

**UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR**

**FACULTAD DE INGENIERÍA Y GESTIÓN  
ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS**



**“IMPLEMENTACIÓN DE UN DATAMART PARA FACILITAR EL ANÁLISIS  
DE DATOS PARA LA TOMA DE DECISIONES EN EL ÁREA DE VENTAS  
DE UNA EMPRESA CAMILA VIALI”**

**TRABAJO DE SUFICIENCIA PROFESIONAL**

Para optar el Título Profesional de

**INGENIERO DE SISTEMAS**

**PRESENTADO POR EL BACHILLER**

SANTOS DAVILA, CHRISTIAN JOSEPH

**Villa El Salvador  
2018**

## **DEDICATORIA**

Dedicado a mis padres y hermanas que siempre estuvieron a mi lado durante la vida universitaria brindándome su comprensión y apoyo en todo momento.

## **AGRADECIMIENTO**

Un especial agradecimiento a Dios por encaminarme al logro de mis metas y a toda persona, profesores, familiares y amigos que me ayudaron en este camino profesional.

# ÍNDICE

INTRODUCCIÓN .....	1
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA .....	2
1.1 DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA .....	2
1.2 JUSTIFICACIÓN DEL PROBLEMA .....	3
1.3 DELIMITACIÓN DEL PROYECTO .....	3
1.3.1 Temporal.....	3
1.3.2 Espacial .....	4
1.4 FORMULACIÓN DEL PROBLEMA.....	4
1.4.1 Problema general.....	4
1.4.2 Problemas específicos .....	4
1.5 FORMULACIÓN DE OBJETIVOS .....	4
1.5.1 Objetivo general.....	4
1.5.2 Objetivos específicos .....	4
CAPÍTULO II: MARCO TEÓRICO.....	5
2.1 ANTECEDENTES DE LA INVESTIGACIÓN.....	5
2.2 BASES TEÓRICAS .....	7
2.2.1 DATOS E INFORMACIÓN.....	7
2.2.2 SISTEMAS DE INFORMACIÓN .....	8
2.2.2.1 Definición.....	9
2.2.2.2 Sistemas de procesamiento de transacciones .....	10
2.2.2.3 Sistemas de información MIS Y DSS .....	10
2.2.2.4 Ventajas.....	11
2.2.3 OLTP VS OLAP .....	11
2.2.3.1 OLTP (On-line Transaction Processing) .....	11
2.2.3.2 OLAP (On-line Analytical Processing) .....	12
2.2.3.3 Diferencias entre OLTP Y OLAP .....	12
2.2.4 METODOLOGÍA .....	13
2.2.4.1 METODOLOGÍA HEFESTOS.....	13
2.2.4.2 PASOS Y APLICACIÓN METODOLÓGICA .....	15
2.2.4.2.1 ANÁLISIS DE REQUERIMIENTO .....	15
2.2.4.2.1.1 Preguntas de Negocio.....	15
2.2.4.2.1.2 Indicadores y Perspectivas .....	17
2.2.4.2.1.3 Modelo Conceptual .....	18
2.2.4.2.2 ANÁLISIS DE DATA SOURCES .....	19
2.2.4.2.2.1 Hechos e Indicadores .....	19
2.2.4.2.2.2 Mapeo .....	19
2.2.4.2.2.3 Granularidad .....	20
2.2.4.2.2.4 Modelo Conceptual Ampliado .....	21
2.2.4.2.3 MODELO LOGICO DEL DW .....	22
2.2.4.2.3.1 Tipología .....	22
2.2.4.2.3.2 Diseño de dimensiones.....	23
2.2.4.2.3.3 Tablas de hechos.....	24
2.2.4.2.3.4 Uniones.....	25
2.2.4.2.4 INTEGRACIÓN DE DATOS .....	25
2.2.4.2.4.1 Carga Inicial.....	25
2.2.4.2.4.2 Actualización.....	27

2.2.5 DATAWAREHOUSE .....	28
2.2.5.1 DEFINICION .....	28
2.2.5.2 ARQUITECTURA .....	28
2.2.5.2.1 OLTP (On Line Transaction Processing) .....	29
2.2.5.2.2 LOAD MANAGER .....	29
2.2.5.2.2.1 Extracción .....	30
2.2.5.2.2.2 Transformación .....	31
2.2.5.2.2.3 Carga .....	32
2.2.5.2.3 DATAWAREHOUSE MANAGER .....	32
2.2.5.2.3.1 Base de datos multidimensional .....	33
2.2.5.2.3.2 Tablas de Dimensiones.....	34
2.2.5.2.3.3 Tablas de Hechos .....	35
2.2.5.2.3.4 El esquema en estrella.....	36
2.2.5.2.3.5 Esquema Copo de Nieve .....	38
2.2.5.2.3.6 ROLAP .....	39
2.2.5.2.3.7 MOLAP .....	40
2.2.5.2.3.8 HOLAP .....	41
2.2.5.2.4 QUERY MANAGER .....	41
2.2.5.2.5 HERRAMIENTAS DE CONSULTA Y ANÁLISIS .....	42
2.2.5.2.5.1 Reportes y Consultas .....	44
2.2.5.2.5.2 OLAP .....	44
2.2.3.2.5.3 Data Mining.....	45
2.2.5.2.6 EIS .....	45
2.2.6 DATAMART .....	46
2.2.6.1 Características.....	46
2.2.6.2 Componentes de un Datamart.....	46
2.2.7 TOMA DE DECISIÓN .....	49
2.2.7.1 Proceso de toma de decisiones.....	49
2.2.7.2 Etapas de la Toma de Decisiones. ....	50
2.3 DEFINICIÓN DE TÉRMINOS BÁSICOS.....	52
CAPÍTULO III: DESARROLLO DEL OBJETIVO DE TRABAJO DE SUFICIENCIA 57	
3.1 ANALISIS DE REQUERIMIENTO.....	57
3.2.1 Preguntas de Negocio .....	57
3.1.2 indicadores y Perspectivas .....	58
3.1.3 Modelo Conceptual .....	59
3.2 ANALISIS DE DATA SOURCES.....	60
3.2.1 Hechos e indicadores .....	60
3.2.2 Mapeo.....	61
3.2.3 Granularidad .....	62
3.2.4 Modelo conceptual ampliado .....	67
3.3 MODELO LÓGICO DEL DW .....	68
3.3.1 Tipología .....	68
3.3.2 Tablas de Dimensiones .....	68
3.3.3 Tablas de Hechos .....	72
3.3.4 Uniones.....	72
3.4 INTEGRACION DE DATOS.....	74
3.4.1 Carga Inicial.....	74
3.4.2 Cubo Multidimensional.....	79
3.4.3 Resultados de Implementación.....	83

3.4.4 KPI ..... 87  
3.4.5 Cronograma y Presupuesto ..... 88  
CONCLUSIONES..... 90  
RECOMENDACIONES ..... 91  
BIBLIOGRAFÍA ..... 92  
ANEXOS ..... 96

## ÍNDICE DE FIGURAS

<b>FIGURA 1:</b> FASES DE LA METODOLOGÍA HEFESTOS (FUENTE: DATAPRIX) ..	13
<b>FIGURA 2:</b> EJEMPLO DE PERSPECTIVAS E ÍNDICES (FUENTE: DATAPRIX) ...	18
<b>FIGURA 3:</b> MODELO CONCEPTUAL GENERAL (FUENTE: DATAPRIX).....	18
<b>FIGURA 4:</b> EJEMPLO MODELO CONCEPTUAL (FUENTE: DATAPRIX) .....	19
<b>FIGURA 5:</b> EJEMPLO DE MAPEO (FUENTE: DATAPRIX) .....	20
<b>FIGURA 6:</b> EJEMPLO MODELO CONCEPTUAL AMPLIADO (FUENTE:DATAPRIX) .....	22
<b>FIGURA 7:</b> EJEMPLO DEL DISEÑO DE DIMENSIONES (FUENTE: DATAPRIX) .	23
<b>FIGURA 8:</b> EJEMPLO DE JERARQUÍA DE DIMENSIÓN (FUENTE: DATAPRIX)..	24
<b>FIGURA 9:</b> EJEMPLO GENERACIÓN TABLA DE HECHOS (FUENTE: DATAPRIX) .....	24
<b>FIGURA 10:</b> EJEMPLO DE LA UNIÓN DE TABLAS DIMENSIONALES Y DE HECHOS (FUENTE: DATAPRIX) .....	25
<b>FIGURA 11:</b> EJEMPLO DE CARGA DE TABLAS DIMENSIONALES (FUENTE: DATAPRIX) .....	26
<b>FIGURA 12:</b> EJEMPLO DE SENTENCIAS SQL UTILIZADAS (FUENTE: DATAPRIX) .....	27
<b>FIGURA 13:</b> ESTRUCTURA OLTP. (FUENTE:DATA WAREHOUSING- BERNABEU).....	29
<b>FIGURA 14:</b> CUBO MULTIDIMENSIONAL. (FUENTE:DATA WAREHOUSING- BERNABEU).....	33
<b>FIGURA 15:</b> TABLAS DIMENSIONALES (FUENTE: DATAPRIX).....	35
<b>FIGURA 16:</b> RELACIÓN TABLAS DIMENSIONALES Y TABLA HECHOS (FUENTE: DATAPRIX) .....	36
<b>FIGURA 17:</b> MODELO ESTRELLA (FUENTE: DATAPRIX).....	37
<b>FIGURA 18:</b> EJEMPLO MODELO ESTRELLA (FUENTE: DATAPRIX) .....	37
<b>FIGURA 19:</b> MODELO COPO DE NIEVE (FUENTE: DATAPRIX) .....	38
<b>FIGURA 20:</b> PROCESO DE CONSULTA Y ANÁLISIS FUENTE: DATA WAREHOUSING-BERNABEU).....	43
<b>FIGURA 21:</b> IMPLEMENTACIÓN MODELO CONCEPTUAL (FUENTE PROPIA) ..	59
<b>FIGURA 22:</b> IMPLEMENTACIÓN DEL MAPEO (FUENTE PROPIA) .....	61

<b>FIGURA 23:</b> IMPLEMENTACIÓN MODELO CONCEPTUAL AMPLIADO .. (FUENTE PROPIA).....	68
<b>FIGURA 24:</b> TABLA DIMCLIENTE (FUENTE PROPIA) .....	69
<b>FIGURA 25:</b> TABLA DIMEMPLEADO (FUENTE PROPIA) .....	69
<b>FIGURA 26:</b> TABLA DIMLOCAL (FUENTE PROPIA).....	70
<b>FIGURA 27:</b> TABLA DIMPRENDA (FUENTE PROPIA) .....	70
<b>FIGURA 28:</b> TABLA DIMITIPOPAGO (FUENTE PROPIA).....	71
<b>FIGURA 29:</b> TABLA DIMITIEMPO (FUENTE PROPIA).....	71
<b>FIGURA 30:</b> IMPLEMENTACIÓN TABLA DE HECHOS (FUENTE PROPIA) .....	72
<b>FIGURA 31:</b> IMPLEMENTACIÓN DE UNIONES (FUENTE PROPIA).....	73
<b>FIGURA 32:</b> CARGA DIMCLIENTE (FUENTE PROPIA).....	74
<b>FIGURA 33:</b> CARGA DIMEMPLEADO (FUENTE PROPIA) .....	75
<b>FIGURA 34:</b> CARGA DIMPRENDA (FUENTE PROPIA).....	75
<b>FIGURA 35:</b> CARGA DIMLOCAL (FUENTE PROPIA) .....	76
<b>FIGURA 36:</b> CARGA DIMITIPOPAGO (FUENTE PROPIA) .....	77
<b>FIGURA 37:</b> CARGA DIMITIEMPO (FUENTE PROPIA) .....	77
<b>FIGURA 38:</b> CARGA FACVENTA (FUENTE PROPIA) .....	78
<b>FIGURA 39:</b> CREACIÓN DEL CUBO (FUENTE PROPIA) .....	79
<b>FIGURA 40:</b> AGREGAR TABLA FACVENTA (FUENTE PROPIA).....	79
<b>FIGURA 41:</b> CREA DIMENSIONES (FUENTE PROPIA) .....	80
<b>FIGURA 42:</b> CREAR JERARQUÍAS (FUENTE PROPIA).....	80
<b>FIGURA 43:</b> AGREGAR TABLA DIMENSIONAL (FUENTE PROPIA) .....	81
<b>FIGURA 44:</b> CREAR NIVELES (FUENTE PROPIA).....	81
<b>FIGURA 45:</b> DIMENSIONES DEL CUBO (FUENTE PROPIA).....	82
<b>FIGURA 46:</b> CREACIÓN DE MÉTRICAS (FUENTE PROPIA).....	82
<b>FIGURA 47:</b> REPORTE DE CANTIDAD DE PRENDAS POR MODELO, COLOR Y TIENDAS ESPECÍFICAS (FUENTE PROPIA).....	83
<b>FIGURA 48:</b> GRÁFICO GENERAL DE LOS COLORES MÁS VENDIDOS DE UN MODELO ESPECÍFICO (FUENTE PROPIA).....	84
<b>FIGURA 49:</b> GRÁFICO DE DETALLE DE CANTIDADES DE PRENDA (FUENTE PROPIA).....	84
<b>FIGURA 50:</b> REPORTE DE VENTAS GENERALES POR TIENDA (FUENTE PROPIA).....	85

<b>FIGURA 51: GRÁFICO DE RECONOCIMIENTO RÁPIDO DE DETALLES</b>	
(FUENTE PROPIA).....	85
<b>FIGURA 52: REPORTE DE MONTO DE VENTAS POR EMPLEADO</b>	
(FUENTE PROPIA).....	86
<b>FIGURA 53: KPI MONITOREO DE VENTAS (FUENTE PROPIA)</b> .....	87
<b>FIGURA 54: KPI PRENDAS MÁS VENDIDAS (FUENTE PROPIA)</b> .....	88
<b>FIGURA 55: CRONOGRAMA DEL PROYECTO (FUENTE PROPIA)</b> .....	88

## INDICE DE TABLAS

<b>TABLA 1:</b> DIFERENCIA ENTRE OLTP Y OLAP (FUENTE PROPIA).....	12
<b>TABLA 2:</b> FORMATO ENCUESTA (FUENTE PROPIA) .....	16
<b>TABLA 3:</b> PRESUPUESTO (FUENTE PROPIA).....	89

## INTRODUCCIÓN

Como es de saber la tecnología está creciendo a pasos agigantados y por tales motivos es esencial que las organizaciones se adapten para cumplir con las competencias que exige el mercado y no quedar obsoletos.

“Camila Viali”, una empresa dedicada al rubro textil, productora y comercializadora de prendas de vestir elegantes para damas, cuenta con 6 puntos de ventas en Lima y 1 un punto de venta en Guayaquil; la situación actual del problema es referente a los reportes que solicita el área de ventas como por ejemplo los reportes mensuales y anuales de ventas, cantidad de prendas vendidas en un lapso de tiempo por determinados puntos de ventas, etc. estos demoran un tiempo de 2 días, generar dicho reporte por el motivo de que se cuenta con dos base de datos distintas entre Lima y Guayaquil lo que genera demora en la integración de los datos y se necesita agilizar estos procesos para tomas de decisiones futuras.

Ante esta situación se propuso la implementación de un Datamart que se ha desarrolla bajo la metodología Hefestos que cuenta con los pasos detallados de análisis y construcción de un Datamart, se logró obtener un modelo dimensional adecuado para la empresa y en base a esto se pudo implementar el Datamar con la herramienta Pentaho, esta herramienta es open source y cuenta con un conjunto de programas de Inteligencia de negocio los cuales pasaremos a nombrar: Spoon herramienta de Integración de datos (ETL), Schema Workbench necesario para la creación del cubo dimensional y Pentaho Business Analytics para visualización de los reportes.

El presente proyecto se divide en 3 capítulos; Capítulo I: Planteamiento del problema, en esta parte se definieron las problemáticas y los objetivos que se necesitan cumplir; Capitulo II: Marco teórico esencial para dar a conocer en base a que se va a desarrollar el proyecto y el Capítulo III: Desarrollo del objetivo de trabajo de suficiencia, este capítulo muestra cómo se realizó la implementación del proyecto con en base a la metodología y herramientas utilizadas.

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

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

“Camila Viali” es una empresa del rubro textil, donde su principal actividad es la comercialización de prendas de vestir para damas, ésta tiene puntos de venta en Lima y Guayaquil, además cuenta con bases de datos transaccionales de ventas donde la base de datos de Lima es distinta a la de Guayaquil lo que genera que la obtención y análisis de la información demore más tiempo, el encargado de la administración de la base de datos estima que el tiempo aproximado para generar los reportes con los procesos actuales son de 2 días, 1 día para integrar la información de las bases de datos de Ecuador con las de Perú y otro día preparar las hojas de cálculos en tablas dinámicas.

Se realizaron 2 entrevistas en el mes de diciembre 2017, primero al Jefe de ventas Sergio Saldaña que declara que el envío de prendas se realiza semanalmente a Ecuador y del total enviado al mes, el 10% son retornados al Perú por depreciación, esto se debe a que no se realiza un análisis de las prendas con los datos históricos que cuenta la empresa, por ejemplo saber cuáles son las prendas más vendidas en cierto periodo ayudaría reducir el porcentaje de retorno de

prendas del Ecuador. En la segunda entrevista el Jefe de diseño Alberto Salomón menciona que para la realización de los diseños lo hacen en base a tendencias de revistas internacionales, pero no se han enfocado en un estudio de los modelos, colores y tallas de prendas que más han preferido los clientes en cada temporada (3 meses) de años anteriores lo que le podría generar un ahorro de costes al saber las preferencia de los clientes.

## **1.2 JUSTIFICACIÓN DEL PROBLEMA**

Se puede decir que con el desarrollo y aplicación de esta herramienta se ayudara a mejorar el rendimiento y la eficiencia en el análisis de datos para tomar mejores decisiones en procesos de ventas, diseño, etc. Lo más notable con esta implementación será que la empresa podrá usar data más confiable y ayudara a realizar una predicción más certera.

La tecnología de un Datamart resulta sumamente importante para las organizaciones para el análisis de sus datos pero esta pueden resultar ser muy costosas, ante esa dificultad existe Pentaho una herramientas de análisis e informes integrado para procesos de negocio como, minería de datos, ETL, etc. además de ser gratuita (GPL).

## **1.3 DELIMITACIÓN DEL PROYECTO**

### **1.3.1 Temporal**

El proyecto tendrá una duración de dos meses y medio, y empezará el 15 de diciembre del 2017 y finalizará el 20 de febrero del 2018.

### **1.3.2 Espacial**

El proyecto se realizara en Lima metropolitana, en el distrito de Santa Anita  
- Av. María Parado de Bellido Nro. 920 Co. Universal. Situada en la oficina  
Principal de "Camila Viali"

## **1.4 FORMULACIÓN DEL PROBLEMA**

### **1.4.1 Problema general**

¿Cómo implementar un Datamart para facilitar el análisis de los datos para tomar las mejores decisiones en el área de ventas de la empresa 'Camila Viali'?

### **1.4.2 Problemas específicos**

¿Cómo obtener los requerimientos de los interesados?

¿Cómo se obtendrá el modelo multidimensional?

¿Cómo se obtendrá los datos desde la base de datos origen?

## **1.5 FORMULACIÓN DE OBJETIVOS**

### **1.5.1 Objetivo general**

Implementar un Datamart para facilitar el análisis de datos para tomar las mejores decisiones en el área de ventas de la empresa 'Camila Viali'.

### **1.5.2 Objetivos específicos**

- Recopilación de requerimientos para la presentación de reportes por parte de los interesados.
- Analizar y diseñar el modelo multidimensional.
- Construcción del proceso de extracción de datos (ETL)

## **CAPÍTULO II: MARCO TEÓRICO**

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

Nacionales

Rojas (2014). Realizó la investigación: “Implementación de un Datamart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el Departamento de Finanzas de la Contraloría general de la república” esta investigación mejoró el promedio de rapidez de la obtención de la información, también se obtuvieron los reportes para estimaciones anuales, trimestrales y mensuales con respecto a las órdenes de servicios realizados.

Durand (2014). Realizó la investigación: “Desarrollo de un Datamart para mejorar la toma de decisiones en el área de ventas de la corporación Furukawa”, los resultados que se obtuvieron fue la visualización de reportes de ventas por cada sub área durante el periodo 2012, se pudo conocer cuáles son los vendedores que más venden en cada uno de sus respectivas áreas y monitorear cuales son

los clientes que más compraron durante el año 2012 que será de importancia para el área de marketing.

Moreno (2013). Realizó la investigación: “Análisis, Diseño e Implementación de Datamart para las áreas de ventas y recursos humanos de una empresa dedicada a la exportación e importación de productos alimenticios”, los resultados obtenidos tras la ejecución del proyecto son los reportes de ventas que mediante los filtros correspondientes se pueden evaluar la cantidad de devoluciones que se tuvo por cada cliente en una determinada fecha, también se logró evaluar el porcentaje de horas trabajadas por cada empleado y evaluar si cubren todas sus horas que deberían laborar.

Villanueva (2008). Realizó la investigación: “Análisis, Diseño e Implementación de un Datawarehouse de soporte de decisiones para un hospital del sistema de salud público” ayudó a que el hospital tenga toda su información consolidada y ordenada en un solo lugar, entre sus resultados que se obtuvo fue la generación del reporte de ingreso por cada servicio el cual se corroboró la integridad de los datos y que el desarrollo de cada uno de los pasos para desarrollar los reportes ayudaron a una mejor toma de decisión a las autoridades pertinentes.

#### Internacionales

Guevara y Vera (2009). Realizó la investigación: “Desarrollo de una solución de inteligencia de negocios para un departamento de finanzas dentro del sector petroquímico” se tuvieron como resultado que con el desarrollo de inteligencia de negocios se disminuía el tiempo de la realización de la extracción de los datos y

la realización de los reportes en un 90%, gracias a la correcta definición de las métricas o indicadores se pudo cubrir todas las preguntas por parte del departamento de finanzas de manera satisfactoria

Hernández (2008). Realizó la investigación: “Diseño y Construcción de un Datamart para la mantención de indicadores de sostenibilidad de la industria del salmón” en esta investigación se concluyó que la arquitectura de un Datawarehouse era lo más conveniente ya que necesitaban una herramienta que fuera lo suficientemente flexible frente a cambios.

Aimacaña (2013). Realizó la investigación: “Análisis, Diseño e Implementación de un Datamart académico usando tecnología de BI para la Facultad de Ingeniería, Ciencias Físicas y Matemática”. Tiene por finalidad que se filtren datos desde el ambiente operacional, al depósito del Datamart. Se han creado indicadores que ayudan a obtener la información de una manera más rápida que tener que pasar por un proceso largo como lo hacían hoy en día de forma manual, optimizando tiempo de respuesta y desarrollo.

## **2.2 BASES TEÓRICAS**

### **2.2.1 DATOS E INFORMACIÓN**

Datos

Un dato es una representación formalizada de entidades o hechos, adecuada para la comunicación, interpretación y procesamiento por medios humanos o automáticos. El dato es un material de valor escaso o nulo para un individuo en una situación concreta; es una representación simbólica que

por sí misma no reduce la dosis de ignorancia o el grado de incertidumbre de quien tiene que tomar una decisión. Sakora (2002, p.24)

### Información

La información es un dato o un conjunto de datos evaluados por un individuo concreto que trabaja, en un momento dado, sobre un problema específico, para alcanzar un objetivo determinado. La información se genera a partir de un grupo de datos seleccionados para reducir la dosis de ignorancia o el grado de incertidumbre de quien debe adoptar una decisión. Sakora (2002, p.24)

Podemos decir que los datos y la información están muy relacionada pero no siempre un conjunto de datos generara información, todo esto dependerá de la persona que solicite esta información o quien la revise, por ejemplo un conjunto de datos de fórmulas químicas para crear un nuevo compuesto no le será de información al personal de limpieza ya que no sabrían cómo interpretar a diferencia del químico que a pesar de ver lo mismo que el personal de limpieza sabe interpretar estos datos.

## **2.2.2 SISTEMAS DE INFORMACIÓN**

En situaciones de nuestra vida ha sido necesario obtener información para tomar alguna decisión o realizar una acción de algo, como por ejemplo para comprar un ordenador portátil hay que saber que marcas son las más recomendables, que hardware es óptimo y necesario, que sistema

operativos nos convendría utilizar análogamente para las pequeñas empresas es necesario.

### **2.2.2.1 Definición**

Los sistemas de información según Oz (2008, p.11) está formado por todos los componentes que colaboran para procesar los datos y producir información. Casi todos los sistemas de información empresariales están integrados por muchos subsistemas con metas secundarias, todas las cuales contribuyen a la meta principal de la organización.

También se puede definir al sistema de información como el conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. En un sentido amplio, un sistema de información no necesariamente incluye equipo electrónico (*hardware*). Sin embargo, en la práctica se utiliza como sinónimo de “sistema de información computarizado”. Cohen y Asín (2000 p.4).

En una organización todos los procesos que tienen generan una fuente de información, es de esta manera que las organizaciones tienen una gran variedad de sistemas para la generación de esta, y si se trabajara de manera separada, un solo sistema no brindaría la información necesaria para la toma de decisión de algún hecho.

### **2.2.2.2 Sistemas de procesamiento de transacciones**

Es un sistema computarizado que efectúa y registra las transacciones diarias de rutina necesarias para realizar negocios, como introducir pedidos de ventas, reservaciones de hoteles, nómina, registro de empleados y envíos. El principal propósito de los sistemas en este nivel es responder a las preguntas de rutina y rastrear el flujo de transacciones por toda la organización. ¿Cuántas piezas están en el inventario? ¿Qué ocurrió con el pago del Sr. Smith? Para responder a estos tipos de preguntas, por lo general la información debe estar fácilmente disponible, actualizada y precisa. Kanneth y Jane Laudon (2012,p.45-46)

### **2.2.2.3 Sistemas de información MIS Y DSS**

Sistemas de información gerencial (MIS) designa una categoría específica de sistemas de información que dan servicio a la gerencia de nivel medio. Los MIS proveen a los gerentes de este nivel reportes sobre el desempeño actual de la organización. Esta información se utiliza para supervisar y controlar la empresa, además de predecir su desempeño en el futuro. (p.47)

Sistemas de soporte de decisiones (DSS) Brindan apoyo a la toma de decisiones que no es rutinaria. Se enfocan en problemas que son únicos y cambian con rapidez. Tratan de responder a preguntas como éstas: ¿Cuál sería el impacto en los itinerarios de producción si se duplicaran las ventas en el mes de diciembre? ¿Qué ocurriría con nuestro rendimiento sobre la inversión si se retrasara el itinerario de una fábrica por seis meses? (p.48).

#### **2.2.2.4 Ventajas**

Para Laudon (2012, p. 537) las ventajas de sistemas de información son las siguientes:

- Mejora en el uso de los activos
- Mejora en el control de los recursos
- Mejora en la planificación organizacional
- Mejora en la flexibilidad organizacional
- Más información oportuna
- Aumento en el aprendizaje organizacional
- Mejora en la toma de decisiones
- Mejora en las operaciones
- Mejor imagen corporativa.

#### **2.2.3 OLTP VS OLAP**

##### **2.2.3.1 OLTP (On-line Transaction Processing)**

Se caracteriza por un gran número de transacciones cortas en línea (INSERT, UPDATE, DELETE). El énfasis principal para los sistemas OLTP se basa en el procesamiento de consultas muy rápido, manteniendo la integridad de los datos en entornos de acceso múltiple y una efectividad medida por el número de transacciones por segundo. En la base de datos OLTP hay datos detallados y actuales, y el esquema utilizado para almacenar bases de datos transaccionales es el modelo de entidad (generalmente 3NF). Datawarehouse4u (2009, ¶ 2)

### 2.2.3.2 OLAP (On-line Analytical Processing)

Se caracteriza por un volumen relativamente bajo de transacciones. Las consultas suelen ser muy complejas e implican agregaciones. Para los sistemas OLAP, un tiempo de respuesta es una medida de efectividad. Las aplicaciones OLAP son ampliamente utilizadas por las técnicas de Data Mining. En la base de datos OLAP hay datos históricos agregados, almacenados en esquemas multidimensionales (por lo general, esquema en estrella). Datawarehouse4u (2009, ¶ 3)

### 2.2.3.3 Diferencias entre OLTP Y OLAP

OLTP	OLAP
Orientado a lo operativo (procesos)	Orientado a temas
Predomina la actualización	Predomina la consulta. Datos históricos
Se accede a pocos registros	Procesos masivos, se accede a muchos registros
Datos altamente normalizados	Datos desnormalizados
Estructura relacional	Estructura multidimensional
Tiempos de respuesta muy rápidos (pocos registros)	Tiempos de respuesta rápidos (respuesta datos masivos)
Estructura estática	Estructura dinámica, abundantes cambios

Tabla 1: Diferencia entre OLTP y OLAP (Fuente Propia)

Podemos decir que cada uno tiene sus ventajas para las funcionalidades donde se desempeñan pero para este caso la OLAP sería la más indicada para realizar análisis de datos más exhaustivamente, ya que sirven para la toma de decisiones que son de importancia para las organizaciones.

## 2.2.4 METODOLOGIA

### 2.2.4.1 METODOLOGÍA HEFESTOS

Es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos. Se puede destacar que esta metodología agrega retroalimentación utilizados en diversos países. La idea principal, es comprender cada uno de los pasos que se realizara, para no tener complicaciones y atrasos en su ejecución. Mediante el siguiente gráfico se podrá seguir paso a paso los puntos a seguir de esta metodología.



Figura 1: Fases de la Metodología Hefestos (Fuente: Dataprix)

Esta metodología cuenta con las siguientes características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para Data Warehouse como para Data Mart

## **2.2.4.2 PASOS Y APLICACIÓN METODOLÓGICA**

### **2.2.4.2.1 ANÁLISIS DE REQUERIMIENTO**

En primera instancia se deberá identificar los requerimientos de los usuarios a través de preguntas que expliciten los objetivos de su organización. Luego se analizarán estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. Finalmente se confeccionará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso.

Es importante recordar que esta metodología es utilizada para construir un DataWarehouse o un Datamart a la vez, es decir en caso de que se construyan más de un datamart, para cada uno de ellos por separado se tendría que utilizar la esta metodología .

#### **2.2.4.2.1.1 Preguntas de Negocio**

El primer paso comienza con la recolección de las necesidades de información, el cual puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc.

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 estrategias de la empresa, y que facilitará una eficaz y eficiente toma de decisiones.

A continuación se presenta un modelo de entrevista que será utilizado para la recolección de información del personal estratégico de la empresa

INFORMACION GENERAL
Nombre del entrevistado: _____
Cargo: _____ Fecha de entrevista: ____/____/____
Lugar de encuesta: _____

<ol style="list-style-type: none"><li>1. ¿Cuáles son las necesidades más prioritarias que tiene el área al que pertenece?</li><li>2. ¿Cómo evaluaría la entrega de reportes actualmente?</li><li>3. ¿En qué tiempo le gustaría tener los reportes para su análisis y evaluación?</li><li>4. ¿Qué tan variables es el tiempo para los reportes que requiere?</li><li>5. ¿Cuáles son los indicadores que usted considera más importantes para el análisis de la información?</li></ol>
--

Tabla 2: Formato Encuesta (Fuente Propia)

Como ejemplo se muestran dos repuestas estratégicas para el proceso de ventas de la organización

- Se desea conocer cuántas unidades de cada producto fueron vendidas a sus clientes en un periodo determinado. O en otras palabras: *“Unidades vendidas de cada producto a cada cliente en un tiempo determinado”*.

- Se desea conocer cuál fue el monto total de ventas de productos a cada cliente en un periodo determinado. O en otras palabras: *“Monto total de ventas de cada producto a cada cliente en un tiempo determinado”*.

#### **2.2.4.2.1.2 Indicadores y Perspectivas**

Una vez que se han establecido las preguntas claves, y se ha obtenido la información necesaria se debe proceder a su descomposición y análisis de la misma para identificar los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

Para ello, se debe tener en cuenta que los indicadores, para que sean realmente efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc.

En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar, que el Tiempo es muy comúnmente una perspectiva.

Aquí tenemos dos ejemplos de posibles enunciados o preguntas que se identificaron para el análisis de las estrategias de una empresa, donde se pueden identificar las perspectivas e indicadores.



Figura 2: Ejemplo de perspectivas e índices  
(Fuente: Dataprix)

### 2.2.4.2.1.3 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 ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y explicado con facilidad.

La representación gráfica del modelo conceptual es la siguiente:

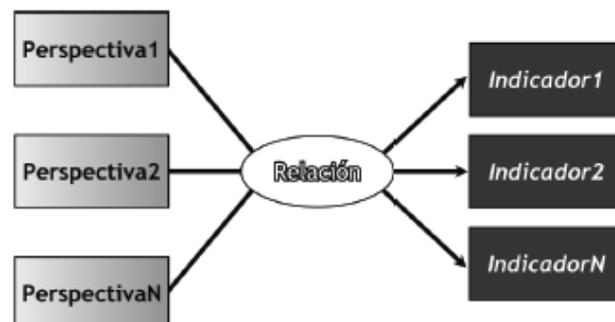


Figura 3: Modelo conceptual General  
(Fuente: Dataprix)

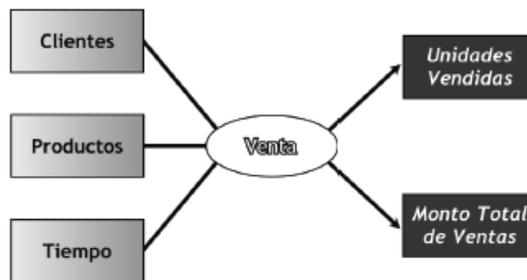


Figura 4: Ejemplo modelo conceptual  
(Fuente: Dataprix)

A la izquierda se colocan las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación, constituye el proceso o área de estudio elegida. De dicha relación y entrelazadas con flechas, se desprenden los indicadores o medidas, estos se ubican a la derecha del esquema.

## 2.2.4.2.2 ANÁLISIS DE DATA SOURCES

### 2.2.4.2.2.1 Hechos e Indicadores

En este paso se deberán explicitar cómo se calcularán los Indicadores, definiendo los siguientes conceptos para cada uno de ellos:

- Hecho/s que lo componen, con su respectiva fórmula de cálculo. Por ejemplo: Hecho1 + Hecho2
- Función de agregación que se utilizará. Por ejemplo: SUM, AVG, COUNT, etc.

### 2.2.4.2.2.2 Mapeo

El objetivo de este análisis, es el de examinar los OLTP disponibles que contengan la información requerida, como así también sus

características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

En el caso de los indicadores, deben explicitarse como se calcularán, y más aún si son fórmulas u operaciones complejas.

La idea es, que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

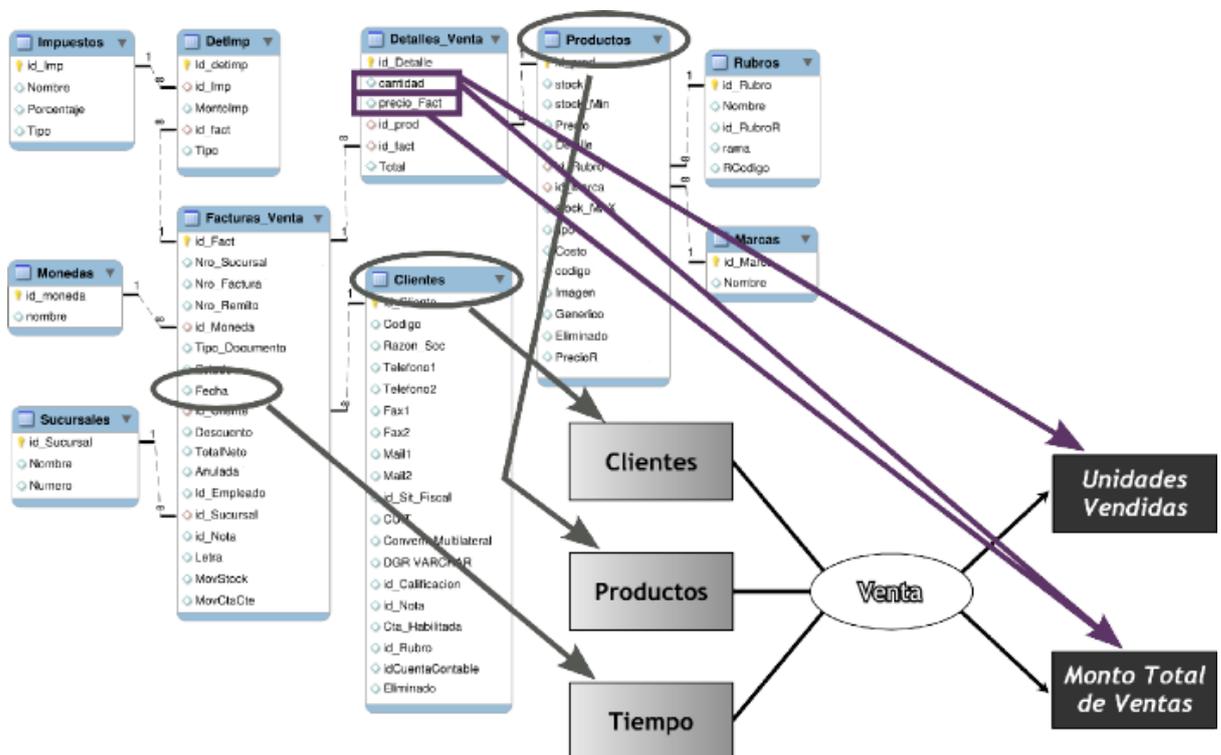


Figura 5: Ejemplo de Mapeo (Fuente: Dataprix)

### 2.2.4.2.2.3 Granularidad

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

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

Luego de exponer frente al usuario los datos existentes, explicando su significado, valores posibles y características, este debe decidir cuáles son los que considera relevantes para consultar los indicadores y cuáles no.

Con respecto a la perspectiva “Tiempo”, es muy importante definir el ámbito mediante el cual se agruparán o sumaran los datos. Este punto es fundamental precizarlo con claridad, debido a que, determinará la granularidad de la información encontrada en el DW. Sus campos posibles pueden ser: día de la semana, quincena, mes, trimestres, semestre, año, etc.

#### **2.2.4.2.2.4 Modelo Conceptual Ampliado**

Finalmente, y con el fin de graficar los resultados obtenidos, se ampliará el modelo conceptual expuesto anteriormente, colocando bajo cada perspectiva los campos o atributos elegidos.

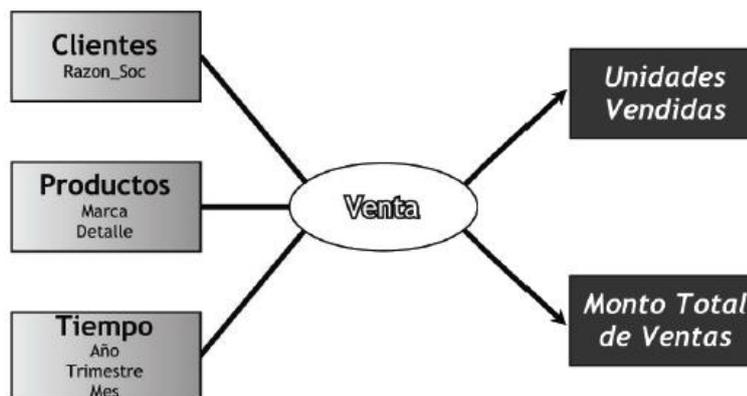


Figura 6: Ejemplo modelo conceptual Ampliado  
(Fuente: Dataprix)

### 2.2.4.2.3 MODELO LOGICO DEL DW

Luego de haber obtenido el modelo conceptual se deberá elaborar el modelo lógico, en esta parte se tendrá que escoger el esquema que más se adapte a las necesidades y requerimientos del usuario, entre los esquemas que podemos elegir tenemos al esquema estrella, copo de nieve o constelación.

#### 2.2.4.2.3.1 Tipología

Debemos seleccionar el tipo de Esquema que mejor se adapte a los requerimientos y necesidades de los usuarios. El Modelo Lógico seguirá este esquema.

Es muy importante definir objetivamente si se empleará un Esquema en Estrella, Copo de Nieve o Constelación, ya que esta decisión afectará considerablemente la elaboración del Modelo Lógico.

### 2.2.4.2.3.2 Diseño de dimensiones

Lo primero que se hará será crear las dimensiones del mismo, para ello se tomará cada perspectiva con sus atributos relacionados y se les realizará el siguiente proceso:

- Se elegirá un nombre que identifique la dimensión.
- Se añadirá un campo que represente su clave principal.
- Se redefinirán los nombres de los atributos si es que no son lo bastante explicativos.

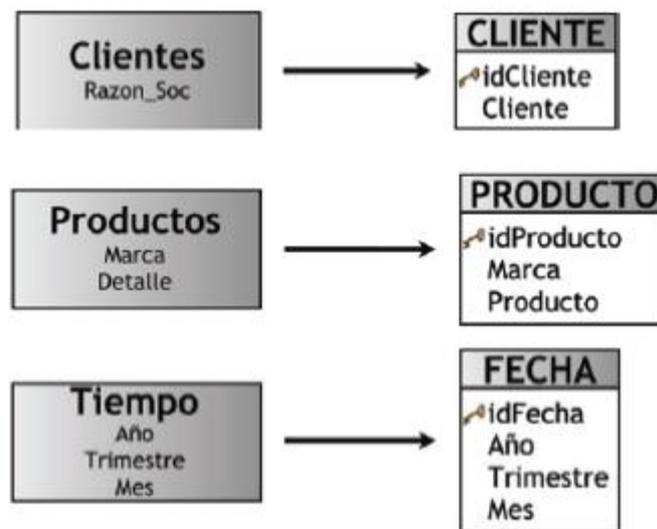


Figura 7: Ejemplo del diseño de dimensiones  
(Fuente: Dataprix)

Para los esquemas en estrella y constelación, se deberán especificar las jerarquías que existirán dentro de cada tabla de dimensión, teniendo siempre presente cual es el objetivo de las mismas. Para representar las jerarquías en el modelo lógico, se deberán colocar los atributos pertenecientes a las jerarquías en sus respectivas tablas, en orden descendente y acompañado con un número ordinal encerrado entre corchetes.

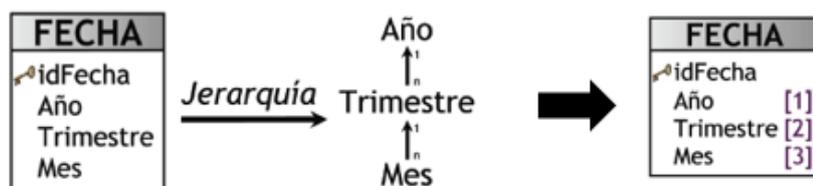


Figura 8: Ejemplo de jerarquía de dimensión  
(Fuente: Dataprix)

### 2.2.4.2.3.3 Tablas de hechos

Para los esquemas en estrella y copo de nieve, se realizará lo siguiente:

- Al igual que las dimensiones, se le deberá asignar un nombre a la tabla de hechos que en este caso represente la información analizada, área de investigación, negocio enfocado, etc.
- Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada dimensión que se utilizará para generar las consultas.
- Se renombrarán los hechos o indicadores si es que no llegasen a ser lo suficientemente explícitos.

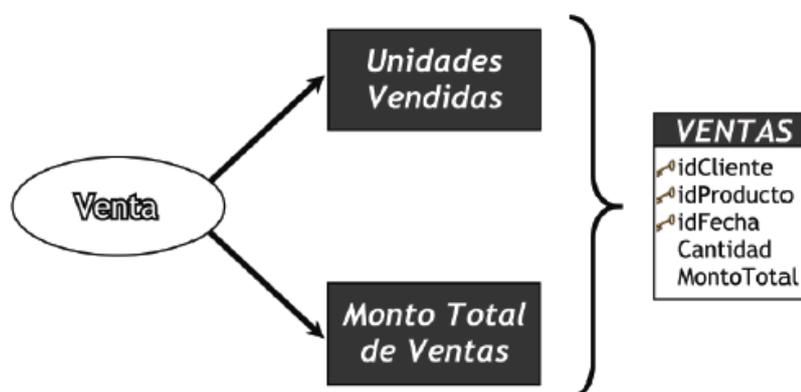


Figura 9: Ejemplo generación tabla de hechos  
(Fuente: Dataprix)

#### 2.2.4.2.3.4 Uniones

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

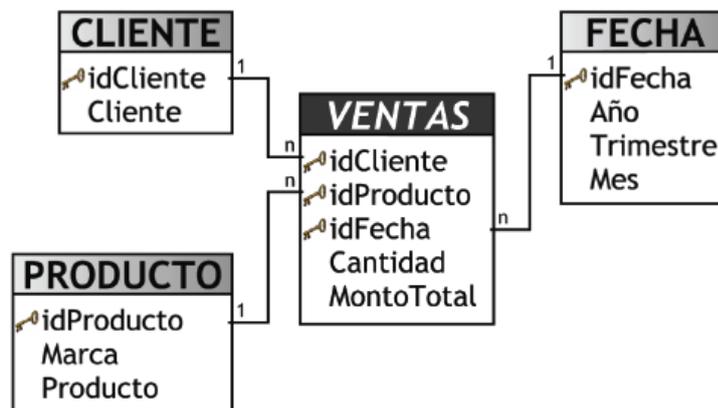


Figura 10: Ejemplo de la unión de tablas dimensionales y de hechos (Fuente: Dataprix)

#### 2.2.4.2.4 INTEGRACIÓN DE DATOS

Una vez construido el Modelo Lógico, se deberá proceder a probarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc.

##### 2.2.4.2.4.1 Carga Inicial

Debemos en este paso realizar la Carga Inicial del DW, poblando el modelo construido en pasos anteriores. Para lo cual debemos llevar adelante una serie de tareas básicas, tales como asegurar la limpieza y calidad de los datos, procesos ETL, etc.

Primero se cargarán los datos de las Dimensiones y luego los de las tablas de Hechos. En el caso en que se esté utilizando un Esquema Copo de Nieve, cada vez que existan Jerarquías de

Dimensiones, se comenzarán cargando las tablas de Dimensiones del nivel más general al más detallado. Esto se debe a la existencia de claves foráneas y se realiza para evitar problemas de rechazo de datos por parte del SGBD.

Concretamente, en este paso se deberán registrar en detalle las acciones llevadas a cabo con los diferentes Softwares de Integración de datos. Por ejemplo, es común que sistemas ETL trabajen con Pasos y Relaciones, en donde cada Paso realiza una tarea en particular del Proceso ETL y cada Relación indica hacia donde debe dirigirse el flujo de datos.

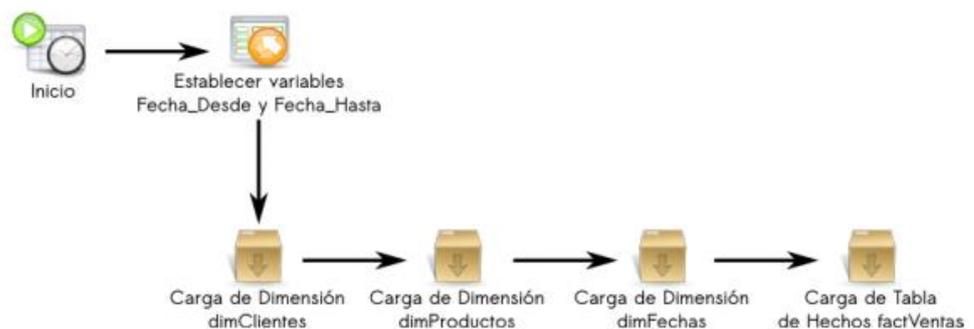


Figura 11: Ejemplo de carga de Tablas dimensionales (Fuente: Dataprix)

### Carga de Dimensión dimClientes

Obtener datos de Datasource: obtiene a través de una consulta SQL los datos del Datasource necesarios para cargar la tabla de Dimensión dimitentes. Se tomará como fuente de entrada la tabla Clientes del Data Source mencionado anteriormente. Se consultó con los usuarios y se averiguó que deseaban tener en cuenta solo aquellos clientes que NO estén eliminados y que tengan su cuenta habilitada.

Es importante destacar que aunque existían numerosos movimientos de clientes que en la actualidad NO poseen su cuenta habilitada o que figuran como eliminados, se decidió NO incluirlos debido a que el énfasis está puesto en analizar los datos a través de aquellos clientes que NO se encuentren en estas condiciones.

A continuación se expondrá la sentencia SQL configurada en este Paso:

```
SELECT
    Clientes.id_Cliente AS idCliente,
    Clientes.Razon_Soc AS cliente
FROM
    Clientes
WHERE
    (Clientes.Eliminado <> 1)
AND (Clientes.Cta_Habilitada <> 0)
ORDER BY
    Clientes.id_Cliente,
    Clientes.Razon_Soc
```

Figura 12: Ejemplo de sentencias SQL utilizadas  
(Fuente: Dataprix)

#### 2.2.4.2.4.2 Actualización

Cuando se haya ejecutado la carga inicial del DW, se deben establecer las políticas y estrategias de actualización periódica.

Entonces, se deben llevar a cabo las siguientes acciones:

- Determinar el proceso de limpieza de datos y calidad de datos, definir los procesos ETL, etc. que deberán realizarse para actualizar los datos del DW.

## **2.2.5 DATAWAREHOUSE**

### **2.2.5.1 DEFINICION**

Ramos (2016, p5) menciona : "Es el conjunto de estrategias, tecnologías y metodologías que nos ayuda a convertir los datos en información de calidad, y dicha información en conocimiento que nos permita una toma de decisiones más acertadas y que nos ayude así a mejorar nuestra competitividad."

Entre otras definiciones tenemos a Cano (2007, p22) menciona que "El objetivo básico de la Business Intelligence es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones".

Se puede decir que business intelligence es un englobado de muchas herramientas, conocimientos y estrategias que son útiles para que una organización pueda responder a los cambios competitivos que se genera en el mercado donde están inmersos, respondiendo con tomas de decisiones más rápidas y certeras.

### **2.2.5.2 ARQUITECTURA**

En las arquitecturas de Inteligencia de negocio existen factores y variantes de las cuales hay se mostraran en el siguiente gráfico y se detallarán a continuación.

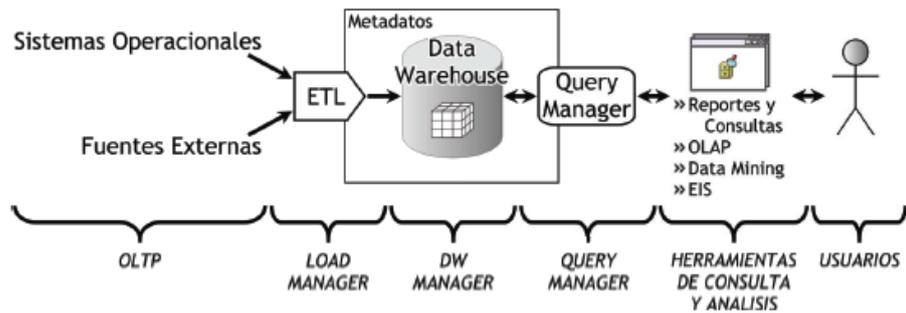


Figura 13: Estructura OLTP. (Fuente:Data Warehousing-Bernabeu)

### 2.2.5.2.1 OLTP (On Line Transaction Processing)

Representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer. Como ya se ha mencionado, estas fuentes de información, son de características muy disímiles entre sí, en formato, procedencia, función, etc.

Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran:

- Archivos de textos.
- Hipertextos.
- Hojas de cálculos.
- Informes semanales, mensuales, anuales, etc.
- Bases de datos transaccionales. (Bernabeu, 2007, p20)

### 2.2.5.2.2 LOAD MANAGER

Para poder extraer los datos desde los OLTP, para luego manipularlos, integrarlos y transformarlos es necesario contar con algún sistema que se encargue de ello. Precisamente los ETL (Extracción, Transformación y Carga) son los que cumplirán con tal fin. Tal y como

sus siglas lo indican, los ETL, extraen datos de las diversas fuentes que se requieran, los transforman para resolver posibles problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado se procede a su carga en el depósito de datos.

En síntesis, las funciones específicas de los ETL son tres:

- Extracción.
- Transformación.
- Carga. (p21)

#### **2.2.5.2.2.1 Extracción**

Es aquí, en donde, basándose en las necesidades y requisitos del usuario, se exploran las diversas fuentes OLTP que se tengan a disposición, y se extrae la información que se considere relevante al caso.

La dificultad para obtener esta información dependerá de cómo se ha manejado la información mencionando un ejemplo se podría decir que más accesible y en menos tiempo se podría extraer los datos de un Sistema gestor de base de datos mediante consultas SQL que los almacenados en texto, hipertextos, hojas de cálculos, etc.

Una vez que los datos son seleccionados y extraídos, se guardan en un almacenamiento intermedio, lo cual permite, entre otras ventajas:

- Manipular los datos sin interrumpir ni paralizar los OLTP, ni tampoco el DW.
- Almacenar y gestionar los metadatos que se generarán en los procesos ETL.
- Facilitar la integración de las diversas fuentes, internas y externas.(p21 -22)

#### **2.2.5.2.2.2 Transformación**

Esta función es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el DW. Estas acciones se llevan a cabo, debido a que pueden existir diferentes fuentes de información, y es vital conciliar un formato y forma única, definiendo estándares, para que todos los datos que ingresarán al DW estén integrados.

Los casos más comunes en los que se deberá realizar integración, son los siguientes:

- Codificación.
- Medida de atributos.
- Convenciones de nombramiento.
- Fuentes múltiples.

Además de lo antes mencionado, esta función se encarga de realizar el proceso de Limpieza de Datos.(p.22)

#### **2.2.5.2.2.3 Carga**

Este proceso es el responsable de cargar la estructura de datos del DW con: Aquellos datos que han sido transformados y que residen en el almacenamiento intermedio.

Aquellos datos de los OLTP que tienen correspondencia directa con el depósito de datos.

Se debe tener en cuenta, que los datos antes de moverse al almacén de datos, deben ser analizados con el propósito de asegurar su calidad, ya que este es un factor clave, que no debe dejarse de lado.(p.25)

#### **2.2.5.2.3 DATAWAREHOUSE MANAGER**

El DW Manager presenta las siguientes características y funciones:

- Transforma e integra los datos fuentes y del almacenamiento intermedio en un modelo adecuado para la toma de decisiones.
- Gestiona el depósito de datos y lo organiza en torno a una base de datos multidimensional.
- Permite realizar todas las funciones de definición y manipulación del depósito de datos, para poder soportar todos los procesos de gestión del mismo.

- Es el encargado de ejecutar y definir las políticas de particionamiento.
- Realiza copias de resguardo incremental o total de los datos.
- Se constituye típicamente al combinar un SGBD con aplicaciones y software propietarios.
- Posee un repositorio de datos propio.
- Gestiona y mantiene metadatos.(p.26-27)

### 2.2.5.2.3.1 Base de datos multidimensional

Base de datos multidimensional proveen una estructura que permite tener acceso flexible a los datos, para explorar y analizar sus relaciones, resultados consiguientes.

Estas se pueden visualizar como un cubo multidimensional, en donde las variables asociadas existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa la medida, indicador o el hecho que se está evaluando.

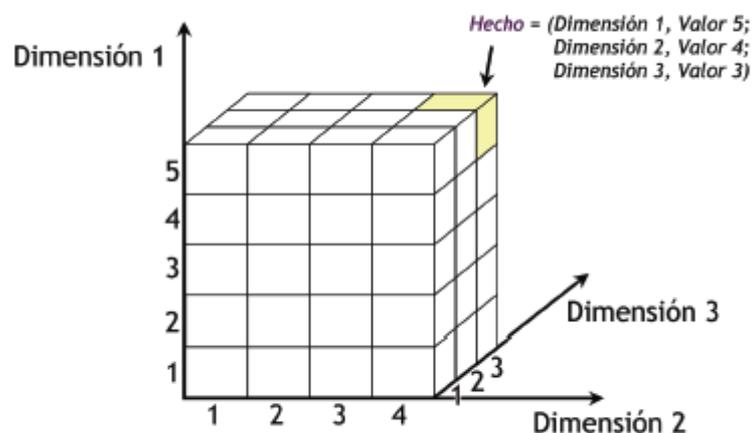


Figura 14: Cubo Multidimensional. (Fuente:Data Warehousing-Bernabeu)

Las bases de datos multidimensionales implican tres variantes posibles de modelamiento, que permiten realizar consultas de soporte de decisión:

- Esquema en estrella<sup>5</sup> (Star Scheme).
- Esquema copo de nieve<sup>6</sup> (Snowflake Scheme).
- Esquema constelación<sup>7</sup> o copo de estrellas (Starflake Scheme).

Los mencionados esquemas pueden ser implementados de diversas maneras, que, independientemente al tipo de arquitectura, requieren que toda la estructura de datos este desnormalizada o semi desnormalizada, para evitar desarrollar uniones (Join) complejas para acceder a la información, con el fin de agilizar la ejecución de consultas. Los diferentes tipos de implementación son los siguientes:

- Relacional – ROLAP<sup>8</sup>.
- Multidimensional – MOLAP<sup>9</sup>.
- Híbrido – HOLAP<sup>10</sup>.(p.27-28)

#### **2.2.5.2.3.2 Tablas de Dimensiones**

Las tablas de dimensiones definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio.

Representan los ejes del cubo, y los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos.

GEOGRAFIA	PRODUCTOS	CLIENTES	FECHAS
id_Geografía País Provincia Ciudad Barrio	id_Producto Rubro Tipo NombreProducto	id_Cliente NombreCliente	id_Fecha Año Trimestre Mes Día

Figura 15: Tablas Dimensionales (Fuente: Dataprix)

Como se puede observar, cada tabla posee un identificador único y al menos un atributo que describe los criterios de análisis relevantes de la organización, estos son por lo general de tipo texto. Usualmente la cantidad de tablas de dimensiones, aplicadas a un tema de interés en particular, varían entre tres y quince. (p.28-29)

### 2.2.5.2.3.3 Tablas de Hechos

Las tablas de hechos contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones.

Los hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones.

El registro del hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este.

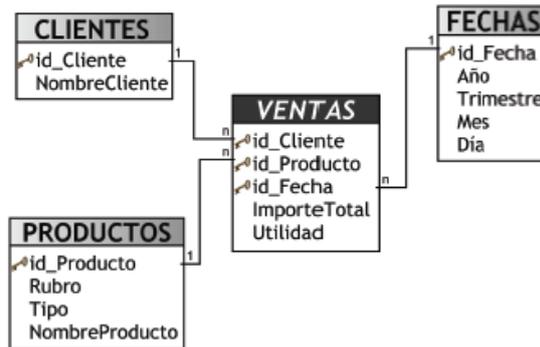


Figura 16: Relación tablas dimensionales y tabla hechos (Fuente: Dataprix)

Los hechos son todas aquellas sumalizaciones o acumulaciones preestablecidas que residen en una tabla de hechos para agilizar las consultas y permitir que los datos puedan ser accedidos por las diferentes dimensiones.

Las sumalizaciones no están referidas solo a sumas, sino que también a promedios, mínimos, máximos, totales por sector, porcentajes, fórmulas predefinidas, etc, dependiendo de los requerimientos de información del negocio, como por ejemplo:

- $\text{ImporteTotal} = \text{SUM}(\text{precioProducto} * \text{cantidadVendida})$
- $\text{Rentabilidad} = \text{SUM}(\text{utilidad} / \text{PN})$
- $\text{PromedioGeneral} = \text{AVG}(\text{notasFinales})$
- $\text{CantidadVentas} = \text{COUNT}(\text{idVenta})$ .(p.31-32)

#### 2.2.5.2.3.4 El esquema en estrella

Consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves.

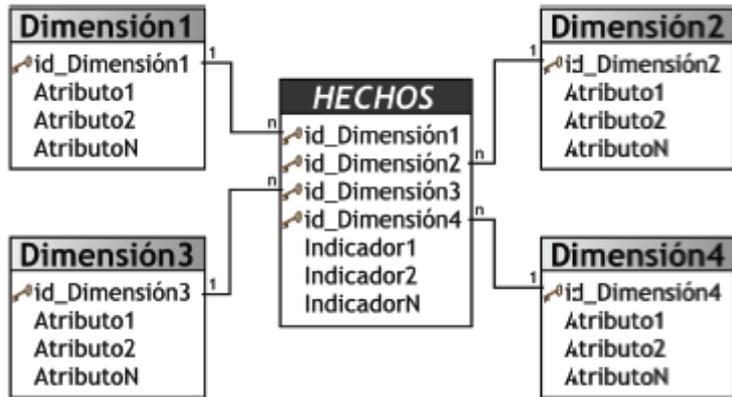


Figura 17: Modelo estrella (Fuente: Dataprix)

El modelo que se muestra a continuación debe estar totalmente desnormalizado, es decir que no puede presentarse en tercera forma normal (3ra FN), es por ello que por ejemplo, la dimensión “PRODUCTOS” contiene los atributos “Rubro”, “Tipo” y “NombreProducto”.

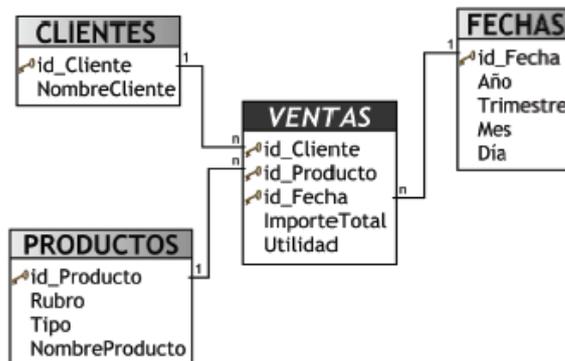


Figura 18: Ejemplo modelo estrella (Fuente: Dataprix)

Las ventajas que trae aparejada la desnormalización, son las de obviar uniones (Join) entre las tablas cuando se realizan consultas, procurando así un mejor tiempo de respuesta y una mayor sencillez con respecto a su utilización. El punto en contra, es que

se genera un cierto grado de redundancia, pero el ahorro de espacio no es significativo.

A continuación se destacarán algunas características de este modelo, que ayudarán a comprender mejor el porqué de sus ventajas:

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

#### 2.2.5.2.3.5 Esquema Copo de Nieve

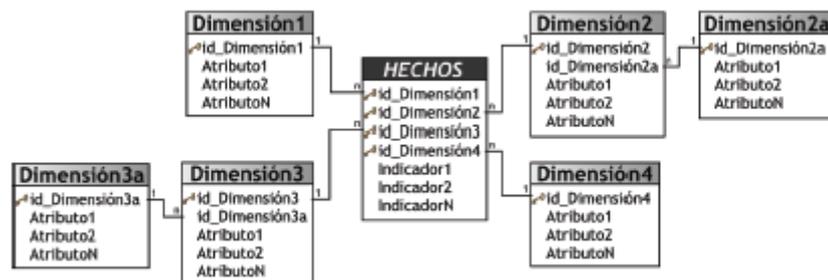


Figura 19: Modelo copo de nieve (Fuente: Dataprix)

Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas.

Una de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de las dimensiones y proveer un esquema que sustente los requerimientos de diseño. Otra razón es que es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

#### **2.2.5.2.3.6 ROLAP**

Este tipo de organización física se implementa sobre tecnología relacional, pero disponen de algunas facilidades para mejorar el rendimiento.

Es decir, ROLAP (Relational On Line Analytic Processing) cuenta con todos los beneficios de una SGBD Relacional a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor de DW.

Entre las características más importantes y sobresalientes de ROLAP, se encuentran las siguientes:

- Almacena la información en una base de datos relacional.
- Posee tres capas lógicas: de almacenamiento, de análisis y de presentación.
- Utiliza índices de mapas de bits.
- Utiliza índices de Join.
- Posee técnicas de particionamiento de datos.

- Posee optimizadores de consultas.
- Cuenta con extensiones del SQL (drill-up, drill-down, etc).

Como se aclaró anteriormente, el almacén de datos se organiza a través de una base de datos multidimensional, sin embargo, puede ser soportado por un SGBD Relacional.

Para lograr esto se utilizan los diferentes esquemas, en estrella, copo de nieve y constelación, los cuales transformarán el modelo multidimensional y permitirán que pueda ser gestionado por un SGBD Relacional, ya que solo se almacenarán tablas.

#### **2.2.5.2.3.7 MOLAP**

El objetivo de los sistemas MOLAP (Multidimensional On Line Analytic Processing) es almacenar físicamente los datos en estructuras multidimensionales de manera que la representación externa y la interna coincidan. Para ello, se dispone de estructuras de almacenamiento específicas (Arrays) y técnicas de compactación de datos que favorecen el rendimiento del depósito de datos.

- Las principales características de MOLAP son:
- Posee tecnología optimizada para consultas y análisis, basada en el modelo multidimensional.
- Cuenta con un motor especializado.

- Provee herramientas limitadas y propietarias.
- No es adecuada para muchas dimensiones.
- Construye y almacena datos en estructuras multidimensionales.

Las herramientas de consulta y exploración OLAP<sup>13</sup>, presentan estas estructuras multidimensionales.

#### **2.2.5.2.3.8 HOLAP**

HOLAP (Hybrid On Line Analytic Processing) constituye un sistema híbrido entre MOLAP y ROLAP, que combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional.

#### **2.2.5.2.4 QUERY MANAGER**

Este componente realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, tales como Join y agregaciones, y de consultas propias del análisis de datos, como drill-up y drill-down. Query Manager recibe las consultas del usuario, las aplica a las tablas correspondientes y devuelve los resultados obtenidos.

Cabe aclarar que una consulta a un DW, generalmente consiste en la obtención de medidas o indicadores a partir de los datos de una

tabla de hechos, restringida por las propiedades o condiciones de las dimensiones seleccionadas.

Las operaciones que se pueden realizar sobre modelos multidimensionales y que son las que verdaderamente les permitirán a los usuarios explorar y desmenuzar los datos en busca de respuestas, son:

- Drill-down.
- Drill-up.
- Drill-across.
- Roll-across.
- Pivot.
- Page.

#### **2.2.5.2.5 HERRAMIENTAS DE CONSULTA Y ANÁLISIS**

Las herramientas de consulta y análisis son sistemas que permiten al usuario realizar la exploración de datos del DW. Básicamente constituyen el nexo entre el depósito de datos y los usuarios.

A través de una amigable interfaz gráfica y una serie de simples pasos, el usuario genera consultas que son enviadas desde la herramienta de consulta y análisis al Query Manager, este a su vez realiza la extracción de información al DW Manager y devuelve los resultados obtenidos a la herramienta que se los solicitó. Luego, estos resultados son expuestos ante el usuario en formatos que le son familiares.

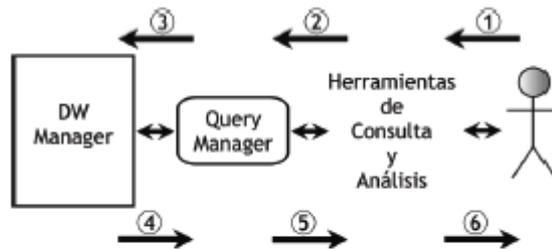


Figura 20: Proceso de consulta y análisis  
Fuente: Data Warehousing-Bernabeu)

El mismo, se lleva a cabo a través de seis pasos sucesivos:

1. El usuario selecciona o establece que datos desea obtener del DW, mediante las interfaces de la herramienta que utilice.
2. La herramienta recibe el pedido del usuario, construye la consulta y la envía al Query Manager.
3. El Query Manager ejecuta la consulta en las tablas del DW.
4. El Query Manager obtiene los resultados de la consulta.
5. El Query Manager envía los datos a las herramientas de consulta y análisis.
6. Las herramientas presentan al usuario la información requerida.

Existen diferentes tipos de herramientas de consulta y análisis, y de acuerdo a la necesidad, tipos de usuarios y requerimientos del negocio, se deberán seleccionar las más propicias al caso. Entre ellas se destacan las siguientes:

- Reportes y Consultas
- OLAP
- Data Mining

- EIS

#### **2.2.5.2.5.1 Reportes y Consultas**

Se han desarrollado varias herramientas para la producción de consultas y reportes, que ofrecen a los usuarios, a través de pantallas gráficas intuitivas, la posibilidad de generar informes avanzados y detallados del área de interés del negocio que se esté analizando. El usuario solo debe seguir una serie de simples pasos, como por ejemplo seleccionar opciones de un menú, presionar tal o cual botón para especificar los elementos de datos, sus condiciones, criterios de agrupación y demás atributos que se consideren significativos.

#### **2.2.5.2.5.2 OLAP**

El procesamiento analítico en línea OLAP (On Line Analytic Processing), es la componente más poderosa de los DW, ya que es el motor de consultas especializado de la bodega.

Las herramientas OLAP, son una tecnología de software para análisis en línea, administración y ejecución de consultas, que permiten inferir información del comportamiento del negocio.

Su principal objetivo es el de brindar rápidas respuestas a complejas preguntas, para interpretar la situación del negocio y tomar decisiones. Cabe destacar que lo que es realmente interesante en OLAP, no es la ejecución de simples consultas

tradicionales, sino la posibilidad de utilizar operadores tales como drill-up, drill-down, etc, para explotar profundamente la información.

#### **2.2.3.2.5.3 Data Mining**

Esta herramienta constituye una poderosa tecnología con un gran potencial que ayuda y brinda soporte a los usuarios, con el fin de permitirles analizar y extraer conocimientos ocultos y predecibles a partir de los datos almacenados en un DW o en un OLTP. Claro que es deseable que la fuente de información sea un DW, por todas las ventajas que aporta.

La integración con el depósito de datos facilita que las decisiones operacionales sean implementadas directamente y monitoreadas. A medida que el DW crece con nuevos requerimientos y resultados, la organización puede minar las mejores prácticas y aplicarlas en futuras decisiones.

#### **2.2.5.2.6 EIS**

EIS (Executive Information System) proporciona medios sencillos para consultar, analizar y acceder a la información de estado del negocio. Además, pone a disposición facilidades para que el usuario pueda conseguir los datos buscados rápidamente, empleando el menor tiempo posible para comprender el uso de la herramienta.

## **2.2.6 DATAMART**

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

Se podría decir que un Datamart es un pequeño datawarehouse para una cantidad específica de usuarios ya que su alcance abarca solamente un área o departamento en específico, incluyendo las variables e indicadores para el proceso de toma de decisiones.

### **2.2.6.1 Características**

- Son pobladas por usuarios finales.
- Se optimizan en función a procesos transaccionales.
- Se actualizan constantemente.
- Contienen mucha información de detalle.

### **2.2.6.2 Componentes de un Datamart**

- **Fuentes de Datos**

Son las que alimentan de información al DataMart, están diseñadas para registrar grandes cantidades de transacciones. Entre ella tenemos la

base de datos OLTP (Una base de datos para soportar procesos transaccionales).

- **Procesos de extracción, transformación y carga de datos (ETL)**

Los datos se encuentran almacenados en base de datos destinados al registro de transacciones. Es necesario extraer y transformar los datos antes de cargar los resultados en el DataMart.

- **DataWarehouse**

Un DataMart es una solución que, compartiendo tecnología con el DataWarehouse (pero con contenidos específicos, volumen de datos más limitado y un alcance histórico menor), permita dar soporte a una empresa pequeña, un departamento o área de negocio de una empresa grande.

- **Herramientas de Explotación**

Las herramientas de consulta al igual que la mayoría de herramientas visuales, permiten apuntar y dar un click a los menús y botones para especificar los elementos de datos, condiciones, criterios de agrupación y otros atributos de una solicitud de información.

- **Herramientas de base de datos multidimensionales / OLAP**

Las primeras soluciones OLAP (On Line Analytical Processing), estuvieron basadas en bases de datos multidimensionales (MDDBS). Un cubo estructural (dos veces un hipercubo o un arreglo multidimensional) almacenaba los datos para que se puedan manipular

intuitivamente y claramente ver las asociaciones a través de dimensiones múltiples

- **Sistemas de información ejecutivos**

Las herramientas de sistemas de información ejecutivos (Executive Information Systems - EIS), proporcionan medios sumamente fáciles de usar para consulta y análisis de la información confiable. Generalmente se diseñan para el usuario que necesita conseguir los datos rápidamente, pero quiere utilizar el menor tiempo posible para comprender el uso de la herramienta.

- **Herramientas de Data Mining**

Data Mining es una categoría de herramientas de análisis open-end. En lugar de hacer preguntas, se toma estas herramientas y se pregunta algo "interesante", una tendencia o una agrupación peculiar, por ejemplo. El proceso de Data Mining extrae los conocimientos guardados o información predictiva desde el DataMart sin requerir pedidos o preguntas específicas. Las herramientas Mining usan algunas de las técnicas de computación más avanzadas para generar modelos y asociaciones como redes neuronales, detección de desviación, modelamiento predictivo y programación genética

## 2.2.7 TOMA DE DECISIÓN

“La toma de decisiones es el proceso de identificación de un problema u oportunidad y la selección de una alternativa de acción entre varias existentes, es una actividad diligente clave en todo tipo de organización.”

Para B. Emily (2005), citado por Toainga M.(2014, p6)

### 2.2.7.1 Proceso de toma de decisiones.

Debido a la inestabilidad de los ambientes y el riesgo de actuar sin contar con los elementos necesarios, los directivos de las organizaciones necesitan manipular grandes volúmenes de datos, convertirlos en información y generar el conocimiento necesario para cumplir con los objetivos propuestos.

Este proceso debe abarcar las cuatro funciones administrativas fundamentales: planeamiento, organización, conducción y control.

**La Planeación:** Selección de misiones y objetivos así como de las acciones para cumplirlas. Esto implica “Toma de decisión”. ¿Cuáles son los objetivos de la organización, a largo plazo? ¿Qué estrategias son mejores para lograr este objetivo?

**Organización:** Establecimiento de la estructura que desempeñan los individuos dentro de la organización.

¿Cómo deben diseñarse los puestos?

¿Cuándo debe una organización instrumentar una estructura diferente?

**Dirección:** Esta función requiere que los administradores influyan en los individuos para el cumplimiento de las metas organizacionales y grupales. ¿Cómo manejo a un grupo de trabajadores que parecen tener una motivación baja? ¿Cuál es el estilo de liderazgo más eficaz para una situación dada?

**Control:** Es la medición y corrección del desempeño individual y organizacional de manera tal que se puedan lograr los planes. ¿Qué actividades en la organización necesitan ser controladas? ¿Cómo deben controlarse estas actividades?. Taoinga (2014,p7)

#### **2.2.7.2 Etapas de la Toma de Decisiones.**

El proceso de toma de decisiones, incluye una serie de ocho etapas que comienza por identificar el problema y los criterios de decisiones y por ponderados.

##### **Etapas 1. Identificar un problema.**

Se empieza con un problema, la identificación será de acuerdo al observador ya que dicha acción es subjetiva. Lo que el tomador de decisiones ve como un problema para otro no será un problema; y es aquí don una mala decisión puede traer efectos muy negativos a la organización.

## **Etapa 2. Identificar los criterios de decisión.**

Cuando el tomador de decisiones descubre la problemática, tiene que saber cuáles son los criterios de toma de decisiones más oportunas que hay que tener en cuenta para solucionar dicha dificultad.

## **Etapa 3. Asignar pesos a los criterios.**

Si el tomador de decisiones estableció unos criterios en la segunda etapa, tiene que saber priorizarlas en orden de importancias.

## **Etapa 4. Desarrollar las alternativas.**

En esta etapa, el tomador de decisiones se realizara una lista de las alternativas posibles que solucionen el problema. No se valoran las opciones, solo se muestran enumeradas.

## **Etapa 5. Analizar las alternativas.**

Después de enumerar las alternativas, el tomador de decisiones tiene que examinar sin saltarse las opciones, se calcula dependiendo de las alternativas creadas en las etapas 1 y 2, con esta acción de comparar se identifican las ventajas y desventajas de cada alternativa.

## **Etapa 6. Seleccionar una alternativa**

“Esta etapa radica en optar la mejor alternativa entre todas las seleccionadas, después de a ver priorizado a cada uno de las opciones de las decisiones y de examinar todas las alternativas posibles, se elige a la que generó mayor peso o importancia.

### **Etapas 7. Implementar la alternativa.**

En esta etapa se ejecuta las decisiones, es donde se da la información a los involucrados y tener su aceptación con ella, los usuarios que implementan las decisiones en la causa, lo realizan con la expectativa que los resultados sean los esperados.

### **Etapas 8. Evaluar la eficacia de la decisión.**

En esta etapa que es la última lo que se quiere es valorar los resultados de las decisiones, se quiere saber si se solucionó la problemática. Se realizaran las preguntas ¿Se logró lo que se buscaba dar solución con la alternativa seleccionada en la etapa 6 y puesta en funcionamiento en la etapa 7”.

## **2.3 DEFINICIÓN DE TÉRMINOS BÁSICOS**

- Metadatos:

La definición más concreta de los metadatos es “datos acerca de los datos” y sirven para suministrar información sobre los datos producidos. Los metadatos consisten en información que caracteriza datos, describen el contenido, calidad, condiciones, historia, disponibilidad y otras características de los datos. Los Metadatos permiten a una persona ubicar y entender los datos, incluyen información requerida para determinar qué conjuntos de datos existen para una localización geográfica particular, la información necesaria para determinar si un conjunto de datos es apropiado para fines específicos, la información requerida para recuperar o conseguir un conjunto ya identificado de datos y la información requerida para procesarlos y utilizarlos. GeolDEP (2014).

- Hipertexto:

En informática, es el nombre que recibe el texto que en la pantalla de un dispositivo electrónico conduce a otro texto relacionado. La forma más habitual de hipertexto en informática es la de hipervínculos o referencias cruzadas automáticas que van a otros documentos (lexias). Si el usuario selecciona un hipervínculo el programa muestra el documento enlazado. EcuRed (2018)

- Datamart:

Es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información. Sinnexus(2016)

- Drill-down:

Es una de las herramientas más significativas y útiles del sistema, ya que permite al usuario desglosar cualquier dato de un informe hasta el comprobante original, para conocer el detalle de cómo fue calculado. Usado comúnmente en los sistemas de información gerencial o de análisis de información, es la habilidad para poder navegar de lo general a lo particular en la información presentada. Para obtener esta característica en los sistemas de información se requiere un alto grado de integración de la información, además de la herramienta que permita hacer el drill down. Stephany(2012).

- Drill-up:

Es una operación que permite apreciar los datos con un menor nivel de detalle, subiendo por una Jerarquía definida en un Cubo. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis.

- Roll-across:

Es una operación que trabaja de forma similar a drill-up, con la diferencia de que rollacross NO se hace sobre una Jerarquía, sino que su forma de ir de lo específico a lo general es quitar un Atributo de la consulta, eliminando de esta manera un criterio de análisis.

- Pivot:

Es una operación que permite seleccionar el orden de visualización de los Atributos e Indicadores, con el objetivo de analizar la información desde diferentes puntos de vista.

- Page:

Es una operación que presenta el Cubo dividido en secciones, a través de los valores de un Atributo, como si se tratase de páginas de un libro. Page es muy útil cuando las consultas devuelven muchos registros y es necesario desplazarse por los datos para poder verlos en su totalidad.

- Normalizar:

Modelo Lógico es el proceso mediante el cual se transforman datos complejos a un conjunto de estructuras de datos más pequeñas, que además de ser más simples y más estables, son más fáciles de mantener. También se puede entender la normalización como una serie de reglas que sirven para ayudar a los diseñadores de bases de datos a desarrollar un esquema que minimice los problemas de lógica. Cada regla está basada en la que le antecede. La normalización se adoptó porque el viejo estilo de poner todos los datos en un solo lugar, como un archivo o una tabla de la base de datos, era ineficiente y conducía a errores de lógica cuando se trataban de manipular los datos. MysqlHispano (2003).

- SGBD:

Es un programa orientado a la gestión y diseño de bases de datos, permitiendo su creación, modificación, atributos e interfaz. Además de permitir el desarrollo y construcción de las bases de datos, un SGBD permite operar directamente en las tablas de la base de datos, haciendo posible la navegación y visualización de los registros almacenados en las tablas de la misma, su edición, búsqueda, inserción y eliminación. En sí mismo el SGBD no sólo actúa como una herramienta de construcción de BD sino como un interfaz que permite interactuar y explotar sus contenidos. Para ello un SGBD está compuesto por diversos subsistemas a saber; el motor de la base de datos, el sistema de definición de datos, el sistema de manipulación y gestión, el sistema de herramientas y aplicaciones y el módulo de administración. Manuel (2014)

- Jerarquías de Dimensiones:

Es un constructor de datos utilizado para agrupar datos en intervalos basados en valores. Una jerarquía de dimensiones puede contener varios niveles, cada uno de los cuales con su propio conjunto de grupos. Los grupos de cada nivel inferior se deben acumular en grupos en niveles superiores. IBM (2018).

- ETL

Son las siglas en inglés de Extraer, Transformar y Cargar (Extract, Transform and Load). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio. Espinosa (2010).

- OLTP:

(Procesamiento de Transacciones En Línea) es una clase de programa que facilita y gestiona aplicaciones orientadas a transacciones, típicamente para ingreso de datos y retención de transacciones en varias industrias, incluyendo banca, aerolíneas, correo, supermercados y productores. Probablemente el producto OLTP instalado más distribuido es el CICS (Sistema de Control de Información de Clientes) de IBM.

## **CAPÍTULO III: DESARROLLO DEL OBJETIVO DE TRABAJO DE SUFICIENCIA**

### **3.1 ANALISIS DE REQUERIMIENTO**

#### **3.2.1 Preguntas de Negocio**

Se ha realizado entrevistas a dos representantes importantes para la toma de decisión del proceso de ventas de la empresa (Ver encuesta en Anexo1) donde se obtuvo la información de los requerimientos necesarios para el análisis y la identificación de las variables y perspectivas que serán necesarios para poder realizar el Datamart, lo que se pudo obtener fue lo siguiente.

#### Requerimientos obtenidos:

- La cantidad de prendas vendidas a cada cliente en un tiempo determinado.
- El monto total de ventas de las prendas en un tiempo determinado.
- Comparativo de monto de venta por tiempo determinado.
- El monto total de ventas por empleado en un tiempo determinado.
- El monto total de ventas por tipo de pago en un tiempo determinado.

- La cantidad de prendas vendidas en un tiempo determinado.
- La cantidad de prendas vendidas en un local en un tiempo determinado

### 3.1.2 indicadores y Perspectivas

Para este paso, ya obtenida la información se realiza la identificación de los indicadores y perspectivas, donde los indicadores es lo que se quiere medir y las perspectivas son las entidades mediante el cual se va obtener información, para el caso de base de datos seria las tablas.

- Las cantidades vendidas de prendas a cada cliente en un tiempo determinado.

**Indicadores**

**Perspectivas**

- El monto total de ventas, de las prendas en un tiempo determinado.

**Indicadores**

**Perspectivas**

- Comparativo de monto total de venta, en un tiempo determinado

**Indicadores**

**Perspectivas**

- El monto total de ventas por empleado en un tiempo determinado.

**Indicadores**

**Perspectivas**

- Las cantidades vendidas de prendas en un tiempo determinado.

**Indicadores**

**Perspectivas**

- Las cantidades vendidas de prendas en un local en un tiempo determinado.

**Indicadores**

**Perspectivas**

- El monto total de ventas por tipo de pago en un tiempo determinado
- Indicadores**
**Perspectivas**

Los Indicadores son:

- Cantidades Vendidas
- Monto total de ventas

Y las Perspectivas de análisis son:

- Prendas
- Cliente
- Tiempo
- Local
- Empleados
- TipoPagos

### 3.1.3 Modelo Conceptual

De los indicadores y perspectivas encontrados, se realizó un esquema conceptual colocando al lado izquierdo todas las perspectivas, al lado derecho los indicadores y en el centro el proceso que se trabajó, este modelo sirve para poder dar una mejor explicación y facilidad de entendimiento a los usuarios interesados.

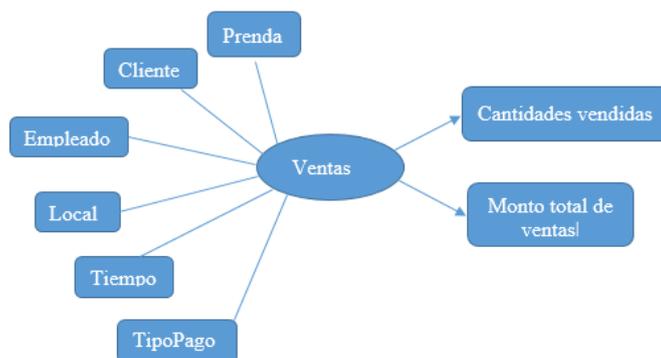


Figura 21: Implementación modelo conceptual (Fuente Propia)

## 3.2 ANALISIS DE DATA SOURCES

### 3.2.1 Hechos e indicadores

Se definirán las fórmulas que se deberá efectuar para el cálculo cada uno de los indicadores. A continuación se muestran los indicadores con su respectiva aclaración de porque se eligió dicha fórmula o función.

**Indicador:** Cantidades Vendidas:

- Hechos: Cantidades Vendidas
- Función de agregación: SUM

**Aclaración:** el Indicador Cantidades Vendidas representa la sumatoria de las unidades que se han vendido de una prenda en particular.

**Indicador:** Monto Total de Ventas

- Hechos:  $(\text{Cantidades Vendidas}) * (\text{Precio de Venta})$
- Función de agregación: SUM

**Aclaración:** el Indicador Monto Total de Ventas representa la sumatoria del monto total que se ha vendido de cada prenda, y se obtiene al multiplicar las Cantidades vendidas, por su respectivo precio de venta.

### 3.2.2 Mapeo

Se muestra la base de datos de la empresa, donde el proceso de ventas está representada por el siguiente Diagrama de Entidad Relación.

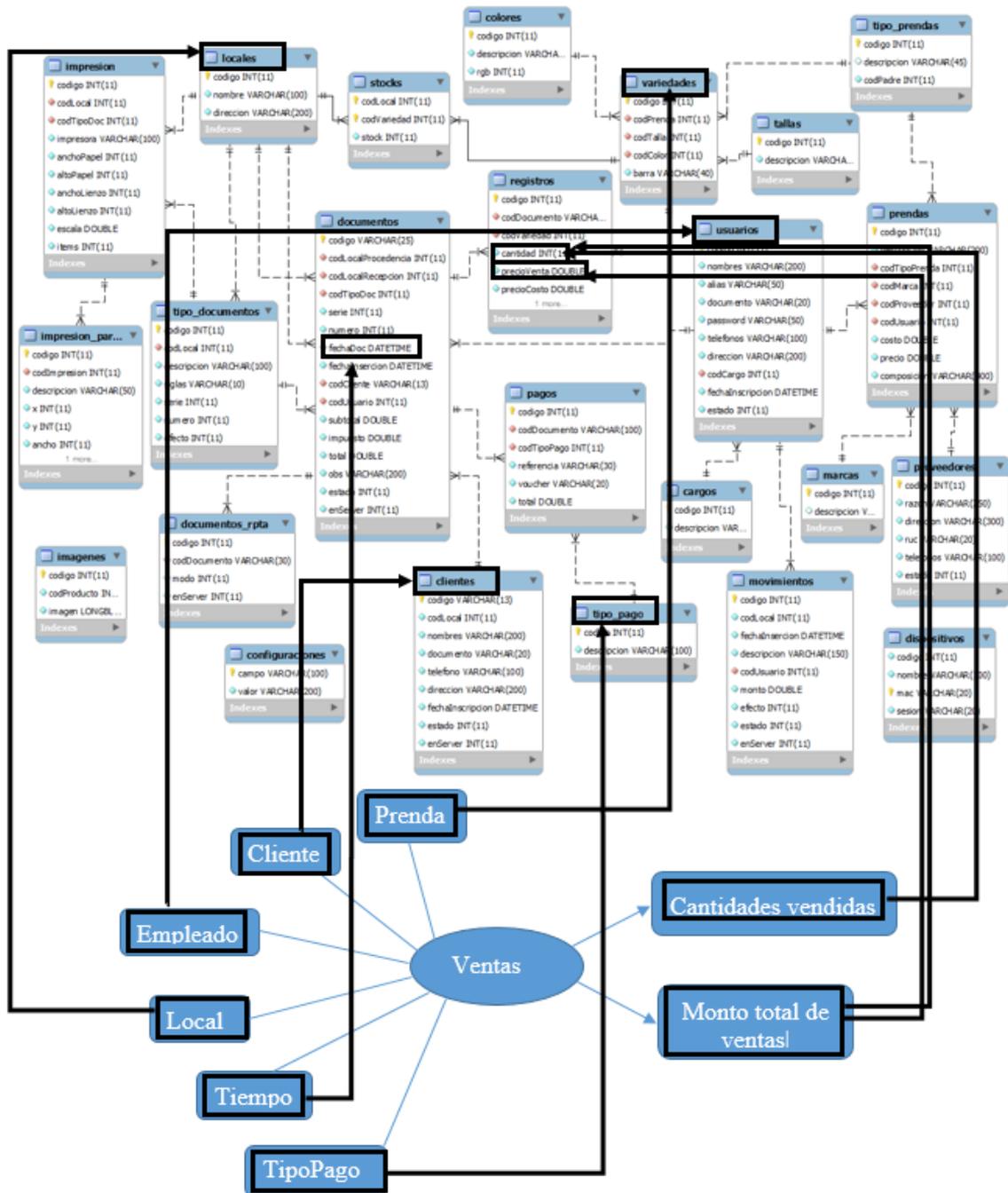


Figura 22: Implementación del Mapeo (Fuente Propia)

El mapeo realizado se pudo identificar lo siguiente:

- La perspectiva prenda se relaciona con la tabla variedades.
- La perspectiva cliente se relaciona con la tabla clientes.
- La perspectiva empleado se relaciona con la tabla usuarios.
- La perspectiva local se relaciona con la tabla locales.
- La perspectiva pago se relaciona con la tabla pagos.
- La perspectiva Tiempo se relaciona con el campo fechaDoc de la tabla Documentos debido a que es la fecha principal del proceso ventas.

El indicador Cantidades de ventas se relaciona con el campo cantidad de la tabla registros, teniendo como fórmula de cálculo lo siguiente:

- $\text{Sum}(\text{cantidad})$ .

El indicador Monto de ventas se relaciona con los campos cantidad y precioVenta de la tabla registros, teniendo como fórmula de cálculo lo siguiente:

- $\text{Sum}(\text{cantidad} * \text{precioVenta})$ .

### **3.2.3 Granularidad**

Se analizaron los datos de los atributos de las tablas que tiene relación con las perspectivas encontradas, primero se detallaron cada atributo para que los usuarios interesados puedan identificar y decidir cuáles son los indicados para poder hacer la consulta de los indicadores. Para el caso de la perspectiva tiempo se deberá analizar en qué periodos se requerirá ver la información.

A continuación se mostrara los detalles de los atributos de las tablas a analizar:

Con respecto a la Perspectiva **Prenda**, los datos disponibles son los siguientes:

- codigo: Es la clave primaria de la tabla variedades y es auto incrementables.
- codPrenda: Es la clave foránea que representa el modelo de una prenda.
- codTalla: Es la clave foránea que representa la talla de una prenda.
- codColor: Es la clave foránea que representa color de una prenda.
- barra: Es el código completo de una prenda especifica (modelo+color+talla).

Con respecto a la Perspectiva **Cliente**, los datos disponibles son los siguientes:

- codigo: Es la clave primaria e identificador único de un cliente.
- codLocal: Es la clave foránea que representa el local donde el cliente fue registrado.
- nombres: Representa el nombre completo del cliente.
- documento: Es el DNI o ruc de un cliente.
- teléfono: Es el teléfono de contacto del cliente.
- dirección: Representa la dirección de donde viene el cliente.
- fechaInscripcion: Representa la fecha en el que fue registrado un cliente.

- estado: Número que representa si el cliente está activo en el sistema.
- enServer: Número que representa si el registro ha subido al servidor.

Con respecto a la Perspectiva **Empleado**, los datos disponibles son los siguientes:

- codigo: Es la clave primaria de la tabla usuarios, es auto incrementable.
- nombres: Representa el nombre completo de un usuario.
- alias: Representa el apodo o alias del usuario.
- documento: Es el DNI del usuario.
- password: Es la contraseña de acceso del usuario.
- teléfonos: Es el teléfono de contacto con el usuario.
- dirección: representa la dirección de donde viene el usuario.
- codCargo: Es clave foránea que representa el cargo del usuario.
- fechaInscripcion: Representa la fecha en el que fue registrado el usuario.
- estado: número que representa si el cliente está activo en el sistema.

Con respecto a la Perspectiva **local**, los datos disponibles son los siguientes:

- codigo: Es la clave primaria de la tabla locales y es autoincrementable.
- nombre: Es el nombre del local.
- direccion: Es la dirección física del local.

Con respecto a la Perspectiva **tipoPago**, los datos disponibles son los siguientes:

- código: Es la clave primaria de la tabla tipo\_pago y es autoincrementable.
- descripción: Es el nombre de las formas de pagos que puede realizar.

Con respecto a la Perspectiva **tiempo**, los datos disponibles son los siguientes:

- Año
- Semestre
- Cuatrimestre
- Trimestre
- Número de mes
- Nombre del mes
- Quincena
- Semana
- Número de día
- Nombre del día

Ya habiendo presentado la información de los detalles de los campos de las perspectivas que se van a utilizar se proseguirá a seleccionar los campos que serán de interés para el análisis de los datos y se realizara la des normalización de las tablas.

### **Perspectiva Prenda**

- codigo: Hace referencia a la clave primaria de la tabla variedades.
- codPrenda: Hace referencia al codigo del modelo de prenda de la tabla variedades.
- codTipoPrenda: Hace referencia al tipo de prenda de la tabla prendas.
- codTalla: Hace referencia a la talla de la prenda de la tabla tallas.
- codColor: Hace referencia al color de la prenda de la tabla colores.
- barra: Hace referencia al código de barra de una prenda de la tabla variedades.

### **Perspectiva Cliente**

- codigo: Hace referencia a la clave primaria de la tabla clientes.
- codLocal: hace referencia al local donde fue registrado el cliente.
- documento: Hace referencia al DNI o ruc del cliente.

### **Perspectiva Empleado**

- codigo: Hace referencia a la clave primaria de la tabla usuarios.
- alias: Hace referencia al alias que tiene cada usuario.
- documento: Hace referencia al DNI del cliente.
- codCargo: Hace referencia al codigo del cargo que tiene un usuario.

### **Perspectiva Local**

- codLocal: Hace referencia a la clave primaria de la tabla locales.
- descripcion: Hace referencia al nombre de un local.

### **Perspectiva TipoPago**

- codPago: Hace referencia a la clave primaria de la tabla tipo\_pago.
- descripcion: Hace referencia al nombre del tipo de pago.

### **Perspectiva Tiempo**

- Año
- Semestre
- Trimestre
- Número de mes
- Nombre del mes
- Quincena
- Semana
- Número de día
- Nombre del día

### **3.2.4 Modelo conceptual ampliado**

Luego de haber culminado varios pasos se realizara un modelo conceptual más detallado, donde se podrá identificar fácilmente a grandes razgos los atributos seleccionados para las consultas indicadas de los indicadores asi como las funciones que realizaran cada indicador.



Figura 23: Implementación modelo conceptual ampliado  
(Fuente Propia)

### 3.3 MODELO LÓGICO DEL DW

#### 3.3.1 Tipología

Para este proyecto se utilizara el Esquema en estrella por es más simple de implementar y está acorde con los requerimientos identificados.

#### 3.3.2 Tablas de Dimensiones

Se realizaron las tablas dimensiones en base a las perspectivas de modelo conceptual ampliado, en este paso es posible cambiar los nombres de los atributos si no son lo suficientemente entendibles.

## Perspectiva Cliente

- La nueva tabla de Dimensión tendrá el nombre dimcliente.
- Se le agregará una clave principal con el nombre idCliente.
- Se modificará el nombre del campo codigo por codCliente.
- Se modificará el nombre del campo codLocal por Local.

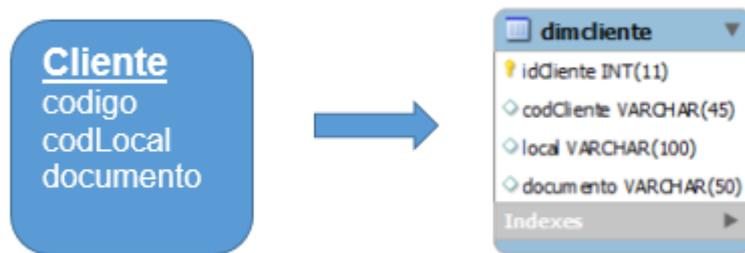


Figura 24: Tabla dimCliente (Fuente Propia)

## Perspectiva Empleado

- La nueva tabla de Dimensión tendrá el nombre dimempleado.
- Se le agregará una clave principal con el nombre dimempleado
- Se modificará el nombre del campo codigo por codEmpleado.
- Se modificará el nombre del campo codCargo por cargo.

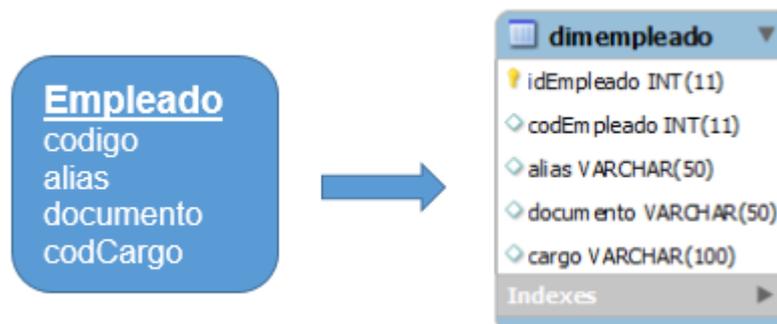


Figura 25 Tabla dimEmpleado (Fuente Propia)

## Perspectiva Local

- La nueva tabla de Dimensión tendrá el nombre dimlocal.
- Se le agregará una clave principal con el nombre idlocal.
- Se modificará el nombre del campo codigo por codLocal.
- Se modificará el nombre del campo descripcion por nombre.



Figura 26 Tabla dimLocal (Fuente Propia)

## Perspectiva Prenda

- La nueva tabla de Dimensión tendrá el nombre dimprenda.
- Se le agregará una clave principal con el nombre idPrenda.
- Se modificará el nombre del campo codigo por codPrenda.
- Se modificará el nombre del campo codPrenda por modelo.
- Se modificará el nombre del campo codTipoPrenda por tipoPrenda.
- Se modificará el nombre del campo codTalla por talla.
- Se modificará el nombre del campo codColor por color.

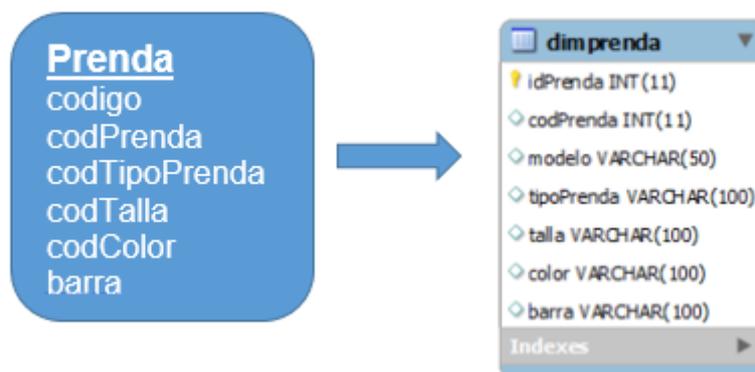


Figura 27: Tabla dimPrenda (Fuente Propia)

## Perspectiva TipoPago

- La nueva tabla de Dimensión tendrá el nombre dimtipopago.
- Se le agregará una clave principal con el nombre idtipopago.
- Se modificará el nombre del campo codigo por codTipoPago.
- Se modificará el nombre del campo descripcion por nombre.

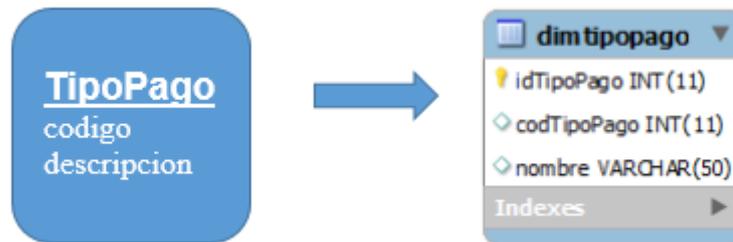


Figura 28: Tabla dimTipoPago (Fuente Propia)

## Perspectiva Tiempo

- La nueva tabla de Dimensión tendrá el nombre dimtiempo.
- Se le agregará una clave principal con el nombre idtiempo.
- Se agregara un campo con el nombre fechaDoc.

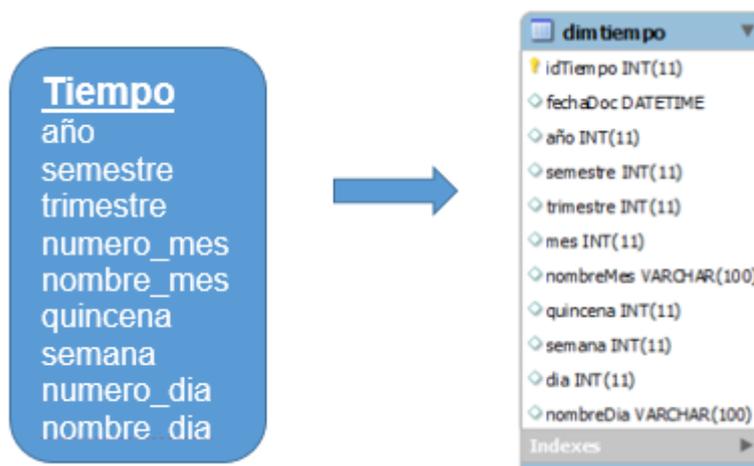


Figura 29: Tabla dimTiempo (Fuente Propia)

### 3.3.3 Tablas de Hechos

Una vez que se realizaron las tablas Dimensiones se pudo diseñar la tabla de hechos donde se definirán las claves primarias en relación con las tablas dimensionales y se agregaran los campos de hechos tantos como indicadores se hayan definido.

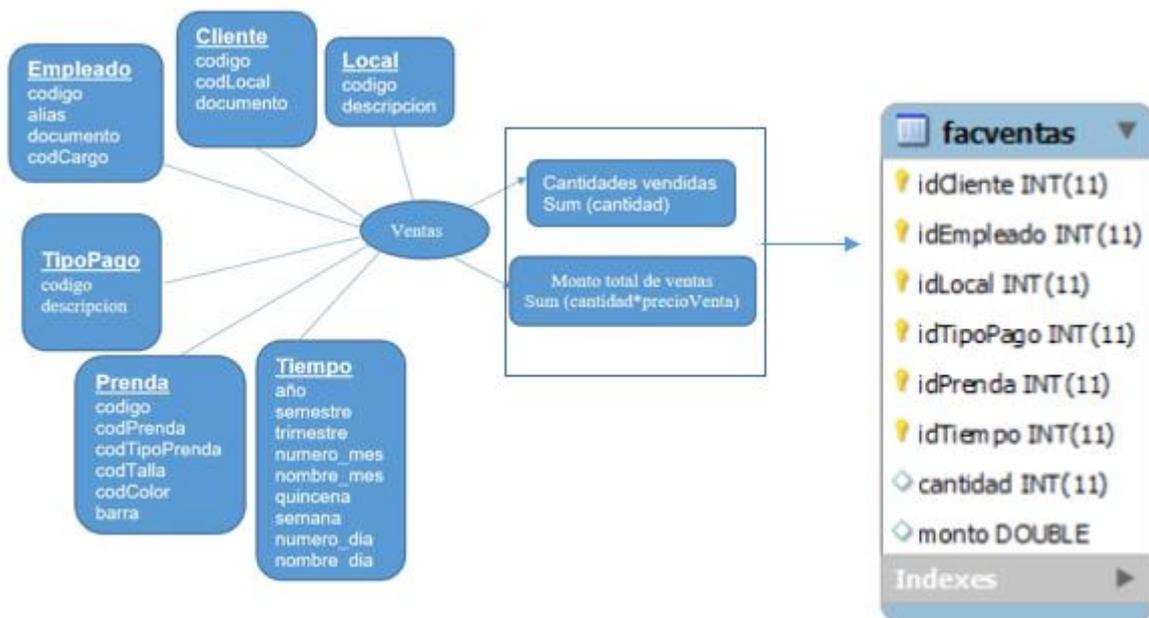


Figura 30: Implementación tabla de hechos (Fuente Propia)

### 3.3.4 Uniones

Luego del diseño de las tablas dimensionales y la tabla de hechos se realizó la unión de todas estas mediante la topología de estrella ya antes mencionado.

- La tabla de Hechos tendrá el nombre factVentas.
- Su clave principal será la combinación de las claves principales de las tablas de dimensiones antes definidas: idCliente, idEmpleado, idPrenda, idTipoPago, idLocal, idTiempo.

➤ Se crearán dos Hechos, que se corresponden con los dos Indicadores:

- Cantidades Vendidas será renombrado como cantidad y
- Monto Total de Ventas será renombrado como monto.

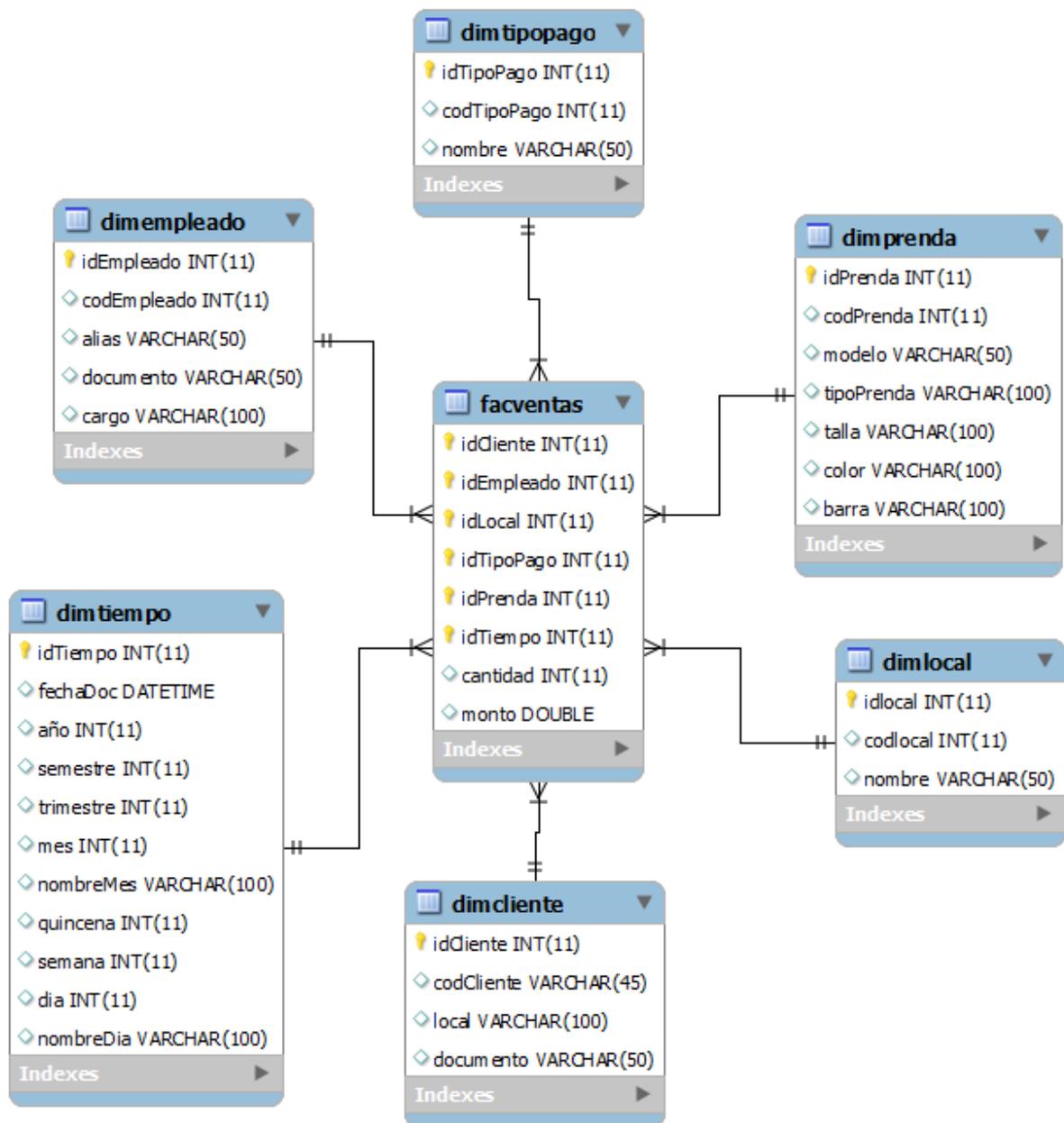


Figura 31: Implementación de Uniones (Fuente Propia)

## 3.4 INTEGRACION DE DATOS

### 3.4.1 Carga Inicial

En el proceso de carga se utilizó la herramienta Spoon de la suit de Pentaho, Se realizó la cargar de información de las dos bases de datos, una llamada “camilaViali” de Guayaquil y la otra “Wardrobe” de Lima. A continuación se muestran las transformaciones que se realizaron para cargar datos a cada una de las dimensiones.

#### Carga dimCliente

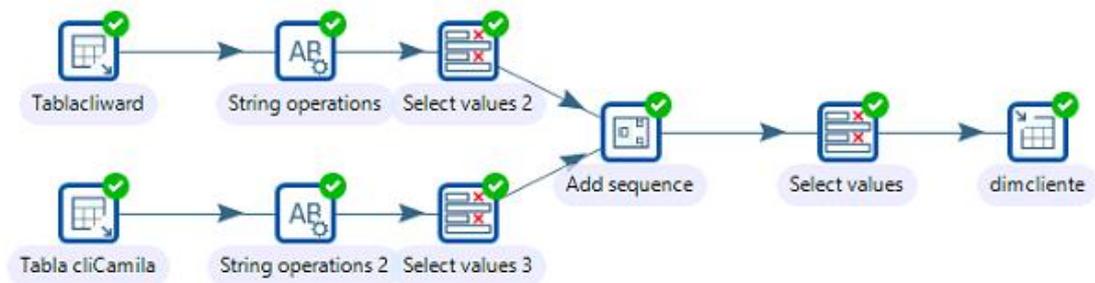


Figura 32: Carga dimCliente (Fuente Propia)

#### Sentencia SQL utilizada

##### Base Wardrobe

```
select c.codigo as codCliente,l.nombre as local,c.documento as documento
from wardrobe.clientes c, wardrobe.locales l
where c.codLocal=l.codigo
```

##### Base Camilaviali

```
select codigo as codCliente,'ecuador' as local, codigo as documento
from camilaviali.clientes
```

## Carga dimEmpleado

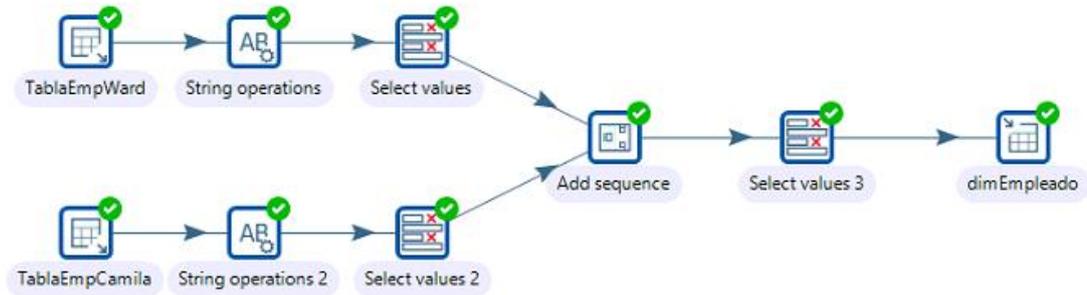


Figura 33: carga dimEmpleado (Fuente Propia)

## Base Wardrobe

```
select u.codigo as codEmpleado,u.alias as alias,u.documento as documento,c.descripcion as cargo
from wardrobe.usuarios u,wardrobe.cargos c
where u.codCargo=c.codigo
```

## Base Camilaviali

```
select u.codigo as codEmpleado,u.nombres as alias,u.codigo as documento,c.descripcion as cargo
from camilaviali.usuarios u,camilaviali.areastrabajo c
where u.codarea=c.codigo
```

## Carga dimPrenda

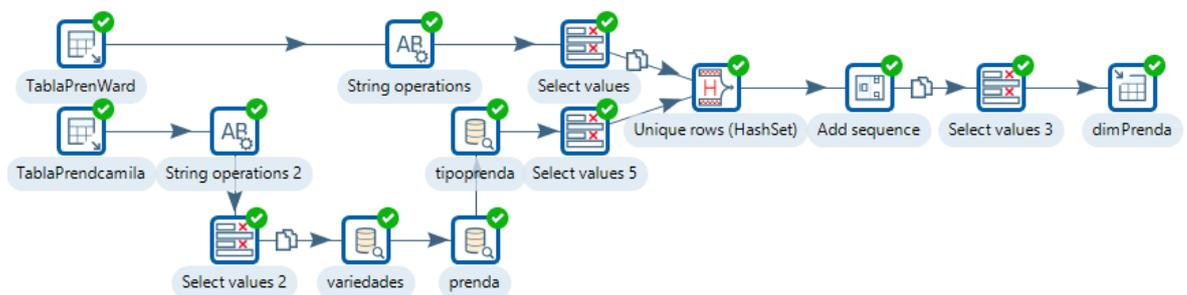


Figura 34: Carga dimPrenda (Fuente Propia)

## Sentencia SQL utilizada

### Base Wardrobe

```
select v.codigo as codPrenda,p.descripcion as
modelo,tp.descripcion as tipoPrenda,t.descripcion as
talla,c.descripcion as color,v.barra as barra
from wardrobe.variedades v,wardrobe.prendas
p,wardrobe.tipo_Prendas tp ,wardrobe.tallas t,wardrobe.colores c
where v.codPrenda=p.codigo
and v.codTalla=t.codigo
and v.codColor=c.codigo
and p.codTipoPrenda=tp.codigo
```

### Base Camilaviali

```
select p.codigoalternativo as modelo, t.descripcion as
talla,c.descripcion as color,p.codigobarra as barra
from camilaviali.productos p ,camilaviali.talla
t,camilaviali.color c where p.codTalla=t.codigo and
p.codColor=c.codigo
```

## Carga DimLocal

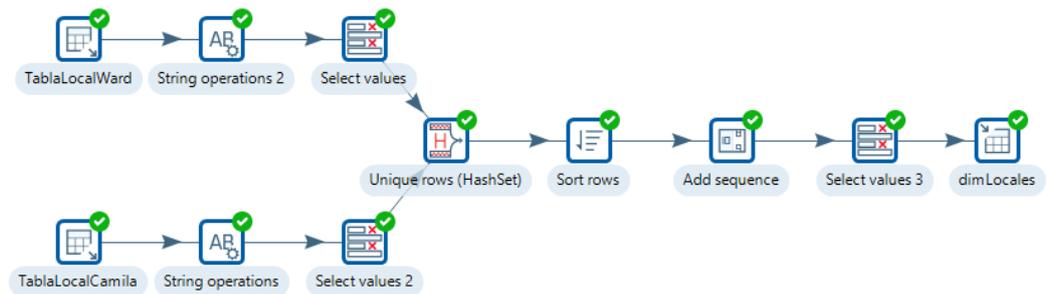


Figura 35: Carga dimLocal (Fuente Propia)

## Sentencia SQL utilizada

### Base Wardrobe

```
select codigo as codLocal,nombre as nombre
from wardrobe.locales
```

### Base Camilaviali

```
select codigo as codLocal,nombre as nombre
from camilaviali.empresa
```

## Carga DimTipoPago

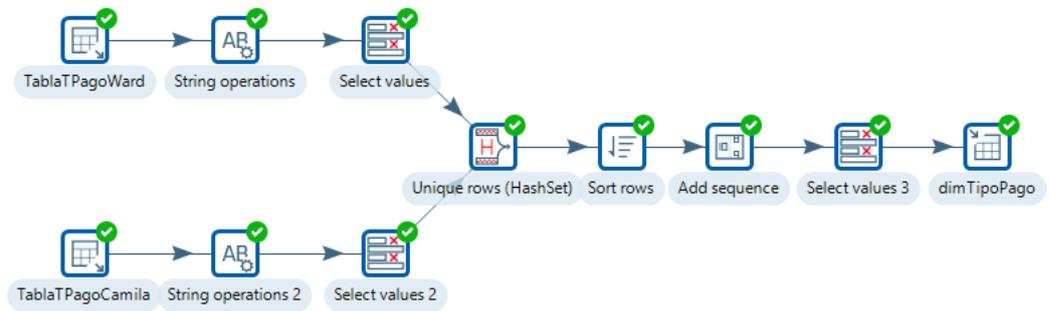


Figura 36: Carga dimTipoPago (Fuente Propia)

## Sentencia SQL utilizada

### Base Wardrobe

```
select codigo as codTipoPago, descripcion as nombre  
from wardrobe.tipo_pago
```

### Base Camilaviali

```
select codigo as codTipoPago, descripcion as nombre  
from camilaviali.tarjetas
```

## Carga DimTiempo

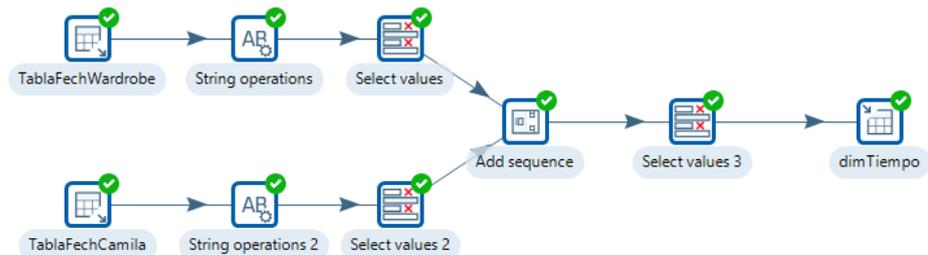


Figura 37: Carga dimTiempo (Fuente Propia)

## Sentencia SQL utilizada

### Base Wardrobe

```
select distinct fechadoc as fechaDoc,  
year(fechadoc) as año,  
case when MONTH (fechadoc)<=6 then 1 else 2 end as semestre,  
QUARTER(fechadoc) as trimestre,  
MONTH(fechaDoc) as mes,  
MONTHNAME(fechadoc) as nombreMes,  
case when DAYOFMONTH (fechadoc)<=15 then 1 else 2 end as  
quincena,  
WEEK(fechadoc) as semana,
```

```

DAYOFMONTH(fechaDoc) as dia,
DAYNAME(fechaDoc) as nombreDia
FROM wardrobe.documentos

```

## Base Camilaviali

```

select distinct fecha as fechaDoc,
year(fecha) as año,
case when MONTH (fecha)<=6 then 1 else 2 end as semestre,
QUARTER(fecha) as trimestre ,
MONTH(fecha) as mes,
MONTHNAME(fecha) as nombreMes,
case when DAYOFMONTH (fecha)<=15 then 1 else 2 end as quincena,
WEEK(fecha) as semana,
DAYOFMONTH(fecha) as dia,
DAYNAME(fecha) as nombreDia
FROM camilaviali.ventas

```

## FactVenta

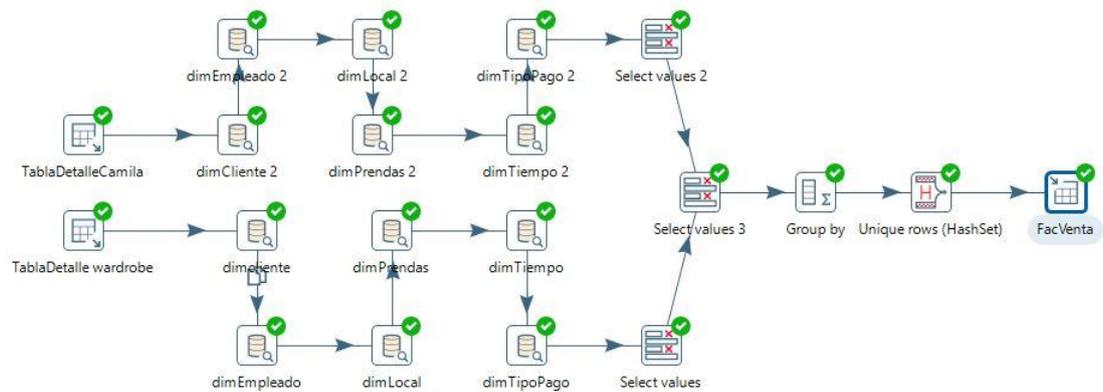


Figura 38: Carga FacVenta (Fuente Propia)

## Sentencia SQL utilizada

### Base Wardrobe

```

select
d.codCliente,d.codUsuario,d.codTienda,p.codTarjeta,r.codigoProducto
,d.fecha,r.cantidad,r.cantidad*r.precio AS monto
from camilaviali.Ventas d ,camilaviali.registros
r,camilaviali.tipopago p
where d.codigogeneral=r.codigoEncabezado
and d.codigogeneral=p.codigoEncabezado and d.estado=0

```

### Base Camilaviali

```

select
d.codCliente,d.codUsuario,d.codLocalProcedencia,p.codTipoPago,r.cod
Variedad,d.fechaDoc,r.cantidad,r.cantidad*r.precioVenta
from wardrobe.documentos d ,wardrobe.registros r,wardrobe.pagos p
where d.codigo=r.codDocumento and d.codigo=p.codDocumento and
d.estado=0

```

### 3.4.2 Cubo Multidimensional

En esta parte se mostrará los pasos a seguir de cómo se realizó la creación del cubo para el análisis de los datos de la base de datos dimensional. Para ello se utilizó la herramienta Schema Workbench de Pentaho.

- Se crea el cubo con el nombre “Cubo Ventas” dentro del schema llamada “Ventas Camila”

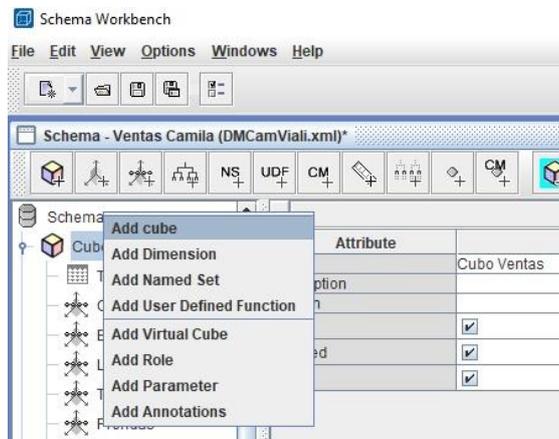


Figura 39: Creación del cubo (Fuente Propia)

- Se agrega una tabla dentro del cubo y se selecciona la tabla de hechos “facventas”.

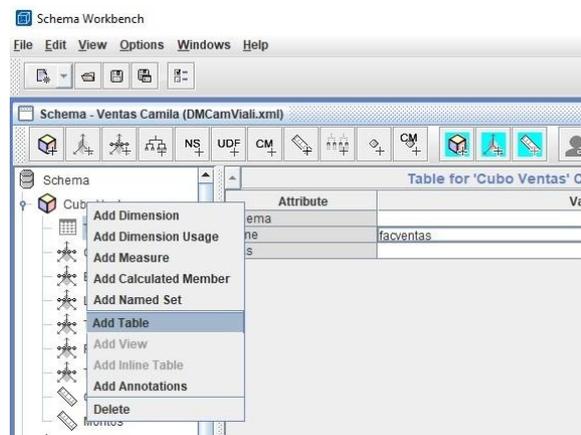


Figura 40: Agregar tabla facventa (Fuente Propia)

- Se agrega una dimensión y se le nombra “dimCliente”, esto se realiza para cada uno de las dimensiones que se identificaron.

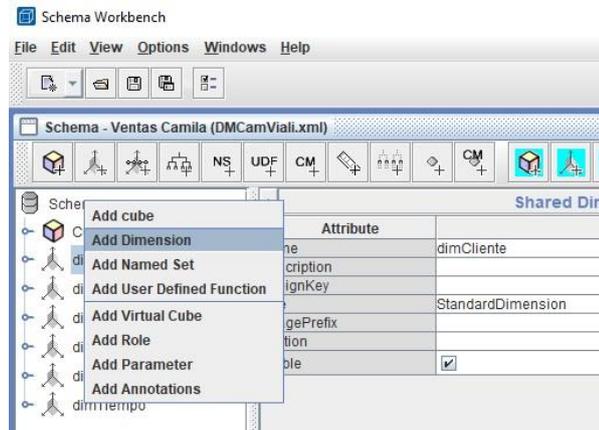


Figura 41: Crea dimensiones (Fuente Propia)

- Sobre la dimensión creada se agrega la jerarquía y se le coloca el nombre de “clientes” y se selecciona la llave primaria de la dimensión cliente.

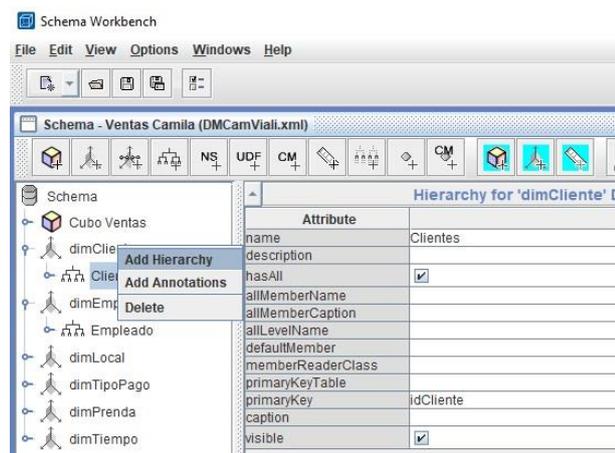


Figura 42: Crear Jerarquías (Fuente Propia)

- Dentro de la jerarquía “Clientes” se agrega una tabla y se selecciona la dimensión que le pertenece en este caso el dimCliente de la base de datos dimensional.

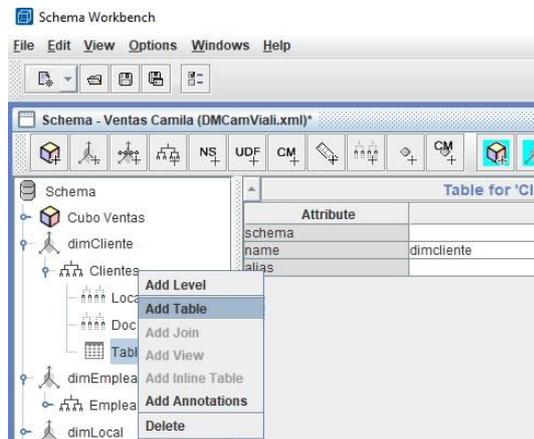


Figura 43: Agregar tabla dimensional (Fuente Propia)

- Luego dentro la de la misma jerarquía “Clientes” se agrega un nuevo nivel, se le nombró “Local Clientes” se selecciona la columna correspondiente, se determina el tipo de dato y se marca la opción uniqueMembers.

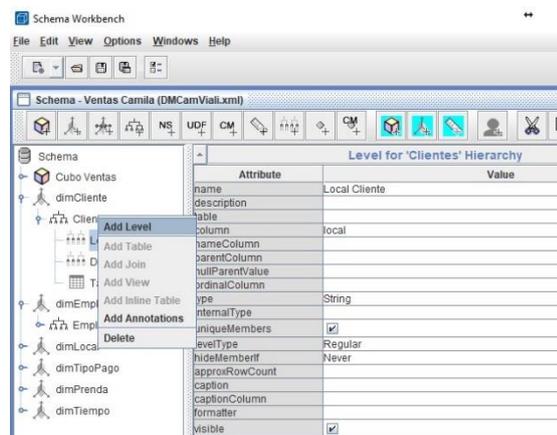


Figura 44: Crear niveles (Fuente Propia)

- Dentro del cubo se agregan todas las dimensiones que tenemos y las referenciamos con las claves foráneas correspondientes para cada una y se selecciona la dimensión correspondientes que se crearon fuera del cubo es decir en el schema.

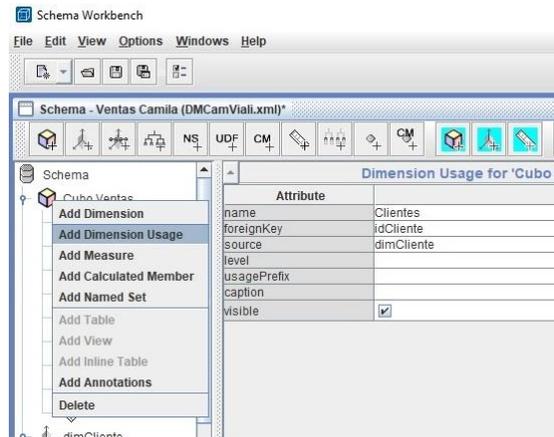


Figura 45: Dimensiones del cubo (Fuente Propia)

- y por último se agregan las métricas correspondientes en este caso se le nombró como “cantidades”, se eligió la función “sum” y se selecciona la columna correspondiente de la “tabla facventa”

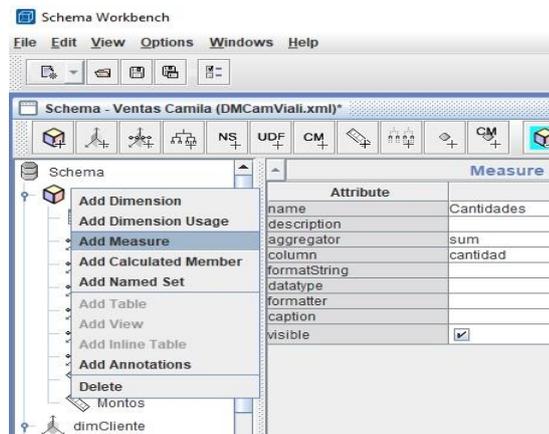


Figura 46: creación de métricas (Fuente Propia)

### 3.4.3 Resultados de Implementación

A continuación se mostrara los resultados que se obtuvieron tras la implementación de un Datamart para el área de venta por las necesidades de los distintos interesados para mejores tomas de decisiones.

La herramienta utilizada para visualizar los reportes es Pentaho Business Analytics (Saiku Analytics) esta herramienta es muy dinámica e intuitiva, permite ver los reportes en tablas y poder exportarlas rápidamente en formato Excel, además sus reportes gráficos son muy dinámicos y rápidos.

Uno de los reportes que se requerían era evaluar las cantidades de prendas vendidas dentro de un periodo determinado en este caso se solicitó que se muestre la cantidad de pantalones vendidos del modelo “P0213” en sus diferentes colores en cada trimestre del año 2017 para las tiendas de “Plaza Norte” y “Sebastián Barranca”. De lo requerido se obtuvo un reporte en tablas y dos reportes gráficos Mostrados a continuación.

Año		2017							
Trimestre		1		2		3		4	
Local		Plaza Norte 2	Sebastián Barranca						
Modelo	Color	Cantidades	Cantidades	Cantidades	Cantidades	Cantidades	Cantidades	Cantidades	Cantidades
P0213	ARENA	13	8	5	5	33	20	16	7
	AZUL	389	235	268	209	367	146	508	318
	BEIGE	209	119	124	107	183	108	304	173
	BLANCO	162	126	103	87	115	56	260	136
	CEMENTO	53	30	18	18	40	32	81	53
	MARRON	113	88	78	61	103	58	221	155
	NEGRO	583	356	450	360	386	219	793	441
	PALO ROSA	18	15	23	15	24	9	40	20
	PLOMO	54	44	28	16	56	25	98	70
	VINO	135	98	122	118	131	70	267	173

Figura 47: Reporte de cantidad de prendas por modelo, color y tiendas específicas (Fuente Propia)



Figura 48: Gráfico general de los colores más vendidos de un modelo específico (Fuente Propia)

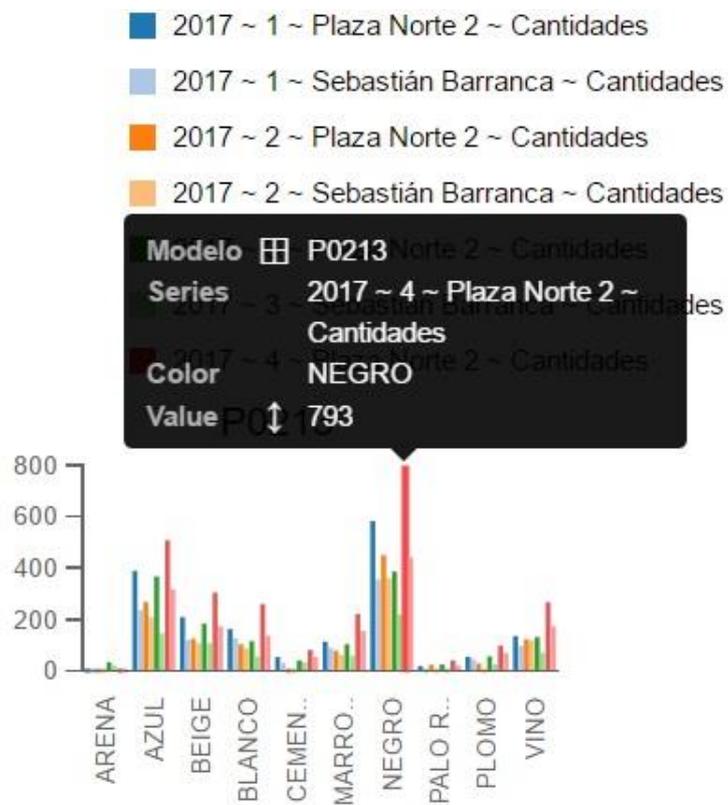


Figura 49: Gráfico de detalle de cantidades de prenda (Fuente Propia)

Otro de los reportes requeridos era determinar el monto total de ventas de prendas en un tiempo determinado, para este requerimiento se precisó que se querían visualizar las ventas anuales de cada una de las tiendas de Perú y Ecuador.

El resultado obtenido fue el reporte de ventas en tablas y un reporte grafico simple pero de gran información visual ya que a simple vista se sabe que tienda y en qué año se vendió más, lo que podrían evaluar cuales fueron los antecedentes que afectaron ese año.

Año	2014	2015	2016	2017	2018
Local	Montos	Montos	Montos	Montos	Montos
Bellavista	-	-	131.244,1	616.367,25	69.113
CentroPark	123	134.379,5	188.386,7	135.875,3	11.488
El Polo	-	-	-	211.269	95.276
El Rey	-	-	180.490	669.973,616	45.046,9
Mall del Sur	-	-	141.560,5	833.997,003	108.477,3
Plaza Norte 2	-	-	349.692,55	1.650.344	191.888,4
Sebastián Barranca	-	-	137.695,2	1.084.873,45	125.650

Figura 50: Reporte de ventas generales por tienda (Fuente Propia)

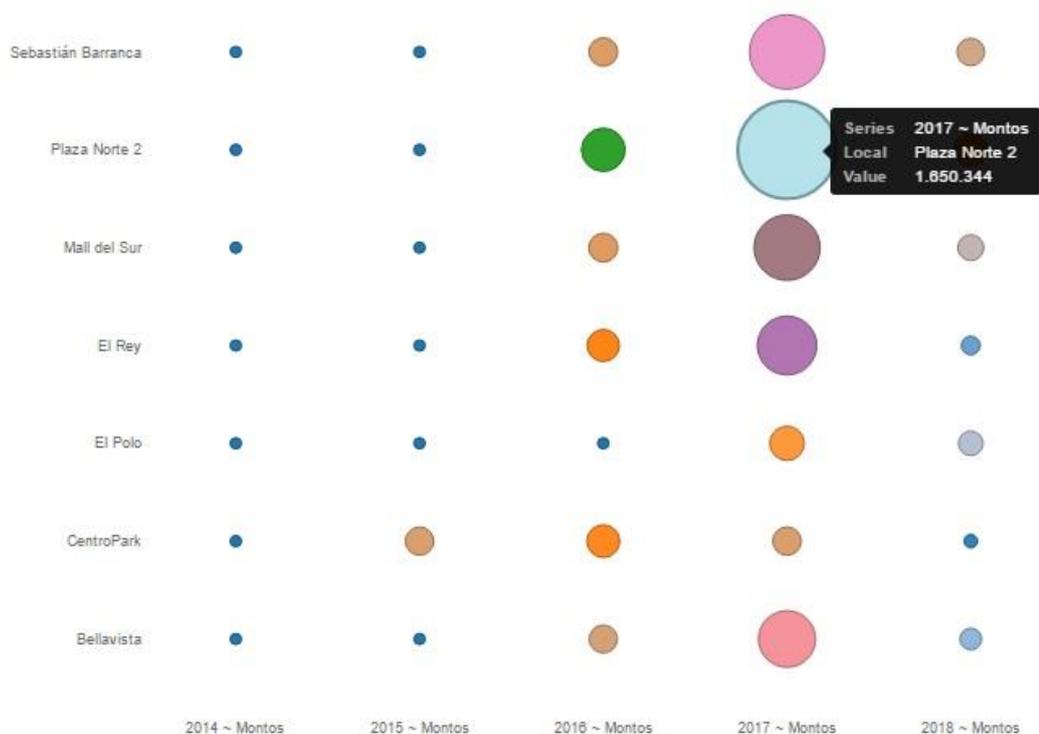


Figura 51: Gráfico de reconocimiento rápido de detalles (Fuente Propia)

El siguiente reporte se realizó para determinar el monto total de ventas que realizó un empleado en un tiempo determinado para este caso se solicitó que se mostraran todos los empleados que realizaron ventas en la tienda de “Plaza Norte” en el año 2017. El resulta fue el siguiente:

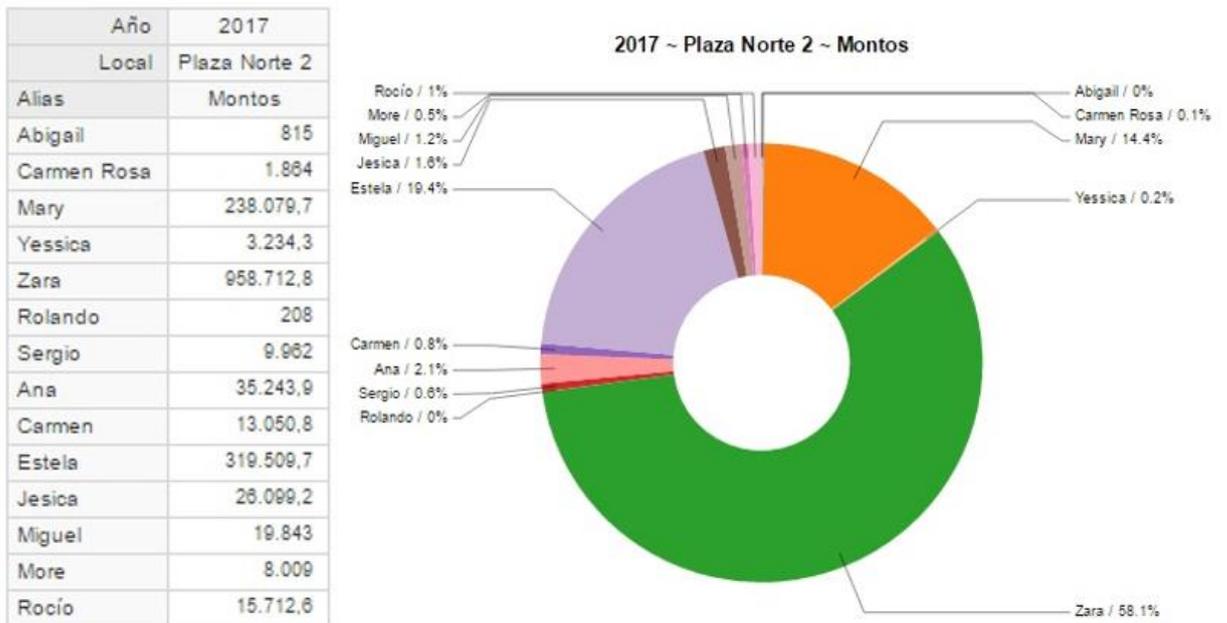


Figura 52: Reporte de monto de ventas por empleado (Fuente Propia)

### 3.4.4 KPI

#### **Métrica orientada a medir el estado del desarrollo de las ventas en un determinado tiempo**

En esta métrica se evalúa el rendimiento de las ventas, para este caso en un periodo de un mes, se podrá visualizar cómo va el desarrollo y ver si van cumpliendo con las cuotas de ventas establecidas en todo caso se podría tomar medidas o acciones en el caso de que el rendimiento no sea el favorable.



Figura 53 KPI Monitoreo de ventas (Fuente Propia)

#### **Métrica orientada obtener el máximo de prendas vendidas en un periodo determinado.**

Esta Métrica evalúa el movimiento de determinadas prendas, con estos indicadores se puede realizar una mejor distribución de inventario, y no se incurriría a la adquisición de productos que no tienen movimientos en las ventas y se optarían en potenciar a los que realmente prefiere los clientes.

## INDICADORES DE VENTAS

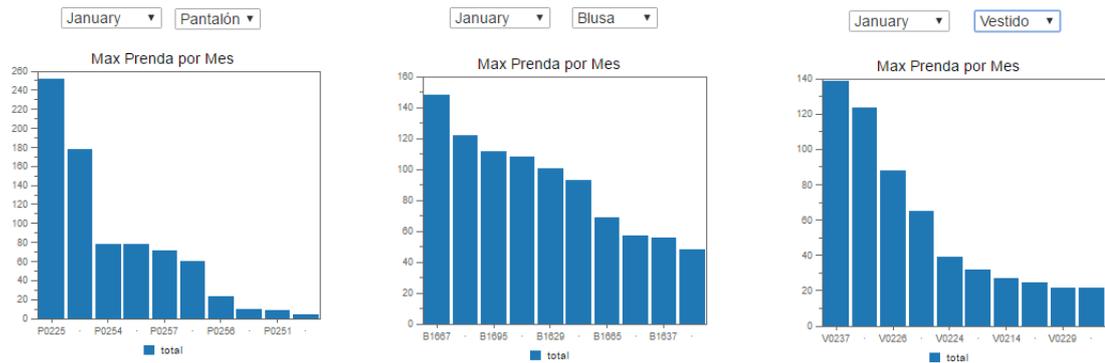


Figura 54 KPI Prendas más vendidas (Fuente Propia)

### 3.4.5 Cronograma y Presupuesto

#### Cronograma

El cronograma realizado está centrada específicamente en el desarrollo de la implementación del proyecto, estimando el tiempo aproximado y describiendo las dependencias que cada uno de las fases que tiene la metodología Hefestos, cada una de las actividades de las distintas fases se realizaron de manera satisfactoria y culminadas a un 100%

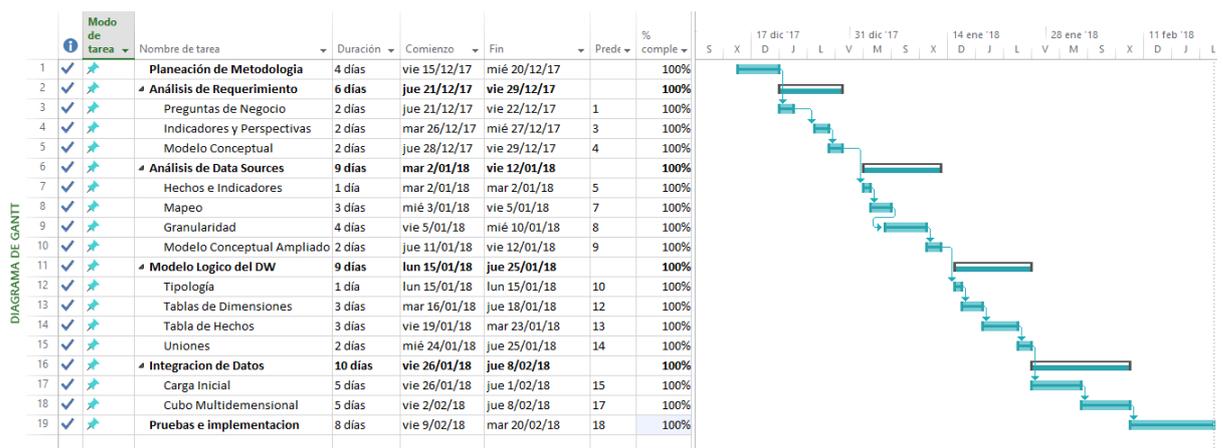


Figura 55 Cronograma del Proyecto (Fuente Propia)

## Presupuesto

En temas de costos para la implementación de este proyecto en tema de software es 0 puesto que se han utilizado software Open Source, como lo son Mysql y la suite de Pentaho. Donde sí se ha realizado costos es en el pago de personal y equipos de hardware, a continuación se mostrara una tabla con los costos que se ha tenido para la implementación del proyecto.

Concepto	Costo (S/)	Descripcion
Mysql	S/ 0	Open Source
Pentaho Business Analytics	S/ 0	Open Source
Schema Workbench	S/ 0	Open Source
Spoon Data Integration	S/ 0	Open Source
Pago a Personal	S/ 3500	Pago por implementación del proyecto al autor del proyecto
Laptop	S/3300	Compra de laptop con recursos necesarios para soportar herramientas de software
<b>Total</b>	<b>S/6800</b>	

Tabla 3 Presupuesto (Fuente Propia)

## CONCLUSIONES

1. Se logró diseñar correctamente el Datamart gracias a la metodología Hefestos, debido al análisis en cada uno de sus pasos y sus buenas practicas, ahora se puede revisar los reportes muchos más rápido y con mayor detalle, además que es interactiva para los usuarios.
2. Se realizó satisfactoria la recolección de información por parte de los interesados mediante un formato de encuesta que dieron a conocer cuáles eran las estrategias gerenciales más importantes para poder analizar la información y plasmarlos dentro del Datamart.
3. Se logró obtener un buen diseño del modelo Dimensional por la identificación correcta de cada uno de los factores que eran necesarios, como las perspectivas e indicadores y la relación que guardaban con la base de dato.
4. Se ejecutó correctamente los procesos de extracción, transformación y carga (ETL) gracias a Pentaho, una herramienta de BI completa y que una de sus ventajas es ser open source lo que ha llevado a minimizar costos de compra de licencias.
5. Se logró reducir en un 80% la entrega de reportes, además que resultó interactivo para los usuarios el modo de obtención de información y la rapidez con la que se puede recolectar dicha información requerida. Además los resultados de la implementación permitieron estimar en el caso de las venta anuales cuales son las metas que se debe seguir este año 2018 y que factores se deberían tomar en cuenta para lograr los objetivos de la empresa.

## RECOMENDACIONES

1. Se recomienda que se construyan y adicionen nuevos modelos de las demás áreas que cuenta la empresa para poder tener mejores resultados de información y realizar mejores tomas de decisiones.
2. Se debe tener una administración y seguimiento continuo de todo el proceso ETL hacia el Datamart para contar con información actualizada que sea de beneficio a los interesados.
3. Se recomienda generar documentación que respalde los procesos de implementación que se realizó para las futuras integraciones de Datamart de las demás áreas de la empresa.
4. Se debería implementar una solución de sistema web para el monitoreo desde cualquier lugar por parte de los usuarios.

## BIBLIOGRAFÍA

1. Roja, A. (2014). Implementación de un Datamart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el Departamento de Finanzas de la Contraloría general de la república. Tesis de Titulación. Facultad de Ingeniería y Arquitectura. Universidad San Martín de Porres. Chiclayo, Perú.
2. Durand, A. (2014) Desarrollo de un Datamart para mejorar la toma de decisiones en el área de ventas de la corporación Furukawa. Trabajo de Investigación. Facultad de Ingeniería de Sistemas y Administración de Empresas. Universidad Nacional Tecnológica de Lima Sur. Lima, Perú.
3. Moreno, R. (2013) Análisis, Diseño e Implementación de Datamarts para las áreas de ventas y recursos humanos de una empresa dedicada a la exportación e importación de productos alimenticios. Tesis de Titulación, Facultad de Ciencias e Ingeniería. Pontificia Universidad Católica del Perú. Lima, Perú.
4. Villanueva, A. (2008) Análisis, Diseño e Implementación de un Datawarehouse de soporte de decisiones para un hospital del sistema de salud público. Tesis de Titulación. Facultad de Ciencias e Ingeniería. Pontificia Universidad Católica del Perú. Lima, Perú.
5. Contreras, L y Vera, S. (2009). Desarrollo de una solución de inteligencia de negocios para un departamento de finanzas dentro del sector petroquímico. Trabajo de Investigación. Facultad de Ciencias. Universidad Central de Venezuela. Caracas, Venezuela.
6. Hernández, L. (2008). Diseño y Construcción de un Datamart para la mantención de indicadores de sostenibilidad de la industria del salmón. Tesis

de Maestría. Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile. Santiago de Chile, Chile.

7. Aimacaña, D. (2013). Análisis, Diseño e Implementación de un Datamart académico usando tecnología de BI para la Facultad de Ingeniería, Ciencias Físicas y Matemática. Trabajo de Investigación. Facultad de Ingeniería, Ciencias Físicas y Matemática. Universidad Central del Ecuador. Quito, Ecuador.
8. Oz E. (2008). Administración de los sistemas de información. Santa fe, México: Ediciones Cengage Learning.
9. Arjonilla S. J.; Medina, J. A. (2013). La gestión de los sistemas de información en la empresa. Madrid, España: Ediciones Pirámide.
10. Laudon K. C.; Laudon J. P. (2012). Sistemas de información Gerencial. Naucalpan de Juárez, México: Ediciones Pearson.
11. Soroka R. (2002). Sistema de Información en la era Digital. Argentina: Ediciones OSDE.
12. Cohen, D.; Asin, E. (2000). Sistemas de información para los negocios. México: Ediciones Mc GRAW- HILL.
13. Ramos S. (2016). Business Intelligence (BI) & Analytics. El arte de convertir datos en conocimiento. Albatra, España: Ediciones SolidQ Press.
14. Cano J. (2007). Business intelligence; competir con información. Madrid, España: Ediciones Banesto, Fundación Cultur.
15. Berbabeu R. (2007). DATA WAREHOUSING: Investigación y Sistematización de Conceptos – HEFESTO: Metodología propia para la Construcción de un Data Warehouse. Córdoba, Argentina.

16. Datawarehouse4u. (2009). OLTP vs. OLAP. En Datawarehouse4u. Consultado el 20 de diciembre del 2017. Disponible en <http://datawarehouse4u.info/OLTP-vs-OLAP.html>
17. GeoIDEP (2014). Que son los Metadatos. En GeoIDEP. Consultado el 10 de enero del 2018. Disponible en <http://www.geoidep.gob.pe/metadatos/que-son-los-metadatos>
18. EcuRed (2018). Hipertexto. En EcuRed. Consultado el 10 de enero del 2018. Disponible en <https://www.ecured.cu/Hipertexto>
19. Sinnexus(2016). Datamart. En Sinnexus. Consultado el 10 de enero del 2018. Disponible en [http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx)
20. Stephany(2012). Analisis de drilldown. Consultado el 10 de enero del 2018. Disponible en <http://stefymichel.blogspot.pe/2012/05/analisis-de-drill-down.html>
21. MySQLHispano(2003). Normalización de base de datos. Consultado el 10 de enero del 2018. Disponible en <http://www.eet2mdp.edu.ar/alumnos/MATERIAL/MATERIAL/info/infonorma.pdf>
22. Manuel (2014). Los sistemas de gestión de bases de datos SGBD. En Fundamentos y Diseño de Base de datos. Consultado el 10 de enero del 2018. Disponible en <http://ccdoc-basesdedatos.blogspot.pe/2014/02/los-sistemas-de-gestion-de-bases-de.html>
23. IBM (2018). ¿Qué es jerarquía de dimensiones. En IBM. Consultado el 10 de enero del 2018. Disponible en [https://www.ibm.com/support/knowledgecenter/es/SSCVKV\\_9.1.2/Campaign/DimensionHierarchyAdmin/What\\_is\\_a\\_dimension\\_hierarchy.html](https://www.ibm.com/support/knowledgecenter/es/SSCVKV_9.1.2/Campaign/DimensionHierarchyAdmin/What_is_a_dimension_hierarchy.html)

24. Espinosa (2010). Herramientas ETL. ¿Que son, para que valen?. Productos mas conocidos. ETL´s Open Source. En Dataprix. Consultado el 10 de enero del 2018. Disponible en <http://www.dataprix.com/blogs/respinosamilla/herramientas-etl-que-son-para-que-valen-productos-mas-conocidos-etl-s-open-sour>
25. Rouse (2012). OLTP (Procesamiento de Transacciones En Línea). En SearchDataCenter. Consultado el 10 de enero del 2018. Disponible en <http://searchdatacenter.techtarget.com/es/definicion/OLTP-Procesamiento-de-Transacciones-En-Linea>
26. Dataprix (2009). Metodología Hefestos. Recuperado de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/hefesto-metodologia-propia-para-la-construccion-un-data-wareh>.
27. Toaing, M. (2014), "Construcción de un Datamart orientado a las ventas para la toma de decisiones en la empresa Amevet cia. Ltda". Proyecto de investigación. Facultad de Ingeniería en Sistemas Electrónica e Industrial. Universidad Técnica de Ambato. Ambato, Ecuador.
28. Campomanes, J. (2017 ), "Datamart en el Proceso de Toma de Decisiones de Ventas de la empresa Industria del calzado el Lobo s.a.c". Tesis de Titulación. Universidad Cesar Vallejo. Lima, Perú.

## ANEXOS

### Anexo 1 Entrevistas

#### INFORMACION GENERAL

Nombre del entrevistado: Sergio Saldaña  
Cargo: Jefe de Ventas Fecha de entrevista: 13/12/2017  
Lugar de encuesta: Oficinas de "Camila Viali"

1. ¿Cuáles son las necesidades más prioritarias que tiene el área al que pertenece?

Como cualquier área de ventas la prioridad es vender cada día más, lo que se necesita es información relevante para poder realizar mejores estrategias de ventas, como por ejemplo realizar un comparativo de las ventas mensuales de años anteriores para saber cuotas aproximadas para los meses actuales, además hay datos de Ecuador que no están integrados con los datos de Perú.

2. ¿Cómo evaluaría la entrega de reportes actualmente?

Actualmente es un poco tardía la información por el motivo de que se tiene dos sistemas distintos uno para las ventas de Guayaquil y otra para Lima.

3. ¿En qué tiempo le gustaría tener los reportes para su análisis y evaluación?

Lo más rápido posible, entre una a dos horas sería lo ideal.

4. ¿Qué tan variables es el tiempo para los reportes que requiere?

Es bastante variable porque existen diferentes reportes y cada uno es más complejo que otro, el que más tarda es el reporte mensual por el problema de tener dos sistemas diferentes.

5. ¿Cuáles son los requerimientos que usted considera más importantes para el análisis de la información?

La cantidad de prendas vendidas a cada cliente en un periodo determinado.  
El monto total de ventas de las prendas en un periodo determinado.  
Comparativo de monto de venta por periodos determinado  
El monto total de ventas por empleado en un tiempo determinado.  
El monto total de ventas por tipo de pago en un tiempo determinado.

Fuentes Propia

## INFORMACION GENERAL

Nombre del entrevistado: Alberto Salomón

Cargo: Jefe de Diseño Fecha de entrevista: 13/12/2017

Lugar de encuesta: Oficinas de "Camila Viali"

1. ¿Cuáles son las necesidades más prioritarias que tiene el área al que pertenece?

En esta área se necesita que diseñen prendas lo más actuales a la moda pero manteniendo el estilo elegante que se nos caracteriza, pero no solo basta con eso, se necesitaría cuáles son las preferencias de los clientes en cuanto a modelos, tallas y colores, esto brindaría una mejor toma de decisiones al momento de decidir diseñar una prenda.

2. ¿Cómo evaluaría la entrega de reportes actualmente?

Para los reportes simples no hay demasiada demora, pero para los que son más complejos como evaluar las prendas en distintos periodos de tiempo el reporte demora mucho más.

3. ¿En qué tiempo le gustaría tener los reportes para su análisis y evaluación?

Me gustaría que estuviera en lapso de horas, es decir que el reporte esté listo para el mismo día.

4. ¿Qué tan variables es el tiempo para los reportes que requiere?

Es regularmente variable, depende del reporte que se solicite

5. ¿Cuáles son los indicadores que usted considera más importantes para el análisis de la información?

La cantidad de prendas vendidas en un periodo de tiempo.

La cantidad de prendas vendidas en un local en un periodo determinado.

Fuente Propia