

Función de Sumarización: COUNT.

El indicador representa la cantidad de nuevos clientes.

Monto Crédito:

Hechos: Monto Crédito

Función de Sumarización: SUM.

El indicador representa la sumatoria del monto de crédito otorgado (cupo) de todas las tarjetas de crédito que se han vendido.

Facturación

Número de Clientes:

Hechos: Número de Clientes.

Función de Sumarización: COUNT.

El indicador representa la cantidad de clientes que realizaron consumos.

Monto Consumo:

Hechos: Monto Consumo

Función de Sumarización: SUM.

El indicador representa la sumatoria del monto de consumo utilizado de cada tarjeta de crédito.

Morosidad

Número de Clientes:

Hechos: Número de Clientes.

Función de Sumarización: COUNT

El indicador representa la cantidad de clientes que se encuentran en mora.

Monto en Mora:

Hechos: Monto en Mora

Función de Sumarización: SUM.

El indicador representa la sumatoria del monto en mora de cada tarjeta de crédito.

Cierres

Número de Clientes:

Hechos: Clientes que cerraron su Tarjetas.

Función de Sumarización: COUNT.

El indicador representa la cantidad de clientes que han cerrado su tarjeta de crédito.

Monto Crédito:

Hechos: Monto Crédito

Función de Sumarización: SUM.

El indicador representa la sumatoria del monto de crédito otorgado (cupo) de todas las tarjetas de crédito que se han cerrado.

4.1.8 Establecer correspondencias

En este punto se analizará que la base de datos relacional disponible contenga la información necesaria y las características de dicha información, de modo de poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos existentes. El escenario ideal es que todos los elementos del modelo conceptual tengan correspondencia en las fuentes de datos de la entidad.

Para el caso práctico de la entidad en estudio, los procesos elegidos: Ventas, Facturación, Morosidad y Cierres están representados por el diagrama de entidad relación descrito en la figura 19:

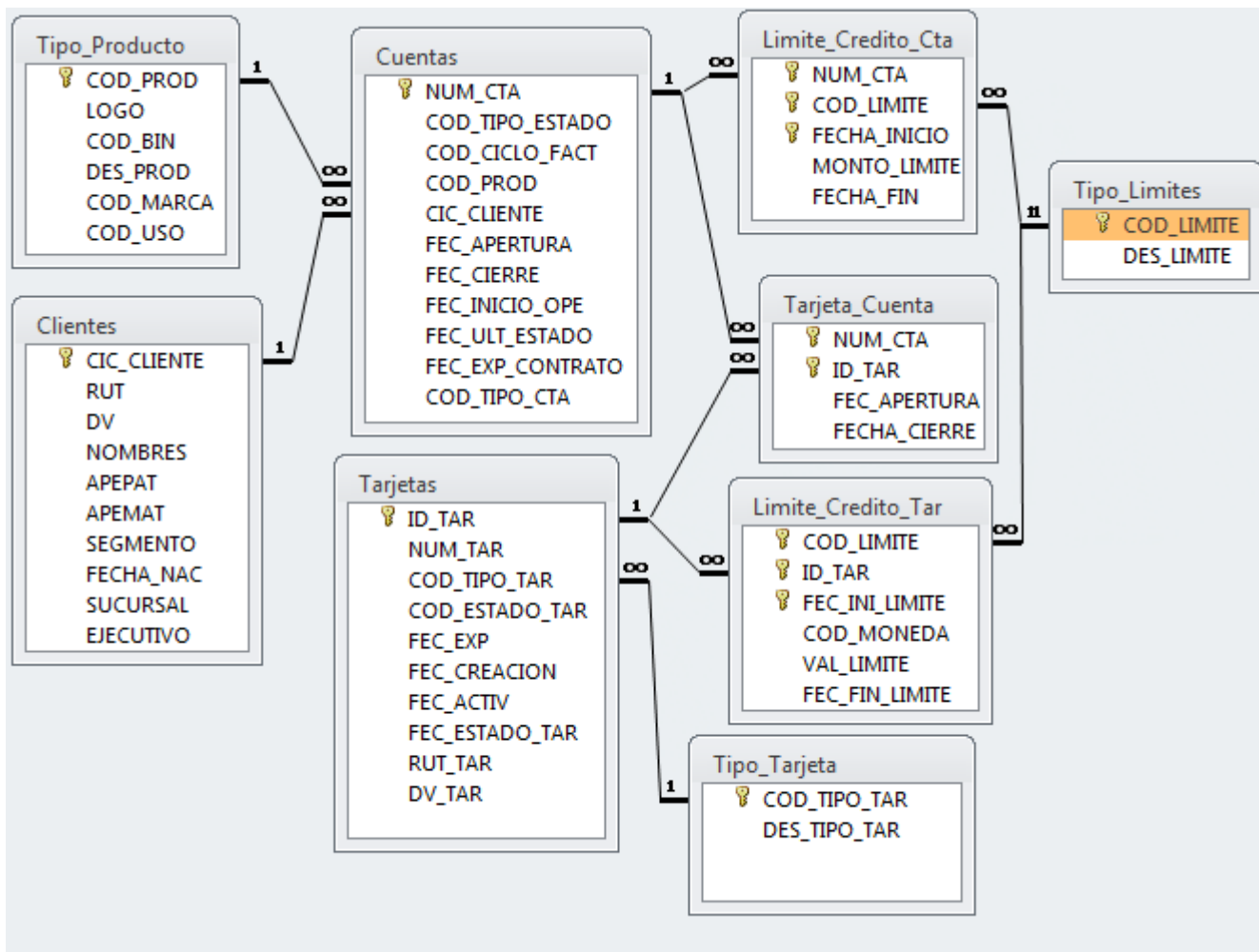


Figura 19, Modelo base de datos relacional tarjetas de crédito.

El siguiente paso a realizar será visualizar la correspondencia entre los dos modelos, tal como se muestra en la figura 20:

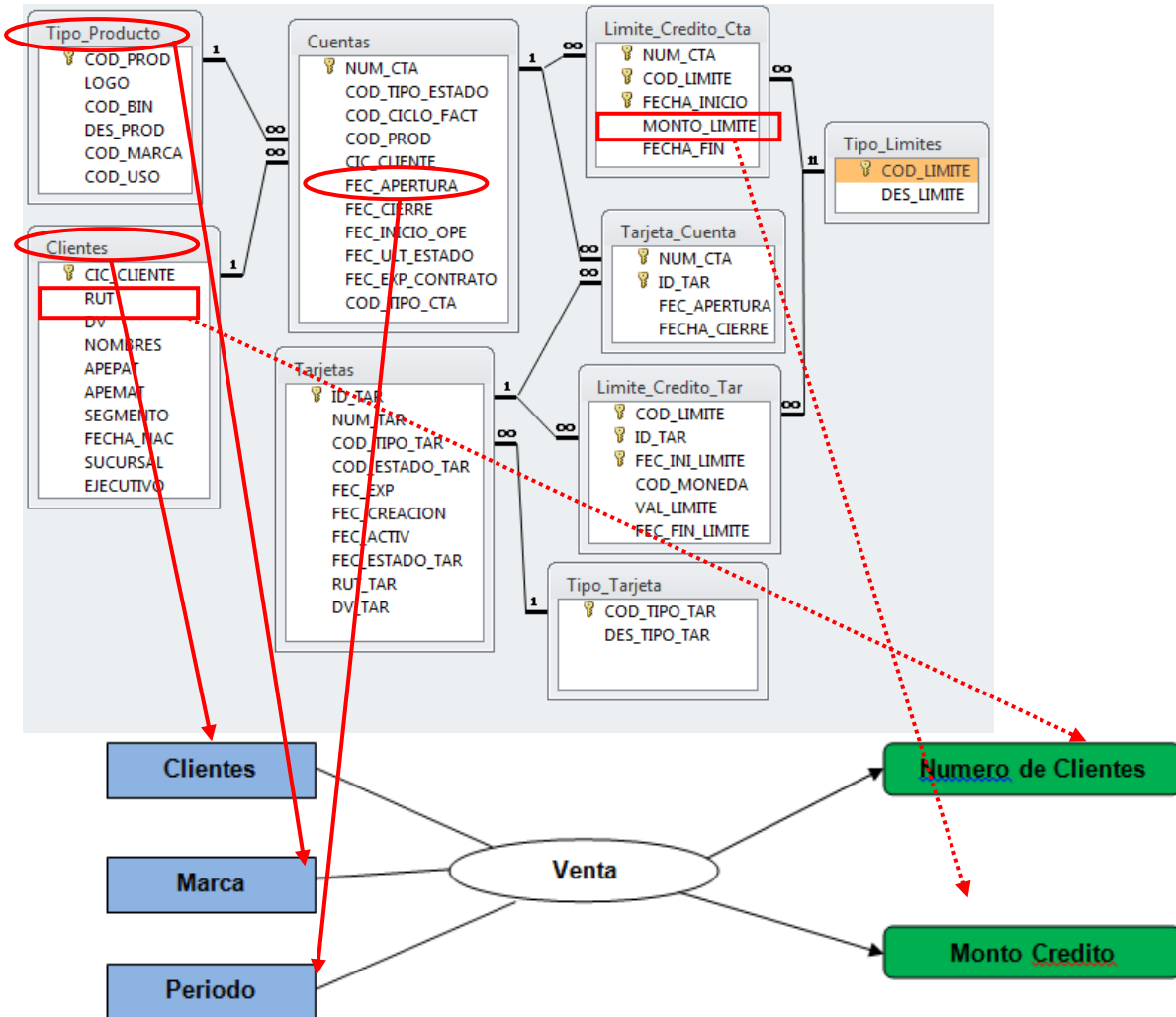


Figura 20, Correspondencia entre modelo relacional y conceptual.

A continuación se expondrán las relaciones encontradas:

- La tabla “Clientes” con la perspectiva “Clientes”.
- La tabla “Tipo_Producto” con la perspectiva “Marca” (esto debido a que dicha tabla contiene la descripción de cada tipo de tarjeta).
- El campo “FEC_APERTURA” de tabla “Cuentas” con la perspectiva “Tiempo” (esto debido a que dicha fecha representa la fecha principal en el proceso relacionado con la venta).

- El campo “RUT” de la tabla “Clientes” con el indicador “número de clientes”, puesto que el contador (count) de este campo indicara la cantidad de nuevos clientes.
- El campo “Monto Limite” de la tabla “Limite_Credito_Cta” con el indicador “Monto Crédito” puesto que este último contiene el cupo (límite de crédito) otorgado a cada Tarjeta de Crédito.

Diagrama Entidad Relación Facturación

En la figura 21 se muestra el diagrama de entidad relación correspondiente al proceso de facturación:

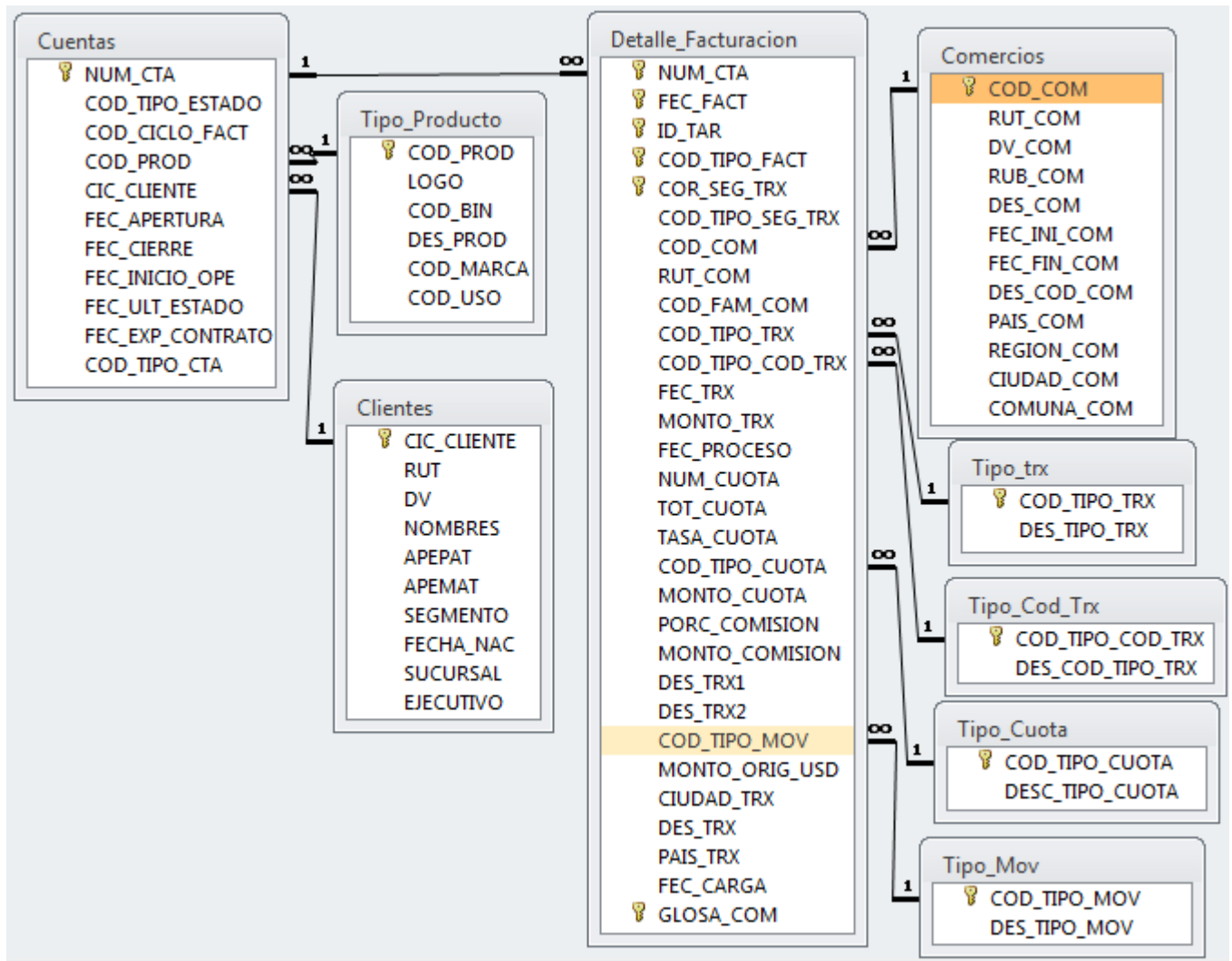


Figura 21, Diagrama entidad relación facturación tarjetas de crédito.

La figura 22 muestra la correspondencia que existe entre los dos modelos:

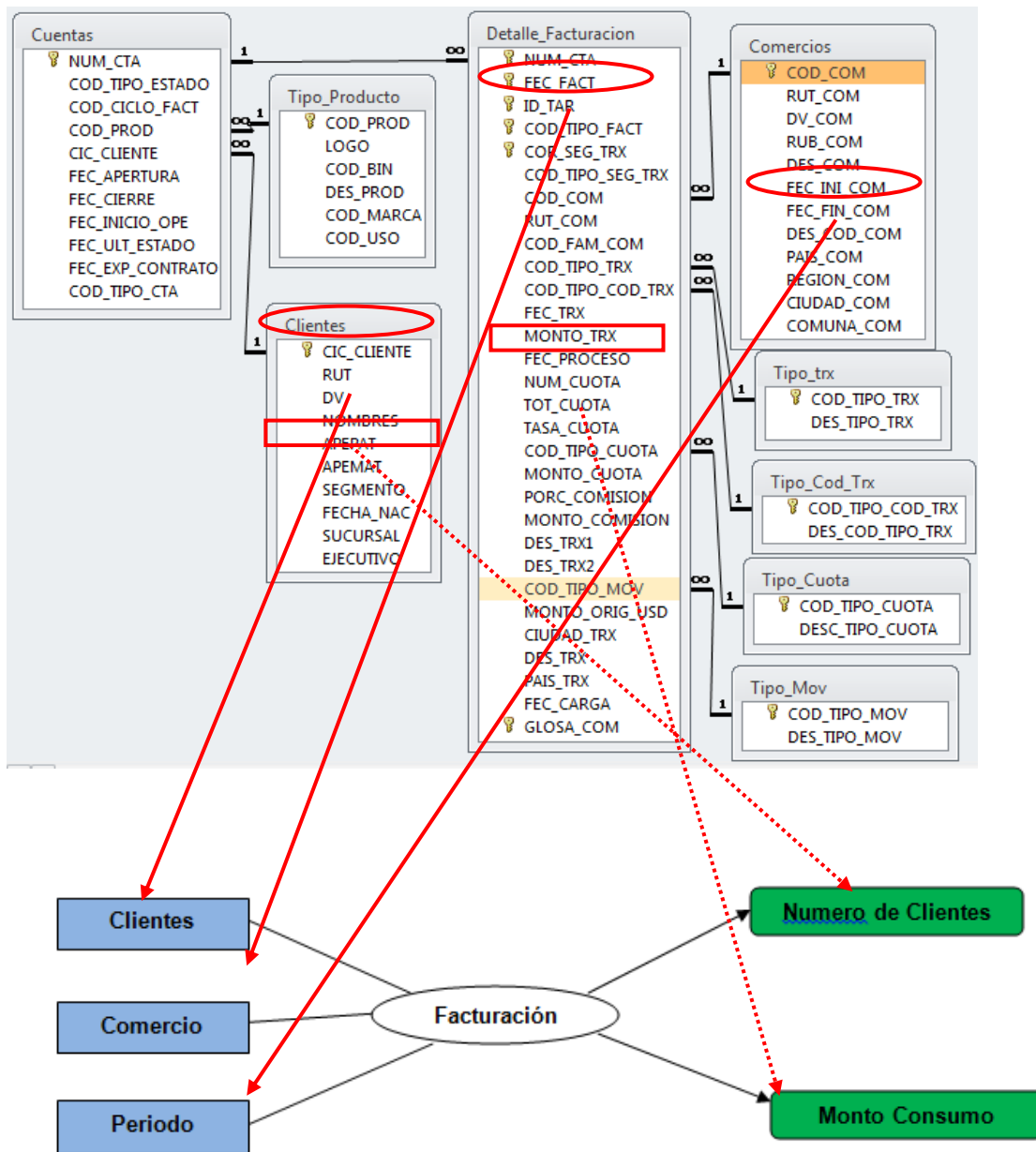


Figura 22, Correspondencia entre modelo relacional y conceptual de facturación.

Las relaciones encontradas son las siguientes:

- La tabla “Clientes” con la perspectiva “Clientes”.
- El campo “FEC_FACT” de tabla “Detalle_Facturación” con la perspectiva “Periodo” (esto debido a que dicha fecha representa la fecha principal en el proceso relacionado con la facturación)
- El campo “DES_COM” de la tabla “Comercios” con la perspectiva “Comercio”.
- El campo “MONTO_TRX” de la tabla “Detalle_Facturacion” con el indicador monto consumo, puesto que la suma (sum) de este campo indicara el monto de los consumos.
- El campo “RUT” de la tabla “Clientes” con el indicador “número de clientes”, puesto que el contador (count) de este campo indicara la cantidad de nuevos clientes.

Diagrama Entidad Relación Mora

La figura 23 representa gráficamente el diagrama entidad relación del proceso de mora:

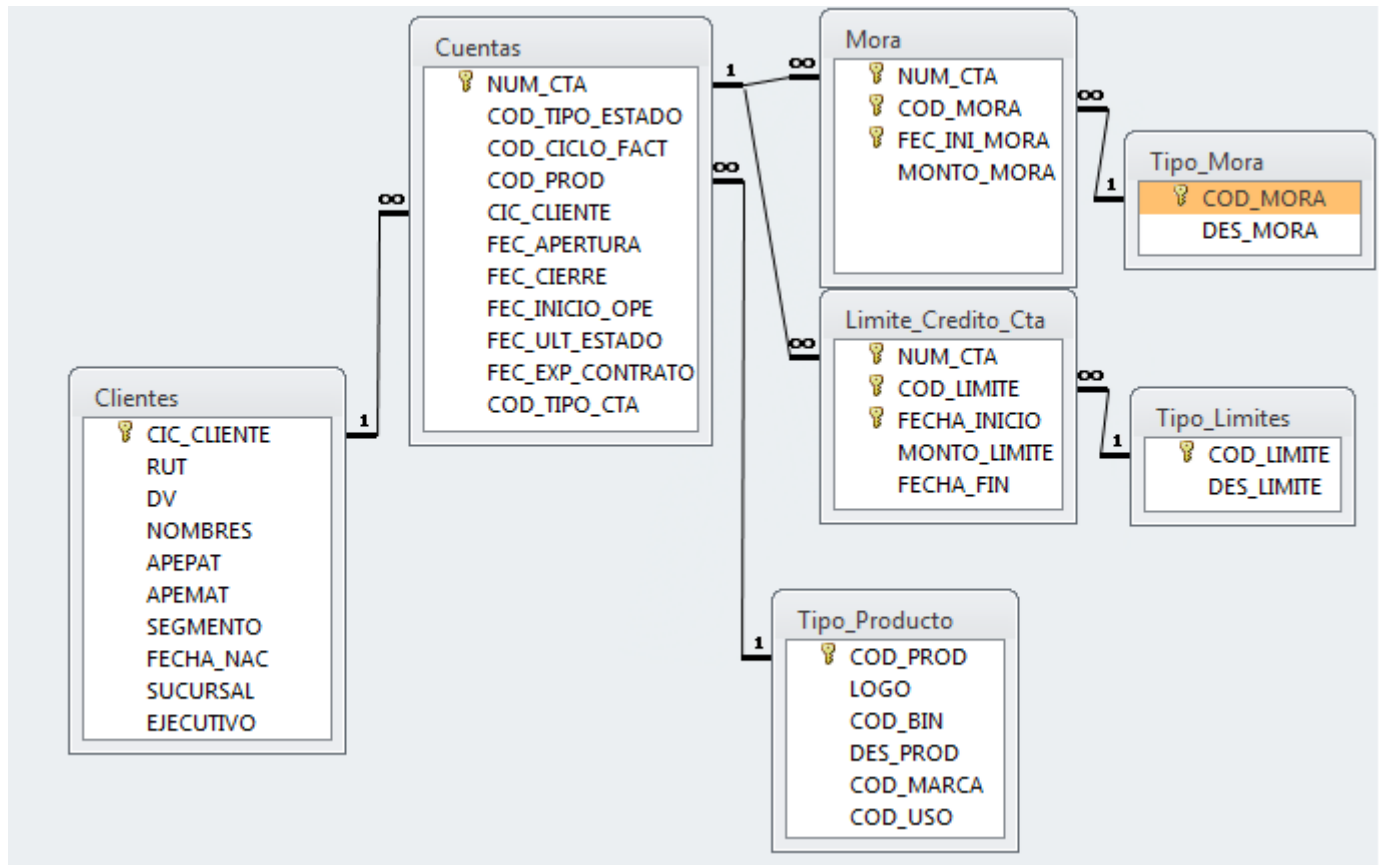


Figura 23, Diagrama entidad relación mora tarjetas de crédito.

En la figura 24 se muestra la correspondencia entre los dos modelos para el proceso de mora:

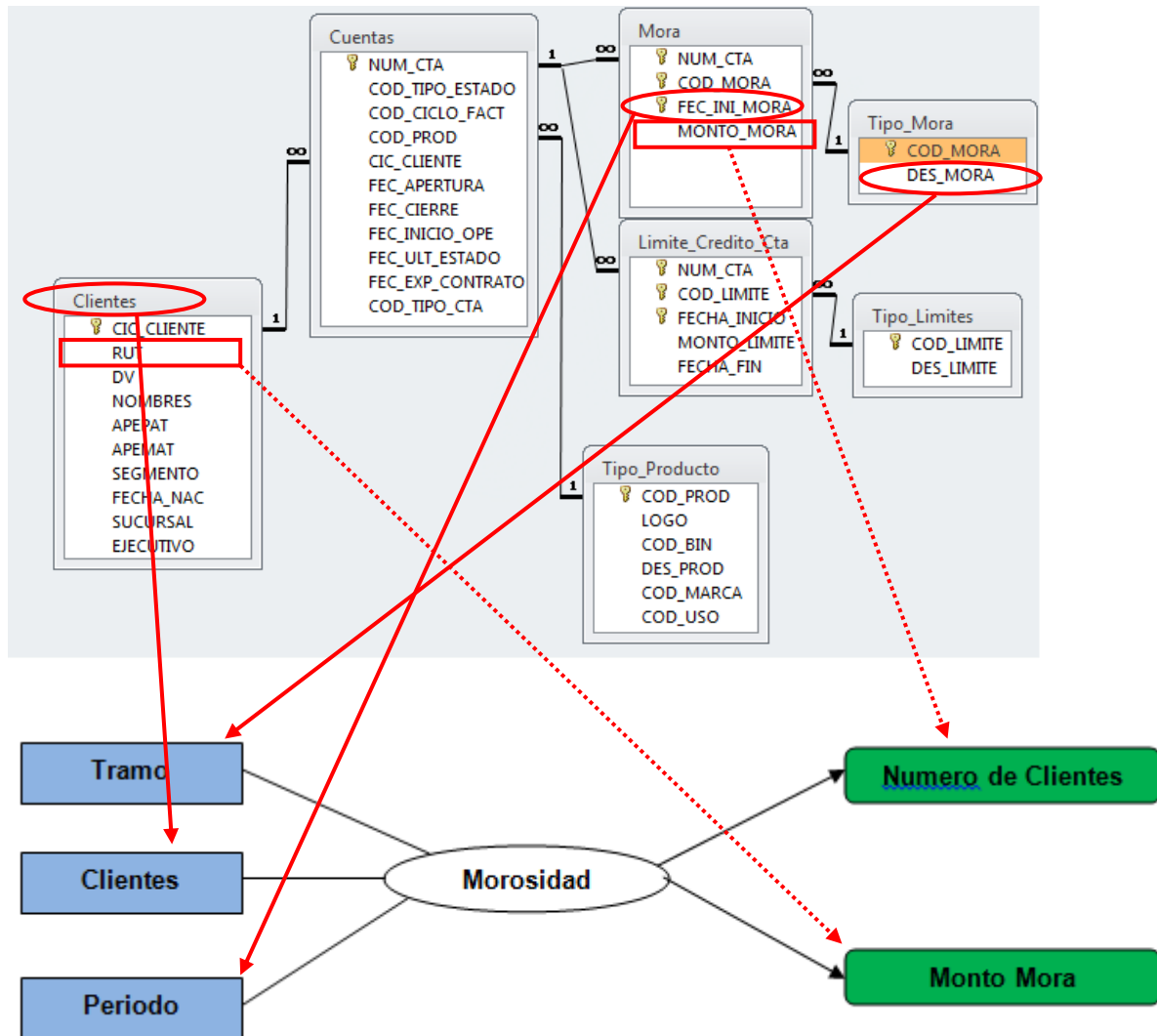


Figura 24, Correspondencia entre modelo relacional y conceptual de mora.

Las relaciones encontradas son las siguientes:

- La tabla “Clientes” con la perspectiva “Clientes”.
- El campo “DES_MORA” de tabla “Tipo_Mora” con la perspectiva “Tramo” (esto debido a que dicho campo representa el tramo de mora en el proceso relacionado con la mora)

- El campo “FEC_INI_MORA” de tabla “Mora” con la perspectiva “Periodo” (esto debido a que dicha fecha representa la fecha principal en el proceso relacionado con la mora).
- El campo “MONTO_MORA” de la tabla “Mora” con el indicador “monto mora”, puesto que la suma (sum) de este campo indicara el monto de la mora.
- El campo “RUT” de la tabla “Clientes” con el indicador “número de clientes”, puesto que el contador (count) de este campo indicara la cantidad de nuevos clientes.

Diagrama Entidad Relación Cierre de Tarjetas

La figura 25 representa gráficamente el diagrama entidad relación del proceso de cierre:

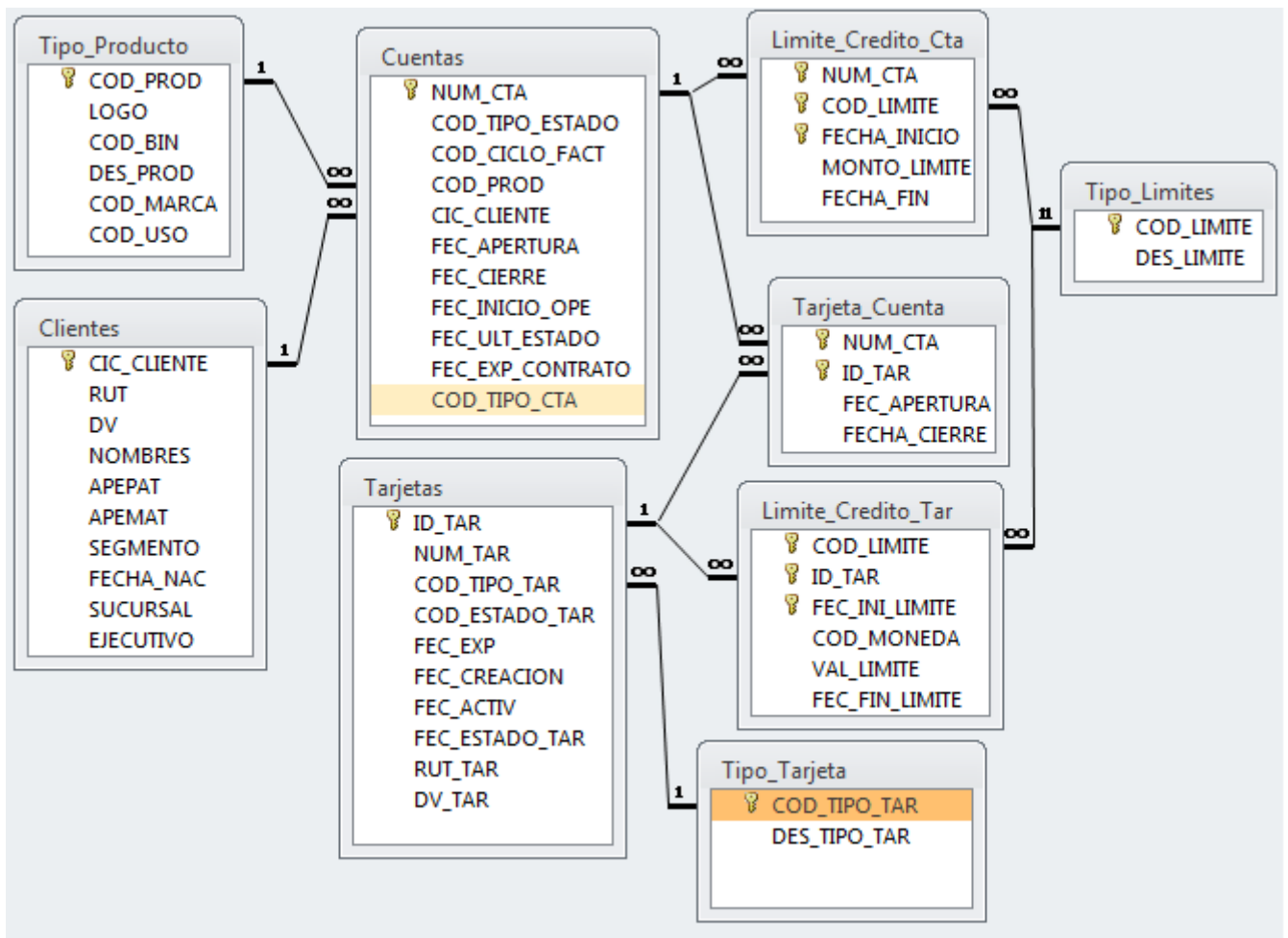


Figura 25, Diagrama entidad relación cierre tarjetas de crédito.

En la figura 26 se muestra la correspondencia entre los dos modelos para el proceso de cierre:

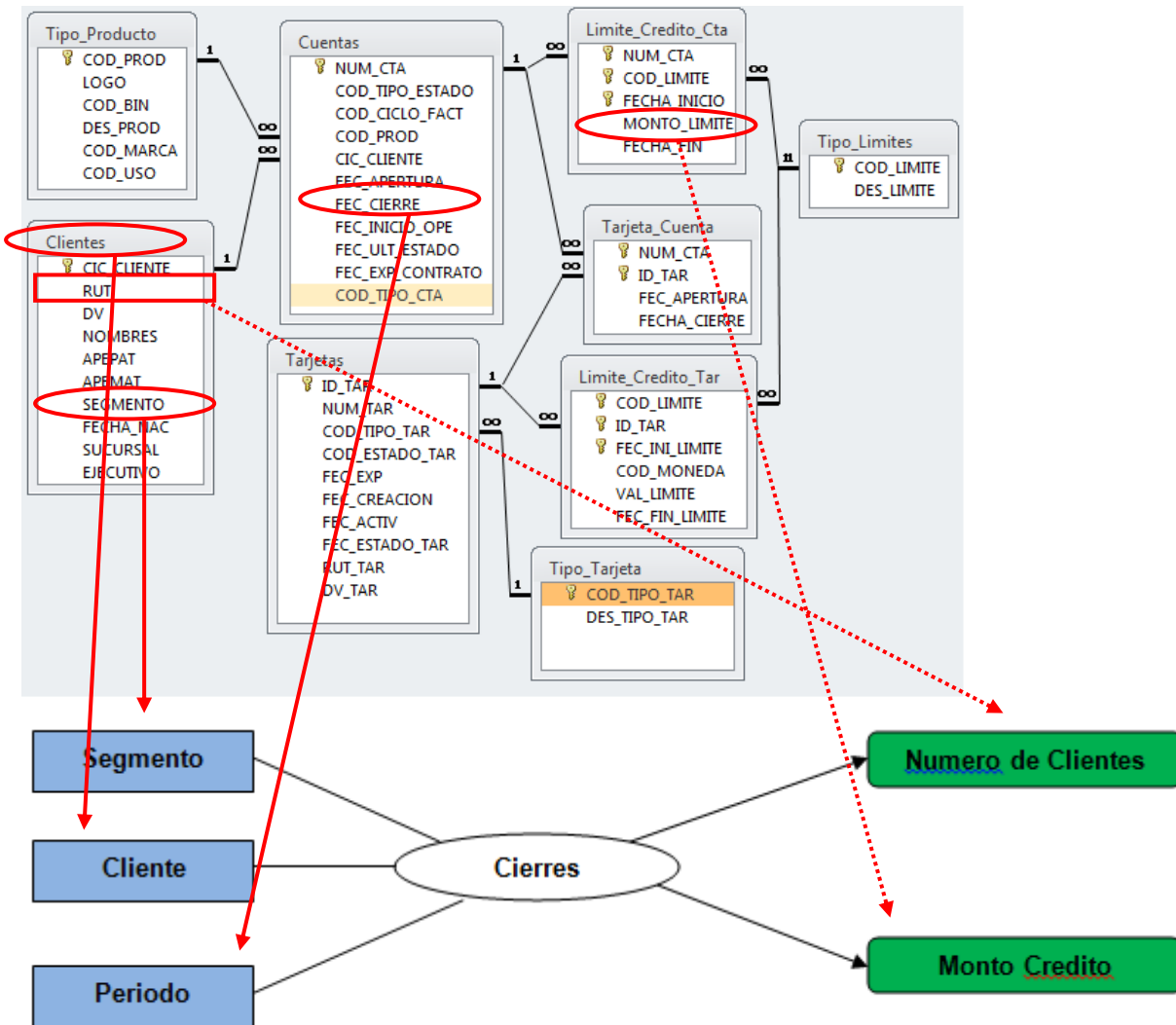


Figura 26, Correspondencia entre modelo relacional y conceptual de cierres.

Las relaciones encontradas son las siguientes:

- La tabla “Clientes” con la perspectiva “Clientes”.
- El campo “SEGMENTO” de la tabla “Clientes” con la perspectiva “Segmento”.

- El campo “FEC_CIERRE” de tabla “Cuentas” con la perspectiva “Periodo” (esto debido a que dicha fecha representa la fecha principal en el proceso relacionado con el cierre de tarjetas)
- El campo “RUT” de la tabla “Clientes” con el indicador “número de clientes”, puesto que el contador (count) de este campo indicará la cantidad de nuevos clientes.

4.1.9 Nivel de Granularidad

En el paso anterior quedaron establecidas claramente las relaciones con la base de datos relacional, como siguiente paso se analizarán y seleccionarán los campos que incluirá cada perspectiva, y por este medio se realizará la manipulación y los filtros de cada uno de los indicadores.

Tomando como base las relaciones establecidas en el paso anterior, se presentará a los usuarios los datos analizados para cada perspectiva. En este sentido es de vital importancia conocer el significado de cada campo y entender el contexto, pudiendo utilizar para este efecto la documentación de los sistemas, diccionario de datos u otra fuente de información existente.

Se expondrá a los usuarios los datos existentes, de forma de explicar claramente el significado y las características de cada campo, esto para que el usuario decida cuales son los más importantes para consultar los indicadores y cuales podemos excluir del modelo.

En relación a la perspectiva “Tiempo” será vital definir los criterios mediante los cuales se agruparan o sumarizaran los datos. En este sentido los campos a considerar serán: día, semana, mes, trimestre y año.

En este punto prestaremos especial atención, ya que esto será clave para determinar la granularidad de la información que almacenara nuestro Data Mart.

Al momento de establecer la correspondencia entre los dos modelos hemos podido visualizar los nombres de los campos, si bien es cierto su contenido puede ser deducido fácilmente, aclararemos el significado de cada uno de ellos:

Proceso Ventas

Perspectiva Clientes

- CIC_CLIENTE: Corresponde al código interno único de cada cliente de la tabla.
- RUT: Es la clave primaria de esta perspectiva e identifica en forma única a cada cliente.
- DV: Corresponde al dígito verificador del Rut.
- NOMBRES: Nombre del cliente.
- APEPAT : Apellido paterno del cliente.
- APEMAT: Apellido materno del cliente.
- SEGMENTO: Corresponde al código de segmento interno definido por la empresa y construido en base a los ingresos del cliente.
- FECHA_NAC: Fecha de nacimiento del cliente.
- SUCURSAL: Nombre de sucursal a la cual pertenece el cliente.
- EJECUTIVO: Nombre del ejecutivo de cuenta asignado al cliente.

Perspectiva Marca : Es la que indica el tipo de producto tarjeta

- COD_PROD: Es la clave primaria de esta perspectiva e identifica en forma única a cada tipo de producto.
- LOGO: Código numérico que indica la marca de cada tarjeta.
- COD_BIN: Corresponde a los 6 primeros dígitos de cada producto.
- DES_PROD: Indica la descripción del producto.
- COD_MARCA: Indica la marca de la tarjeta (Visa, MasterCard, etc.).
- COD_USO: Indica el tipo de uso que tendrá la tarjeta (Nacional, Internacional).

Perspectiva Tiempo : Es el que indica la granularidad del Data Mart y los datos a utilizar serán:

- Día.
- Semana.
- Mes.
- Trimestre.

- Año.

Al terminar la recolección de la información para el proceso de ventas, se consultó con los usuarios cuales serían los indicadores de interés para su gestión y los resultados fueron los siguientes:

Respecto a la perspectiva cliente se considerará el nombre completo del cliente, es decir: APEPAT + APEMAT + NOMBRES.

En la perspectiva marca serán fundamentales los campos que indican la descripción del producto (“DES_PROD”), su marca (“COD_MARCA”) y su tipo de uso (“COD_USO”).

En lo que se refiere a la perspectiva tiempo, serán de vital importancia los campos mes, trimestre y año.

Proceso Facturación

Perspectiva Clientes

- CIC_CLIENTE: Corresponde al código interno único de cada cliente de la tabla.
- RUT: Es la clave primaria de esta perspectiva e identifica en forma única a cada cliente.
- DV: Corresponde al dígito verificador del Rut.
- NOMBRES: Nombre del cliente.
- APEPAT: Apellido paterno del cliente.
- APEMAT: Apellido materno del cliente.
- SEGMENTO: Corresponde al código de segmento interno definido por la empresa y construido en base a los ingresos del cliente.
- FECHA_NAC: Fecha de nacimiento del cliente.
- SUCURSAL: Nombre de sucursal a la cual pertenece el cliente.
- EJECUTIVO: Nombre del ejecutivo de cuenta asignado al cliente.

Perspectiva Comercio: Es la que indica el comercio en el cual fue realizada la compra.

- COD_COM: Corresponde a los 6 primeros dígitos de cada producto.

- RUT_COM: Es la clave primaria de esta perspectiva e identifica en forma única a cada comercio.
- DV_COM: Dígito verificador correspondiente al Rut del comercio.
- RUBRO_COM: Corresponde a la categorización del comercio "Rubro".
- DES_COM: Indica la descripción del comercio (nombre de fantasía).
- FEC_INI_COM: Fecha desde la cual funciona el comercio.
- FEC_FIN_COM: Fecha de término de funcionamiento del comercio.
- DES_COD_COM: Campo no se está utilizando.
- PAIS_COM: Indica el país donde se encuentra ubicado el comercio.
- REGION_COM: Indica la región del país donde se encuentra ubicado el comercio.
- CIUDAD_COM: Indica la ciudad de la región donde se encuentra ubicado el comercio.
- COMUNA_COM: Indica la comuna de la ciudad donde se encuentra ubicado el comercio.

Perspectiva Periodo : Es el que indica la granularidad del Data Mart y los datos a utilizar serán:

- Día.
- Semana.
- Mes.
- Trimestre.
- Año.

Al terminar la recolección de la información para el proceso de facturación, se consultó con los usuarios cuales serían los indicadores de interés para su gestión y los resultados fueron los siguientes:

Respecto a la perspectiva cliente se considerará el nombre completo del cliente, es decir: APEPAT + APEMAT + NOMBRES. En la perspectiva comercio será fundamental el campo que indica el nombre del comercio ("DES_COM").

En lo que se refiere a la perspectiva tiempo, serán de vital importancia los campos mes, trimestre y año.

Proceso Mora

Perspectiva Tramo: Indica el tramo de mora en el que se encuentra el producto.

- COD_MORA: Es la clave primaria de esta perspectiva e identifica en forma única a cada tramo de mora.
- DES_MORA: Indica la descripción del tramo de mora, este campo especifica los días de mora ej. : 1-14 días, 14-30 días, etc.

Perspectiva Periodo : Es el que indica la granularidad del Data Mart y los datos a utilizar serán:

- Día.
- Semana.
- Mes.
- Trimestre.
- Año.

Al terminar la recolección de la información para el proceso de mora, se consultó con los usuarios cuales serían los indicadores de interés para su gestión y los resultados fueron los siguientes:

En la perspectiva tramo será fundamental el campo que indica la descripción del tramo de mora (“DES_MORA”).

En lo que se refiere a la perspectiva tiempo, serán de vital importancia los campos mes, trimestre y año.

Proceso Cierres

Perspectiva Clientes

- CIC_CLIENTE: Corresponde al código interno único de cada cliente de la tabla.
- RUT: Es la clave primaria de esta perspectiva e identifica en forma única a cada cliente.

- DV: Corresponde al dígito verificador del Rut.
- NOMBRES: Nombre del cliente.
- APEPAT: Apellido paterno del cliente.
- APEMAT: Apellido materno del cliente.
- SEGMENTO: Corresponde al código de segmento interno definido por la empresa y construido en base a los ingresos del cliente.
- FECHA_NAC: Fecha de nacimiento del cliente.
- SUCURSAL: Nombre de sucursal a la cual pertenece el cliente.
- EJECUTIVO: Nombre del ejecutivo de cuenta asignado al cliente.

Perspectiva Periodo : Es el que indica la granularidad del Data Mart y los datos a utilizar serán:

- Día.
- Semana.
- Mes.
- Trimestre.
- Año.

Al terminar la recolección de la información para el proceso de cierre, se consultó con los usuarios cuales serían los indicadores de interés para su gestión y los resultados fueron los siguientes:

Respecto a la perspectiva cliente se considerará el tipo de segmento al cual pertenece el cliente.

En lo que se refiere a la perspectiva tiempo, serán de vital importancia los campos mes, trimestre y año.

4.1.10 Modelo Conceptual Ampliado

Finalmente se graficarán los resultados obtenidos, ampliando cada modelo conceptual e indicando bajo cada perspectiva los campos elegidos y bajo cada indicador la forma de cálculo, tal como se indica en la figura 27:

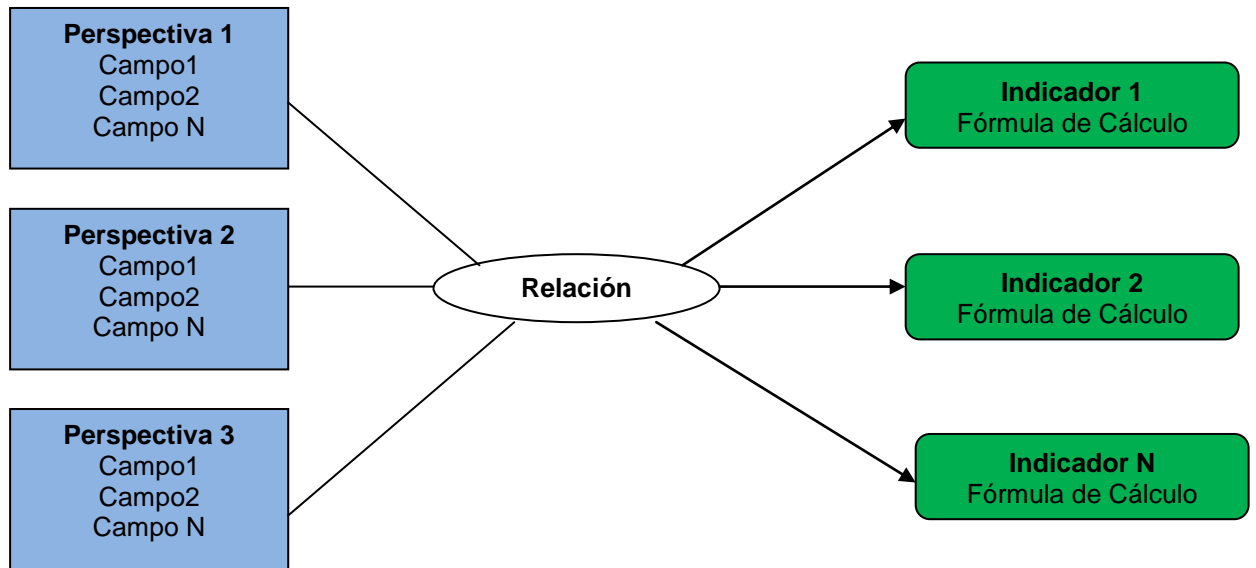


Figura 27, Representación gráfica de un modelo conceptual ampliado.

Tomando como base este gráfico, completaremos el diseño del diagrama conceptual para cada proceso. La figura 28 muestra el diagrama conceptual del proceso de ventas de tarjeta de crédito:

Proceso Ventas

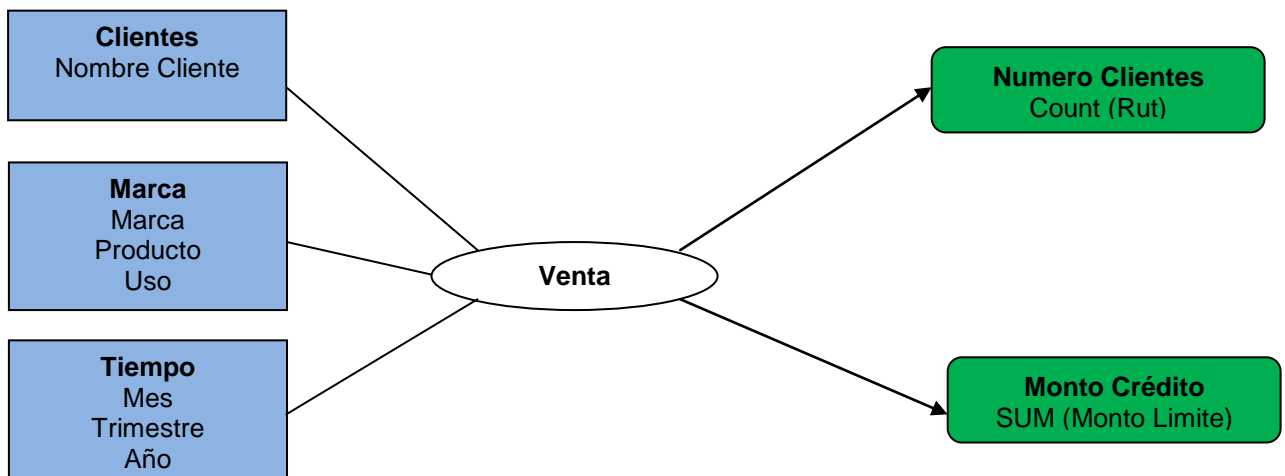


Figura 28, Representación gráfica modelo conceptual ampliado ventas.

Proceso Facturación

La figura 29 representa el diagrama conceptual del proceso de facturación de tarjetas de crédito:

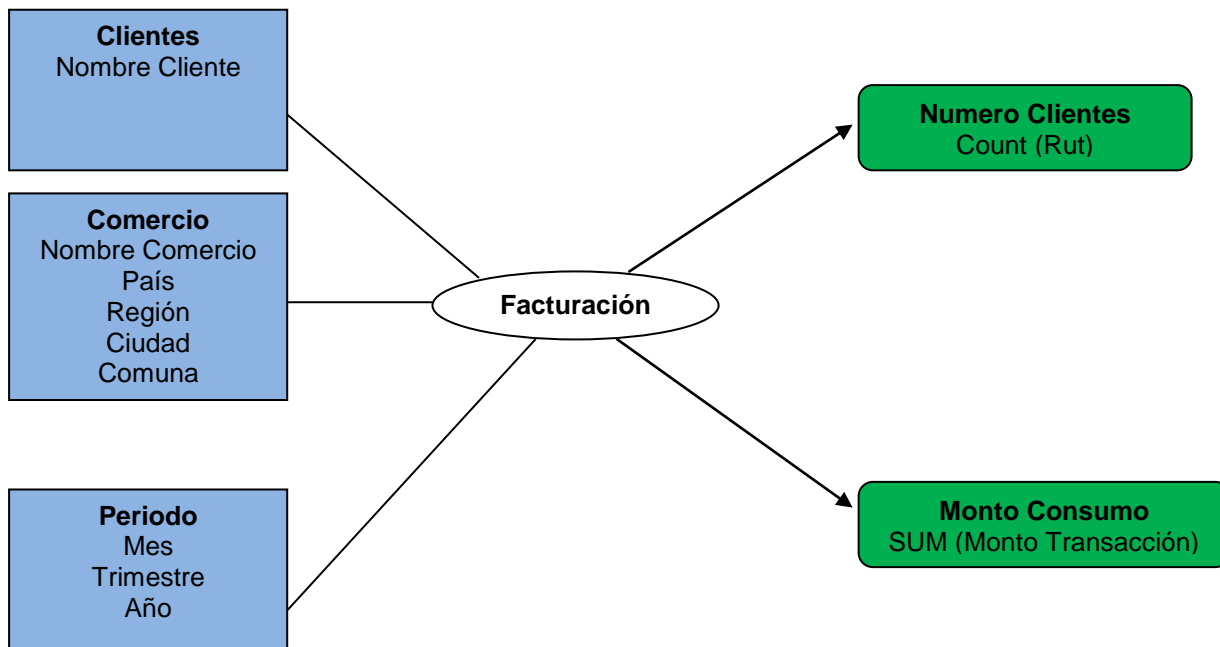


Figura 29, Representación gráfica modelo conceptual ampliado facturación.

Proceso Mora

La figura 30 representa el diagrama conceptual del proceso de mora de tarjetas de crédito:

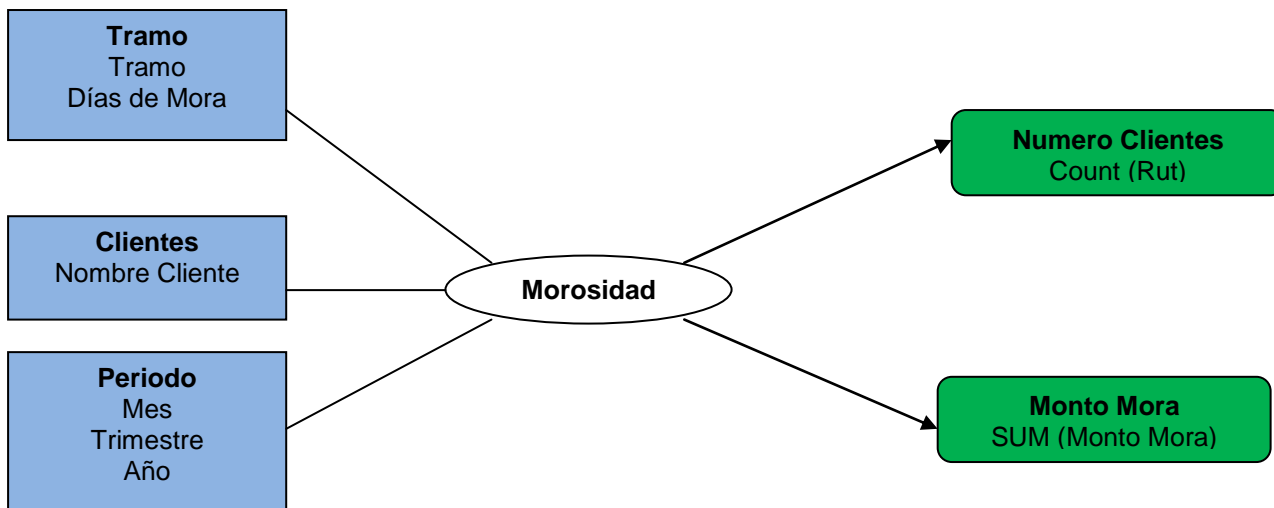


Figura 30, Representación gráfica modelo conceptual ampliado mora.

Proceso Cierres

La figura 31 representa el diagrama conceptual del proceso de cierre de tarjetas de crédito:

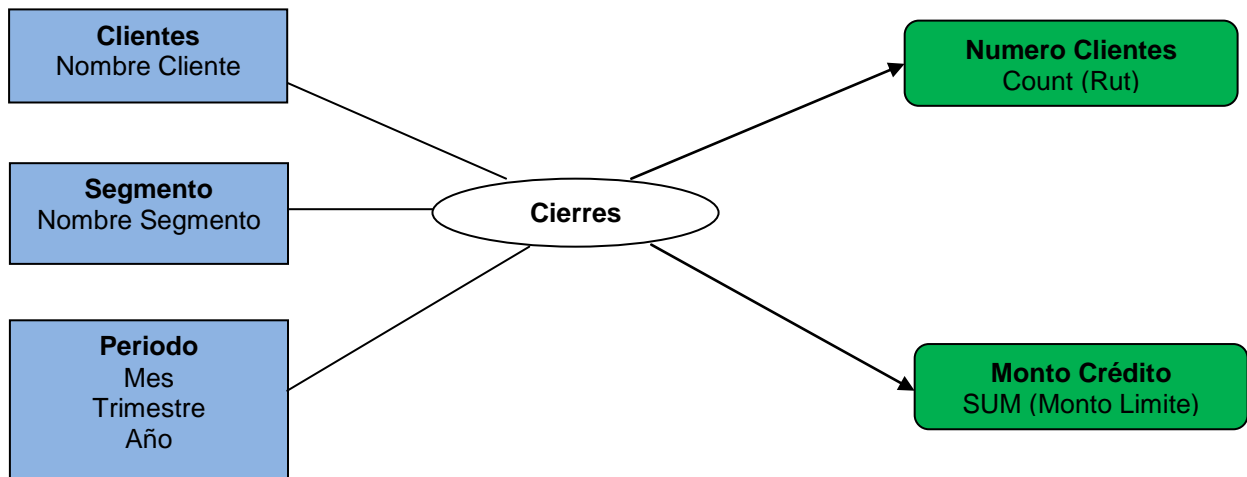


Figura 31, Representación gráfica modelo conceptual ampliado cierres.

4.1.11 Modelo Lógico del Data mart

Como siguiente paso elaboraremos el modelo lógico de la estructura del Data Mart, considerando como base el modelo conceptual previamente diseñado. Para lograr este objetivo será primordial elegir el tipo de modelo que se utilizará, para posteriormente aplicar este al caso en estudio y poder diseñar las tablas de dimensiones y hechos, para finalizar se definirán las uniones necesarias entre las tablas.

4.1.12 Tipo de Modelo Lógico del Data mart

Se seleccionará el tipo de esquema a utilizar, que es el que contendrá la estructura que soportará los datos, adaptándolo a las necesidades del usuario. El esquema seleccionado para la elaboración del modelo es el esquema estrella, esto debido a sus características las cuales se adaptan al caso en estudio además de poseer claras ventajas sobre otros modelos.

4.1.13 Tablas de Dimensiones

En este punto se procederá a diseñar las tablas de dimensiones que conformarán el Data Mart. Para realizar este proceso se tomará cada perspectiva con sus campos relacionados, asignando un nombre único para cada dimensión, se añadirá una clave principal y finalmente se darán nombres a los campos de tal forma que representen fielmente el dato contenido.

La figura 32 representa gráficamente la construcción de una tabla de dimensión en base a una perspectiva:

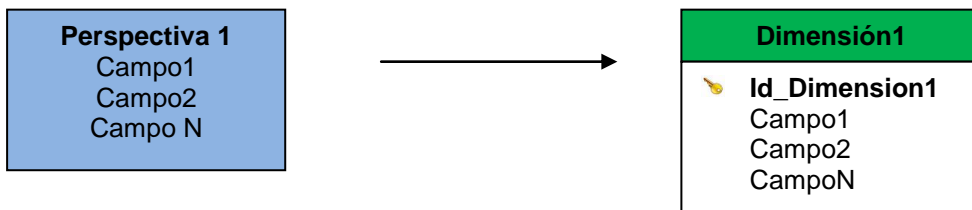


Figura 32, Representación gráfica de una tabla de dimensiones.

Para los esquemas copo de nieve, en caso de encontrar jerarquías en una tabla de dimensión, necesariamente esta debe ser normalizada. Para el caso práctico de la dimensión comercio será importante conocer la ubicación del comercio, para esto se utilizará la tabla de dimensión comercio y se tomará como referencia las relaciones padre hijo entre sus campos, según lo muestra la figura 33:

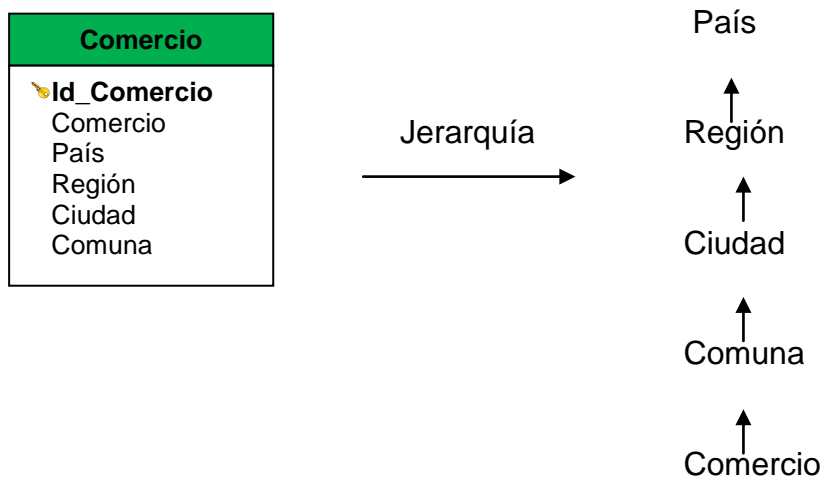


Figura 33, Desglose jerarquía comercio.

Al momento de normalizar la tabla anterior se obtendrán 4 tablas, las cuales se describen en la figura 34:

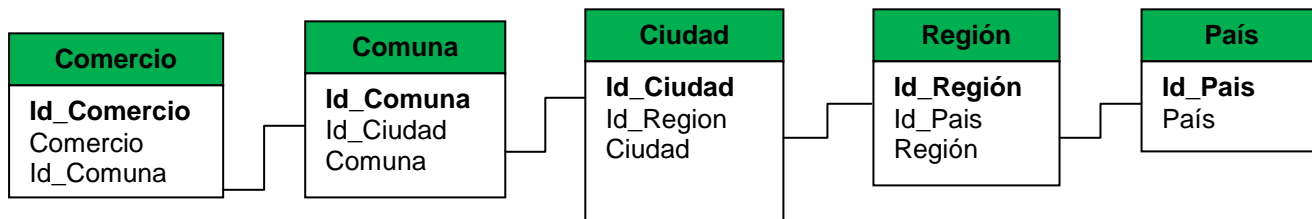


Figura 34, Tablas normalizadas jerarquía comercio.

Dado el caso práctico anterior, se procederá al diseño de las tablas de dimensiones.

Perspectiva Clientes

- La tabla de dimensión se llamará “CLIENTE”.
- Se asignará una clave principal llamada “Id_Cliente”.
- Se modificara el campo “Nombre_Cliente” por “Cliente”.

En la figura 35 se puede apreciar el resultado de esta operación:

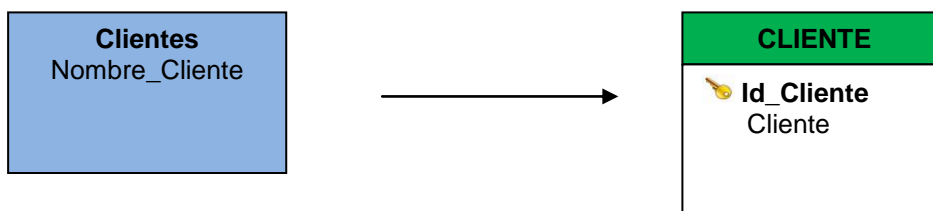


Figura 35, Tabla de dimensión Cliente.

Perspectiva Marca

- La tabla de dimensión se llamará “MARCA”.
- Se asignará una clave principal llamada “Id_Marca”.

El resultado de esta operación se muestra en la figura 36:

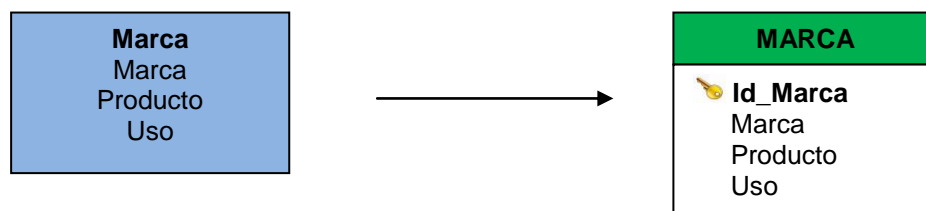


Figura 36, Tabla de dimensión Marca.

Perspectiva Tiempo

- La tabla de dimensión se llamará “PERIODO”.
- Se asignará una clave principal llamada “Id_Periodo”.
- Se mantendrán los nombres de campo.

El resultado de esta operación se muestra en la figura 37:

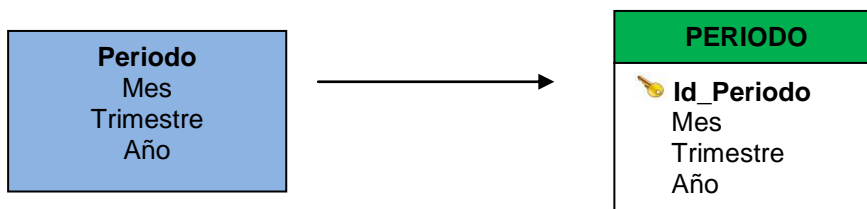


Figura 37, Tabla de dimensión Periodo.

Perspectiva Comercio

- La tabla de dimensión se llamará “COMERCIO”.
- Se asignará una clave principal llamada “Id_Comercio”.
- Se mantendrán los nombres de campo.

El resultado de esta operación se muestra en la figura 38:

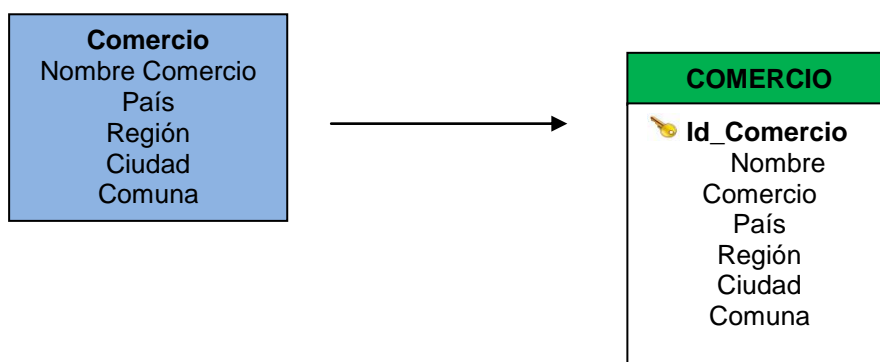


Figura 38, Tabla de dimensión Comercio.

Perspectiva Tramo

- La tabla de dimensión se llamará “TRAMO”.
- Se asignará una clave principal llamada “Id_Tramo”.
- Se mantendrán los nombres de campo.

El resultado de esta operación se muestra en la tabla 39:

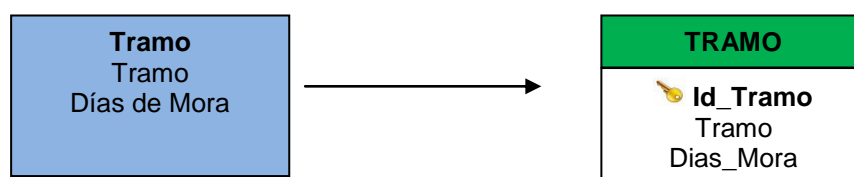


Figura 39, Tabla de dimensión Tramo.

Perspectiva Segmento

- La tabla de dimensión se llamará “SEGMENTO”.
- Se asignará una clave principal llamada “Id_Segmento”.
- Se reemplazará el nombre de campo “Nombre Segmento” por “Segmento”.

El resultado de esta operación se muestra en la figura 40:

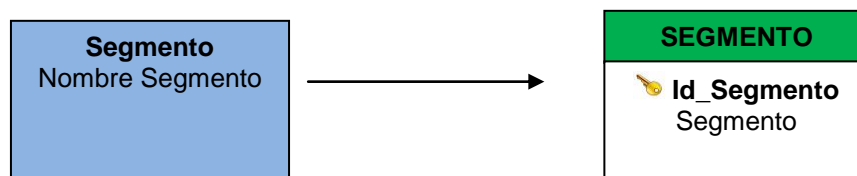


Figura 40, Tabla de dimensión Segmento.

4.1.14 Tablas de Hechos

En este punto se dará inicio a la definición de las tablas de hechos, que permitirán construir los indicadores para el Data Mart.

Los pasos a realizar para la construcción de las tablas de hechos de los esquemas estrella y copo de nieve serán los que se describen a continuación:

- Como primer paso se debe indicar un nombre para la tabla de hechos y que represente la información del negocio.
- Definir una clave primaria y que estará compuesta por las claves primarias de cada una de las tablas de dimensión definidas anteriormente.
- Finalmente se crearán los campos de hechos que serán equivalentes a los indicadores previamente definidos en el modelo conceptual, asignando los mismos nombres.

- Cada esquema resultante quedará de la forma representada en la figura 41:

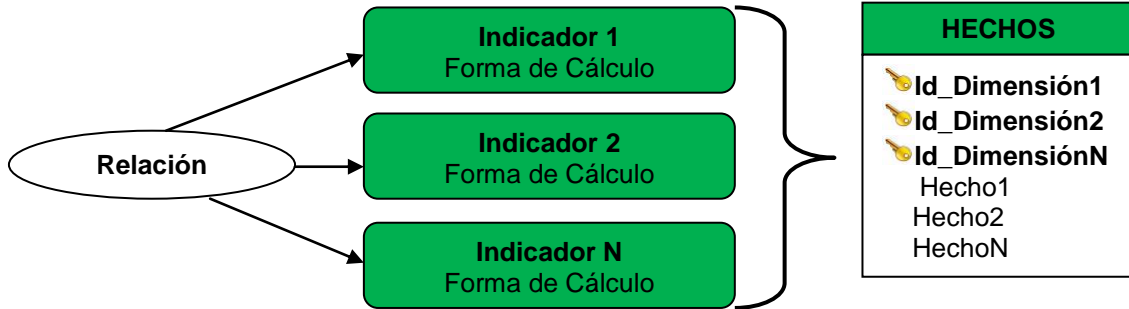


Figura 41, Representación gráfica de una tabla de hechos.

Para el esquema en forma de constelación los pasos a realizar serán los siguientes:

- Las tablas de hechos se confeccionarán considerando como base el análisis de las preguntas realizadas por los usuarios y sus equivalentes indicadores y perspectivas.
- Se asignará un nombre a cada tabla de hecho, además esta contendrá los hechos correspondientes y la clave, la cual estará conformada por la combinación de las claves de las tablas de dimensiones.
- En el diseño de las tablas de hechos, se debe considerar:
 - Si en dos o más preguntas aparecen los mismos indicadores pero con distintas perspectivas de análisis, se crearan tantas tablas de hechos como preguntas cumplan esta condición. Ejemplo:

“Analizar *Indicador1* por *Perspectiva1* y por *Perspectiva2*”

“Analizar *Indicador1* por *Perspectiva2* y por *Perspectiva3*”

Gráficamente se obtendrá lo mostrado en la figura 42:

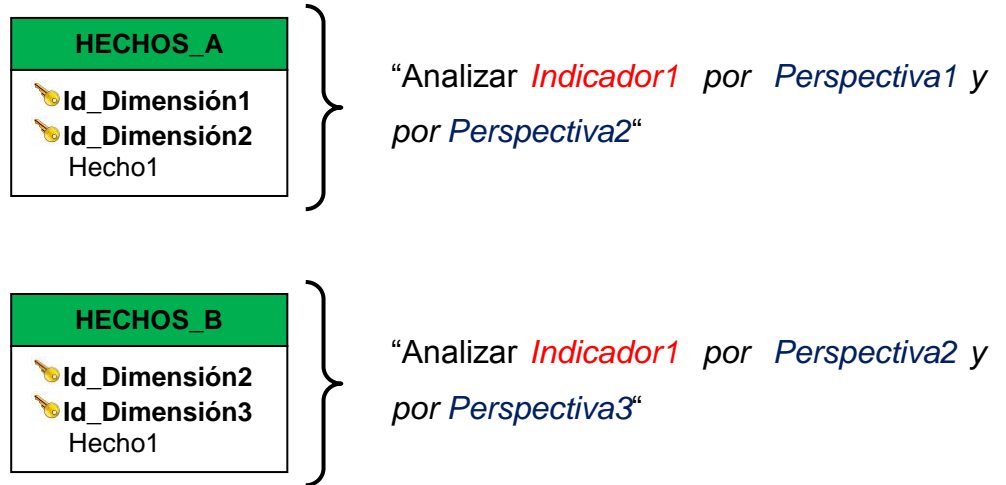


Figura 42, Representación gráfica de una tabla de hechos en esquema constelación.

- Si en dos o más preguntas aparecen diferentes indicadores con distintas perspectivas de análisis, se crearán tantas tablas de hechos como preguntas cumplan esta condición. Ejemplo:

“Analizar *Indicador1* por *Perspectiva1* y por *Perspectiva2*”

“Analizar *Indicador2* por *Perspectiva2* y por *Perspectiva3*”

Gráficamente se obtendrá lo mostrado en la figura 43:

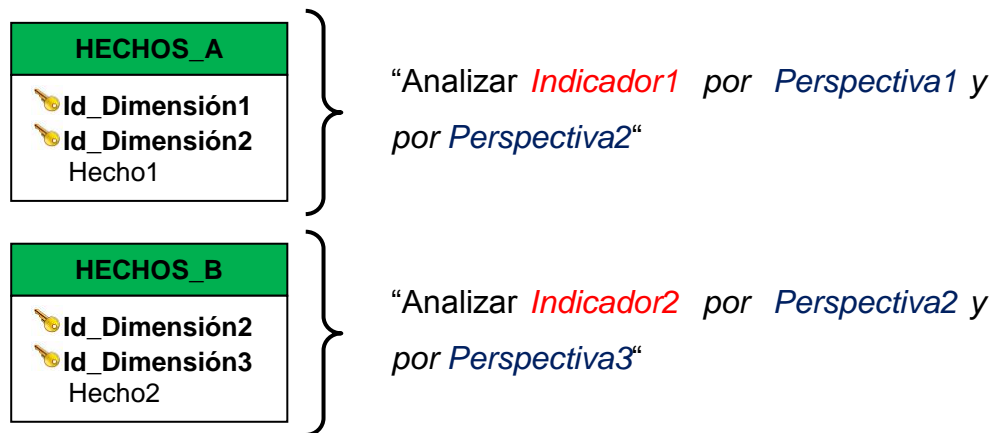


Figura 43, Representación gráfica de distintas tablas de hechos.

Para el caso de que las preguntas cumplan con las condiciones de los dos puntos anteriormente descritos, se procederá a unificar las preguntas que posean diferentes indicadores pero iguales perspectivas, para posteriormente volver a realizar el análisis de las interrogantes. Ejemplo:

“Analizar el *Indicador1* por *Perspectiva1* y por *Perspectiva2*”

“Analizar el *Indicador2* por *Perspectiva1* y por *Perspectiva2*”

En forma unificada quedará de la siguiente forma:

“Analizar el *Indicador1* y el *Indicador2* por *Perspectiva1* y por *Perspectiva2*”

4.1.15 Construcción de Tablas de Hechos

Proceso Ventas

- El nombre de la tabla de hechos será “VENTAS”.
- Se definirá la clave principal como la combinación de las claves de las tablas de dimensiones definidas anteriormente: “Id_Cliente”, “Id_Marca” y “Id_Periodo”.
- Finalmente se crearán dos hechos que equivalen a los 2 indicadores, para el caso “Unidades Vendidas” se cambiará el nombre por “Cantidad”.

La figura 44 representa gráficamente el resultado de esta operación:

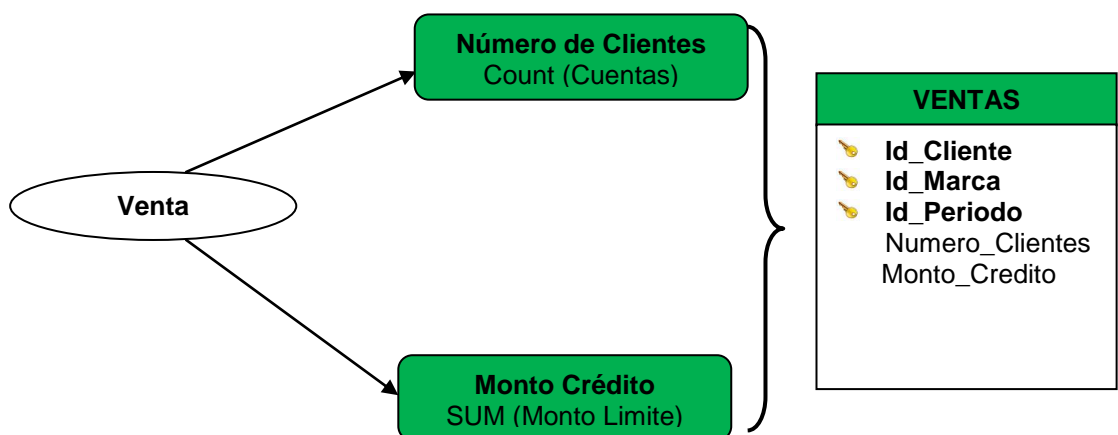


Figura 44, Representación gráfica tabla de hechos Ventas.

Proceso Facturación

- El nombre de la tabla de hechos será “FACTURACION”.
- Se definirá la clave principal como la combinación de las claves de las tablas de dimensiones definidas anteriormente: “Id_Cliente”, “Id_Comercio”, “Id_Periodo”
- Finalmente se crearan 2 hechos que equivalen a los 2 indicadores definidos para este proceso.

La figura 45 representa el resultado del proceso descrito anteriormente:

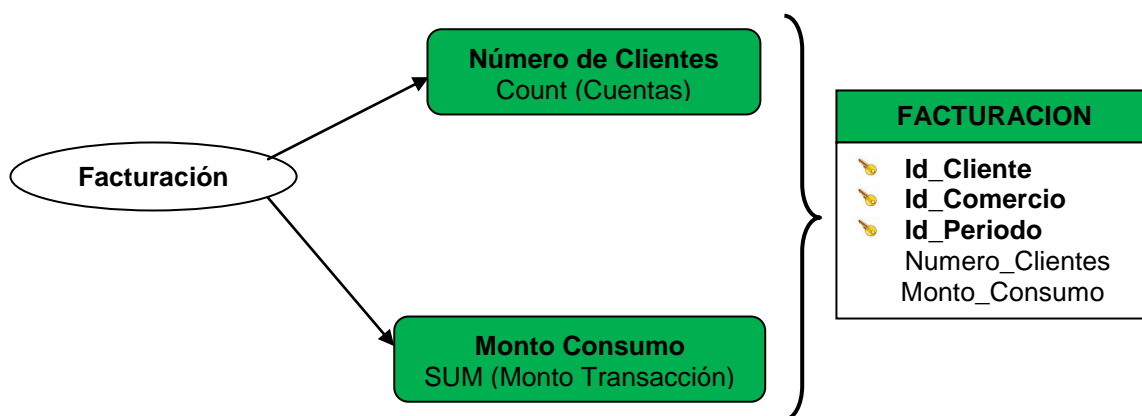


Figura 45, Representación gráfica tabla de hechos Facturación.

Proceso Mora

- El nombre de la tabla de hechos será “MORA”.
- Se definirá la clave principal como la combinación de las claves de las tablas de dimensiones definidas anteriormente: “Id_Cliente”, “Id_Tramo”, “Id_Periodo”
- Finalmente se crearán 2 hechos que equivalen a los 2 indicadores definidos para este proceso.

La figura 46 muestra el resultado de la construcción de la tabla de hechos “Mora”:

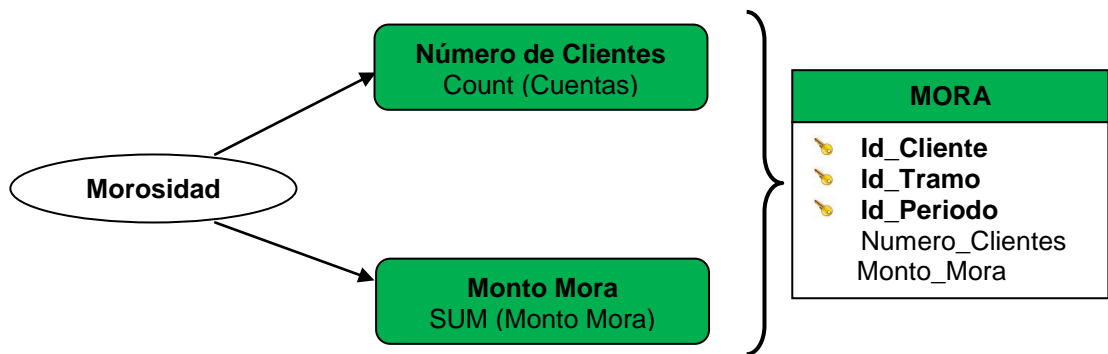


Figura 46, Representación gráfica tabla de hechos Mora.

Proceso Cierres

- El nombre de la tabla de hechos será “CIERRES”.
- Se definirá la clave principal como la combinación de las claves de las tablas de dimensiones definidas anteriormente: “Id_Tramo”, “Id_Periodo”
- Se crearán 2 hechos que equivalen a los 2 indicadores definidos para este proceso.

La figura 47 muestra el resultado de la construcción de la tabla de hechos “Cierres”:

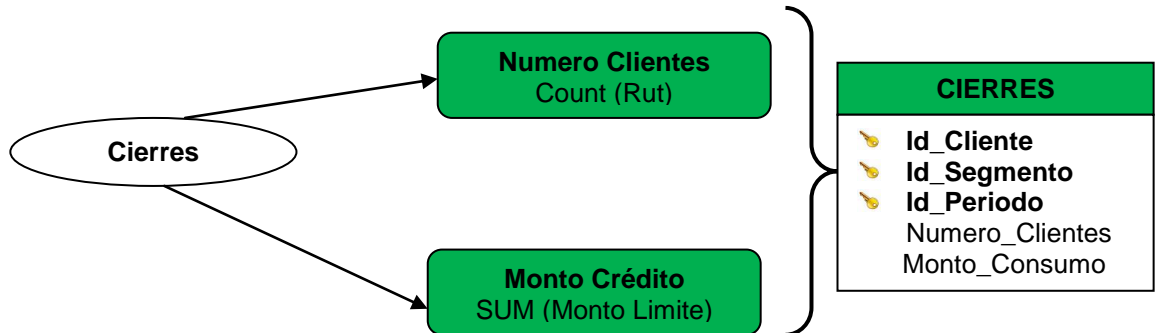


Figura 47, Representación gráfica tabla de hechos “Cierres”.

4.1.16 Uniones

Para los tres tipos de esquemas, se realizarán las uniones que correspondan entre sus tablas de dimensiones y sus tablas de hechos:

Proceso Ventas

La figura 48 muestra el resultado de la unión de tablas de dimensiones y hechos para el proceso de ventas:

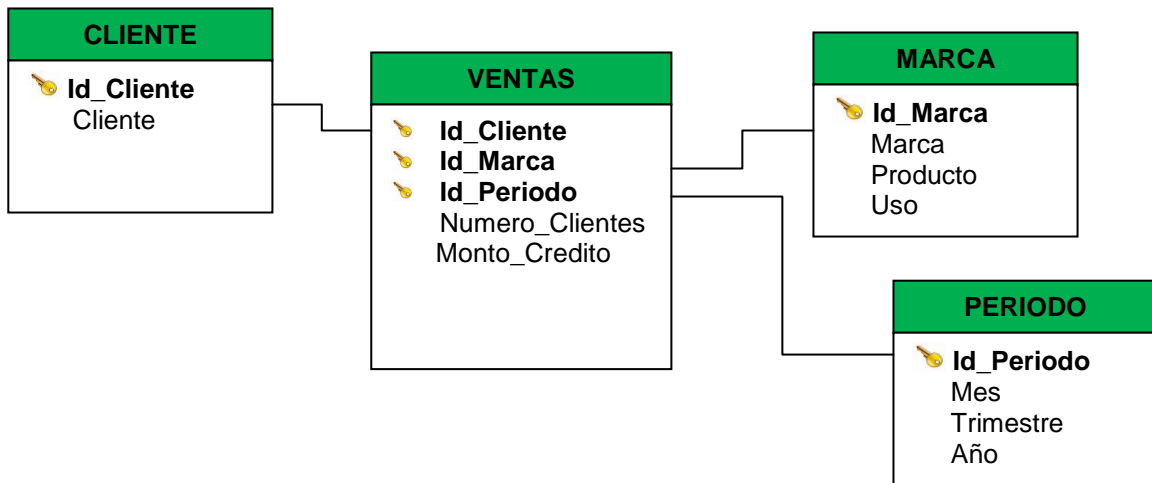


Figura 48, Unión de tablas de dimensiones y hechos para el proceso de ventas.

Proceso Facturación

La figura 49 muestra el resultado de la unión de tablas de dimensiones y hechos para el proceso de facturación:

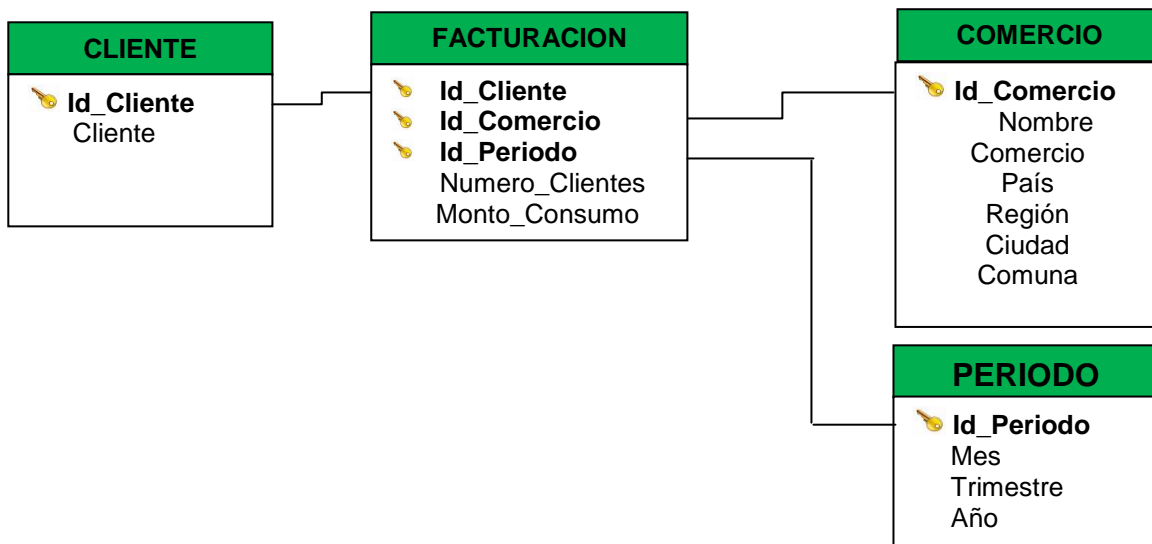


Figura 49, Unión de tablas de dimensiones y hechos para el proceso de facturación.

Proceso Mora

La figura 50 muestra el resultado de la unión de tablas de dimensiones y hechos para el proceso de mora:

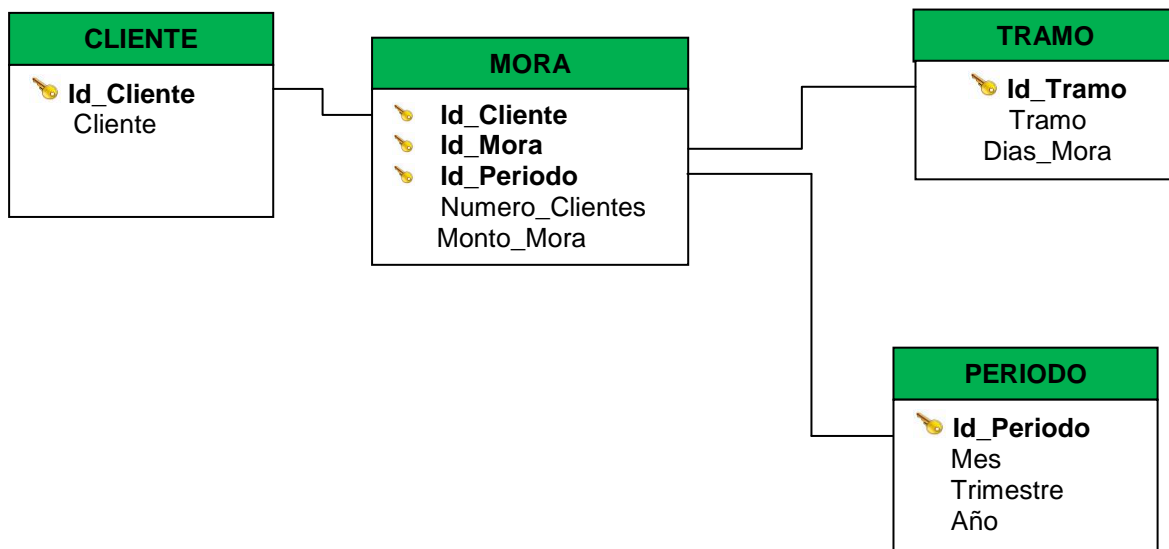


Figura 50, Unión de tablas de dimensiones y hechos para el proceso de mora.

Proceso Cierres

La figura 51 muestra el resultado de la unión de tablas de dimensiones y hechos para el proceso de mora:

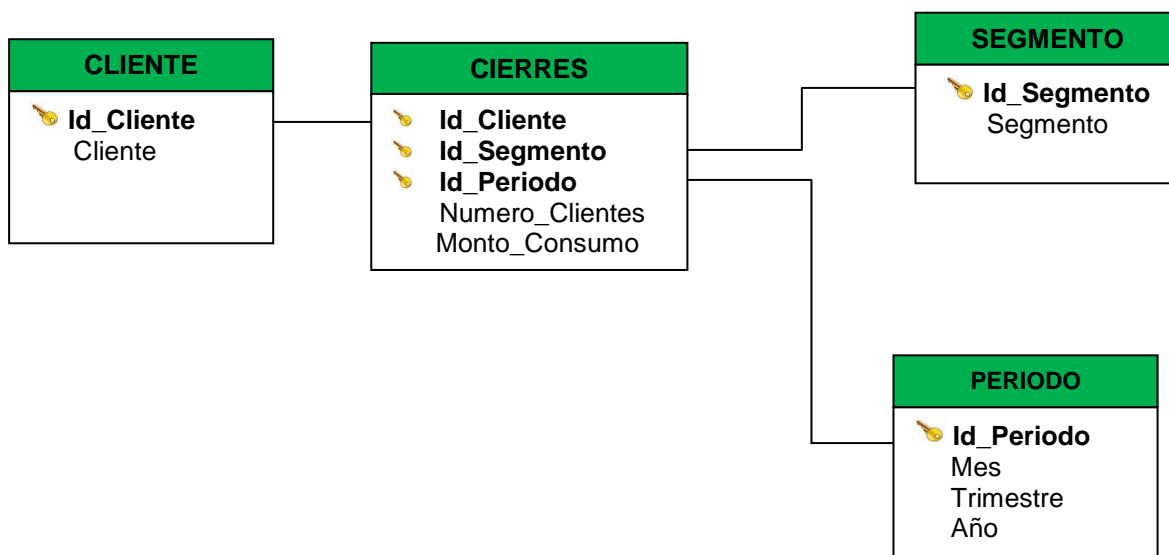


Figura 51, Unión de tablas de dimensiones y hechos para el proceso de cierres.

4.2 ETL y Carga de Datos

Al finalizar el modelo lógico, se realizarán los pasos para agregar los datos del OLTP (Base de Datos relacional) al modelo desarrollado, esto a través de procesos ETL.

Debido a la complejidad de esta actividad, que significa extraer los datos desde el sistema fuente para luego realizar los filtros y una correcta depuración, se utilizará el motor SQL Server que aloja la base de datos relacional y a través de distintas consultas se extraerán los datos de interés para el modelo definido.

Antes de este proceso es necesario realizar una revisión exhaustiva de los datos con el fin de evitar cargar datos con la falta de información, inconsistencias o datos erróneos.

La generación de los ETL implica la visión y el conocimiento de la información que se desea almacenar, estableciendo restricciones y condiciones, lo cual evitará la pérdida de datos importantes para análisis.

El primer paso a realizar será la carga de las dimensiones para luego continuar con las tablas de hechos, privilegiando la correspondencia entre cada dato. En caso de existir jerarquías de dimensiones, se procederá a realizar la carga de dimensiones desde lo más general a lo más detallado, de tal forma de evitar errores de integridad referencial.

Una vez finalizada la carga del Data Mart se establecerán políticas de actualización de datos y frecuencia, detallando claramente los tiempos de carga esperados para cada tabla o dimensión.

Se procederá a general las consultas SQL para la carga de las tablas de dimensiones y hechos:

Tabla de dimensión “CLIENTE”

- La tabla fuente del OLTP para esta dimensión será la tabla “Clientes”.
- Se concatenaran los campos “NOMBRES”, “APEPAT” y “APEMAT” conformando el campo “Cliente”.

```
SELECT
    [RUT] AS Id_Cliente
    , [NOMBRES]+' '+[APEPAT] +' '+ [APEMAT] AS Cliente
FROM
    [BD_TARJETAS_BCO].[dbo].[Clientes]
ORDER BY
    [RUT]
```

Tabla de dimensión “MARCA”

La tabla fuente del OLTP para esta dimensión será la tabla “Tipo_Producto”.

```
SELECT
    [COD_PROD] AS Id_Marca
    , [COD_MARCA] AS Marca
    , [DES_PROD] AS Producto
    , [COD_USO] AS Uso
FROM
    [BD_TARJETAS_BCO].[dbo].[Tipo_Producto]
ORDER BY
    [LOGO]
```

Tabla de dimensión “PERIODO”

Esta es una tabla muy importante en el desarrollo del Data Mart, existen varias herramientas de software que proporcionan diversas opciones para la confección de esta tabla, sin embargo si no se cuenta con ninguna, se realizará importando los datos desde una hoja de cálculo y luego exportándolos a donde se requiera.

A continuación en la figura 52 se aprecia una muestra de los datos generados.

	A	B	C	D
1	Id_Periodo	Mes	Trimestre	Año
2	20130101	01	01	2014
3	20130102	01	01	2014
4	20130103	01	01	2014
5	20130104	01	01	2014
6	20130105	01	01	2014
7	20130106	01	01	2014
8	20130107	01	01	2014
9	20130108	01	01	2014
10	20130109	01	01	2014
11	20130110	01	01	2014
12	20130111	01	01	2014

Figura 52, Datos en tabla de dimensión periodo.

La clave principal es un campo numérico representado por el formato “yyyymmdd”.

Tabla de dimensión “COMERCIO”

La tabla fuente del OLTP para esta dimensión será la tabla “Comercios”.

```
SELECT
    [RUT_COM] AS Id_Comercio
    , [DES_COM] AS Nombre_Comercio
    , [PAIS_COM] AS Pais
    , [REGION_COM] AS Region
    , [CIUDAD_COM] AS Comuna
    , [COMUNA_COM] AS Comuna
FROM
    [BD_TARJETAS_BCO].[dbo].[COMERCIOS]
```


Tabla de dimensión “TRAMO”

La tabla fuente del OLTP para esta dimensión será la tabla “Tipo_Mora”.

```
SELECT
    [COD_MORA] AS Id_Mora
    , [COD_MORA] AS Tramo
    , [DES_MORA] AS Dias_Mora
FROM
    [BD_TARJETAS_BCO].[dbo].[Tipo_Mora]
```

Tabla de dimensión “SEGMENTO”

La tabla fuente del OLTP para esta dimensión será la tabla “Clientes”.

Será necesario agrupar cada uno de los segmentos existentes en tabla de clientes para poder cargar la tabla de segmentos.

```
SELECT
    [SEGMENTO]
FROM
    [BD_TARJETAS_BCO].[dbo].[Clientes]
GROUP BY
    [SEGMENTO]
```

Tabla de hechos “VENTAS”

- Para la creación de esta tabla, se tomará como fuente de datos las tablas “Cuentas”, “Clientes”, “Tipo_Producto” y “Limite_Credito_Cta”
- Un punto importante en este proceso es que será necesario convertir la fecha de apertura a formato numérico “yyyymmdd”.

- Se aplicará una pre agregación a los campos que formarán parte de la tabla de hechos, por este motivo se utilizará la cláusula “GROUP BY” para agrupar los registros a través de las claves primarias de esta tabla.

```
SELECT
CLIENTES.RUT AS Id_Cliente,
TIPO_PRODUCTO.COD_PROD AS Id_Marca,
((YEAR(CUENTAS.FEC_APERTURA)*10000)
      + (MONTH(CUENTAS.FEC_APERTURA))*100 +
(DAY(CUENTAS.FEC_APERTURA))) AS Id_Periodo,
COUNT(CUENTAS.NUM_CTA) AS Numero_Clientes,
SUM(LIMITES.MONTO_LIMITE) AS Monto_Credito
FROM
[BD_TARJETAS_BCO].[dbo].[Cuentas] CUENTAS
INNER JOIN
[BD_TARJETAS_BCO].[dbo].[Clientes] CLIENTES ON
CLIENTES.CIC_CLIENTE=CUENTAS.CIC_CLIENTE
INNER JOIN
[BD_TARJETAS_BCO].[dbo].[Tipo_Producto] TIPO_PRODUCTO ON
TIPO_PRODUCTO.COD_PROD=CUENTAS.COD_PROD
INNER JOIN
[BD_TARJETAS_BCO].[dbo].[Limite_Credito_Cta] LIMITES ON
LIMITES.NUM_CTA=CUENTAS.NUM_CTA
GROUP BY
CLIENTES.RUT,
TIPO_PRODUCTO.COD_PROD,
CUENTAS.FEC_APERTURA
```

Tabla de hechos “FACTURACION”

- Para la creación de esta tabla, se tomará como fuente de datos las tablas “Cuentas”, “Clientes”, “Detalle_Facturacion” y “Comercios”
- Un punto importante en este proceso es que será necesario convertir la fecha de apertura a formato numérico “yyyymmdd”.
- Se aplicará una pre agregación a los campos que formarán parte de la tabla de hechos, por este motivo se utilizará la cláusula “GROUP BY” para agrupar los registros a través de las claves primarias de esta tabla.

```
SELECT
```

```
    CLIENTES.RUT AS Id_Cliente,
    COMERCIOS.RUT_COM AS Id_Comercio,
    ((YEAR(FACTURACION.FEC_TRX)*10000)+
    (MONTH(FACTURACION.FEC_TRX))*100 +
    (DAY(FACTURACION.FEC_TRX))) AS Id_Periodo,
    COUNT(CUENTAS.NUM_CTA) AS Numero_Clientes,
    SUM(FACTURACION.MONTO_TRX) AS Monto_Consumo
```

```
FROM
```

```
    [BD_TARJETAS_BCO].[dbo].[Detalle_Facturacion] FACTURACION
```

```
INNER JOIN
```

```
[BD_TARJETAS_BCO].[dbo].[Cuentas] CUENTAS ON
FACTURACION.NUM_CTA=CUENTAS.NUM_CTA
```

```
INNER JOIN
```

```
[BD_TARJETAS_BCO].[dbo].[Clientes] CLIENTES ON
CLIENTES.CIC_CLIENTE=CUENTAS.CIC_CLIENTE
```

```
INNER JOIN
```

```
[BD_TARJETAS_BCO].[dbo].[COMERCIOS] COMERCIOS ON
COMERCIOS.RUT_COM=FACTURACION.RUT_COM
```

```
GROUP BY
```

```
CLIENTES.RUT, COMERCIOS.RUT_COM, FACTURACION.FEC_TRX
```

Tabla de hechos “MORA”

- Para la creación de esta tabla, se tomará como fuente de datos las tablas “Cuentas”, “Clientes”, “Mora” y “Tipo_Mora”
- Un punto importante en este proceso es que será necesario convertir la fecha de apertura a formato numérico “yyyymmdd”.
- Se aplicará una pre agregación a los campos que formarán parte de la tabla de hechos, por este motivo se utilizará la cláusula “GROUP BY” para agrupar los registros a través de las claves primarias de esta tabla.

```

SELECT
    CLIENTES.RUT AS Id_Cliente,
    MORA.COD_MORA AS Id_Mora,
    ((YEAR(MORA.FEC_INI_MORA)*10000)+
    (MONTH(MORA.FEC_INI_MORA))*100 +
    (DAY(MORA.FEC_INI_MORA))) AS Id_Periodo,
    COUNT(CUENTAS.NUM_CTA) AS Numero_Clientes,
    SUM(MORA.MONTO_MORA) AS Monto_Mora
FROM
    [BD_TARJETAS_BCO].[dbo].[Mora] MORA
INNER JOIN
    [BD_TARJETAS_BCO].[dbo].[Cuentas] CUENTAS ON
MORA.NUM_CTA=CUENTAS.NUM_CTA
INNER JOIN
    [BD_TARJETAS_BCO].[dbo].[Clientes] CLIENTES ON
CLIENTES.CIC_CLIENTE=CUENTAS.CIC_CLIENTE
GROUP BY
    CLIENTES.RUT,
    MORA.COD_MORA,
    MORA.FEC_INI_MORA

```

Tabla de hechos “CIERRES”

- Para la creación de esta tabla, se tomará como fuente de datos las tablas “Cuentas”, “Clientes”, “Tipo_Producto” y “Limite_Credito_Cta”
- Un punto importante en este proceso es que será necesario convertir la fecha de apertura a formato numérico “yyyymmdd”.
- Se aplicará una pre agregación a los campos que formaran parte de la tabla de hechos, por este motivo se utilizará la cláusula “GROUP BY” para agrupar los registros a través de las claves primarias de esta tabla.

```

SELECT
    CLIENTES.RUT AS Id_Cliente,
    CLIENTES.SEGMENTO AS Id_Segmento,
    ( (YEAR(CUENTAS.FEC_CIERRE) *10000)
+ (MONTH(CUENTAS.FEC_CIERRE)) *100 +
    (DAY(CUENTAS.FEC_CIERRE))) AS Id_Periodo,
    COUNT(CUENTAS.NUM_CTA) AS Numero_Clientes,
    SUM(LIMITES.MONTO_LIMITE) AS Monto_Credito
FROM
    [BD_TARJETAS_BCO].[dbo].[Cuentas] CUENTAS
INNER JOIN
    [BD_TARJETAS_BCO].[dbo].[Clientes] CLIENTES ON
    CLIENTES.CIC_CLIENTE=CUENTAS.CIC_CLIENTE
INNER JOIN
    [BD_TARJETAS_BCO].[dbo].[Limite_Credito_Cta] LIMITES ON
    LIMITES.NUM_CTA=CUENTAS.NUM_CTA
GROUP BY
    CLIENTES.RUT,
    CLIENTES.SEGMENTO,
    CUENTAS.FEC_CIERRE

```

La información se actualizará en forma diaria y mensual, esta última contendrá la información del cierre de cada mes.

Los datos de la tabla dimensión “fecha” se cargarán en forma incremental teniendo en cuenta la fecha de última actualización.

Estas cargas se realizarán a modo de prueba durante un período para poder determinar los tiempos de carga y la forma más eficiente de actualización.

4.3 Creación de Cubos Dimensionales

En esta etapa se crearán los cubos multidimensionales que tomarán como base el modelo lógico diseñado anteriormente. El objetivo de la creación de estos cubos es ,principalmente, poder distinguir entre los hechos de una tabla y los indicadores de un cubo, y así mismo entender la diferencia entre los campos de una tabla de dimensión y los atributos de un cubo.

La figura 53 nos muestra cómo se representará el cubo para el proceso de ventas:

Cubo de Ventas

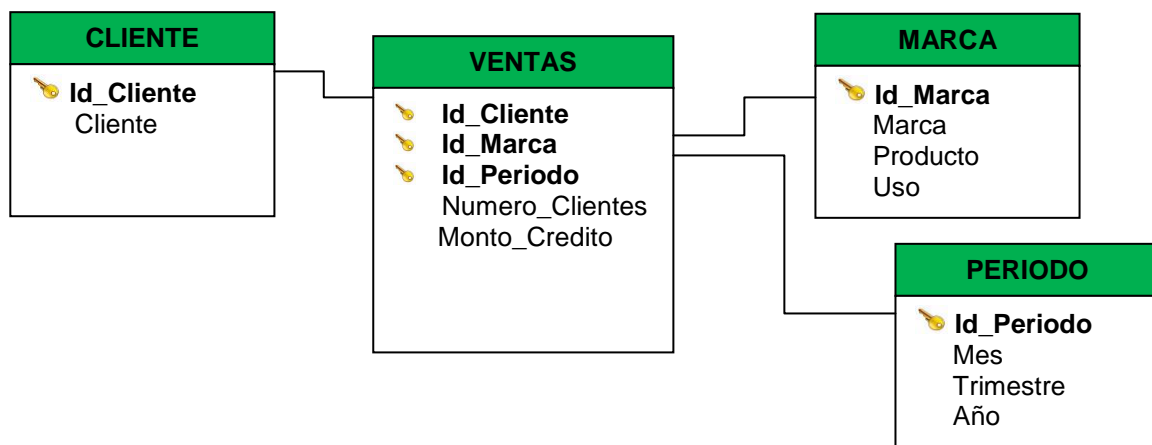


Figura 53, Representación cubo de ventas.

4.3.1 Creación de Indicadores

Se crearán los indicadores que serán incluidos en el “Cubo de Ventas”.

- Tomando como base la tabla de hechos “VENTAS” se aplicará una sumarización al hecho “Numero_Clientes” para crear el indicador :
 - “Número de Clientes”.
- La fórmula a utilizar para la creación de este indicador es:
 - “Número de Clientes”= SUM (VENTAS.Numero_Clientes).
- Tomando como base la tabla de hechos “VENTAS” se aplicará una sumarización al hecho “Monto_Credito” para crear el indicador :
 - “Monto Total Crédito”.

- La fórmula a utilizar para la creación de este indicador es:
 - “Monto Total Crédito”=SUM (VENTAS.Monto_Credito).

En la figura 54 se muestra cómo quedaría conformado el cubo:

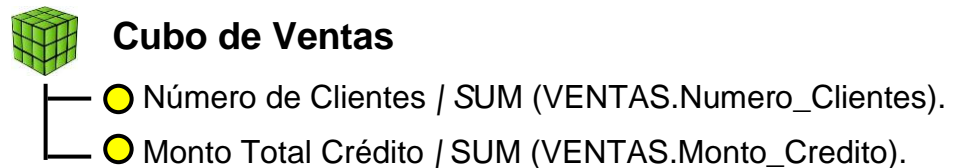


Figura 54, Representación de indicadores para cubo de ventas.

4.3.2 Creación de Atributos

En este punto se agregarán al cubo sus respectivos atributos:

- De la tabla de dimensión “CLIENTE” se tomará el campo “Cliente” para la creación de un atributo llamado:
 - “Clientes”.
- De la tabla de dimensión “MARCA” se tomará el campo “Marca” para la creación de un atributo llamado:
 - “Marcas”.
- De la tabla de dimensión “PERIODO” se tomará el campo “Año” para la creación de un atributo llamado:
 - “Años”.
- De la tabla de dimensión “PERIODO” se tomará el campo “Trimestre” para la creación de un atributo llamado:
 - “Trimestres”.
- De la tabla de dimensión “PERIODO” se tomará el campo “Mes” para la creación de un atributo llamado:
 - “Meses”.

La figura 55 muestra gráficamente la representación de los atributos para la construcción del cubo de ventas:

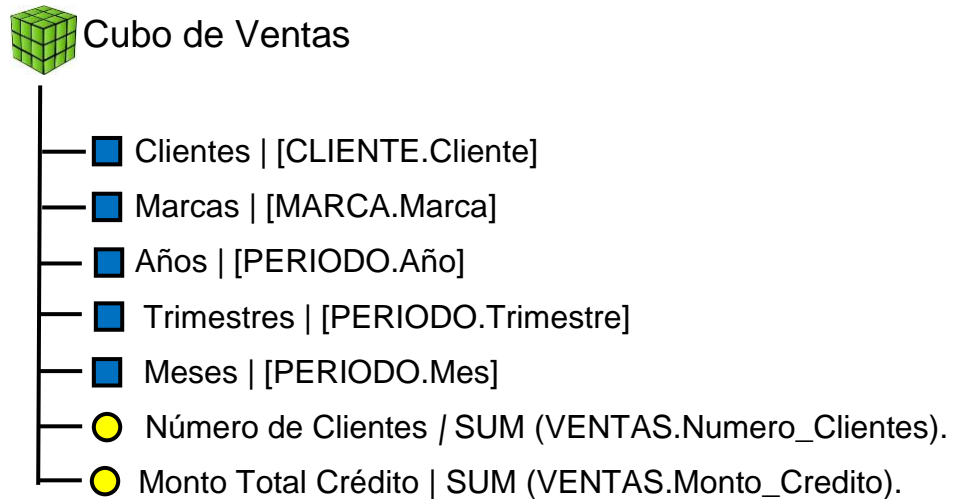


Figura 55, Representación de atributos para cubo de ventas.

4.3.3 Creación de Jerarquías

Como paso final se crearán y agregarán al cubo las jerarquías:

- Se definirá la jerarquía “Jerarquía Productos”, aplicada sobre el atributo “Marca”:
- Un producto pertenece a una sola marca, una marca puede tener uno o más productos.

En la figura 56 podemos apreciar la representación de la jerarquía “Marca”:

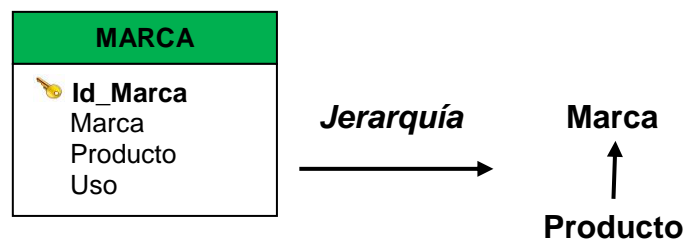


Figura 56, Representación de jerarquía “Marca”.

- Se definirá la jerarquía “Jerarquía Fechas”, aplicada sobre los atributos “Años”, “Trimestres”, “Meses”:
- Un mes del año pertenece a un solo trimestre del año, Un trimestre del año tiene uno o más meses del año.
- Un trimestre del año pertenece solo a un año. Un año tiene uno o más trimestres del año.

En la figura 57 podemos apreciar la representación de la jerarquía “Período”:

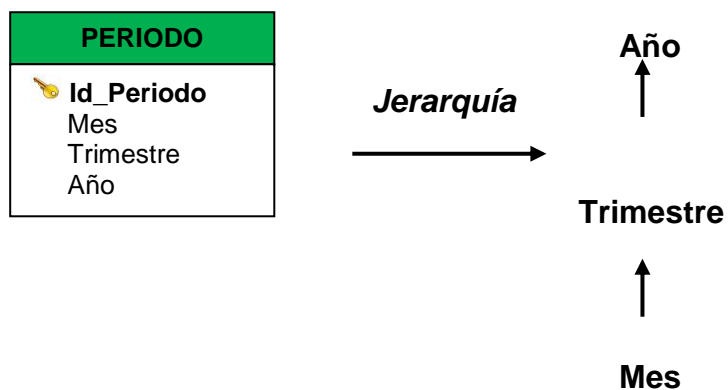


Figura 57, Representación de jerarquía Periodo.

En la figura 58 se puede apreciar cómo quedará conformado el cubo final “Ventas”:

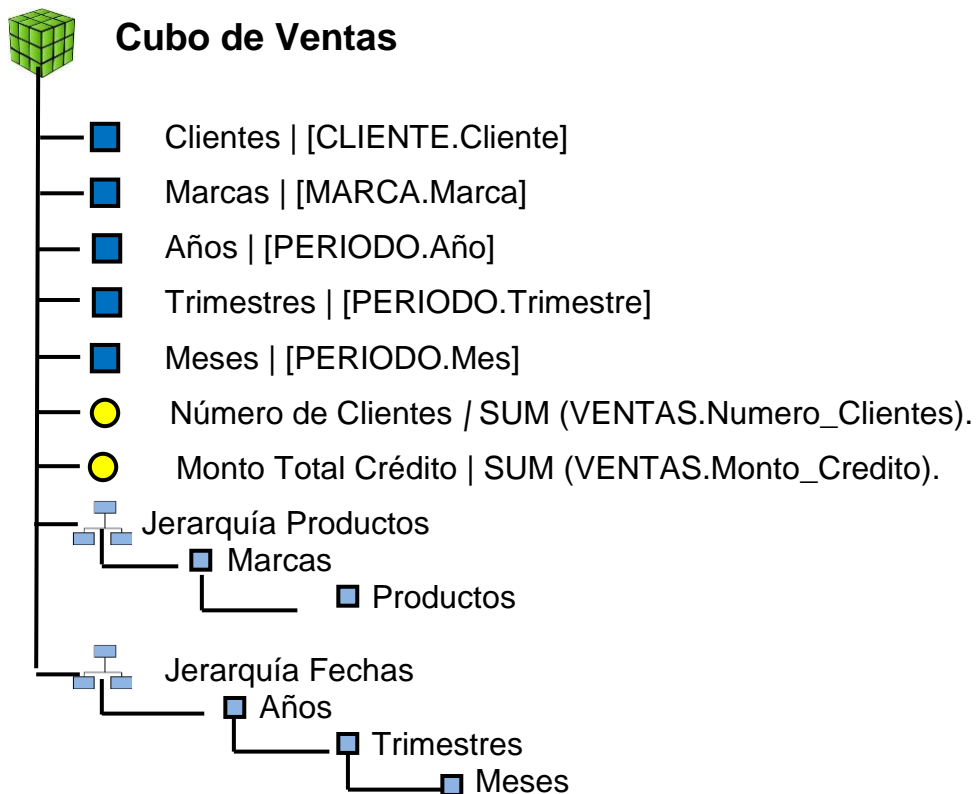


Figura 58, Prototipo final cubo multidimensional de ventas.

4.3.4 Otros Cubos Multidimensionales del Modelo

Tomando como base el modelo lógico desarrollado para este Data Mart se puede crear una gran cantidad de cubos orientados a algún análisis en particular.

Las figuras 59, 60 y 61 muestran la representación final de los cubos para los procesos de facturación, mora y cierres:

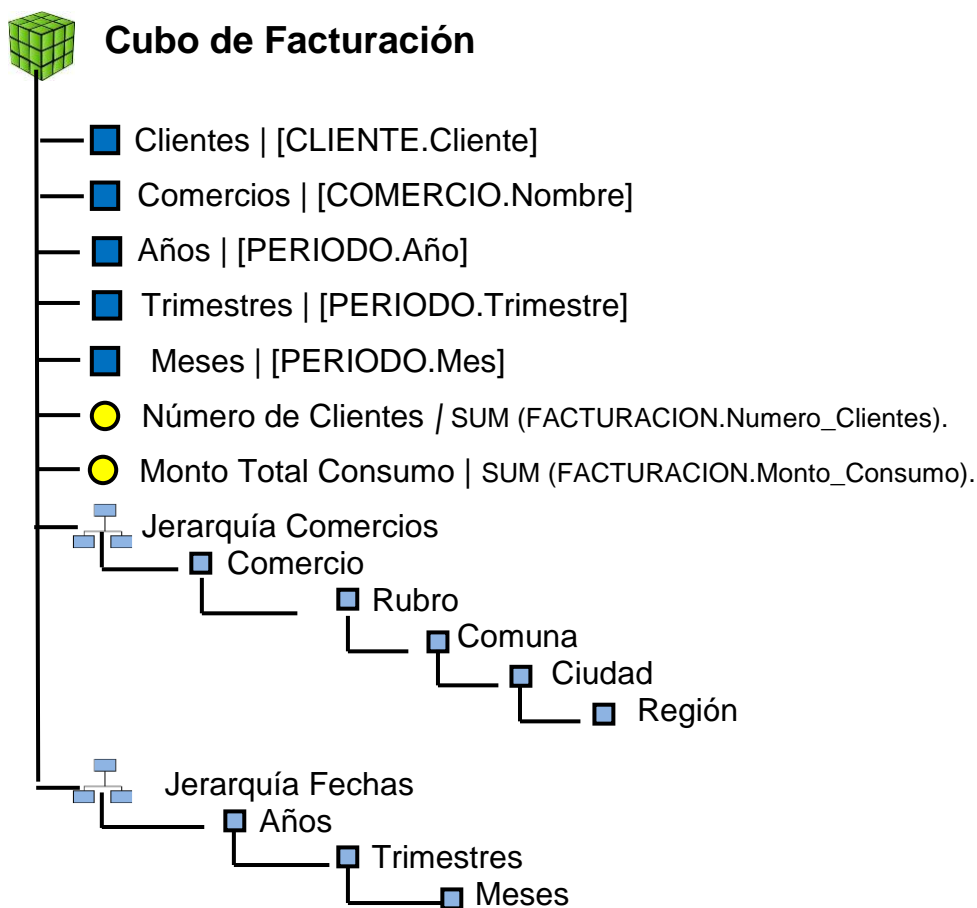


Figura 59, Prototipo final cubo multidimensional de facturación.

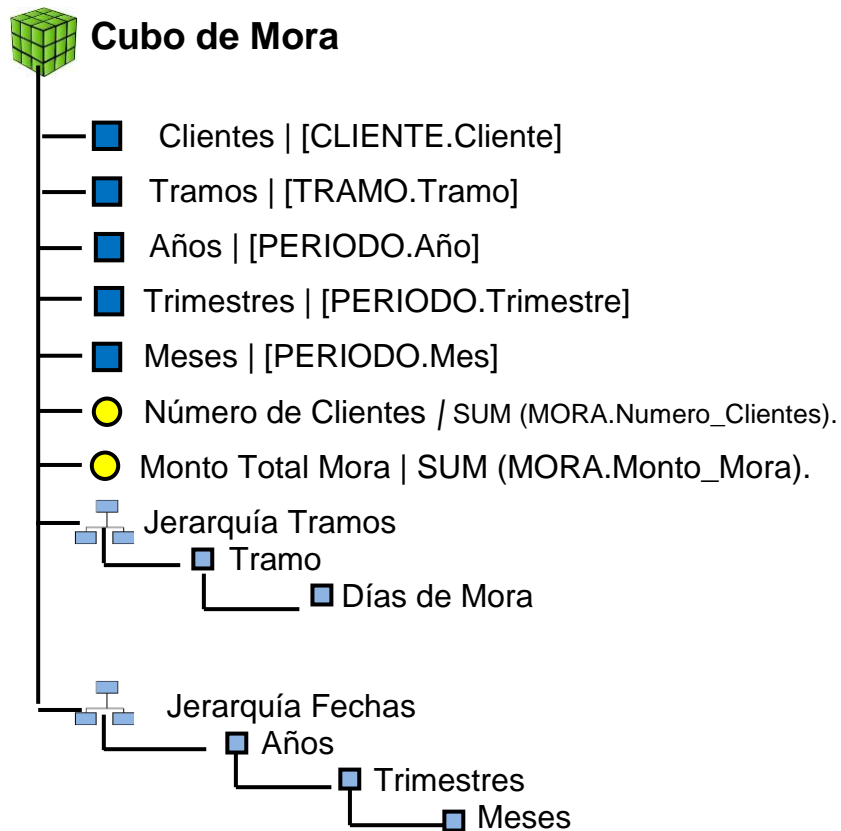


Figura 60, Prototipo final cubo multidimensional de mora.

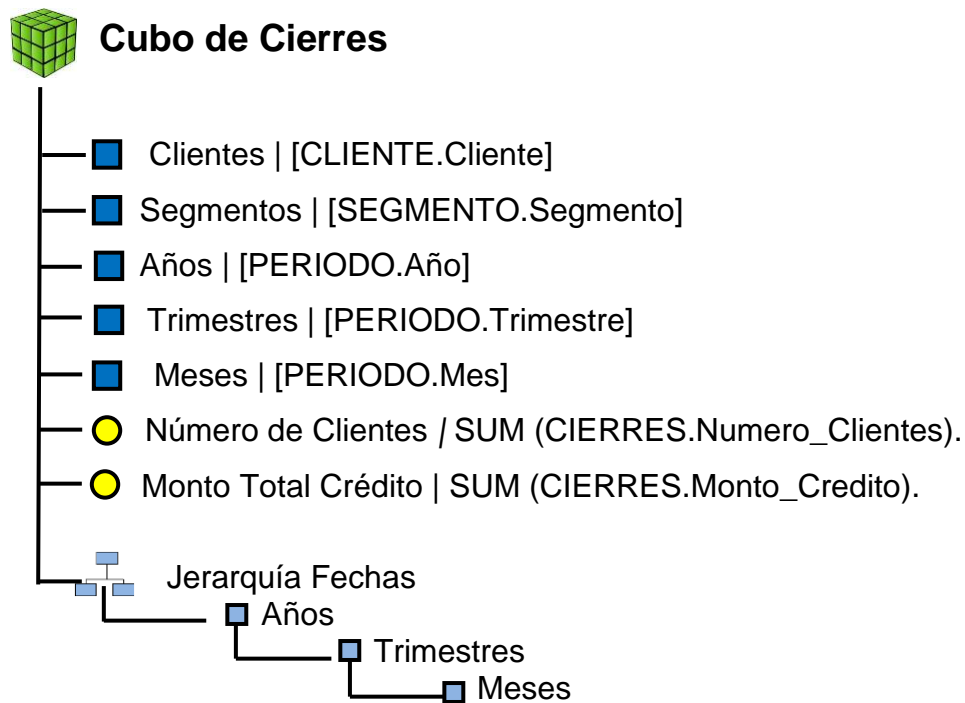


Figura 61, Prototipo final cubo multidimensional de cierres.

4.4 Prototipos de Informe

Es la etapa final del proyecto en la cual se busca responder en forma concreta a las preguntas obtenidas en el análisis de requerimientos, los usuarios que utilizarán el Data Mart podrán contar con una herramienta potente que les permita obtener con exactitud y en forma gráfica información útil para el control y la toma de decisiones estratégicas, utilizando como base los cubos multidimensionales creados.

4.4.1 Informe de Mora

Este prototipo en específico busca responder a las preguntas:

- a) Cantidad de Clientes para cada tramo de mora en un periodo determinado.
- b) Monto en mora por tramos en un periodo determinado.

La figura 62 nos da un ejemplo del prototipo de informe “Informe de mora”:

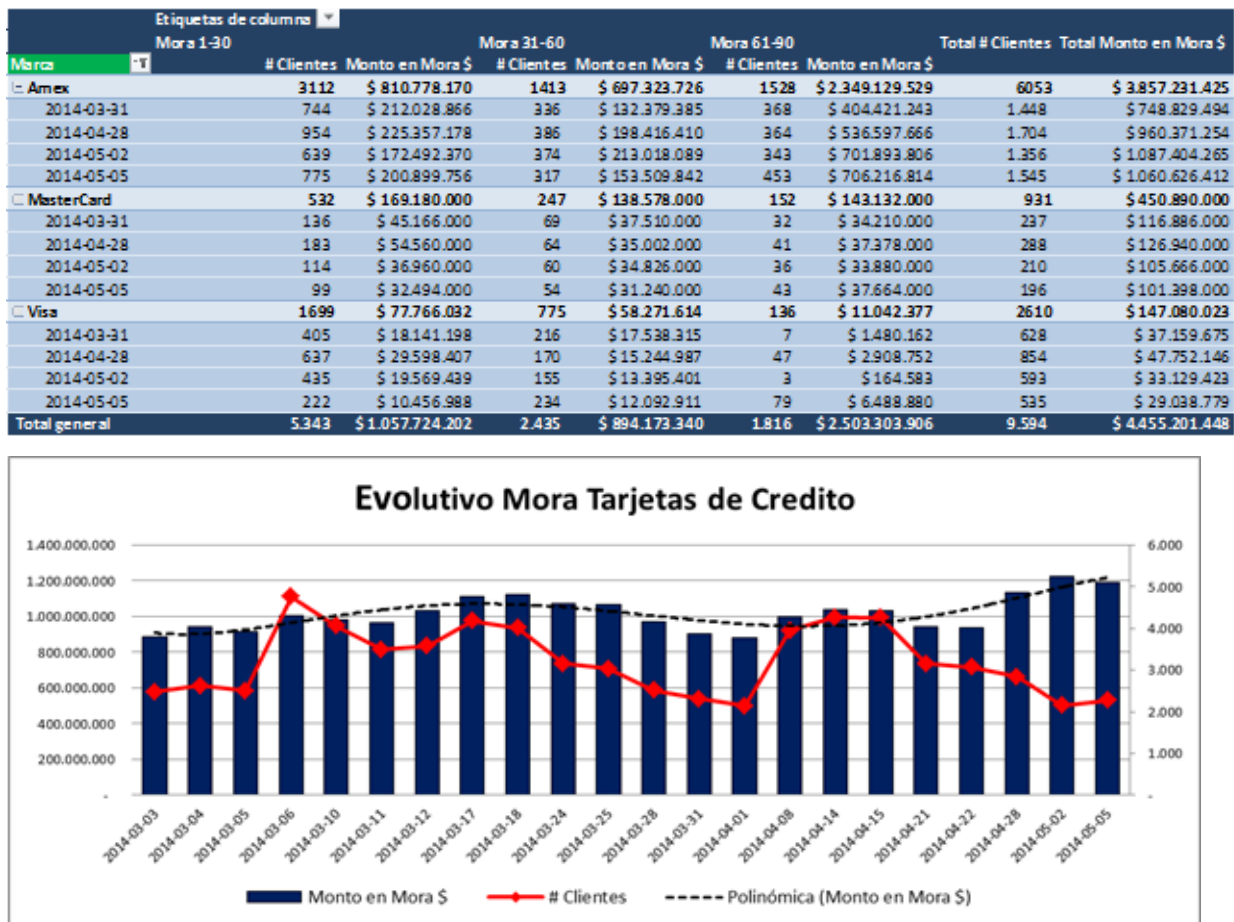


Figura 62, Prototipo de informe de mora.

4.4.2 Informe de Ventas y Consumos

La construcción de este prototipo busca responder a las siguientes preguntas:

- Numero de adquisiciones de clientes por marca en un periodo determinado.
- Monto de crédito asignado a cada cliente en un periodo determinado.
- Monto de consumos por cada cliente en un periodo.

La figura 62 nos da un ejemplo del prototipo de informe “Informe de ventas y consumo”:

Vigente Año Anterior	(Todas)	▼		
Segmento TBanc	(Todas)	▼		
Filtro de Riesgo	(Todas)	▼		
Funcionario	(Todas)	▼		
Segmento Marketing	(Todas)	▼		
Banca	(Todas)	▼		
PANEL DE SEGMENTOS TCR				
	Etiquetas			
Valores	201312	201401	201402	201403
% Cruce Tarjeta de Crédito	79%	79%	79%	79%
# Stock Tarjeta de Crédito	44.676	44.675	44.541	44.705
% Cruce PAT	18%	18%	19%	19%
% Clientes con Compras Nacionales	37%	35%	34%	36%
# Compras Nacionales Tarjeta de Crédito	120.491	101.140	95.900	108.094
Volumen Compras Tarjeta de Crédito (M\$)	5.491.481	4.353.831	3.894.249	5.001.869
% Clientes con Avances en Tarjeta de Crédito	6%	7%	6%	6%
# Avances en Tarjeta de Crédito	4.619	5.093	4.425	4.615
Monto Avances Tarjeta de Crédito (M\$)	674.617	670.591	644.021	640.295
Saldo Promedio Tarjeta de Crédito (M\$)	12.278.814	12.741.031	13.698.652	13.929.671
# Ventas Mes	9.372	9.484	9.290	9.188
Monto Credito Otorgado MMS	119.694	130.501	128.121	115.837
% Clientes con Protestos	0%	0%	0%	0%
Margen Financiero Total (M\$)	1.334.239	1.280.128	1.163.800	1.367.414
Margen Financiero CCT (M\$)	208.679	184.721	166.355	190.077
Margen Financiero Línea de Sobregiro (M\$)	250.272	242.285	218.222	241.991
Margen Financiero Línea de Emergencia (M\$)	5.041	4.666	2.650	4.591
Margen Financiero Tarjeta de Crédito (M\$)	298.912	278.017	254.316	288.935
Comisión Mantenimiento Tarjeta de Crédito (M\$)	24.379	10.669	5.090	3.461
Tarjetas / Clientes	1,16	1,16	1,16	1,18

Figura 63, Prototipo de informe de ventas y consumos.

4.4.3 Informe de Cierres por Sucursal

La construcción de este prototipo busca responder a las siguientes preguntas:

- a) Cantidad de tarjetas cerradas para cada segmento en un periodo determinado.
- b) Monto crédito de tarjetas cerradas para cada segmento en un periodo.

La figura 62 nos da un ejemplo del prototipo de informe “Informe de cierres”:

Con_cct		Tramo Mora		Valores					
(Todas)				Enero		Febrero		Marzo	
FECHA PROCESO 2014-05-05				Operaciones	MONTO_Otorgado	Operaciones	MONTO_Otorgado	Operaciones	MONTO_Otorgado
SUCURSAL	Producto	Operaciones	MONTO_Otorgado	Operaciones	MONTO_Otorgado	Operaciones	MONTO_Otorgado	Operaciones	MONTO_Otorgado
R Norte I	Visa	117	33.075.609	26	15.684.374	38	50.387.267		
	MasterCards	16	6.204.000	7	3.234.000	3	2.222.000		
	Amex	28	1.798.424	24	1.622.306	4	264.764		
R Santiago II	Visa	116	35.846.268	38	21.700.995	41	67.102.967		
	MasterCards	17	5.830.000	7	4.488.000	4	4.422.000		
	Amex	28	1.965.264	21	1.299.383	7	1.165.835		
R Santiago Centro I	Visa	110	29.512.103	26	18.569.504	27	56.351.966		
	MasterCards	11	4.730.000	3	2.354.000	2	1.210.000		
	Amex	24	1.862.284	13	1.600.579	2	826.621		
R Metro Sur	Visa	100	20.087.771	28	11.452.872	29	74.251.459		
	MasterCards	9	3.630.000	3	1.848.000	3	1.716.000		
	Amex	34	1.310.914	32	1.950.882	10	513.500		
R Oriente	Visa	112	26.672.383	48	18.123.226	56	62.118.599		
	MasterCards	9	2.596.000	6	2.904.000	4	2.706.000		
	Amex	32	1.315.268	47	2.806.305	15	1.088.526		
O Lo Barnechea	Visa	117	19.252.498	57	17.960.114	66	40.346.195		
	MasterCards	9	1.870.000	1	858.000	3	1.782.000		
	Amex	68	1.558.037	91	2.395.838	34	2.138.090		
Total general		957	199.116.823	478	130.852.378	348	370.613.789		

Figura 64, Prototipo de informe de cierres.

5 Conclusión

La implementación de tecnologías de Data Warehouse y soluciones de Business Intelligence, aplicadas al negocio de Tarjetas de Crédito de la institución bancaria en estudio, permitirá obtener una serie de beneficios tangibles e intangibles dentro de la organización. Dado que la información es imprescindible para tomar cualquier decisión que afecte el negocio, la creación del Data Mart corporativo es una forma de recopilar los datos necesarios y construir los indicadores que aporten una visión transversal para la toma de decisiones.

Con la construcción del prototipo del Data Mart se logró que los principales usuarios pudieran expresar en forma clara y simple sus necesidades y requerimientos de información respecto a las tareas cotidianas y a lo que periódicamente se requiere de su parte, esto permitió automatizar la construcción de los reportes, lo que claramente se convirtió en un ahorro de tiempo considerable el cual puede ser destinado al análisis y acciones específicas que permitan corregir desviaciones y lograr el logro de las metas de la institución.

Adicionalmente se logró construir un panel para monitorear los KPI⁶ comprometidos con la alta gerencia y tomar acciones correctivas y a tiempo en cuanto estas se produzcan.

El análisis de la información proporcionada por el Data Mart permitió generar acciones de marketing y fidelización mediante campañas de telemarketing dirigidas específicamente a clientes que no utilizan sus productos, esto permitió realizar encuestas a los clientes y poder satisfacer sus expectativas respecto al producto logrando una activación temprana evitando el futuro cierre del producto.

Respecto a la mora en el producto, el análisis de la información proporcionada por el Data Mart permitió detectar que casos son los que representaban el mayor porcentaje de mora del total de cartera y realizar acciones de cobranza telefónica,

⁶ *KPI, del inglés Key Performance Indicator, conocido como Indicador clave de desempeño, (o también Indicador clave de rendimiento) es una medida del nivel del desempeño de un proceso; el valor del indicador está directamente relacionado con un objetivo fijado de antemano. Normalmente se expresa en porcentaje.*

envío de SMS , envío de mails y campañas Blaster⁷, estas mismas acciones permitieron automatizar el trabajo diario del área de cobranzas y maximizar la gestión de recupero de moras.

Con la experiencia de la construcción de este prototipo , fácilmente será posible expandirlo a otras áreas funcionales o de productos , ya que claramente la obtención de resultados concretos es un factor fundamental a la hora de la selección de una herramienta informática para el apoyo a la gestión.

⁷ *Campaña Blaster es un servicio de generación de llamadas pregrabadas, utilizada principalmente para recordatorios de cobranza, campañas políticas, invitación a eventos, alerta, alarmas y encuestas de satisfacción.*

6 Glosario de Términos

Agregación: Actividad de combinar datos desde múltiples tablas para formar una unidad de información más compleja, necesitada frecuentemente para responder consultas del Data Warehouse en forma más rápida y fácil.

Base de Datos: Conjunto de datos no redundantes, almacenados en un soporte informático, organizados de forma independiente de su utilización y accesibles simultáneamente por distintos usuarios y aplicaciones.

Carga: Momento en el cual los datos obtenidos en la fase transformación son cargados en el sistema de destino.

Data Mart: 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 Data Mart puede ser alimentado desde los datos de un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

Data Mining: Minería de datos, conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática.

Data Warehouse: 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.

Dimensión: Vista de datos categóricamente consistente. Todos los miembros de una dimensión pertenecen a un mismo grupo. Entidad independiente dentro del modelo

multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como mecanismo de selección de datos.

Drill down: Técnica de navegación utilizadas por los usuarios para visualizar el detalle de ciertos cálculos en base a jerarquías.

DSS: Sistema de Soporte de Decisiones. Sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la Información histórica.

EIS: Sistema de Información Ejecutiva (Executive information system, EIS por sus siglas en inglés) es una herramienta de Inteligencia empresarial (Business Intelligence, BI), orientada a usuarios de nivel gerencial, que permite monitorear el estado de las variables de un área o unidad de la empresa a partir de información interna y externa a la misma.

Esquema de Estrella: Componente básico del modelo dimensional, que consiste de una tabla central denominada tabla de hechos (fact table) y un conjunto de tablas “satélite” que representan las dimensiones. La tabla de hechos corresponde a un subconjunto del producto cartesiano de las dimensiones, y cada elemento tiene asociado, en general, un conjunto de medidas.

ETL: Extract, Transform and Load (extraer, transformar y cargar) proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, Data Mart, o Data Warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

Extracción: Acción de obtener la información deseada a partir de los datos almacenados en fuentes externas.

Fidelización: Concepto de marketing, 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.

Granularidad: Utilización de las jerarquías de diversas dimensiones dentro de las soluciones de inteligencia de negocios para realizar un análisis avanzado de tipo drill down y arrastrar datos.

GUI: graphical user interface, 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.

Hechos: Medidas numéricas, las cuales describen los resultados (rendimiento) del negocio. Un hecho debe ser un valor cuantificable o medible. Por ejemplo el número de unidades vendidas, ingreso por ventas, etc.

Inteligencia de Negocios: Conjunto de tecnologías y aplicaciones que permiten recopilar, almacenar, analizar y tener acceso a datos, de tal manera que los usuarios de la organización pueden tomar mejores decisiones. Consiste en transformar los datos operacionales de una empresa en información “accionable”, es decir, información que realmente habilite y optimice el proceso de toma de decisiones y la definición de estrategias y acciones encaminadas a mejorar el desempeño del negocio.

Institución Bancaria (Banco): Empresa financiera que se encarga de captar recursos en la forma de depósitos, y prestar dinero, así como la prestación de servicios financieros

Instancia: Realización específica de una clase o prototipo determinado.

Jerarquía: Consiste en los niveles de visualización de una condición dentro de

Inteligencia de negocios.

Marketing: Proceso social y administrativo por el que los grupos e individuos satisfacen sus necesidades al crear e intercambiar bienes y servicios.

Metadato: literalmente «sobre datos», son datos que describen otros datos.

Modelo de negocio: Representación simplificada de la lógica organizacional; es decir, la configuración de recursos de una compañía respecto a la obtención de ingresos y beneficios.

MOLAP: Acrónimo inglés de Multidimensional Online Analytical Processing, es decir, 'procesamiento analítico multidimensional en línea'.

Multidimensionalidad: Capacidad que ofrece una herramienta de Inteligencia de Negocios para analizar la información utilizando distintas dimensiones a la vez.

OLAP: 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: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos. Este sistema es típico de los Data Marts.

OLTP (On-line Transaction Processing): Procesamiento de Transacciones en línea, es un término utilizado para referirse a los sistemas de información de soporte del día a día o sistemas operacionales, enfocados al registro de transacciones realizadas en las diferentes áreas de una empresa.

Query: Es una consulta o una pregunta que se hace a una base de datos o a un sistema de información empresarial.

Reporting: Capacidad que tiene una herramienta de Inteligencia de Negocios para extraer directamente los datos de un sistema de gestión corporativo y de las múltiples bases de datos de la empresa, integrando y formateando dicha información en informes del tipo hojas de cálculo.

Rentabilidad: Beneficio comparado con los recursos propios invertidos para obtener esos beneficios.

Riesgo: Posible pérdida que asume un agente económico como consecuencia del incumplimiento de las obligaciones contractuales que incumben a las contrapartes con las que se relaciona.

ROLAP: Procesamiento Analítico OnLine Relacional, es decir, se trata de sistemas y herramientas OLAP (Procesamiento Analítico OnLine) construidos sobre una base de datos relacional.

RDBMS: Sistema de gestión de bases de datos relacionales, un sistema de gestión de bases de datos relacionales es aquel que sigue el modelo relacional.

SQL: Structured Query Language, lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en ellas.

Sistema de Información: Conjunto de elementos orientados al tratamiento y administración de datos e información, organizados y listos para su uso posterior, generados para cubrir una necesidad u objetivo.

Sumarización: Actividad de incremento de la granularidad de la información en una base de datos. La sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.

Tabla Dimensional: Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad a esta estrategia.

Transformación: Cualquier operación realizada sobre los datos para que puedan ser cargados en el Data Warehouse o se puedan migrar de éste a otra base de datos.

Tarjeta de Crédito: Instrumento material de identificación del usuario, que puede ser una tarjeta de plástico con una banda magnética, un microchip y un número en relieve. Es emitida por un banco o entidad financiera que autoriza a la persona a cuyo favor es emitida, a utilizarla como medio de pago en los negocios adheridos al sistema, mediante su firma y la exhibición de la tarjeta.

7 Índice de Figuras

Figura 01, Ejemplo de componentes de un Data Warehouse.	18
Figura 02, Representaciones OLTP y Data Warehouse.	22
Figura 03, Ejemplo de arquitectura de un Data Warehouse.	23
Figura 04, Representación gráfica de un hipercubo.	27
Figura 05, Ejemplo de esquema en estrella.	29
Figura 06, Ejemplo de esquema copo de nieve.	30
Figura 07, Ejemplo de Tabla de relaciones.	31
Figura 08, Representación esquema estrella para ventas.	36
Figura 09, Ejemplo esquema copo de nieve.	37
Figura 10, Representación gráfica etapas del proyecto.	50
Figura 11, Descripción de tareas para cada etapa del proyecto.	51
Figura 12, Descripción del flujo de tareas para el proyecto.	52
Figura 13, Carta Gantt del proyecto Data Mart Tarjetas de Crédito.	52
Figura 14, Representación gráfica de un modelo conceptual.	57
Figura 15, Representación gráfica del modelo conceptual Ventas.	58
Figura 16, Representación gráfica del modelo conceptual Facturación.	58
Figura 17, Representación gráfica del modelo conceptual Morosidad.	58
Figura 18, Representación gráfica del modelo conceptual Cierres.	59
Figura 19, Modelo base de datos relacional tarjetas de crédito.	62
Figura 20, Correspondencia entre modelo relacional y conceptual	63
Figura 21, Diagrama entidad relación facturación tarjetas de crédito.	64
Figura 22, Correspondencia de modelo relacional y conceptual de facturación	65
Figura 23, Diagrama entidad relación mora tarjetas de crédito.	67
Figura 24, Correspondencia entre modelo relacional y conceptual de mora.	68
Figura 25, Diagrama entidad relación cierre tarjetas de crédito.	69
Figura 26, Correspondencia entre modelo relacional y conceptual de cierres.	70
Figura 27, Representación gráfica de un modelo conceptual ampliado.	77
Figura 28, Representación gráfica modelo conceptual ampliado ventas.	77
Figura 29, Representación gráfica modelo conceptual ampliado facturación.	78
Figura 30, Representación gráfica modelo conceptual ampliado mora.	78
Figura 31, Representación gráfica modelo conceptual ampliado cierres.	79
Figura 32, Representación gráfica de una tabla de dimensiones.	80
Figura 33, Desglose jerarquía comercio.	80
Figura 34, Tablas normalizadas jerarquía comercio.	81
Figura 35, Tabla de dimensión Cliente.	81
Figura 36, Tabla de dimensión Marca.	82
Figura 37, Tabla de dimensión Periodo.	82
Figura 38, Tabla de dimensión Comercio.	83
Figura 39, Tabla de dimensión Tramo.	83
Figura 40, Tabla de dimensión Segmento.	84
Figura 41, Representación gráfica de una tabla de hechos.	85

Figura 42, Representación gráfica de una tabla de hechos en esquema constelación.	86
Figura 43, Representación gráfica de distintas tablas de hechos.	86
Figura 44, Representación gráfica tabla de hechos Ventas.	87
Figura 45, Representación gráfica tabla de hechos Facturación.	88
Figura 46, Representación gráfica tabla de hechos Mora.	89
Figura 47, Representación gráfica tabla de hechos Cierres.	89
Figura 48, Unión de tablas de dimensiones y hechos para el proceso de ventas.	90
Figura 49, Unión de tablas de dimensiones y hechos para el proceso de facturación.	90
Figura 50, Unión de tablas de dimensiones y hechos para el proceso de mora.	91
Figura 51, Unión de tablas de dimensiones y hechos proceso de cierres.	91
Figura 52, Datos en tabla de dimensión periodo.	94
Figura 53, Representación cubo de ventas.	101
Figura 54, Representación de indicadores para cubo de ventas.	102
Figura 55, Representación de atributos para cubo de ventas.	103
Figura 56, Representación de jerarquía Marca.	103
Figura 57, Representación de jerarquía Periodo.	104
Figura 58, Prototipo final cubo multidimensional de ventas.	105
Figura 59, Prototipo final cubo multidimensional de facturación.	106
Figura 60, Prototipo final cubo multidimensional de mora.	107
Figura 61, Prototipo final cubo multidimensional de cierres.	104
Figura 62, Prototipo de informe de mora.	108
Figura 63, Prototipo de informe de ventas y consumos.	109
Figura 64, Prototipo de informe de cierres.	110

8 Bibliografía

Libros

“Business Intelligence: Competir con información”, Autor: Josep Lluís Cano Giner, Año 2011, Editorial ESADE, Primera Edición.

“Sistemas de Bases de Datos, Un enfoque práctico para diseño, implementación y gestión”, Autor: Thomas M. Connolly, Carolyn E. Begg, Año 2005, Editorial Addison-Wesley, Cuarta Edición.

“The Data Warehouse Lifecycle Toolkit”, Autor: Ralph Kimball, Margy Gross, Warren Thornthwaite, Joy Mundy, Bob Becker, Año 2008, Editorial Wiley Publishing, Primera Edición.

Internet

<http://www.monografias.com/trabajos90/DataWarehouse-kimball-y-sql-2005/DataWarehouse-kimball-y-sql-2005.shtml>

<http://www.slideshare.net/IGTSlide/el-mercado-de-tarjetas-de-credito-en-chile-la-oferta-hacia-los-consumidores>

<http://bi-businessintelligence.blogspot.com/2008/11/manual-para-la-adquisicin-de-un-sistema.html>

<http://blog.classora.com/2013/06/25/bases-de-datos-multidimensionales-olap-vs-oltp/>

http://www.sinnexus.com/business_intelligence/Data Warehouse.aspx

<http://bddimensionales.wikispaces.com/Modelado+Dimensional>