

**Inteligencia de negocios para ventas en una empresa  
manufacturera**

**Nicolás Arias G.–Santiago Cáliz N.**

**Trabajo de grado para optar al título en Ingeniería Informática**

**Directora**

**Claudia Cecilia Noraña Pérez Ingeniera de Sistemas, Jefa  
del Área de TI de Socoda S.A**



**Escuela de Ingeniería de Antioquia  
Ingeniería Informática  
Envigado  
2012**

## **AGRADECIMIENTOS**

Queremos agradecer a los integrantes del área de sistemas de Socoda S.A por su ayuda, sin quienes este proyecto nunca se habría podido llevar a cabo. Adicionalmente muchas gracias a Jorge Weiman Durango, Director de Gestión Integral de Socoda. Ya que él fue la persona que más confió y apoyó el proyecto.

Por último un agradecimiento especial para Claudia Cecilia Noreña nuestra asesora de tesis, su conocimiento y experiencia es lo que aseguró que cumpliéramos con los objetivos trazados.

# CONTENIDO

	pág.
1 PRELIMINARES.....	15
1.1 Planteamiento del problema .....	15
1.2 Objetivos del proyecto .....	15
1.2.1 Objetivo General.....	15
1.2.2 Objetivos Específicos .....	15
1.3 Marco de referencia.....	16
2 Metodología.....	17
3 Requisitos del Sistema .....	18
4 Diseño .....	20
4.1 Arquitectura .....	20
4.2 Modelo Dimensional .....	21
4.2.1 Atributos de las dimensiones .....	23
4.3 ETL.....	28
4.3.1 <i>Jobs</i> Extracción .....	30
4.3.2 <i>Jobs</i> Transformación .....	32
4.3.3 <i>Jobs</i> Carga .....	36
4.4 Base de datos ETL.....	37
5 Selección de la herramienta de visualización.....	40
6 Ejemplo de Reporte .....	42
7 Resultados .....	44
8 Bibliografía .....	47
9 ANEXOS .....	48
9.1 Informe de reunión Ana María Franco .....	48

Inteligencia de Mercados Socoda S.A.....	48
Objetivos y responsabilidades .....	48
Reportes actuales.....	48
Análisis .....	48
Inconvenientes .....	49
9.2    Informe de reunión Adriana Jaramillo .....	50
Directora Canal de Construcciones e Instituciones Socoda S.A.....	50
Objetivos y responsabilidades .....	50
Clientes .....	50
Análisis .....	50
Reportes actuales.....	51
Inconvenientes .....	51
9.3    Informe de reunión Andrés Delgado .....	52
Director Canal de Exportaciones Socoda S.A .....	52
Análisis .....	52
Reportes actuales.....	52
9.4    Informe de reunión Andrés Restrepo .....	53
Director Canal de Distribuidores Socoda S.A .....	53
Objetivos y responsabilidades .....	53
Criterios de desempeño.....	53
Reportes Actuales .....	53
Análisis .....	54
Inconvenientes .....	54
9.5    Informe de reunión Iván Hincapié .....	55
Gerente Comercial Socoda S.A .....	55

Objetivos y responsabilidades .....	55
Criterios de desempeño.....	55
Reportes Actuales .....	55
Análisis.....	56
Inconvenientes .....	56
Anexos .....	56
9.6 Informe de reunión Carolina Velásquez.....	57
Gerente de Mercadeo Socoda S.A.....	57
Objetivos y responsabilidades .....	57
Reportes Actuales .....	57
Análisis.....	57
Inconvenientes .....	58
Anexos .....	58
9.7 Informe de reunión Diana Sanchez.....	59
Directora Canal de Cadenas Socoda S.A.....	59
Objetivos y responsabilidades .....	59
Clientes .....	59
Reportes actuales.....	59
Análisis.....	59
Inconvenientes .....	60
9.8 Anexo a Diseño de Jobs de ETL .....	61

## Lista de tablas

	pág.
Tabla 1-Dimensión Producto.....	23
Tabla 2-Dimensión Cliente.....	24
Tabla 3-Dimensión Vendedor .....	24
Tabla 4-Dimensión Moneda .....	25
Tabla 5-Dimensión Servicio .....	25
Tabla 6-Dimensión Fecha.....	25
Tabla 7-Dimensión Auditoría .....	26
Tabla 8-Fact Ventas .....	26

## Lista de figuras

	pág.
Ilustración 1-Kimball Lifecycle.....	17
Ilustración 2-Arquitectura global de BI .....	20
Ilustración 3-Modelo Dimensional .....	22
Ilustración 4-Jobs ETL Histórico .....	29
Ilustración 5-Extracción Clientes.....	<b>¡Error! Marcador no definido.</b>
Ilustración 6-Extracción Moneda.....	<b>¡Error! Marcador no definido.</b>
Ilustración 7-Extracción Productos.....	<b>¡Error! Marcador no definido.</b>
Ilustración 8-Extracción Servicios .....	<b>¡Error! Marcador no definido.</b>
Ilustración 9-Extracción Vendedores.....	<b>¡Error! Marcador no definido.</b>
Ilustración 10-Extracción Ventas.....	<b>¡Error! Marcador no definido.</b>
Ilustración 11-Extracción Notas de Cartera.....	<b>¡Error! Marcador no definido.</b>
Ilustración 12-Extracción Componentes Combo-Factura .....	<b>¡Error! Marcador no definido.</b>
Ilustración 13-Modelo dimensional de ETL .....	38
Ilustración 14-Ejemplo de Reporte.....	41

## Lista de anexos

	pág.
Anexo 1 - Ana María Franco .....	48
Anexo 2 - Adriana Jaramillo .....	50
Anexo 3 - Andrés Delgado .....	52
Anexo 4 - Andrés Restrepo .....	53
Anexo 5 - Iván Hincapié .....	55
Anexo 6 - Carolina Velásquez .....	57
Anexo 7 - Diana Sanchez .....	59
Anexo 8-Carga Clientes Incremental .....	61
Anexo 9-Carga Clientes .....	61
Anexo 10-Carga Moneda Incremental .....	61
Anexo 11-Carga Moneda .....	61
Anexo 12-Carga Producto Incremental .....	61
Anexo 13-Carga Producto .....	61
Anexo 14-Carga Servicios Incremental .....	61
Anexo 15-Carga Servicios .....	61
Anexo 16-Carga Vendedores Incremental .....	61
Anexo 17-Carga Vendedores .....	61
Anexo 18-Carga Ventas Incremental .....	61
Anexo 19-Carga Ventas .....	61
Anexo 20-Extracción Clientes Incremental .....	62
Anexo 21-Extracción Clientes .....	62
Anexo 22-Extracción Componentes Combo-Factura .....	62



Anexo 23-Extracción Moneda Incremental.....	62
Anexo 24-Extracción Moneda.....	62
Anexo 25-Extracción Notas de Cartera Incremental .....	62
Anexo 26-Extracción Notas de Cartera.....	62
Anexo 27-Extracción Productos Incremental .....	62
Anexo 28-Extracción Productos.....	62
Anexo 29-Extracción Servicios Incremental.....	62
Anexo 30-Extracción Servicios .....	62
Anexo 31-Extracción Vendedores Incremental .....	62
Anexo 32-Extracción Vendedores.....	62
Anexo 33-Extracción Ventas Incremental .....	63
Anexo 34-Extracción Ventas.....	63
Anexo 35-Master Histórico.....	63
Anexo 36-Master Incremental.....	63
Anexo 37-Transformación Clientes.....	63
Anexo 38-Transformación Moneda.....	63
Anexo 39-Transformación Productos.....	63
Anexo 40-Transformación Servicios .....	63
Anexo 41-Transformación Vendedores.....	63
Anexo 42-Transformación Ventas-Notas Cartera.....	63
Anexo 43-Transformación Ventas-Productos.....	63
Anexo 44-Transformación Ventas-Servicios .....	63

# GLOSARIO

**AD-HOC QUERYING:** Consulta o *query* que no está limitada por la herramienta que se está usando, o por unas definiciones previas. El usuario es el que decide que es lo que quiere consultar y como desea consultarlo. Este concepto se aplica mucho a análisis operacionales sobre una base de datos.

**ÁREA DE STAGING:** Área de almacenamiento de datos temporal. En un proceso de integración de datos es el lugar donde se van almacenando resultados que muestran diferentes pasos del proceso que se está ejecutando. El staging sirve como un back up en caso de cualquier error, además puede ser usado para verificar que un determinado proceso de transformación de datos si este funcionando como estaba previsto.

**BI:** Plataforma de inteligencia de negocios. Ayuda a la empresa a tener un acceso fácil a sus datos y provee los medios para que se puedan tomar decisiones basadas en los mismos y con ayuda de funciones estadísticas si se necesita. Además ayuda a mejorar la calidad de los datos gracias al proceso de ETL.

**ETL:** *Extract-Transform-Load* (Extraer-Transformar-Cargar). Proceso que se encarga de extraer información de las fuentes de datos operacionales, transformar dichos datos para que estén en la estructura necesaria, verificar la calidad de los datos siguiendo reglas específicas del negocio, y por último cargar dicha información en una base de datos que normalmente es el data warehouse o bodega de datos.

**DATA WAREHOUSE:** Gran bodega de datos que incluye toda la información que es valiosa para una compañía a la hora de tomar decisiones. Los datos que están contenidos en el *data warehouse* normalmente vienen de diferentes fuentes de datos operacionales que son usadas en procesos de negocio, la información es integrada y estructurada de manera que pueda ser consultada fácilmente. Un *data warehouse* se compone de varios data marts, que finalmente son varias bases de datos modeladas como bases de datos en estrella.

**DATA MART:** Cada uno de los almacenes que conforman un *data warehouse*, cada uno de éstos está enfocado en un proceso de negocio específico todos se relacionan entre sí para sintetizar los datos de la empresa. **DIMENSIÓN:** Las dimensiones son un conjunto de tablas del *data warehouse* que se utilizan para filtrar, agrupar y etiquetar los datos numéricos o mediciones almacenadas. Normalmente solo incluyen atributos textuales y además incluyen jerarquías para hacer sencilla la navegación sobre los datos. Un ejemplo de una Jerarquía en una dimensión fecha sería Año->Mes->Día. Se caracterizan por tener gran cantidad de columnas.

**DIMENSIÓN DEGENERADA:** Las dimensiones degeneradas son dimensiones que solo están compuestas de un atributo, por lo que para evitar construir una tabla completa de dimensión se añaden directamente a la tabla donde se encuentran los atributos numéricos o las mediciones relacionadas con dicha dimensión. Normalmente son códigos

operacionales que no tienen cabida en una de las otras tablas dimensiones previamente consideradas. Ejemplos: Orden de Venta, Numero de Factura.

**FACT:** Las tablas *Fact* son las que incluyen las mediciones o atributos numéricos que van a ser analizados. Son específicas de un proceso de negocio. Mientras que una dimensión puede ser utilizada en diferentes procesos, las tablas *facts* reflejan un proceso en especial, se caracterizan por tener gran cantidad de filas y no muchas columnas. Pueden incluir dimensiones degeneradas y además conectan las mediciones con las dimensiones relacionadas.

**GRANULARIDAD:** La granularidad denota el mayor nivel de detalle asociado a una tabla; es lo que ayuda a determinar que información tiene sentido y que información no se va a incluir en dicha tabla. Cada *fact* y cada dimensión tiene una granularidad asociada. Ejemplo: La granularidad de la dimensión de clientes es un registro por cada cliente, podemos bajar en el nivel de detalle agrupando los clientes por el atributo departamento.

**NATURAL KEY:** Código operacional que está relacionado con un registro de una dimensión, este código es administrado por los sistemas que están relacionados con la fuente de datos de la cual viene la dimensión. Normalmente hace parte del *data warehouse* para hacer más fácil el proceso de ETL. En la base de datos original el natural key identifica los registros de la dimensión y es el atributo que los vuelve únicos.

**SURROGATE KEY:** Es el identificador o clave primaria de una dimensión, se utiliza para relacionar los registros de las tablas de *facts* con las dimensiones. Normalmente son números enteros que son asignados por el motor de base de datos en el momento en el que se ingresan los registros a la tabla de dimensión.

## RESUMEN

La empresa Socoda S.A. cuenta con dificultades para la creación y manipulación de reportes que muestren la realidad respecto a las ventas. Esto causado porque su ERP(sistema de información gerencial) no cuenta con buenas capacidades para la generación de reportes, y además los datos que están ingresados dentro de la base de datos no se encuentran en el estado óptimo, los sistemas usados dentro de la empresa son bastante permisivos y los usuarios han terminado ingresando información descuidadamente.

Con base en estos problemas se propuso diseñar un BI (sistema de inteligencia de negocios), que ayude a facilitar el proceso de análisis de la información y toma de decisiones para las áreas de mercadeo y comercial de la empresa, que son las que están directamente relacionadas con el proceso de ventas, además de contribuir a mejorar la calidad en los datos.

El sistema se diseñó siguiendo los planteamientos de Ralph Kimball y de diferentes miembros de Kimball University, la cual es una organización dedicada a la consultoría en el área de inteligencia de negocios en el plano mundial. La metodología Kimball es la más utilizada para el desarrollo de este tipo de sistemas en el mundo.

Con base en esa metodología, primero se ejecutó una toma de requisitos a través de entrevistas con personajes representativos de cada una de las áreas descritas para tener una definición clara del sistema que están necesitando y de los aspectos importantes que se deben tomar en cuenta a la hora de realizar el diseño.

Con la ayuda de los requisitos recolectados y del área sistemas de la compañía se comenzó con el diseño de la arquitectura del BI. Para este diseño se tuvieron en cuenta las recomendaciones que da la metodología Kimball, la capacidad económica de la empresa, y por último se tomó en cuenta el hardware y software existente dentro de Socoda para que fuera reutilizado dentro de lo posible. En la etapa final de la arquitectura de evaluaron diferentes aplicaciones para el análisis de información, porque Socoda consideró importante contar con una herramienta que los asistiera en esta tarea. Se realizaron evaluaciones de algunas herramientas entre las cuales se hizo la selección.

Posteriormente se realizó el diseño dimensional del data warehouse. Se comenzó por definir cuáles serían las dimensiones involucradas. Por cada dimensión se examinó que información se tenía almacenada en el ERP y además que información nueva se quería adicionar. Con este conocimiento se identificaron los atributos de cada una de esas dimensiones. Además cuales otros miembros iban a formar parte de la tabla de *facts* como dimensiones degeneradas. Para concluir el diseño dimensional se establecieron los *facts*, que consisten en las ventas, cantidades vendidas y descuentos relacionados.

Por último se creó la estructura del ETL usando la base de datos del ERP de Socoda más el modelo dimensional anteriormente descrito como bases del sistema.

Palabras clave: ETL, Data warehouse, Modelo dimensional, ERP, BI

## ABSTRACT

Socoda S.A. is a company that has encountered difficulties in the creation and handling of reports that represent the current state of their sales. The reason behind this is mainly because their ERP doesn't include quality features for report creation and also because the data stored in their database isn't in optimal state, the systems used within the company are too permissive and the users have ended up entering information in a sloppy way.

Based on these problems, a BI (Business Intelligence system) design was proposed, which would help facilitate the process of report creation and administration for marketing and commercial departments inside the company. Both of these departments are the ones directly involved in the sales process. The system would also contribute to the quality assurance of Socoda's Data.

The system was designed following Ralph Kimball's and other members of the Kimball University's guidelines. Kimball University is an organization dedicated to Business Intelligence consulting at an international level for the development of this kind of applications.

Based on that methodology, the requirements for the system were elicited through interviews with different individuals from the aforementioned departments for a clear definition of the system they are expecting and the important aspects that must be taken into account when designing it.

Based on these requirements, and with the support of the people working in the IT department at Socoda, the system architecture was designed. For the system creation, besides the aforementioned guidelines, the company's current economical capability and its technological infrastructure were taken into account. Finally, an evaluation of different commercial products that the company could afford is included for them to be part of the project.

Afterwards, the dimensional design for the data warehouse was made. The first step was to define what were the dimensions involved. As a result of this, product, combo, date, salesperson, salesperson-client, currency, service, customer, offer and auditing dimensions were identified. For each dimension, an analysis was made to identify what information was already in the ERP and what was missing. Using this knowledge, the attributes needed for each dimensions were established, also what elements would be part of the facts tables as degenerate dimensions. In order to finish the dimensional design, the fact tables were established, they consisted of sales, quantity sold and related discounts.

Lastly, the ETL structure was created using Socoda's ERP database and the aforementioned dimensional model as the building blocks for the extraction, transformation and loading system.

## INTRODUCCIÓN

Socoda S.A es una empresa manufacturera que cuenta con información valiosa para la toma de decisiones, sin embargo la visualización y el análisis de la información es un problema debido a la falta de funcionalidad de los sistemas existentes para hacerlo. El sistema de inteligencia de negocios ayuda a tener una verdadera visión global de la compañía y mejora las decisiones tomadas a nivel gerencial gracias a la conocimiento que se gana de los procesos de negocio.

Adicionalmente el ETL que es uno de los componentes más importantes del BI, se preocupa por garantizar la calidad de los datos, evitando incoherencias y problemas en el futuro análisis de la información almacenada dentro del BI.

En este trabajo de grado se encontrarán los siguientes elementos:

- Una descripción de los requisitos recolectados que fueron la base para la creación del BI.
- Un resumen de los principios de la metodología Kimball que rigen el diseño de estas plataformas.
- Una explicación detallada de todo el proceso que se siguió para el diseño del sistema incluyendo arquitectura, ETL y modelado dimensional.
- Una evaluación de diferentes herramientas para la visualización y análisis de datos.

# 1 PRELIMINARES

## 1.1 Planteamiento del problema

El problema consiste en el diseño de un sistema de inteligencia de negocios para el proceso de ventas de la compañía que permita una visualización de los datos de alto nivel de una manera ágil y amigable que no requiera de conocimientos técnicos para su correcto uso

## 1.2 Objetivos del proyecto

### 1.2.1 Objetivo General

Diseñar una solución de inteligencia de negocios para el proceso de ventas de Socoda S.A, que posibilite una toma efectiva de decisiones.

### 1.2.2 Objetivos Específicos

- Establecer los *facts* o estadísticas que el área de ventas de Socoda desea manejar para que sean incluidas dentro del sistema y así puedan guiar las decisiones del área.
- Realizar el diseño del modelo dimensional de la base de datos que soportará el proyecto.
- Establecer las dimensiones específicas del área de ventas que van a ser usadas dentro del sistema. Las dimensiones son los elementos que se van a convertir en filtros o criterios de búsqueda de las estadísticas.
- Diseñar un proceso de ETL que permita extraer, transformar y cargar los datos desde la base de datos del ERP de Socoda a la bodega de datos y así contar con un sistema actualizado.
- Sugerir una herramienta de visualización de datos que proporcione las herramientas necesarias para generar los reportes que el usuario considere convenientes de forma fácil e intuitiva.

### 1.3 Marco de referencia

Para la realización de este trabajo de grado, se hace uso de la teoría de modelado dimensional desarrollada por Ralph Kimball, pionero en el tema de data warehousing e inteligencia de negocios y fundador de Kimball Group, una empresa dedicada al entrenamiento y consultoría para la construcción de data warehouses. La aplicación de dicha teoría consiste en el modelado de los procesos de negocio de la empresa en términos de “facts” y “dimensiones”. Los primeros son los datos de la empresa sobre los cuales se pueden hacer diferentes operaciones matemáticas para obtener nueva información a partir de ellos. Por otro lado, las dimensiones contienen información que puede ser usada posteriormente como filtros a la hora de visualizar los datos y generar reportes.

En la práctica, los *facts* y dimensiones son tablas dentro de una base de datos y cada una tiene sus propias características específicas: Los *facts* son tablas que contienen muy pocas columnas y una gran cantidad de filas, las columnas son, por lo general, un identificador, claves foráneas a las tablas de dimensiones y unos pocos datos “agregables”, mientras que las dimensiones son tablas con un gran número de columnas y pocos datos, en comparación con la cantidad de datos de los facts. Las columnas de las dimensiones, representan filtros por los cuales se puede pasar la información para obtener una síntesis de la misma en casos relevantes, mientras que las de los facts representan datos que se pueden agrupar u operar de diferentes maneras para la obtención de totales y estadísticas.

La base de datos que conforma un *data warehouse* es una base de datos de tipo analítico, optimizada para la extracción de datos. Ésta debe alimentarse de los sistemas transaccionales de la compañía, pero en muchos de los casos la información arrojada por estos sistemas se encuentra incompleta o en un formato que no es el adecuado para su análisis en la solución de inteligencia de negocios. Debido a esto nace la necesidad de un proceso de extracción, transformación y carga (ETL), en el que se extraen los datos del sistema transaccional, se hacen las modificaciones pertinentes a los datos para que tengan el formato y la estructura adecuada para su inserción en la data warehouse y finalmente se vuelcan a la misma.

Finalmente, a través de una herramienta de visualización de datos, los usuarios pueden generar reportes acerca del proceso de negocio haciendo uso de los filtros disponibles y agrupando la información relevante en las formas que consideren pertinentes para finalmente poder tomar decisiones sobre el rumbo que debe tomar el proceso ya mencionado.



## 2 Metodología

El desarrollo del proyecto se rigió por el método Kimball para la creación de sistemas de inteligencia de negocios. Esta metodología se caracteriza por centrarse en el negocio y sus necesidades como base para el desarrollo del sistema. Según el método los sistemas BI se crean de una manera iterativa atacando procesos del negocio separadamente, en un orden que está dado por la prioridad que se defina para cada uno. De esta forma la complejidad no es tan grande y en muchas ocasiones el desarrollo de un proceso anterior ayuda al de otros.

A continuación se muestra una gráfica del “Kimball Lifecycle” que es el ciclo de vida de un proyecto según la teoría que se aplicó.

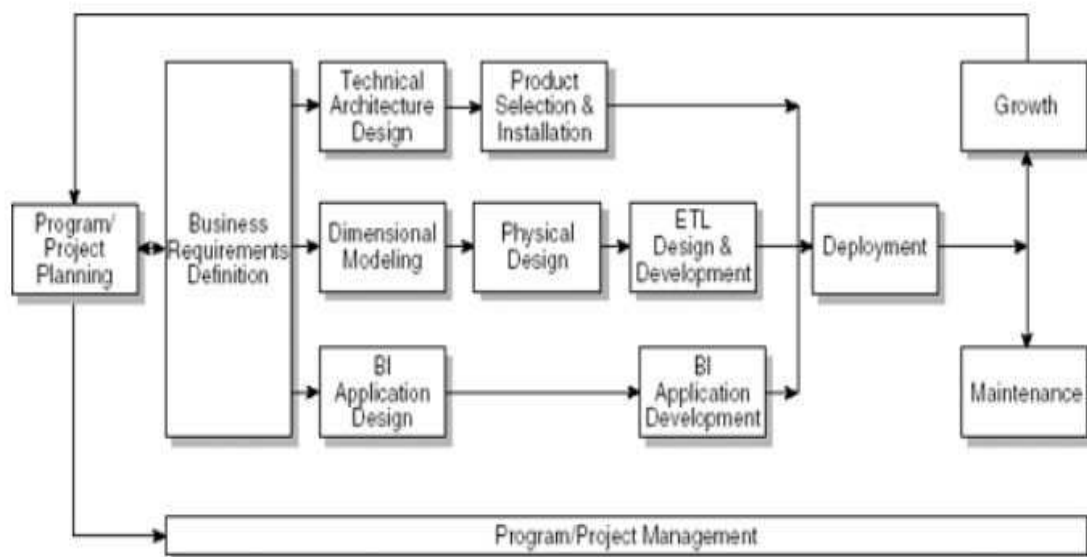


Ilustración 1-Kimball Lifecycle<sup>1</sup>

Siguiendo con el esquema presentado anteriormente, comenzamos con una planeación general con la jefa de TI de Socoda y con el Director de Gestión Integral, para definir entrevistas, cronograma, plan de acción, etc.

---

<sup>1</sup>Release Management of Data warehouses. Kimball Lifecycle <http://www.konings.biz/> (21 Diciembre de 2011)

Luego de esta reunión se empezaron a realizar diferentes entrevistas para la toma de requisitos. En Socoda existen cinco canales de ventas, cada canal tiene un director asignado que es quién que se encarga de dirigir a los vendedores y toda la gestión relacionada con el canal. Cada director fue entrevistado para conocer las particularidades de las ventas en cada canal, también se entrevistaron el gerente del área comercial, la gerente de mercadeo y los auxiliares que ayudan en la creación de reportes y estadísticas de la empresa.

De cada entrevista se creó un informe que se les enviaba a los entrevistados para retroalimentación. Además se grabaron dichas reuniones para poder consultarlas de ser necesario.

Terminada esta etapa de toma de requisitos, empezó el modelado de diseño dimensional y adicionalmente se comenzaron a evaluar diferentes herramientas para la visualización de los datos. Uno de los deseos más importantes de Socoda consistía en reducir la dependencia del área de sistemas para la creación y consulta de reportes de ventas, por lo que se analizaron herramientas haciendo énfasis en la usabilidad de la misma además de los requisitos, precio, robustez, etc.

Respecto al modelo dimensional, en primer lugar se determinó la granularidad de la tabla de facts, luego se identificaron las dimensiones asociadas así como los *facts* o mediciones que iban a formar parte del *data mart*. Por último se crearon las relaciones entre dimensiones y *facts* para terminar el modelo.

Para finalizar con el proyecto se estructuró el proceso ETL usando una herramienta open source llamada Talend Open Studio. Se crearon diferentes *jobs* o subprocesos de integración de datos. Por cada dimensión se crearon tres subprocesos, uno por cada tarea ETL (Extraer, Transformar y Cargar). Además se crearon trabajos adicionales para los facts y otras necesidades de Socoda.

### **3 Requisitos del Sistema**

Para esta primera etapa, se identificaron las personas dentro de Socoda que tienen poder de decisión sobre el proceso del negocio de ventas, ya que éstos son los potenciales usuarios del sistema de inteligencia de negocios. La toma de requisitos se centró en las áreas de mercadeo y comercial de la compañía, y se realizaron entrevistas con los cinco directores de canal correspondientes a los canales de cadenas, exportaciones, distribuidores, Estado y concesiones y constructores. Para terminar de cubrir el área comercial, se entrevistó finalmente al gerente del área y por sugerencia suya se hizo una última reunión con las personas que actualmente generan los reportes a partir de información extraída directamente del ERP de la compañía. Con respecto al área de mercadeo, se entrevistaron dos personas, la gerente del área y su asistente.

Siguiendo la metodología Kimball, estas entrevistas se llevaron a cabo en reuniones de una duración de aproximadamente una hora, en la que se les dio una breve explicación sobre los objetivos del proyecto y se cuestionó a los entrevistados sobre su capacidad

actual de generar reportes, los análisis que podían lograr, las carencias de los reportes con los que cuentan y qué expectativas tenían sobre el proyecto.

Un elemento importante que es necesario resaltar sobre las entrevistas es su corta duración, contrario a lo que se podría esperar de éstas, resultaron siendo más productivas debido a que fue posible persuadir a las personas para que sacaran tiempo de sus agendas para ser parte activa del proyecto en lugar de simplemente llenar rápidamente un formulario y seguir con su trabajo. Por otro lado, el hecho de que el tiempo sea tan limitado permite mantener foco de la reunión en lo que es realmente importante y se logran así requisitos más claros y concisos que finalmente se traduce en un sistema de calidad.

Las necesidades generales de reportes de los usuarios con respecto a las ventas fueron bastante básicas, algo que era de esperarse ya que estas personas no han tenido experiencia con un sistema de este tipo y no conocen su potencial. Se identificó principalmente la necesidad de visualizar las ventas tanto en unidades vendidas como en dinero, el filtrado de la información relevante para cada uno de ellos y la separación de algunos elementos agrupados que Socoda llama “combos”, que es algo que no se puede lograr con los reportes actuales.

Se identificó un requisito durante la reunión con la directora del canal de cadenas, Diana Sánchez, que, a pesar de ser relevante para el proyecto, está totalmente fuera del alcance del mismo. Para el canal de cadenas, es más importante conocer el inventario que sale de cada cadena y llega a un consumidor final que lo que Socoda les suministra. Como esta información es externa a la compañía y no se encuentra almacenada en la base de datos del ERP, no se puede incluir en el proyecto.

La etapa de toma de requisitos en un proyecto de inteligencia de negocios se considera culminada cuando, al entrevistar a los potenciales usuarios, éstos no manifiesten nada nuevo además de que los usuarios aprueben los informes que se les envían con la síntesis de las reuniones y no haya sugerencias de más personas para entrevistar porque podrían ser valiosas para el proyecto.

## 4 Diseño

El diseño de la plataforma se enmarco en 3 grandes bloques que son: Arquitectura, Modelo Dimensional y ETL.

### 4.1 Arquitectura

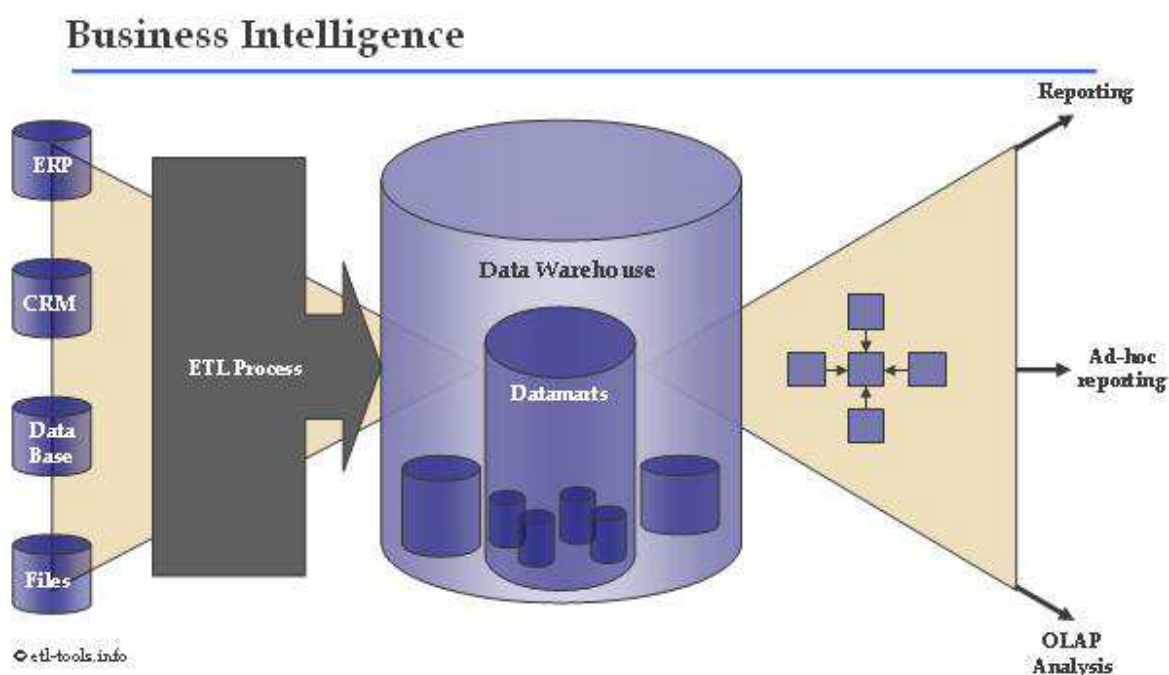


Ilustración 2-Arquitectura global de BI<sup>2</sup>

La arquitectura está dividida en una serie de elementos que componen el sistema; en primer lugar están las fuentes de datos que en el caso de Socoda S.A sólo es el ERP, puesto que toda la información relacionada con las ventas reside en su base de datos. El segundo elemento es el ETL. Su diseño como se dijo anteriormente se realizó con un producto open source que se llama Talend, el cual

---

<sup>2</sup> Business Intelligence. Arquitectura Global de BI <http://www.bizguts.com/tag/business-intelligence/> (11 de Septiembre de 2011)

produce código java internamente para la creación de los procesos de integración de datos.

Se eligió Talend por su facilidad de uso, por el conocimiento en java por parte de los integrantes del equipo de trabajo, y además por la movilidad que brinda java al ser un lenguaje multiplataforma. En la próxima sección se explica en detalle el diseño del ETL, y todos sus componentes.

El elemento siguiente dentro de la arquitectura es el *Data Warehouse*, que básicamente es la base de datos donde residen los datos de la organización. Más adelante se explicará en detalle la estructura de este elemento, en la sección de Modelo Dimensional.

Con respecto al motor de base de datos que se podría usar en el sistema se sugiere usar Microsoft SQL Server 2008. Según el área de sistemas de la empresa, actualmente cuentan con una licencia de uso. SQL Server es un motor robusto que cuenta con una gama de herramientas que pueden servir mucho en su implementación; por eso se recomienda usarlo.

El último elemento de la arquitectura es la aplicación para el análisis de datos y la creación de reportes y *dashboards*, la herramienta más ocionada para ocupar este lugar se llama Tableau Software. En un capítulo de este trabajo de grado se explica cómo se seleccionaron las herramientas que compondrían el sistema de inteligencia de negocios.

## 4.2 Modelo Dimensional

El modelo dimensional es el corazón de un data warehouse; de este elemento depende la velocidad, claridad y veracidad en los datos que van a ser consultados más adelante. El modelo dimensional fue diseñado con base en las necesidades expresadas por las personas entrevistadas en la fase de toma de requisitos, de manera que sí cumpliera con las expectativas de los futuros usuarios del sistema de inteligencia de negocios.

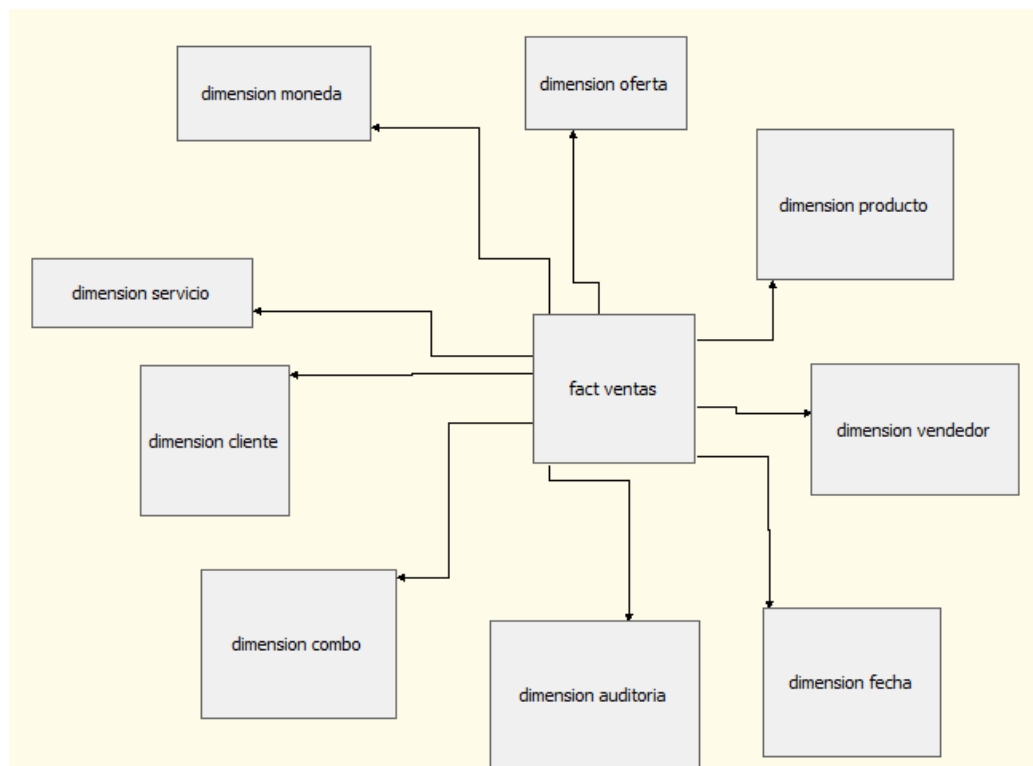
Para el modelo del diseño dimensional existen 4 pasos básicos que se deben seguir: elegir un proceso del negocio, elegir la granularidad, establecer las dimensiones y por último los *facts* asociados al proceso que se está modelando.

Con respecto a la elección del proceso de negocio no hubo mucho que discutir o decidir, desde un principio se había dicho que solo se modelaría el proceso de ventas.

La granularidad, por otro lado, fue algo completamente diferente, porque depende del proceso como tal y de los datos que tengan almacenados. Cuando se está decidiendo que granularidad se va a usar en el modelo, siempre se debe optar por usar la menor posible, de manera que no se pierdan datos y toda la información

relacionada con la venta tenga sentido. Para el caso que se iba a analizar era bastante claro que la granularidad consistía en cada ítem de factura, ya que en una venta es el mínimo nivel de detalle que se tiene respecto a la transacción como tal.

Después de determinar la granularidad del modelo se empezaron a identificar las dimensiones que iban a formarlo. A continuación se puede ver un gráfico que muestra cómo se estructuró el modelo.



### Ilustración 3-Modelo Dimensional

La dimensión moneda, fecha, cliente y vendedor hablan por sí solas, simplemente indican cuál es la unidad monetaria con la que se realizó la venta, cuál fue la fecha, a qué cliente se le vendió y por último que vendedor de la compañía fue el que la realizó.

Las otras dimensiones requieren un poco más de explicación, la dimensión de oferta indica que una venta se realizó con una promoción o una oferta especial. Esta se incluyó en el modelo para que en Socoda puedan analizar la efectividad de sus promociones y con base en esto tomen decisiones.

Lo que debería haber sido únicamente dimensión producto, se separó en Servicio, Producto y Combo. Esto justificado en que primero que todo Socoda ofrece servicios de instalaciones, transporte, arrendamientos, entre otros, que no son productos ni combos y que deben ser tomados en cuenta en las ventas, por lo que se crea la dimensión de servicios para tener disponibilidad frente a esa información.

La diferencia entre producto y combo radica en la necesidad de analizar las ventas según los productos vendidos no combos, los combos vendidos y también productos que incluyen los componentes de combos. Actualmente las personas que se encargan de realizar los reportes de ventas no tienen la capacidad de generar reportes que separen los combos en los productos que lo componen, por eso se crea esta separación para habilitar esta capacidad y que, de esta forma, puedan analizar realmente qué productos son los que más se venden, sea a través de combos o como producto independiente.

La última dimensión que falta por explicar es la dimensión de auditoría. Esta dimensión existe para hacer un seguimiento de cómo están llegando los datos al data warehouse de Socoda; en ella existen datos como versión del ETL, fecha de extracción, número de errores encontrados, entre otros.

#### 4.2.1 Atributos de las dimensiones

Estos son los atributos que contiene cada dimensión. A partir de este momento se analiza detalladamente el ERP de la compañía ya que se establecen las integraciones definiendo para cada dimensión la tabla de donde se obtendrían, el tipo de dato asociado y el nombre de la columna en la tabla origen en el ERP de Socoda:

**Tabla 1-Dimensión Producto**

Producto	Tablas Asociadas: pt_mstr	
Nombre Atributo	Tipo de Dato	Columna en el origen de datos
Línea de Producto	varchar(50)	pt_prod_line
Grupo Producto	varchar(50)	pt_group
Tipo de Producto	varchar(40)	pt_part_type
Segmento Producto	varchar(24)	pt__chr01
Estado Producto	varchar(24)	pt_status
Color Producto	varchar(24)	pt_promo
Alto	decimal(10, 3)	pt_height
Ancho	decimal(10, 3)	pt_width
Profundo	decimal(10, 3)	pt_size
Referencia Producto	varchar(18)	pt_part
Código de Barras Producto	varchar(45)	pt_article
Descripción Producto	varchar(60)	pt_desc1+" "+pt_desc2
Pertenezco A Combo	char(2)	
Producto Especial	varchar(20)	pt_draw
Unidad de Medida	varchar(45)	pt_um

La dimensión de combo tiene la misma estructura que la de producto ya que finalmente es otro producto que se vende, simplemente se cambian los nombres de los atributos para diferenciarlos de los de producto, para evitar confusiones en los análisis que incluyan información de las dos dimensiones.

**Tabla 2-Dimensión Cliente**

<b>Cliente</b>	<b>Tablas Asociadas: cm_mstr + ad_mstr</b>	
<b>Nombre Atributo</b>	<b>Tipo de Dato</b>	<b>Columna en el origen de datos</b>
Clase de Cliente	varchar(40)	cm_class
Región Cliente	varchar(24)	cm_region
Identificación Cliente	varchar(18)	ad_gst_id
Ciudad Cliente	varchar(30)	ad_city
País Cliente	varchar(24)	ad_country
Departamento Cliente	varchar(24)	ad_state
Nombre Cliente	varchar(28)	cm_sort
Canal Cliente	varchar(40)	cm_type
Negocio Cliente	varchar(30)	cm_char
Razón Social Cliente	varchar(70)	ad_name
Tipo de Identificación Cliente	varchar(18)	ad_pst_id
Tipo de Persona Cliente	varchar(20)	ad_misc3_id
Código Cliente	varchar(8)	cm_addr

**Tabla 3-Dimensión Vendedor**

<b>Vendedor</b>	<b>Tablas Asociadas: sp_mstr</b>	
<b>Nombre Atributo</b>	<b>Tipo de Dato</b>	<b>Columna en el origen de datos</b>
Nombre Vendedor	varchar(35)	sp_sort
Zona Vendedor	varchar(24)	sp_territory
Codigo_Vendedor	varchar(6)	sp_addr



**Tabla 4-Dimensión Moneda**

Moneda	Tablas Asociadas: exr_rate	
Nombre Atributo	Tipo de Dato	Columna en el origen de datos
Nombre Moneda	varchar(25)	exr_curr1
Fecha Medida Moneda	datetime	exr_start_date
Valor en Pesos Moneda	float	exr_rate2

**Tabla 5-Dimensión Servicio**

Servicio	Tablas Asociadas: trl_mstr	
Nombre Atributo	Tipo de Dato	Columna en el origen de datos
Descripción Servicio	varchar(24)	trl_desc
Natural key servicio	varchar(5)	trl_code

**Tabla 6-Dimensión Fecha**

Fecha	
Nombre Atributo	Tipo de dato
Fecha	date
Día	tinyint
Número_De_Mes	tinyint
Año	smallint
Día_De_La_Semana	varchar(15)
Mes	varchar(15)
Día_De_Fiesta	varchar(2)
Fin_De_Semana	varchar(2)
Día_Del_Año	smallint
Semana	tinyint
Semestre	tinyint
Bimestre	tinyint

Trimestre	tinyint
-----------	---------

Esta dimensión no tiene tabla origen, su contenido sería generado por el equipo de inteligencia de negocios, su finalidad es proveer al usuario de la posibilidad de filtrar sus reportes por fechas.

**Tabla 7-Dimensión Auditoría**

Auditoría	
Nombre Atributo	Tipo de Dato
Calidad Cualitativa	varchar(15)
Calidad Cuantitativa	int
Compleitud Cualitativa	varchar(15)
Compleitud Cuantitativa	int
Validez Cualitativa	varchar(15)
Validez Cuantitativa	int
Pruebas Fallidas	varchar(2)
Fecha Hora de Extracción	datetime
Fecha Hora de Limpieza	datetime
Versión ETL	varchar(45)

Esta dimensión es poblada con datos que son proveídos por el proceso ETL, donde se ve reflejado el resultado de su ejecución.

La dimensión oferta no está incluida en esta sección debido a que la compañía no tiene una estructura clara para la creación de promociones, además no tiene registro de productos se han vendido en promoción o no, por lo que simplemente se incluyó en el diseño global del modelo pero no está detallada, en el momento en que se tenga más información y se estandarice el proceso de la creación de promociones se pueden agregar los atributos que se consideren pertinentes.

Para terminar con el proceso de modelado dimensional, los facts o mediciones, y las dimensiones degeneradas son los siguientes:

**Tabla 8-Fact Ventas**

Ventas	Tablas Asociadas: ih_hist + idh_hist		
	Nombre Atributo	Tipo de Dato	Nombre Columna
Ventas Brutas	float	Calculado	Precio Unitario * Cantidad
Cantidad Vendida	float	idh_qty_inv	

Precio Unitario	float	Calculado	$idh\_price / ( 1 - ( idh\_disc\_pct / 100) )$
Porcentaje Descuento Elemento	float	idh_disc_pct	
Valor Descuento Elemento	float	Calculado	idh_disc_pct * Precio Unitario * Cantidad Vendida
Porcentaje Descuento Factura	float	ih_disc_pct	
Valor Descuento Factura	float	Calculado	(Ventas Brutas - Valor Descuento Elemento)*Porcentaje Descuento Factura
Ventas Netas	float	Calculado	Ventas Brutas - Valor Descuento Elemento - Valor Descuento Factura
Numero Factura	varchar(8)	idh_inv_nbr	Dimensión Degenerada
Orden de Venta	varchar(8)	idh_nbr	Dimensión Degenerada
Compañía	varchar(8)	idh_domain	Dimensión Degenerada
Id Producto	varchar(18)	idh_part	
Id Combo	varchar(18)	idh_part	
Id Vendedor	varchar(8)	idh_slspn	
Id Fecha	date	ih_inv_date	
Id Servicio	varchar(2)	ih_trl1_cd, ih_trl2_cd, ih_trl3_cd	
Id Cliente	varchar(18)		

Id Moneda	varchar(3)	ih_curr	
Id Oferta			
Id Auditoria	smallint	No aplica	El etl la asigna

Los facts fueron determinados por la información que se tiene y se quiere analizar. En esta tabla también se observan las dimensiones degeneradas que no son más que atributos que también sirven como filtros y agrupaciones en los datos, sin embargo normalmente son códigos operacionales que están relacionados con el proceso, pero no tienen cabida en ninguna de las dimensiones que se han considerado. No justifica crear una tabla completa de dimensión para un solo atributo y por eso se incluyen directamente en tabla de los facts.

Por último están los atributos que empiezan con las letras "Id" que se muestran en la tabla, estos constituyen los identificadores de cada dimensión tal cual como existen en el ERP de Socoda, la idea es que en el sistema se remplacen por números enteros que se convertirán en las "surrogate keys" del BI, que simplemente son nuevos identificadores pero que son administrados por el sistema de inteligencia de negocios. Los "Ids" originales de Socoda se almacenan también en el data warehouse, pero solo se utilizan para identificar como se debe realizar la relación entre fact y dimensión. Cuando la identificación se ha terminado se enlazan utilizando los "surrogate keys".

### 4.3 ETL

El ETL como se dijo anteriormente, es un proceso que se realiza para la integración y sincronización de la base de datos del ERP con el *data mart*, con lo cual no sólo se cumple con el objetivo de ingresar la información a éste, sino que también contribuye al aseguramiento debe asegurar la calidad de la información contenida allí.

En esta etapa del proyecto se requiere realizar un análisis profundo de la operación del ERP ya que el BI debe reflejar el resultado de las transacciones que está entrando en el mismo. Si el BI no refleja correctamente esos movimientos el proyecto nunca podrá tener éxito.

Siguiendo con la metodología, este sistema se diseñó haciendo una separación de los diferentes procesos según sus funciones: (Extracción, Transformación y Carga).

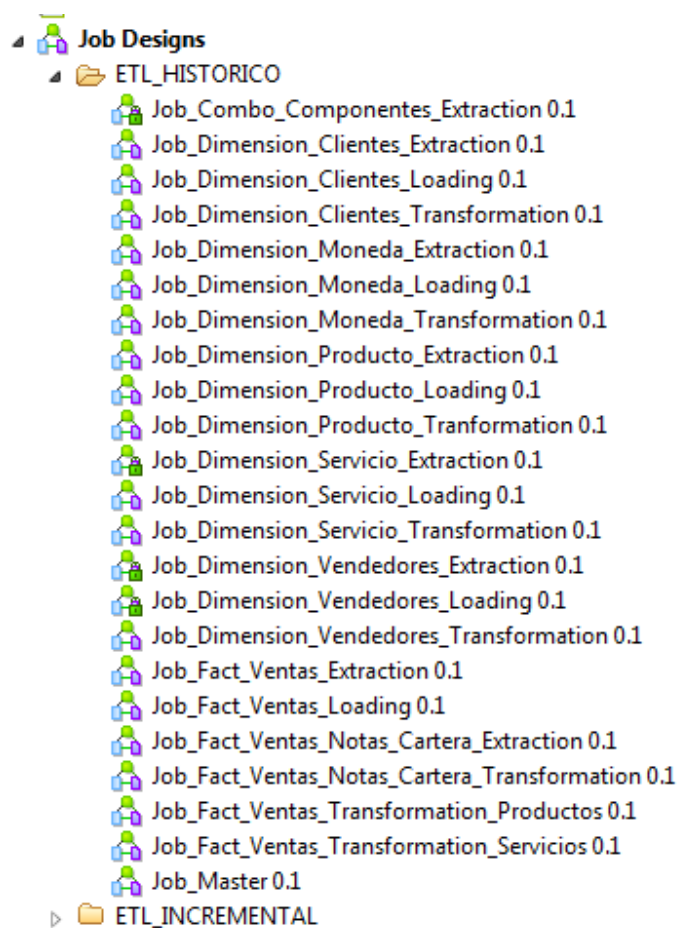
Cuando se diseña un ETL para cualquier proceso del negocio, en realidad se deben realizar dos diseños, uno para la carga histórica que es un volcado de toda la información que se necesita de los orígenes de datos hacia el *data warehouse*. Y por otro lado se crea un diseño incremental que transfiere los registros cambiados o nuevos que no se encuentran todavía en la base de datos para mantener el sistema de inteligencia de negocios actualizado.

Se empezó atacando el ETL histórico porque serviría de base para hacer el incremental. El histórico e incremental básicamente se diferencian en su extracción, las

transformaciones y cargas de los datos son prácticamente iguales. Como se explicó antes, el histórico extrae toda la información de las tablas origen, mientras que el incremental solo extrae los registros de dimensiones que están identificados en unas tablas especiales que se crearían en la base de datos transaccional de Socoda, y que se alimentan mediante *triggers* para que sólo contengan los registros actualizados y nuevos.

Por parte de los *facts*, las consultas de extracción se modificaron para que sólo solicitaran la información del día en curso. El ETL está diseñado para ejecutarse automáticamente todos los días cerca de las 11 de la noche por lo que en ese momento se asume que no hay más actividad en la compañía y todos los datos necesarios van a ser extraídos. Las tablas de las que se extraen las dimensiones se borran automáticamente todos los días para no realizar un doble procesamiento de datos.

Para cada dimensión se creó un *job*, que es la manera como Talend llama a un subproceso de integración de datos. El nombre de cada Job incluye una palabra clave para mostrar cuál de las tres funciones básicas descritas anteriormente está realizando. Además de indicar qué dimensión o fact tiene como objetivo dicho Job, a continuación se puede ver el listado de Jobs diseñado para el ETL histórico del proceso de ventas.



**Ilustración 4-Jobs ETL Histórico**

Como se puede ver en la imagen se incluyeron algunos *Jobs* adicionales para transformaciones, y procesos intermedios que eran necesarios dada la lógica del negocio. Cada uno de estos *Jobs* se muestra a continuación:

#### 4.3.1 **Jobs Extracción**

Los *Jobs* de extracción tienen el objetivo de traer datos de la base de datos origen y almacenarlos en el área de trabajo para la transformación (también llamada *Staging*), que para este caso serán archivos planos en formato csv separados por punto y coma. Esta área se justifica en su utilidad para auditoría, verificación de errores y recuperación en caso de cualquier problema. En estos *Jobs* se realiza el conteo y almacenamiento de las estadísticas de extracción de datos que se almacenan en la base de datos del ETL. Obviamente cada subproceso de extracción obtiene los datos de tablas diferentes y la estructura de los archivos está enmarcada por los datos que se extraen, que más adelante se insertarán en las dimensiones o *facts* que se mencionaron anteriormente.

- Extracción Clientes: El objetivo de este *Job* es la extracción de los clientes de la base de datos del ERP. La información de esta dimensión se encuentra dividida en dos tablas que se llaman *cm\_mstr* y *ad\_mstr*. Adicionalmente en Socoda existe una división en varias compañías, por lo que se incluye en el ETL el concepto de compañía en los clientes para hacer una buena diferenciación de los mismos.

Modelos:

- Anexo 20-Extracción Clientes Incremental
- Anexo 21-Extracción Clientes
- Extracción Moneda: Esta extracción es mucho más sencilla que la de clientes, no se tienen varias tablas y el concepto de compañía no se toma en cuenta. Similarmente al proceso de los clientes, se crea el *staging* y se almacenan las estadísticas de extracción.

Modelos:

- Anexo 23-Extracción Moneda Incremental
- Anexo 24-Extracción Moneda
- Extracción Productos: La extracción de productos es similar a la de clientes en que se debe tomar en cuenta el manejo de compañías. En la parte inferior de la imagen que muestra el proceso se puede ver una pequeña extracción adicional; este pequeño subproceso cumple con la función de extraer todos los productos que forman parte de combos, para hacer posible asignación del atributo “pertenezco a combo”, de cada producto.

Modelos:

- Anexo 27-Extracción Productos Incremental

- Anexo 28-Extracción Productos
- Extracción Servicios, Extracción Vendedores , Extracción Ventas y Extracción notas de cartera : Estos procesos no tienen ningún elemento relevante por lo que no se realiza una explicación detallada.

Modelos:

- Anexo 29-Extracción Servicios Incremental
- Anexo 30-Extracción Servicios
- Anexo 31-Extracción Vendedores Incremental
- Anexo 32-Extracción Vendedores
- Anexo 33-Extracción Ventas Incremental
- Anexo 34-Extracción Ventas
- Anexo 25-Extracción Notas de Cartera Incremental
- Anexo 26-Extracción Notas de Cartera
- Extracción Componentes Combo-Factura: La siguiente extracción es la más complicada de todas las que se diseñó. Como su nombre lo indica consiste en la consulta de todos los combos que se han vendido y su descomposición en productos, lo que incluye la cantidad de cada componente, la asignación de costos, precio unitario, el valor de las ventas por cada componente de combo y la cantidad de combo que corresponde a la cantidad original del combo que se vendió. La inclusión de factura en el nombre del proceso se da porque esta descomposición no es estándar, es decir algunas veces un combo puede estar compuesto por unos productos y otras veces puede estarlo por otros. Por eso cada factura que posea uno o varios combos puede tener una composición diferente a otra que incluya el mismo combo. Esta extracción se creó para poder realizar esa separación en la transformación de las facturas.

En esta extracción se involucran no solo productos y combos sino que también se usan tablas financieras, tablas de almacenes de costos y tablas con información de historiales de facturas; ahí radica su complejidad y sus múltiples consultas. Toda esta información se unió en un solo staging para hacer más ágil y sencilla la transformación de los datos en el Job Transformación Ventas-Productos que se explicara en el próximo apartado.

Modelos:

- Anexo 27-Extracción Productos Incremental
- Anexo 28-Extracción Productos

### 4.3.2 Jobs Transformación

Los procesos de transformación se caracterizan por cambiar la estructura de la información para que se acomode a las dimensiones y a los facts que se establecieron en el modelo dimensional, garantizando siempre la coherencia frente a las transacciones del ERP. Además realizan diferentes tareas que están asociadas a la verificación de calidad de los datos como se explica en el marco teórico de este trabajo de grado.

El ETL incluye tareas de completitud y validación de datos; las de completitud verifican que atributos importantes no estén en blanco, mientras que las de validación se aseguran que los valores de un determinado atributo si sean los esperados, normalmente los valores esperados están almacenados en maestros que dentro de nuestro sistema son llamados "*Decoding*", esos maestros contienen todos los valores aceptados por el sistema, para un atributo en especial. En caso de que exista algún error ya sea de completitud o validación el método aplicado consiste en almacenarlo y reemplazar el valor del atributo por un indicador que muestre que los datos se corregirán más adelante. El ETL no debe detenerse puesto que se perdería la actualización de muchos datos, por eso se guarda el error, y se continúa con el proceso normal de ejecución. Los errores son almacenados en archivos planos y en la base de datos del ETL para su consulta y posterior corrección por parte del área de TI.

En todos los *Jobs* de transformación se parte del staging de la extracción respectiva y los datos se van transformando hasta que se almacenan en otro archivo de *staging* para guardar el estado de los mismos antes de su carga en la base de datos.

#### 4.3.2.1 Transformaciones de dimensiones

- Transformación Clientes: Los clientes constituyen la dimensión con más pruebas de validación y completitud, por ese motivo esta transformación es compleja y su tamaño es bastante grande. Se realizan pruebas en 7 atributos diferentes para que la calidad de los datos sea la indicada. Los atributos mencionados anteriormente son: Canal, Ciudad, Departamento, Clase de cliente, Región, Tipo de persona y Negocio.

La extracción de los clientes explicada anteriormente culminaba con un *staging* que estaba dividido en 4 archivos: *cm\_mstr\_socoda*, *cm\_mstr\_IYR*, *ad\_mstr\_socoda* y *ad\_mstr\_IYR*. En el comienzo de este *Job* se integran los datos provenientes de esas 4 fuentes, haciendo prioritarias las versiones de clientes de las diferentes compañías. Además se unen los atributos de la tabla *cm\_mstr* con los de *ad\_mstr*, para tener una fila por cada cliente con todos los atributos necesarios. Si existe algún cliente que no esté presente en las dos tablas, dicho cliente no formara parte de la dimensión. En este caso se almacena esta inconsistencia tanto en planos como en la base de datos del ETL, y se continúa con el cliente siguiente.

Después del proceso de integración de las fuentes de los clientes siguen las pruebas de validación y completitud. Terminadas las pruebas el resultado final se



almacena en un nuevo archivo de *staging* para que sea cargado a tablas en el siguiente proceso.

Modelos:

- Anexo 37-Transformación Clientes

- Transformación Moneda: Una de las transformaciones más sencillas del ETL. Solo incluye una verificación de que el nombre de moneda si esté de acuerdo al los estándares de la compañía, posteriormente se llevan los datos a *staging*. De tener ese campo vacío o si tiene un nombre invalido, se almacena el error y se reemplaza por "Sin Nombre de Moneda" para que sea corregido por el área de TI más adelante.

Modelos

- Anexo 38-Transformación Moneda

- Transformación Productos: La transformación de productos comienza con una integración similar a la de los clientes, en el hecho de que contiene la lógica para determinar la versión indicada del producto según la compañía que tiene asociada. Tal como en los clientes, la versión prioritaria de un producto es la que pertenece a la compañía principal, los productos de las demás compañías sólo pasan a través del ETL si no se encuentran en la principal.

Este *Job* contiene pruebas para los atributos grupo, tipo, color y línea de producto. El comportamiento de la lógica es similar al de los otros procesos, es decir almacenamiento del error y continuar.

La mayor particularidad de esta transformación recae en el hecho que incluye la determinación si un producto es o no combo, además indica si un producto es usado o no como componente por lo menos de un combo de los que se ofrecen. Por eso la salida de esta transformación son 4 archivos de *staging*: De de combos, de productos ordinarios, de productos que son componentes de combos marcados con pertenencia a combo y de productos que son componentes de combo pero que están marcados como no pertenecientes a combos para aquellas ventas en las que se vendan como productos ordinarios.

Modelos

- Anexo 39-Transformación Productos

- Transformación Servicios: La transformación más sencilla de todas, simplemente se organizan los atributos en el orden y estructura adecuada para la subida a base de datos.

Modelos

- Anexo 40-Transformación Servicios

- Transformación Vendedores: Similar a la de moneda, se verifica el atributo zona del vendedor y se crea el *staging* que corresponde.

#### Modelos

- Anexo 41-Transformación Vendedores

#### 4.3.2.2 Transformaciones de *Facts*

Las transformaciones de facts son diferentes a las de dimensiones. Estas adicionalmente contienen una asignación de las claves foráneas que unen los registros con las dimensiones adecuadas. En este trabajo de grado solo se incluye una tabla de *fact* que contiene los registros de las ventas, sin embargo para que fuera más ordenado y sencillo el diseño del ETL se crearon 3 procesos según el tipo de registro que se está transformando (Ventas-Productos, Ventas-Notas Cartera y Ventas-Servicios). Cada tipo de registro tiene sus particularidades y en ellas radican las variaciones en la transformación de los datos.

- Transformación Ventas-Productos: En esta transformación se procesan los registros de ventas de los productos de la compañía. De todos los procesos de ETL es el más complejo y que más información manipula. El *Job* comienza con la extracción de los registros del archivo de *staging* y luego filtra los registros para que solo pasen por el aquellos que son ventas de productos o combos.

Después de tener separados los registros que deben pasar por la transformación se calculan el precio unitario, ventas brutas, valor descuento factura, valor descuento elemento y ventas netas basados en la cantidad, porcentaje descuento elemento, porcentaje descuento factura y una columna que se llama *idh\_price* que es como un precio unitario, pero que ya tiene descuento asociado. Las fórmulas para calcular estas variables son:

Ventas Brutas	Precio Unitario * Cantidad
Precio Unitario	Porcentaje Descuento Elemento / ( 1 - (Porcentaje Descuento Elemento / 100) )
Valor Descuento Elemento	Porcentaje Descuento Elemento * Precio Unitario * Cantidad Vendida
Valor Descuento Factura	(Ventas Brutas - Valor Descuento Elemento)*Porcentaje Descuento Factura
Ventas Netas	Ventas Brutas - Valor Descuento Elemento - Valor Descuento Factura

La unión de todas las variables mencionadas conforma los facts que definen las ventas.

Con los *facts* listos la actividad que falta es la asignación de claves foráneas para unir los *facts* con las respectivas dimensiones. En la tabla origen existen columnas con códigos operacionales que cumplen con este mismo objetivo, pero para que el *data warehouse* funcione correctamente la base de datos no debe usar esos mismos códigos, por eso se reemplazan por una clave foránea propia del BI que mejora el desempeño y hace fácil la administración.

La tarea de reemplazo normalmente es muy rutinaria, para las dimensiones Cliente, Vendedor y Moneda el proceso es exactamente igual, se busca en las dimensiones ya procesadas el código operacional que viene de la tabla origen, luego se reemplaza esa columna por otra que contenga un entero que corresponde a la generada por el BI. En caso de que se presente un error se sigue con el procedimiento habitual, la única variación respecto el procesado de las dimensiones, es que como estamos asignando números enteros, se asigna -1 o -2, que son filas especiales en las dimensiones que reflejan las inconsistencias en los procesos de reemplazo.

Las dimensiones Combo y Producto tienen un tratamiento especial. Tomando en cuenta que cada fila corresponde a un ítem de factura, en caso de combos se debe reemplazar la fila que viene de la tabla *idh* por un número *x* de filas, donde *x* corresponde a la cantidad de elementos que componen el combo. Para hacer esta asignación primero se verifica si la fila contiene un combo, de no contenerlo es porque es un producto ordinario y se realiza el reemplazo normal. De ser un combo se buscan los componentes que lo componen y se adicionan al flujo las filas respectivas.

Las filas nuevas pasan por un proceso en el que se recalculan los *facts* asignándoles nuevos valores para que sean coherentes. Cada *fact* se multiplica por el cociente entre el costo del componente de combo y el costo del combo como tal, de esa forma los valores si representan verdaderamente el combo dentro de la tabla de facts. Cada fila de componente también tiene la clave foránea del combo para que pueda analizarse desde el punto de vista de productos y combos, según desee el usuario final.

Las Ventas de productos no tienen asociado un servicio específico, por lo que todos los registros apuntan a una fila especial dentro de la dimensión que las etiqueta como venta de producto o combo.

Para finalizar con el *Job* se ajustan las dimensiones degeneradas Unidad de Negocio, Unidad de Negocio Combo y se culmina con el proceso.

## Modelos

- Anexo 43-Transformación Ventas-Productos

- Transformación Ventas-Notas Cartera: Las notas de cartera representan ajustes que se realizan en la contabilidad de la empresa, pero que afectan el valor de las estadísticas de ventas. Están relacionadas con las mismas dimensiones que las ventas de productos a excepción de la dimensión producto/combo ya que no tienen asociado un producto o combo en especial. Para mostrar este hecho todas las notas de cartera apuntan a la una fila especial en su Id Producto e Id Combo. Dicha fila no contiene información de un producto o combo, sino que tiene en cada columna un valor “Nota de Cartera” o “NC” dependiendo de la longitud asociada.

La transformación como tal es muy similar a la de Ventas-Productos. Como ya se explicó se elimina todo el subproceso de asignación de productos y combos porque no es necesario. El resto del proceso es prácticamente igual.

Las notas de cartera tienen un número de documento asociado dentro del ERP que se almacena en el campo Factura y Orden de Venta explicados anteriormente. También vale la pena resaltar el hecho de que no tienen descuentos, por lo que todos los atributos relacionados tienen un valor de 0% o 0 según sea el caso.

Tal como en las ventas de productos la dimensión servicio no aplica para las notas de cartera, por eso se les asigna otra fila especial que las etiqueta como Notas de Cartera en esa dimensión.

Modelos

- Anexo 42-Transformación Ventas-Notas Cartera

- Transformación Ventas-Servicios: Los servicios constituyen un caso similar a las notas de cartera, no tienen un producto/combo asociado, no tienen descuentos, pero si están asociados a facturas y a órdenes de venta dentro del ERP.

Son los únicos registros que de verdad hacen uso de la dimensión servicio. El resto de dimensiones y facts se calculan y asignan normalmente.

Modelos

- Anexo 44-Transformación Ventas-Servicios

### 4.3.3 Jobs Carga

Los procesos de carga son muy similares entre sí. En el diseño del ETL todos los procesos de carga histórica son exactamente iguales, consisten en la extracción de los datos de un archivo plano de *staging*, y un volcado a tabla objetivo.

Por parte del incremental si hay algunas diferencias, todo lo relacionado con ventas se carga como en el histórico. Pero en lo que se refiere a dimensiones el comportamiento no es el mismo. Puede darse el caso que no sean registros nuevos para la dimensión sino que son cambios en los registros existentes, por eso primero se busca el registro en la tabla correspondiente, si el registro es encontrado se actualizan todos sus atributos, de lo contrario se agrega a la tabla.

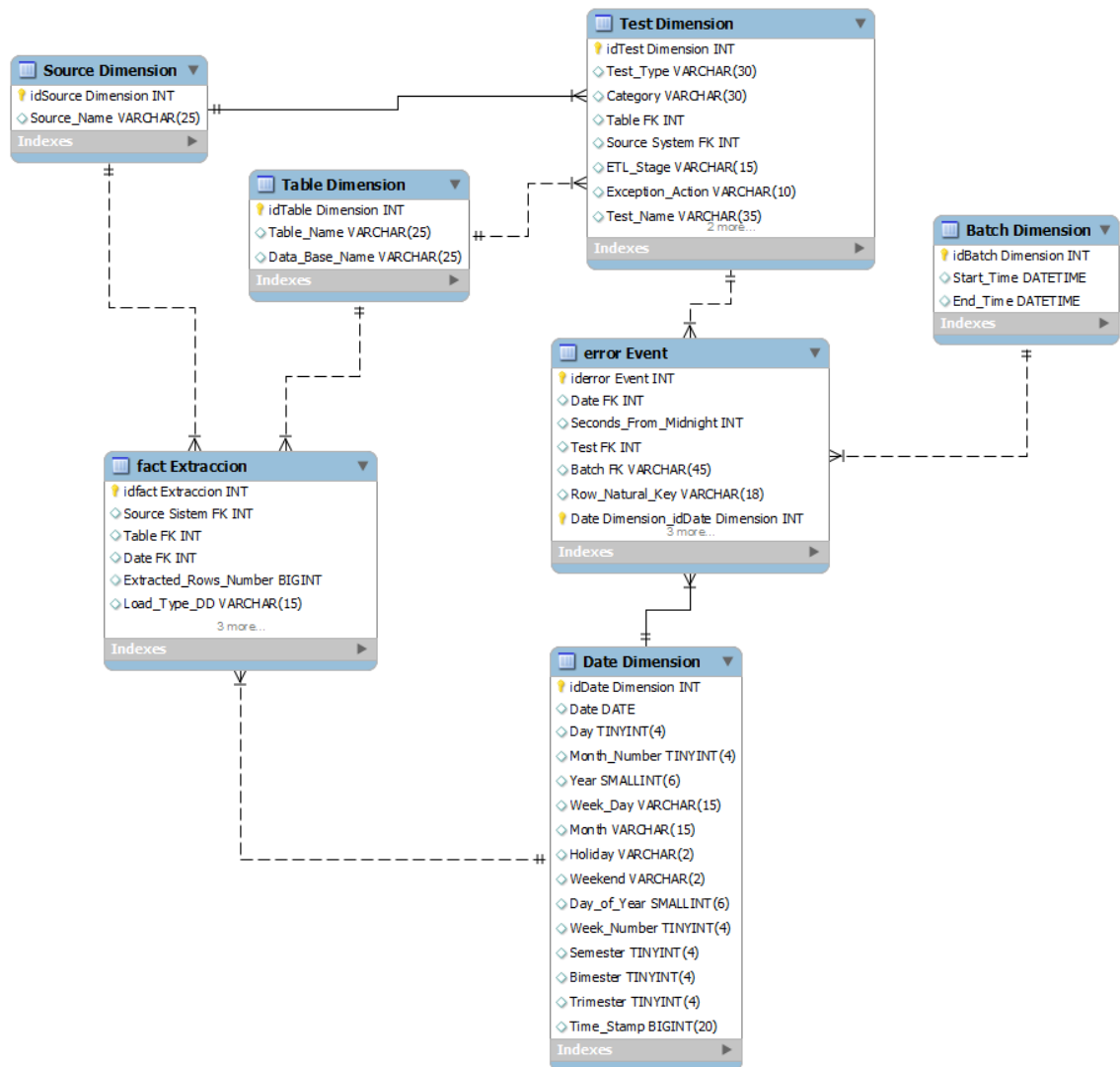
## Modelos

- Anexo 8-Carga Clientes Incremental
- Anexo 9-Carga Clientes
- Anexo 10-Carga Moneda Incremental
- Anexo 11-Carga Moneda
- Anexo 12-Carga Producto Incremental
- Anexo 13-Carga Producto
- Anexo 14-Carga Servicios Incremental
- Anexo 15-Carga Servicios
- Anexo 16-Carga Vendedores Incremental
- Anexo 17-Carga Vendedores
- Anexo 18-Carga Ventas Incremental
- Anexo 19-Carga Ventas

### **4.4 Base de datos ETL**

A partir del proceso de ETL, se produce información valiosa tanto para la compañía como para el equipo de BI. Esta información es almacenada en una base de datos independiente que no será accedida por los usuarios finales, sino por los miembros del área de TI y el equipo de inteligencia de negocios.

La siguiente figura corresponde al modelo de la base de datos:



### Ilustración 5-Modelo dimensional de ETL

Esta base de datos se realizó siguiendo un modelo dimensional en estrella, al igual que la base de datos de ventas y se compone de los siguientes elementos:

- **Fact Extracción:** Donde se almacena el número de filas extraídas de cada una de las tablas de los sistemas fuente. Esta información es útil para que el equipo de BI valide que la cantidad de datos que se están extrayendo sí estén dentro del promedio de operaciones diarias de la compañía, de no ser así, debe hacerse una revisión de los *Jobs* para identificar la falla.
- **Error Event:** Donde se almacena cada uno de los errores que se detecten en el proceso de ETL, producto de errores en los sistemas de origen. Esta tabla es un beneficio secundario del proceso y es de gran utilidad para la compañía, ya que

permite el aseguramiento de la calidad de la información almacenada en la base de datos del ERP. Los miembros del área de TI pueden extraer de estos registros la tabla y la fila donde está almacenado el dato erróneo y pueden proceder a corregirlo.

- Source dimension: Contiene la lista de las fuentes de datos de donde se extraerá la información.
- Table dimension: Contiene las tablas que hay en las fuentes de datos.
- Date dimension: Al igual que en el datamart, tiene almacenadas las fechas con el fin de que los reportes puedan tener un marco de tiempo definido.
- Batch dimension: Identifica cada corrida del proceso y registra su hora de inicio y de finalización
- Test dimension: Contiene una fila por cada filtro que se le aplica a los datos extraídos para validar que éstos estén correctos. Esta tabla permite que se identifique exactamente en qué parte de la fila donde se produjo un error, se encuentra la información incorrecta.

## 5 Selección de la herramienta de visualización

Uno de los componentes más importantes de un sistema de inteligencia de negocios es la herramienta de visualización, ya que ésta permite que el usuario acceda a la información que se refinó y organizó en las etapas anteriores del proceso de construcción del sistema. Con el fin de sugerir una herramienta de este tipo para que los empleados de Socoda accedan al data warehouse, se evaluaron tres productos: Pentaho, Jasper y Tableau. Pentaho y Jasper son las dos suites más utilizadas en el entorno para la construcción de sistemas de inteligencia de negocios, ambas cuentan con una versión comunitaria totalmente gratuita y una versión comercial. Por otro lado, Tableau es un producto relativamente joven que apenas está incursionando en el mercado.

Inicialmente se evaluaron de forma paralela tanto Pentaho como Jasper, explorando las capacidades que ofrecían para el proceso de ETL y para la visualización de datos. Los criterios de evaluación que se tuvieron en cuenta mientras se probaban estas dos soluciones fueron principalmente la usabilidad a la hora de crear reportes, precio y rendimiento. Posteriormente se descartó el de rendimiento, ya que éste depende más de la estructura de la base de datos y de las optimizaciones que se realicen en ella.

Al finalizar el período de prueba gratuito que ofrecen las compañías fabricantes de ambos software, los resultados fueron muy satisfactorios con ambas herramientas, las dos ofrecen formas fáciles de creación de reportes sin necesidad de conocimientos técnicos, que es lo que se busca con un sistema de este tipo, e incluyen una gran variedad de herramientas para apoyar el proceso de creación, mantenimiento y utilización del sistema. Cuando se evaluaron las versiones comunitarias de ambos productos, se identificó que las características claves que se habían identificado en las versiones de prueba comerciales no estaban presentes en ellas, la creación de reportes era mucho más complicada que en las otras versiones, lo que hizo que se descartaran inmediatamente éstas como opción.

Se procedió entonces a contactar a las compañías para una cotización de las herramientas. Se enviaron correos a ambas empresas, pero sólo se obtuvo respuesta por parte de Jaspersoft, los fabricantes de Jasper y se tuvo una conferencia a través de Skype con Julio Chez, asesor de ventas para Latinoamérica pero desafortunadamente, el precio de la herramienta está por fuera del presupuesto de Socoda e incluye muchos componentes que no son necesarios para el estado actual del proyecto.

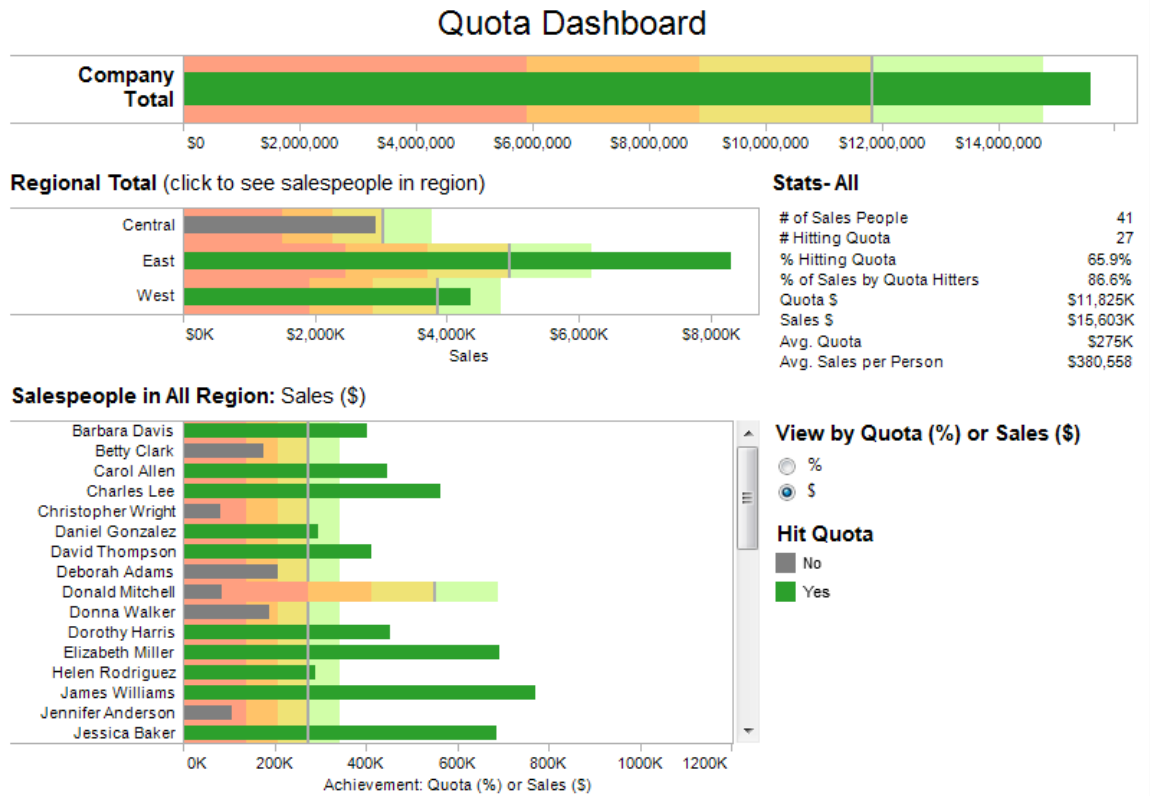
Después se estudiaron herramientas especializadas únicamente en *ad-hoc querying*, es decir, en la creación de reportes “al vuelo” de forma fácil y rápida, de acuerdo con los objetivos iniciales del proyecto y siempre teniendo en cuenta las necesidades de los usuarios finales del sistema. Entre las herramientas evaluadas, se seleccionó una herramienta llamada Tableau, producida por una empresa Estadounidense llamada Tableau Software con base en Seattle, Washington. Al igual que con Jaspersoft, se contactó via Skype al encargado de ventas para Latinoamérica, llamado Fred DeWorken,



quien hizo una demostración en vivo del software, haciendo énfasis en su facilidad de uso y su versatilidad.

Se le deja entonces a Socoda, la sugerencia de adquisición de licencias de Tableau, tanto en su versión *desktop*, utilizada para la generación de reportes, como la versión *server* para publicación y consumo de los mismos.

## 6 Ejemplo de Reporte



### Ilustración 6-Ejemplo de Reporte<sup>3</sup>

La ilustración 14 muestra cómo se ve un reporte en la herramienta de visualización de datos propuesta. En este reporte en particular se muestran las ventas de una empresa ficticia organizadas por vendedor, por región y totales. Cada uno de estos valores es comparado con un presupuesto establecido por la misma compañía y el cumplimiento o no cumplimiento del presupuesto es lo que determina el color de las barras. Puede verse como son verdes cuando se cumple y grises en caso contrario. Adicionalmente se ven diferentes colores que sirven para identificar hasta qué punto se cumplió con el presupuesto en cada barra. Rojo denota el 50% o menos, naranja 50-75%, amarillo 75-100% y verde 100-125%.

Más que concentrarse en los datos vale la pena resaltar la calidad de los gráficos, la facilidad de análisis que da la herramienta. Y el dinamismo que ofrece al permitir modificar

<sup>3</sup> Quota Dashboard <http://www.tableausoftware.com/solutions/sales-management> (27 de Mayo de 2012)

la visualización cambiando los valores por porcentajes con el control "View by Quota (%) or Sales (\$)".

## 7 Resultados

Desde la planeación del proyecto, se definió que el *fact* alrededor del cuál iba a construirse el *data mart* que operaría el sistema de inteligencia de negocios serían las ventas de la compañía. Tras un análisis de la información disponible en la base de datos del ERP de Socoda, se estableció que la granularidad correcta para dicho *fact* sería cada línea de la factura, es decir, en el *fact*, habrá una fila por cada línea que componga las facturas de ventas en Socoda. Adicionalmente, en la tabla del *fact* serán almacenadas notas de cartera, que son elementos análogos a los elementos facturados pero en lugar de representar ventas, representan devoluciones o movimientos de cartera que afectan las ventas. Y servicios que también son facturados y por lo tanto afectan las estadísticas de ventas.

Para la base de datos de ETL, se usa el *fact* extracción para llevar registro de cada vez que se ejecuta el proceso y almacenar la información de extracción de cada tabla y cada sistema de origen. También está el *fact* de eventos de error para que el área de TI de Socoda y el equipo de BI puedan analizar y corregir las fallas y problemas de calidad de información que se presenten

Una vez definidos los *facts*, a partir de las entrevistas con los potenciales usuarios del sistema de BI en Socoda, se identificaron cuáles serían los filtros que ellos usarían para la generación de reportes una vez el sistema se encuentre operando, cada uno de acuerdo a los intereses particulares de cada usuario. Se establecieron las siguientes dimensiones: producto, vendedor, fecha, combo, cliente, servicio, moneda, oferta y auditoría. El diseño que se describe a continuación incluye todas las dimensiones, excepto la de oferta y auditoría, la primera es necesaria para el área de mercadeo pero actualmente los procesos de negocio de la empresa no cuentan con datos almacenados para ella. Se prescinde de la dimensión de auditoría porque en este momento el equipo de TI de Socoda no la considera necesaria.

Según la teoría de Kimball, se definieron también las dimensiones para la base de datos de ETL así: *source*, *table*, *date*, *batch* y *test*.

Contando con los *facts* y las dimensiones se procedió a realizar el modelo dimensional del datamart de ventas, se determinó la mejor forma de relacionar los elementos para que se pueda acceder de una forma clara y coherente a la información, usando los *facts* como elementos con los que se pueden realizar operaciones matemáticas y agrupaciones y las dimensiones como filtros para darle sentido a éstas. Este diseño fue validado en numerosas ocasiones con los usuarios y el personal de TI de Socoda. De estas validaciones surgieron varios ajustes que llevaron a un mejor entendimiento de la arquitectura de información de la empresa así como de sus procesos de negocio. Finalmente se llegó a un diseño que fue satisfactorio para todos y que asegura la conformidad de los clientes y usuarios con la aplicación.

Luego de realizar el diseño y conocer a fondo la forma en la que se encuentra almacenada la información en la base de datos del ERP de Socoda, se realizó la tarea más compleja del diseño del sistema, que es el diseño del proceso de ETL, para esto fue empleada una herramienta especializada llamada Talend. Fueron diseñados dos procesos: ETL histórico y ETL incremental, éstos tienen propósitos completamente diferentes aunque en esencia son muy similares. Idealmente, el primero se correrá una sola vez para cargar toda la información que haya en la base de datos hasta el día en el que se ejecutó y el segundo se correrá diariamente para identificar las novedades y trasladarlas al *data mart*.

Dentro de todo el proyecto de diseño del *data mart*, el diseño del proceso de ETL es posiblemente el que más tiempo consume, debido a la cantidad de tareas y pruebas que se deben realizar durante éste.

El último elemento que compone un sistema de BI es la herramienta de visualización de Datos. Se escogió Tableau software, la opción que mejor se acomoda a las necesidades y el presupuesto de Socoda, ellos quedaron satisfechos con la capacidad del producto luego de ver las demostraciones tanto del personal de ventas de Tableau, como del equipo de desarrollo de BI y, de llevarse a término el proyecto, están dispuestos a invertir en licencias del producto.

## CONCLUSIONES Y CONSIDERACIONES FINALES

- Afrontar un proyecto en la vida real es muy enriquecedor para cualquier profesional que está empezando su carrera. La teoría siempre presenta escenarios ideales en los que todos los procesos están estandarizados y toda la información de una compañía se encuentra en un estado óptimo y al alcance de la mano. Sin embargo la realidad es completamente diferente.
- Este trabajo de grado demuestra que en muchas organizaciones existe gran cantidad de información que si se tiene disponible en todo momento es valiosa para evitar la toma de decisiones basadas en la intuición. El BI ayuda a que se tenga un acceso y poder sobre la información, lo que tiene como última consecuencia decisiones bien fundamentadas
- Las organizaciones siempre van a estar concentradas en las actividades que les brindan valor, el almacenamiento de los datos provenientes de los procesos de negocio normalmente es una actividad secundaria que supuestamente no trae un beneficio y por eso es subestimado dentro de muchas compañías. Lo que trae como resultado información incoherente y con problemas de calidad. Un sistema de inteligencia de negocios ayuda a los clientes a darse cuenta de este tipo de problemas y los ayuda a corregirlos.
- Un sistema de Inteligencia de negocios tiene el objetivo principal de mejorar la toma de decisiones del cliente, por medio de un acceso fácil a la información que es importante para este fin. Saber qué información es importante para ellos y que no es importante es otro factor que puede ser determinante en el éxito de un proyecto de BI.
- Para realizar un trabajo de este tipo es vital la interacción entre el equipo de desarrollo y el área de sistemas del cliente. Sin un conocimiento profundo del ERP y de las diferentes fuentes de datos, el diseño del ETL se hace imposible.
- Se recomienda informarse bien cuando se va a realizar un proyecto de este tipo. Muchos ingenieros de sistemas tienden a pensar que con solo el conocimiento de una herramienta específica es suficiente para afrontar un proyecto completo, sin embargo si en realidad se quiere construir una solución robusta, lo ideal es tener unos conceptos claros, conocer las mejores prácticas para el desarrollo de estos sistemas y luego si concentrarse en las herramientas para hacer el mejor uso de las últimas.

## 8 Bibliografía

*Hackathorn, R. (17 de June de 2010). Defining Advanced Analytics. Recuperado el 21 de March de 2011, de Beye Network: <http://www.b-eye-network.com/view/14021>*

*Kimball, R. (2008). The data warehouse lifecycle toolkit. John Wiley & sons.*

*Kimball, R., & Ross, M. (2002). The data warehouse toolkit. New York: John Wiley & Sons inc.*

*Mello, A. d. (15 de April de 2011). 5 Best Practices for Telling Great Stories with Data. Recuperado el 15 de April de 2011, de Beye Network: [http://research.b-eye-network.com/data/document.do?res\\_id=1203504036\\_523&=asrcEM\\_DWP](http://research.b-eye-network.com/data/document.do?res_id=1203504036_523&=asrcEM_DWP)*

*Ralph, K. (2004). The Data warehouse ETL Toolkkit. John Wiley & sons.*

*Stolte, D. C. (25 de April de 2011). A New Approach to Business Intelligence: Rapid-Fire BI. Recuperado el 25 de April de 2011, de Beye Network: [http://research.b-eye-network.com/data/document.do?res\\_id=1300917628\\_54&=asrcEM\\_DWP](http://research.b-eye-network.com/data/document.do?res_id=1300917628_54&=asrcEM_DWP)*

## **9 ANEXOS**

### **9.1 Informe de reunión Ana María Franco**

#### **Inteligencia de Mercados Socoda S.A**

##### **Anexo 1 - Ana María Franco**

#### **Objetivos y responsabilidades**

Anteriormente encargada de la generación de los reportes para toda la compañía. Sin embargo actualmente sus funciones solo incluyen los reportes específicos de mercadeo y el pronóstico de la demanda, los reportes que son para el área comercial son realizados por José Gabriel. El área de mercadeo se enfoca más en el campo de los productos y no tanto de clientes como lo hace al área comercial. Aparte de los reportes, están encargados de la clasificación de los artículos y de hacer un seguimiento a las ventas según esas clasificaciones que se están estableciendo.

#### **Reportes actuales**

Se reciben reportes diariamente, pero éstos tienen poca relevancia en el análisis y son usados únicamente para realizar seguimiento. Los informes más importantes para mercadeo son los de cierre de mes. Además de los reportes recibidos de la parte comercial, que consisten en ventas por tipo de cliente mensual y ventas por región entre otros, se generan en el área reportes por línea de producto, segmento (Por ejemplo Mejoramiento del Hogar: Esencial, Práctico e Innova, son diferentes para cada Negocio), grupo y tipo. Sin embargo la clasificación de los diferentes productos actualmente se está volviendo a establecer.

#### **Análisis**

Se quieren conocer los datos de las ventas de Combos con un detalle mínimo, abarcando hasta los componentes individuales que hacen parte del mismo. También se desea poder trabajar con las ventas no surtidas, que son las unidades que no fueron enviadas en un pedido debido a que no se alcanzaron a producir antes de la fecha límite de envío.

Es necesario tener en cuenta el manejo de los precios dependiendo el cliente al que se está vendiendo, con el fin de evaluar precios promedio, precio máximo y precio mínimo.

Se deben incluir los descuentos dentro de los datos para poder tomar decisiones basados en los resultados de los mismos.



Se necesita analizar las ventas en un determinado periodo, por segmento, por grupo y por tipo, también se debe poder analizar las ventas por referencia, es decir sin involucrar ninguna otra clasificación.

Con estos informes, se busca conocer el crecimiento, el cumplimiento del presupuesto la participación de las ventas de las diferentes referencias, grupos, segmentos y tipos.

### **Inconvenientes**

La principal dificultad que se tiene actualmente con los reportes es la falta de gráficas para representar los datos de forma diferente, ésto es debido a que es una tarea que consume mucho tiempo en su generación, también se mencionó que existen análisis que sería bueno hacerlos para conocer mejor el comportamiento de algunas referencias o grupos de productos de acuerdo a los tipos de clientes o zonas, pero no se realizan actualmente debido a que su elaboración es muy complicada.

No se están separando los cargos adicionales de las ventas para efectos de revisión contra el presupuesto.

## **9.2 Informe de reunión Adriana Jaramillo**

### **Directora Canal de Construcciones e Instituciones Socoda S.A**

#### **Anexo 2 - Adriana Jaramillo**

#### **Objetivos y responsabilidades**

Los objetivos de la directora de canal consisten en desarrollar el negocio de constructores y el negocio institucional enfocado a mejoramiento del hogar, materiales y acabados. También incluyen definir estrategias para que el negocio crezca desde el punto de vista de ventas, garantizar la rentabilidad esperada, garantizar los presupuestos y garantizar un recaudo de cartera entre otros.

#### **Clientes**

Los clientes de este canal son constructores e instituciones como colegios o entidades y no incluyen particulares. Existe una clasificación extra dentro de los institucionales que son colegios, centros comerciales, entre otros. Estos clientes se clasifican por "tipo de cliente" (Institucional o mejoramiento del hogar), también se clasifican por su área de influencia en locales, regionales y nacionales. Actualmente se encuentran en la labor de migrar esta clasificación únicamente a locales y nacionales.

A pesar de que los clientes de este canal son empresas constructoras, se registra como cliente cada uno de los proyectos que éstas realizan. Lo que tienen en común todos los proyectos para identificarlos como pertenecientes a un cliente es el NIT sin embargo, en algunas ocasiones cambian de NIT para la realización de obras aliándose con otras empresas.

#### **Análisis**

La dirección de canal desea conocer como ha sido el desempeño en ventas y la rentabilidad de cada constructora o de cada categoría de cliente mencionada anteriormente, filtrando también por zona y tipo de producto.

Además del análisis del desempeño de los diferentes clientes, es importante conocer cómo es el comportamiento de un vendedor, para esto se deben conocer los de presupuestos de cada uno de los vendedores y del canal. A partir de estos datos, la directora de canal puede analizar cómo está un vendedor con respecto a su presupuesto actual y con respecto a un período anterior.

Es fundamental realizar los análisis con base en las unidades y el dinero. El presupuesto lo crea mercado a nivel de producto para que luego el área comercial lo convierta a canales, vendedores y tipo de cliente.

Por último se debe tener la posibilidad de comparar los datos de las ventas históricamente ya sea a nivel anual, mensual, semestral etc.

### **Reportes actuales**

En este momento, los reportes para esta dirección se realizan por demanda y se demoran aproximadamente tres días en llegar.

La directora no consulta datos desde el ERP de la compañía (QAD), debido a la dificultad que representa obtenerlos desde ahí a pesar de contar con un usuario en el sistema, tampoco realiza cálculos adicionales ni transformaciones sobre los informes recibidos.

### **Inconvenientes**

Los reportes recibidos no se pueden visualizar de forma fácil para la directora del canal debido a que los reportes en dinero y en unidades se encuentran en hojas diferentes y lo óptimo sería tenerlos juntos. Ella tiene el conocimiento para realizar esta transformación pero no lo hace porque resulta laborioso.

Actualmente no hay una buena clasificación de los productos en Socoda, lo que hace más complicado el análisis.

### **9.3 Informe de reunión Andrés Delgado**

#### **Director Canal de Exportaciones Socoda S.A**

##### **Anexo 3 - Andrés Delgado**

#### **Análisis**

La necesidad más importante del área de exportaciones consiste en poder tener los datos de las ventas involucrando el País donde se vendieron los productos, el canal al cual pertenecen los clientes y la línea del producto también.

Las ventas que se manejan en exportaciones cuentan con descuentos, combos y el precio puede variar dependiendo del cliente, por lo que estos factores se deben tomar en cuenta a la hora de realizar la base de datos. También se debe resaltar la importancia de la posibilidad de la visualizar las ventas no solo en pesos si no también en dólares, ya que así es que se venden los productos en otros países.

Se tiene presupuesto por producto, por canal y por país. Por lo que se deben incluir presupuestos para hacer comparaciones en los tres niveles.

#### **Reportes actuales**

Está recibiendo un reporte con las ventas mostrando las exportaciones, las ventas por línea y los presupuestos asociados, no está la discriminación por país o canal. Actualmente los datos que les entregan en los reportes no se encuentran listos para hacer el análisis. Se necesitan hacer unas operaciones extras para llegar la forma deseada. Los reportes se reciben diariamente, sin embargo lo más importante es el cierre de mes.

## **9.4 Informe de reunión Andrés Restrepo**

### **Director Canal de Distribuidores Socoda S.A**

#### **Anexo 4 - Andrés Restrepo**

#### **Objetivos y responsabilidades**

El canal se encarga de administrar los distribuidores de los productos de Socoda a nivel nacional. Ellos trabajan con depósitos, ferreterías, entre otros.

El canal tiene un grupo de vendedores a nivel nacional los cuales son los encargados de estar en contacto con los diferentes distribuidores, y por lo tanto el director del canal tiene la responsabilidad de estar al tanto del trabajo de los mismos. Que incluye visitas y un cumplimiento de presupuestos para asegurar las ventas de la compañía en este canal.

Los objetivos a largo plazo del canal están plasmados en el plan "Saeta", que busca que los distribuidores más grandes de las diferentes zonas se encarguen de realizar las ventas a otros más pequeños, ofreciéndoles el mismo precio que les ofrece Socoda, esto con el fin de disminuir y centralizar el esfuerzo de ventas en las diferentes regiones. Un aspecto que hay que tener en cuenta para el análisis del plan es identificar siempre quién es el que efectúa la venta a los distribuidores pequeños para poder identificar y controlar cuando uno de los compradores cambia su proveedor por otro de los clientes de Socoda o si éste se perdió como cliente.

#### **Criterios de desempeño**

Los indicadores de desempeño que se manejan actualmente para los vendedores son el cumplimiento de presupuesto, el recaudo de cartera y el cumplimiento de visitas programadas a los clientes.

#### **Reportes Actuales**

Actualmente, se reciben informes diarios, pero el que realmente tiene peso y se analiza más a fondo es el de cierre de mes. Si se solicitan reportes adicionales, se pueden tardar medio día en recibirlos.

Los informes son filtrados por canal y se resumen por vendedor, en general no se requiere realizar transformaciones adicionales.

Adicionalmente, se envía a los vendedores de forma semanal los resultados acumulados de la facturación comparados con el presupuesto, para que evalúen su desempeño hasta la fecha y propongan posibles alternativas para mejorarlo en caso de ser necesario.

Por su parte, los vendedores pasan un informe indicando qué visitas van a realizar durante la semana y con qué propósito. También deben enviar un informe mensual indicando los resultados de su zona, analizando las causas de éstos.

A cada una de estas visitas, los vendedores deben llevar las ventas del cliente, agrupadas de forma trimestral y mensual, comparándolas con el año anterior, además de los informes de cartera del cliente.

Lo ideal es que al cliente se le muestre el comportamiento de ventas del portafolio, separándolo por grupo, por tipo de producto y por referencia, haciendo una comparación con las ventas en períodos anteriores y comparándolo con el promedio de ventas de la zona. En ningún momento se le puede mostrar a un cliente el comportamiento del otro cliente, solamente se debe tratar el comportamiento de la zona en general.

### **Análisis**

Los distribuidores se dividen entre Aliados y Mayoristas por lo que esta clarificación debe formar parte de la información relacionada con los clientes. Se necesita conocer las ventas realizadas en las diferentes zonas, involucrando también los vendedores, los presupuestos asociados y todo esto debe poder visualizarse a nivel histórico, incluyendo mínimo los últimos dos años de operación.

Se deben poder realizar comparativos entre diferentes años, meses, semestres, entre otros. Mirando el comportamiento de las referencias en las diferentes zonas y comparando las ventas de los clientes pertenecientes a dichas zonas.

Los vendedores realizan visitas a los clientes con diferentes objetivos, la idea es poder tener estadísticas de cuantas visitas a realizado cada vendedor, a que cliente las ha realizado, con qué objetivo era dicha visita. En resumen toda la información relacionada con estos procesos de visitas, para poder tener claro cómo se está desempeñando el vendedor dentro de la zona que tiene asignada. Y además se debe poder hacer una trazabilidad de estas visitas a las ventas dentro de dicho sector.

### **Inconvenientes**

El principal inconveniente que se identificó es por parte de los vendedores, que no tienen equipos para trabajar.

## **9.5 Informe de reunión Iván Hincapié**

### **Gerente Comercial Socoda S.A**

#### **Anexo 5 - Iván Hincapié**

#### **Objetivos y responsabilidades**

Sus responsabilidades incluyen el monitoreo de los 5 canales que tiene Socoda con todo lo que esto incluye como cumplimientos de presupuestos de ventas y recaudo, rotación de cartera de cada canal entre otros. También está encargado del apoyo técnico relacionado con las cotizaciones que Socoda realiza, y además está liderando el área de apoyo a las ventas que son los que realizan la mayoría de los reportes actualmente.

#### **Criterios de desempeño**

Los principales criterios de desempeño que se tienen para cada uno de los canales son el cumplimiento de presupuesto de ventas, cumplimiento de recaudo de cartera y un medidor extra específico de cada uno de los canales.

#### **Reportes Actuales**

Actualmente se reciben informes de dos asistentes y de José Gabriel, ellos son los encargados de ingresar a QAD y obtener los datos en bruto para la generación de los siguientes reportes:

- Venta por unidad de negocio
- Venta por canal
- Venta por referencia
- Venta por vendedor
- Devoluciones
- Descuentos Financieros
- Cartera por cliente
- Cartera por antigüedad
- Medidor de rotación de cartera
- Margen bruto por grupo de producto por referencia y canal

- Gastos
- PyG por canal
- PyG total

Estos reportes llegan listos y no hace falta hacerles ninguna transformación o realizar cálculos extra para poder analizarlos correctamente.

## **Análisis**

De los anteriores reportes, se analizan los diferentes márgenes por canal y se monitorean los costos y gastos para los diferentes grupos de productos.

El análisis de las devoluciones y descuentos tiene especial importancia para el gerente comercial, ya que de éstos dependerá el margen que tenga cada una de las referencias.

También es valioso conocer los productos que más se venden por cliente desde el punto de vista de referencias y de grupos, conocer quiénes son sus mayores clientes, cuál es el peso de los clientes en las ventas totales en la compañía y qué vendedores son responsables de estas ventas.

Se espera que durante una visita a un cliente, se le pueda mostrar su comportamiento con respecto a las ventas, haciendo énfasis en el portafolio de productos que se le ofrecen.

Para el análisis del comportamiento de los productos, no es necesario desglosar los combos en componentes individuales sino analizarlos como el conjunto de elementos.

## **Inconvenientes**

El único inconveniente identificado fue la clasificación de los combos, todavía no se tiene claro exactamente si deben ser una clasificación aparte o se deben incluir dentro de alguna en especial, dependiendo de los elementos que lo constituyen.

## **Anexos**

Muchos de los reportes que el director comercial usa hoy en día no van a poder ser creados por completo en esta primera etapa de desarrollo del BI. Esto justificado en que el alcance de esta etapa incluye lo estrictamente relacionado con las ventas brutas de la compañía. Más adelante se empezara a trabajar en costos, cartera y elementos similares para poder llegar a los reportes que no se cubran.



## **9.6 Informe de reunión Carolina Velásquez**

### **Gerente de Mercadeo Socoda S.A**

#### **Anexo 6 - Carolina Velásquez**

#### **Objetivos y responsabilidades**

El área de mercadeo se enfoca más en el campo de los productos y no tanto de clientes como lo hace al área comercial. Están encargados de la clasificación de los artículos y de hacer un seguimiento a las ventas según esas clasificaciones que se están estableciendo. También cumplen con la función de estimación de precios de los diferentes productos de acuerdo al movimiento del mercado y la generación de listas de precios para cada uno de los canales de la compañía.

#### **Reportes Actuales**

Los reportes actuales tienen la información que se necesita para el área, sin embargo sería muy útil no estar limitados a tablas y poder realizar gráficos u otros elementos que fueran más dicientes y que permitieran hacer unos análisis más complejos. Además el estar en tablas de excel hace que la información sea muy estática y no se puedan hacer cambios en los reportes fácilmente.

#### **Análisis**

Para obtener el presupuesto de una referencia en un mes determinado, se tiene en cuenta el comportamiento de la misma en los meses corridos y el del mismo período en años anteriores. Para esto se usan los precios promedio de cada una de las referencias después de haber aplicado los diferentes descuentos por canal.

Es de suma importancia realizar análisis entre meses, entre referencias y separar las ventas entre las ventas nacionales y extranjeras, discriminando por países.

Se deben incluir las medidas de los productos dentro de los datos que se analizan ya que para el área esta información es importante.

Una de los aspectos en los que se hizo énfasis es en poder saber si verdaderamente la empresa ha crecido, porque actualmente no se puede llegar a esta conclusión fácilmente por la dificultad para analizar la información.

La plataforma debe permitir realizar un monitoreo de las ofertas o promociones que se están realizando en un momento dado para poder verificar la efectividad de las mimas.

Por último se quiere poder analizar las ventas comparándolas con los descuentos para verificar que no se esté afectando el mercado.

## **Inconvenientes**

Entre los inconvenientes que se encuentran, tienen especial protagonismo las clasificaciones de los combos y la segmentación de los productos, tema en el cual está trabajando el área de mercadeo.

Como se mencionó anteriormente, sería bueno tener gráficas que apoyen los reportes actuales, los cuales son únicamente tablas.

## **Anexos**

Durante la reunión se mencionó que no hace falta realizar una entrevista con Maria Paulina, encargada de salas de venta, ni con Mauricio, encargado de Estado y Concesiones porque ya se cubrió la información necesaria durante las reuniones con Ana María y con Carolina.

## **9.7 Informe de reunión Diana Sanchez**

### **Directora Canal de Cadenas Socoda S.A**

#### **Anexo 7 - Diana Sanchez**

#### **Objetivos y responsabilidades**

El canal de Cadenas espera lograr en el próximo año conseguir dos, máximo tres clientes nuevos.

También aspira conocer el mercado en las diferentes regiones del país para así identificar nuevas oportunidades de crecimiento y mejora.

#### **Clientes**

Actualmente, el canal cuenta únicamente con tres clientes: Sodimac, Corona y Flamingo, de éstos el más importante es Sodimac, el cual representa aproximadamente el 92% de las ventas del canal y el 30% de las ventas en Socoda.

La importancia de los clientes se mide únicamente de acuerdo al total facturado a cada uno y se clasifican según este criterio.

Estos clientes tienen tiendas en varias ciudades, y para QAD, cada una de las tiendas se registra como un cliente diferente. Se necesita conocer cuántos clientes se tienen por ciudad (Clientes) para hacer comparaciones de volumen de ventas entre diferentes ciudades.

#### **Reportes actuales**

La dirección del canal cadenas recibe reportes diariamente, pero éstos tienen poca relevancia en el análisis y son usados únicamente para monitorear cumplimiento del presupuesto. Los informes más importantes para el canal son los de cierre de mes.

Además de los informes de Socoda, la dirección de cadenas requiere la información de evacuación de los clientes. Corona y Flamingo proveen esta información directamente, pero Sodimac lo provee a través del sistema CEN. Esta información es lo más importante para el canal, ya que constituye el principal indicador de toma de decisiones.

#### **Análisis**

La directora ocupa aproximadamente dos días haciendo transformaciones de la información recibida, durante este período realiza análisis de ventas de productos por tamaño, calcula su participación y un peso ponderado. Ella también realiza gráficas para las presentaciones a partir de los datos que obtiene de sus cálculos.

Es necesario realizar una calificación de productos por regular, ofertado y descontinuado para lograr un análisis de crecimiento o decrecimiento de acuerdo a esta calificación para determinar la orientación del negocio. Otra clasificación que se debe considerar agregar es la de segmento económico al que está destinado cada línea de producto y diferenciar si un producto hace parte de los MTO (Productos estándar), o MTS (Productos especiales). También se debe tener en cuenta el color del producto vendido para para conocer cómo se están moviendo los productos no solo por la referencia sino también por este atributo.

La rastreabilidad de los productos es un factor importante. Cuando el producto se encuentra en Socoda tiene un código de barras que lo identifica, sin embargo cuando el producto pasa a una cadena la identificación cambia, por lo que se deben incluir ambos códigos para tener una correcta trazabilidad.

Los combos son otra característica para hacer análisis con las ventas, se requiere conocer que combos se han vendido, que productos dentro de un combo se han vendido y se deben poder realizar todos los filtros ya mencionados como cliente, zona, fecha, tipo de producto, entre otros.

Realiza cálculos de desempeño y rentabilidad por ciudad, y por tienda. Sin embargo, un análisis por zona no es relevante.

Se necesita conocer qué inventario tienen los clientes, si tienen un portafolio adecuado, donde están los productos y si hay rotación.

### **Inconvenientes**

Existen numerosos inconvenientes tanto con la información a la que se puede acceder como a los reportes que recibe la directora del canal.

El inconveniente más importante que se encuentra en este canal es la dificultad de acceso a los datos de evacuación de Sodimac, su cliente más importante. Actualmente sólo se puede acceder a parte de la información a través del sistema CEN y si se quiere acceder a la información completa, es necesario incurrir en un gasto adicional.

También se mencionaron problemas con listas de precios desordenadas y con la clasificación de los productos.

Se mencionó que los presupuestos recibidos no están discriminados por cliente, sin embargo esto no supone un problema ya que la directora se encarga de hacer la distribución del presupuesto.

Además de estos problemas, los reportes que recibe ofrecen poco valor, por lo que se deben hacer transformaciones y cálculos a partir de los volcados crudos de tablas de la base de datos.

Otro inconveniente importante consiste en las devoluciones que realizan los clientes, se debe tener claro que devoluciones se han recibido, de que clientes, cuales productos, etc.

## **9.8 Anexo a Diseño de Jobs de ETL**

[Anexos\Carga Clientes Incremental.png](#)

**Anexo 8-Carga Clientes Incremental**

[Anexos\Carga Clientes.png](#)

**Anexo 9-Carga Clientes**

[Anexos\Carga Moneda Incremental.png](#)

**Anexo 10-Carga Moneda Incremental**

[Anexos\Carga Moneda.png](#)

**Anexo 11-Carga Moneda**

[Anexos\Carga Producto Incremental.png](#)

**Anexo 12-Carga Producto Incremental**

[Anexos\Carga Producto.png](#)

**Anexo 13-Carga Producto**

[Anexos\Carga Servicios Incremental.png](#)

**Anexo 14-Carga Servicios Incremental**

[Anexos\Carga Servicios.png](#)

**Anexo 15-Carga Servicios**

[Anexos\Carga Vendedores Incremental.png](#)

**Anexo 16-Carga Vendedores Incremental**

[Anexos\Carga Vendedores.png](#)

**Anexo 17-Carga Vendedores**

[Anexos\Carga Ventas Incremental.png](#)

**Anexo 18-Carga Ventas Incremental**

[Anexos\Carga Ventas.png](#)

**Anexo 19-Carga Ventas**

[Anexos\Extracción Clientes Incremental.png](#)

**Anexo 20-Extracción Clientes Incremental**

[Anexos\Extracción Clientes.png](#)

**Anexo 21-Extracción Clientes**

[Anexos\Extracción Componentes Combo-Factura.png](#)

**Anexo 22-Extracción Componentes Combo-Factura**

[Anexos\Extracción Moneda Incremental.png](#)

**Anexo 23-Extracción Moneda Incremental**

[Anexos\Extracción Moneda.png](#)

**Anexo 24-Extracción Moneda**

[Anexos\Extracción Notas de cartera Incremental.png](#)

**Anexo 25-Extracción Notas de Cartera Incremental**

[Anexos\Extracción Notas de cartera.png](#)

**Anexo 26-Extracción Notas de Cartera**

[Anexos\Extracción Productos Incremental.png](#)

**Anexo 27-Extracción Productos Incremental**

[Anexos\Extracción Productos.png](#)

**Anexo 28-Extracción Productos**

[Anexos\Extracción Servicios Incremental.png](#)

**Anexo 29-Extracción Servicios Incremental**

[Anexos\Extracción Servicios.png](#)

**Anexo 30-Extracción Servicios**

[Anexos\Extracción Vendedores Incremental.png](#)

**Anexo 31-Extracción Vendedores Incremental**

[Anexos\Extracción Vendedores.png](#)

**Anexo 32-Extracción Vendedores**

[Anexos\Extracción Ventas Incremental.png](#)

**Anexo 33-Extracción Ventas Incremental**

[Anexos\Extracción Ventas.png](#)

**Anexo 34-Extracción Ventas**

[Anexos\Master Historico.png](#)

**Anexo 35-Master Histórico**

[Anexos\Master Incremental.png](#)

**Anexo 36-Master Incremental**

[Anexos\Transformación Clientes.png](#)

**Anexo 37-Transformación Clientes**

[Anexos\Transformación Moneda.png](#)

**Anexo 38-Transformación Moneda**

[Anexos\Transformación Productos.png](#)

**Anexo 39-Transformación Productos**

[Anexos\Transformación Servicios.png](#)

**Anexo 40-Transformación Servicios**

[Anexos\Transformación Vendedores.png](#)

**Anexo 41-Transformación Vendedores**

[Anexos\Transformación Ventas-Notas Cartera.png](#)

**Anexo 42-Transformación Ventas-Notas Cartera**

[Anexos\Transformación Ventas-Productos.png](#)

**Anexo 43-Transformación Ventas-Productos**

[Anexos\Transformación Ventas-Servicios.png](#)

**Anexo 44-Transformación Ventas-Servicios**