

DISEÑO E IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA LA CREACIÓN  
DE INFORMES GERENCIALES

PRESENTADO POR:

DIANA KATERINE GUTIERREZ MENDIETA



FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES  
FACULTAD DE INGENIERÍA  
INGENIERÍA DE SISTEMAS  
BOGOTÁ  
2017

DISEÑO E IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA LA CREACIÓN  
DE INFORMES GERENCIALES

PRESENTADO POR:

DIANA KATERINE GUTIERREZ MENDIETA

*Trabajo de Grado desarrollado  
Como requisito para optar por el título de  
Ingeniero de Sistemas*

DIRECTOR:  
CELIO GIL AROS  
INGENIERO DE SISTEMAS



FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES  
FACULTAD DE INGENIERÍA  
INGENIERÍA DE SISTEMAS  
BOGOTÁ  
2017

## DEDICATORIA

*Dedico este trabajo de grado primeramente a Dios, sin él no podría estar donde estoy, a mi madre Flor Dely Mendieta por enseñarme el valor de las cosas y darme el ejemplo de seguir mis sueños sin importar los obstáculos, a mi hijo Juan Esteban Velandia por ser el motor que mueve mi vida e inspirarme a luchar con más ganas cada día, a mi esposo por la paciencia en todo este proceso, a mis hermanos Jair, Sandra y Viviana por sus aportes, ejemplos y amor durante toda mi vida, a mis sobrinos Valentina y David para quienes quiero ser un buen ejemplo de vida, a mis amigos y compañeros de estudio por los aprendizajes, motivación, apoyo y cariño y por supuesto a mis profesores recuerdo a todos y cada uno con total orgullo por sus enseñanzas me llevo la satisfacción de saber que aprendí con los mejores.*

*Diana Katerine Gutiérrez Mendieta*

## AGRADECIMIENTOS

*Le agradezco en primera medida a Dios por permitirme cumplir mis sueños y reiterarme una vez más que con esfuerzo, dedicación y tomada de su mano puedo lograr cualquier cosa que me proponga. Un muy especial y sincero agradecimiento al Ingeniero Celio Gil por la compañía y asesoramiento durante este proceso y en general a todos mis profesores, los cuales contribuyeron de gran manera en todo mi proceso de aprendizaje, a mis compañeros de estudio les agradezco la ayuda y las enseñanzas, a mis amigos más cercanos les agradezco el apoyo incondicional en cada etapa, a mi hijo y esposo les agradezco el amor y la paciencia durante toda este proceso, a mi madre, hermanos y sobrinos les agradezco su amor, comprensión, ejemplos y enseñanzas que he recibido de ellos durante toda mi vida y por supuesto un gran agradecimiento al que fue mi lugar de entrenamiento, donde llore, reí y me sentí orgullosa al ir culminando cada semestre la Fundación Universitaria Los Libertadores me siento feliz de ser parte de esta comunidad.*

**Nota de Aceptación**

---

---

---

**Jurado**

---

---

## RESUMEN

Se plasma en este documento el proceso para desarrollar una estrategia de inteligencia de negocios basada en un almacén de datos (Data Warehouse) para el área de ventas de la organización CORONA S.A. ya que la toma de decisiones en cualquier organización es una parte de suma importancia para el desarrollo de las actividades y alcance de las metas propuestas, también lo es para la organización CORONA, en donde el crecimiento de la empresa y el alto volumen de información dificultan la toma de decisiones a causa de la insuficiencia en las herramientas utilizadas actualmente.

Analizando y aplicando las técnicas de inteligencia de negocios, representada en las etapas de; Dirigir y planear, identificación de la información, procesamiento de los datos y análisis y producción y, en busca del aprovechamiento de la información, se desarrolló un almacén de datos (DW) por medio de la herramienta de software libre llamada Talen Studio versión 5.2.

Se implementó la metodología Hefesto, la cual está fundamentada en una amplia investigación, es ideal para este desarrollo, ya que se encuentra en una continua evolución y se basa en la completa comprensión de cada paso en el desarrollo de software, además a lo largo del documento se puede apreciar una gran cantidad de figuras y tablas que explican perfectamente cada paso en el desarrollo del proyecto.

Este sistema le permitirá a la alta gerencia y a los usuarios finales de las bases de datos contar con información accesible, veraz y actualizada, con la cual podrán tomar decisiones sin la preocupación de fallas en la información, así mismo este sistema podría ser ampliado a otras áreas de la organización en las cuales se requiera desglosar la información.

Este almacén de datos permite garantizar la veracidad y seguridad en la información de igual manera que su fácil manejo por parte del personal calificado, su facilidad para monitoreo y mantenimiento de acuerdo al crecimiento de la empresa.

## ABSTRACT

The process is reflected in this document to develop a strategy based business intelligence Data Warehouse for the sales of the organization CORONA S.A. since the decision in any organization is a part of paramount importance for the development of activities and scope of the proposed goals, also for the CORONA organization, where the growth of the company and the high volume of information difficult decision making because of the insufficiency of the tools currently used.

Analyzing and applying the techniques of business intelligence, shown in stages; Direct and plan, identifying information, data processing and analysis and production, seeking the use of information, a Data Warehouse (DW) through free software tool called Talen Studio version 5.2 was developed.

The Hefesto methodology, which is based on extensive research, is ideal for this development as it is in continuous development and is based on the complete understanding of each step in the development of software, plus over Implemented document you can see a lot of figures and tables perfectly explain each step in the project.

This system will allow senior management and end of the databases accessible information have real, true time, which can make decisions without worrying about failure information, users likewise this system could be extended to other areas of the organization which requires disaggregating information.

This data store ensures the accuracy and information security just as easy handling by qualified personnel, it's easy to monitor and maintain according to the growth of the company.

## SIGLAS

**BI:** Inteligencia de negocios.

**DM:** Data Mart.

**DW:** Data Warehouse.

**ERP:** Planificación de recursos empresariales.

**ETL:** Extracción, transformación y carga.

**HOLAP:** Proceso analítico híbrido en línea.

**MOLAP:** Proceso analítico multidimensional en línea.

**ODS:** Almacén de datos operativos.

**OLAP:** Proceso analítico en línea.

**OLTP:** Procesamiento de transacciones en línea.

**ROLAP:** Proceso analítico Relacional en línea.

**SCD:** Dimensiones lentamente cambiantes.

**SGBD:** Sistema de gestión de base de datos.

## TABLA DE CONTENIDO

RESUMEN .....	6
ABSTRACT .....	7
SIGLAS .....	8
INTRODUCCIÓN.....	15
1. ASPECTOS DE INVESTIGACIÓN.....	16
1.1 TITULO.....	16
1.2 DESCRIPCIÓN DEL PROBLEMA .....	16
1.3 JUSTIFICACIÓN DEL PROYECTO DE INVESIGACIÓN .....	16
1.4 IMPACTO .....	17
1.5 DELIMITACIÓN .....	17
1.5.1 Espacial.....	18
1.5.2 Cronológica. ....	19
1.5.3 Conceptual.....	20
1.6 RECURSOS .....	21
1.6.1. Recurso humano .....	21
1.6.2. Recursos de software .....	21
1.6.3. Recursos de Hardware .....	22
1.7 METODOLOGIA .....	22
1.8. OBJETIVO GENERAL .....	22
1.8.1. Objetivos Específicos. ....	23
1.9 FORMULACIÓN DE LA PREGUNTA DE INVESTIGACIÓN .....	23
2. MARCO TEORICO .....	24
2.1. ESTADO DEL ARTE .....	24
2.1.1 Antecedentes.....	24
2.1.2 Históricos.....	25
2.2 Bases teóricas .....	46
2.2.1 Consideraciones de diseño.....	46
2.3. Metas a alcanzar .....	54
2.4. Productos a entregar .....	54
2.5. Glosario .....	54
3. DISEÑO E IMPLEMENTACIÓN .....	55

3.1. Planeación y administración del proyecto .....	56
3.1.1 Definición del proyecto .....	56
3.2 Análisis de requerimientos .....	56
3.2.1 Identificar preguntas.....	56
3.2.2 Identificar indicadores y perspectivas .....	56
3.2.3 Modelo Conceptual .....	57
3.3 Análisis de los OLTP.....	58
3.3.1 Conformar Indicadores.....	58
3.3.2 Establecer Correspondencias.....	58
3.3.3 Nivel de Granularidad. ....	60
3.3.4 Modelo conceptual ampliado.....	60
3.4 Modelo lógico del DW .....	61
3.4.1 Tipo de modelo lógico del DW. ....	61
3.4.2 Tablas dimensionales.....	61
3.4.3 Uniones.....	63
3.5 Integración de los datos .....	63
3.5.1 Carga inicial.....	63
3.5.2 Actualización.....	64
3.6 ingeniería de requerimientos .....	67
3.6.1 Definición de los requerimientos del negocio .....	67
3.6.2 Diagramas de modelamiento unificado .....	69
3.7 Diseño arquitectónico .....	71
3.7.1 Back Room. ....	71
3.7.2 Front Room.....	72
3.8 Diseño físico.....	73
3.8.1 Base de datos de origen. ....	73
3.8.2 Base de datos destino (Data Warehouse).....	74
3.9 Diseño y desarrollo del proceso ETL .....	76
3.9.1 Paso 1: Trazar un plan de alto nivel. ....	76
3.9.2 Paso 2: Funcionalidad de la herramienta ETL. ....	76
3.9.3 Pasó 3: Desarrollo de estrategias.....	76
3.9.4 Paso 4: Profundizar en las tablas destino. ....	79
3.9.5 Paso 5: Poblar las tablas de dimensiones con los datos históricos.....	79
3.9.6 Paso 7: Procesamiento incremental de las dimensiones.....	85

3.9.7 Paso 8: Procesamiento incremental de los hechos.....	86
3.9.8 Paso 9: Manejo y automatización del sistema ETL.....	87
3.10 Diseño y desarrollo del DataMart.....	90
3.10.1 Tipo de aplicación BI. ....	90
3.11 Implementación del DataMart.....	91
4. ANALISIS DE RESULTADOS Y CONCLUSIONES .....	94
4.1 Pruebas de la herramienta.....	94
4.1.1 Desarrollo de la aplicación de BI con la herramienta Microsoft Excel 2010.....	94
4.1.2 Despliegue del informe final .....	94
4.1.3 Despliegue de la ETL. ....	95
4.1.4 Despliegue BI. ....	98
4.1.5 Mantenimiento de la plataforma.....	103
4.1.6 Back Room. ....	104
4.1.7 Crecimiento del Data Warehouse .....	105
4.1.8 Evaluar el entorno actual. ....	105
4.1.9 Oportunidades de crecimiento.....	105
4.2 conclusiones .....	106
Bibliografía.....	107

## TABLA DE TABLAS

Tabla 1. Recursos humanos .....	21
Tabla 2. Recursos de Software .....	21
Tabla 3. Recursos de Hardware .....	22
Tabla 4. Matriz DWH.....	73
Tabla 5. Decodificación y Formato de la Dimensión de Tiempo.....	80
Tabla 6. Valores Nulos Tabla de pedidos.....	81
Tabla 7. Valores Nulos Tabla de ventas .....	81
Tabla 8. Valores Nulos Tabla de número de ventas .....	82
Tabla 9. Búsqueda llave subrogada Tabla de pedidos.....	82
Tabla 10. Búsqueda llave subrogada Tabla de ventas .....	83
Tabla 11. Búsqueda llave subrogada Tabla número de ventas.....	83
Tabla 12. Primer informe de ventas.....	102
Tabla 13. Informe TOP 20.....	103

## TABLA DE ILUSTRACIONES

Ilustración 1. Objetivo de la inteligencia de negocios (Dataprix, s.f.) .....	26
Ilustración 2. Arquitectura de BI (Dataprix, s.f.) .....	27
Ilustración 3. Etapas de un BI (Kimball) .....	28
Ilustración 4. Arquitectura de un Data WareHouse (Dataprix, s.f.) .....	32
Ilustración 5. Estructura de un DW (Slid, s.f.) .....	33
Ilustración 6. Proceso de ETL (Dataprix, s.f.) .....	33
Ilustración 7. Proceso de Extracción de ETL (Postech, s.f.) .....	34
Ilustración 8. Proceso de Transformación de ETL (Postech, s.f.) .....	34
Ilustración 9. Proceso de Carga de ETL (Postech, s.f.) .....	35
Ilustración 10. Esquema de estrella (Dataprix, s.f.) .....	36
Ilustración 11. Esquema de Copo de Nieve (Dataprix, s.f.) .....	37
Ilustración 12. Tabla de Hechos (Dataprix, s.f.) .....	38
Ilustración 13. Tabla de Dimensiones (Dataprix, s.f.) .....	39
Ilustración 14. Niveles de la dimensión (Dataprix, s.f.) .....	39
Ilustración 15. Figura de Jerarquía (Dataprix, s.f.) .....	40
Ilustración 16. Figura de Relaciones (Dataprix, s.f.) .....	41
Ilustración 17. Análisis OLAP (Wikispaces, s.f.) .....	43
Ilustración 18. Drill Down .....	45
Ilustración 19. Roll Up (Slid, s.f.) .....	45
Ilustración 20. Metodología Hefesto (Bernau, 2010) .....	55
Ilustración 21. Modelo Conceptual (Dataprix, s.f.) .....	57
Ilustración 22. Modelo Entidad – Relación Almacenes corona SAS .....	59
Ilustración 23. Modelo de correspondencia .....	59
Ilustración 24. Modelo conceptual ampliado .....	61
Ilustración 25. Tabla dimensión de tiempo .....	62
Ilustración 26. Tabla de Hechos .....	62
Ilustración 27. Tabla de Uniones .....	63
Ilustración 28. Modelo Lógico del Cubo .....	64
Ilustración 29. Cubo de Ventas por cliente .....	66
Ilustración 30. Indicadores de Cumplimiento .....	68
Ilustración 31. Diagrama General de caso de uso .....	70
Ilustración 32. Diagrama general de secuencia .....	70
Ilustración 33. Diagrama General de Actividades .....	71
Ilustración 34. Back Room .....	72
Ilustración 35. Front Room del Proyecto .....	73
Ilustración 36. Arquitectura CDC .....	77
Ilustración 37. Transformaciones ETL .....	78
Ilustración 38. Auditoria Error ETL .....	78
Ilustración 39. Proceso ETL dimensión de pedido – carga inicial .....	84
Ilustración 40. Proceso ETL Hecho de ventas – carga inicial .....	84
Ilustración 41. Funcionamiento CDC – Dimensiones .....	85
Ilustración 42. Agregar clave subrogada .....	86
Ilustración 43. Funcionamiento CDC – Hechos .....	86
Ilustración 44. Proceso ETL dimensión de pedido – carga inicial .....	87
Ilustración 45. Proceso ETL Hecho de ventas – carga inicial .....	87

Ilustración 46. Verificar dimensiones.....	88
Ilustración 47. Verificar Hechos.....	88
Ilustración 48. Poblar Dimensiones.....	89
Ilustración 49. Poblar Hechos.....	89
Ilustración 50. Poblar DWH.....	90
Ilustración 51 Tablero de Mando ALCO.....	91
Ilustración 52. Informe Gerencial.....	92
Ilustración 53. Informe top 20 ventas.....	92
Ilustración 51. Ejecución del JOB.....	96
Ilustración 52. Transformaciones de las tablas finales.....	97
Ilustración 53. Programación del administrador de tareas.....	98
Ilustración 54. Valor de Inventario.....	99
Ilustración 55. Valor de Inventario.....	99
Ilustración 56. Valor de Inventario por Negocio.....	100
Ilustración 57. Valor de Inventario por Formato.....	100

## INTRODUCCIÓN

En el transcurso de los años las empresas han entendido la importancia en el manejo de la información para el crecimiento de sus compañías; es por esto que nace un concepto llamado inteligencia de negocios o sus abreviaciones en ingles BI. La inteligencia de negocios es la conversión de los datos operativos de las empresas en información que da soporte a la toma de decisiones. (Diaz, 2012)

Almacenes corona S.A.S es una de las empresas pertenecientes al grupo CORONA donde su campo de actividad es la construcción y remodelación. Es una empresa que lleva más de 35 años en el mercado y se ha mantenido competitiva debido a la calidad de sus productos, actualmente cuenta con 36 puntos de venta a nivel nacional y una facturación mensual aproximada de 21.000.000.000 millones de pesos. Almacenes corona es considerada una empresa grande, debido a sus ingresos y número de empleados esto conlleva a que también se generen mucho flujo de información como lo son los datos de los clientes, los datos de los empleados, la información de los productos que se venden y la información de las ventas que se generan. Toda esta información está alojada en una base de datos en Oracle 11, el cual permite por medio de queries extraer la información, sin embargo, no es muy ágil debido a que se deben generar varias vistas para traer información importante al momento de la toma de decisiones en cuanto a las ventas del negocio. Es por esto que se desarrollara un DATA WAREHOUSE con herramientas de BI para apoyar la toma de decisiones en cuanto a la gerencia de operaciones la cual es la encargada de realizar acciones y crear estrategias de ventas.

Para el desarrollo se utilizarán las herramientas de base de datos que actualmente tiene la compañía Oracle 11g, para el proceso ETL se hará uso de una herramienta llamada Talen Studio Versión 5.2 y para la interfaz con el usuario se usará Excel con sus respectivos complementos por su fácil entendimiento entre los usuarios, en cuanto a la metodología se usará Hefesto la cual trae todas las etapas para el desarrollo de un DATA WAREHOUSE.

# 1. ASPECTOS DE INVESTIGACIÓN

## 1.1 TITULO

DISEÑO E IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA LA GENERACIÓN DE INFORMES GERENCIALES

## 1.2 DESCRIPCIÓN DEL PROBLEMA

Almacenes Corona S.A.S. es una empresa dedicada al negocio del Retail (compra y venta) de productos y servicios para remodelación de hogares y empresas; actualmente la empresa cuenta con “PeopleSoft Versión 9.1”, un ERP, el cual permite generar informes o detalles de la información almacenada y procesada en la base de datos. Información con la cual las áreas de la compañía como: el área Financiera, área Comercial, área de Gestión Humana, área de Mercadeo, área de Logística y área de Operaciones, realizan sus respectivos análisis y consultas para la toma de decisiones, apoyados por el CIAC (Centro De Inteligencia y Análisis Comercial).

Para ayudar en esta labor se cuenta con una arquitectura de datos llamada “cubo de ventas”, visualizada en Microsoft Excel que permite consultar y analizar esta información de forma profunda y detallada de históricos y tendencias que se han presentado en el transcurso del tiempo dentro de la empresa.

Pese a estar la información disponible y actualizada, la facilidad para encontrar o desglosar información a ciertos niveles de detalle se vuelve lento para los usuarios, ya que estas consultas requieren bastante tiempo para llegar a los niveles deseados, los cuales comúnmente se refieren al detalle de códigos de productos asociados a un punto de venta en fechas específicas, esto ligado a toda la información generada en la transacción o proceso generado desde su compra hasta su venta.

De igual manera la estructura en la que está creada la herramienta, visualmente no es muy clara y amigable para los usuarios del área de Operaciones; algunos colaboradores requieren bastante tiempo para entender, adaptarse y manejarla; inconvenientes que el área de inteligencia y análisis comercial tiene que asumir frente a los usuarios y su necesidad de urgencia que no permite dar soluciones oportunas que permitan apoyar el proceso de toma de decisiones en la compañía.

## 1.3 JUSTIFICACIÓN DEL PROYECTO DE INVESTIGACIÓN

En los últimos años Almacenes Corona S.A.S. ha presentado un importante crecimiento a nivel de compañía el cual ha sido acompañado por el arduo trabajo de sus colaboradores en búsqueda del objetivo de suplir necesidades tanto de sus clientes internos como externos, todo apoyado principalmente en el manejo y uso de la información que posee la empresa, información que las gerencias de cada área han podido utilizar para apoyar la toma de decisiones, afirmar el crecimiento de los negocios y mejorar la gestión de los puntos de venta que esta compañía posee.

Es en este ámbito donde surge una necesidad de la compañía y sus líderes, de tener la información precisa en el momento indicado, una labor que no solo se quede en informes para analizar, sino que también incluya un complemento que permita por medio de herramientas de Inteligencia de Negocios y manejo de bases de datos; cambiar el esquema de dichos informes, convertirlos en tableros de mando y reportes dinámicos, concediendo la oportunidad de consultar la información al instante, obtener una mejor perspectiva del negocio con respecto a su comportamiento y agilizar la búsqueda de las mejores opciones para el cumplimiento de las metas establecidas.

Adicionalmente al tener un ERP, actualmente Almacenes Corona S.A.S. También cuenta con herramientas de manejo de datos como lo son los Cubos; los cubos permiten entre otros datos visualizar información de Ventas, márgenes de utilidad, Presupuestos, y verlos en dimensiones de tiempo tanto históricos como actuales.

Con la implementación de un Data Warehouse para la creación de informes gerenciales y tableros de mando, se beneficia a los usuarios de las distintas áreas de la compañía el poder contar con la información de una manera más detallada, oportuna, clara y ágil, que permita tomar decisiones de gestión en un menor tiempo y con un mayor porcentaje de efectividad, contribuyendo para el futuro de la compañía apoyados con el enfoque la siguiente ecuación:

#### Ecuación 1

Datos + Tratamiento de datos = Información

#### Ecuación 2

Información + Interpretación = Conocimiento

Con esta propuesta se busca cambiar y mejorar las herramientas con las que viene trabajando la compañía y sus colaboradores, implementando un Data Warehouse que permita almacenar la información histórica de la compañía en distintas bases de datos, este, complementado con herramientas de inteligencia de negocios (BI), contribuirán a hacer un mejor uso de la información, volviendo más amigable y entendible su visualización, mejorando el proceso de análisis, ayudando a la toma de decisiones y aprovechando los tiempos y oportunidades que la operación y el mercado ofrecen.

## 1.4 IMPACTO

La implementación de una Data Warehouse como parte de una solución de Inteligencia de Negocios que suplirá las necesidades de información de las distintas áreas de la compañía como: Disponibilidad, Análisis y Eficiencia a nivel estratégico que permita el apoyo a la toma de decisiones en beneficio de la empresa.

## 1.5 DELIMITACIÓN

La implementación de un Data Warehouse como medio de almacenamiento de información de la compañía, que permita por medio de un DataMart desarrollado en un software llamado Talend Open Studio V. 5.2 5.2 con el fin de crear y gestionar informes para la gerencia de operaciones en Almacenes Corona S.A.S.

El alcance contempla la revisión, evaluación y aprobación de la propuesta de implementación, por parte del área de tecnología y equipo del centro de inteligencia y análisis comercial (CIAC), quienes serán los administradores y directos responsables del manejo de la herramienta, así como de su mantenimiento y actualización.

Igualmente se contemplarán los proyectos de inversión o actualización de software y ERP que la compañía tenga planificada para los próximos 12 meses, que puedan afectar o aplazar la implementación de la presente propuesta.

#### **1.5.1 Espacial**

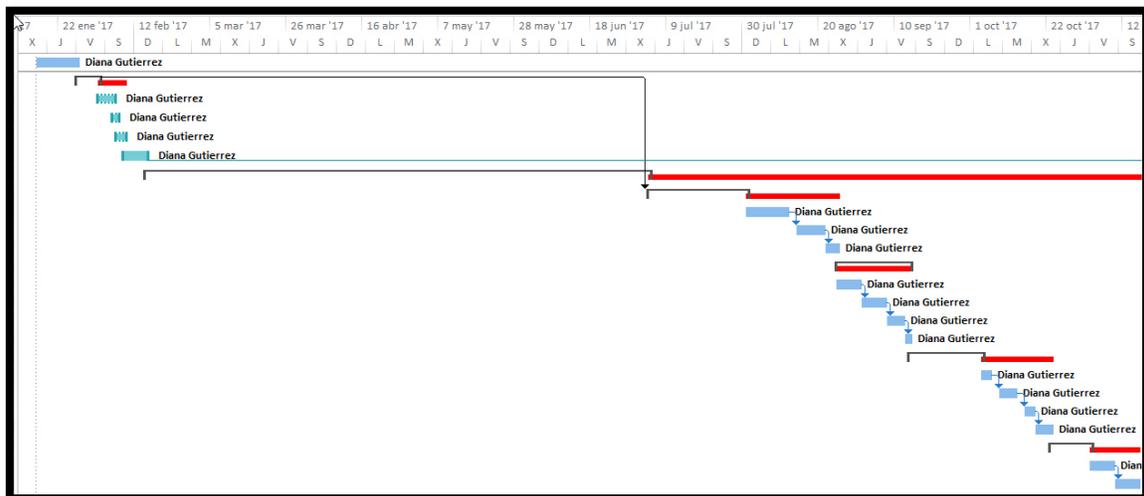
El proyecto será realizado en y para Almacenes corona SAS, usando los recursos de la empresa como servidores y equipos, ya que solamente es allí donde se puede realizar el desarrollo del proyecto, debido a que se necesitan los equipos locales por el tema de permisos.

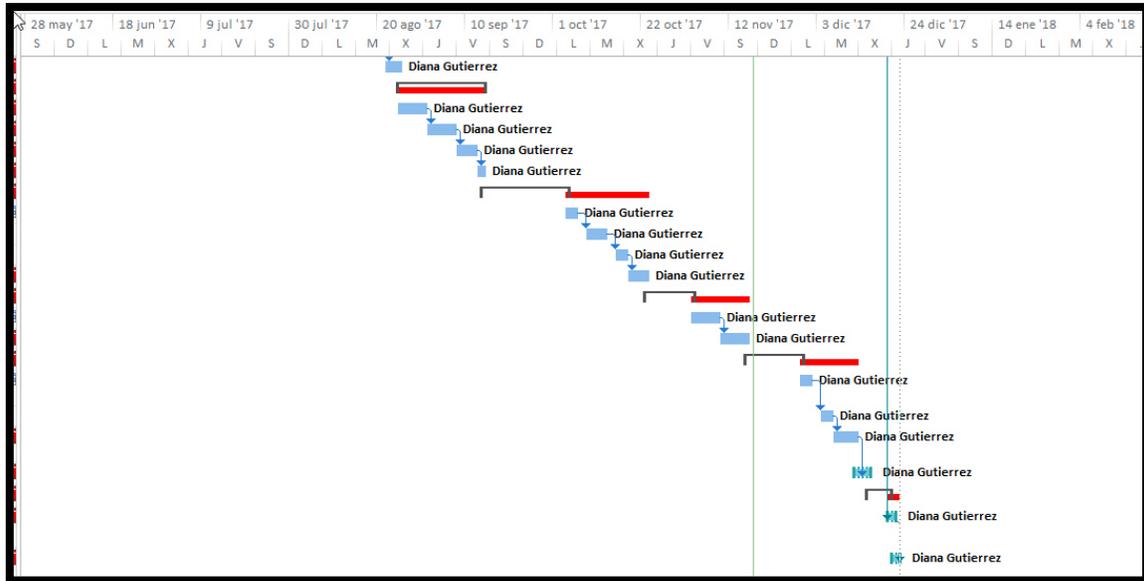
## 1.5.2 Cronológica.

### Cronograma del proyecto

Presentación Proyecto y equipo de trabajo	10 días	lun 16/01/17	vie 27/01/17	Diana Gutierrez
▲ Fase 1: Requerimientos	5 días	vie 27/01/17	<u>jue 2/02/17</u>	Diana Gutierrez
Identificación de requerimientos y riesgos	3 días	jue 2/02/17	<u>lun 6/02/17</u>	Diana Gutierrez
Estimación de recursos	2 días	lun 6/02/17	<u>mar 7/02/17</u>	Diana Gutierrez
Evaluación de la capacidad para desarrollar el proyecto	3 días	mar 7/02/17	<u>jue 9/02/17</u>	Diana Gutierrez
Fase 2: Levantamiento de información	5 días	jue 9/02/17	mié 15/02/17	Diana Gutierrez
▲ Fase 3: Desarrollo del prototipo	100 días	mié 15/02/17	<u>mar 4/07/17</u>	Diana Gutierrez
▲ Fase 3.1: Análisis de requerimientos	20 días	mar 4/07/17	<u>lun 31/07/17</u> 2	Diana Gutierrez
Requerimientos funcionales y no funcionales	10 días	lun 31/07/17	vie 11/08/17	Diana Gutierrez
Identificar preguntas y diferentes perspectivas	6 días	lun 14/08/17	lun 21/08/17 9	Diana Gutierrez
Diseñar un modelo conceptual	4 días	mar 22/08/17	vie 25/08/17 10	Diana Gutierrez
▲ Fase 3.2: Análisis de los OLTP	15 días	vie 25/08/17	<u>jue 14/09/17</u>	Diana Gutierrez
Identificación y conformación de indicadores	5 días	vie 25/08/17	jue 31/08/17	Diana Gutierrez
Establecer correspondencia de información	5 días	vie 1/09/17	jue 7/09/17 13	Diana Gutierrez
Identificar a que nivel de granularidad se requiere la información	3 días	vie 8/09/17	mar 12/09/17 14	Diana Gutierrez
Diseño y construcción del modelo conceptual ampliado	2 días	mié 13/09/17	jue 14/09/17 15	Diana Gutierrez
▲ Fase 3.3: Diseño del Modelo Lógico de DW	15 días	jue 14/09/17	<u>mié 4/10/17</u>	Diana Gutierrez
Diseño del tipo de modelo lógico del DW.	3 días	mié 4/10/17	vie 6/10/17	Diana Gutierrez
Identificación y diseño de las tablas de dimensiones del DW	5 días	lun 9/10/17	vie 13/10/17 18	Diana Gutierrez
Identificación y diseño de las tablas de hechos de DW	3 días	lun 16/10/17	mié 18/10/17 19	Diana Gutierrez
Construcción de las uniones necesarias entras las distintas tablas	3 días	jue 19/10/17	lun 23/10/17 20	Diana Gutierrez
▲ Fase 3.4: Integración de los datos	10 días	lun 23/10/17	<u>vie 3/11/17</u>	Diana Gutierrez
Identificar las cargas iniciales de los datos	5 días	vie 3/11/17	jue 9/11/17	Diana Gutierrez
Creación de políticas de actualización de los datos	5 días	vie 10/11/17	jue 16/11/17 23	Diana Gutierrez
▲ Fase 3.5: Implementación de vistas	10 días	jue 16/11/17	<u>mié 29/11/17</u>	Diana Gutierrez
Identificar los informes que se manejan manualmente y que toman más tiempo en su construcción	3 días	mié 29/11/17	vie 1/12/17	Diana Gutierrez
Análisis y diseño de informes requeridos	3 días	lun 4/12/17	mié 6/12/17 26	Diana Gutierrez
Configuración, despliegue y capacitación a los usuarios finales acerca de la manipulación de la información	4 días	jue 7/12/17	mar 12/12/17 27	Diana Gutierrez
Fase 4: Evaluación del prototipo.	4 días	mar 12/12/17	<u>vie 15/12/17</u> 28	Diana Gutierrez
▲ Fase 5: Construcción del documento.	4 días	vie 15/12/17	<u>mié 20/12/17</u>	Diana Gutierrez
Fase 5.1: Análisis y selección de la información recolectada.	2 días	mié 20/12/17	<u>jue 21/12/17</u> 6	Diana Gutierrez
Fase 5.2: Construcción y digitación del marco teórico.	2 días	jue 21/12/17	<u>vie 22/12/17</u> 31	Diana Gutierrez

### Diagrama Gantt





### 1.5.3 Conceptual

Para definir el alcance del proyecto del desarrollo del DW el cual se utilizará para extraer información importante que permita un análisis tanto a nivel de inventarios de la compañía como en las unidades de negocio (Puntos de venta), su justificación es contar con información útil que soporte las distintas decisiones en función de su análisis.

Para la planeación del proyecto se definieron los roles que intervendrán para cumplir los objetivos del proyecto.

Desarrolladores del proyecto que son las personas que se encargaran de:

- Desarrollar
- Monitorear el avance el proyecto
- Diseñar y modelar
- Analizar el Negocio
- Jefes de Área
- Jefes de zona y áreas de negocio a los cuales se les va ir notificando del avance del proyecto.
- Gerente de operaciones
- Coordinadores de negocio
- Gerencia Comercial

Personal involucrado en el negocio: jefes de punto de venta, aseguradores de producto, analistas logísticos y líderes de entrega.

## 1.6 RECURSOS

### 1.6.1. Recurso humano

ITEM	CANTIDAD DE HORAS	VALOR UNITARIO	VALOR TOTAL
Análisis	60	9000	540.000
Diseño	90	9000	810.000
Desarrolló	115	9000	1.035.000
Pruebas	20	9000	180.000
Implementación	80	9000	720.000
TOTAL	365	45000	3.285.000

Tabla 1. Recursos humanos

### 1.6.2. Recursos de software

ITEM	CANTIDAD	VALOR UNITARIO	VALOR TOTAL
MS Windows Server	1	0	0
MS Windows 2007	1	0	0
MS office 2010	1	0	0
MS SQL server 2005	1	0	0
MYSQL Workbench 5.2	1	0	0
MS Acces 2007	1	0	0
MS Analysis Services	1	0	0
MDM Talend Open Studio 5.2	1	0	0
Oracle 11	1	0	0
People tools 8.5	1	0	0
People 9.1	1	0	0
Acceso OLAP	1		\$ 0

Tabla 2. Recursos de Software

Los Recursos de Software no tienen costo por ser un proyecto Estudiantil la empresa suministra estos recursos.

### 1.6.3. Recursos de Hardware

ITEM	CANTIDAD	VALOR UNITARIO	VALOR TOTAL
PC Intel core I5 de 1.50 GHz	2	1.669.900	3.339.800
Disco 500 GB	1	189.000	189000
Disco externo 500 GB	1	189.000	189000
CD/DVD	10	3000	30000
TOTAL	14	2050900	3720800

Tabla 3. Recursos de Hardware

## 1.7 METODOLOGIA

Hefesto es una metodología propia cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos. Cabe destacar que Hefesto se encuentra en continua evolución y se ha tenido en cuenta como gran valor agregado, todos los feedbacks que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines.

La idea principal es comprender cada paso que se realizara para no tener que caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente que se está haciendo y por qué.

La construcción e implementación de un DW puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán diferentes, lo que se debe tener muy en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue demasiado largas, lo que busca es entregar una primera implementación que satisfaga una parte de las necesidades para demostrar las ventajas del DW.

## 1.8. OBJETIVO GENERAL

Implementar un Almacén de Datos para la toma de decisiones gerenciales en la compañía Almacenes Corona S.A.S.

### **1.8.1. Objetivos Específicos.**

- Evaluar la relevancia y calidad de los datos almacenados en el Data Warehouse, requeridos en la toma de decisiones de las áreas de Almacenes corona S.A.S.

Unificar toda la información de la compañía disponible en el ERP PeopleSoft 9.1 en un almacén de datos (DW).

- Analizar la información obtenida de las distintas áreas de Almacenes Corona S.A.S. Por medio de técnicas de Extracción, Transformación y Carga de datos.

## **1.9 FORMULACIÓN DE LA PREGUNTA DE INVESTIGACIÓN**

¿Cómo con la implementación de un Data Warehouse se solucionaría la problemática actual de Almacenes Corona S.A.S.?

## 2. MARCO TEORICO

### 2.1. ESTADO DEL ARTE

#### 2.1.1 Antecedentes

Luego de realizar investigaciones sobre trabajos o escritos en los que se haya aplicado el concepto de inteligencia de negocios en una compañía y pese a que existen demasiadas fuentes y referencias se optó por tomar más énfasis en dos proyectos que se realizaron en Colombia y en Perú y aun cuando las empresas involucradas pertenecen a industrias muy distintas, el principal concepto manejado es el de implementar un Data WareHouse como herramienta de inteligencia de negocios.

Análisis, Diseño e Implementación de un Data WareHouse de Soporte de Decisiones para un Hospital del Sistema de Salud Público (Basurto, 2008)

La tesis presentada por la Ingeniera Karla Basurto presenta una propuesta diseñada para un hospital del sector público que por medio de la implementación del Data WareHouse (DW) permita soportar el proceso de toma de decisiones de la entidad. Con su proyecto de grado esperaban mejorar la productividad del Hospital en cuanto a la optimización de los recursos del ente, mejorar la calidad de los funcionarios y los pacientes, entre otras, algunas de las conclusiones que se pueden destacar de su tesis son las siguientes:

El trabajo de tesis presenta una solución que los hospitales pueden implementar para satisfacer sus necesidades de gestión, análisis y toma de decisiones. Otorga un panorama de lo que está sucediendo en el hospital y presenta esta información en línea.

Los reportes finales no están limitados a presentar la información calculada en el trabajo de tesis. Si un hospital posee información adicional que desearía presentar en su plataforma de Inteligencia de Negocios, es posible agregar esta información al Data WareHouse para satisfacer esta necesidad. Probablemente sería necesario crear nuevas tablas y agregar nuevos procesos al ETL. El presente trabajo de tesis deja abierta esta posibilidad. Sólo se presentan los procesos básicos que posee un hospital, no se trabaja con procesos adicionales que pudiera contener un hospital en particular.

La creación de un Data WareHouse previa a el desarrollo de los DataMarts, según la arquitectura planteada por Inmon, ayuda a que el hospital tenga toda su información consolidada y ordenada en un solo lugar, lo cual es muy importante en este tipo de organizaciones debido a la sensibilidad e importancia de la información, y brinda coherencia entre todos los Data Marts, pues estos partirían desde una misma fuente de información.

Tener todos los datos consistentes y ordenados en el Data Warehouse brinda una fuente confiable y estandarizada para el desarrollo de futuros Data Marts o para la ampliación del alcance de los existentes, facilitando el desarrollo de estos.

Es muy importante desarrollar una buena fase de análisis para evitar que a lo largo del proyecto surjan problemas que ameriten una reestructuración de los procesos, mapeos o de los reportes mismos. Algunos inconvenientes no saltan a la vista hasta que se tiene el reporte terminado, puesto que saltan incongruencias en los datos del informe o se identifica que los

datos no eran agregables y por lo tanto se está presentando información incorrecta. En estos casos, se debe regresar a los procesos anteriores para resolver el problema.

Utilización de información histórica para decisiones empresariales (Peña, Suarez, 2005)

El trabajo realizado y presentado por los estudiantes de Ingeniería Juan David Peña y Jesús Armando Suarez, consiste en aplicar técnicas de Inteligencia de negocios apoyadas en un Data WareHouse (DW) que permita soportar procesos de toma de decisiones en una compañía para mejorar su productividad y ventaja competitiva en cualquier negocio que se aplique, de este trabajo cabe resaltar las siguientes conclusiones que ayudaron al planteamiento de nuestro proyecto de grado como Ingenieros.

El Data WareHouse es mucho más que un producto. Es una serie de productos relacionados entre sí que por medio de procesos bien estructurados y bien definidos proporcionan información útil para los usuarios. Para que una implementación de estas tenga éxito los usuarios deben aprovechar esta información y ejecutar decisiones y acciones basadas en el conocimiento adquirido.

El desarrollo óptimo del proceso de implementación de la data WareHouse se encuentra estrechamente relacionado con la existencia de facilitadores dentro de la organización para el suministro de la información y los requerimientos.

El uso de software libre en inteligencia de negocios tiene la ventaja de ser desarrollado a la medida de las necesidades del negocio. Por lo tanto, ofrece un análisis más personalizado y con menores restricciones que el llevado a cabo por las herramientas comerciales, a las cuales se debe ajustar el negocio. Con el software libre, la solución es la se ajusta a el negocio.

Hoy en día las empresas y las organizaciones valoran mucho las estrategias que se tengan para la respectiva gestión de la información y esta a su vez se convierte en una herramienta clave para poder competir de una mejor manera con el mercado cambiante. La información que se maneja en una compañía es de vital importancia ya que en ella se fundamenta la toma de decisiones sirviendo como un pilar de crecimiento y gestión para la consecución de sus objetivos, metas, visión y misión de la organización.

Actualmente existen empresas que poseen un gran número de datos, pero que a su vez no cuentan con una información oportuna que les hable, que les de soporte para mejorar a la organización. La mayoría de las empresas cuenta con la información, pero no saben cómo explotarla, no saben que esta información puede ser muy valiosa para mejorar la empresa a nivel competitivo, y es aquí donde se incorpora el concepto de Inteligencia de Negocios, como una solución oportuna para el soporte de la toma de decisiones, mediante la generación y análisis de la información que tiene como objetivo un conocimiento organizacional.

## **2.1.2 Históricos**

### **2.1.2.1. Inteligencia de negocios.**

Business Intelligence (BI) o también conocido como inteligencia de negocios pretende dar una visión a la información y conocimiento a la organización.

Para definir mejor este concepto se partirá de una definición de Garther:

“Es una variedad de aplicaciones y tecnologías para obtener, almacenar, analizar, compartir y proveer acceso a los datos para ayudar a los usuarios de las empresas a tomar mejores decisiones”

Teniendo como base la anterior definición se puede citar una definición más general y practica para proyectos.

“La Inteligencia de Negocios es un proceso, el cual plantea una solución al negocio utilizando un conjunto de metodologías, aplicaciones y tecnologías capaces de transformar los datos e información desestructurada en información estructurada generando un mejor conocimiento para beneficio de la organización.”

Importancia y beneficio la inteligencia de negocios en las organizaciones:

Como se mencionó anteriormente en la introducción, la gran mayoría de las empresas carecen de información y por ende conocimiento, por lo que esto puede ser considerado como una gran amenaza por las organizaciones modernas. En la ejecución de actividades de los negocios se depende mucho de las decisiones que se tomen para que garantice la eficiencia del trabajo. Es por esto que el factor más importante que persiguen las organizaciones, es contar con un conocimiento necesario sobre los negocios y las necesidades de explotar sus datos, por lo que tratan de buscar mecanismos que les ayuden en la consecución de los objetivos, para ello las empresas apuntan a la inteligencia de negocios que tiene como objetivo primordial, eliminar la falta de información y conocimiento en los ambientes empresariales, aumentando su desempeño y competitividad



Ilustración 1. Objetivo de la inteligencia de negocios (Dataprix, s.f.)

Dentro de los beneficios de implementar una solución de este tipo se destaca como su principal objetivo la ayuda de la toma de decisiones que proporciona un sistema inteligente. A continuación, se enumerarán algunos de los beneficios que se pueden conseguir en una compañía que trabaja con Inteligencia de Negocios.

#### 2.1.2.1.1 Beneficios Tangibles

- Aumentar ingresos en las ventas.
- Aumentar la compra de productos frecuentes.

- Disminuir productos poco comercializados.
- Aumentar la producción de los productos más consumidos.
- Aumento de la información de calidad con tiempo de respuesta más cortos.
- Disminuir el gasto de productos innecesarios.

#### 2.1.2.1.2 Beneficios Intangibles

- Menor dependencia de las áreas en solicitud de la información.
- Mayor conocimiento de los empleados sobre la empresa.
- Mejora en el control de los gastos.
- Mejora la atención al cliente.
- Corregir los errores que se han cometido.

#### 2.1.2.1.3 Beneficios estratégicos

- Aumento por parte de los directivos sobre la visión de la empresa.
- Aumento en la gestión de las Gerencias de la compañía.
- Incremento de Habilidades de los directivos en la toma de decisiones.

Arquitectura de la inteligencia de negocios:

Las aplicaciones analíticas incorporan las mejores prácticas tecnológicas orientadas a entregar una solución completa y eficiente, a continuación, se presenta una solución de BI.

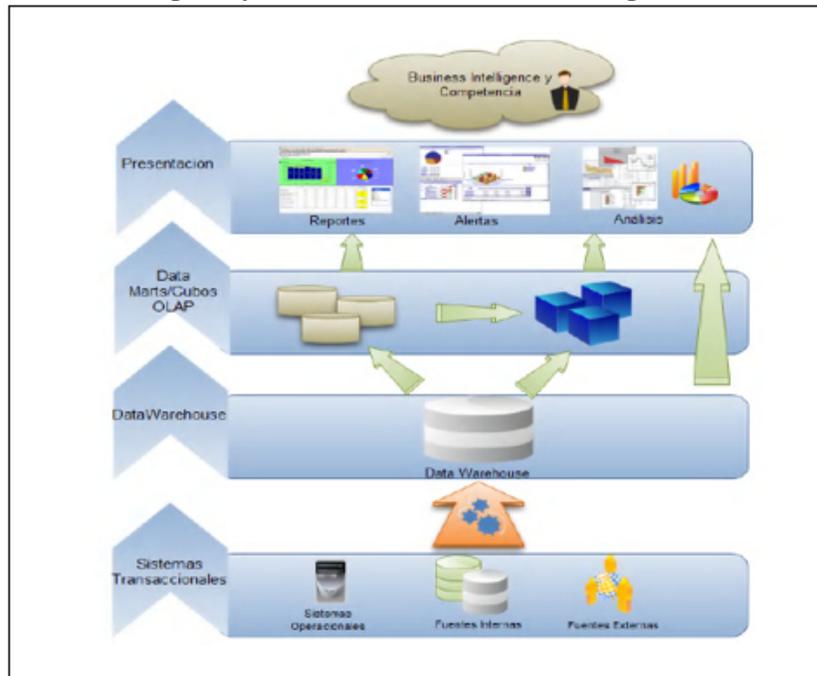


Ilustración 2.Arquitectura de BI (Dataprix, s.f.)

Como se observa en la figura 2, una solución de inteligencia de negocios inicia su fase desde los sistemas de origen (Bases de Datos, ERP, archivos planos, Etc), sobre los que suelen ser necesarios aplicar un proceso de ETL (Extracción, Trasmformación y Carga) para mejorar el

proceso analítico, posteriormente estos datos son almacenados y presentados de forma diferente.

#### 2.1.2.1.4 Etapas de la inteligencia de negocios

El proceso de BI garantiza un ciclo que permite generar información analítica y oportuna con el fin de suplir la necesidad del usuario final.

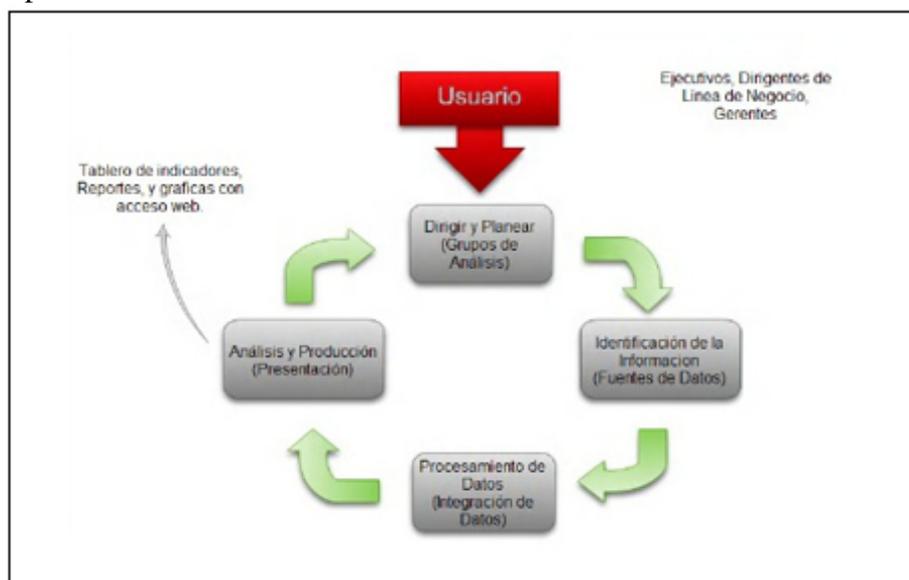


Ilustración 3. Etapas de un BI (Kimball)

En la figura 3 se puede apreciar las etapas de la inteligencia de negocios, las cuales se explicarán más adelante.

##### 2.1.2.1.4.1 Dirigir y planear

La primera etapa y más importante de la Inteligencia de Negocios en la cual se inicia según las necesidades organizacionales, los requerimientos de los usuarios, estas necesidades son planteadas por medio de preguntas como:

- Cuál fue el producto que más generó ventas en cierto periodo de tiempo
- Donde se concentra la mayor parte de la venta
- Cuál es el margen de ganancia de un producto.
- Gracias al análisis de estos cuestionamientos se producen las necesidades a implementar en el DW.

##### 2.1.2.1.4.2 Toma de decisiones.

Se denomina toma de decisiones a una actividad que realizan un grupo de personas en acción a un tema en específico para resolver inconvenientes presentados, estas se presentan en dos clases:

###### 2.1.2.1.4.2.1 Estructuradas

Son los tipos de decisiones que son programadas ya que son decisiones que se repiten mucho y son rutinarias de día a día de la operación.

Por lo general las toman los gerentes muy a menudo y que se rigen por las políticas de la compañía.

#### *2.1.2.1.4.2.2 No Estructuradas*

Se conocen como las que no son programadas debido a que no se presentan de manera frecuente y que para su resolución se requiere tener en cuenta cierta información. Estas tienen que ver con problemas complejos dentro de la organización.

Esta es la primera etapa y luego de ella se procede al planteamiento del proyecto BI y se tienen en cuenta punto a consideración como:

- Se cuenta con el apoyo de las áreas para su desarrollo.
- Se cuenta con el apoyo de las áreas interesadas a las cuales se suplirá la información.
- Se cuenta con el personal capacitado para la implementación.
- La compañía cuenta con la infraestructura necesaria para el desarrollo del proyecto.
- Considerar el tiempo necesario de acuerdo al negocio en el que se pretende dar la solución.

#### *2.1.2.1.5 Identificación de la información.*

Para comenzar con la segunda etapa se debe tener en cuenta lo enunciado en la etapa anterior para poder así plantear los requerimientos del negocio.

##### *2.1.2.1.5.1 Requisitos funcionales*

- Qué tipo de información necesitan las personas.
- Cuáles son los reportes que se requieren para el análisis.
- Qué tipo de reportes son los que desean.

##### *2.1.2.1.5.2 Requisitos de Datos*

- Que datos de la empresa necesitan.
- Que datos son de vital importancia para la compañía.
- Hasta que nivel de detalle se requiere la información.

##### *2.1.2.1.5.3 Requisitos Históricos*

Desde que año necesitan la información:

Finalizando la etapa de los requerimientos se continúa con la recolección de los datos y la identificación de las fuentes de los diferentes sistemas de la empresa, y cuáles de ellas van a ser analizadas, en esta etapa es importante tener en cuenta cuales fueron las necesidades inicialmente planteadas para poder dar inicio a la identificación de los datos.

Para poder identificar los datos es de vital importancia ítems como:

- Saber la ubicación de los datos que van a ser analizados, en que bases de datos se encuentran.
- Se cuentan con manuales o diccionarios de datos de las áreas a analizar.
- Saber que personal que es dueño de la información.
- Conocer el grado de limpieza de los datos a consultar.

- Saber identificar como se evidencian las reglas del negocio en los sistemas de base de datos.

#### *2.1.2.1.5.4 Procesamiento de los datos*

En esta etapa se da lugar al proceso de integración de los datos, es acá donde se diseña y se determina la profundidad de la información, las transformaciones que debe sufrir y las estructuras destino donde se va a almacenar dicha información, en nuestro caso se hará en un Data Warehouse.

Para nuestro caso la integración de los datos se realizara por un proceso llamado ETL (Extracción, Transformación y Carga), el cual es considerado como base para la calidad de la información con que se contara al final de una implementación de un proyecto de inteligencia de negocios, ya que se extraerán los datos de las fuentes identificadas y los necesarios de acuerdo al negocio, luego se realiza la transformación de datos erróneos, inconsistentes alojados en diferentes sistemas para que finalmente sea cargados al DW.

#### *2.1.2.1.5.5 Análisis y producción.*

Esta etapa es donde se da lugar al análisis de la información, que se va a generar a partir de los datos encontrados de una manera estructurada de acuerdo a la necesidad del negocio en la etapa de planteamiento.

Es en esta fase donde mediante la presentación de informes gerenciales, tableros de mando, cubos o minería de datos se hará el despliegue de la información para así apoyar la toma de decisiones de la empresa.

### *2.1.2.2 Data Warehouse*

#### *2.1.2.2.1 Introducción.*

Gracias al conocimiento que ya se adquirió sobre el concepto de Inteligencia de Negocios (BI) se puede apreciar la importancia de la construcción de un Data Warehouse como una solución a este problema.

Un DW es una base de datos accesible a los usuarios en el cual se podrán encontrar registros históricos y actuales de las áreas más importantes de la compañía y de los negocios en específico.

El DW Organiza y aloja datos necesarios para ser utilizados en el proceso analítico dentro de una perspectiva de tiempo definida.

Para ampliar esta definición se tomará una definición de Ralph Kimball y Bill Inmon donde exponen el DW como:

“Una copia de la transacción de datos específicamente estructurado para consulta y análisis” (Kimball)

“Orientado al tema, integrado, de tiempo variante, de colección de datos no volátil en apoyo a la gestión del proceso de toma de decisiones” (Inmon)

Al implementar un DW en una compañía se facilita la consulta de información que a simple vista sin aplicar procesos de limpieza no permite su fácil lectura ni permite la toma de decisiones con facilidad.

#### *2.1.2.2.2 Características*

A continuación, se presentan las características principales de una Data Warehouse en cuanto a sus datos.

- Orientado al tema.
- Integrado.
- De Tiempo Variable.
- No volátil.

#### *Orientado al tema*

Se debe a una característica que tiene el DW ya que clasifica la información de acuerdo al interés de la empresa, por ejemplo, el cliente, sector, almacén.

#### *2.1.2.2.2 Integrado:*

En el DW la información se encuentra de forma integrada ya que se corrigen los problemas con los datos, la inconsistencia de datos, uniformidad, diferente codificación de distintas fuentes. La integración permite estandarizar los datos para su respectivo almacenamiento.

#### *2.1.2.2.3 De tiempo Variante:*

Se refiere a la posibilidad de consultar información de años atrás de manera horizontal.

#### *2.1.2.2.4 No volátil:*

Esta característica se debe a que el DW no sufre operaciones de inserción, eliminación, modificación, tan solo se realizan operaciones que son la carga de los datos y el acceso a los mismos.

2.1.2.2.3 El Data Warehouse como producto presenta las siguientes características.

- Fácil accesibilidad a la información organizacional.
- Información totalizada y detallada.
- Presentación consistente de la información.
- Permite realizar análisis rápidamente.

#### 2.1.2.2.4 Arquitectura de un Data Warehouse.

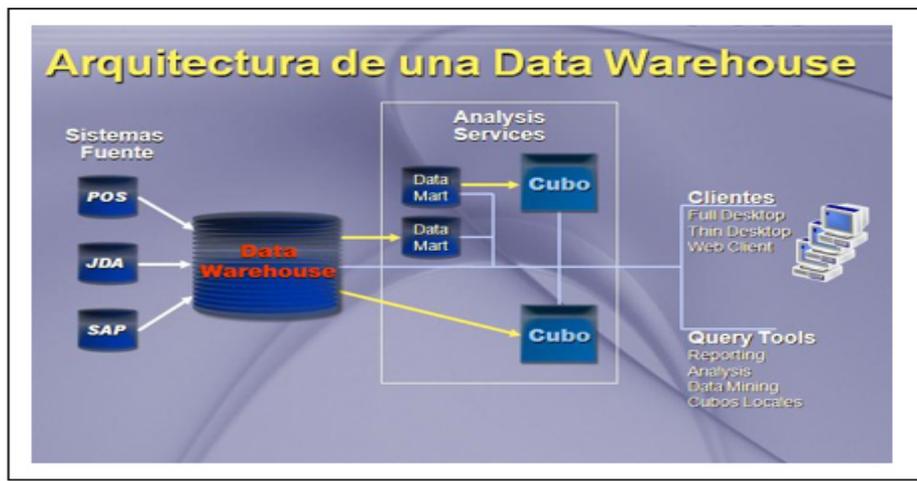


Ilustración 4. Arquitectura de un Data Warehouse (Dataprix, s.f.)

En la figura 4 se puede apreciar la arquitectura de un almacén de datos o Data Warehouse

#### 2.1.2.2.5 Sistemas de origen

Se denominan sistemas de origen a las ubicaciones donde se encuentra la información de interés y que serán objeto de la extracción, transformación, estos orígenes pueden ser bases de datos, ERP, hojas de cálculo, cubos, o archivos planos entre los que se proveerán los datos de interés.

#### 2.1.2.2.6 Stage area (área temporal):

Es un área del DW donde se almacena la información de forma temporal, es allí donde se alojan los datos extraídos de los sistemas origen según las necesidades planteadas del negocio, su función principal es minimizar la afectación de los sistemas de origen ya que cuando se realiza una carga al DW, los orígenes de los datos intervendrán hasta que se realice la próxima carga.

#### 2.1.2.2.7 Ods (operational data store):

Se denominan almacenes de datos operativos y se conocen como un área que da soporte a los sistemas operacionales. Este modelo de datos ODS sigue la estructura relacional y normalizada, característica por la cual se le trata de diferente manera a la de un almacén de datos y con lo que le permite a cualquier herramienta de reporte o sistema operacional acceder a los datos. Los ODS forman parte de una DW en el sentido que el almacén de datos aprovecha los datos que se encuentran previamente integrados a él, permitiendo dar el soporte a todas las transacciones operacionales.

### 2.1.2.2.8 Estructura de DW

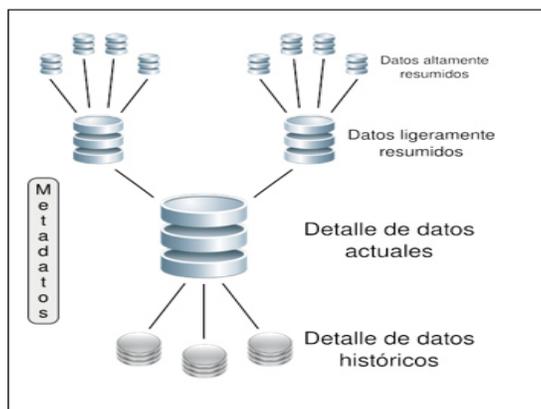


Ilustración 5. Estructura de un DW (Slid, s.f.)

Como se muestra en la figura 5 la estructura de una Data Warehouse es básicamente la referente a la lógica de almacenamiento de los datos, los cuales se caracterizan por tener diferentes niveles de esquematización y detalle como se aprecia en la figura 5, en la que se identifican los niveles de los datos, los metadatos del repositorio. Teniendo en cuenta estos aspectos se puede decir que la estructura del DW es de la siguiente manera.

### 2.1.2.2.9 Proceso de ETL (extracción, transformación y carga).

Existe un conjunto básico de procesos de suma importancia detrás de una implementación de un DW, que permitan garantizar la calidad de los datos que en ella se almacenarán.

A este proceso se le conoce con el nombre de ETL o proceso de integración de datos que cumple con la función principal de organizar e integrar el flujo desde múltiples fuentes hacia un destino definido que es el almacén de datos. El proceso de ETL brinda un soporte a la gestión de datos a realizar, obteniendo calidad en los mismos dentro de un almacén. Dicho proceso se observa a continuación:

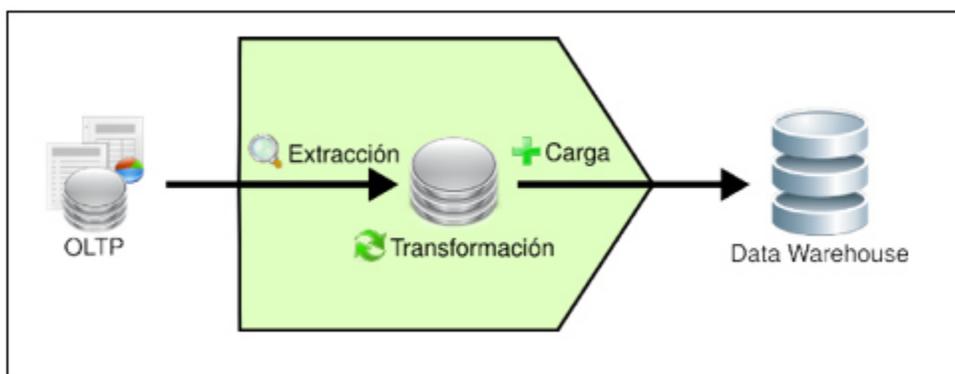


Ilustración 6. Proceso de ETL (Dataprix, s.f.)

Como se muestra en la figura 6, todo proceso que se lleva a cabo dentro de un DW se debe especificar en qué tiempos se debe realizar, lo que garantiza que se mantenga al día con los datos del almacén. Aunque esto inicialmente se define a las necesidades de la organización

y de los usuarios finales, ya que con esto se pueden definir cargas a nivel diario, semanal o mensual.

Este proceso general se encuentra dividido en tres subprocesos que se definen a continuación:

#### 2.1.2.2.9.1 Extracción

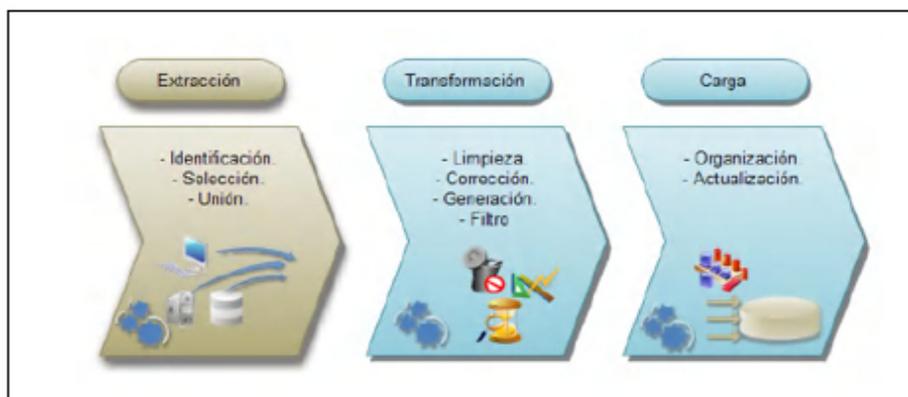


Ilustración 7. Proceso de Extracción de ETL (Postech, s.f.)

En la figura 7 la extracción hace referencia a la ubicación de los datos, los cuales pueden ser recogidos de diferentes fuentes como archivos planos, cubos, Bases de datos, ERP, etc.

Las extracciones de los datos se almacenarán en un área temporal o Stage, vale la pena aclarar que solo se extraerán los datos necesarios, es decir de acuerdo a los requerimientos que se plantearon inicialmente ya que al hacer la extracción nos podremos encontrar con datos que no son necesarios.

#### 2.1.2.2.9.2 Transformación

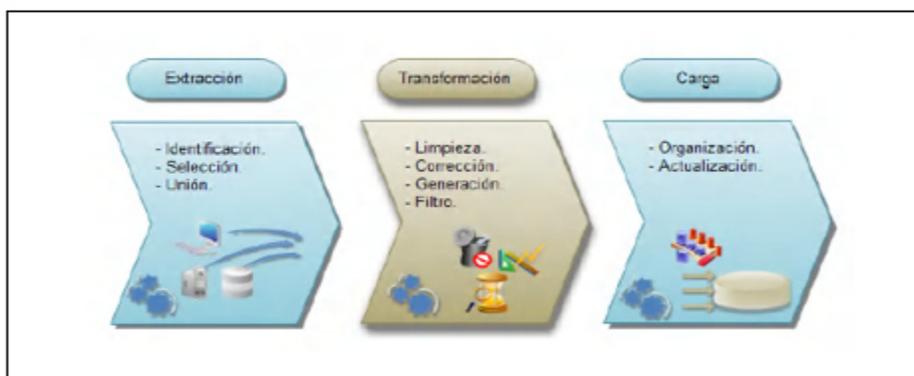


Ilustración 8. Proceso de Transformación de ETL (Postech, s.f.)

Como se aprecia en la figura 8, este es el proceso que más capacidad de análisis y labor requiere con respecto a los otros dos ya que en esta etapa es donde se realiza el refinamiento de los datos que inicialmente han sido extraídos de las distintas fuentes, por lo que se especificara pasos de acuerdo a los datos que van a ser tratados dando un valor agregado a

los usuarios. Este proceso incluye corrección de errores, decodificación, borrado de campos y datos que no son necesarios, generación de claves, niveles de agregación de la información, Etc.

#### 2.1.2.2.9.3 Carga



Ilustración 9. Proceso de Carga de ETL (Postech, s.f.)

Como se muestra en la figura 9 y en la parte de carga, este es el último proceso que se caracteriza por la carga de la información extraída y transformada al DW de los orígenes escogidos anteriormente, es en este paso donde se escogen métodos y maneras para la carga de los datos con el fin de tener siempre la información actualizada.

Finalizado el proceso de ETL, se pretende contar con datos relevantes para el negocio, los mismos que deben ser sin ningún tipo de codificación, es decir datos transparentes y entendibles a los usuarios.

#### 2.1.2.2.10 Modelo dimensional de una DW.

Un DW adopta un modelo dimensional en su estructura de almacenamiento por ser un esquema de estrella o copo de nieve, lo que permite maximizar el rendimiento de las consultas.

A continuación, se nombra una definición de un modelo dimensional.

“Es una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del negocio (Carmen Gloria Wolff)”.

Un diseño dimensional en estrella es muy distinto a un esquema de base de datos operacional, en este último los datos están altamente normalizados para soportar frecuentes actualizaciones y para mantener la integridad referencial, en cambio en un diseño de un DW los datos están desnormalizados o redundantes para proporcionar acceso inmediato a los datos sin tener que realizar una gran cantidad de relaciones.

#### 2.1.2.2.10.1 Esquema estrella

Se considera la técnica más popular para el diseño de un DW, es un paradigma en el cual el único objeto en el centro (es conocido como una tabla de hechos), se encuentra conectado radialmente con otros objetos alrededor llamados tablas de dimensiones formando así una estrella.

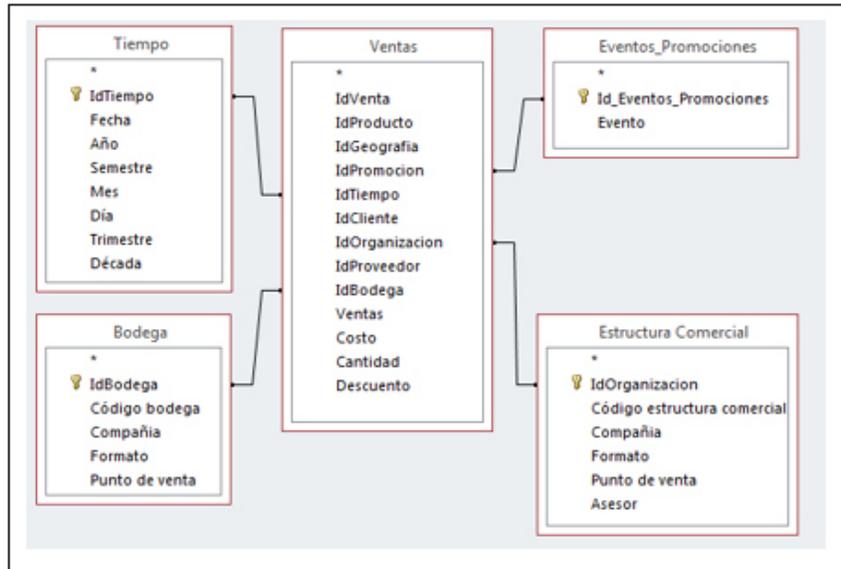


Ilustración 10. Esquema de estrella (Dataprix, s.f.)

Como se muestra en la figura 10, este tipo de esquema es el más utilizado en el desarrollo de los DW debido a que los datos se encuentran des normalizados y por ende las consultas que se pueden realizar son mucho menos complejas ya que no existe la necesidad de realizar muchas relaciones entre las tablas.

### 2.1.2.2.10.2 Esquema de copo de nieve.

Este esquema copo de nieve es una extensión del esquema de estrella, donde cada punta se explota en más puntas y su denominación se debe a que el diagrama de asemeja a un copo de nieve.

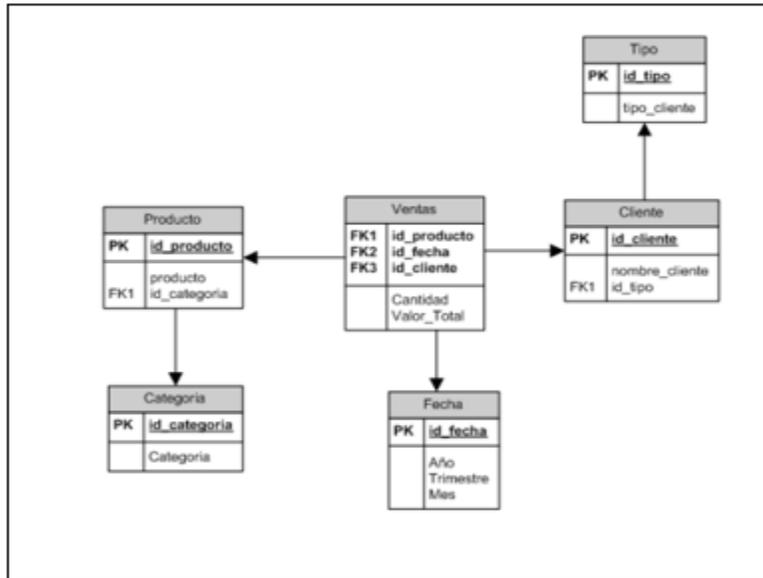


Ilustración 11. Esquema de Copo de Nieve (Dataprix, s.f.)

En este esquema se normalizan dimensiones para eliminar la redundancia, permitiendo que los datos de las dimensiones se agrupen en múltiples tablas en lugar de una tabla grande.

### 2.1.2.2.10.3 Tablas de Hechos.

La tabla de hechos, es una tabla central en el esquema dimensional, que contiene valores de las medidas del negocio a ser analizadas como se observa a continuación, en un modelo de tipo de datos de estrella. Este tipo de tabla representa el hecho o actividad del negocio como, por ejemplo, las ventas, movimientos, pedidos, clientes, Etc. Por lo general estos datos son de tipo numérico y pueden agruparse en valor total, las medidas de referencia pueden ser, por ejemplo, cantidad vendida, costo, precio. Es decir, son los indicadores que permitan medir los hechos que se realizan en el negocio.

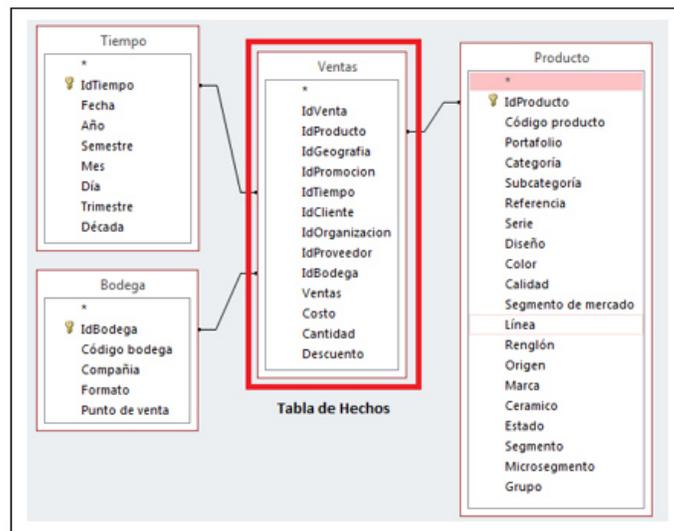


Ilustración 12. Tabla de Hechos (Dataprix, s.f.)

Una tabla de hechos está dividida por 2 tipos de atributos, estos son: las llaves foráneas provenientes de otras dimensiones y por los indicadores o medidas del hecho.

#### 2.1.2.2.10.4 Dimensiones

Son las tablas que describen a la tabla de hechos, mediante los atributos descriptores que poseen de acuerdo a un tema específico del negocio como, por ejemplo: clientes, productos, Ubicación geográfica.

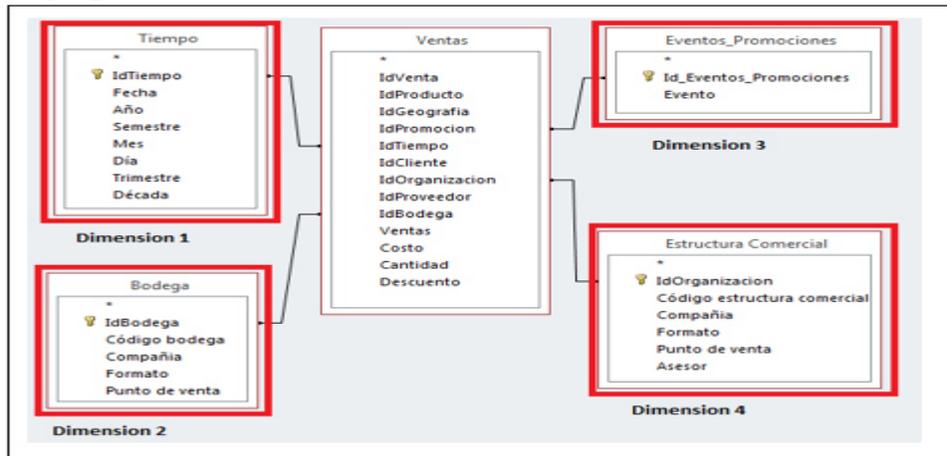


Ilustración 13. Tabla de Dimensiones (Dataprix, s.f.)

Como se muestra en la figura 13, al igual que las tablas de hechos, las tablas de dimensiones también cuentan con 2 tipos de atributos: una clave primaria única para cada registro y la de los atributos descriptores de ellas.

#### 2.1.2.2.10.4 Nivel

En las dimensiones se pueden especificar características (Atributos) en particular de acuerdo a cada dimensión, Estos atributos necesitan ser ordenados o agrupados con un fin específico.

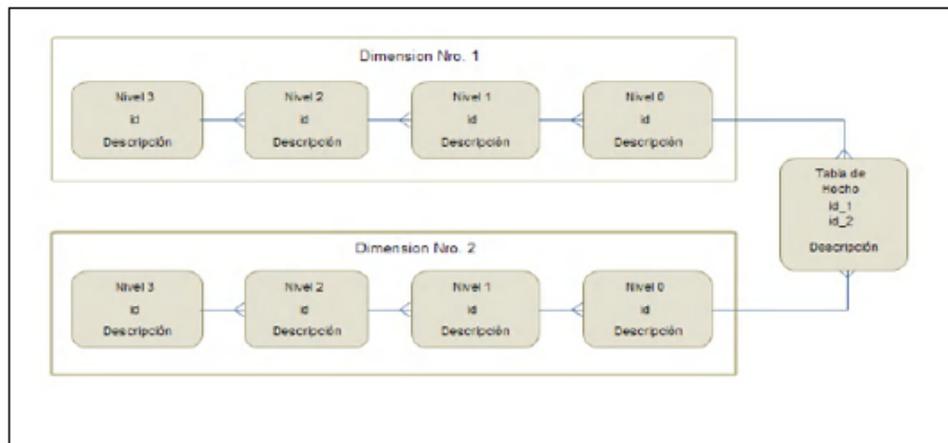


Ilustración 14. Niveles de la dimensión (Dataprix, s.f.)

Para garantizar que existan registros únicos en cada dimensión es necesario contar con una clave única para cada nivel de agregación, la misma que se recomienda que sea secuencial.

#### 2.1.2.2.10.5 Jerarquía

Las jerarquías son estructuradas lógicas para categorizar los datos de acuerdo a los niveles definidos, la jerarquía también ayuda a definir la navegación por la información que va a tener la información en un orden definido.

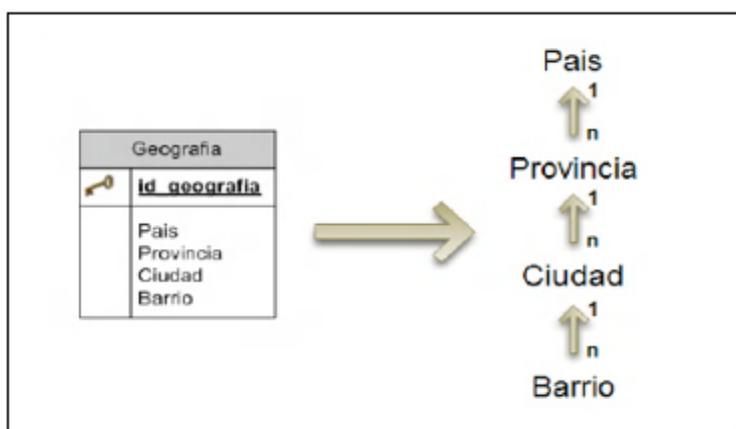


Ilustración 15. Figura de Jerarquía (Dataprix, s.f.)

#### 2.1.2.2.11 Dimensiones lentamente cambiantes (SCD).

Estas dimensiones hacen referencia al destino de la información en las tablas y se pueden clasificar de la siguiente manera:

Tipo 0 (SCD): En este nivel no se tienen en cuenta los cambios históricos, de alguna manera u otra la información va a ser sobrescrita, pero otra no. En esta fase no se tiene en cuenta la planificación.

Tipo 1 (SCD): En esta fase no se guarda información, la nueva se sobrescribe la antigua, principalmente la escritura se realiza por errores en la calidad de datos. Este tipo de dimensiones es fácil de mantener y son usadas cuando la información histórica no es importante.

Tipo 2 (SCD): En esta etapa toda la información histórica se almacena en el DW. Cuando sucede un cambio se crea una nueva entrada con su fecha y llave subrogada apropiadas. A partir de ese momento será el valor usado para las futuras entradas.

Tipo 3 (SCD): En esta etapa, toda la información es almacenada en el DW, en esta fase se crean nuevas columnas con los valores antiguos y los actuales son reemplazados con los nuevos.

### 2.1.2.2.12 Granularidad.

La granularidad en el Data Warehouse se define el nivel de detalle con el cual se almacena la información.

Algunos ejemplos:

- Datos de las ventas registradas en un punto de venta por día.
- Datos de clientes que nos compraron en un lapso de tiempo específico.
- A mayor nivel de detalle existe mayor posibilidad analítica ya que los mismos datos podrán ser resumidos o totalizados.
- Los datos con granularidad fina (Nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa.

### 2.1.2.2.13 Relaciones

Las relaciones definidas son de vital importancia ya que mantienen la integridad de la información referenciando datos entre distintas tablas, en este caso relacionando dimensiones con los hechos, como por ejemplo la relación de Producto – Ventas, Bodega – Pedidos

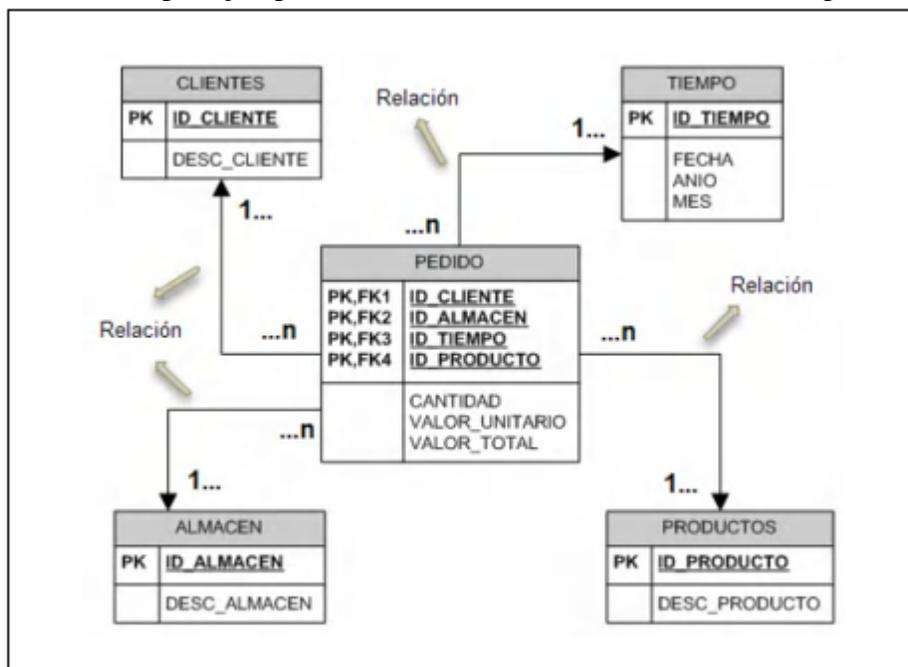


Ilustración 16. Figura de Relaciones (Dataprix, s.f.)

#### 2.1.2.2.14 Llaves subrogadas.

Se caracterizan por formar parte de las dimensiones y de los hechos, al ser un campo único o identificado único de la tabla de dimensión, lo que permite que se mantenga la integridad en el modelo dimensional con el hecho. Las claves subrogadas son independientes a las claves primarias de los orígenes de datos, por lo general estas son creadas por los procesos ETL.

Estas llaves deben ser de tipo (INTEGER) para que se garantice el rendimiento del modelo dimensional.

#### 2.1.2.2.15 Explotación del Data Warehouse.

La explotación de un DW, es prácticamente en donde se pueden identificar los resultados que puede llegar a brindar la implementación de este tipo, ya que como se ha venido observando durante el trabajo la creación de un almacén de datos no es la solución completa de un proyecto de BI, y es solamente el medio para dar como resultado a la inteligencia del negocio, este objetivo es cumplido solamente si se lleva a cabo la adecuada explotación de los datos almacenados en un DW, es decir mediante un interfaz gráfica se puede hacer un constante monitoreo del negocio, como por ejemplo como van las ventas de un almacén para el caso de Almacenes Corona, como estamos con respecto al inventario, poder plasmar un top 20 de los productos más vendidos en general en la compañía o si se desea a nivel específico en un punto de venta. Entre los más utilizados podemos destacar los siguientes.

#### 2.1.2.3.16 Reportes y consultas.

Dentro de cualquier organización es de vital importancia los reportes que se puedan obtener acerca del estado actual del negocio a un nivel detallado, así como también consultas a usuarios, distintas gerencias o si desea realizar algo específico, a estas consultas se le conoce con el nombre ad-hoc.

Las consultas trabajan tanto en el detalle como en el nivel de agregación de la información. Al realizar este tipo de explotación desde un almacén de datos se optimiza el entorno de los reportes e informes dado que el DW mantiene una estructura y una tecnología mucho más apropiada que permita tener la información en menos tiempo y con uno solo clic.

Los sistemas de consultas y reportes no basados en almacenes de datos se caracterizan por la complejidad de las consultas, los altos tiempos de espera y algunas veces con los mismos ERP implementados en la compañía ya que comparte el mismo entorno, esto se debe a la alta normalización de los datos.

### 2.1.2.3 Minería de datos

“Este es un proceso de descubrir correlaciones significativas, patrones de tendencia de cribado a través de grandes cantidades de datos almacenados en repositorios. La minería de datos utiliza tecnologías de reconocimiento de patrones, así como técnicas de estadística y matemáticas”. (Gartner 2009)

Teniendo en cuenta el anterior concepto se puede definir la minería de datos como una automatización de la búsqueda de patrones relevantes en un base de datos, por ejemplo, un patrón podría ser cuentas mujeres casadas con hijos tienen auto propio, en este caso la información puede ser muy valiosa para un gerente de una empresa automotriz.

El proceso de minería de datos se integra con el sistema para el almacenamiento de cantidades masivas de datos (DW) de la empresa y presenta la información dirigida a los usuarios con la capacidad de decisión en el negocio. Para que la minería de datos sea exitosa en una empresa es necesario que estas dispongan de datos de calidad (Completos, estructurados, relevantes, históricos, actualizados) para el propósito del DW.

### 2.1.2.4 Análisis OLAP

OLAP significa (On line Analytical Processing) o proceso analítico en línea, esto surge como contraste a OLTP (On Line Transactional Processing) que define a los sistemas de ambientes transaccionales.

El análisis multidimensional (Análisis OLAP), parte de una visión de la información como dimensiones del negocio, en la que hay que tener en cuenta que se debe olvidar lo que son las tablas y campos, dando mayor énfasis a lo que son las dimensiones y las medidas.

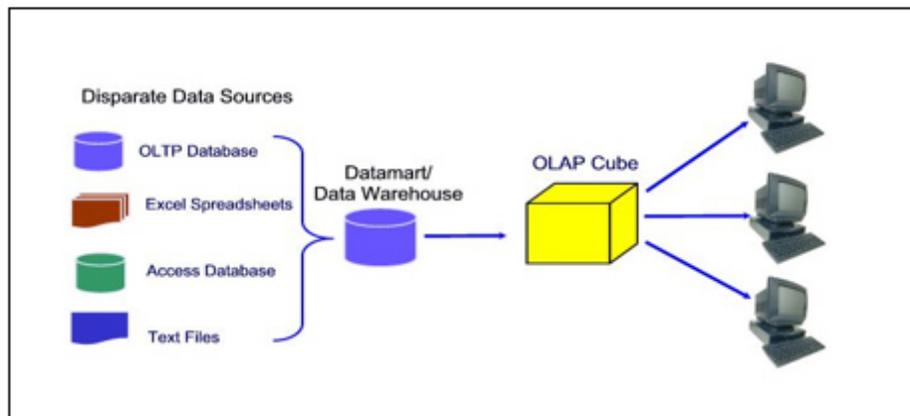


Ilustración 17. Análisis OLAP (Wikispaces, s.f.)

En la figura 17, se observa la funcionalidad del análisis OLAP

Esta herramienta ofrece a los usuarios el acceso a los almacenes de datos DW para que puedan tener las funcionalidades de análisis avanzados que requieren los usuarios finales y analistas, por lo que es considerada una herramienta ideal para sacar el mayor provecho de los datos almacenados en un repositorio.

Las herramientas OLAP se caracterizan por dividirse en 3 tipos de acuerdo a la manera de almacenar los datos, estos se nombran a continuación.

#### 2.1.2.4.1 ROLAP (RELATIONAL OLAP)

Es una arquitectura en la que se almacenan los datos en un motor de base de datos relacional, pero de igual manera se proporciona la funcionalidad analítica. A través de esta implementación se soporta de mejor manera las capacidades OLAP con respecto a las bases de datos relacionales, en el sentido que realiza consultas directas a la base de datos de la manera multidimensional caracterizada por la arquitectura.

#### 2.1.2.4.2 MOLAP (multidimensional OLAP).

En este tipo de arquitectura los datos se almacenan de manera dimensional en un servidor de base de datos multidimensional, permitiendo optimizar los tiempos de respuesta en la información, ya que al ser totalizada o agregada ayuda mucho a los datos calculados por adelantado.

#### 2.1.2.4.3 HOLAP (HYBRID OLAP).

Es una arquitectura Híbrida que se caracteriza por combinar las arquitecturas anteriores ROLAP y MOLAP para brindar una mejor solución y desempeño. Este tipo mantiene los registros al detalle en la base de datos relacional mientras que mantiene las agregaciones en un almacén MOLAP separado.

#### 2.1.2.4.4 Operaciones y funcionalidades de OLAP

En los diferentes análisis OLAP se implementan funcionalidades u operaciones las cuales permiten interpretar la información de distintas maneras, a continuación, se definen las operaciones más importantes.

#### 2.1.2.4.4.1 Drill Down

Se entiende por desglosar una métrica de lo general a lo particular por la jerarquía de sus dimensiones

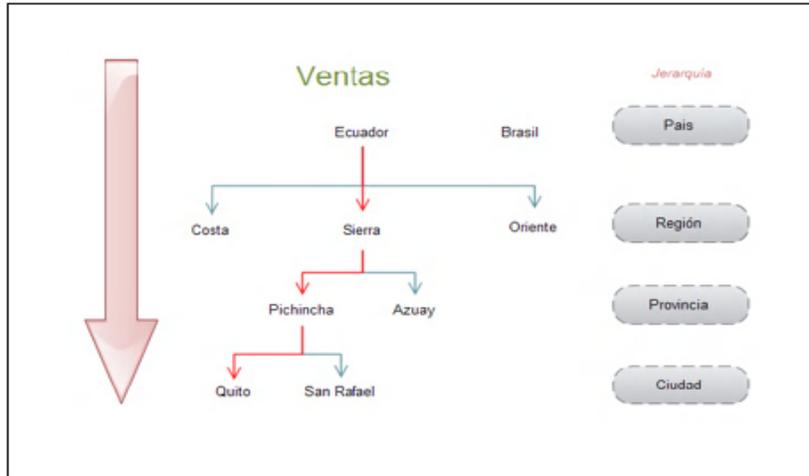


Ilustración 18. Drill Down

La figura 18, muestra la jerarquía descendente de las dimensiones en Drill Down.

#### 2.1.2.4.4.2 ROLL UP

Agregar una métrica de lo particular a lo general por la jerarquía ascendente de sus dimensiones.

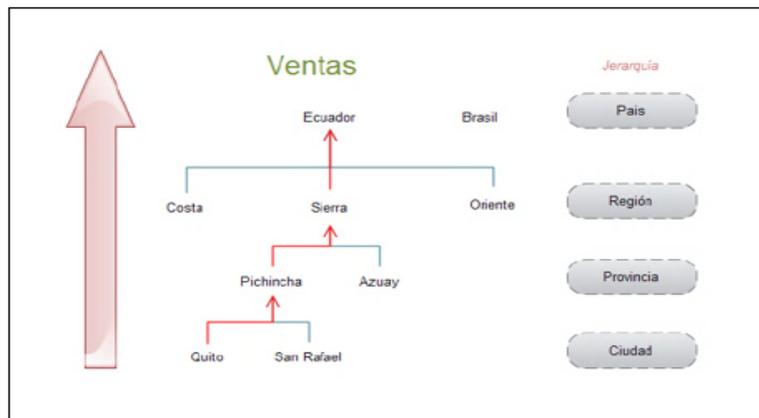


Ilustración 19. Roll Up (Slid, s.f.)

La figura 19, muestra la jerarquía ascendente de las dimensiones en Drill Down.

Existen dos tipos de operaciones de acuerdo a la cantidad de dimensiones utilizadas para el cubo.

DICE: Obtener un Sub cubo fijando dos o más de sus dimensiones.

SLICE: Obtener un Sub cubo fijando una de sus dimensiones.

## 2.2 Bases teóricas

### 2.2.1 Consideraciones de diseño

#### 2.2.1.1 Tamaño del DW.

Dependiendo del negocio, el volumen de datos y el alcance del proyecto, el tamaño del DW puede variar considerablemente, por lo cual, es una buena práctica tener esto en cuenta al momento de diseñar el depósito y al determinar los recursos físicos, los tiempos de desarrollo y los respectivos costos inherentes.

De acuerdo al tamaño del depósito de datos, se lo puede clasificar como:

Personal: si su tamaño es menor a 1 Gigabyte.

$DW < 1 \text{ GB}$

Pequeño: si su tamaño es mayor a 1 Gigabyte y menor a 50 Gigabyte.

$1 \text{ GB} < DW < 50 \text{ GB}$

Mediano: si su tamaño es mayor a 50 Gigabyte y menor a 100 Gigabyte.

$50 \text{ GB} < DW < 100 \text{ GB}$

Grande: si su tamaño es mayor a 100 Gigabyte y menor a 1 Terabyte.

$100 \text{ GB} < DW < 1 \text{ TB}$

Muy grande: si su tamaño es mayor a 1 Terabyte.

$DW < 1 \text{ TB}$

#### 2.2.1.2 Tiempo de construcción.

Diversos autores resaltan la importancia del factor tiempo en la construcción de un DW, por lo cual se ha considerado interesante exponer tres frases seleccionadas al respecto:

El 70% del tiempo total dedicado al proyecto se insume en definir el problema y en preparar la tabla de datos”.

“Estime el tiempo necesario, multiplíquelo por dos y agregue una semana de resguardo”.

” Regla 90 – 90”: el primer 90% de la construcción de un sistema absorbe el 90% del tiempo y esfuerzo asignados; el último 10% se lleva el otro 90% del tiempo y esfuerzo asignado.

#### 2.2.1.3 Implementación.

Las implementaciones de los depósitos de datos varían entre sí de forma considerable, teniendo en cuenta las herramientas de software que se empleen, los modelos que se utilicen, recursos disponibles, SGBD que lo soporten, herramientas de análisis y consulta, entre otros.

#### 2.2.1.4 Performance.

Cuando se diseñan los ETL, es muy importante que los mismos sean lo más eficientes posible, ya que una vez que se tenga un gran volumen de datos, el espacio en disco se volverá

fundamental y los tiempos incurridos en el procesamiento y acceso a la información serán esenciales, y más aún si el DW es considerado o tomado como un sistema de misión crítica. También es muy importante configurar correctamente el SGBD en el que se almacene y mantenga el DW, así como lo es elegir las mejores estrategias para modelar las diferentes estructuras de datos que se utilizarán.

Para mejorar la performance del DW, se pueden llevar a cabo las siguientes acciones sobre el DW y las estructuras de datos (cubos multidimensionales, Business Models, etc):

Prestar especial atención a los tipos de datos utilizados, por ejemplo, para valores enteros pequeños conviene utilizar INT con el fin de no asignar tamaños de datos mayores a los necesarios. Esto toma vital importancia cuando se aplica en las claves primarias, debido a que formarán parte de la tabla de hechos que es la que contiene el volumen del almacén de datos.

- Utilizar Claves Subrogadas.
- Utilizar técnicas de indexación.
- Utilizar técnicas de particionamiento.
- Crear diferentes niveles de totalización.
- Crear vistas materializadas.
- Utilizar técnicas de administración de datos en memoria caché.
- Utilizar técnicas de multiprocesamiento, con el objetivo de agilizar la obtención de resultados, a través de la realización de procesos en forma concurrente.

#### **2.2.1.4 Mantenimiento.**

Un punto muy importante es mantener en correcto funcionamiento el DW, ya que a medida que pase el tiempo, este tenderá a crecer significativamente, y surgirán cambios, tanto en los requerimientos como en las fuentes de datos.

### **2.2.1.5 Impactos.**

Al implementar un DW, es fundamental que los usuarios del mismo participen activamente durante todo su desarrollo, debido a que son ellos los que conocen en profundidad su negocio y saben cuáles son los resultados que se desean obtener. Además, es precisamente en base a la utilización que se le dé, que el depósito de datos madurará y se adaptará a las situaciones cambiantes por las que atraviese la empresa. Los usuarios, al trabajar junto a los desarrolladores y analistas podrán comprender más en profundidad sus propios sistemas operacionales, con todo lo que esto implica.

Con la implementación del DW, los procesos de toma de decisiones serán optimizados, al obtener información correcta al instante en que se necesita, evitando pérdidas de tiempo y anomalías en los datos. Al contar con esta información, los usuarios tienen más confianza en las decisiones que tomarán y en adición a ello, poseerán una base sustentable para justificarlas.

Usualmente, los DW integrarán fuentes de datos de diversas áreas y sectores de la empresa, esto tiene como beneficio contar con una sola fuente de información centralizada y común para todos los usuarios. Esto posibilitará que en las diferentes áreas se compartan los mismos datos, lo cual conducirá a un mayor entendimiento, comunicación, confianza y cooperación entre las mismas. El DW introducirá nuevos conceptos tecnológicos y de inteligencia de negocios, lo cual requerirá que se aprendan nuevas técnicas, herramientas, métodos, destrezas, formas de trabajar, etc.

### **2.1.1.5 DM como sub proyectos.**

Al diseñar e implementar DM como partes de un proyecto DW, se debe tener en cuenta que el análisis que se efectuará, los modelos que intervendrán y el alcance, deben ser globales, con el fin de determinar, por ejemplo, tablas de dimensiones comunes entre las diferentes áreas de trabajo. Esto evitará que se realicen tareas repetidas, ahorrando tiempo y enfocándose en la consolidación, unificación y centralización de la información de los diferentes sectores.

#### **2.1.1.5.1 Teoría de grafos**

Para evaluar la validez de la estructura lógica del depósito de datos, puede emplearse la teoría de grafos, la cual afirma que su estructura será correcta sí y solo sí está conformada únicamente por trayectorias cíclicas. Si se encuentran trayectorias cíclicas, deberán ser transformadas para que las consultas al DW sean válidas y confiables. Una trayectoria cíclica, es aquella que sólo tiene una forma de recorrido (en un solo sentido). Una trayectoria cíclica, es aquella que se puede recorrer en dos o más secuencias diferentes.

#### **2.1.1.5.2 Elección de columnas.**

Cuando se seleccionan los campos que integrarán el DW, se debe tener en cuenta lo siguiente: Se deben descartar aquellos campos cuyos valores tengan muy poca variabilidad. Se deben descartar los campos que tengan valores diferentes para cada objeto, por ejemplo, el número de D.N.I. cuando se analizan personas.

En los casos en que no existan jerarquías dentro de alguna tabla de dimensión, en la cual la cantidad de registros que posee la misma son demasiados, es conveniente, conjuntamente con los usuarios, definirlos. Pero, si llegase a suceder que no se encontrase ningún criterio por el cual jerarquizar los campos, es una buena práctica crear jerarquías propias. El objetivo de llevar a cabo esta acción, es la de poder dividir los registros en grupos, propiciado de esta manera una exploración más amena y controlable.

Entonces, lo que se realizará será crear una nueva jerarquía a partir de los campos disponibles: Además, se pueden aplicar algunas de las acciones que se expondrán a continuación sobre los valores de los campos que se incluirán en el depósito de datos:

**Factorizar:** se utiliza para descomponer un valor en dos o más componentes. Por ejemplo, el campo “código de producto” perteneciente a un producto está formado por tres identificadores separados por guiones medios, que representan su rubro, marca y tipo (“idRubro-idMarca-idTipo”), entonces este campo puede factorizarse y separarse en tres valores independientes (“idRubro”, “idMarca” e “idTipo”).

**Estandarizar:** se utiliza para ajustar valores a un tipo de formato o norma preestablecida. Por ejemplo, se puede emplear este método cuando se desea que todos los campos del tipo texto sean convertidos a mayúscula.

**Codificar:** es utilizado para representar valores a través de las reglas de un código preestablecido. Por ejemplo, en el campo “estado” se pueden codificar sus valores, “0” y “1”, para transformarlos en “Apagado” y “Encendido” respectivamente.

**Discretizar:** es empleado para convertir un conjunto continuo de valores en uno discreto. Por ejemplo, cuando se especificaron los tamaños del DW se realizó esta operación.

**Claves primarias en tablas de Dimensiones.**

Al momento de añadir la clave principal a una tabla de dimensión, se puede establecer:

1. Una única columna que sea clave primaria e identifique unívocamente cada registro.
2. Varias columnas que sean clave primaria e identifiquen en conjunto, unívocamente cada registro.

La primera opción requiere menos espacio de almacenamiento en el DW y permite que las consultas SQL sean más sencillas. La segunda opción requiere más espacio de almacenamiento en el DW, provoca que las consultas SQL sean más complejas y por consiguiente hace que se demore más tiempo en procesar los resultados. Sin embargo, esta última alternativa hace que los procesos ETL sean menos complejos y más eficientes.

Más allá de estas dos grandes opciones, es totalmente recomendable la utilización de Claves Subrogadas.

### 2.1.1.5.3 Relación muchos a muchos

Siempre que sea posible, se debe evitar mantener en el DW tablas de dimensiones con relaciones muchos a muchos entre ellas, ya que esta situación puede, entre otros inconvenientes, provocar la pérdida de la capacidad analítica de la información y conducir a una totalización incorrecta de los datos.

Para explicar esta problemática, se tomará como ejemplo la relación existente entre ríos y provincias, es decir: Una provincia tiene uno o más ríos, y un río pertenece a una o más provincias.

Cuando existe este tipo de relación (muchos a muchos) entre dos o más tablas, se pueden realizar diferentes acciones para solventar esta situación. Una posible solución, sería llevar a cabo los siguientes pasos:

1. Crear una tabla de dimensión por cada entidad que pertenece a la relación. Cada una de estas tablas no debe incluir ninguna correspondencia a las demás. En este caso se crearán dos tablas de dimensiones, DIM\_RIOS (correspondiente a la entidad "RIOS") y DIM\_PROV (correspondiente a la entidad "PROVINCIAS").

2. Crear otra tabla de dimensión (en este caso DIM\_RELACION), que sea hija de las tablas de dimensiones recientemente confeccionadas (en este caso DIM\_RIOS y DIM\_PROV), que estará compuesta de los siguientes campos:

Clave principal: dato auto numérico o autoincrementable (en este caso "id\_dim\_Relacion").

Claves foráneas: se deben añadir cada una de las columnas que representan la clave principal de las tablas de dimensiones en cuestión (en este caso "id\_dim\_Rio" y "id\_dim\_Prov").

Otros campos de información adicional.

Incluir el campo clave principal creado en el paso anterior (en este caso "id\_dim\_Relacion") en la tabla de hechos.

Otra posible solución sería agregar las dos claves primarias de las tablas de dimensiones DIM\_RIOS y DIM\_PROV en la tabla de hechos.

Existen otras soluciones para solventar esta brecha, pero la primera propuesta posee mucha performance, ya que:

Elimina la relación muchos a muchos.

Solo se necesita un campo clave en la tabla de Hechos.

La relación entre las tablas resultantes es simple y fácil de visualizar.

La única desventaja es en cuanto a los procesos ETL, ya que se aumenta su complejidad y tiempo de proceso.

#### 2.1.1.5.4 Claves Subrogadas.

Las claves existentes en los OLTP se denominan claves naturales; en cambio, las claves subrogadas son aquellas que se definen artificialmente, son de tipo numérico secuencial, no tienen relación directa con ningún dato y no poseen ningún significado en especial.

Lo anterior, es solo una de las razones por las cuales utilizar claves subrogadas en el DW, pero se pueden definir una serie de ventajas más:

- Ocupan menos espacio y son más performantes que las tradicionales claves naturales, y más aún si estas últimas son de tipo texto.
- Son de tipo numérico entero (auto numérico o secuencial).
- Permiten que la construcción y mantenimiento de índices sea una tarea sencilla.
- El DW no dependerá de la codificación interna del OLTP.
- Si se modifica el valor de una clave en el OLTP, el DW lo tomará como un nuevo elemento, permitiendo de esta manera, almacenar diferentes versiones del mismo dato.
- Permiten la correcta aplicación de técnicas SCD (Dimensiones lentamente cambiantes)

Esta clave subrogada debe ser el único campo que sea clave principal de cada tabla de dimensión.

Una forma de implementación sería, a través de la utilización de herramientas ETL, mantener una tabla que contenga la clave primaria de la tabla del OLTP y la clave subrogada correspondiente a la dimensión del DW.

En la tabla de dimensión Tiempo, es conveniente hacer una excepción y mantener un formato tal como "yyyy-mm-dd", ya que esto provee dos grandes beneficios:

- Se simplifican los procesos ETL.
- Brinda la posibilidad de realizar particiones de la tabla de hechos a través de ese campo
- Dimensiones lentamente cambiantes.

Las dimensiones lentamente cambiantes o SCD (Slowly Changing Dimensions) son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, o implique a un solo registro o la tabla completa.

Cuando ocurren estos cambios, se puede optar por seguir alguna de estas dos grandes opciones:

- Registrar el historial de cambios.
- Reemplazar los valores que sean necesarios.

Inicialmente se plantean tres estrategias a seguir cuando se tratan las SCD: tipo 1, tipo 2 y tipo 3; pero a través de los años la comunidad de personas que se encargaba de modelar bases de datos profundizó las definiciones iniciales e incluyó varios tipos SCD más, por ejemplo: tipo 4 y tipo 6.

A continuación, se detallará cada tipo de estrategia SCD:

SCD Tipo 1: Sobrescribir.

SCD Tipo 2: Añadir fila.

SCD Tipo 3: Añadir columna.

SCD Tipo 4: Tabla de Historia separada.

SCD Tipo 6: Híbrido.

Cabe destacar que existe un SCD Tipo 0, que representa el no tener en cuenta los cambios que pudieran llegar a suceder en los datos de las dimensiones y por consiguiente no tomar medidas al caso.

De acuerdo a la naturaleza del cambio se debe seleccionar qué Tipo SCD se utilizará, en algunos casos resultará conveniente combinar varias técnicas.

Es importante señalar que, si bien hay diferentes maneras de implementar cada técnica, es indispensable contar con claves subrogadas en las tablas de dimensiones para aplicar poder aplicar dichas técnicas.

Al aplicar las diferentes técnicas SCD, en muchos casos se deberá modificar la estructura de la tabla de dimensión con la que se esté trabajando, por lo cual estas modificaciones son recomendables hacerlas al momento de modelar la tabla; aunque también puede hacerse una vez que ya se ha modelado y contiene datos, para lo cual al añadir por ejemplo una nueva columna se deberá especificar los valores por defecto que adoptarán los registros de la tabla.

SCD Tipo 1: Sobrescribir.

Este tipo es el más básico y sencillo de implementar, ya que, si bien no guarda los cambios históricos, tampoco requiere ningún modelado especial y no necesita que se añadan nuevos registros a la tabla.

En este caso cuando un registro presente un cambio en alguno de los valores de sus campos, se debe proceder simplemente a actualizar el dato en cuestión, sobrescribiendo el antiguo.

Usualmente este tipo es utilizado en casos en donde la información histórica no sea importante de mantener, tal como sucede cuando se debe modificar el valor de un registro porque tiene errores de ortografía.

SCD Tipo 2: Añadir fila.

Esta estrategia requiere que se agreguen algunas columnas adicionales a la tabla de dimensión, para que almacenen el historial de cambios.

Las columnas que suelen agregarse son:

- Fecha Inicio: fecha desde que entró en vigencia el registro actual. Por defecto suele utilizarse una fecha muy antigua, ejemplo: "01/01/1000".
- Fecha Fin: fecha en la cual el registro actual dejó de estar en vigencia. Por defecto suele utilizarse una fecha muy futurista, ejemplo: "01/01/9999".
- Versión: número secuencial que se incrementa cada nuevo cambio. Por defecto suele comenzar en "1".
- Versión actual: especifica si el campo actual es el vigente. Este valor puede ser en caso de ser verdadero: "true" o "1"; y en caso de ser falso: "false" o "0".

Entonces, cuando ocurra algún cambio en los valores de los registros, se añadirá una nueva fila y se deberá completar los datos referidos al historial de cambios.

SCD Tipo 3: Añadir columna.

Esta estrategia requiere que se agregue a la tabla de dimensión una columna adicional por cada columna cuyos valores se desea mantener un historial de cambios.

SCD Tipo 4: Tabla de Historia separada.

Esta técnica se utiliza en combinación con alguna otra y su función básica es almacenar en una tabla adicional los detalles de cambios históricos realizados en una tabla de dimensión.

Esta tabla histórica indicará por ejemplo que tipo de operación se ha realizado (Insert, Update, Delete), sobre qué campo y en qué fecha.

El objetivo de mantener esta tabla es el de contar con un detalle de todos los cambios, para luego analizarlos y poder tomar decisiones acerca de cuál técnica SCD podría aplicarse mejor.

SCD Tipo 6: Híbrido.

Esta técnica combina las SCD Tipo 1, 2 y 3.

Se denomina SCD Tipo "6", simplemente porque:  $6 = 1 + 2 + 3$ .

Dimensiones Degeneradas.

El término Dimensión Degenerada, hace referencia a un campo que será utilizado como criterio de análisis y que es almacenado en la tabla de hechos.

Esto sucede cuando un campo que se utilizará como criterio de análisis posee el mismo nivel de granularidad que los datos de la tabla de hechos, y que por lo tanto no se pueden realizar agrupaciones o totalizaciones a través de este campo. Los "números de orden", "números de ticket", "números de transacción", etc., son algunos ejemplos de dimensiones degeneradas

La inclusión de estos campos en las tablas de hechos, se lleva a cabo para reducir la duplicación y simplificar las consultas.

Se podría plantear la opción de simplemente incluir estos campos en una tabla de dimensión, pero en este caso estaríamos manteniendo una fila de esta dimensión por cada fila en la tabla de hechos, por consiguiente, obtendríamos la duplicación de información y complejidad, que precisamente se pretende evitar.

Dimensiones Clustering.

Las dimensiones Clustering, son aquellas que están relacionadas a dos o más dimensiones y que brindan información diferente a cada una de ellas.

Para solucionar esta situación pueden llevarse a cabo diferentes estrategias, cada una de las cuales trae aparejadas sus ventajas y desventajas, por lo cual dependiendo cual sea el contexto se elegirá entre una y otra.

A continuación, se destaca una solución a esta situación:

- Se pueden incluir todos los campos de la dimensión Clustering en cada tabla de dimensión con que se relacione y eliminar luego la dimensión Clustering.

### 2.3. Metas a alcanzar

Desarrollar un tablero de mando que le permita a los usuarios de la gerencia de operaciones de Almacenes corona S.A.S tener la información de ventas actualizada y disponible para que a los usuarios puedan tomar decisiones de una manera ágil y oportuna.

El despliegue de informe final se realizará por medio de una herramienta de fácil manejo y conocida por los usuarios finales.

### 2.4. Productos a entregar

El producto final que se entregara al área de Gerencia de Operaciones es un tablero de mando que permita a los usuarios tener información actualizada y veraz de la venta de cada uno de los almacenes y de segregar la información desde lo más general hasta lo más específico.

### 2.5. Glosario

**Clave primaria:** campo o a una combinación de campos que identifica de forma única a cada fila de una tabla.

**Clave foránea:** la clave foránea identifica una columna o grupo de columnas en una tabla.

**Clave subrogada:** suelen utilizarse especialmente en tablas de dimensión versionadas o históricas, conocidas.

**Cubo OLAP:** es una base de datos multidimensional, en la cual el almacenamiento físico de los datos se realiza en un vector multidimensional.

**Dato:** es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc) de un atributo o variable.

**Dependencia:** es una conexión entre uno o más atributos.

**Dimensión:** es una colección de objetos relacionados, denominados atributos, que se pueden usar para proporcionar información sobre los datos de hechos de uno o varios cubos.

**FeedBack:** retroalimentación.

**Granulidad:** es una característica importante que define el nivel en una tabla de hechos.

**Tabla de Hecho:** es la tabla central de un esquema dimensional.

**Registro:** fila, renglón o tupla.

**Relación:** tabla o archivo.

**Reporte:** es un informe que organiza y exhibe la información contenida en una base de datos.

**Repositorio:** es el sitio centralizado donde se almacena y mantiene la información de la base de datos.

**Totalizacion:** muestra los datos de una manera más resumida, permitiendo, precisamente, calcular valores agregados.

### 3. DISEÑO E IMPLEMENTACIÓN

Desarrollo de un Data Warehouse basado en la metodología de Hefesto:

Hefesto es una metodología propia cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos. Cabe destacar que Hefesto se encuentra en continua evolución y se ha tenido en cuenta como gran valor agregado, todos los feedbacks que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines. (Bernau, 2010)

La idea principal es comprender cada paso que se realizara para no tener que caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente que se está haciendo y por qué. (Bernau, 2010)

La construcción e implementación de un DW puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán diferentes, lo que se debe tener muy en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue demasiado largas, lo que busca es entregar una primera implementación que satisfaga una parte de las necesidades para demostrar las ventajas del DW. (Bernau, 2010)

Esta metodología se puede describir en el siguiente gráfico:



Ilustración 20. Metodología Hefesto (Bernau, 2010)

Como se aprecia en la Figura 20 la metodología Hefesto se compone de 4 fases o etapas las cuales se explicarán a continuación.

### 3.1. Planeación y administración del proyecto

#### 3.1.1 Definición del proyecto

En la definición del proyecto se identifica el escenario para poder determinar el alcance del proyecto, en el sentido que se pueda apreciar de donde se origina la demanda dentro de la empresa.

Los posibles escenarios originados dentro de la compañía Almacenes Corona son:

Demanda de un sector del negocio: Ejemplo las Gerencias de la empresa.

Demasiada cantidad de información: Se caracteriza porque la información proviene de muchas áreas del negocio.

En busca de la demanda: Es un escenario en el cual se encuentran involucrados los altos gerentes y el presidente de la organización.

### 3.2 Análisis de requerimientos

Lo primero que se hará será identificar los requerimientos de los usuarios a través de preguntas que expliquen los objetivos de la organización, luego se analizan estas preguntas a fin de especificar cuáles serán los indicadores y perspectivas que serán tomadas para la construcción del DW.

Es muy importante tener en cuenta que la metodología Hefesto se puede aplicar bien sea para construir un Data Warehouse o un Data Mart a la vez, es decir si se desea construir un Data Mart se deberá aplicar la metodología dos veces una por cada Data Mart del mismo modo, un ejemplo sería si se desea analizar dos áreas de la empresa, Ventas y Compras.

#### 3.2.1 Identificar preguntas.

El objetivo principal de esta fase es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y las estrategias de la empresa y que facilitara una eficaz y eficiente toma de decisiones.

Una forma de asegurarse de que se ha realizado un buen análisis, es corroborar que el resultado del mismo haga explícitos los objetivos estratégicos planteados por la compañía.

La idea central es que se formulen preguntas complejas sobre el negocio que incluyan variables de análisis que se consideren relevantes, ya que estas son las que permiten estudiar la información desde distintas perspectivas.

Un punto importante que se debe tener muy en cuenta es que la información debe estar soportada de alguna manera por un OLTP, ya que de otra forma no se podrá elaborar el DW.

#### 3.2.2 Identificar indicadores y perspectivas

Una vez se ha establecido las preguntas del negocio se procede a su descomposición para descubrir los indicadores que se utilizar y las perspectivas de análisis que intervendrán.

Para que sean realmente efectivos se debe tener en cuenta que los indicadores son en general valores numéricos y representan lo que se desea analizar concretamente, para nuestro trabajo se analizarán: saldos de clientes, promedios de venta, cantidades vendidas, fórmulas de indicadores.

Por otro lado, las perspectivas se refieren a los objetos mediante los cuales se requiere examinar los indicadores con el fin de responder las preguntas planteadas, Ejemplo de ello sería: información de los clientes, puntos de venta a nivel nacional, productos, proveedores.

### 3.2.3 Modelo Conceptual

En esta etapa se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior.

A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto para luego poder trabajar sobre él, además al poseer un alto nivel de definición de los datos es posible que sea presentado a los usuarios explicando con facilidad.

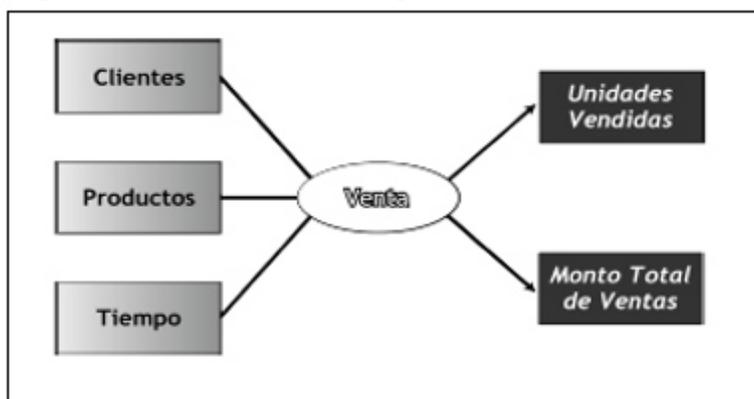


Ilustración 21. Modelo Conceptual (Dataprix, s.f.)

En la figura 21, se puede apreciar el modelo conceptual del negocio.

A la izquierda se pueden apreciar las perspectivas seleccionadas que serán unidas por un ovalo centra que representa el nombre de la relación, esta relación constituye el proceso o área de estudio elegida. De dicha relación y entrelazadas con flechas se desprenden los indicadores, estos se ubican a la derecha del esquema.

Como se puede observar en el modelo conceptual cuáles serán las variables que se utilizaran para analizarlos y cuál es su relación existente.

### 3.3 Análisis de los OLTP

Luego de estos pasos entraremos a analizar las fuentes OLTP que nos ayudaran a determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las diferentes fuentes de los datos, luego se definirán que campos se incluirán en cada perspectiva y finalmente se ampliara el modelo con la información obtenida.

#### 3.3.1 Conformar Indicadores.

En este paso se debe explicar cómo se van a calcular los indicadores definiendo los siguientes conceptos:

Hecho: estos se componen con su respectiva formula, por ejemplo,  $Venta\ neta = Venta - Costo$

Funciones de suma que se utilizaran para su respectiva agregación, por ejemplo, Sum, Count, Avg,

Para el caso de Almacenes corona un ejemplo es:

Venta Neta

Hechos.  $Venta - Costo$

Función = Sum

#### 3.3.2 Establecer Correspondencias

El objetivo de este paso es el de examinar los OLTP que contengan información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y los orígenes de datos.

La idea es que todos los elementos del modelo conceptual estén correspondidos en los OLTP. Para nuestro caso el proceso de venta está representado de la siguiente manera.

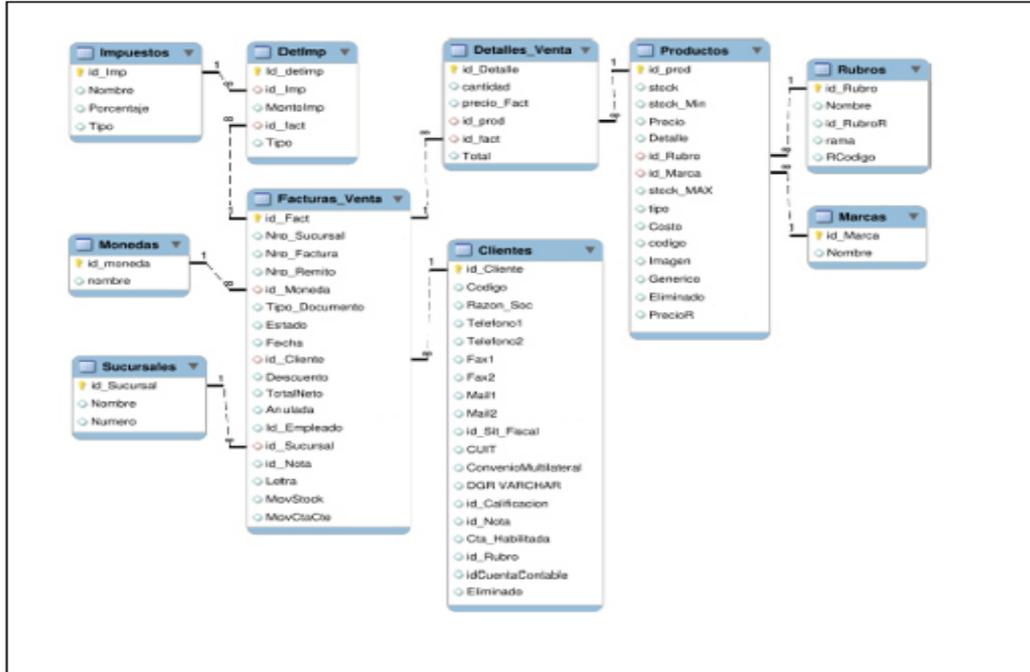


Ilustración 22. Modelo Entidad – Relación Almacenes corona SAS

En la figura 22, se puede apreciar el modelo entidad relación de la base de datos de Almacenes Corona S.A.S.

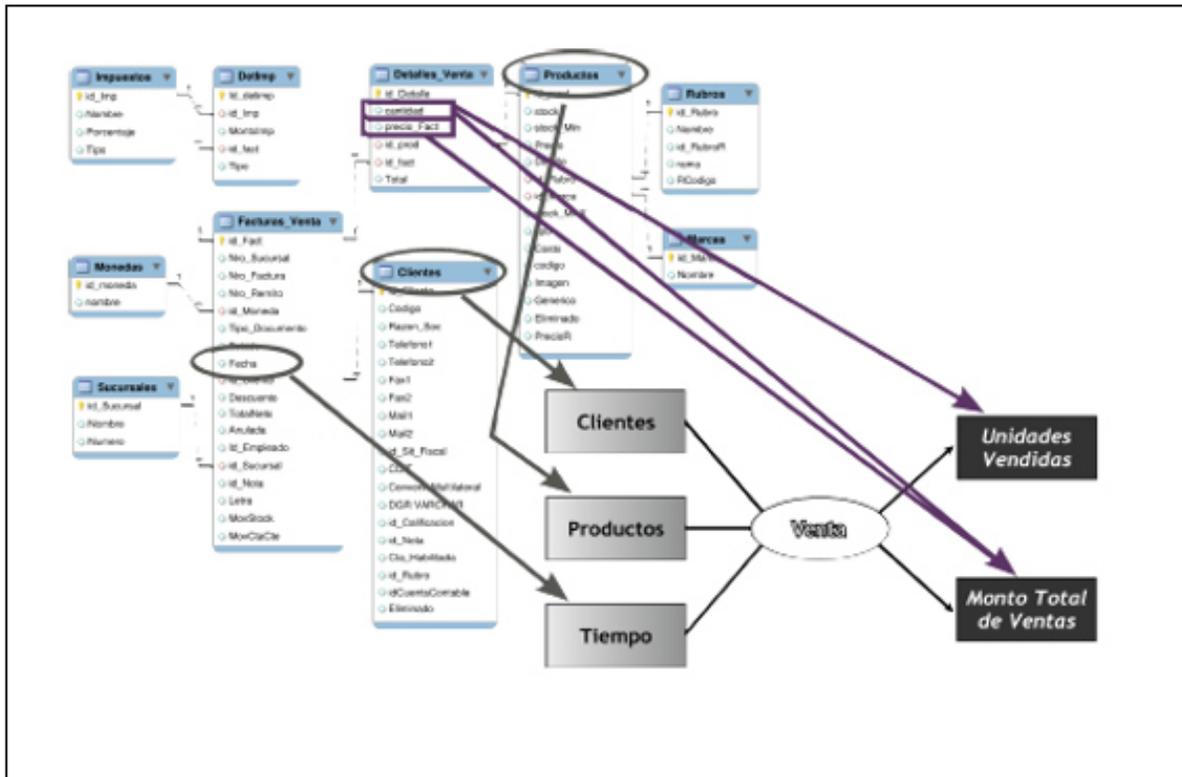


Ilustración 23. Modelo de correspondencia.

Como se puede apreciar en la figura 23, si juntamos los dos modelos se pueden identificar las siguientes relaciones:

- La tabla de producto se relaciona con la perspectiva de producto.
- La tabla clientes se relaciona con la perspectiva de clientes.
- El campo cantidad de la tabla de detalles de venta con el indicador de unidades vendidas.

### **3.3.3 Nivel de Granularidad.**

Una vez se han establecido las relaciones con los OLTP, se deben seleccionar los campos que contienen cada perspectiva, ya que será a través de estos por los que se examinarán y filtrarán los indicadores.

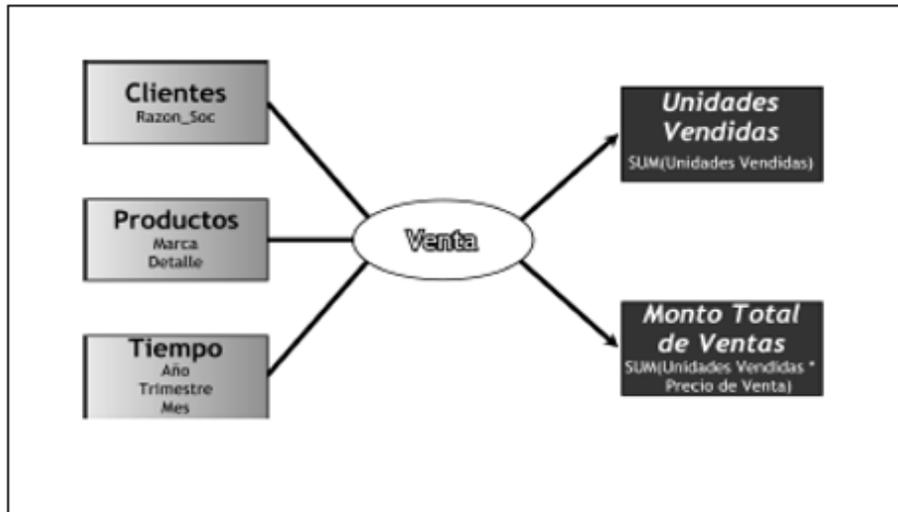
Para ello basándose en las correspondencias establecidas en el paso anterior, se deben presentar a los usuarios los datos de análisis disponibles para cada perspectiva. Es muy importante conocer en detalle que significa cada campo y valor de los datos encontrados en los OLTP, por lo cual es conveniente investigar su sentido a través de diccionarios de datos o reuniones con los encargados del sistema.

Luego de exponer frente a los usuarios los datos existentes, explicando su significado, valores posibles y características, estos deben decidir cuáles se consideran relevantes para consultar los indicadores.

Con respecto a la perspectiva del tiempo es muy importante definir el ámbito mediante el cual se agruparán o sumarán los datos. Los campos posibles pueden ser: Día de la semana, cada 15 días, mensualmente etc.

### **3.3.4 Modelo conceptual ampliado.**

En este paso con el fin de graficar los resultados obtenidos en los pasos anteriores se ampliará el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y debajo de cada indicador su respectiva fórmula de cálculo.



**Ilustración 24. Modelo conceptual ampliado**

### 3.4 Modelo lógico del DW

A continuación, se explicará el modelo lógico de la estructura del DW teniendo como base el modelo conceptual que ya ha sido creado, para ello primero se definirá el tipo de modelo que se utilizara y luego se llevaran a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hecho. Finalmente se realizarán las uniones pertinentes.

#### 3.4.1 Tipo de modelo lógico del DW.

Se debe seleccionar cual será el tipo de esquema que se utilizara para presentar la estructura del depósito de datos que se adapte mejor a los requerimientos de los usuarios, es muy importante definir si se empleara un esquema de estrella o copo de nieve ya que esta decisión afectara considerablemente el modelo lógico.

#### 3.4.2 Tablas dimensionales.

En este paso se diseñarán las tablas de dimensiones que harán parte del DW.

Para los dos tipos de esquemas, cada perspectiva definida en un modelo conceptual constituirá una tabla de dimensión. Para ello deberá tomarse cada perspectiva con sus campos relacionados y realizar el siguiente proceso.

- Se le dará un nombre que identifique la tabla de dimensión.
- Se creará un campo de clave primaria.
- Se definen los nombres de los campos.

Para nuestro caso definimos la dimensión de tiempo.

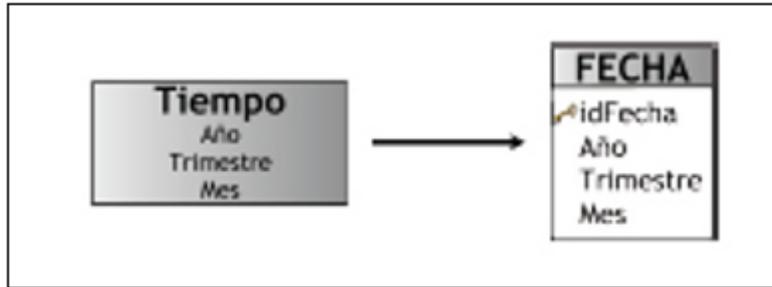


Ilustración 25. Tabla dimensión de tiempo

En la figura 25, se aprecia la dimensión del tiempo que se aplicara al modelo conceptual

### 3.4.3 Tabla de Hechos

En este paso se definirán las tablas de hechos, que son las que contienen los hechos que construirán los indicadores de estudio.

Para los dos tipos de esquemas se realiza lo siguiente:

- Se le dará un nombre a la tabla de hecho que represente la información analizada.
- Se definirá su clave primaria que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
- Se crearán tanto los campos de hechos como de indicadores que se hayan definido en el modelo conceptual y se les asignarán los mismos nombres que estos.

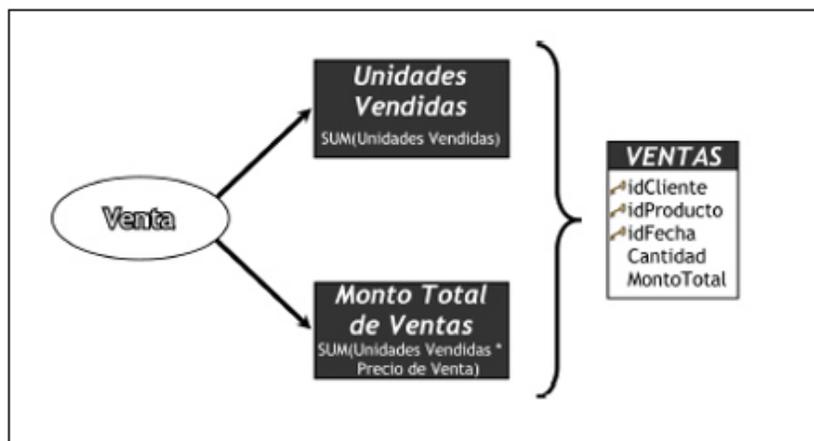


Ilustración 26. Tabla de Hechos

En la figura 26, se muestra la tabla de hechos con los indicadores de estudio.

### 3.4.3 Uniones.

Para los dos esquemas se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

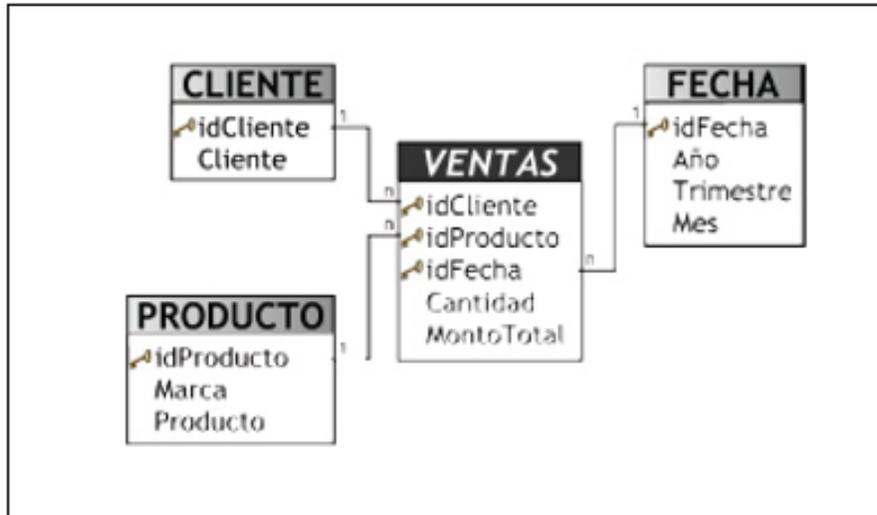


Ilustración 27. Tabla de Uniones

En la figura 27, se puede observar las uniones correspondientes.

## 3.5 Integración de los datos

Una vez construido el modelo lógico, se deberá proceder a llenarlo con datos, utilizando técnicas de limpieza y calidad de datos ETL, luego se definirán las reglas y políticas para su respectiva actualización, así como también procesos que la llevarán a cabo.

### 3.5.1 Carga inicial

En este paso debemos realizar la carga inicial de los datos de la carga inicial al DW poblando el modelo que hemos construido anteriormente. Para esta tarea debemos llevar a cabo técnicas de limpieza de datos como en nuestro caso ETL.

Las realizaciones de estas tareas pueden contener una lógica de datos compleja en algunos casos, para facilidad de esto existe software que permite hacer este proceso con calidad, para nuestro desarrollo utilizaremos una herramienta llamada Talen Studio versión 5.2

Primero se cargarán los datos a las dimensiones ya creadas y luego a las tablas de hechos teniendo en cuenta siempre la correcta correspondencia entre cada elemento, en el caso de aplicar un modelo de copo de nieve, cada vez que existan jerarquías de dimensiones se comenzarán cargando tablas de dimensiones del nivel general al más detallado.

Concretamente en este paso se deberá registrar el detalle de las acciones con el diferente software, para nuestro caso es común que los sistemas ETL trabajen con pasos y relaciones

en donde cada paso realiza una tarea en particular del proceso de ETL y cada relación indica hacia donde debe dirigirse el flujo de datos. Es importante tener presente que al cargar los datos en las tablas de hechos pueden utilizarse pre agregaciones ya sea al nivel de granularidad de la misma o a otros niveles diferentes.

### 3.5.2 Actualización

Cuando se haya cargado totalmente el DW, se deben establecer políticas y estrategias de actualización y refresco de los datos.

Una vez llevado a cabo esto se tienen que llevar a cabo las siguientes acciones:

Especificar tareas de limpieza de datos, calidad de datos, ETL que se deben aplicar al DW. Especificar en forma general y detallada las acciones que deberá realizar el Software, en nuestro caso Talend Open Studio V. 5.2.

Creación de cubos multidimensionales

A continuación, se creará un cubo multidimensional de ejemplo que será llamado cubo de ventas que estará basado en el modelo lógico diseñado anteriormente.

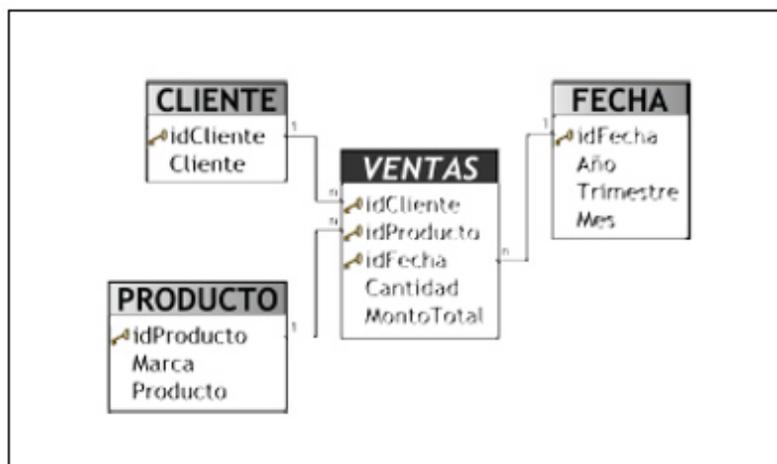


Ilustración 28. Modelo Lógico del Cubo

La creación de este cubo tiene como objetivo:

- Creación de cubos multidimensionales.
- Propiciar la correcta distinción entre hechos de una tabla de hechos e indicadores de un cubo.
- Propiciar la correcta distinción entre campos de una tabla de dimensión y atributos del cubo.

#### 3.5.2.1 Creación de indicadores.

En esta fase se crearán dos indicadores que serán incluidos en el cubo de Ventas

De la tabla hechos de Ventas se sumarán el hecho de cantidad para crear el indicador nombrado como Unidades Vendidas que utilizara la siguiente formula:

$$\text{Unidades Vendidas} = \text{Sum} (\text{Ventas} * \text{Cantidad})$$

De la tabla hechos de ventas se sumará el hecho Monto Total para crear el indicador denominado Monto Total de ventas que se calculará de la siguiente formula:

$$\text{Monto Total de Ventas} = \text{Sum} (\text{Ventas} * \text{Monto Total}).$$

### **3.5.2.2 Creación de los atributos.**

Ahora se crearán y agregarán al cubo cinco atributos:

- De la tabla de dimensión cliente se tomará el campo cliente para la creación del atributo nominado clientes.
- De la tabla de dimensión de producto se tomará el campo Importado para la creación del atributo Importado.
- De la tabla de dimensión producto se tomará el campo Código de Producto para indicar la Referencia del producto.
- De la tabla de la dimensión Tiempo se tomará el campo año para la creación del atributo denominado Año.
- De la tabla de dimensión de tiempo se tomará el campo mes para la creación del atributo llamado Mes.

### **3.5.2.3 Creación de las Jerarquías.**

Finalmente se crearán y agregarán al cubo 2 jerarquías.

Se definió la jerarquía de Productos que se aplicara sobre los atributos que se tiene anteriormente, Producto, en donde un producto (Referencia) puede pertenecer a un solo negocio de la compañía ya bien sea (Baños, Cocinas, Materiales y acabados y Revestimientos).

Se definió la jerarquía tiempo que se aplicará sobre los atributos Año y Mes, en la cual un mes del año pertenece a un solo trimestre y un año tiene uno o más meses.

Otros ejemplos de cubos multidimensionales:

A partir del modelo lógico planteado podrían haberse creado una gran cantidad de cubos, cada uno de ellos estaría orientado a un tipo de análisis en específico, como se explicó antes estos cubos pueden coexistir sin ningún problema.

Los cubos se pueden trabajar con Microsoft Excel, mediante tablas dinámicas.

Para empezar a trabajar con cubos simplemente se abre un archivo de Excel con un cubo ya pre construido y empezar a conformar la información que se quiere trabajar.

A continuación, se ejemplificar algunos cubos que se tiene para una necesidad en particular.

### Cubo de Ventas – PPTO

Contiene la información de ventas y presupuestos cargados en people mes a mes, el nivel de detalle de esta información está por Categoría – Punto de venta – día; no se puede ver por asesor ni por cliente

### Cubo de Ventas – PPTO Oficial – Origen

Maneja el mismo nivel de detalle del Cubo Ventas – PPTO pero adicionalmente esta por origen, la información de presupuesto es la oficializada a la corporación.

Cubo de Ventas:

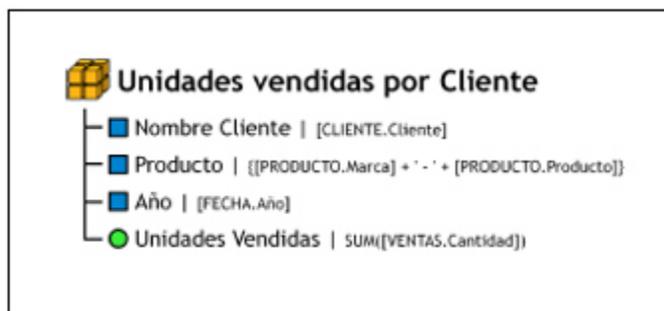


Ilustración 29. Cubo de Ventas por cliente

Como se observa en la figura 29, el cubo de ventas contiene la información de ventas a nivel de detalle producto – punto de venta – Asesor - Cliente – día.

## 3.6 ingeniería de requerimientos

### 3.6.1 Definición de los requerimientos del negocio

Para la definición de los requerimientos se realizaron 4 actividades que ayudaron a determinar las necesidades de los usuarios del negocio, estas son:

- Realizar con el jefe de tecnología la petición de servidores que tuvieran SQL SERVER instalado para la creación del DW.
- Reunión con Gerente de operaciones.
- Realizar un estudio de los reportes que actualmente se están haciendo de manera manual, sistemas relacionados con procesos del negocio. En este punto vale la pena recalcar la importancia de los reportes estadísticos ya que es a través de ellos es que se puede recopilar valiosa información del estado actual de la compañía.
- Conocer de los procesos que se manejan al interior de la compañía y de los negocios para entender su contexto, funcionalidad e importancia en el desarrollo.

Teniendo en cuenta las actividades anteriores se obtuvieron los siguientes requerimientos por parte de las áreas.

#### 3.6.1.1 Proceso de inventarios

La gerencia Logística nos dio a conocer la información que le interesa conocer en los informes y sobre la cual tomaran decisiones.

- Ingresos de nuevas referencias al portafolio de productos con sus respectivos costos.
- Movimiento de Inventarios entre bodegas y puntos de venta.
- Inventario en el Kardex.
- Cargos de fletes.
- Pedidos de referencias de acuerdo al portafolio de productos.
- Referencias de poca rotación.
- Cantidad de referencias compradas a los distintos proveedores.
- Cantidad de productos defectuosos, rotos o extraviados.

##### 3.6.1.1.1 Indicadores

Nuevos ingresos de mercancía a los puntos de venta y las bodegas centrales mensualmente y su respectivo costo.

Distribución del 100 % del inventario por punto de venta a nivel nacional y su respectivo costo.

Mencionado estos indicadores por la gerencia logística manifestaron las posibles consultas para la presentación de los informes que se especifican a continuación.

Ingresos por proveedores y por bodegas: este indicador permite ver el valor total de referencias ingresadas a los puntos de venta por proveedores.

Egresos por Bodega: este indicador permite ver el valor total de egresos que se producen por cada una de las bodegas.

Existencia: Permite ver y evaluar cuantas referencias existentes hay en el inventario por punto de venta y por cada una de las bodegas.

Movimiento de Mercancía: Permite ver cuáles son las referencias que se están en movimiento por punto de venta y por bodegas.

Compras por proveedor: Permite ver el valor total de la compra a nivel de proveedor.

Control de pedidos: Permite ver las cantidades de cada una de las referencias a nivel de pedido.

### 3.6.1.2 Proceso de punto de venta.

Para la compañía de Almacenes Corona el negocio por punto de venta es un factor importante del negocio y es la fuente principal de ingresos; debido a que no existe una herramienta existente que permita estar monitoreando este proceso se realizó una reunión con el Gerente de operaciones, jefes de zona y jefes de punto de venta para reunir los requerimientos de la información que quisieran conocer a diario, las cuales se exponen a continuación:

- Total, de ventas por punto de venta y total compañía.
- Top 20 de las referencias más vendidas por categoría.
- Análisis de la información de ventas por punto de venta con un histórico de mínimo un año para poder ver el comportamiento del punto.
- Rotación de Inventario.
- Valor de la venta de cada una de las referencias.

### 3.6.1.2 Indicadores

Ventas totales por punto de venta en un determinado tiempo, este sería un indicador que represente el cumplimiento del punto de venta con respecto al presupuesto pactado por mes. Top 20 por categoría, este indicador permite ver cuáles son las referencias más vendidas por categoría, lo que tiene como objetivo evaluar cuáles de esos productos son de alta rotación y cuales son de inventario de lento movimiento.

Indicadores claves de rendimiento (KPI)

Nivel de Cumplimiento. Este indicador representa el valor de evaluación de los puntos de venta en función al cumplimiento propuesto mensualmente.

Indicadores de cumplimiento

% Cumplimiento	Nivel de Cumplimiento (KPI)
>= 90 %	5
>= 80 % y < 90 %	4
>= 70 % y < 80 %	3
>= 60 % y < 70 %	2
< 60 %	1

Ilustración 30. Indicadores de Cumplimiento

En la Figura 30, se menciona la información anterior que es de vital importancia conocer a los jefes de zona y a los jefes de punto de venta. Se tomarán las consultas para la visualización que se especificaran a continuación:

- Ventas y cantidad vendidas por referencia.
- Cliente
- Ambiente
- Negocio
- Punto de venta
- Factura
- Cotización
- Asesor

Esta descripción permite visualizar los productos más vendidos, el valor de la venta y el número de cotizaciones que se convirtieron en facturas por cliente.

Número de pedidos cerrados por asesor y por punto de venta:

esta descripción permite ver el nivel de productividad por asesor de acuerdo al presupuesto establecido.

Estos requerimientos a nivel de mercancía y de puntos de venta se especificarán a nivel de tiempo en los siguientes niveles de medición.

- Anual
- Semestral
- Trimestral
- Mensual
- Diario

Adicional a esto por parte de las gerencias de operaciones y logísticas se especificó que tuvieran datos actualizados y también históricos empezando desde el año 1999 al actual para el tema de productos matriculados en el portafolio y el valor total de las ventas.

No se definieron indicadores claves para el rendimiento del tema de inventarios ya que actualmente no cuentan con sistemas automatizados que permitan ver estos indicadores.

Para el tema de las ventas se establecieron indicadores que permitan ver el rendimiento por punto de venta.

### **3.6.2 Diagramas de modelamiento unificado**

#### **3.6.2.1 Diagrama General de caso de uso**

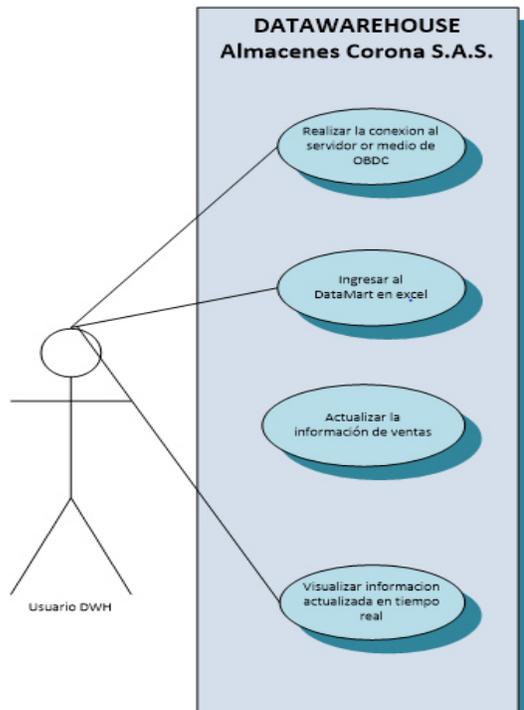


Ilustración 31. Diagrama General de caso de uso

### 3.6.2.2 Diagrama general de Secuencia

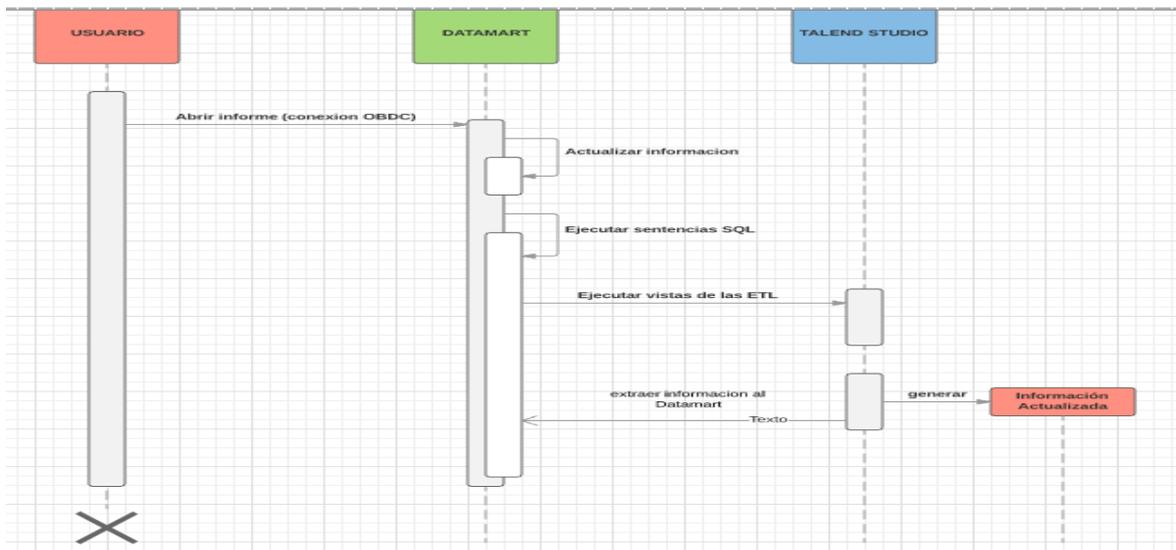


Ilustración 32. Diagrama general de secuencia

### 3.6.2.3 Diagrama general de Actividades

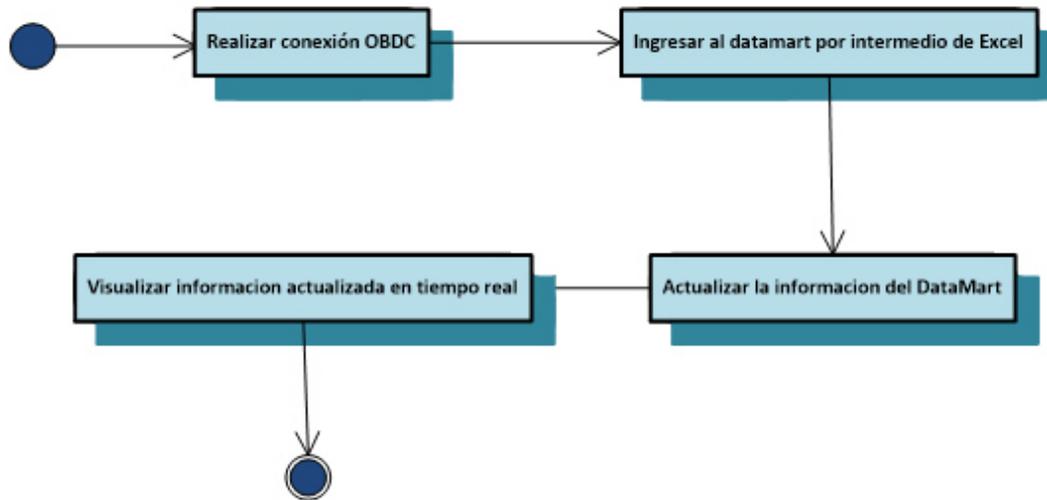


Ilustración 33. Diagrama General de Actividades

## 3.7 Diseño arquitectónico

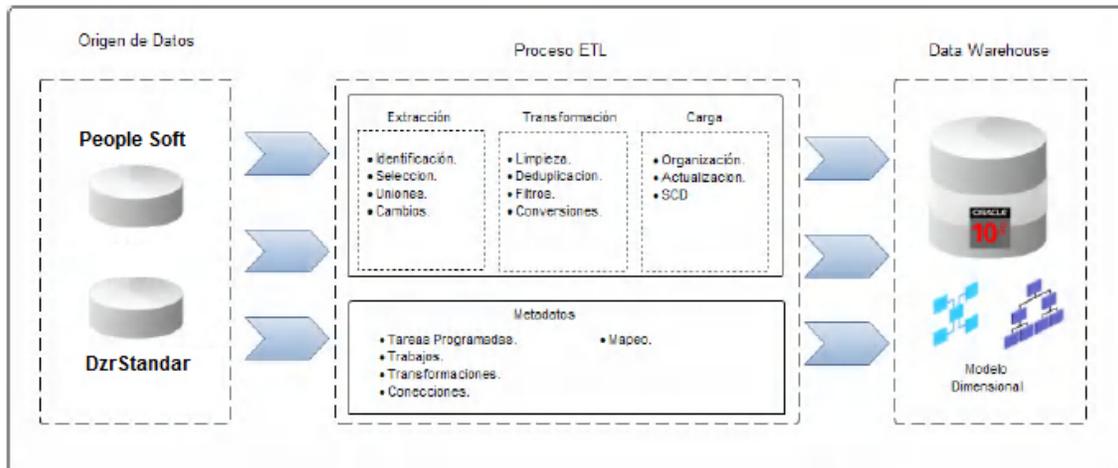
En el diseño de la arquitectura se definieron dos ambientes, el primero de ellos es el Back room y el segundo que es Front Room, los mismos que serán detallados a continuación:

### 3.7.1 Back Room.

En este ambiente se especificarán el proceso de ETL así como los orígenes de datos, como se aprecia en la siguiente figura se detalla el diseño para la realización del back Room que consta del origen de los datos hasta su almacenamiento.

#### 3.7.1.1 Origen de datos

La base de datos de origen se encuentra almacenada sobre distintas plataformas, Oracle 11 g, SQL server 2008, como se puede observar en la en la figura 31 se establecen dos orígenes de datos PeopleSoft y DzsStandar , la primera se utilizara cuando el proyecto se encuentre en producción mientras que la segunda se utilizara para traer información histórica y para las respectivas pruebas. Cabe la pena recalcar que los metadatos (esquemas, tablas, archivos planos y funciones) de ambas bases de datos son similares.



**Ilustración 34. Back Room**

### 3.7.1.2 ETL y Metadatos

En cuanto al proceso de ETL se realizó el mapeo de los datos origen, que serán especificados posteriormente en el modelamiento dimensional, después se desarrolló el proceso de extracción, seguido de la transformación de los datos origen, dependiendo de los cambios se les debe adecuar, y finalmente se procedió con la carga hacia el DW.

Los metadatos son las especificaciones de las transformaciones de la información, mapeos, tareas programadas que se almacenaron en el repositorio del DW como el soporte al proceso ETL.

Se especificó con más detalle el proceso ETL y los metadatos del proyecto en la sección de diseño y desarrollo de la ETL.

### 3.7.1.3 Destino de datos (DW)

El destino que es el almacén de datos, mantiene el modelo dimensional que será especificado en la sección de diseño y modelamiento dimensional.

La base de datos destino está separada físicamente de las bases de datos origen, por motivo de rendimiento y para poder separar el ambiente de producción con el diseño del DW.

### 3.7.2 Front Room.

En esta sección se especifica la arquitectura del DW mediante las aplicaciones dirigidas al usuario final, como se puede apreciar en la figura 32, se parte del DWH en donde se encuentra los datos que ya han sido tratados y almacenados, posteriormente se procede a la presentación de acuerdo al tipo de aplicación de usuario final que se utilice. Para el desarrollo de este proyecto se implementará una arquitectura de ambiente analítico para poder diseñar los cubos OLAP e informes desde orígenes Microsoft query y conexiones directas al SQL server para

cargar las distintas vistas. Así como también el DW sirve como soporte para las consultas directas que se necesitan en la compañía, el desarrollo de los informes y los cubos será detallado más adelante en la sección de diseño y desarrollo de la aplicación BI.

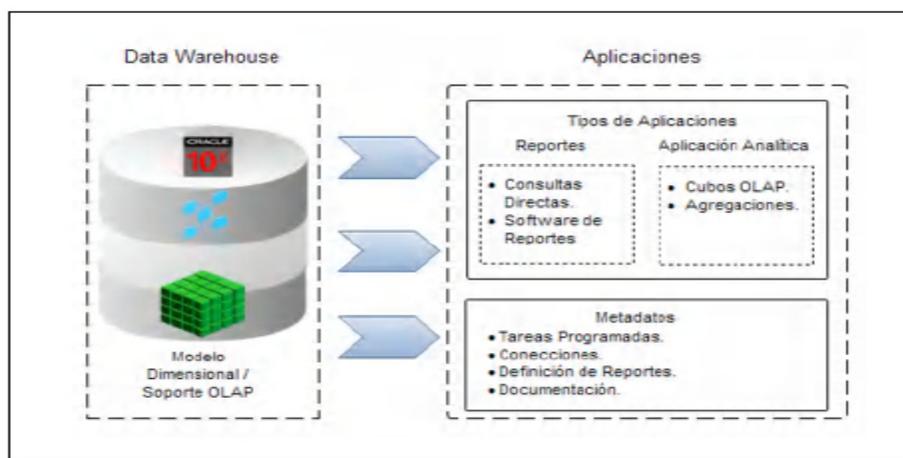


Ilustración 35. Front Room del Proyecto

### 3.8 Diseño físico

El diseño físico se realizó en función de la base de datos ya mencionada, Oracle 11g, como son los esquemas, espacios de las tablas, etc., los scripts de la creación para la base de datos origen se encuentra en el anexo 1.

Proceso de negocio	Tablas de hechos	Granularidad	Área	Producto	Bodega	PDV	Factura	Ítem	Pedido	Proveedor	Asesor	Tiempo
Inventarios	Pedidos Movimientos	Registro por transacción de movimiento	X	X	X	X		X	X	X		X
PVD	Ventas	Registro por cada transacción de venta		X	X	X	X	X	X	X	X	X

Tabla 4. Matriz DWH

#### 3.8.1 Base de datos de origen.

Primero se empezaron especificando los scripts para la creación de la base de datos origen, lo que permitirá llevar a cabo la técnica de extracción CDC (Change Data Capture) la misma

que se explica en el diseño de la ETL para su mayor entendimiento, así como la funcionalidad por la que se optó.

#### **3.8.1.1 Espacio de la tabla.**

Se crea un tablespace para almacenar los datos que han cambiado en la base de datos origen.

#### **3.8.1.2 Esquema.**

Se crea un esquema o un usuario a la base de datos de origen que permita administrar los recursos que fueron agregados en ella, así como también a través del cual permitan leer los datos para la extracción del proceso de ETL. Los privilegios o permisos asignados para la base son:

- Lectura de datos.
- Creación de tablas (Para su esquema).
- Creación de cualquier disparador (Trigger).

#### **3.8.1.3 Tablas.**

Se utiliza una tabla de Estadísticas para poder evaluar el correcto funcionamiento de las ETL en el ambiente de producción.

#### **3.8.1.4 Índices.**

Se asigna un índice único para la clave primaria de la tabla de Estadísticas.

#### **3.8.1.5 Secuencia.**

Se utiliza una secuencia incremental para el campo de la clave primaria de la tabla de estadísticas.

#### **3.8.1.6 Disparadores.**

Disparador de la tabla de Estadísticas:

Se crea un disparador para que vaya aumentando en uno la secuencia para cada registro en la tabla.

##### **3.8.1.6.1 Disparadores de la tabla de negocio**

Se crean disparadores para las diferentes tablas que serán utilizadas en el proceso ETL. Mediante los disparadores se guardarán las claves de los registros nuevos o que han cambiado en las tablas de dimensiones.

#### **3.8.2 Base de datos destino (Data Warehouse).**

Primero se empieza con las especificaciones de la base destino, es decir el Data Warehouse.

### **3.8.2.1 Espacio de Tabla (Table Space).**

Se crean 2 tablas paca que se mencionan a continuación:

- Repositorio Data Warehouse: En este espacio de tablas se alojan todos los metadatos utilizados por la herramienta ETL para todo el proceso de integración de datos.
- Data Warehouse: En este espacio se alojan los datos propios del almacén, es decir todas las dimensiones, hechos, métricas, índices, etc.

### **3.8.2.2 Tablas.**

Se crean las tablas del Data Warehouse, tanto lo que serán las dimensiones, así como también los hechos. Además, se crea una tabla para monitorear en caso de que existan errores en la ejecución de la ETL, lo que permite llevar un control de errores.

### **3.8.2.3 Índices.**

Se definen los índices para cada tabla de dimensión, Hecho y log de errores.

### **3.8.2.4 Restricciones (Constraints).**

Se crean las restricciones, lo que son las claves primarias y foráneas, tanto para las tablas de dimensiones como para las de hechos y para el log de errores.

### **3.8.2.5 Secuencia.**

Se crea una secuencia para el incremento de registros de la tabla log de error.

### **3.8.2.6 Disparadores.**

Se genera un disparador para aumentar la secuencia en uno para cada registro nuevo en la tabla de log error.

## 3.9 Diseño y desarrollo del proceso ETL

### DESARROLLO DEL PLAN ETL

A continuación, se elabora un mapa gráfico, el cual permitirá tener una mayor idea de cómo se ejecutará el proceso de ETL.

#### **3.9.1 Paso 1: Trazar un plan de alto nivel.**

Las gráficas que muestran el proceso de cada una de las ETL se pueden observar en la parte 2 del Anexo 2.

#### **3.9.2 Paso 2: Funcionalidad de la herramienta ETL.**

Se renombro este paso en función de que ya se especificó la herramienta ETL a utilizar por lo que a continuación se detalla las funcionalidades a utilizar.

Para la elaboración del proceso ETL se utilizaron dos herramientas del software ETL de Talend Open Studio V. 5.2 estas son.

- Spoon, que permite diseñar las transformaciones y trabajos de manera gráfica y amigable.
- Kitchen, permite ejecutar el proceso periódicamente scripts batch, de acuerdo al sistema operativo que se maneje.

#### **3.9.3 Pasó 3: Desarrollo de estrategias.**

- Extracción de los datos: la extracción de los datos se realiza desde múltiples bases de datos que se encuentran en el ambiente de producción, PeopleSoft 9.1 que maneja bases de Oracle 11g, Sql Server donde se encuentra la historia de los datos de la compañía, Archivos planos. Estos esquemas a utilizar ya se especificaron anteriormente en el modelamiento dimensional.
- Captura de los Datos: como parte de la extracción de los datos se procedió a implementar una técnica la cual garantice que solo extraiga los datos que hayan cambiado en la base de datos origen. La manera como se hará este proceso se muestra en el grafico siguiente.

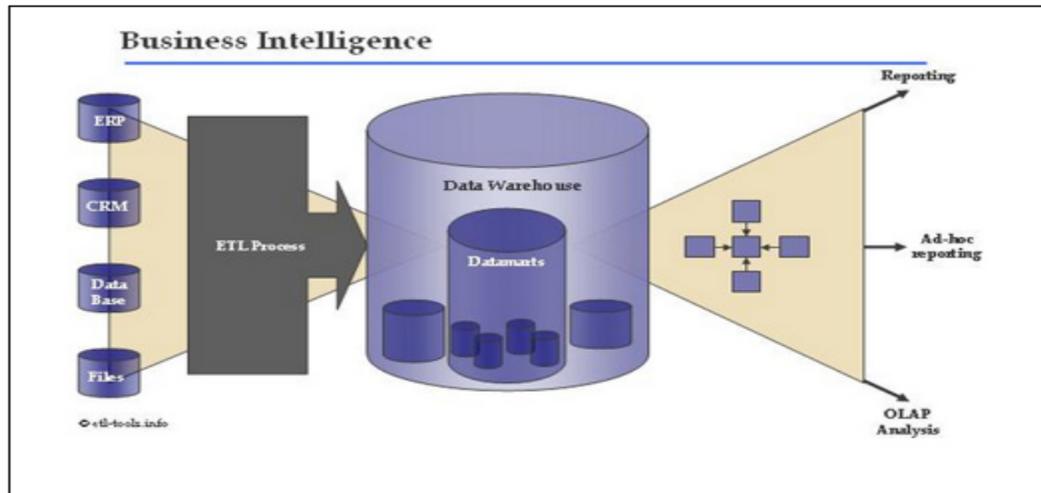


Ilustración 36. Arquitectura CDC

Cualquier inserción de datos o actualización sobre las tablas origen producirá que se lance un disparador, el cual guardará las llaves primarias en una tabla de estadísticas.

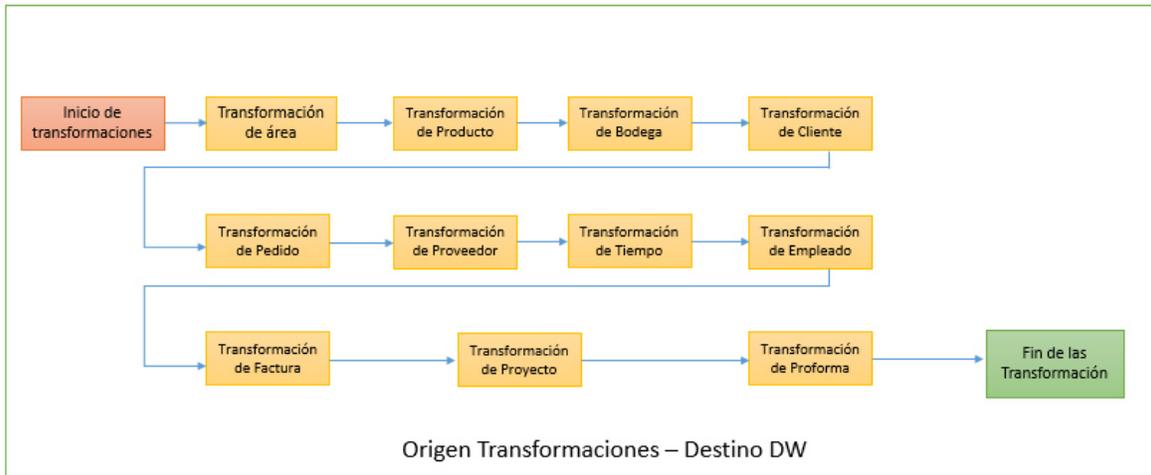
Las tablas de estadísticas están en la base de datos SQL Server recolectando todos los cambios, estos mismos son extraídos por la herramienta ETL para procesar los datos origen que han sufrido los cambios o que serán en su defecto registros nuevos.

- Administración de cambios en los atributos de las dimensiones: como se especificó en los requerimientos iniciales y el modelo dimensional, los datos que se encuentran en los almacenes (DW) serán actualizados, por lo que los atributos de las tablas destino serán sobrescritos con los datos actuales, de acuerdo a esto se especificara un SDC de tipo 1.

Para los atributos a los cuales se especificó este cambio, se le puede apreciar en la sección de modelamiento dimensional en la que se indica el SDC por atributo de acuerdo a cada tabla ya sea de dimensión o hechos.

- Carga de datos se enfoca en dos tipos:
  1. Carga Inicial: se realiza cuando se llena por primera vez la tabla del almacén destino, por ser la primera carga dura mucho más tiempo para llenar en su totalidad el almacén debido a que extraerá todos los datos de la base origen.
  2. Carga Incremental: se realiza una vez las tablas destino están llenas con los datos que han sido modificados o que sean nuevos en la base origen, garantizando una mayor eficiencia del procesamiento de los datos. La carga incremental se ejecuta una vez haya finalizado el proceso de la carga inicial.
- Frecuencia de la Carga: La frecuencia de la carga de los datos se especifica según los requerimientos iniciales, es decir, se puede ver la información, diaria, semanal, mensual, trimestral, según lo requiera el usuario.

- Secuencia de los procesos de transformaciones: la secuencia de las transformaciones depende de las tablas destino y la dependencia entre sus relaciones con las demás tablas. En función a este punto la secuencia de las transformaciones se hace de la siguiente manera.

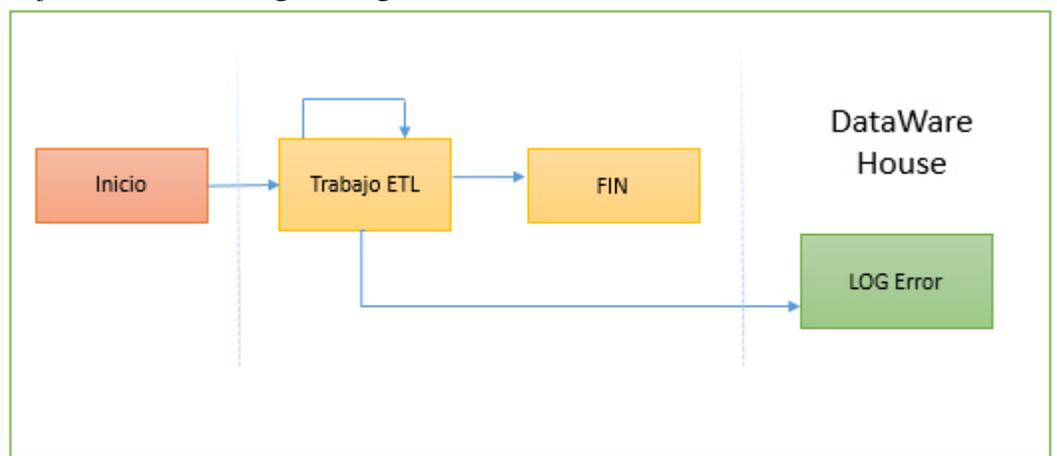


**Ilustración 37. Transformaciones ETL**

Primero se realizan todas las transformaciones para poblar las dimensiones, a continuación, se seguirá con los hechos ya que como se pudo apreciar en el modelo dimensional los hechos mantienen una integridad referencial en las tablas de dimensiones.

- Registro de Errores: como parte del proceso de la ETL se pretende realizar auditoria de errores para poder verificar y estar al tanto sobre los errores que lleguen a producirse y poder identificar el origen del error para tomar las medidas de corrección.

La manera como se pretende realizar el manejo de errores se puede interpretar de una mejor manera en el siguiente gráfico:



**Ilustración 38. Auditoria Error ETL**

#### **3.9.4 Paso 4: Profundizar en las tablas destino.**

A continuación, se especifica de manera más detallada los niveles y jerarquía por cada dimensión, estos se pueden apreciar en el Anexo 2 parte 3

#### **DESARROLLO DE LA CARGA INICIAL**

Aquí se especifica la carga de datos que se da por primera instancia en el Data Warehouse

#### **3.9.5 Paso 5: Poblar las tablas de dimensiones con los datos históricos.**

##### **3.9.5.1 Extracción de Datos:**

La extracción de los datos se hace mediante sentencias SQL en las tablas de datos origen, esto como resultado devolverá todos los registros de acuerdo a los filtros hechos de las tablas consultadas debido a que es el proceso de la carga inicial.

Las consultas desarrolladas para cargar los datos iniciales se pueden apreciar en la parte del anexo 1.

##### **3.9.5.2 Transformaciones.**

Los datos antes de ser cargados en el almacén son transformados según lo requerido, para lo cual se especificó ciertas transformaciones que se mencionan a continuación de acuerdo a cada dimensión.

###### **3.9.5.2.1 Valores Nulos.**

Los valores nulos serán reemplazados por valores que tengan un significado, y los cuales fueron especificados por el personal de la empresa. Estos valores tomados son de acuerdo a los valores origen que se extraigan y se pueden apreciar en el Anexo 2 Parte 4.1

###### **3.9.5.2.2 Agregar registro sin valor a las dimensiones.**

Se agrega un registro a cada dimensión, el cual tiene un valor que permita identificar hechos de los que no tienen especificación con la dimensión, por ejemplo: un pedido al cual no se le haya identificado un operador (Asesor) o un pedido que no tenga la identificación de la bodega. Estos parámetros fueron especificados con la ayuda de las personas involucradas en el proceso del negocio se detallarán en el Anexo 2 parte 4.2

###### **3.9.5.2.3 Decodificación de los Datos Origen.**

Como parte de la transformación esta es la etapa de la decodificación de los datos de origen que se encuentran almacenados de manera codificada, para soportar los ambientes transaccionales. En el Anexo 2 parte 4.3. se especifica las dimensiones para las cuales se decodificaron los datos origen.

### 3.9.5.2.4 Carga de Datos.

La carga inicial de los datos se realiza utilizando SCD de tipo 1 ya que para el análisis de los datos se necesitará que estos estén actualizados según los requerimientos anteriormente especificados.

### 3.9.5.2.5 Poblar las dimensiones estáticas.

La dimensión estática se considera la del tiempo, ya que en su carga inicial de los datos no se realiza ningún tipo de modificación o actualización. El contenido de esta tabla se generó en codificación de la fecha en un formato corto DD-MM-YYYY, lo que permitió obtener los nombres de los días, mes entre otros como se muestra a continuación.

<b>Formato de Fecha Seleccionado</b>	<b>Valor de Dato Generado</b>	<b>Campo Destino</b>	<b>Valor de dato Codificado de Destino</b>
DD/MM/YYYY	01/01/1999....31/12/2016	key_Tiempo	0,1,2,3,4....n
DD/MM/YYYY	01/01/1999....31/12/2016	Fecha	01/01/1999....31/12/2016
YYYY	1999 - 2016	Año	1999 - 2016
MM	1,2	Semestre	S1, S2
MM	1,2,3,4	Trimestre	Tr1, Tr2, Tr3, Tr4
MM	1,2,3,4,5,6,7,8,9,10,11,12	Mes Nombre	Enero - Diciembre
DD	1,2,3,4,5,6,7	Día_Nombre	Lunes - Domingo

**Tabla 5. Decodificación y Formato de la Dimensión de Tiempo**

La dimensión del tiempo no cuenta con un campo origen ya que no es extraído de alguna fuente de datos, por lo que se genera desde una herramienta ETL.

3.9.5.2.6 Desarrollo de las ETL iniciales de dimensiones con la herramienta Talend Open Studio V. 5.2. El desarrollo y modelo de las ETL iniciales por medio de la herramienta Talend Studio V 5.2 se detallarán en la Parte 5 del Anexo 2.

### 3.9.5.2.7 Realizar carga histórica de la tabla de hechos.

La carga de datos en las tablas de hechos al igual que es las dimensiones se realizará utilizando SCD de tipo 1 según los requerimientos de los usuarios.

### 3.9.5.2.8 Extracción de datos.

Las consultas que se desarrollaron en sql server necesarias para poblar los hechos se especifican en la sección del anexo 2. Para este proceso se tiene en cuenta tanto la extracción del hecho desde el sistema ERP (PeopleSoft 9.1) así como también las claves subrogadas de las dimensiones para asignar los hechos.

### 3.9.5.2.9 Transformación.

Los datos antes de ser cargados al DW serán transformados según lo requieran, para lo cual se especifican las transformaciones que se mencionan a continuación de acuerdo a cada hecho.

### 3.9.5.2.10 Valores Nulos.

Los valores nulos para un hecho son tratados en función y tomando en cuenta los registros sin valor que se les dio a las dimensiones, por lo que, para un registro nulo en una tabla de hecho, existe un registro en su respectiva dimensión, en la cual existe un registro con valor especificado es caso de existir entidad o dimensión para ese hecho.

Las métricas no necesitan ser transformadas ya que por defecto desde la fuente de origen los campos asociados para el cálculo son de características no nulas.

### Pedidos

Tabla Origen	Campo Origen	Valor a reemplazar por valor nulo
T_PEDIDO	NO_PEDIDO	0
	PROYECTO_CODIGO	0
	PERIODO_CODIGO	0
	NO_CD	0
	NO_BODEGA	0
T_DET_PEDIDOS	NO_ARTICULO	0
	NO_REG_PEDIDO	0
T_DET_ORDEN_COMPRA	NO_ORDEN	0
	NO_PROVEEDOR	0

Tabla 6. Valores Nulos Tabla de pedidos

### Ventas

Tabla Origen	Campo Origen	Valor a reemplazar por valor nulo
TPROFORMA	ID_PROFORMA	0
	CIU	0
	ID_CREADOR	0
TDETPROFORMA	NO_REG	0
	ID_PERIODO	0
	ID_SUCURSAL	0
	ID_PRODUCTO	0
	ID_PROYECTO	0
	ID_BODEGA	0
	ID_FACTURA	0

Tabla 7. Valores Nulos Tabla de ventas

## Número de Ventas

Tabla Origen	Campo Origen	Valor a reemplazar por valor nulo
TPROFORMA	CIU	0
	ID_CREADOR	0
TDETPROFORMA	ID_SUCURSAL	0
	ID_FACTURA	0

**Tabla 8. Valores Nulos Tabla de número de ventas**

### 3.9.5.2.11 Búsqueda de claves subrogadas en las dimensiones.

Como parte de las transformaciones para llenar las tablas de hechos, un proceso fundamental es el reemplazo de las claves primarias de las tablas de origen por las claves subrogadas primarias de las tablas destino (Dimensiones), esto permite mantener la integridad referencial entre las tablas del DWH.

Para poder observar de mejor manera se presenta un cuadro a continuación especificando la llave primaria origen a reemplazar por la clave subrogada de la dimensión.

## Pedidos

Campos Origen	Campos clave origen	Campos clave subrogados
NO_PEDIDO, PROYECTO_CODIGO, PERIODO_CODIGO, NO_CD, NO_BODEGA, NO_PRODUCTO, NO_ORDEN, NO_PROVEEDOR, FECHA_PEDIDO	FECHA_PEDIDO	LLAVE_TIEMPO
	NO_PRODUCTO	LLAVE_PRODUCTO
	NO_BODEGA	LLAVE_BODEGA
	NO_CD	LLAVE_AREA
	NO_ORDEN	LLAVE_ORDEN_COMPRA
	PERIODO_CODIGO	
	NO_PROVEEDOR	LLAVE_PROVEEDOR
	NO_PEDIDO	LLAVE_PEDIDO
	PERDIDO_CODIGO	
PROYECTO_CODIGO	LLAVE_PROYECTO	

**Tabla 9. Búsqueda llave subrogada Tabla de pedidos**

## Ventas

Campos Origen	Campos clave origen	Campos clave subrogados
FECHA, CIU, ID_SUCURSAL, ID_CENTRO_GESTION, ID_FACTURA, ID_PERIODO, ID_PRODUCTO, ID_PROYECTO, ID_PROFORMA, NO_REG, ID_PERIODO, ID_CREADOR	FECHA	LLAVE_TIEMPO
	CIU	LLAVE_CLIENTE
	ID_SUCURSAL	LLAVE_BODEGA
	ID_CENTRO_GESTION	LLAVE_CENTRO_GESTION
	ID_FACTURA	LLAVE_FACTURA
	ID_PERIODO	
	ID_PRODUCTO	LLAVE_PRODUCTO
	ID_PROYECTO	LLAVE_PROYECTO
	ID_PROFORMA	LLAVE_PROFORMA
ID_PERIODO		

Tabla 10. Búsqueda llave subrogada Tabla de ventas

## Numero Ventas

Campos Origen	Campos clave origen	Campos clave subrogados
CIU, ID_CREADOR, ID_SUCURSAL, FECHA, ID_FACTURA, PERIODO_CODIGO	FECHA	LLAVE_TIEMPO
	CIU	LLAVE_CLIENTE
	ID_SUCURSAL	LLAVE_SUCURSAL
	ID_FACTURA	LLAVE_FACTURA
	ID_PERIODO	
	ID_CREADOR	LLAVE_EMPLEADO

Tabla 11. Búsqueda llave subrogada Tabla número de ventas

### 3.9.5.3 Carga de datos.

La carga de datos se realiza utilizando SCD de tipo 1 de acuerdo a los requerimientos planteados al inicio del proyecto, al igual que la carga de la dimensión los datos actualizados.

#### 3.9.5.3.1 Desarrollo de la ETL Inicial de Hechos con la herramienta Talend Open Studio V. 5.2.

A continuación, se mostrará mediante imágenes el desarrollo del proceso ETL de acuerdo a cada hecho.

## Hecho de Pedidos

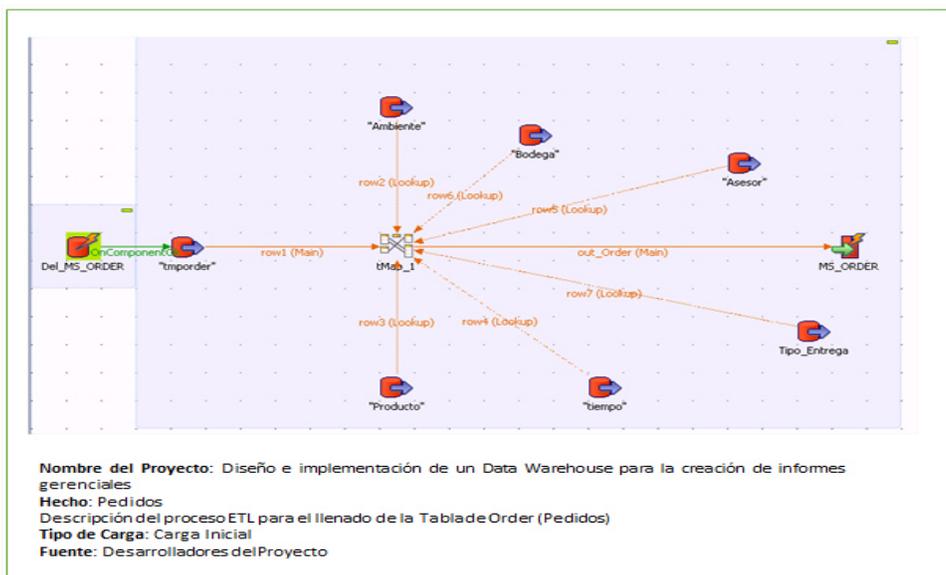


Ilustración 39. Proceso ETL dimensión de pedido – carga inicial

## Hecho de Ventas

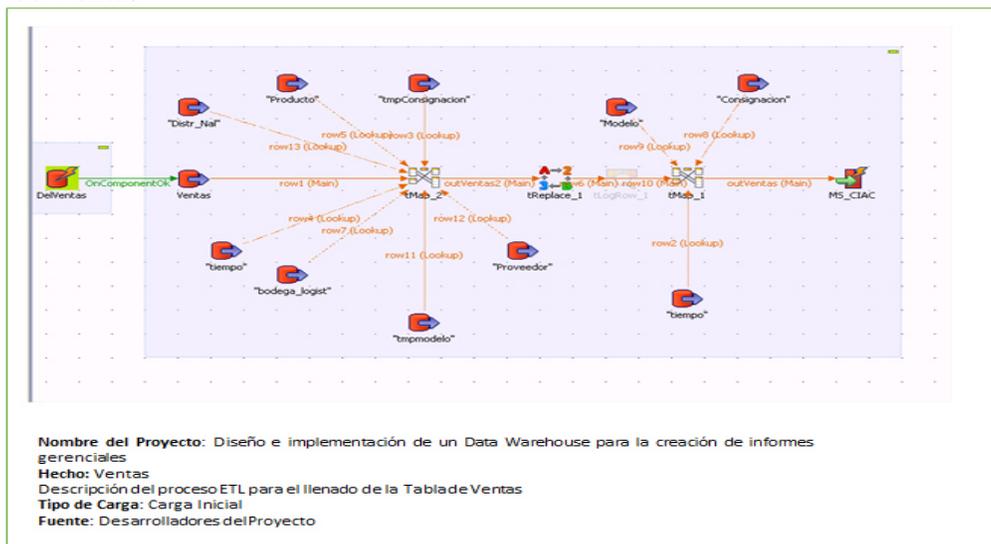


Ilustración 40. Proceso ETL Hecho de ventas – carga inicial

## DESARROLLO DE LA CARGA INCREMENTAL

La carga incremental se da lugar a partir de la segunda ejecución del proceso y carga de datos, para el desarrollo incremental se manejan las mismas políticas que se implementaron para la inicial, pero además de esto se especifica a continuación una nueva funcionalidad en cuestión de extracción de datos.

Como parte del desarrollo de la carga incremental es fundamental tomar en cuenta los datos que son nuevos o que han cambiado su fuente de origen. Esto nos permitirá que la extracción de ellos se realice únicamente sobre filas nuevas o sobre aquellas que hayan sufrido modificaciones.

### 3.9.6 Paso 7: Procesamiento incremental de las dimensiones.

En la carga incremental de las dimensiones se toma en cuenta las transformaciones especificadas en la carga inicial, como el manejo de valores nulos, registros sin valores, búsqueda y agregado de claves subrogadas, etc. Pero se adicionará la nueva funcionalidad sobre extracciones de datos que sean nuevos o que hayan sufrido cambios.

#### 3.9.6.1 Identificar filas nuevas con cambios.

Para implementar esta arquitectura se realizó la extracción de filas de base de datos origen, cuyas claves primarias se encuentran registradas en la tabla de estadísticas, para un mejor entendimiento se mostrará un gráfico a continuación que especifica cómo se realizar la identificación de cambios para cada una de las dimensiones.

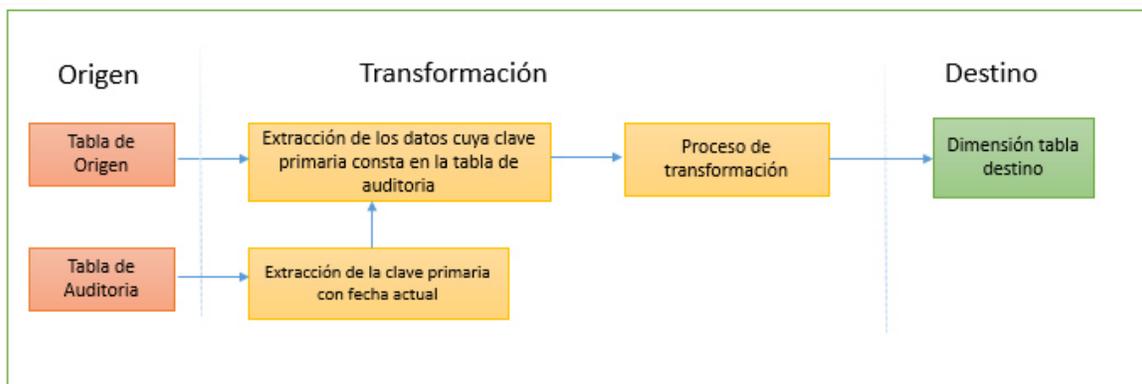


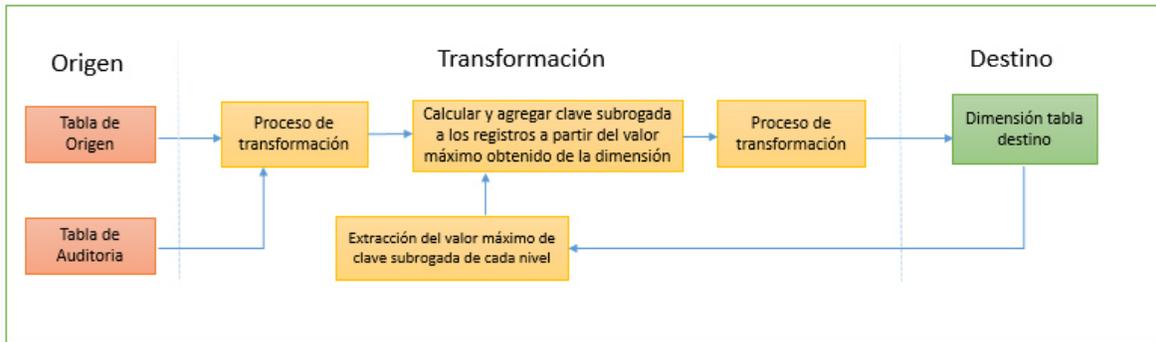
Ilustración 41. Funcionamiento CDC – Dimensiones

#### 3.9.6.2 Extracción de claves subrogadas de las dimensiones.

Este apartado se agregó como parte de la metodología debido a la importancia técnica sobre la función que tiene en el proceso incremental en las dimensiones.

Como parte de la transformación se especifica la asignación de las claves subrogadas para cada nivel de la dimensión. En función a esto y teniendo en cuenta que para el proceso incremental solo se extraerán los datos nuevos o modificados, se diseñó la manera de asignar claves identificando el valor máximo de la clave subrogada del nivel actual que se encuentre en la dimensión. Esto permitirá que en caso de ser datos nuevos se asigne una clave subrogada adecuada para la carga de estos datos en las dimensiones evitando duplicidad en las claves y así poder omitir errores de carga, y en caso de datos modificados no repercutirá en los datos ya que se actualizarán los registros de cada dimensión.

A continuación, se mostrará un gráfico que ilustra mejor su funcionamiento.



**Ilustración 42. Agregar clave subrogada**

### 3.9.6.3 Desarrollo de la ETL Incremental de las dimensiones con la herramienta Talend Open Studio V. 5.2.

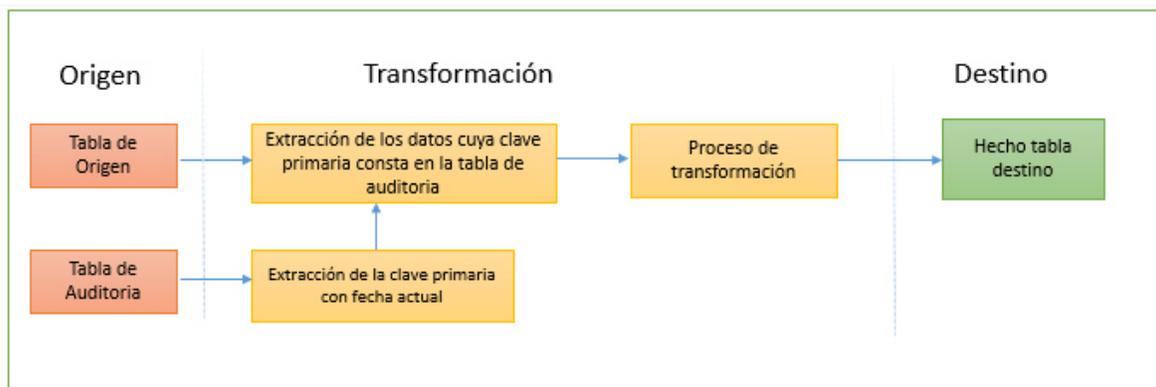
Los diseños para poblar las dimensiones se pueden apreciar en la parte 5 del Anexo 2.

### 3.9.7 Paso 8: Procesamiento incremental de los hechos.

Para la carga incremental de los hechos se toma en cuenta las transformaciones específicas en la carga inicial, como el manejo de valores nulos, la asignación de claves subrogadas, etc. Pero se adicionará la nueva funcionalidad sobre la extracción de datos que sean nuevos o que hayan sufrido cambios.

#### 3.9.7.1 Identificar filas nuevas y con cambios.

En la carga incremental para los hechos en cuanto a identificar los registros de la misma idea que las dimensiones como se aprecia en el siguiente grafico



**Ilustración 43. Funcionamiento CDC – Hechos**

#### 3.9.7.2 Extracción de claves subrogadas de filas de cambio.

Se agregó este punto por la importancia técnica y la cual merece ser destacada. En la carga incremental después de haber extraído únicamente los datos nuevos o aquellos que han sufrido cambios se optimiza mucho el tiempo y rendimiento en el proceso de ETL de los datos, pero al momento de agregarles la clave subrogada de acuerdo a cada dimensión se extraerán los datos de las mismas, haciendo que el rendimiento dependa de la cantidad de datos que existan en las dimensiones. Por lo que para optimizar aún más el proceso de ETL se diseñó una manera en la cual solo se extraigan los registros pertenecientes o relacionados al nuevo registro.

### 3.9.7.3 Desarrollo ETL Incremental de hechos con la herramienta Talend Open Studio V. 5.2.

En las imágenes que se muestran a continuación se muestran las transformaciones incrementales de los hechos.

Hecho de Pedidos

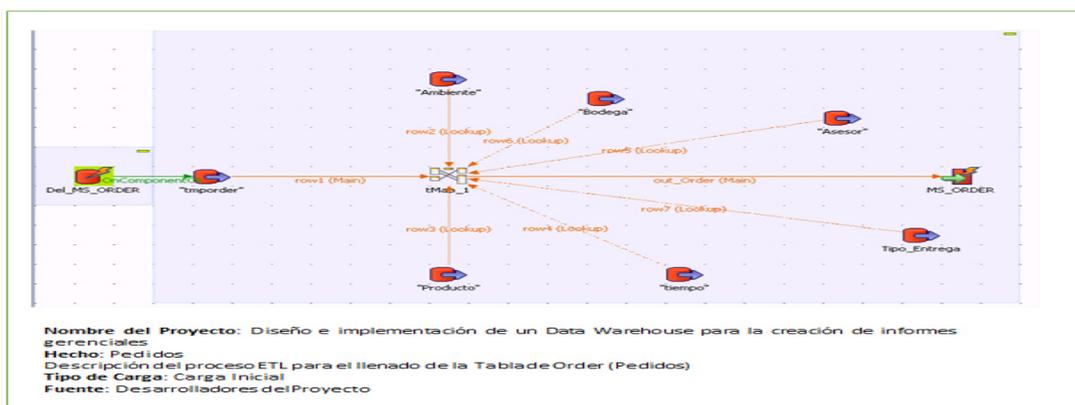


Ilustración 44. Proceso ETL dimensión de pedido – carga inicial

Hecho de Ventas

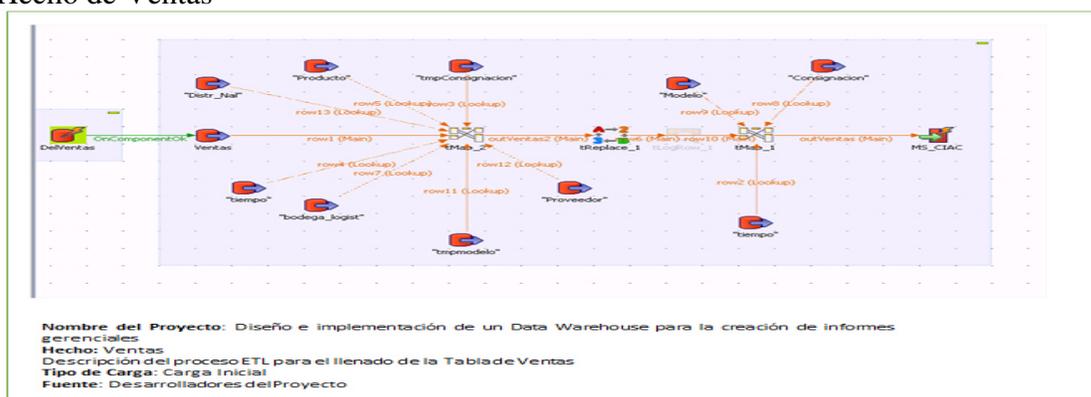


Ilustración 45. Proceso ETL Hecho de ventas – carga inicial

### 3.9.8 Paso 9: Manejo y automatización del sistema ETL

Se realizaron dos actividades empleando un orden para la secuencia de las transformaciones en conjunto con la programación de la tarea para que se ejecute según la periodicidad que se necesite.

### 3.9.8.1 Trabajos ETL.

Para los trabajos de ETL se plantearon los siguientes modelos, tanto para las dimensiones como para los hechos.

Verificar las dimensiones

Se empleará un trabajo el cual valide si existe la tabla de dimensión en el DWH como se muestra en el siguiente gráfico.

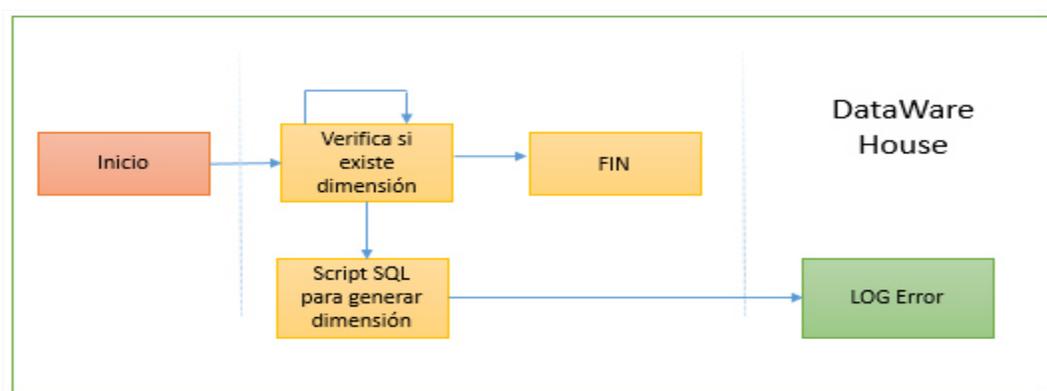


Ilustración 46. Verificar dimensiones

Verificar hechos

Al igual que las dimensiones, para las tablas de hechos también se realizará una validación de existencia como se aprecia en la siguiente figura.

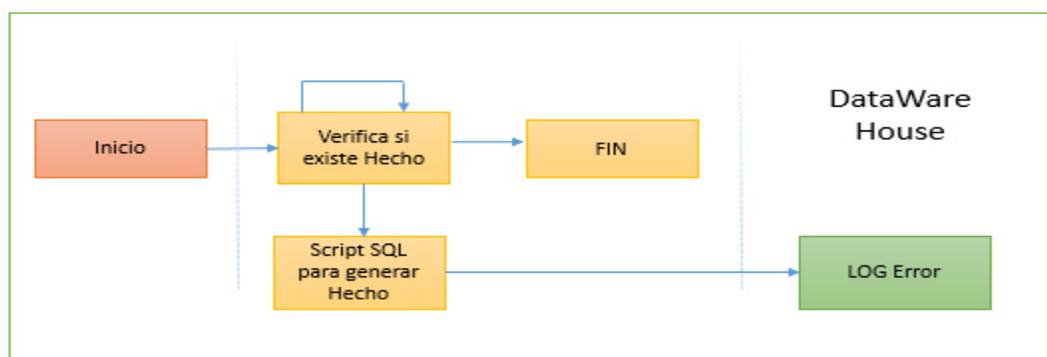


Ilustración 47. Verificar Hechos

Poblar dimensiones

Para poder poblar las dimensiones se realiza un condicional, en caso de haber registros en la tabla se hará una carga de tipo incremental de los datos por medio de un componente especial de la herramienta Talend Open Studio V. 5.2 que permite realizar esta acción, de lo contrario se realizará una carga inicial. A continuación, se ilustra el proceso.

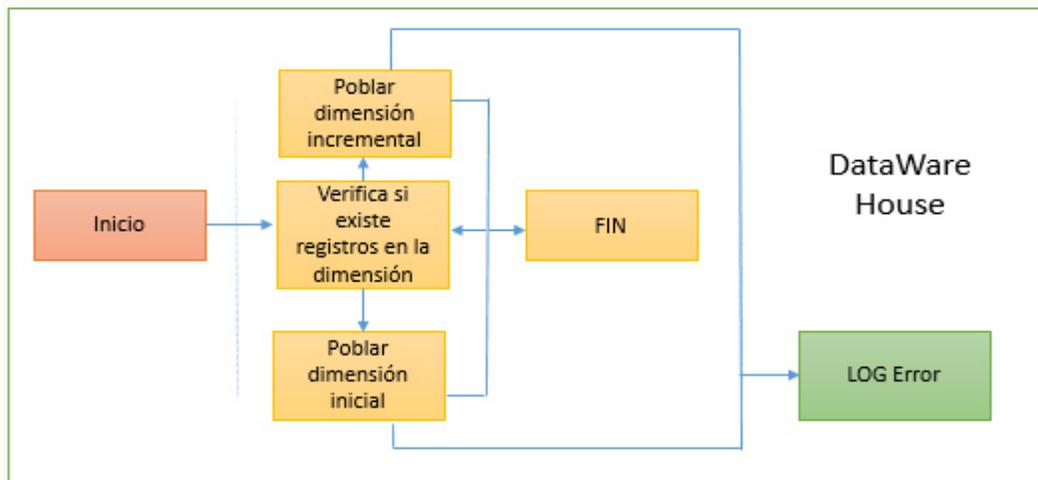


Ilustración 48. Poblar Dimensiones

### Poblar Hechos

Para poblar las tablas de hechos se maneja el mismo esquema que las dimensiones como se muestra a continuación.

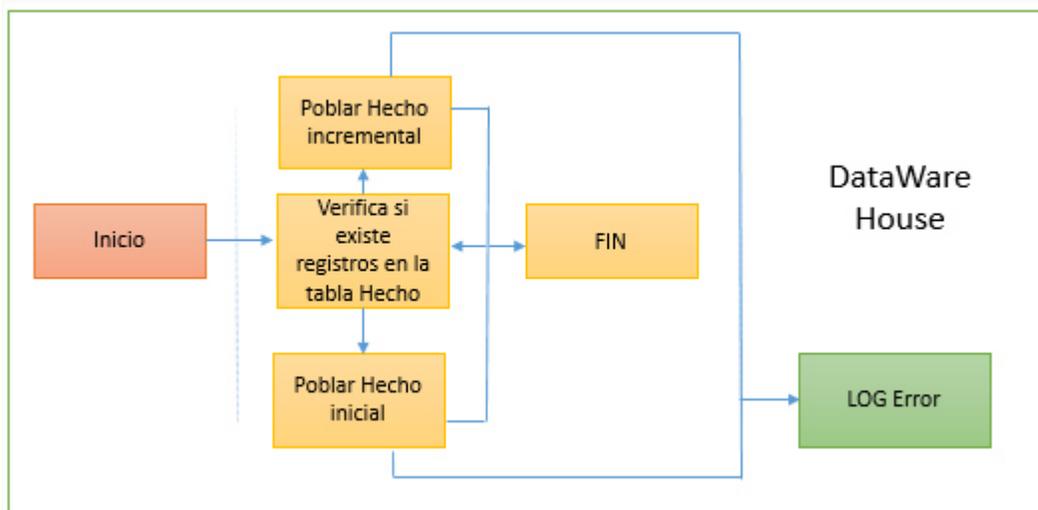
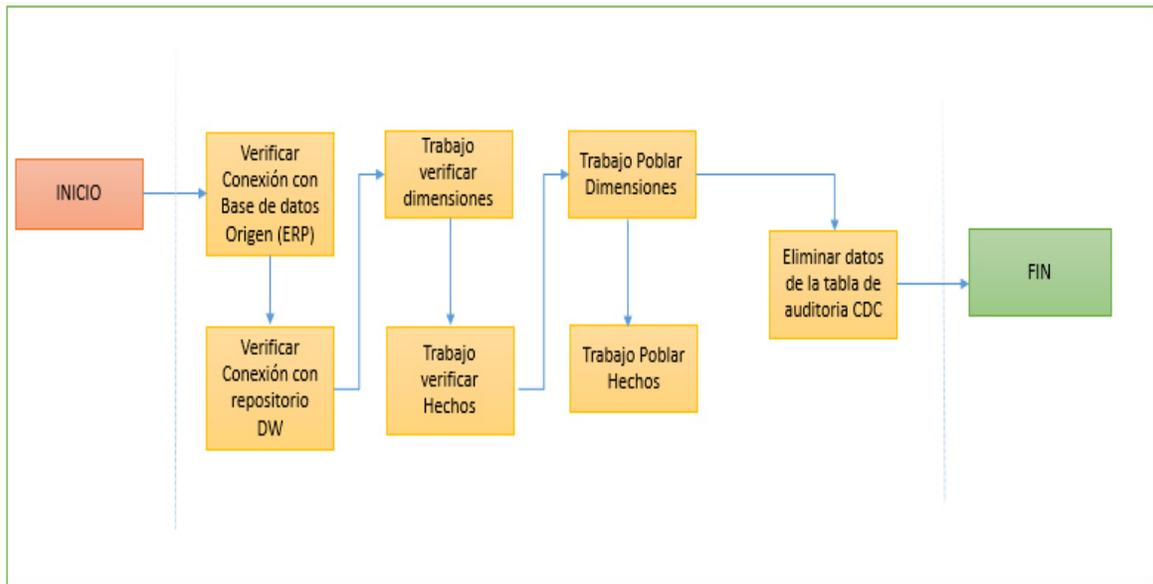


Ilustración 49. Poblar Hechos

### Poblar Data Warehouse

Finalmente se realizará un trabajo el cual englobe los trabajos anteriores lo que nos permitirá seguir una secuencia para el llenar del DW, a continuación, se ilustra en un gráfico.



**Ilustración 50. Poblar DWH**

### 3.9.8.2 Programar trabajos.

El trabajo principal que es poblar el DWH es el que establecerá el horario en el cual se ejecutaran las tareas programadas para su respectiva ejecución. Al ser Talend Open Studio V. 5.2 una herramienta ajena al ERP PeopleSoft 9.1 se tienen que exportar los JOBS construidos para el llenado y actualización de las dimensiones por medio de un programa .BAT que se crea cuando se exporta un JOB desde la herramienta, esto nos permitirá programar su ejecución desde el programador de tareas de Windows Server.

## 3.10 Diseño y desarrollo del DataMart

### 3.10.1 Tipo de aplicación BI.

La aplicación BI al cual está enfocado este proyecto es:

- Construir un reporte por medio de Microsoft Excel que permita visualizar en tiempo real la información de ventas de los puntos de venta que tiene la compañía Almacenes Corona S.A.S. de una manera fácil, coherente que contenga indicadores de cumplimiento y margen y que permita un nivel de granularidad específico.

### 3.10.2 Preparar el desarrollo de la aplicación.

Se configurará la herramienta BI Talend Open Studio V. 5.2 para poder acceder a la información de DWH, así como también a los Metadatos de la compañía para que le herramienta permita actualizar los datos para el reporte final.

#### 3.10.2.1 Poblar los metadatos de la herramienta BI.

Dentro de la compañía Almacenes Corona S.A.S. se encuentra en funcionamiento la herramienta BI Talend Open Studio V. 5.2 que será soportada por los desarrolladores del Área de CIAC (Centro de Análisis de Información de Almacenes Corona) que diseñaron el reporte por medio de Microsoft Excel 2010 y que son los encargados del mantenimiento y actualización de la herramienta.

### 3.10.2.2 Crear los metadatos de negocio.

Se crearán los metadatos de las distintas áreas del negocio tanto para los puntos de venta como para el área comercial por medio de Microsoft Excel 2010 que por su facilidad de manejo genera una interface final lo más entendible posible para nuestros usuarios.

### 3.11 Implementación del DataMart

El siguiente es el diseño del tablero de mando y del informe gerencial para la toma de decisiones:

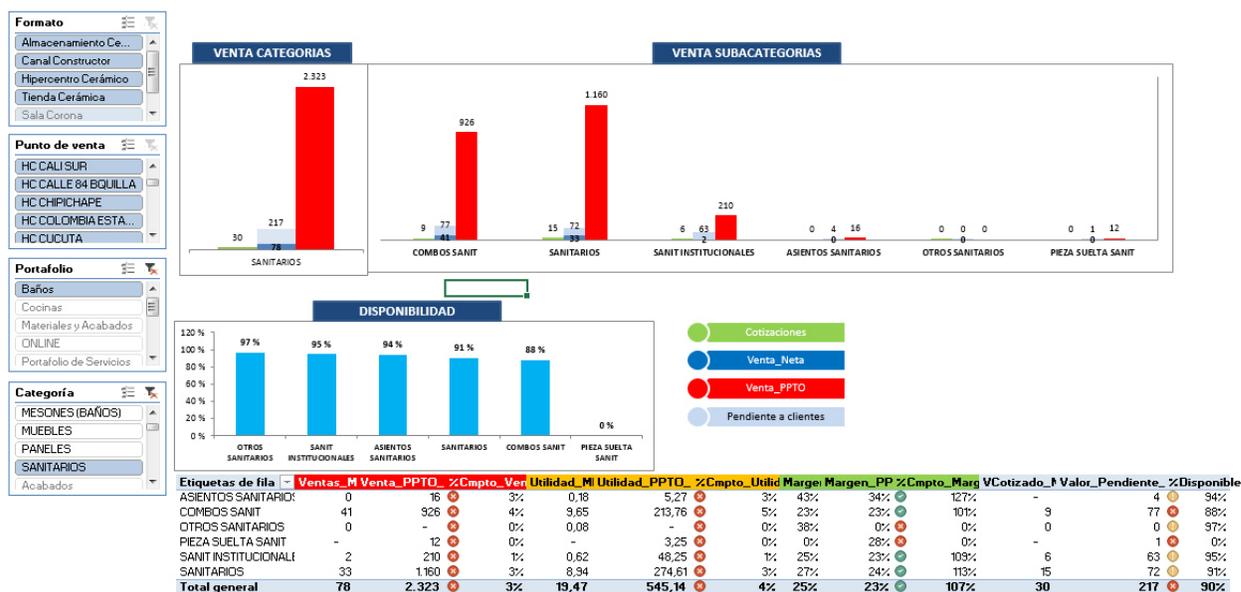


Ilustración 51 Tablero de Mando ALCO

En la ilustración 51, está el diseño del tablero de mando que se implementó en Almacenes corona SAS, en donde se encuentra la información y comportamiento de la compañía por formato, punto de venta, portafolio, categorías y subcategorías de manera gráfica, en donde los usuarios de una manera visual pueden consultar la información de interés para apoyar el proceso de toma de decisiones.

Punto de venta		INCIDENTAS															
Nuevas CC		(Total)															
Nuevas Prevector		(Total)															
Bodega		IN04															
Datos																	
Partida	Categoría	Referencia	Modelo	Ventas	Costo	3M	Pre	Rotaci	Val	Dispo	Val	Teori	Qtv	Dispo	3M	Pre	
Fluoriminister	BAÑOS	10236100	SANITARIO MONTECARLO NOVO AL HET EN CAJA	224,955.917	94,408.574	212,579.258	2,36%	6,27%	61,224.630	95.8%	11,220	28,044	34,561.476	28.0%	-2.0%	57	773,372.648
				40,027.763	162,315.999	92,212.699	1,22%	3,17%	-2,184.936	-34.4%	3,972	28.4%	27.4%	25.2%	-2.0%	47	228,216.559
				47,251.477	104,317.825	69,371.644	1,52%	3,18%	-22,628.936	-34.4%	2,886	39.0%	39.4%	39.4%	-0.4%	56	117,896.017
				7,876.392	27,899.894	10,284.074	222	3.4%	-2,507.772	-24.2%	245	31.5%	20.2%	31.2%	1.3%	51	32,636.296
				13,445.917	47,444.971	11,522.015	2,82%	7.5%	-1,232.490	-26.2%	1,174	33.6%	33.6%	29.8%	-0.4%	40	151,142.467
				41,529.410	45,459.147	23,744.349	370	9.2%	12,325.040	49.0%	42	35.2%	35.2%	37.2%	2.0%	10	108,148.231
				32,740.477	62,710.526	35,233.331	640	8.8%	-2,489.654	-7.0%	109	39.4%	31.0%	29.4%	-0.4%	56	32,075.507
				24,025.678	47,170.651	25,416.259	1,070	6.7%	3,289.476	15.4%	913	35.4%	37.0%	37.8%	-1.0%	66	65,759.424
				9,980.252	25,830.592	16,430.833	96	3.4%	-4,450.241	-39.2%	4	20.2%	15.4%	24.8%	8.8%	2	1,573.461
				4,550.516	8,596.417	2,340.276	163	1.2%	2,168.180	92.4%	20	23.5%	31.4%	25.7%	-7.0%	53	11,021.971
				3,064.622	7,969.739	14,700.776	64	1.2%	-3,393.347	-39.2%	9	30.2%	36.0%	46.4%	14.2%	26	5,175.769
				2,400.142	4,178.915	5,097.951	11	0.2%	-1,699.710	-29.8%	169	21.4%	19.8%	22.3%	1.4%	-	0.0%
				-	459.476	-	4	0.0%	-	0.0%	-	0.0%	23.3%	0.0%	-	16	247.681
				-	201.776	-	11	0.0%	-	0.0%	-	0.0%	0.0%	0.0%	-	30	140,000.000
				93,571.446	105,325.735	56,228.459	567	8.78%	-2,497.011	-2.3%	57,322	23.3%	24.0%	35.8%	-0.2%	29	156,557.278
				4,462.314	20,793.091	11,075.149	727	9.7%	-1,642.963	-14.6%	375	22.4%	22.2%	25.9%	-3.4%	31	41,264.794
				5,002.339	4,402.860	3,399.141	320	3.00%	1,242.240	33.2%	211	10.2%	25.2%	32.4%	-4.2%	52	1,501.615
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000
				1,221.117	1,022.364	720.147	85	0.45%	911.171	299.8%	295	46.4%	52.4%	49.3%	-4.0%	195	3,994.000

referencias que no se están vendiendo con tanta frecuencia y a partir de esta información seguir reforzando la toma de decisiones en la compañía.

## 4. ANALISIS DE RESULTADOS Y CONCLUSIONES

### 4.1 Pruebas de la herramienta.

Se realizan pruebas de funcionamiento de las dimensiones a partir de la herramienta Talend Open Studio V. 5.2 que permita identificar posibles errores en la carga y actualización de los datos con el fin de presentar información real y coherente, esta tarea se apoya por medio de vistas creadas en SQL server que permitan validar que la información generada sea coherente con la del ERP PeopleSoft 9.1.

#### 4.1.1 Desarrollo de la aplicación de BI con la herramienta Microsoft Excel 2010.

Se crearán los niveles del reporte de usuario final que serán los jefes de puntos de venta a nivel nacional de la compañía Almacenes Corona S.A.S.

Creación de las áreas de negocio:

- Conexión con la base de datos INFORMES que se encuentra alojada en un servidor SQL Server llamado AEBOG3102 con el nombre de conexión INFORMESSQLSERVER
- Se darán permisos de consulta a cada uno de los usuarios finales para que puedan acceder a consultar y actualizar los informes sobre la base de datos Informes alojada en el SQL Server.

En el Manual de Usuario final conexión se detallará el paso a paso de como configurar la conexión (DNS) para que el usuario final pueda acceder a actualizar y consultar información en las plantillas de gestión de Punto de Venta.

#### 4.1.2 Despliegue del informe final

Antes de realizar el despliegue y salida a producción del desarrollo se realizarán las siguientes actividades correspondientes a la base de datos del ERP PeopleSoft 9.1 y el DataWare House.

- Creación de espacios de tablas y esquemas para la base de datos del destino del DW, para el repositorio donde se alojaron los metadatos por medio de las representaciones del negocio en las dimensiones y hechos.

Adicional y con el fin de brindar el soporte para los usuarios finales se deben tener en cuenta los siguientes aspectos:

- Personal técnico capacitado en el desarrollo y mantenimiento de la herramienta DW, este soporte lo brindaran los Analistas del área del CIAC.
- Administradores de Base de datos que puedan conceder permisos a los usuarios y hacer mantenimiento de la misma.
- Personal capaz de interpretar los requerimientos de los usuarios finales.
- Personal capacitado en sistemas de información.
- Personal altamente capacitado en el conocimiento de las áreas de la compañía.

Recursos a Nivel de Hardware.

- Dos servidores con sistema operativo Windows Server 2008 que permitan: el primero de ellos AEBOG3102 es donde se encuentra alojado el SQL Server y donde están las bases de datos del DW, es en este servidor donde los usuarios finales entraran a consultar los datos de los informes de ventas y es en esta dónde se autorizaran los permisos para cada una de las base de datos. En el segundo servidor AEBOG3106 es donde se encuentra a instalada la herramienta BI Talend Open Studio V. 5.2 es ahí donde se harán las modificaciones de los JOBS y donde se ejecutarán las tareas programadas para que se mantenga actualizada la información del Data Warehouse.

#### Recursos de Software

- Aplicación para monitorear y administrar la base de datos.
- Herramienta de BI que permitir administrar y programar los JOBS de manipulación de datos Origen – Destino.
- SQL Server 2005 donde se crearon las vistas para los usuarios finales.
- Microsoft Excel 2010 para la presentación del informe final.

#### **4.1.3 Despliegue de la ETL.**

##### Creación del repositorio

Se crea el repositorio y la estructura de almacenamiento de la herramienta ETL mediante tablas en las cuales se registran los trabajos de procesamiento de cada uno de los JOBS ejecutados (Tabla de Estadísticas ubicada en el servidor AEBOG3102 en la base de datos Stage). A continuación, se podrá ver cómo queda almacenado el LOG de cada uno de los JOBS procesados.

moment	pid	father_pid	root_pid	system_pid	project	job	job_repository_id	job_version	context	origin	message_type	message	duration
1/6/2015 7:45:56 PM	HTeeE	HTeeEx	HteeEx	129492	REPORTING	ETL_Cotizac	XJA0ADvEeGQep4W0a6Eg	0.1	Default	<NULL>	end	success	55204
1/6/2015 7:45:01 PM	HTeeE	HTeeEx	HteeEx	129492	REPORTING	ETL_Cotizac	XJA0ADvEeGQep4W0a6Eg	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:35:34 PM	L84F5	L84F5	L84F5	116908	REPORTING	Load_Cotizac	7DvEYPS-EeCj05Q6wQNA9	0.1	Default	<NULL>	end	success	331051
1/6/2015 7:30:03 PM	L84F5	L84F5	L84F5	116908	REPORTING	Load_Cotizac	7DvEYPS-EeCj05Q6wQNA9	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 2:26:11 PM	CSpleq	CSpleq	CSpleq	98972	REPORTING	variabilidad	YP7pF6EeSQ7UPELW13Q	0.1	Default	<NULL>	end	success	81202
1/6/2015 2:26:11 PM	CSpleq	CSpleq	CSpleq	98972	REPORTING	variabilidad	YP7pF6EeSQ7UPELW13Q	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 2:20:30 PM	02csVx	02csVx	02csVx	98228	REPORTING	variabilidad	YP7pF6EeSQ7UPELW13Q	0.1	Default	<NULL>	end	failure	1687
1/6/2015 2:20:29 PM	02csVx	02csVx	02csVx	98228	REPORTING	variabilidad	YP7pF6EeSQ7UPELW13Q	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 2:20:01 PM	HhH2E	HhH2E6	HhH2E6	96644	REPORTING	variabilidad	YP7pF6EeSQ7UPELW13Q	0.1	Default	<NULL>	end	failure	594
1/6/2015 2:20:00 PM	HhH2E	HhH2E6	HhH2E6	96644	REPORTING	variabilidad	YP7pF6EeSQ7UPELW13Q	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 2:07:33 PM	KSh1g	KSh1gB	KSh1gB	104560	REPORTING	tmp_var	YDroMIFEeSMCSxF3eMv5g	0.1	Default	<NULL>	end	success	85762
1/6/2015 2:06:07 PM	KSh1g	KSh1gB	KSh1gB	104560	REPORTING	tmp_var	YDroMIFEeSMCSxF3eMv5g	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 2:05:28 PM	Ma5rC	Ma5rCX	Ma5rCX	121112	REPORTING	tmp_var	YDroMIFEeSMCSxF3eMv5g	0.1	Default	<NULL>	end	failure	2031
1/6/2015 2:05:26 PM	Ma5rC	Ma5rCX	Ma5rCX	121112	REPORTING	tmp_var	YDroMIFEeSMCSxF3eMv5g	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 11:33:29 AM	GD4Sh	GD4Sh8	GD4Sh8	121636	REPORTING	dimensiones	GklJYQOEeGvZP1_ddBw	0.1	Default	<NULL>	begin	<NULL>	18672
1/6/2015 11:33:10 AM	GD4Sh	GD4Sh8	GD4Sh8	121636	REPORTING	dimensiones	GklJYQOEeGvZP1_ddBw	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 10:22:07 AM	UzvfuW	UzvfuW	UzvfuW	104020	CIAC	ETL_Order_M	RWNjWwCBEe57i-ggpt5Qbw	0.1	Default	<NULL>	end	success	18094
1/6/2015 10:21:49 AM	UzvfuW	UzvfuW	UzvfuW	104020	CIAC	ETL_Order_M	RWNjWwCBEe57i-ggpt5Qbw	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 10:11:02 AM	YAdDuZ	YAdDuZ	YAdDuZ	92094	CIAC	Load_Invent	Wt6dQA_4EeC0SLa8m_0vH	0.1	Default	<NULL>	end	success	23465
1/6/2015 10:10:38 AM	YAdDuZ	YAdDuZ	YAdDuZ	92094	CIAC	Load_Invent	Wt6dQA_4EeC0SLa8m_0vH	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 10:00:44 AM	DDKng	DDKngB	DDKngB	102980	REPORTING	Job_Matrimo	81FtG4EeOMwLQy5D1DGQ	0.1	Default	<NULL>	end	success	43282
1/6/2015 10:00:01 AM	DDKng	DDKngB	DDKngB	102980	REPORTING	Job_Matrimo	81FtG4EeOMwLQy5D1DGQ	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 8:26:01 AM	sj38t	sj38t	sj38t	128404	REPORTING	CP_Promedio	G100EeSRKdgt_Ty0A	0.1	Default	<NULL>	end	success	59847
1/6/2015 8:25:01 AM	sj38t	sj38t	sj38t	128404	REPORTING	CP_Promedio	G100EeSRKdgt_Ty0A	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 8:07:06 AM	0j0k7u	0j0k7u	0j0k7u	123472	REPORTING	MS_Desucuent	efF1EFD-EeOIHU-ahzQ-A	0.1	Default	<NULL>	end	success	4740
1/6/2015 8:07:01 AM	0j0k7u	0j0k7u	0j0k7u	123472	REPORTING	MS_Desucuent	efF1EFD-EeOIHU-ahzQ-A	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 8:06:35 AM	t19Tc	t19TcB	t19TcB	121768	REPORTING	Informe_Imp	n_bEESIEeSny-n_3857A	0.1	Default	IMSSgOutputBulE	end	success	2205
1/6/2015 8:06:33 AM	t19Tc	t19TcB	t19TcB	121768	REPORTING	Informe_Imp	n_bEESIEeSny-n_3857A	0.1	Default	IMSSgOutputBulE	end	success	93639
1/6/2015 8:06:33 AM	t19Tc	t19TcB	t19TcB	121768	REPORTING	Informe_Imp	n_bEESIEeSny-n_3857A	0.1	Default	IMSSgOutputBulE	begin	<NULL>	<NULL>
1/6/2015 8:06:28 AM	t19Tc	t19TcB	t19TcB	121768	REPORTING	Informe_Imp	n_bEESIEeSny-n_3857A	0.1	Default	IMSSgOutputBulE	begin	<NULL>	<NULL>
1/6/2015 8:05:01 AM	t19Tc	t19TcB	t19TcB	121768	REPORTING	Informe_Imp	n_bEESIEeSny-n_3857A	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:40:02 AM	gfyqrq	gfyqrq	gfyqrq	104380	REPORTING	DMart_Comen	B1U0H4EeOUC5Mmb6Cg	0.1	Default	<NULL>	end	success	3344
1/6/2015 7:55:01 AM	gfyqrq	gfyqrq	gfyqrq	104380	REPORTING	DMart_Comen	B1U0H4EeOUC5Mmb6Cg	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:48:02 AM	z5KWx	z5KWx0	z5KWx0	121104	REPORTING	ETL_DataMar	DUFgsOdeEomxus9a8A	0.1	Default	<NULL>	end	failure	1297
1/6/2015 7:48:01 AM	z5KWx	z5KWx0	z5KWx0	121104	REPORTING	ETL_DataMar	DUFgsOdeEomxus9a8A	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:45:43 AM	tb5gY2	tb5gY2	tb5gY2	121816	REPORTING	Load_Disponi	FMJCOHMEeOMxdvH7WN	0.1	Default	<NULL>	end	success	42641
1/6/2015 7:45:01 AM	tb5gY2	tb5gY2	tb5gY2	121816	REPORTING	Load_Disponi	FMJCOHMEeOMxdvH7WN	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:40:09 AM	s2U5kx	s2U5kx	s2U5kx	118244	REPORTING	SLO3	JoyCOT5EeKloaq7TEK5KQ	0.1	Default	<NULL>	end	success	6660
1/6/2015 7:40:02 AM	s2U5kx	s2U5kx	s2U5kx	118244	REPORTING	SLO3	JoyCOT5EeKloaq7TEK5KQ	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:25:24 AM	N5s00A	N5s00A	N5s00A	120664	REPORTING	DMart	xogM4EeKuj9QvzjMg	0.1	Default	<NULL>	end	success	22032
1/6/2015 7:25:02 AM	N5s00A	N5s00A	N5s00A	120664	REPORTING	DMart	xogM4EeKuj9QvzjMg	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:21:11 AM	Ev3vm0	Ev3vm0	Ev3vm0	115308	REPORTING	DMart_Main	mgegQM3EeKuJ9QvzjMg	0.1	Default	<NULL>	end	success	70267
1/6/2015 7:21:11 AM	Ev3vm0	Ev3vm0	Ev3vm0	115308	REPORTING	DMart_Main	mgegQM3EeKuJ9QvzjMg	0.1	Default	IMSSgOutputBulE	end	success	2375
1/6/2015 7:21:09 AM	Ev3vm0	Ev3vm0	Ev3vm0	115308	REPORTING	DMart_Main	mgegQM3EeKuJ9QvzjMg	0.1	Default	IMSSgOutputBulE	begin	<NULL>	<NULL>
1/6/2015 7:21:09 AM	Ev3vm0	Ev3vm0	Ev3vm0	115308	REPORTING	DMart_Main	mgegQM3EeKuJ9QvzjMg	0.1	Default	IMSSgOutputBulE	end	success	4765
1/6/2015 7:21:04 AM	Ev3vm0	Ev3vm0	Ev3vm0	115308	REPORTING	DMart_Main	mgegQM3EeKuJ9QvzjMg	0.1	Default	IMSSgOutputBulE	begin	<NULL>	<NULL>
1/6/2015 7:20:30 AM	pp50Ta	pp50Ta	pp50Ta	118348	CIAC	ETL_Pendient	6J7gF6EeKQCbRr5FUA	0.1	Default	<NULL>	end	success	29256
1/6/2015 7:20:01 AM	pp50Ta	pp50Ta	pp50Ta	118348	CIAC	ETL_Pendient	6J7gF6EeKQCbRr5FUA	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:20:01 AM	Ev3vm0	Ev3vm0	Ev3vm0	115308	REPORTING	DMart_Main	mgegQM3EeKuJ9QvzjMg	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:18:14 AM	Lw4n1A	Lw4n1A	Lw4n1A	117140	REPORTING	Factura_Esta	6JdyCQ2Eek_J0FX0ba3rg	0.1	Default	<NULL>	end	success	11610
1/6/2015 7:18:02 AM	Lw4n1A	Lw4n1A	Lw4n1A	117140	REPORTING	Factura_Esta	6JdyCQ2Eek_J0FX0ba3rg	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:07:34 AM	nppCQ6	nppCQ6	nppCQ6	115008	CIAC	ETL_Disponib	H_b-AHEeEK1MvvpW_xv9Gw	0.1	Default	<NULL>	end	success	152439
1/6/2015 7:05:02 AM	nppCQ6	nppCQ6	nppCQ6	115008	CIAC	ETL_Disponib	H_b-AHEeEK1MvvpW_xv9Gw	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:00:56 AM	O2qD5T	O2qD5T	O2qD5T	113120	CIAC	MS_Inventari	mwA3H4EeH59Q3UwTDw	0.1	Default	<NULL>	end	success	54798
1/6/2015 7:00:45 AM	Gg9t5	O2qD5T	O2qD5T	113120	CIAC	MS_De_Inven	r8E4EDWEeOb5ZMwvP8R	0.1	Default	<NULL>	end	success	1062
1/6/2015 7:00:44 AM	Gg9t5	O2qD5T	O2qD5T	113120	CIAC	MS_De_Inven	r8E4EDWEeOb5ZMwvP8R	0.1	Default	<NULL>	begin	<NULL>	<NULL>
1/6/2015 7:00:44 AM	Tz2oH	O2qD5T	O2qD5T	113120	CIAC	MS_Distribuci	xmb54DzEeO6D7R05cV6w	0.1	Default	<NULL>	end	success	11000

Ilustración 54. Ejecución del JOB

Como se puede apreciar en la figura 51, nos muestra la hora de ejecución del JOB, a qué horas comenzó, a qué horas término y además de esto el estado en el cual termino y su respectiva duración.

Guardar o desplegar transformaciones de las bases de datos:

Se guardan las trasformaciones de las tablas finales en la base de datos INFORMES en el servidor AEBOG302 como se muestra en la siguiente imagen.

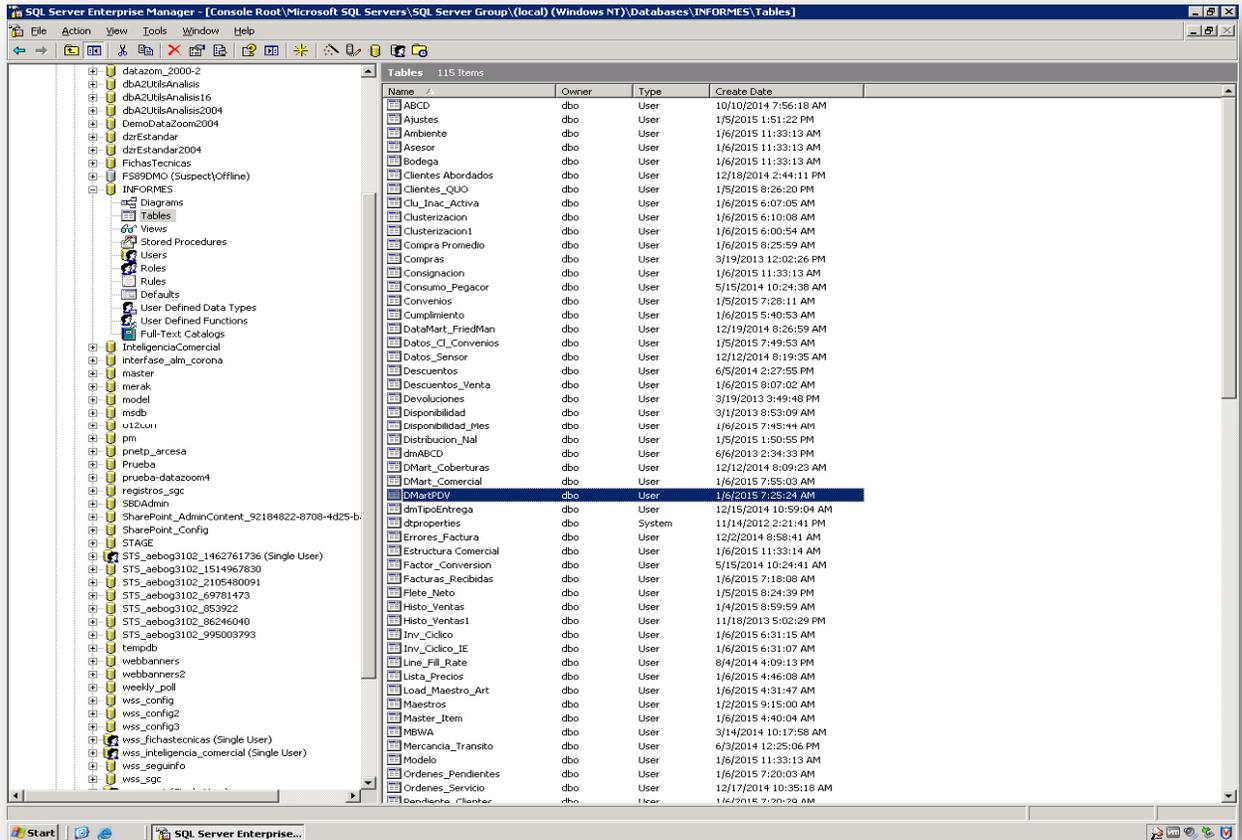


Ilustración 55. Transformaciones de las tablas finales

## Ejecución Inicial

Por medio del trabajo principal se automatiza el resto de los JOBS en el servidor ABOG3106 y se programa por medio de administrador de tareas como se muestra a continuación:

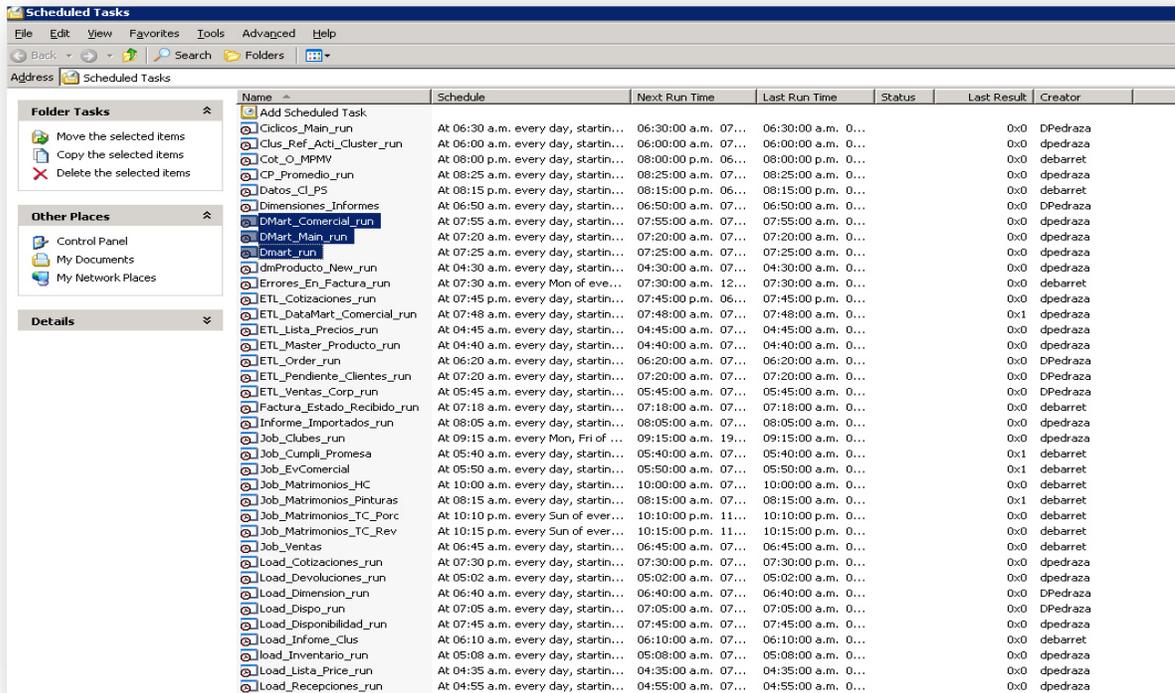


Ilustración 56. Programación del administrador de tareas

### 4.1.4 Despliegue BI.

Teniendo en cuenta los requerimientos de los usuarios finales se procede a la construcción de los Jobs que alimentaran las dimensiones de los informes finales para cada punto de venta.

#### 4.1.4.1 Inventarios.

##### Indicadores

- Indicadores de número de referencias de productos por punto de venta.
- Valor del inventario actual
- Valor del inventario de lento movimiento
- Cantidad de ventas por referencia y punto de venta
- Días de inventario por referencia
- % Valor del inventario

- Cantidad de inventario
- Referencia por Modelo

Valor de Inventario con Corte noviembre 2017

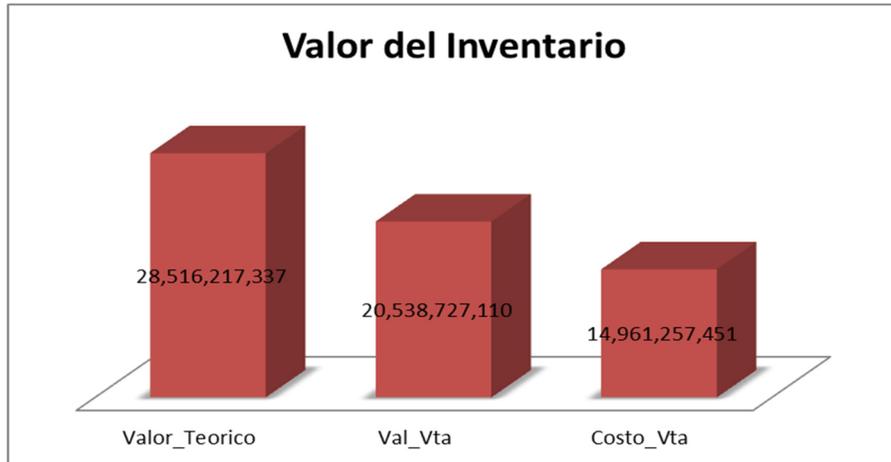


Ilustración 57. Valor de Inventario

Valor de Inventario por origen con Corte noviembre 2017

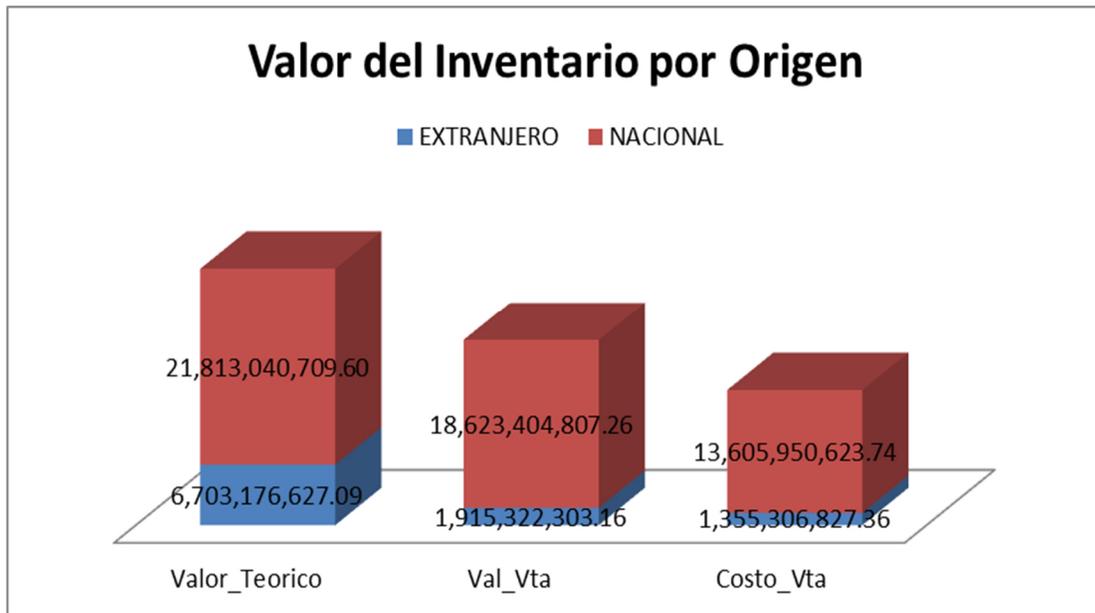


Ilustración 58. Valor de Inventario

Valor de Inventario por Negocio con Corte noviembre 2017

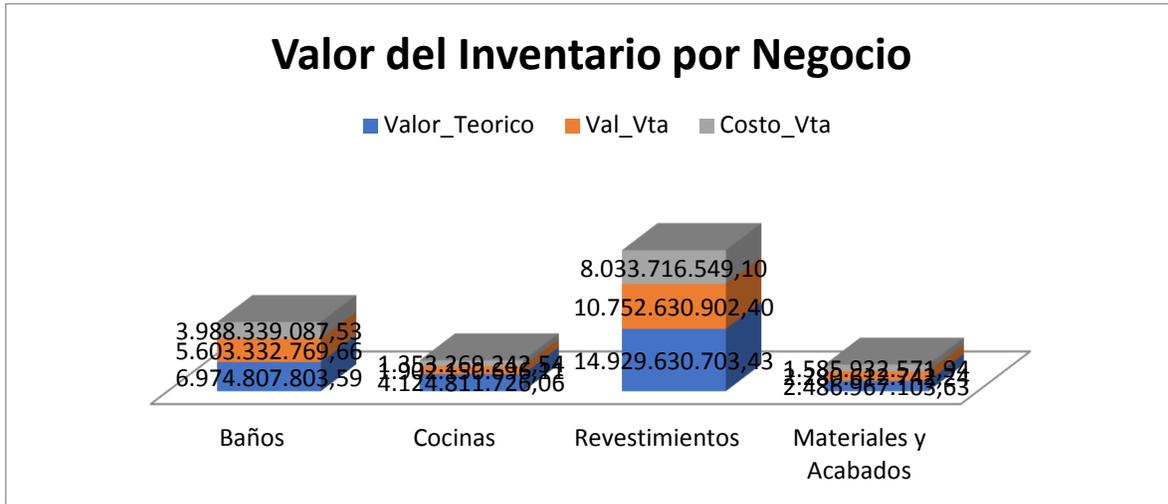


Ilustración 59. Valor de Inventario por Negocio

Valor de Inventario por Formato con Corte noviembre 2017

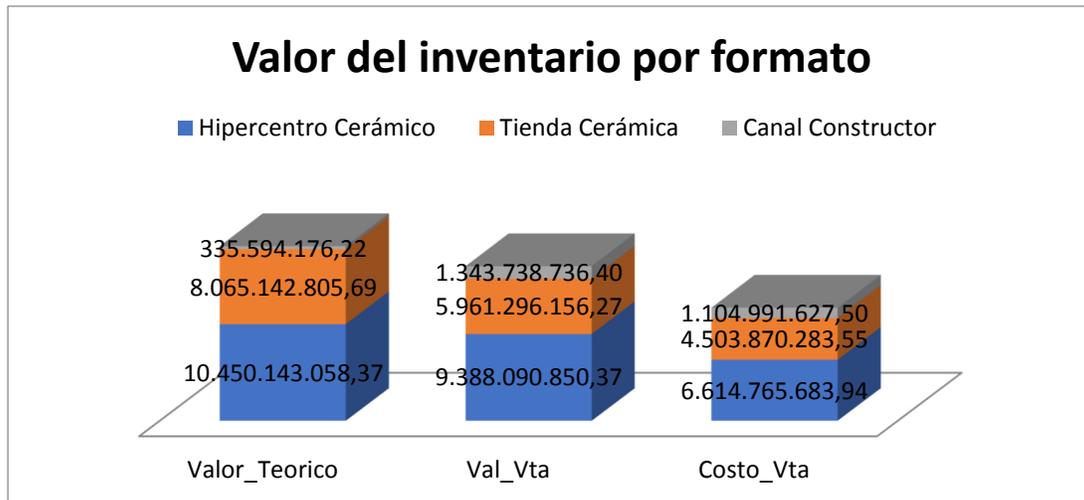


Ilustración 60. Valor de Inventario por Formato

En las figuras anteriores se puede apreciar el inventario que existe en la compañía Almacenes Corona S.A.S.

Puntos de Venta  
Indicadores

- Valor total en ventas del primer día del mes al anterior día de la consulta.
- Comparación de las ventas a la fecha actual y un año antes.
- Indicadores de margen de crecimiento.
- Valor del inventario por punto de venta.

- Top 20 de las referencias más vendidas por portafolio.
- Margen de utilidad por referencia.
- Venta Neta
- Venta promedio de los últimos 3 meses
- Cantidad promedio por referencia
- Valor teórico del inventario
- Valor promedio del inventario
- Valor disponible
- Cantidad Teórica
- Cantidad reservada

#### Mostrar Indicadores

#### Consultas

Para las consultas se realizarán dos informes que se plantearon al inicio del proyecto que sirven para apoyar el proceso de toma de decisiones de los puntos de venta y se construirán de la siguiente manera:

- Informe de Ventas: Este primer informe contiene la información de ventas de un punto de venta que permitirá al usuario final visualizar la siguiente información en tiempo real y poder tener así una visión de cómo va su gestión diaria en cumplimiento con los presupuestos establecidos a principios de mes, este informe tiene los siguientes campos:
  - Punto de Venta
  - Bodega
  - Negocio
  - Id Referencia con descripción
  - Modelo
  - Exhibición
  - Venta Neta
  - Cantidad
  - % Margen de utilidad
  - Días de inventario
  - Valor del Inventario
  - Días de Inventario

- % Valor del inventario
- Cantidad de Inventario

A continuación, se mostrara el diseño del primer informe de Ventas.

A	B	C	D	E	F	G	M	Q	T	U	V	W
Punto de venta	(Todas)	-										
Negocio_CC	(Todas)	-										
						Datos						
Portafolio	Categoria	Subcategoria	Referencia	Modelo	Exhibicion	Venta	Cantidad	%Margen	Dias Inventario	Valor Inventario	% Val Inv	Cant. Inventario
Revestimientos	PISOS					3,922,111,478	210,461	22.7%	0	42,170	2.5%	3
	PAREDES					1,834,867,151	103,068	23.6%	-	-	0.0%	-
	DECORADOS Y COMPLEMENTOS					653,220,024	108,870	36.6%	0	149,509	8.7%	9
	FACHADAS					162,211,628	6,134	30.2%	-	-	0.0%	-
Baños	SANITARIOS					1,633,188,917	7,891	24.3%	-	-	0.0%	-
	GRIFERIAS (BAÑOS)					655,829,699	16,152	34.2%	0	57,722	3.4%	2
	MUEBLES					366,631,800	24,356	33.0%	-	-	0.0%	-
	LAVAMANOS					283,032,279	2,427	32.1%	0	948,870	55.2%	9
	ACCESORIOS (BAÑOS)					219,061,453	4,125	37.7%	-	-	0.0%	-
	BAÑERAS					93,773,197	14,174	30.6%	-	-	0.0%	-
	PANELES					60,024,939	161	41.4%	-	-	0.0%	-
	ESPEJOS					54,270,424	1,978	27.3%	-	-	0.0%	-
	DIVISIONES					49,348,240	12,343	25.9%	-	-	0.0%	-
	CABINAS					26,465,639	117	3.5%	-	-	0.0%	-
	MESONES (BAÑOS)					5,200,491	953	24.9%	-	-	0.0%	-
Materiales y Acabados	PEGANTES E INSTALACIÓN					1,154,717,050	1,023,118	29.6%	-	-	0.0%	-
	PINTURAS Y ACABADOS					188,642,989	19,991	28.3%	-	-	0.0%	-
	MOLDURAS					31,822,638	8,975	54.4%	-	-	0.0%	-
	IMPERMEABILIZANTES					23,296,008	2,450	22.5%	-	-	0.0%	-
Cocinas	GABINETES					144,364,303	26,700	29.5%	-	-	0.0%	-
	GASODOMESTICO					113,385,624	248	17.5%	-	-	0.0%	-
	GRIFERIA (COCINA)					110,205,145	3,692	36.7%	-	-	0.0%	-
	VAJILLAS					78,303,992	14,477	19.0%	-	-	0.0%	-
	ELECTRODOMESTICOS					77,893,740	216	10.5%	0	521,569	59.3%	1
	LAVAPLATOS					71,626,980	549	41.1%	-	-	0.0%	-
	MESONES (COCINA)					68,124,587	281	22.7%	-	-	0.0%	-
	ACCESORIOS (COCINA)					15,211,250	313	23.3%	-	-	0.0%	-
<b>Total PDV</b>						<b>12,146,841,656</b>	<b>1,614,220</b>	<b>26.5%</b>	<b>0</b>	<b>1,719,840</b>	<b>100.0%</b>	<b>24</b>

Tabla 12. Primer informe de ventas

- TOP 20: Este segundo informe contiene la información de ventas por referencia más vendida en el punto de venta que permitirá al usuario final revisar cuál de sus referencias son las que más rotación tienen y las que dejan una mayor ganancia con respecto al margen de utilidad, este informe tiene los siguientes campos:
  - Punto de Venta
  - Bodega
  - Negocio
  - Id Referencia con descripción
  - Modelo
  - Exhibición
  - Venta Neta
  - Venta Promedio a nivel de referencia
  - Cantidad Promedio
  - Días de Inventario



Se realiza un constante monitoreo diario de los reportes implementados con el fin de validar que los datos desplegados sean coherentes con los del ERP PeopleSoft 9.1, para ello se realizaran las siguientes acciones:

- Talend Open Studio V. 5.2 nos permite administrar las ETL que alimentaran a los informes finales, esto con el fin de contemplar cambios en los campos de las tablas, nuevos cálculos de campos y mejoras en la presentación de los informes.
- SQL Server, permite modificar los campos de las tablas finales con nuevos requerimientos que hagan los usuarios: creación, eliminación y actualización de nuevos campos para visualización de nueva información.
- Microsoft Excel 2010, esta herramienta nos permite mantener optimizada la interface del usuario final cuando se modifica algún campo de las tablas del SQL Server.

#### **4.1.5.3 Manejo de la seguridad.**

En esta sección es donde se identificara los roles y permisos que van a tener los usuarios finales que utilizaran los reportes, para ello es muy importante asegurar el uso de los reportes a los usuario indicados ya que por ser esta información de vital importancia para la compañía no cualquier persona debería tener acceso, en el manual de usuario final permisos se detallara la manera de cómo se dan permisos a los usuarios sobre la base de datos de INFORMES que se encuentran en el servidor AEBOG3102.

#### **4.1.6 Back Room.**

##### **4.1.6.1 Monitoreo del sistema ETL.**

Otra parte del mantenimiento de la plataforma es estar pendiente de la ejecución de las tareas programadas por la herramienta Talend Open Studio V. 5.2 las cuales son las que alimentaran el DWH, por ende, se debe realizar a diario un constante monitoreo a la tabla de auditoria (Estadísticas) que nos permitan evaluar que las ejecuciones de las tareas programadas concluyan con éxito. De las cuales se pueden destacar las siguientes:

- Verificación de ejecución exitosa de las tareas programadas.
- Ejecución de los JOBS manualmente con el fin de validar su exitosa ejecución.
- Validación con los usuarios finales acerca de los reportes y su correcto funcionamiento.

##### **4.1.6.2 Administración de espacios en disco.**

Según los datos que se cargaran diariamente en nuestro DW se tienen que hacer un constante monitoreo de cómo se encuentran los espacios en las tablas que manejan en SQL Server, por lo cual se debe aumentar los espacios en disco cuando estén superiores al 80% con el fin de anticiparnos a los problemas.

##### **4.1.6.3 Respaldo y recuperación.**

Teniendo en cuenta que el Data Warehouse tiene información histórica y actual de la compañía con una antigüedad máxima de 10 años es importante tomar medidas para respaldar las bases de datos en discos duros que se encuentran en el servidor y se realizara con una periodicidad semanal. En dado caso que se dañe el almacén de datos se corresponderá a hacer nuevamente la carga inicial de los datos por medio de las ETL ya creadas en la herramienta Talend Open Studio V. 5.2.

Lo que si se debe tener muy presente es hacer constantemente un respaldo de los JOBS que se generan por la herramienta Talend Open Studio V. 5.2 para que cuando ocurra una falla con la herramienta solamente se importen nuevamente los JOBS.

#### **4.1.7 Crecimiento del Data Warehouse**

En este apartado lo que se busca es que a un largo plazo se pueda disponer de la información almacenada con el fin de llevar un crecimiento del Data Warehouse para futuros desarrollos en la compañía Almacenes Corona S.A.S.

Para poder llevar a cabo un crecimiento del DWH se deben de tener en cuenta los siguientes Items:

- Comunicación constante con los usuarios para evaluar nuevos requerimientos.
- Evaluar el nivel de satisfacción con los usuarios.

#### **4.1.8 Evaluar el entorno actual.**

Para poder evaluar el crecimiento y evolución del DW se debe evaluar el ambiente actual donde se encuentra la plataforma con el fin de saber si puede sufrir un aumento que permita suplir las necesidades del negocio y las de los usuarios finales.

- Datos coherentes almacenados en el DW.
- Actualización de datos de manera oportuna.
- Facilidad de acceso a la información de los usuarios finales.
- Dimensiones y hechos organizados.

#### **4.1.9 Oportunidades de crecimiento.**

##### **4.1.9.1 Pequeñas Mejoras**

Las pequeñas mejoras hacen referencia a cambios en la plataforma que se tiene actualmente y que diariamente surgen por medio de los usuarios. Estas mejoras se catalogan como CREACIONES, borrado y actualizaciones de nuevos campos en las distintas tablas de hechos y dimensiones.

#### 4.1.9.2 Grandes Mejoras

Las grandes mejoras se enmarcan en los cambios de datos del negocio que no se encuentran plasmados en el Data Warehouse, es decir por ejemplo agregar nuevas dimensiones, nuevos hechos que son representativos para el negocio.

#### 4.2 conclusiones

- La organización CORONA S.A. cuenta ahora con una base sustentable para justificar las decisiones en el área de ventas, soportado en una herramienta confiable que facilita el trabajo de los usuarios de las bases de datos.
- La investigación juega un papel importante en la inteligencia de negocios y esta a su vez garantiza el análisis y desarrollo de un proyecto para lograr la entrega de un producto de calidad.
- La implementación de la metodología Hefesto facilita el desarrollo de proyectos de Data Warehouse y junto con la participación activa de los usuarios finales de la base de datos se pueden lograr los objetivos propuestos.
- Es posible unificar toda la información de una empresa contenida en varias bases de datos, por medio de un Data Warehouse.
- A futuro este almacén de datos se podría expandir a otras áreas de la organización.
- En una organización las viejas transacciones se cambian por el análisis de información, olvidando las tablas y campos, dando mayor énfasis a las relaciones.

## BIBLIOGRAFÍA

- Bernabeu, Ricardo Darío, (2010). Hefesto metodología para la construcción de un Data Warehouse.
- Cano, Josep Luis, Business Intelligence competir con Información.
- Connolly, Thomas M. (2005) sistemas de bases de datos, un enfoque práctico para diseño, implementación y gestión: ed. Addison-Wesley.
- Gunderloy, Mike. & Jorden, Joseph, (2006). Sql server 2005: ed. Anaya multimedia.
- Imhoff, Claudia., Galemmo, Nicholas., & Geinger, Jonathan, (2003) mastering data warehouse design, relational and dimensional techniques: ed. Wiley.
- Kaplan, Robert S. (2000). Cuadro de mando integral (the balanced scorecard): ed. Gestión.
- Kepner, Charles. (1992) Análisis de problemas y toma de decisiones: Ed. McGraw-Hill.
- Peña Juan David (2005), Utilización de información histórica para decisiones empresariales. Bogotá - Colombia
- Porter, Michael E. (2000) estrategia competitiva, técnicas para el análisis de los sectores industriales y de la competencia.
- Pressman, Roger S. (2001) Ingeniería del Software, Un enfoque práctico: Ed MacGraw-Hill.
- Rizzi, Mauricio. (2006) “sistemas avanzados de base de datos con soporte para la toma de decisiones”, universidad católica de córdoba (ucc).
- Robotham, Sebastián, Rodríguez, (2012). estrategia para la implementación y administración inteligente de Data Warehouse (dw).
- s+Emergentes+de+Base+de+Datos