

ESCUELA POLITÉCNICA DEL EJÉRCITO

DPTO. DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

**IMPLEMENTACIÓN DE UN DATAWAREHOUSE PARA
EL INSTITUTO GEOGRÁFICO MILITAR**

Previa a la Obtención del Título de:

INGENIERO EN SISTEMAS E INFORMÁTICA

POR:

SR. ANDRÉS FABIÁN DUQUE GÁLVEZ

SANGOLQUÍ, Abril del 2010

CERTIFICACIÓN

Certifico que el presente trabajo fue realizado en su totalidad por el Sr. Andrés Fabián Duque Gálvez como requerimiento parcial a la obtención del título de Ingenieros en Sistemas e Informática.

27 de Abril de 2010.

Ing. Carlos Rojas

DEDICATORIA

El proyecto de tesis está dedicado especialmente a mi familia, mis padres FABIÁN DUQUE Y ANA LUCÍA GÁLVEZ, y a mis hermanos DANIEL DUQUE y ANITA GABRIELA DUQUE, que con su apoyo incondicional supieron darme fuerzas para seguir adelante en la consecución de la tesis. Por su cariño y amor que día a día me hacen llegar, y que fueron mis pilares desde siempre para no decaer en los retos que me pone la vida. Gracias a mi familia que como núcleo fueron la fuente de inspiración durante toda mi vida, y de aquí en adelante en mi carrera profesional. A su ardua labor de inculcarme los valores necesarios para formarme como persona.

A mis abuelitos que con su preocupación y cariño me daban ánimos en el desarrollo constante del proyecto.

A toda mi familia que me brinda su apoyo en las buenas y en las malas.

ANDRÉS FABIÁN DUQUE GÁLVEZ.

AGRADECIMIENTOS

Agradezco primeramente a Dios por iluminarme durante cada paso que he dado en mi vida, en cada caída y en cada triunfo.

A mi familia que me apoyaron en toda circunstancia que se me presentó, a su constante cariño y amor que me dieron, y que sirvió de camino durante este recorrido universitario.

A mis profesores ING. CARLOS ROJAS y MSC. GABRIEL CHIRIBOGA por su amistad y guía para el desarrollo del proyecto de tesis.

A mis amigos por su amistad y apoyo durante toda nuestra carrera universitaria y en la vida.

A mí apreciada ESPE donde fui formado y a mis maestros que fueron más que docentes, por impartir el conocimiento durante todo el tiempo que duro mi carrera.

ANDRÉS FABIÁN DUQUE GÁLVEZ

ÍNDICE DE CONTENIDOS

CAPÍTULO I: INTRODUCCIÓN.....	- 1 -
1.1. ANTECEDENTES	- 1 -
1.2. PLANTEAMIENTO DEL PROBLEMA.....	- 1 -
1.3. JUSTIFICACIÓN	- 2 -
1.4. OBJETIVOS	- 3 -
1.4.1. OBJETIVO GENERAL	- 3 -
1.4.2. OBJETIVOS ESPECÍFICOS	- 3 -
1.5. ALCANCE	- 3 -
1.6. HIPÓTESIS DE TRABAJO.....	- 4 -
CAPÍTULO II: MARCO TEÓRICO.....	- 5 -
2.1. BUSINESS INTELLIGENCE	- 5 -
2.1.1. INTRODUCCIÓN.	- 5 -
2.1.2. DEFINICIÓN.	- 6 -
2.1.3. IMPORTANCIA Y BENEFICIO DE BUSINESS INTELLIGENCE EN LA ORGANIZACIÓN.	- 6 -
2.1.4. ARQUITECTURA DE BUSINESS INTELLIGENCE	- 8 -
2.1.5. ETAPAS DE BUSINESS INTELLIGENCE	- 10 -
2.1.5.1. DIRIGIR Y PLANEAR	- 10 -
2.1.5.1.1. TOMA DE DECISIONES.....	- 11 -
2.1.5.2. IDENTIFICACIÓN DE INFORMACIÓN.....	- 12 -
2.1.5.3. PROCESAMIENTO DE DATOS.....	- 13 -
2.1.5.4. ANÁLISIS Y PRODUCCIÓN	- 14 -
2.2. DATA WAREHOUSE.....	- 14 -
2.2.1. INTRODUCCIÓN	- 14 -
2.2.2. CARACTERÍSTICAS	- 15 -
2.2.3. ARQUITECTURA DE UN DATAWAREHOUSE	- 17 -
2.2.3.1. SISTEMAS ORIGEN.....	- 17 -
2.2.3.2. STAGING AREA (AREA TEMPORAL)	- 17 -
2.2.3.3. ODS (OPERATIONAL DATA STORE)	- 18 -
2.2.3.4. DATA WAREHOUSE	- 18 -
2.2.3.5. DATA MARTS	- 19 -
2.2.4. ESTRUCTURA.....	- 19 -
2.2.5. PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA	- 21 -
2.2.5.1. EXTRACCIÓN.....	- 22 -
2.2.5.2. TRANSFORMACIÓN	- 23 -
2.2.5.3. CARGA	- 23 -
2.2.6. MODELO DIMENSIONAL DE UN DATAWAREHOUSE.....	- 24 -
2.2.6.1. ESQUEMA ESTRELLA.....	- 25 -
2.2.6.2. ESQUEMA COPO DE NIEVE.....	- 26 -
2.2.6.3. TABLAS	- 27 -
2.2.6.3.1. HECHOS.....	- 27 -
2.2.6.3.2. DIMENSIONES	- 28 -
2.2.6.4. NIVEL	- 28 -
2.2.6.5. JERARQUÍA	- 29 -
2.2.6.6. DIMENSIONES LENTAMENTE CAMBIANTES (SCD)	- 30 -
2.2.6.7. GRANURALIDAD	- 30 -
2.2.6.8. RELACIONES	- 31 -
2.2.6.9. LLAVES SUBROGADAS	- 32 -
2.2.7. EXPLOTACIÓN DEL DATAWAREHOUSE	- 32 -

2.2.7.1.	REPORTES Y CONSULTAS.....	- 32 -
2.2.7.2.	MINERÍA DE DATOS.....	- 33 -
2.2.7.3.	ANÁLISIS OLAP	- 34 -
2.2.7.3.1.	ROLAP (RELATIONAL OLAP)	- 35 -
2.2.7.3.2.	MOLAP (MULTIDIMENSIONAL OLAP)	- 36 -
2.2.7.3.3.	HOLAP (HYBRID OLAP)	- 36 -
2.3.	DESARROLLO DE UN DATAWAREHOUSE BASADO EN LA METODOLOGIA RALPH	
KIMBALL.....		- 39 -
2.3.1.	PLANEACIÓN Y ADMINISTRACIÓN DEL PROYECTO.....	- 40 -
2.3.2.	DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO.....	- 43 -
2.3.3.	DISEÑO TÉCNICO DE LA ARQUITECTURA.....	- 46 -
2.3.3.1.	Back Room	- 46 -
2.3.3.2.	Front Room	- 46 -
2.3.4.	SELECCIÓN E INTALACIÓN DEL PRODUCTO	- 47 -
2.3.5.	MODELAMIENTO DIMENSIONAL.....	- 48 -
2.3.5.1.	Paso 1 – Seleccionar el Proceso de Negocio.	- 48 -
2.3.5.2.	Paso 2 – Declaración de granularidad.....	- 49 -
2.3.5.3.	Paso 3 – Identificación de Dimensiones.....	- 49 -
2.3.5.4.	Paso 4 – Identificación de Hechos.....	- 50 -
2.3.5.5.	DataWarehouse Bus Matrix	- 50 -
2.3.6.	DISEÑO FÍSICO	- 50 -
2.3.7.	DISEÑO Y DESARROLLO DEL PROCESO ETL.....	- 51 -
2.3.7.1.	Paso 1 – Trazar un plan de alto nivel.	- 51 -
2.3.7.2.	Paso 2 – Seleccionar la herramienta ETL.....	- 51 -
2.3.7.3.	Paso 3 – Desarrollar estrategias por defecto.	- 52 -
2.3.7.4.	Paso 4 – Profundizar la tabla objetivo.....	- 52 -
2.3.7.5.	Paso 5 – Poblar las tablas de Dimensiones con los datos históricos.....	- 52 -
2.3.7.6.	Paso 6 – Realizar la carga histórica de la Tabla de Hechos.	- 53 -
2.3.7.7.	Paso 7 – Procesamiento incremental de la Dimensiones.	- 53 -
2.3.7.8.	Paso 8 – Procesamiento incremental de los Hechos.....	- 53 -
2.3.7.9.	Paso 9 – Operación y Automatización del Sistema ETL.....	- 53 -
2.3.8.	DISEÑO Y DESARROLLO DE APLICACIÓN DE BI	- 54 -
2.3.8.1.	Tipo de aplicación de BI	- 54 -
2.3.8.2.	Preparar el desarrollo de la aplicación.....	- 55 -
2.3.8.3.	Crear las aplicaciones.....	- 55 -
2.3.9.	DESPLIEGUE	- 55 -
2.3.9.1.	Despliegue ETL	- 56 -
2.3.9.2.	Despliegue de BI.....	- 56 -
2.3.10.	MANTENIMIENTO Y CRECIMIENTO.....	- 56 -
2.4.	HERRAMIENTAS PARA EL DESARROLLO DEL DATAWAREHOUSE	- 56 -
2.4.1.	MOTOR DE BASE DE DATOS.....	- 57 -
2.4.2.	HERRAMIENTA ETL.	- 57 -
2.4.3.	HERRAMIENTA DE PRESENTACIÓN.....	- 58 -
2.5.	LINEA DE NEGOCIO	- 59 -
2.5.1.	INVENTARIOS.....	- 59 -
2.5.2.	PUNTO DE VENTAS	- 60 -
CAPÍTULO III: DESARROLLO E IMPLEMENTACIÓN DEL DATAWAREHOUSE PARA EL IGM		
.....		- 61 -
3.1.	PLANEACIÓN Y ADMINISTRACIÓN DEL PROYECTO.	- 61 -
3.1.1.	DEFINICIÓN DEL PROYECTO.....	- 61 -
3.1.2.	ALCANCE Y JUSTIFICACIÓN DEL PROYECTO	- 61 -
3.2.	DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO	- 62 -

3.2.1.	Proceso de Inventarios	- 63 -
3.2.2.	Proceso de Punto de Ventas.....	- 66 -
3.3.	DISEÑO TÉCNICO DE LA ARQUITECTURA.....	- 70 -
3.3.1.	BACK ROOM.....	- 70 -
3.3.2.	FRONT ROOM	- 72 -
3.4.	SELECCIÓN DEL PRODUCTO.....	- 73 -
3.4.1.	Plataforma DBMS.	- 73 -
3.4.2.	Herramienta ETL.	- 73 -
3.4.3.	Herramienta de BI.	- 76 -
3.5.	ESTÁNDARES DE DESARROLLO PARA EL PROYECTO	- 78 -
3.5.1.	Estándares de Tablas.	- 78 -
3.5.2.	Estándares de Campos.	- 78 -
3.5.3.	Estándares de Índices.	- 79 -
3.5.4.	Estándares de Disparadores (Triggers)	- 80 -
3.5.5.	Estándares ETL.....	- 80 -
3.6.	MODELAMIENTO DIMENSIONAL.....	- 80 -
3.6.1.	PASO 1: SELECCIÓN DEL PROCESO DE NEGOCIO.....	- 81 -
3.6.2.	PASO 2: DEFINICIÓN DE LA GRANULARIDAD.....	- 82 -
3.6.2.1.	Inventarios	- 82 -
3.6.2.2.	Punto de Ventas.....	- 82 -
3.6.3.	PASO 3: IDENTIFICAR LAS DIMENSIONES Y MAPEAR LOS DATOS ORIGEN.....	- 82 -
3.6.3.1.	Dimensión Tiempo	- 83 -
3.6.3.2.	Dimensión Área.	- 84 -
3.6.3.3.	Dimensión Artículo.....	- 85 -
3.6.3.4.	Dimensión Bodega.	- 86 -
3.6.3.5.	Dimensión Centro de Gestión	- 87 -
3.6.3.6.	Dimensión Cliente.	- 88 -
3.6.3.7.	Dimensión Empleado.	- 89 -
3.6.3.8.	Dimensión Factura	- 90 -
3.6.3.9.	Dimensión Ítem.....	- 91 -
3.6.3.10.	Dimensión Movimiento.....	- 92 -
3.6.3.11.	Dimensión Orden de Compra.....	- 93 -
3.6.3.12.	Dimensión Pedido.	- 94 -
3.6.3.13.	Dimensión Proforma	- 95 -
3.6.3.14.	Dimensión Proveedor	- 96 -
3.6.3.15.	Dimensión Proyecto.....	- 97 -
3.6.3.16.	Dimensión Sucursal.....	- 98 -
3.6.4.	PASO 4: IDENTIFICAR LOS HECHOS Y MAPEAR LOS DATOS ORIGEN.....	- 99 -
3.6.4.1.	Pedidos.....	- 99 -
3.6.4.2.	Movimientos.	- 101 -
3.6.4.3.	Ventas	- 102 -
3.6.4.4.	Número de Ventas	- 103 -
3.6.5.	DISEÑO DEL MODELO DIMENSIONAL INVENTARIOS.....	- 104 -
3.6.6.	DISEÑO DEL MODELO DIMENSIONAL PUNTO DE VENTAS	- 105 -
3.6.7.	DISEÑO DEL MODELO DIMENSIONAL TOTAL	- 106 -
3.6.8.	DATA WAREHOUSE BUS MATRIX FINAL	- 107 -
3.7.	DISEÑO FÍSICO.....	- 108 -
3.7.1.	BASE DE DATOS ORIGEN.....	- 108 -
3.7.1.1.	ESPACIO DE TABLA (TABLESPACE).....	- 108 -
3.7.1.2.	ESQUEMA	- 108 -
3.7.1.3.	TABLAS	- 109 -
3.7.1.4.	ÍNDICES.....	- 109 -
3.7.1.5.	RESTRICCIONES (CONSTRAINTS)	- 109 -

3.7.1.6.	SECUENCIA	- 109 -
3.7.1.7.	DISPARADORES (TRIGGERS)	- 109 -
3.7.2.	BASE DE DATOS DESTINO (DATAWAREHOUSE).....	- 110 -
3.7.2.1.	ESPACIO DE TABLA (TABLESPACE).....	- 110 -
3.7.2.2.	TABLAS	- 110 -
3.7.2.3.	ÍNDICES.....	- 110 -
3.7.2.4.	RESTRICCIONES (CONSTRAINTS)	- 110 -
3.7.2.5.	SECUENCIA	- 111 -
3.7.2.6.	DISPARADORES (TRIGGERS)	- 111 -
3.8.	DISEÑO Y DESARROLLO DEL PROCESO ETL.....	- 111 -
3.8.1.	PASO 1: TRAZAR UN PLAN DE ALTO NIVEL	- 111 -
3.8.2.	PASO 2: FUNCIONALIDAD DE LA HERRAMIENTA ETL	- 120 -
3.8.3.	PASO 3: DESARROLLAR ESTRATEGIAS POR DEFECTO	- 120 -
3.8.4.	PASO 4: PROFUNDIZAR TABLA DESTINO	- 124 -
3.8.5.	PASO 5: POBLAR LAS TABLAS DE DIMENSIONES CON LOS DATOS HISTÓRICOS.....	- 128 -
3.8.5.1.	Extracción de Datos.....	- 128 -
3.8.5.2.	Trasformaciones.....	- 129 -
3.8.5.2.1.	Valores Nulos	- 129 -
3.8.5.2.2.	Agregar registro sin Valor a las Dimensiones.....	- 133 -
3.8.5.2.3.	Decodificación de datos Origen	- 137 -
3.8.5.3.	Carga de Datos	- 139 -
3.8.5.4.	Poblar Dimensión Estática	- 139 -
3.8.5.5.	Desarrollo ETL Inicial de Dimensiones con la Herramienta de Pentaho	- 140 -
3.8.6.	PASO 6: REALIZAR LA CARGA HISTÓRICA DE LA TABLA DE HECHOS	- 146 -
3.8.6.1.	Extracción de Datos.....	- 146 -
3.8.6.2.	Transformaciones.....	- 146 -
3.8.6.2.1.	Valores Nulos	- 146 -
3.8.6.2.2.	Búsqueda de Claves Subrogadas de Dimensiones	- 148 -
3.8.6.3.	Carga de Datos	- 150 -
3.8.6.4.	Desarrollo ETL Inicial de Hechos con la Herramienta de Pentaho.	- 150 -
3.8.7.	PASO 7: PROCESAMIENTO INCREMENTAL DE LA DIMENSIONES	- 155 -
3.8.7.1.	Identificar Filas nuevas y con cambios	- 156 -
3.8.7.2.	Extracción de Claves Subrogadas de Dimensiones	- 156 -
3.8.7.3.	Desarrollo ETL Incremental de Dimensiones con la Herramienta de Pentaho. ..	- 157 -
3.8.8.	PASO 8: PROCESAMIENTO INCREMENTAL DE LOS HECHOS.....	- 163 -
3.8.8.1.	Identificar Filas Nuevas y con Cambios	- 163 -
3.8.8.2.	Extracción de Claves Subrogadas de Filas de cambio	- 164 -
3.8.8.3.	Desarrollo ETL Incremental de Hechos con la Herramienta de Pentaho	- 164 -
3.8.9.	PASO 9: OPERACIÓN Y AUTOMATIZACIÓN DEL SISTEMA ETL	- 169 -
3.8.9.1.	Trabajos ETL	- 169 -
3.8.9.2.	Programar Trabajos.....	- 171 -
3.8.9.3.	Desarrollo ETL de los Trabajos con la Herramienta de Pentaho	- 172 -
3.9.	DISEÑO Y DESARROLLO DE APLICACIÓN DE BI	- 176 -
3.9.1.	TIPO DE APLICACIÓN DE BI	- 176 -
3.9.2.	PREPARAR EL DESARROLLO DE LA APLICACIÓN	- 176 -
3.9.2.1.	POBLAR LOS METADATOS DE LA HERRAMIENTA DE BI	- 176 -
3.9.2.2.	CREAR LOS METADATOS DE NEGOCIO.....	- 176 -
3.9.2.3.	TESTEAR LA HERRAMIENTA DE BI	- 177 -
3.9.3.	CREAR LAS APLICACIONES	- 177 -
3.9.4.	DESARROLLO DE APLICACIÓN DE BI CON LA HERRAMIENTA DISCOVERER DE ORACLE .-	- 177 -
-		
3.10.	DESPLIEGUE.....	- 180 -
3.10.1.	DESPLIEGUE ETL	- 182 -
3.10.2.	DESPLIEGUE BI	- 184 -
3.10.2.1.	Inventarios	- 184 -

3.10.2.2.	Puntos de Venta.....	- 191 -
3.11.	MANTENIMIENTO.....	- 194 -
3.11.1.	FRONT ROOM	- 194 -
3.11.1.1.	Mantenimiento de la Aplicación	- 194 -
3.11.1.2.	Manejo de Seguridad	- 195 -
3.11.2.	BACK ROOM.....	- 196 -
3.11.2.1.	Monitorear el Sistema ETL	- 196 -
3.11.2.2.	Administración de Espacio en Disco.....	- 197 -
3.11.2.3.	Respaldo y Recuperación	- 197 -
3.12.	CRECIMIENTO.....	- 198 -
3.12.1.	EVALUAR EL ENTORNO ACTUAL	- 199 -
3.12.2.	PRIORIZAR OPORTUNIDADES PARA EL CRECIMIENTO.....	- 199 -
3.12.2.1.	PRIORIDAD A PEQUEÑAS MEJORAS	- 199 -
3.12.2.2.	PRIORIDAD A MAYORES INICIATIVAS.....	- 199 -
3.12.3.	GESTIONAR EL CRECIMIENTO ITERATIVO	- 200 -
CAPITULO IV: CONCLUSIONES Y RECOMENDACIONES.....		- 201 -
CAPÍTULO V: BIBLIOGRAFÍA Y FUENTES DE CONSULTA		- 207 -

LISTADO DE TABLAS

Tabla 3.1: Indicador de Cumplimiento.....	68
Tabla 3.2: Comparación OWB vs PDI.....	74
Tabla 3.3: Esquemas – Base de datos Origen.....	81
Tabla 3.4: Valores Nulos – Dimensión Área.....	129
Tabla 3.5: Valores Nulos – Dimensión Artículo.....	129
Tabla 3.6: Valores Nulos – Dimensión Bodega.....	130
Tabla 3.7: Valores Nulos – Dimensión Centro de Gestión.....	130
Tabla 3.8: Valores Nulos – Dimensión Cliente.....	130
Tabla 3.9: Valores Nulos – Dimensión Empleado.....	131
Tabla 3.10: Valores Nulos – Dimensión Factura.....	131
Tabla 3.11: Valores Nulos – Dimensión Ítem.....	131
Tabla 3.12: Valores Nulos – Dimensión Movimiento.....	131
Tabla 3.13: Valores Nulos – Dimensión Orden de Compra.....	132
Tabla 3.14: Valores Nulos – Dimensión Pedido.....	132
Tabla 3.15: Valores Nulos – Dimensión Proforma.....	132
Tabla 3.16: Valores Nulos – Dimensión Proveedor.....	132
Tabla 3.17: Valores Nulos – Dimensión Proyecto.....	133
Tabla 3.18: Valores Nulos – Dimensión Sucursal.....	133
Tabla 3.19: Registro sin valor – Dimensión Área.....	133
Tabla 3.20: Registro sin valor – Dimensión Artículo.....	134
Tabla 3.21: Registro sin valor – Dimensión Bodega.....	134
Tabla 3.22: Registro sin valor – Dimensión Centro de Gestión.....	134
Tabla 3.23: Registro sin valor – Dimensión Cliente.....	134
Tabla 3.24: Registro sin valor – Dimensión Empleado.....	135
Tabla 3.25: Registro sin valor – Dimensión Factura.....	135
Tabla 3.26: Registro sin valor – Dimensión Ítem.....	135
Tabla 3.27: Registro sin valor – Dimensión Movimiento.....	135
Tabla 3.28: Registro sin valor – Dimensión Orden de Compra.....	136
Tabla 3.29: Registro sin valor – Dimensión Pedido.....	136
Tabla 3.30: Registro sin valor – Dimensión Proforma.....	136
Tabla 3.31: Registro sin valor – Dimensión Proveedor.....	136

Tabla 3.32: Registro sin valor – Dimensión Proyecto.....	137
Tabla 3.33: Registro sin valor – Dimensión Sucursal.....	137
Tabla 3.34: Decodificación Factura.....	137
Tabla 3.35: Decodificación Ítem.....	138
Tabla 3.36: Decodificación Orden de Compra.....	138
Tabla 3.37: Decodificación Pedido.....	138
Tabla 3.38: Decodificación Proforma.....	139
Tabla 3.39: Decodificación y Formato Dimensión Tiempo.....	139
Tabla 3.40: Valores Nulos – Hecho Pedidos.....	147
Tabla 3.41: Valores Nulos – Hecho Movimientos.....	147
Tabla 3.42: Valores Nulos – Hecho Ventas.....	147
Tabla 3.43: Valores Nulos – Hecho Número de Ventas.....	148
Tabla 3.44: Búsqueda clave subrogada – Hecho Pedidos.....	149
Tabla 3.45: Búsqueda clave subrogada – Hecho Movimientos.....	149
Tabla 3.46: Búsqueda clave subrogada – Hecho Ventas.....	149
Tabla 3.47: Búsqueda clave subrogada – Número de Ventas.....	150

LISTADO DE FIGURAS

Figura 2.1: Objetivo de BI.....	7
Figura 2.2: Arquitectura de BI.....	9
Figura 2.3: Proceso de BI.....	10
Figura 2.4: Arquitectura de un Data Warehouse.....	17
Figura 2.5: Estructura de un Data Warehouse.....	19
Figura 2.6: Proceso ETL.....	21
Figura 2.7: Subproceso ETL – Extracción.....	22
Figura 2.8: Subproceso ETL – Transformación.....	23
Figura 2.9: Subproceso ETL – Carga.....	23
Figura 2.10: Esquema Estrella.....	25
Figura 2.11: Esquema Copo de Nieve.....	26
Figura 2.12: Tabla de Hecho.....	27
Figura 2.13 - Tablas de Dimensiones.....	28
Figura 2.14: Niveles en la Dimensión.....	29
Figura 2.15 - Jerarquía en la Dimensión.....	29
Figura 2.16: Relaciones en un Modelo.....	31
Figura 2.17: OLAP (On-Line Analytical Processing).....	35
Figura 2.18: Cubo OLAP.....	37
Figura 2.19: Operación OLAP - Drill Down.....	38
Figura 2.20: Operación OLAP - Roll Up.....	38
Figura 2.21: The Kimball Lyfecycle Diagram.....	39
Figura 2.22: DataWarehouse Bus Matrix.....	50
Figura 3.1: Back Room – Proyecto.....	71
Figura 3.2: Front Room – Proyecto.....	72
Figura 3.3: Arquitectura - Pentaho Data Integration.....	75
Figura 3.4: Arquitectura Discoverer.....	77
Figura 3.5: Modelo Dimensional – Inventarios.....	104
Figura 3.6: Modelo Dimensional - Punto de Ventas.....	105
Figura 3.7: Modelo Dimensional - Data Warehouse.....	106
Figura 3.8: Plan ETL – Área.....	111

Figura 3.9: Plan ETL – Artículo.....	111
Figura 3.10: Plan ETL – Bodega.....	112
Figura 3.11: Plan ETL – Centro de Gestión.....	112
Figura 3.12: Plan ETL – Cliente.....	112
Figura 3.13: Plan ETL – Empleado.....	112
Figura 3.14: Plan ETL – Factura.....	113
Figura 3.15: Plan ETL – Ítem.....	113
Figura 3.16: Plan ETL – Movimiento.....	113
Figura 3.17: Plan ETL – Orden de Compra.....	113
Figura 3.18: Plan ETL – Pedido.....	114
Figura 3.19: Plan ETL – Proforma.....	114
Figura 3.20: Plan ETL – Proyecto.....	114
Figura 3.21: Plan ETL – Proveedor.....	115
Figura 3.22: Plan ETL – Sucursal.....	115
Figura 3.23: Plan ETL – Tiempo.....	115
Figura 3.24: Plan ETL – Pedidos.....	116
Figura 3.25: Plan ETL – Movimientos.....	117
Figura 3.26: Plan ETL – Ventas.....	118
Figura 3.27: Plan ETL – Numero de Ventas.....	119
Figura 3.28: Arquitectura CDC (Change Data Capture).....	121
Figura 3.29: Secuencia de las Transformaciones.....	123
Figura 3.30: Auditoria – Error ETL.....	124
Figura 3.31: Transformación Dimensión Área.....	140
Figura 3.32: Transformación Dimensión Artículo.....	140
Figura 3.33: Transformación Dimensión Bodega.....	141
Figura 3.34: Transformación Dimensión Centro de Gestión.....	141
Figura 3.35: Transformación Dimensión Cliente.....	141
Figura 3.36: Transformación Dimensión Empleado.....	142
Figura 3.37: Transformación Dimensión Factura.....	142
Figura 3.38: Transformación Dimensión Ítem.....	142
Figura 3.39: Transformación Dimensión Movimiento.....	143
Figura 3.40: Transformación Dimensión Orden de Compra.....	143

Figura 3.41: Transformación Dimensión Pedido.....	143
Figura 3.42: Transformación Dimensión Proforma.....	144
Figura 3.43: Transformación Dimensión Proveedor.....	144
Figura 3.44: Transformación Dimensión Proyecto.....	145
Figura 3.45: Transformación Dimensión Sucursal.....	145
Figura 3.46: Transformación Dimensión Tiempo.....	145
Figura 3.47: Transformación Hecho Pedidos.....	151
Figura 3.48: Transformación Hecho Movimientos.....	152
Figura 3.49: Transformación Hecho Ventas.....	153
Figura 3.50: Transformación Hecho Número de Ventas.....	154
Figura 3.51: Funcionamiento CDC – Dimensiones.....	156
Figura 3.52: Agregar Clave Subrogada.....	157
Figura 3.53: Transformación Dimensión Área.....	157
Figura 3.54: Transformación Dimensión Artículo.....	158
Figura 3.55: Transformación Dimensión Bodega.....	158
Figura 3.56: Transformación Dimensión Centro de Gestión.....	158
Figura 3.57: Transformación Dimensión Cliente.....	159
Figura 3.58: Transformación Dimensión Empleado.....	159
Figura 3.59: Transformación Dimensión Factura.....	160
Figura 3.60: Transformación Dimensión Ítem.....	160
Figura 3.61: Transformación Dimensión Movimiento.....	160
Figura 3.62: Transformación Dimensión Orden de Compra.....	161
Figura 3.63: Transformación Dimensión Pedido.....	161
Figura 3.64: Transformación Dimensión Proforma.....	162
Figura 3.65: Transformación Dimensión Proveedor.....	162
Figura 3.66: Transformación Dimensión Proyecto.....	162
Figura 3.67: Funcionamiento CDC – Hechos.....	163
Figura 3.68: Extraer y Agregar Clave Subrogada.....	164
Figura 3.69: Transformación Hecho Pedidos.....	165
Figura 3.70: Transformación Hecho Movimientos.....	166
Figura 3.71: Transformación Hecho Ventas	167
Figura 3.72: Transformación Hecho Numero de Ventas.....	168

Figura 3.73: Verificar Dimensiones.....	169
Figura 3.74: Verificar Hechos.....	169
Figura 3.75: Poblar Dimensiones.....	170
Figura 3.76: Poblar Hechos.....	170
Figura 3.77: Poblar DataWarehouse.....	171
Figura 3.78: Trabajo Verificar Dimensiones.....	172
Figura 3.79: Trabajo Verificar Hechos.....	173
Figura 3.80: Trabajo Poblar Dimensiones.....	174
Figura 3.81: Trabajo Poblar Hechos.....	175
Figura 3.82: Trabajo Poblar DataWarehouse.....	175
Figura 3.83: Selección de Dimensiones para EUL.....	178
Figura 3.84: EUL's de Inventarios y Puntos de Venta.....	178
Figura 3.85: Campos ocultos de Dimensiones.....	179
Figura 3.86: Niveles y Jerarquía Dimensión Tiempo.....	180
Figura 3.87: Resultado de Script de creación de repositorio.....	182
Figura 3.88: Progreso de metadatos (transformación) guardados.....	183
Figura 3.89: Catálogo de Transformaciones y Trabajos.....	183
Figura 3.90: Ejecución Inicial.....	184
Figura 3.91: Ingresos Anuales.....	185
Figura 3.92: Ingresos Mensuales.....	185
Figura 3.93: Gráfico de barras Ingresos Anuales.....	186
Figura 3.94: Gráficos de tendencia de Ingresos Mensuales.....	186
Figura 3.95: Consumo Anual.....	187
Figura 3.96: Gráfico de tendencia del Consumo Anual.....	187
Figura 3.97: Ingresos por Proveedor con gráfico de barras.....	188
Figura 3.98: Movimiento de Ingreso con gráfico de barras.....	188
Figura 3.99: Egresos a Bodega con gráfico de barras.....	189
Figura 3.100: Existencia de Grupo de Artículo con gráfico de barras....	189
Figura 3.101: Movimientos por Grupo de Artículo con gráfico de barras.	190
Figura 3.102: Compras por Proveedor en orden descendente con gráfico de barras.....	190
Figura 3.103: Pedidos por Grupo de Artículo, Área y Proyectos con	191

gráfico de barras.....	
Figura 3.104: Nivel de Cumplimiento de Empleados.....	192
Figura 3.105: Total en Ventas por Sucursal con gráfico de barras.....	192
Figura 3.106: Cantidad de Ítems vendidos por Cliente con gráfico de barras.....	193
Figura 3.107: Número de Ventas por Sucursal con gráfico de barras....	193
Figura 3.108: Permiso de Acceso por usuario al EUL.....	196
Figura 3.109: Respaldo de Transformaciones y Trabajos como Archivos *.ktr y *.kjb.....	198

LISTADO DE ANEXOS

ANEXO A: Flujo de Proceso de Inventarios.....	209
ANEXO B: Esquema de Inventarios.....	210
ANEXO C: Esquema de Puntos de Venta.....	211
ANEXO D: Tablas Esuqemas: Financiero, Proyectos, Recursos Humanos.....	212
ANEXO E: Scripts – DataWarehouse.....	213
ANEXO F: Scripts – Base de Datos Origen.....	213
ANEXO G: Scripts – Extracción de Datos origen.....	213
ANECO H: Scripts – Extracción de Datos Destino para carga Incremental.....	213
ANEXO I: Diccionario de Datos.....	213

RESUMEN

El proyecto de tesis de Ingeniería está orientado a proporcionar una herramienta informática que permita soportar la toma de decisiones mediante la generación de información analítica y de manera oportuna dentro del Instituto Geográfico Militar.

Tomando en cuenta la problemática mencionada, se desarrollará un ambiente de DataWarehouse con el fin de suplir estas falencias, se abarcará los negocios de Inventarios y Puntos de Venta, dirigido a los directivos de los mismos.

Para la Metodología de desarrollo se estableció usar la propuesta de Ralph Kimball para este tipo de proyectos de implementación de DataWarehouse/Business Intelligence, debido a que guía y permite el desarrollo durante todo el ciclo de vida empezando desde la planificación del proyecto hasta la finalización, en conjunto con el Mantenimiento y Administración del Proyecto.

Para el desarrollo se convino como plataforma de Base de Datos Oracle 10g, para el proceso ETL la herramienta libre (Open Source) Pentaho Data Integration (Kettle), y para la parte de interfaz con el usuario de negocio Oracle Discoverer Plus 9.

Con la implementación del concepto de DataWarehouse y Business Intelligence dentro del Instituto Geográfico Militar se está inculcando un ambiente Gerencial dentro de la Institución, para una administración eficaz de los directivos encargados de negocio, obteniendo como producto, información útil del negocio a analizar.

CAPÍTULO I: INTRODUCCIÓN

1.1. ANTECEDENTES

El Instituto Geográfico Militar del Ecuador es una institución técnica y científica, encargada de la elaboración de la cartografía nacional y del archivo de datos geográficos de nuestro país.

El Instituto Geográfico Militar cuenta con varios Departamentos los cuales contribuyen a la institución, en ellas se maneja información que ayuda a la toma de decisiones por parte de los Gerentes y Directivos de cada área.

Para el apoyo de toma de decisiones los Gerentes y Directivos, requieren Información de sus distintas áreas y estas a su vez pedir al Departamento de Sistemas de la Institución lo cual es un proceso complejo para el sencillo efecto de contar con información precisa y resumida, este proceso produce pérdida de tiempo, información no oportuna, y la falta de disponibilidad del Departamento de Sistemas para apoyar a los Gerentes y Directivos que requieren la Información para soportar los Procesos de la Institución.

1.2. PLANTEAMIENTO DEL PROBLEMA

Dentro del Instituto Geográfico Militar (IGM) se trata de llevar la información de acuerdo a diferentes niveles, Operativo, Táctico, Gerencial, la misma que se necesita llegue a tiempo o que no se necesite realizar peticiones a las áreas involucradas para obtener respuesta. También la institución cuenta con mecanismos que le permiten dar a conocer los datos existentes dentro de la misma como por ejemplo mediante el desarrollo de reportes de acuerdo a una

necesidad específica, pero no existe información para ser presentada a nivel Táctico, Estratégico - Gerencial, por lo que dentro de la Institución existe malestar por parte de la Gerencia y Autoridades al no contar con información útil y necesaria en el momento oportuno, la cual apoya en la gestión de toma de decisiones.

Los datos dentro del Instituto se encuentran divididos de acuerdo a cada área, por lo que no se cuenta con información útil y analítica la cual brinde apoyo a la toma de decisiones por parte de los Directivos y de la Gerencia.

En el Instituto se ha implementado la gestión por procesos, es por esto la importancia de información útil, la cual sea generada de acuerdo a este esquema.

1.3. JUSTIFICACIÓN

Se planteará una solución basada en tecnologías de información mediante la implementación de un almacén de datos (DataWarehouse), para obtener un producto que permita generar Información tratada y útil, la cual apoyará el análisis del negocio (procesos) que se requieran, como son: Punto de Ventas e Inventario, lo que permitirá dar soporte a los Directivos en la toma de decisiones.

Dentro del Instituto se aumentará la capacidad de generar y recoger datos, los cuales serán identificados, categorizados y resumidos para aportar con información analítica, la misma que podrá ser utilizada eficiente y oportunamente; esto apoyará la toma de decisiones permitiendo incrementar la credibilidad institucional.

La solución del DataWarehouse a implementar se la desarrollará utilizando como base la plataforma Oracle.

1.4. OBJETIVOS

1.4.1. OBJETIVO GENERAL

IMPLEMENTAR UN DATA WAREHOUSE PARA APOYAR LA TOMA DE DECISIONES DENTRO DEL INSTITUTO GEOGRÁFICO MILITAR.

1.4.2. OBJETIVOS ESPECÍFICOS

- Implementar el Data Warehouse para generar información analítica de Puntos de Venta e Inventarios.
- Utilizar la metodología de Ralph Kimball como guía para el desarrollo del Proyecto.
- Utilizar herramientas ORACLE compradas por el IGM, como base tecnológica para el desarrollo.
- Implementar herramienta(s) libre en el desarrollo por motivo del Decreto Ejecutivo Nro. 1014 emitido el 10 de Abril de 2008 por la Presidencia de la República del Ecuador, que establece que toda Institución Pública debe utilizar Software Libre en sus sistemas informáticos.

1.5. ALCANCE

El alcance del presente Proyecto, es la implementación de un Data Warehouse para el Instituto Geográfico Militar.

El producto a implementar abarca:

- Inventarios.
- Puntos de venta.

Se realizara la implementación del Data Warehouse capaz de soportar las necesidades de información de los 2 negocios.

1.6. HIPÓTESIS DE TRABAJO

La implementación de un Data Warehouse como parte de una solución de Business Intelligence brindará información útil, disponible, analítica y oportuna, a nivel Táctico, Estratégico - Gerencial del IGM, lo que permitirá el apoyo a la toma de decisiones, en beneficio de la Institución.

CAPÍTULO II: MARCO TEÓRICO

2.1. BUSINESS INTELLIGENCE

2.1.1. INTRODUCCIÓN.

Hoy en día en las Empresas y Organizaciones, la Gestión de información es un elemento clave para poder competir de mejor manera en un mercado muy cambiante, y de una sociedad de negocio. La información que se maneja es de vital importancia, porque en ella se fundamenta la toma de decisiones, sirviendo de pilar para el crecimiento y gestión, como también de guía en la consecución de los objetivos que persiguen las Empresas.

Además cada vez es más importante saber qué es lo que sucede con los negocios que en ellas se gestiona, y de cómo mejorarlas.

Por otro lado existen Organizaciones las cuales poseen un sin número de datos, pero que a su vez no cuentan con información, esto se debe por el gran problema de que no saben qué hacer ni cómo utilizar esos datos en beneficio propio, y como resultado, hoy en día gran parte del sector empresarial no sabe que cuentan con información valiosa capaz de ser utilizada para mejorar la competitividad del negocio.

Aquí es donde nace el concepto de Business Intelligence o Inteligencia de Negocios, como solución para el soporte de toma de decisiones, mediante la generación y análisis de información, y con la que se busca el conocimiento organizacional.

2.1.2. DEFINICIÓN.

Business Intelligence (*BI*) o también conocido mediante su traducción al español como Inteligencia de Negocio, pretende dar una visión de información y conocimiento en las Organizaciones.

Para definir Business Intelligence se partirá de la definición del glosario de términos de Gartner:

“Es una variedad de aplicaciones y tecnologías para obtener, almacenar, analizar, compartir y proveer acceso a los datos para ayudar a los usuarios de las empresas a tomar mejores decisiones.”¹

Tomando como base la definición anterior y después de haber analizado el significado de Business Intelligence, se cita una definición más general y desde un punto de vista más práctico:

Business Intelligence o Inteligencia de Negocio es un proceso, el cual plantea soluciones de negocio, utilizando un conjunto de metodologías, aplicaciones y tecnología, capaz de transformar los datos e información desestructurada, en información estructurada, generando conocimiento para soportar la toma de decisiones en beneficio del negocio y de la Organización.

2.1.3. IMPORTANCIA Y BENEFICIO DE BUSINESS INTELLIGENCE EN LA ORGANIZACIÓN.

Como se mencionó anteriormente en la introducción, la gran mayoría de las empresas carecen de información y por ende de conocimiento, por lo que esto puede ser considerado como una gran amenaza para las organizaciones

¹ Glosario de Gartner, BI (Business Intelligence), www.gartner.com, Junio 2009, traducido del idioma inglés al español. Gartner es una Consultora internacional especializada en Tecnologías de Información y Comunicación.

modernas. En la ejecución de las actividades de los negocios se depende mucho de las decisiones que se tomen, para que garantice la eficiencia del trabajo. Es por eso que el factor más importante que persiguen las Organizaciones, es contar con el conocimiento necesario sobre los negocios y de la(s) necesidad(es) de explotar sus datos, por lo que tratan de buscar mecanismos que les ayuden en la consecución de estos objetivos; para ello las Empresas apuntan a *Business Intelligence* que tiene por objetivo primordial, eliminar la falta de información y de conocimiento en los ambientes empresariales, aumentando su desempeño y competitividad, como se muestra en la Figura Nro. 2.1.

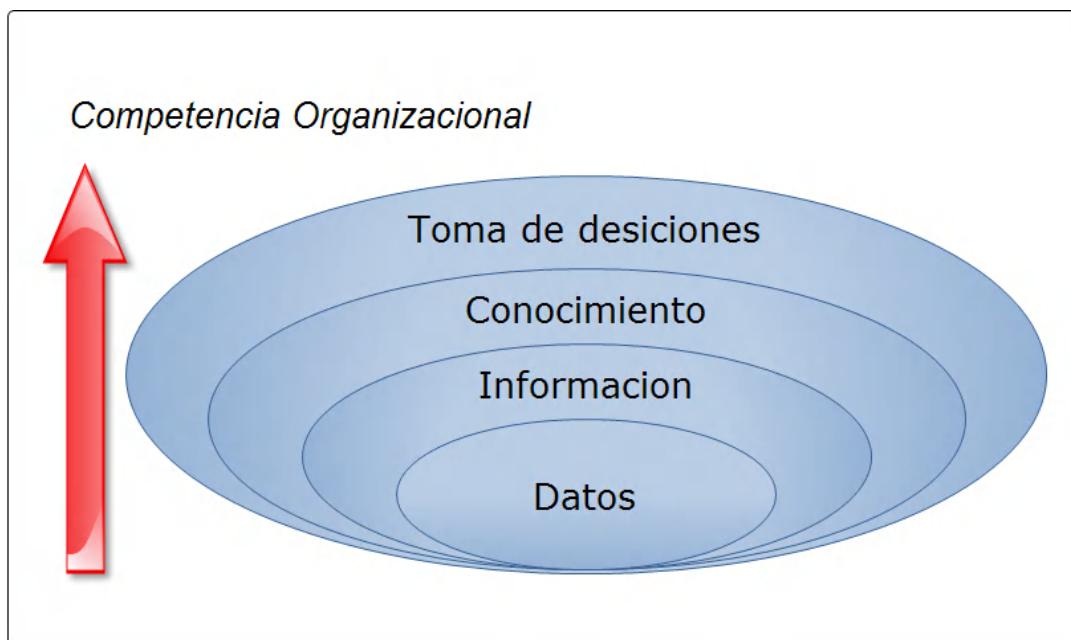


Figura 2.1: Objetivo de BI

Dentro de los beneficios que conlleva la implementación de una solución de este tipo, el principal beneficio es que ayuda a tomar decisiones en función de la información que produce una aplicación de BI.

A continuación se darán unos ejemplos de beneficios que suelen ser producto de la implementación de Business Intelligence en las Organizaciones, entre las cuales la mejora de competitividad es punto a resaltar.

- Beneficios Tangibles
 - Aumentar Ingresos de Ventas.
 - Aumentar compras de Artículos de mayor uso o necesidad.
 - Eliminar sobreproducción de bienes.
 - Aumentar producción de bienes con más Ventas.
 - Aumento en la información de calidad con tiempos de respuesta cortos.
 - Disminuir el gasto innecesario de productos.
- Beneficios Intangibles
 - Menor dependencia a los sistemas que se desarrolle.
 - Mayor conocimiento de los empleados sobre la Organización.
 - Mejora de la atención al cliente.
 - Aumento en el control de costos.
 - Aprender de los errores que se han cometido en el pasado.
- Beneficios Estratégicos
 - Aumento por parte de los Directivos en la visión del(los) negocio(s).
 - Aumento en la gestión de los Directivos.
 - Incremento de habilidades de los Directivos en la toma de decisiones.

2.1.4. ARQUITECTURA DE BUSINESS INTELLIGENCE

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

muestra en la Figura 2.2, la Arquitectura de una solución de Business Intelligence.

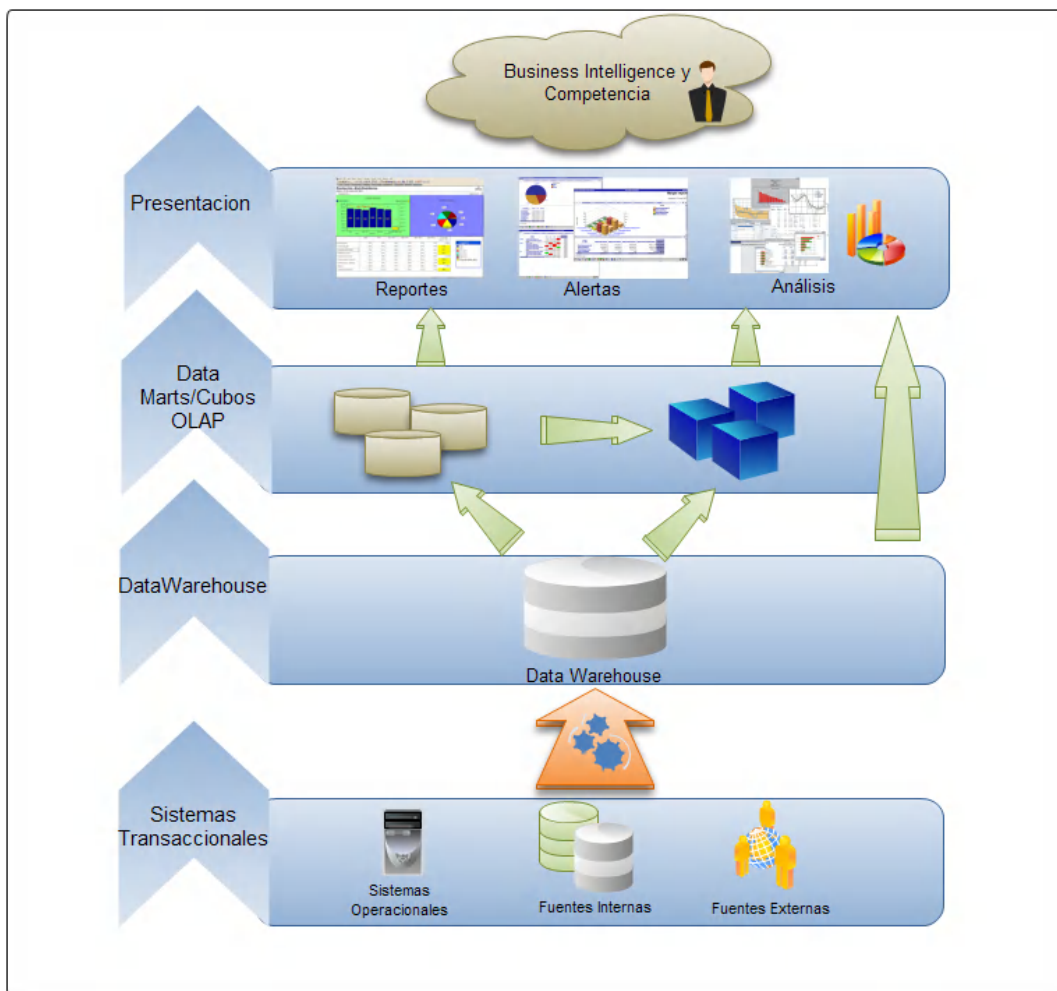


Figura 2.2: Arquitectura de BI

Como se observa en la figura anterior una solución de Business Intelligence inicia desde los sistemas de origen de una organización (bases de datos, ERPs, archivos de texto...), sobre los que suele ser necesario aplicar una transformación estructural para optimizar su proceso analítico (datos de calidad), posteriormente estos datos son almacenados, después pueden ser estructurados específicamente y finalmente ser presentados de diferente manera.

2.1.5. ETAPAS DE BUSINESS INTELLIGENCE

El proceso de Business Intelligence garantiza un ciclo, que permite generar información analítica constante, es decir obtener la información oportuna que el usuario busca.

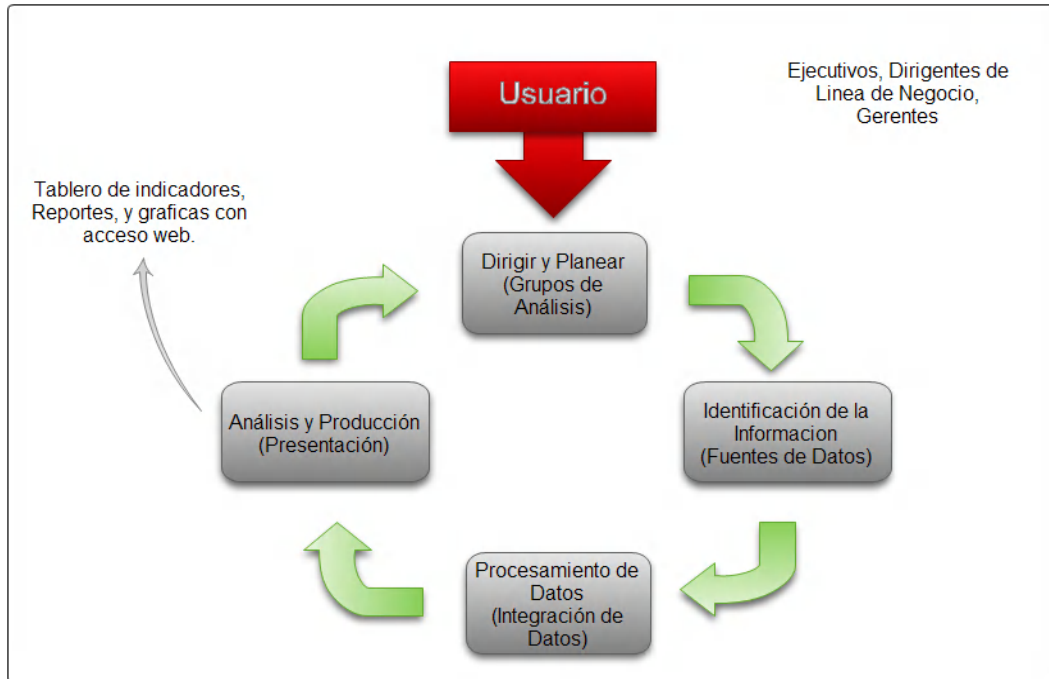


Figura 2.3: Proceso de BI²

2.1.5.1. DIRIGIR Y PLANEAR

Como se observo en la Figura 2.3, la primera etapa es en la cual se inicia y se finaliza el proceso de Business Intelligence mediante la continuidad de la misma y según las necesidades Organizacionales, la cual se inicia desde la redacción de las necesidades de los usuarios y/o Organización. Estas necesidades son planteadas o generadas en el análisis a preguntas como por ejemplo: ¿Qué producto se vendió más en el primer semestre?, ¿Cuál es el margen de ganancia entre un mes y el otro?.

² Lee Wittschen, 2004

El análisis de las preguntas, produce la redacción de necesidades que se plasmarán en una solución de BI.

Así mismo al finalizar un ciclo del proceso de BI, se obtiene la respuesta a preguntas que van a ser sujetas a análisis por parte de los usuarios, y las cuales servirán de soporte a la toma de decisiones.

2.1.5.1.1. TOMA DE DECISIONES

Definición

Es una actividad intelectual realizada por todas las personas para seleccionar una acción u objeto entre varias opciones o alternativas, en beneficio de resolver diferentes situaciones que se presenten.

Las decisiones pueden ser:

Estructuradas

También son conocidas como programadas, ya que son decisiones repetitivas, lo que se convierte en rutinas del día a día.

Por lo general los Gerentes que realizan con frecuencia este tipo de decisiones no necesitan diseñar algún tipo de solución que les ayuden, por el hecho que son decisiones de manera rutinaria y simplemente se rigen a acciones establecidas en una Organización por ejemplo reordenar inventario mínimo.

No Estructuradas

También conocidas como no programadas, debido a que las decisiones tomadas se presentan de una manera no rutinaria y no se puede predecir ni predefinir las necesidades de información. Tienen que ver con problemas complejos que involucran toda la corporación. Ejemplo: Aumento de producción de un artículo en función de ser el de mayor venta a clientes.

En esta primera etapa se procede también a realizar el planeamiento del proyecto de BI y tomar puntos en consideración como por ejemplo:

- Si se cuenta o no con el patrocinador del negocio.
- Se cuenta con el apoyo de las personas interesadas a las cuales va a llegar una solución de BI.
- Se cuenta con los recursos necesarios para una solución de BI.
- Considerar el tiempo necesario de acuerdo al negocio, en el que se pretende proponer la solución.

2.1.5.2. IDENTIFICACIÓN DE INFORMACIÓN

Para empezar con la segunda etapa se debe tomar en cuenta lo planeado en la etapa anterior. Posterior al planeamiento se inicia con lo que son los requerimientos del negocio, como:

Requisitos Funcionales:

- ¿Qué tipo de información necesitan las personas inmersas al negocio?
- ¿En base a que reportes las personas realizan los análisis del negocio?
- ¿Qué tipo de reportes son los que desean?

Requisitos de Datos:

- ¿Qué datos del negocio necesitan?
- ¿Qué datos son considerados críticos para el negocio?
- ¿Hasta qué nivel general y de detalle es requerido los datos?

Requisitos Históricos:

- ¿Desde qué año necesitan ser almacenados los datos?

Finalizado la toma de requerimientos, se da énfasis a la recolección de los datos y a la identificación de la misma en las fuentes de los diferentes sistemas dentro de la organización, y las que van a ser analizadas, aquí es importante tomar en cuenta la etapa anterior desde el punto de vista de cuales fueron las necesidades del negocio planteadas, y poder dar inicio a la identificación de los datos.

Para poder identificar los datos es recomendable tomar en cuenta puntos como:

- Saber donde reside los datos que van a ser analizados, en qué Sistema, en que base de datos o archivos.
- Se cuenta con manuales o diccionarios de datos de los negocios a los que se va a analizar.
- Saber que persona se encuentra encargada sobre el negocio, es decir quien tiene autoridad sobre esas fuentes de datos.
- Saber que tan limpio se encuentran los datos en las fuentes de origen.
- Saber identificar como se evidencia las reglas de negocio en los sistemas y base de datos.

2.1.5.3. PROCESAMIENTO DE DATOS

En esta etapa se da lugar al proceso de integración de datos, aquí se analiza, diseña y determina en profundidad los datos relevantes, las transformaciones a realizar, y las estructuras de destino como es un Data Warehouse (Almacén de datos).

La integración de los datos se lo realiza mediante un proceso principal que es llamado ETL (Extracción, Transformación y Carga), el cual es considerado

como base para la calidad de información con que se va a contar al final de una implementación de Business Intelligence, ya que se extraerán los datos debidamente identificados y los necesarios de acuerdo al negocio, posteriormente se realizará la transformación de datos erróneos, inconsistentes alojados en diferentes sistemas transaccionales, para que finalmente estos datos sean cargados en un DataWarehouse.

2.1.5.4. ANÁLISIS Y PRODUCCIÓN

En esta etapa se da lugar al análisis de la información, que va a ser generada a partir de los datos que se encuentran almacenados de una manera estructurada, de acuerdo al negocio especificado en la etapa de planeamiento.

Aquí se podrá visualizar el resultado de Business Intelligence mediante la presentación de la información del negocio, como es por ejemplo mediante la aplicación de reportes, tableros o cuadros de mando, así como también de análisis de la información como son los cubos OLAP, o minería de datos (Data mining) en busca de tendencia o patrones del negocio.

2.2. DATA WAREHOUSE

2.2.1. INTRODUCCIÓN

Teniendo una mejor perspectiva acerca de Business Intelligence, y mediante la apreciación de la Arquitectura de BI, ya se puede tener una idea del papel que juega un almacén de datos (Data Warehouse) en una solución de este tipo.

Un Data Warehouse es una base de datos accesible por los usuarios el cual tiene un registro de datos históricos y actuales acerca de todas las entidades

importantes que se encuentran en la empresa y de acuerdo a negocios específicos.

El Data Warehouse organiza y aloja los datos necesarios, para ser utilizados en el procesamiento analítico dentro de una perspectiva de tiempo.

Según Ralph Kimball³ y Bill Inmon⁴ un Data Warehouse es:

“Una copia de la transacción de datos específicamente estructurado para consulta y análisis.” (Ralph Kimball, 1996)

“Orientado al tema, integrado, de tiempo variante, de colección de datos no volátil en apoyo a la gestión del proceso de toma de decisiones.” (Bill Inmon, 1995)

El desarrollo de un DataWarehouse facilita información útil como por ejemplo las ventas obtenidas mensualmente, las cuales pueden ser interpretadas mediante herramientas graficas, o la cantidad de Existencia de un artículo.

2.2.2. CARACTERÍSTICAS

A continuación se muestran las características principales de un DataWarehouse en cuanto a sus datos.

- ✓ Orientado al tema.
- ✓ Integrado.
- ✓ De tiempo Variante.
- ✓ No volátil.

³ Ralph Kimball, conocido en todo el mundo como un innovador, escritor, educador, orador y consultor en el campo de almacenamiento de datos. <http://www.kimballgroup.com/html/about.html> - 06 de julio de 2009, traducido del idioma Inglés al Español.

⁴ Bill Inmon, experto de renombre mundial, autor y orador de almacenamiento de datos, es ampliamente reconocido como el "padre de almacenamiento de datos." <http://www.inmoncif.com/about/> - 06 de julio de 2009, traducido del idioma Inglés al Español.

Orientado al tema:

Se debe esta característica debido a que en el DataWarehouse la información se clasifica de acuerdo a aspectos de interés para la empresa, como por ejemplo cliente, vendedor, producto, venta.

Integrado:

En el Data Warehouse la información se encuentra integrada, esta integración puede ser vista a problemas de datos con que se puede encontrar en una base de datos operacional, errores como de inconsistencia de datos, uniformidad, diferente codificación de datos en múltiples fuentes. Con la integración, cualquier tipo de dato será estandarizado de manera general y así será alojado en el almacén.

De tiempo variante:

Se refiere así debido al horizonte de tiempo con que funciona un DataWarehouse.

No Volátil:

Se debe a que el DataWarehouse no sufre las operaciones como inserción, eliminación, modificación, tan solo realizan dos operaciones que son la carga de datos y el acceso a los mismos

Un DataWarehouse como producto presenta las siguientes características:

- ✓ Fácil accesibilidad a la información organizacional.
- ✓ Información sumariada y detallada.

- ✓ Presentación consistente de la información organizacional.
- ✓ Permite realizar análisis rápidamente.

2.2.3. ARQUITECTURA DE UN DATAWAREHOUSE

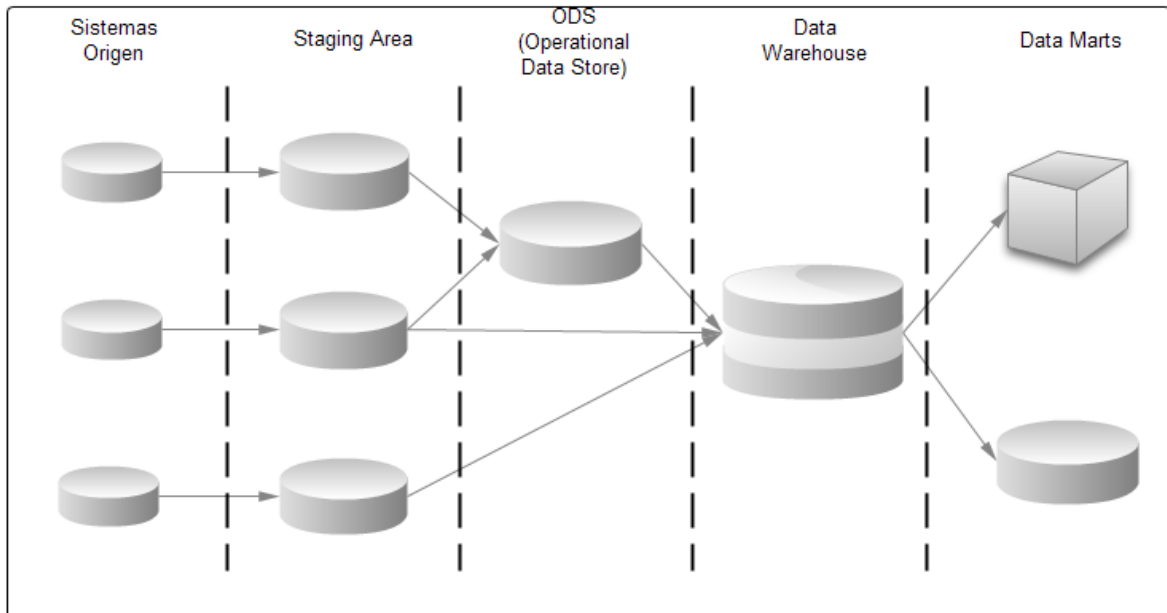


Figura 2.4: Arquitectura de un Data Warehouse⁵

2.2.3.1. SISTEMAS ORIGEN

En cuanto a los Sistemas Origen, son en donde se encuentra los datos de interés y que serán objeto de extracción, estos orígenes pueden ser desde múltiples lugares como por ejemplo las bases de datos, así como también hojas de cálculo, archivos planos, Sistemas ERP (Enterprise Resource Planning), entre otros que serán los que proveerán los datos de interés.

2.2.3.2. STAGING AREA (AREA TEMPORAL)

Es un área en la que se almacenará de manera temporal todos los datos que son extraídos de los sistemas origen según las necesidades planteadas del

⁵ Tomado de <http://www.dataprix.com/arquitectura-data-warehouse-areas-datos-nuestro-almacen-corporativo>

negocio. La función primordial es minimizar la afectación a los sistemas origen, porque cuando se realiza la carga al Data Warehouse, los orígenes de datos no intervendrán hasta que se realice una próxima carga.

2.2.3.3. ODS (OPERATIONAL DATA STORE)

ODS ó Almacén de datos operativos es un área que da soporte a los sistemas operacionales. El modelo de datos del ODS sigue una estructura relacional y normalizada, característica por la cual se la trata de diferente manera a la de un Almacén de Datos (DW), y con lo que le permite que cualquier herramienta de reporte o sistema operacional pueda consultar sus datos. El ODS forma parte de un Data Warehouse, en el sentido de que el Almacén de Datos aprovecha los datos que previamente se encuentran integrados en él, permitiendo también dar soporte a todas las transacciones operacionales.

Los ODS se lo puede considerar como un sistema origen desde el punto de vista que va a ser sujeto a la extracción de los datos que serán cargados en el Data Warehouse. Vale recalcar que si el Almacén no es demasiado grande, o el nivel de exigencia no es muy elevado en cuanto a los requerimientos o consultas operacionales, se puede prescindir de la intervención del ODS.

2.2.3.4. DATA WAREHOUSE

Los datos a ser cargados en el Almacén provendrán del Staging Area así como también puede ser desde el ODS. Aquí los modelos de datos no serán tan normalizados como lo es en los sistemas origen y en el ODS, aquí sucede lo contrario y se realizan técnicas de desnormalización.

2.2.3.5. DATA MARTS

Son considerados pequeños almacenes ya que poseen información específica que se obtiene desde el Data Warehouse, un Data Mart es más personal ya que puede llegar a construirse a partir de las necesidades en particular de un usuario o a un tema en específico, así como también por ejemplo los cubos OLAP sobre cierta información que el usuario requiera, dándole una perspectiva analítica sobre los datos.

2.2.4. ESTRUCTURA

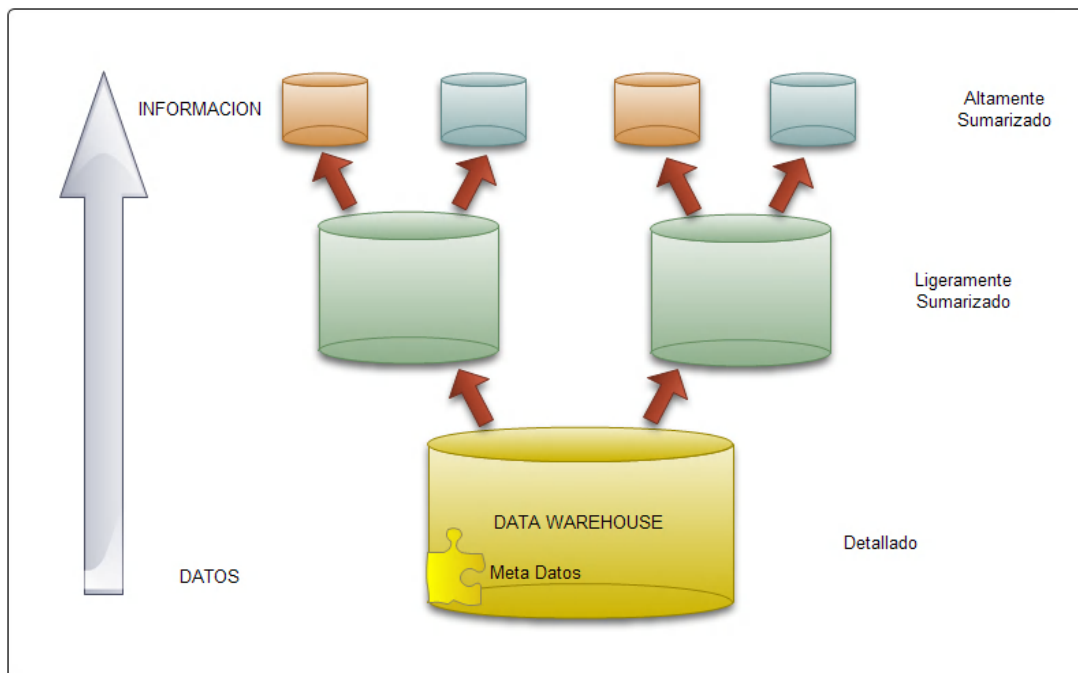


Figura 2.5: Estructura de un Data Warehouse

La estructura de un Data Warehouse, es básicamente referente a la lógica de almacenamiento de los datos, los cuales se caracterizan por tener diferentes niveles de esquematización y de detalle, cómo se muestra en la Figura 2.5, en la que se puede identificar los niveles de datos, además de los metadatos (metadata) del repositorio. Tomando en cuenta estos aspectos, se estructura al Data Warehouse de la siguiente manera:

DATOS

Los datos son claves en un almacén, es lo que genera la información del negocio, estos datos por lo general se encontrarán previamente tratados e integrados.

METADATOS

Son datos estructurados y codificados que describen características de instancias, conteniendo información para ayudar a identificar, descubrir, valorar y administrar las instancias descritas. Estos metadatos por ejemplo contienen:

- Algoritmos utilizados en las transformaciones de datos.
- Mapeo de los datos del proceso de integración de datos.
- Funciones, procedimientos almacenados, vistas q sirven al desarrollo del DataWarehouse.
- Especificaciones y datos propios del repositorio (datos de sistema).

Con lo descrito anteriormente, se puede dar una idea general en la que los metadatos permiten mantener información de la procedencia de los datos, la periodicidad de refresco, su fiabilidad, forma de cálculo, etc., relativa a los datos de nuestro almacén.

Así mismo en cuanto a los datos, poco, ligera y altamente resumizados, esto se puede identificar de mejor manera en ambientes multidimensionales como los cubos OLAP ya que mediante esta implementación, de explotación de un DataWarehouse se puede apreciar, agregados, navegación de niveles, etc.

2.2.5. PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA

Existe un conjunto básico de procesos de suma importancia detrás de una arquitectura de DW, que garantizan la calidad de datos que en ella se almacenarán.

Este proceso de Extracción, Transformación y carga, también conocido como ETL o proceso de integración de datos, cumple con la función principal de organizar e integrar el flujo de datos desde múltiples fuentes, hacia un destino que es el almacén de datos (DataWarehouse). El proceso ETL brinda soporte a la Gestión de Datos que se va a realizar, obteniendo calidad de los mismos dentro de un almacén como se puede apreciar en la Figura Nro. 2.6.

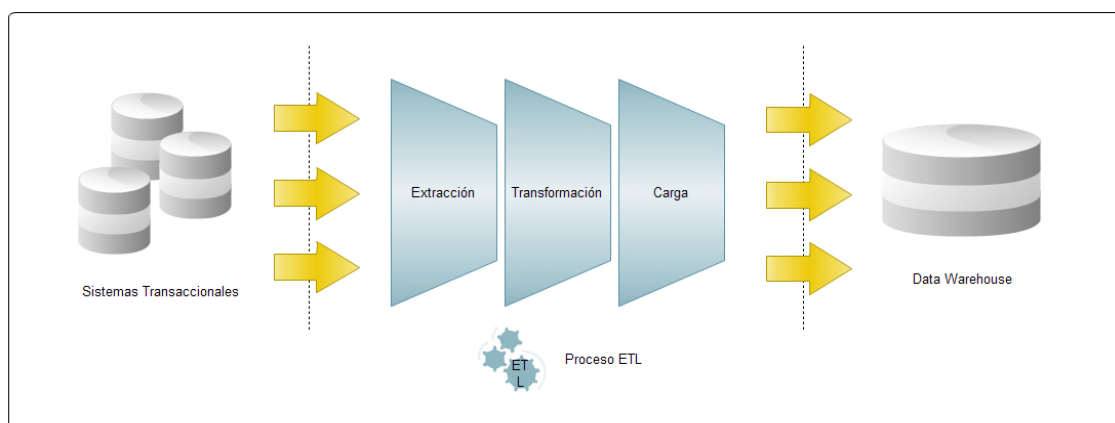


Figura 2.6: Proceso ETL

Todo el proceso que se lleva a cabo, se debe especificar los tiempos en los cuales se deberá realizar el mismo, lo que garantiza que se mantenga al día los datos en el almacén. Aunque esto va a ser definido de acuerdo a las necesidades de la Organización, ya que por ejemplo se puede definir cargas diarias, así como semanales o mensuales.

Este proceso general se encuentra subdividido en 3 subprocesos fundamentales como se detalla a continuación.

2.2.5.1. EXTRACCIÓN

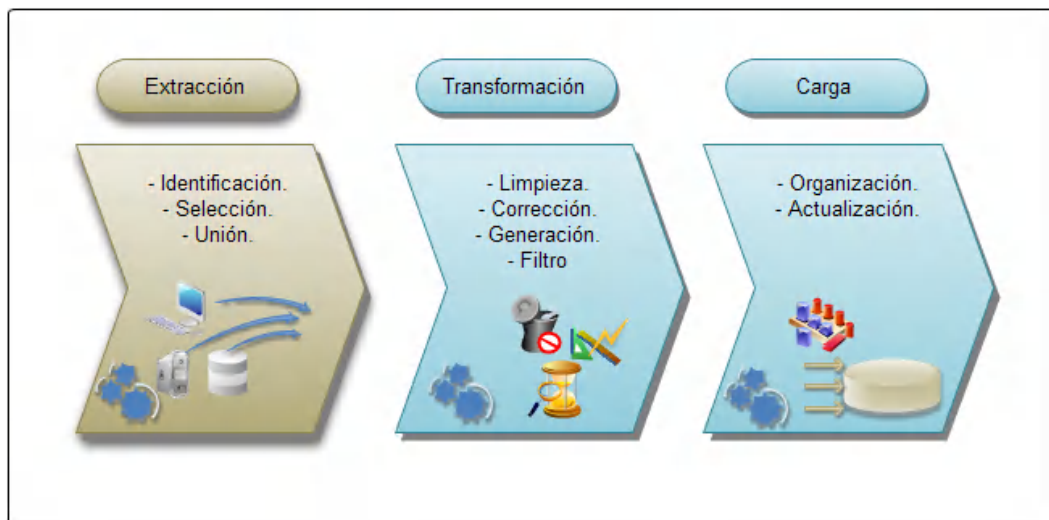


Figura 2.7: Subproceso ETL - Extracción

La extracción se refiere a la adquisición de los datos, los cuales pueden ser recogidos de diferentes fuentes, como archivos planos, hojas de cálculo, ó bases de datos.

La extracción de los datos se almacenarán en una área temporalmente o Staging Area, vale recalcar que solo se extraerán datos necesarios, es decir de acuerdo a lo que se haya especificado en los requerimientos, ya que en el ambiente transaccional se encuentra gran cantidad innecesaria de datos, por lo que es indispensable la extracción de los mismos, y que serán útiles en el ambiente del DataWarehouse.

2.2.5.2. TRANSFORMACIÓN

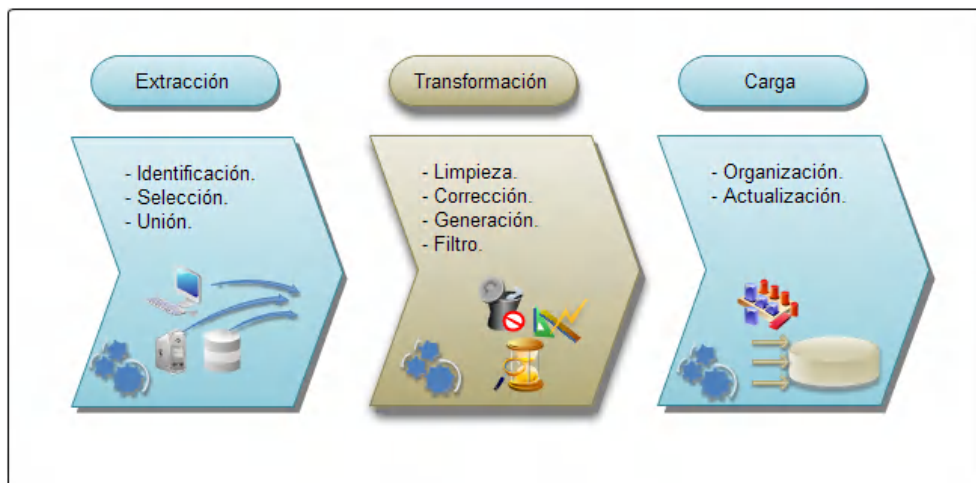


Figura 2.8: Subproceso ETL - Transformación

Es el subproceso más laborioso con respecto a los otros dos, debido a que en esta etapa se realiza el refinamiento de los datos que han sido extraídos de las diferentes fuentes, por lo que aquí se especificará pasos de acuerdo a los datos que van a ser tratados dando valor para los usuarios. Este proceso incluye corrección de errores, decodificación, borrado de campos que no son de interés, generación de claves, agregación de información, etcétera, lo que es más conocido como limpieza de los datos fuentes.

2.2.5.3. CARGA

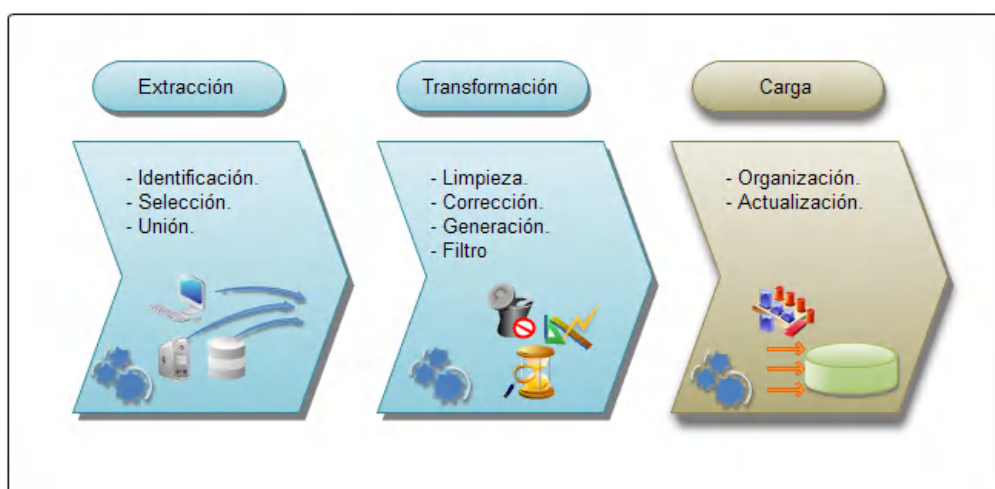


Figura 2.9: Subproceso ETL - Carga

El último subproceso se caracteriza por realizar la carga hacia el DataWarehouse, los datos que previamente han sido extraídos y tratados en los dos subprocesos anteriores para contar con datos de calidad, ahora se procederá a realizar la carga de los mismos a un nuevo ambiente que es el de almacén de datos, para ello es importante implementar métodos y/o maneras de carga de datos con el fin de controlar por ejemplo datos actualizados ó históricos.

Bien, finalizado todo el proceso ETL, lo que se pretende es contar con datos relevantes para el negocio, los mismos que deben ser de valor sin ningún tipo de codificación, es decir datos transparentes y entendibles por los usuarios finales. Ya contado con la calidad de datos en el ambiente del almacén se termina un ciclo del proceso ETL.

2.2.6. MODELO DIMENSIONAL DE UN DATAWAREHOUSE

Un DataWarehouse adopta un *Modelo Dimensional* en su estructura de almacenamiento, caracterizado por ser un esquema en estrella o copo de nieve, lo que permite maximizar el rendimiento de las consultas.

A continuación una definición de Modelo Dimensional:

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

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

2.2.6.1. ESQUEMA ESTRELLA

Es la técnica de diseño más popular usada para un DataWarehouse. Es un paradigma en el cual un único objeto en el centro (conocido como tabla de hecho) se encuentra conectado radialmente con otros objetos circundantes llamados tabla de dimensiones formando una estrella.

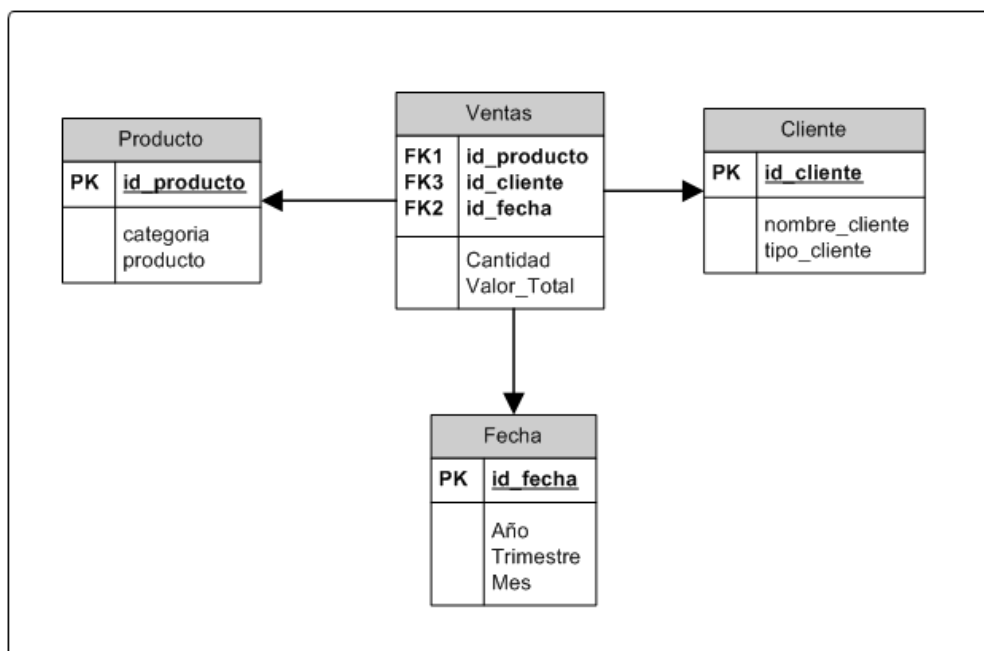


Figura 2.10: Esquema Estrella

Este tipo de esquema es la más utilizada en un ambiente de DataWarehouse debido a q que los datos se encuentran desnormalizados y por ende las

consultas que se pueden realizar son menos complejas ya que no existe la necesidad de realizar muchas relaciones entre tablas.

2.2.6.2. ESQUEMA COPO DE NIEVE

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

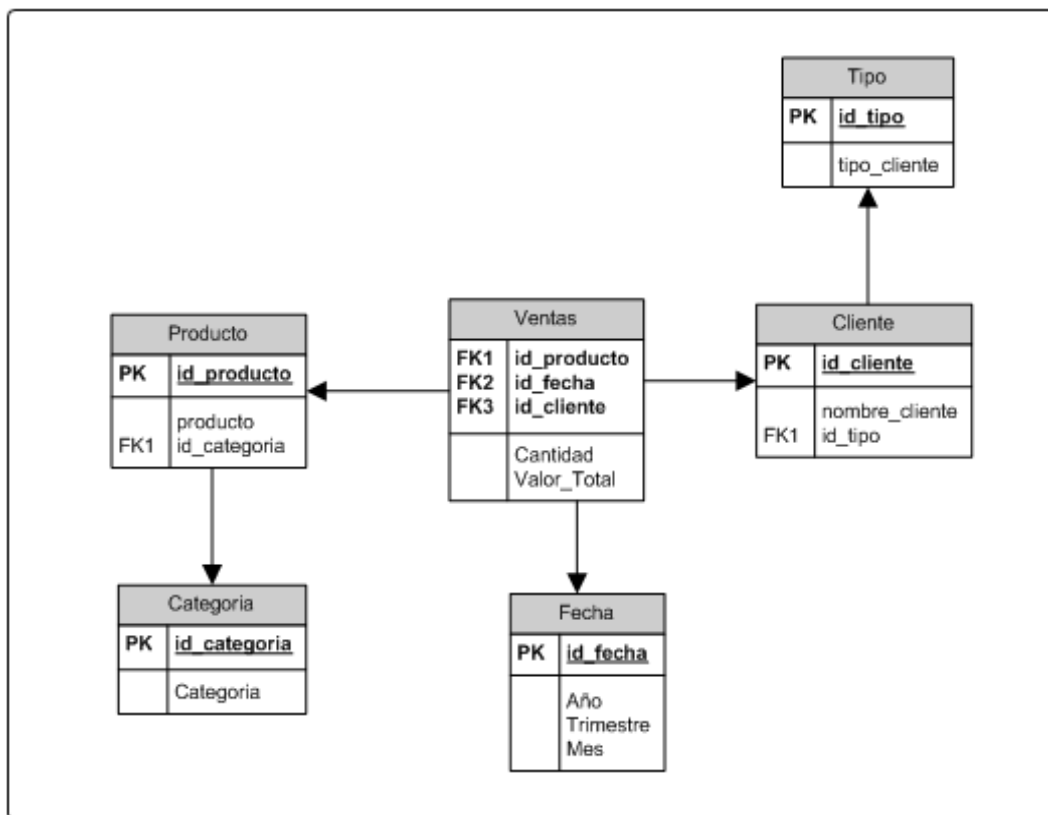


Figura 2.11: Esquema Copo de Nieve

Como se puede observar en la Figura Nro. 2.8, en el esquema copo de nieve se normalizan dimensiones para eliminar redundancia, permitiendo que los datos de las dimensiones se agrupen en múltiples tablas en lugar de una tabla grande.

Tras la especificación del tipo de esquema que se implemente, un modelo de DataWarehouse está estructurado de la siguiente manera:

2.2.6.3. TABLAS

2.2.6.3.1. HECHOS

Tabla de hechos (fact table), es la tabla central de un esquema dimensional (en estrella o en copo de nieve), que contiene los valores de las medidas de negocio a ser analizadas como se puede observar en la Figura 2.12 en un modelo de datos tipo estrella. Este tipo de tabla representa el hecho o actividad del negocio, como por ejemplo, Ventas, Movimientos, Pedidos, etc. Por lo general estos datos son numéricos y pueden agruparse (agregación) en un valor total, las medidas pueden ser por ejemplo, cantidad vendida, costo, precio unitario, etc. Es decir son los indicadores que permitirán medir los hechos que se realizan en el negocio.

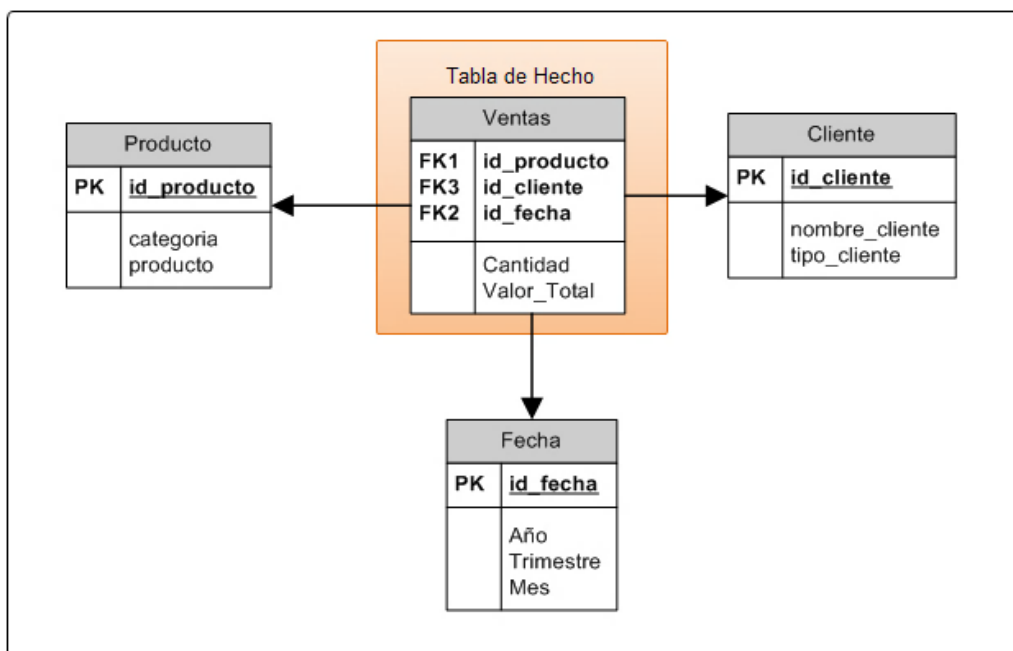


Figura 2.12: Tabla de Hecho

Una tabla de hecho está dividida por 2 tipos de Atributos, estos son: las claves foráneas provenientes de las Dimensiones, y por los indicadores o medidas del Hecho.

2.2.6.3.2. DIMENSIONES

Son tablas que describen a la tabla de hecho, mediante atributos descriptores que poseen de acuerdo a un tema específico del negocio como por ejemplo: Clientes, Productos, Ubicación Geográfica, etc. Esto se puede apreciar de mejor manera a continuación.

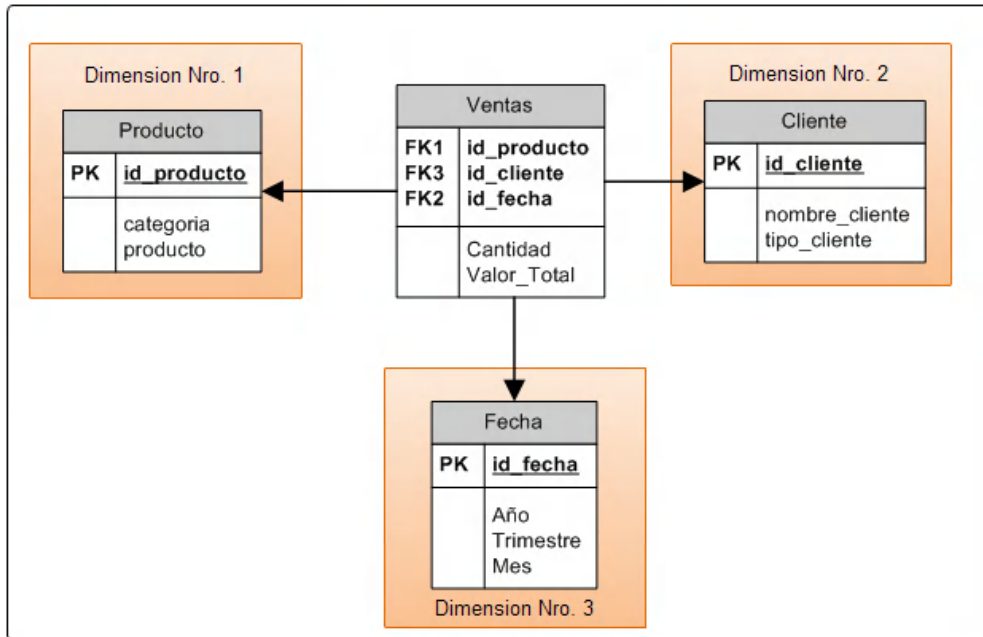


Figura 2.13 - Tablas de Dimensiones

Al igual que la tabla de hecho, en la dimensión se cuenta con 2 tipos de Atributos estos son: la clave primaria, única para cada registro, y de los atributos descriptores de ellas.

Los atributos por lo general son campos textuales ó descriptores numéricos cortos.

2.2.6.4. NIVEL

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

puede observar en la Figura 2.14, en la cual dos dimensiones se encuentran agrupadas de cierta manera, a estas agrupaciones son llamadas Niveles.

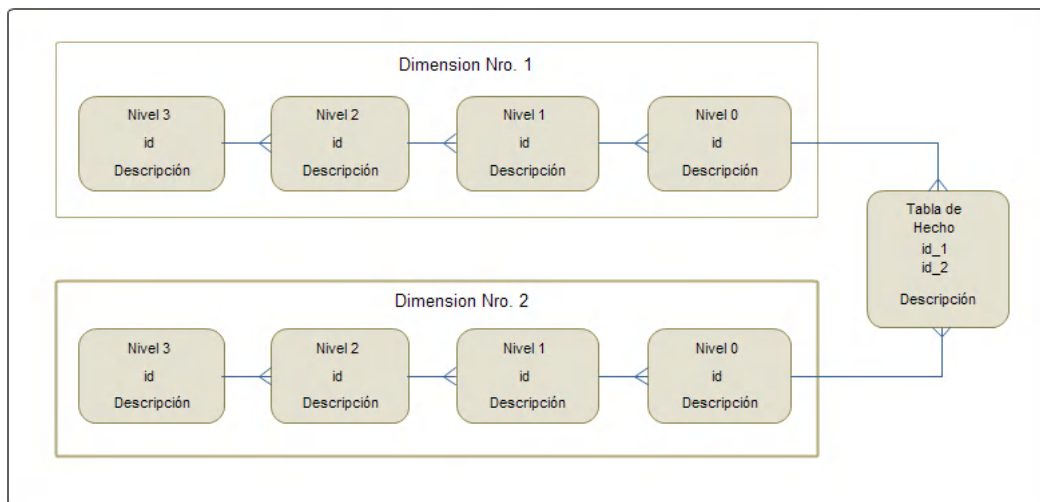


Figura 2.14: Niveles en la Dimensión

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

2.2.6.5. JERARQUÍA

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

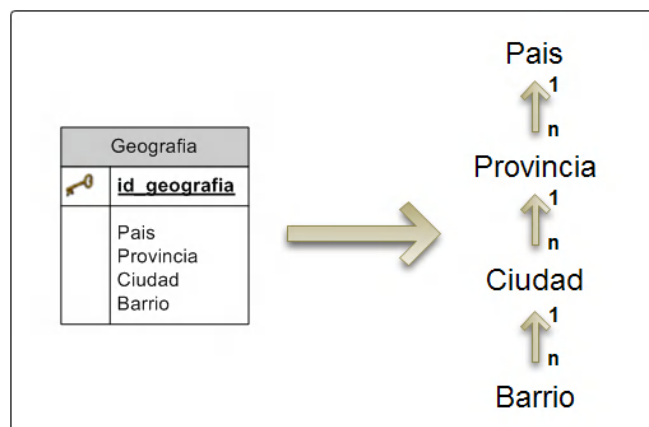


Figura 2.15 - Jerarquía en la Dimensión

2.2.6.6. DIMENSIONES LENTAMENTE CAMBIANTES (SCD)

Respecto este punto los registros de las tablas destino (dimensiones) se pueden clasificar de acuerdo al tipo de carga como:

- **Tipo 0 (SCD):** No se tiene en cuenta la gestión de los cambios históricos y no se realiza esfuerzo alguno. De manera que alguna información será sobrescrita, pero otra no. Sin planificación alguna.
- **Tipo 1 (SCD):** No se guardan históricos. La nueva información sobrescribe la antigua siempre. Principalmente la sobre escritura se realiza por errores de calidad de datos. Este tipo de dimensiones es fácil mantener y son usadas cuando la información histórica no es importante.
- **Tipo 2 (SCD):** Toda la información histórica se guarda en el DataWarehouse. Cuando hay un cambio se crea una nueva entrada con su fecha y clave subrogada apropiadas. A partir de ese momento será el valor usado para las futuras entradas. Las antiguas usaran el valor anterior.
- **Tipo 3 (SCD):** Toda la información histórica se guarda en el DataWarehouse. En este caso se crean nuevas columnas con los valores antiguos y los actuales son remplazados con los nuevos.

2.2.6.7. GRANURALIDAD

La granularidad es el nivel de detalle en que se almacena la información.

Por ejemplo:

- Datos de ventas o compras de una empresa, pueden registrarse día a día

- Datos pertinentes a pagos de sueldos o cuotas de socios, podrán almacenarse a nivel de mes.

A mayor nivel de detalle, mayor posibilidad analítica, ya que los mismos podrán ser resumidos o sumariados.

Los datos con granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario.

La granularidad está centrada en las medidas es decir en la tabla de hecho.

2.2.6.8. RELACIONES

Las relaciones mantienen la integridad de la información del negocio, referenciando datos entre diferentes tablas, en este caso, relacionando las Dimensiones con los Hechos, como por ejemplo la relación Producto – Ventas, Almacén – Pedidos, etc.

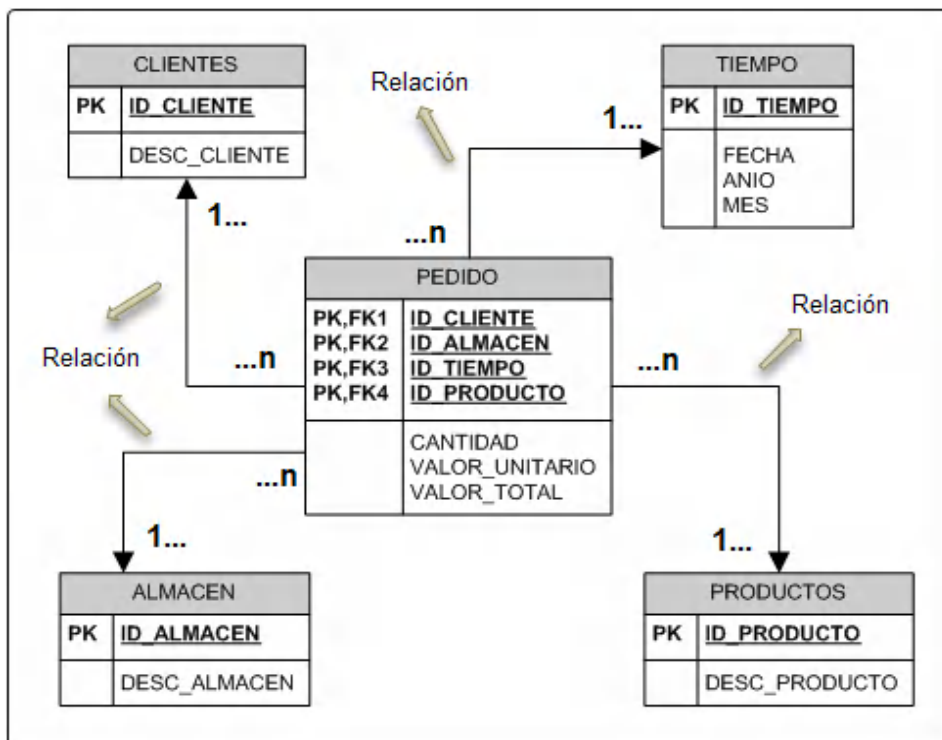


Figura 2.16: Relaciones en un Modelo

2.2.6.9. LLAVES SUBROGADAS

Conocidas también como *surrogate keys*, son caracterizadas por formar parte de las Dimensión y de los hechos, al ser un campo único o identificador único de la tabla (Dimensión), lo que permite que se mantenga la integridad en el modelo dimensional (con el hecho). Las claves subrogadas son independientes a las claves primarias de los orígenes de datos, por lo general estas son creadas en los procesos ETL.

Las claves subrogadas deben ser enteras ya que esto garantiza el rendimiento dentro del modelo dimensional.

2.2.7. EXPLOTACIÓN DEL DATAWAREHOUSE

La explotación de un DataWarehouse, es prácticamente en donde se puede identificar los resultados que puede llegar a brindar una implementación de este tipo, ya que como se ha podido identificar un almacén de datos no es el fin en sí de una solución de BI, y es solamente el medio para dar como resultado a la Inteligencia de Negocio, objetivo que se lleva a cabo con la adecuada explotación de los datos almacenados en el DataWarehouse. Es decir mediante una interfaz grafica se puede apreciar y monitorear el Negocio, como por ejemplo revisar como van las ventas en diferentes sucursales, o analizar cuál es el producto más vendido, o cantidad de artículos a comprar, realizadas por la Institución. Entre los medios más utilizados para aprovechar los datos de un DataWarehouse, se citarán los siguientes:

2.2.7.1. REPORTES Y CONSULTAS

Dentro de toda Organización, es de suma importancia los reportes que se pueden obtener acerca de lo que está pasando con el negocio a la cual esté

especificada, así como también acerca de consultas que los usuarios y empleados, como también directivos, desean realizar sobre algo en específico lo que es también conocido como Consulta ad-hoc.

Las consultas o informes libres trabajan tanto sobre el detalle como sobre las agregaciones de la información.

Realizar este tipo de explotación en un almacén de datos supone una optimización del tradicional entorno de informes (reporting), dado que el Data Warehouse mantiene una estructura y una tecnología mucho más apropiada para este tipo de solicitudes.

Los sistemas de "Consultas & Reportes", no basados en almacenes de datos se caracterizan por la complejidad de las consultas, los altísimos tiempos de respuesta y la interferencia con otros procesos informáticos que compartan su entorno, esto es debido a la alta normalización de sus datos.

2.2.7.2. MINERÍA DE DATOS

“Es el proceso de descubrir correlaciones significativas, los patrones y tendencias de cribado a través de grandes cantidades de datos almacenados en los repositorios. Minería de datos utiliza las tecnologías de reconocimiento de patrones, así como las técnicas estadísticas y matemáticas.”(Gartner, 2009) 6

Tomando en cuenta el concepto anterior, se puede definir que el data mining es una automatización de la búsqueda de patrones relevantes en una base de

⁶ Glosario de Gartner, Data Mining, www.gartner.com, Julio 2009, traducido del idioma inglés al español.

datos. Por ejemplo, un patrón podría ser que los hombres casados sin hijos conducen coches deportivos en el doble de ocasiones que los casados con hijos. En este caso, esta información puede ser muy valiosa para un director de marketing de una empresa automotora.

El proceso de minería se integra con el sistema para el almacenamiento de cantidades masivas de datos (DataWarehouse) de la empresa y presenta la información dirigida a usuarios con capacidad de decisión en el negocio. Para que la minería de datos pueda obtener información útil de las bases de datos de la empresa, es necesario que éstas dispongan de datos de calidad (completos, estructurados, relevantes, históricos, actualizados, etcétera), propósito que cumple el DataWarehouse.

Ejemplos de preguntas que podría responder un proceso de Data Mining son:

- ¿Qué consumidores es más probable que anulen su contrato de telefonía móvil?
- ¿Cuál es la probabilidad que un cliente compre por lo menos 60 dólares del último catálogo de venta por correo?

2.2.7.3. ANÁLISIS OLAP

OLAP (On-Line Analytical Processing) o Procesamiento Analítico en línea, surge como contraste a OLTP (On-Line Transactional Processing) que define a los sistemas de ambientes transaccionales.

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

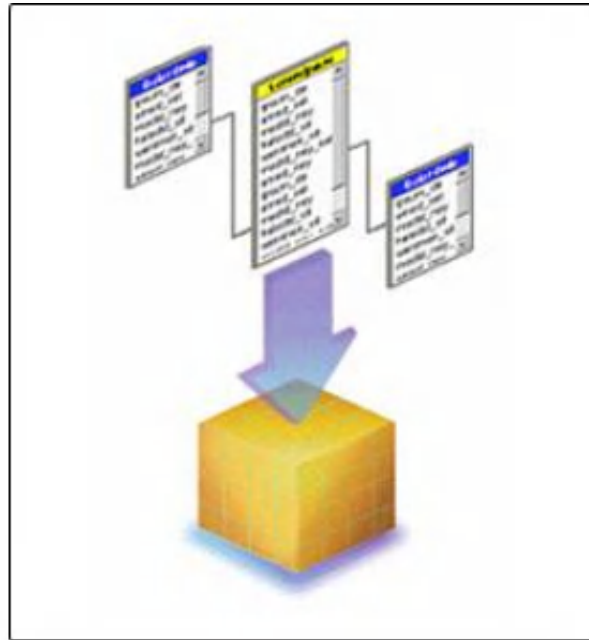


Figura 2.17: OLAP (On-Line Analytical Processing)

Las herramientas OLAP ofrecen a los usuarios, acceso a los almacenes de datos (DataWarehouse) para que puedan obtener las funcionalidades de análisis avanzados que requieren los usuarios finales y analistas, por lo que es considerada la herramienta ideal para sacar el mayor provecho a los datos almacenados en un repositorio.

Las herramientas OLAP se caracterizan por subdividirse en 3 tipos de acuerdo a la manera de almacenar los datos, estos son:

2.2.7.3.1. ROLAP (RELATIONAL OLAP)

Arquitectura en la que se almacenan los datos en un motor de base de datos relacional, pero de igual manera se proporciona la funcionalidad analítica. A través de esta implementación se soporta de mejor manera las capacidades

OLAP con respecto a las bases de datos relacionales, en el sentido que realiza consultas directas a la base de datos, e igualmente presenta los datos de la manera multidimensional caracterizada por la Arquitectura. Los esquemas más comunes sobre los que se trabaja son estrella ó copo de nieve. La arquitectura está compuesta por un servidor de datos relacional y el motor OLAP.

2.2.7.3.2. MOLAP (MULTIDIMENTIONAL OLAP)

En este tipo de Arquitectura los datos se almacenan de manera dimensional en un servidor de base de datos multidimensional, permitiendo optimizar los tiempos de respuesta en la información, ya que al ser sumariada y/o agregada ayuda mucho a los datos calculados por adelantado como por ejemplo los totales, lo que aumenta el desempeño de análisis.

2.2.7.3.3. HOLAP (HYBRID OLAP)

La arquitectura OLAP híbrida (HOLAP), se caracteriza por combinar las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

Para tener un idea de cómo funciona la Tecnología OLAP se puede hacer una analogía con la representación de un cubo, y es aquí donde nace el término de los Cubos de Información, que no son otra cosa más que el análisis

multidimensional de los datos, como se puede apreciar en la Figura 2.15 a continuación.

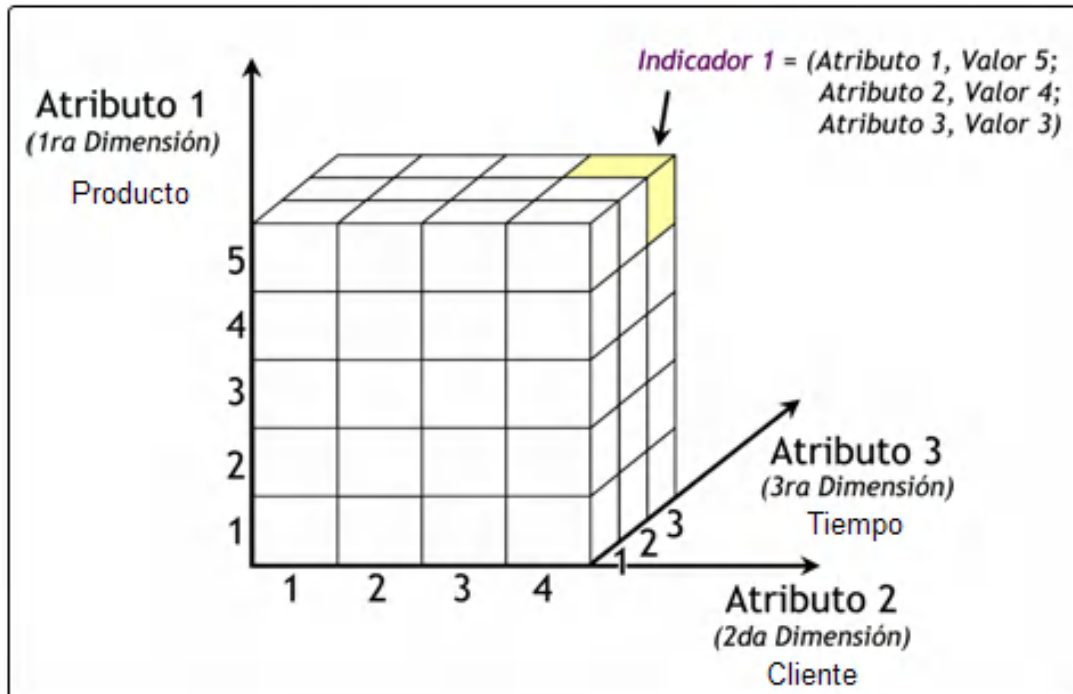


Figura 2.18: Cubo OLAP

Con lo anterior tratado se puede analizar desde una mejor perspectiva el concepto de OLAP e identificar en cómo se puede definir la Arquitectura OLAP de acuerdo a tipo de almacenamiento de datos.

En un análisis OLAP se implementan funcionalidades las cuales permiten interpretar la información de diferentes maneras, estas funciones u operaciones entre las más importantes son:

- DRILL DOWN

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

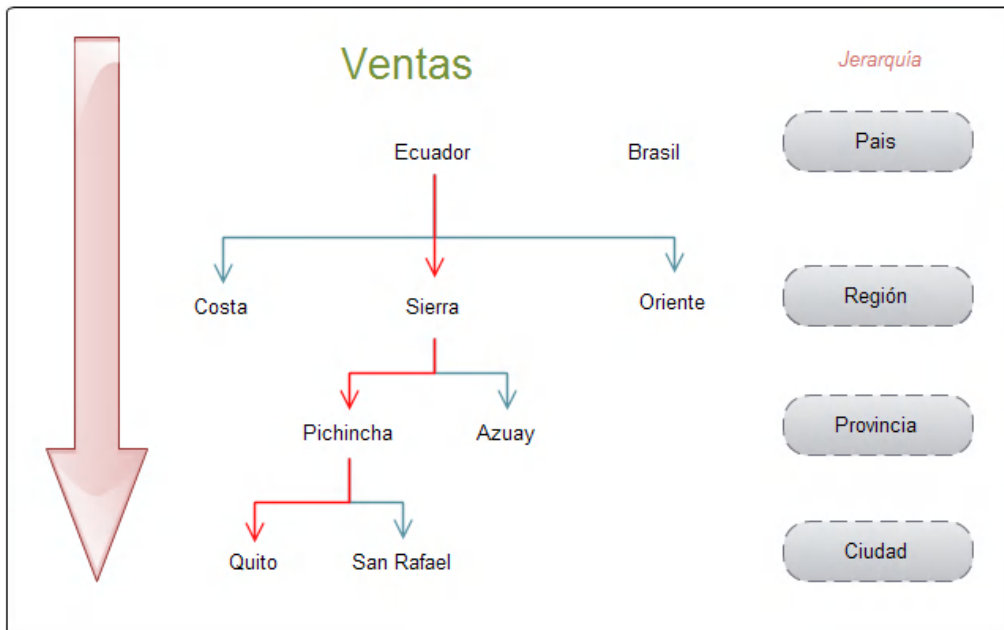


Figura 2.19: Operación OLAP - Drill Down

- ROLL UP

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

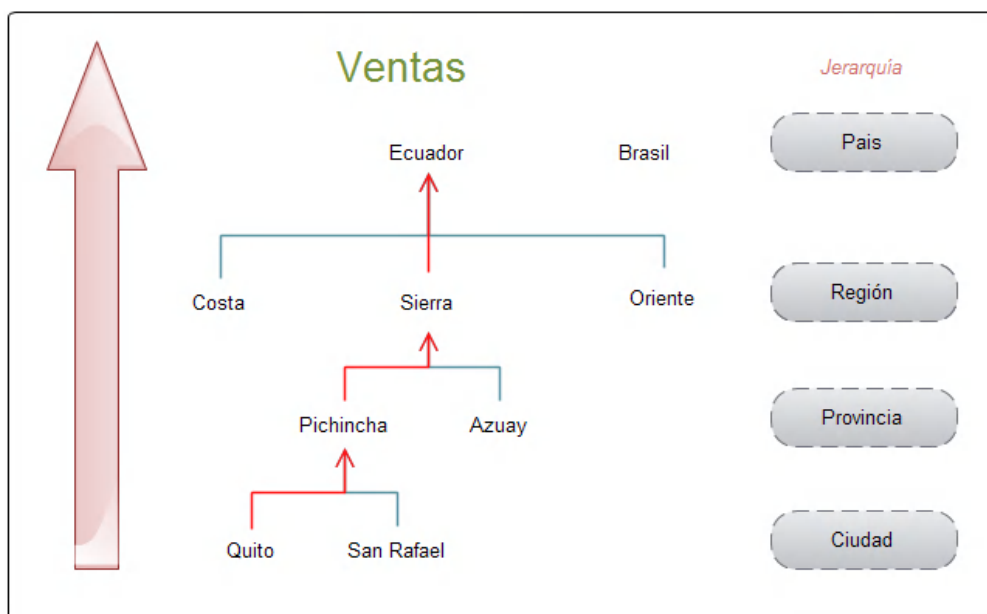


Figura 2.20: Operación OLAP - Roll Up

Existen dos tipos de operaciones de acuerdo a la cantidad de Dimensión utilizadas para un cubo.

- DICE
Obtener un sub-cubo fijando dos o más de sus dimensiones.
- SLICE
Obtener un sub-cubo fijando una de sus dimensiones.

2.3. DESARROLLO DE UN DATAWAREHOUSE BASADO EN LA METODOLOGIA RALPH KIMBALL

Ralph Kimball plantea una metodología para el Desarrollo de soluciones de BI/DWH, que garantiza buenas prácticas para la implementación de un proyecto de este tipo, mediante la consecución de una serie de pasos a seguir durante toda la Administración del Proyecto, como se muestra en la figura a continuación.

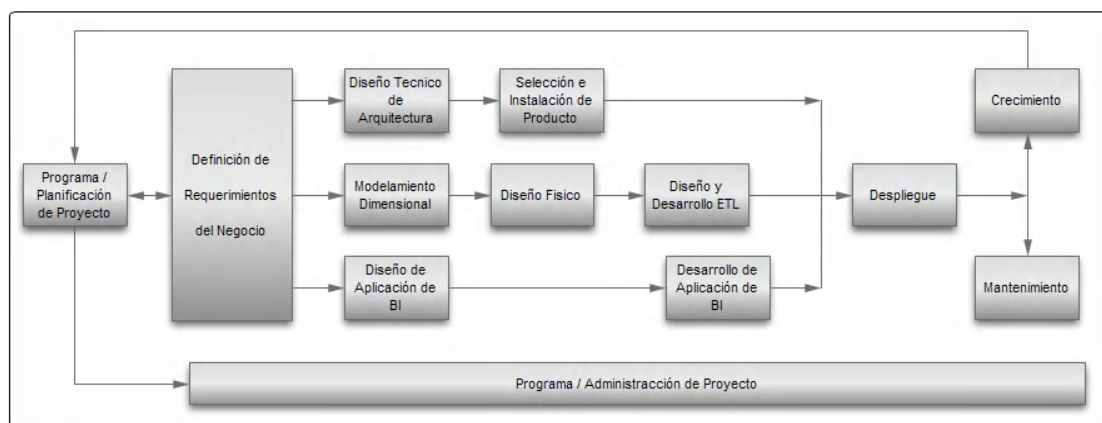


Figura 2.21: The Kimball Lyfecycle Diagram⁷

En base a esta metodología, se tomará los puntos con mayor importancia para que sean especificados, y se realizará ciertas adecuaciones para este proyecto según las necesidades que se presenten.

⁷ Ralph Kimball, The Data Warehouse Lifecycle Toolkit - Second Edition, pag. 3.

2.3.1. PLANEACIÓN Y ADMINISTRACIÓN DEL PROYECTO

Definición del proyecto

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

Los posibles escenarios originados dentro de una empresa son:

- Demanda de un sector del negocio.- Un Sector en específico como por ejemplo los directivos del negocio.
- Demasiada demanda de la información.- Se caracteriza por provenir de más de un sector del negocio, por ejemplo aparte de los directivos del negocio, los de la empresa y ciertos gerentes relacionados con el negocio.
- En busca de demanda.- Escenario en el que se encuentra involucrado el presidente o directivo principal de la empresa.

Determinar preparación de la empresa para un proyecto de DataWarehouse

Son cinco factores importantes según la metodología de Ralph Kimball, las que se deben tomar en cuenta para poder iniciar un proyecto de Almacén de Datos, estos son:

- Patrocinio de la Gerencia del negocio.
- Motivación del Negocio.
- Acompañamiento del departamento de tecnología y de negocio.
- Presencia de cultura Analítica.
- Factibilidad.

Desarrollo del alcance preliminar

Se centra en definir de manera preliminar el alcance del proyecto, para poder hacerlo, la base debe ser los requerimientos del negocio. Para la definición del alcance es importante seguir los siguientes parámetros:

Desarrollo de la Justificación del Negocio

En la Justificación se deben especificar e identificar el costo y beneficio que está relacionado con el proyecto.

Los componentes a ser considerados de inversiones y costos son los siguientes:

- Costos de Hardware.- Se toma en cuenta el costo de acuerdo a los equipos que serán utilizados para el proyecto.
- Costos de Software.- Tomar en cuenta el costo de Software que se necesitara para el proyecto, como software para ETL, reportería, entre otros.
- Recursos Internos.- Recursos con los que la organización cuenta para el proyecto.
- Recursos Externos.- En caso de necesitar algún tipo de consultoría o servicio externo a la organización.
- Costos de Mantenimiento.- La mayoría de productos de hardware y software necesitan de un mantenimiento.
- Soporte.- Todo lo que es el soporte a los usuarios ya sean estos a través de help desk.
- Costos de Crecimiento.- Se debe considerar nuevos requerimientos, mejoras tecnológicas de los productos relacionados de software.

Se considera también los retornos y beneficios a obtener, como:

- Incremento de ganancias.
- Incremento en los niveles de servicio a los clientes.
- Aumento de información para la empresa, y su rápido acceso a ella.

Planeación del Proyecto

Se debe especificar un nombre para el proyecto, así como los roles de la(s) persona(s) que participara en el proyecto, dependiendo la participación de cada una y requerimientos que se hayan especificado.

Los principales roles para el proyecto son:

- Patrocinadores de Negocio.
- Gerente o Administrador del proyecto.
- Líder de Negocio del proyecto.
- Analista de negocio.
- Arquitecto de Datos
- Modelador de datos.
- Administrador de Base de Datos.
- Diseñador de Proceso de ETL.
- Desarrolladores de aplicaciones de BI para usuarios.
- Educador o capacitador del Almacén de datos.

Administración del Proyecto

Aquí se deberá considerar las reuniones con el equipo de trabajo, el monitoreo de estado del proyecto realizado periódicamente, el alcance y estrategias de comunicación ya sea con el equipo de trabajo o con usuarios externos, en

estos últimos esencialmente la manera de tratar por ejemplo cuando llegan nuevos requerimientos a los cuales hay que ver la manera de manejar estas situaciones.

Como entregable de la etapa es el alcance y justificación del Proyecto.

2.3.2. DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO

Los requerimientos de los usuarios del negocio ayudan a definir el diseño e implementación del DataWarehouse, es por esto que desde una perspectiva macro los requerimientos son el centro del negocio y por ende factor de éxito para un satisfactorio desarrollo de un DataWarehouse.

Algo importante a tener en cuenta es saber identificar y asociar los requerimientos desde una perspectiva macro y así como también desde pequeñas perspectivas.

Para empezar se debe hablar con los usuarios del negocio, sobre sus trabajos, objetivos, e intentar conocer cómo y en base a que toman decisiones, al contrario lo que no se debe hacer es preguntar a los usuarios que datos quieren que aparezca o conste en el Data Warehouse, ya que al conocer su objetivos, decisiones, metas tendremos una mayor perspectiva de que desean o de qué tipo de información manejan más frecuencia.

Preparar para la entrevista

Se especifican los roles del equipo entrevistador. Es aconsejable que el mismo equipo maneje las entrevistas tanto con los usuarios del negocio como el equipo de tecnologías de la empresa.

En el equipo los roles principales son, el del líder de entrevista que su principal función es dirigir el interrogatorio, y el otro rol que es el de una persona que tome nota durante las entrevistas (“Escribano”).

Investigación de la organización

Antes de poder iniciar el proceso de levantamiento de requerimientos, se deben analizar los reportes anuales de la organización, para poder determinar en base a que se toman las decisiones y estrategias. Otro factor importante es analizar el negocio y conocer las fortalezas y debilidades.

Selección de los entrevistados.

Para la correcta recolección de requerimientos es importante identificar y seleccionar las personas representativas de cada área de la organización, lo que también es posible identificarlos mediante un organigrama de la organización.

Es fundamental dirigirse hacia los directivos o gerentes de cada negocio, en si a las personas con mayor nivel en la organización debido a que estas personas cuentan con un conocimiento mayor del negocio en el que están involucrados.

Llevar a cabo las entrevistas

Se empieza por las preguntas más generales y seguir con las específicas. La idea es conseguir del personal y directivos, la información suficiente del negocio, objetivos y visión de cada departamento.

También es recomendable realizar entrevistas al personal del departamento de tecnologías, para poder tener presente con que se cuenta para el desarrollo del proyecto.

Revisar los informes y los análisis existentes

Algo importante es tomar en cuenta los reportes que están implementados en la organización, ya que pueden existir reportes analíticos o estadísticos que ayudan al análisis de los requerimientos, y servir también de base para las entrevistas en el sentido de que ayudan a tener la idea de cómo se maneja la información en la organización. Vale recalcar que se debe distinguir entre los reportes y el sistema analítico (Data Warehouse) que se va a implementar.

Concluir la entrevista

Como el resultado de la entrevista es llegar a una conclusión, se pide a cada entrevistado acerca de su o sus criterios de éxito con respecto al proyecto. Por supuesto, cada criterio debe ser medible y cuantificable.

Finalmente se debe agradecer a los entrevistados por sus ideas y hacerles saber lo que pasará posteriormente y lo que será su participación.

Revisar el resultado de la entrevista y Análisis

Es importante tener conocimiento de los sistemas operativos de la empresa, para ayudar a determinar la factibilidad de implementar los requerimientos recolectados.

Se debe revisar los reportes claves y analizar la información recogida de las entrevistas, porque generalmente esto proporciona una traducción aproximada al modelo de datos dimensional, ya que por ejemplo las tablas de dimensiones suministrarán los valores de los títulos de informe y los encabezados de columna o fila superior, mientras que los reportes de medidas numéricas se representan como hechos. Esta especificación es útil para examinar los informes fuera de línea (offline) para una mayor comprensión sobre la dimensionalidad del negocio y las necesidades analíticas de los usuarios.

Para finalizar es importante documentar los requerimientos obtenidos para su aprobación, con lo cual esto pasará a ser el entregable de esta etapa.

2.3.3. DISEÑO TÉCNICO DE LA ARQUITECTURA

Se detalle el ambiente de integración, junto con las tecnologías a utilizar, y el cómo va a funcionar, uniendo tres factores que son: los requerimientos, el entorno técnico, y las estrategias de diseño a implementar.

Para determinar el diseño técnico se basará en dos divisiones que son:

2.3.3.1. Back Room

Se implementa la integración de datos mediante el Proceso de ETL, desde el Origen de los datos hasta la carga de los mismos en el Data Warehouse que es la base destino.

2.3.3.2. Front Room

Parte pública y presentable del Data Warehouse, en la cual se implementa las aplicaciones necesarias para poder presentar al usuario final con el fin de que puedan analizar los datos almacenados en el Data Warehouse.

En esta etapa, el entregable es Diseño de la Arquitectura de todo el proceso de integración de los datos y de la presentación de los mismos, con el fin de obtener un esquema gráfico del funcionamiento del Data Warehouse.

2.3.4. SELECCIÓN E INTALACIÓN DEL PRODUCTO

Mediante las especificaciones tratadas en el diseño técnico de la arquitectura, se visualiza de mejor manera el ambiente en el cual se va a desarrollar el proyecto.

Se evalúa las herramientas a utilizar de acuerdo a cada sección ó área del proceso de desarrollo. Con lo cual estará dirigido para:

- Plataforma DBMS.

Existen dos tipos, motor relacional y motor OLAP, lo que se deberá tener en cuenta cuales son las necesidades, para seleccionar una ó ambas opciones.

- Herramienta ETL.

Se evalúa la herramienta a utilizar, tomando en cuenta aspectos importantes, como si es funcional, y si cumple las posibles necesidades a presentarse en el desarrollo, como por ejemplo si la herramienta puede generar claves subrogadas y hacer búsquedas de ellas, entre otras; esto ayudará determinar la herramienta a utilizar.

- Herramienta BI.

Para seleccionar la mejor opción de una herramienta de Front room ó de presentación, se debe tomar en cuenta la funcionalidad, versatilidad y manejo amigable para el usuario final.

En esta etapa como entregable, son las herramientas a utilizar en cada sección de desarrollo.

2.3.5. MODELAMIENTO DIMENSIONAL

El Modelamiento Dimensional es un técnica de diseño lógico para estructurar datos de manera intuitiva, dirigida a los usuarios de negocio, y permitiendo mayor rendimiento en las consultas.

El modelo dimensional es estrictamente diferente a los modelos relacionales, ya que en el primero los datos se caracterizan por ser redundantes mientras que en el segundo se implementa el Modelamiento 3NF (Third Normal Form) la cual busca eliminar los datos redundantes.

En el modelo dimensional se especifican las medidas y las características. Las medidas son usualmente numéricas y son referidas a los *Hechos*, mientras que las características de estas medidas son representadas por las *Dimensiones*, describiendo por ejemplo quien, que, cuando, de la medida.

La estructura característica del modelo dimensional es el de la semejanza a la de una estrella, es por esto que se es llamado también modelo en estrella.

Grafico de lo dicho se puede referenciar a lo que se trató anteriormente en la sección *2.2.6 Modelo Dimensional y esquemas de un DataWarehouse*.

Cuatro Pasos para el proceso de diseño Dimensional

2.3.5.1. Paso 1 – Seleccionar el Proceso de Negocio.

El primer paso es determinar el proceso que va a ser modelado. Esto es un poco formalismo ya que la selección del proceso es básicamente referente a los requerimientos que se tomaron, a través de ello ya se puede determinar

cuál será el proceso a que se va a modelar. Los procesos de negocio son la parte fundamental del DataWarehouse, y es aconsejable construirlo de manera iterativa es decir un proceso por momento. Vale recalcar que cada proceso de negocio puede producir una o varias tablas de hechos.

2.3.5.2. Paso 2 – Declaración de granularidad.

El segundo paso es definir claramente el nivel de granularidad lo que es lo mismo un registro de la tabla de hecho y decir exactamente la representación de la medida de la tabla de hechos. Un ejemplo es:

- Una medición de inventario realizado cada mes para cada producto en un almacén. ó
- Medición de cantidad de ventas a un cliente de un producto en una tienda.

El grado de granularidad es fundamental, debido a que de ella depende el volumen de datos que se almacenarán en el Data Warehouse, es decir a menor granularidad será mayor el detalle, y a mayor granularidad será menor el detalle y por ende menor volumen de los datos.

2.3.5.3. Paso 3 – Identificación de Dimensiones.

El tercer paso es determinar las dimensiones aplicables a la tabla de hechos en el nivel de granularidad que se detalló. Al agregar nuevas dimensiones los atributos de estas deben cumplir con la misma granularidad que se haya definido.

2.3.5.4. Paso 4 – Identificación de Hechos.

El paso final es identificar los hechos o medidas del negocio. Para la mayoría de los procesos orientados a transacciones, solo hay algunos hechos fundamentales, que miden por ejemplo la cantidad o monto.

2.3.5.5. DataWarehouse Bus Matrix

Es una Matriz Procesos versus Dimensiones comunes, la cual permite tener un panorama más claro del DataWarehouse, al mostrar las dimensiones, hechos, de acuerdo al proceso de negocio.

Proceso de Negocio / Evento	Dimensiones Comunes								
	Artículo	Cliente	Factura de Venta	Item	Proyecto	Proveedor	Sucursal	Empleado	Tiempo
Inventarios / Pedidos	x			x	x				x
Punto de Ventas / Ventas		x	x	x	x		x	x	x

Figura 2.22: DataWarehouse Bus Matrix⁸

En la etapa de Modelamiento Dimensional, como entregable es el Diseño de la Base de Datos del Data Warehouse, es decir el modelo de las tablas, junto con la Matriz de Procesos/Hechos vs. Dimensiones.

2.3.6. DISEÑO FÍSICO

Esta etapa se centra en definir la estructura física, como:

- Configuración del entorno de base de datos.
- Creación de espacios de tablas para Datos y Metadatos.
- Creación de tablas, para el DataWarehouse y para las bases origen en caso de ser necesario.
- Creación de llaves primarias y foráneas.

⁸ Referencia de: Ralph Kimball, The Data Warehouse Lifecycle Toolkit - Second Edition, pag. 90

- Creación de índices en las tablas.
- Desarrollo de tablas temporales (Staging Area) en caso de ser necesario.

En esta etapa, el entregable son los scripts, que son generados al finalizar el diseño y desarrollo físico.

2.3.7. DISEÑO Y DESARROLLO DEL PROCESO ETL

El desarrollo del Sistema ETL es un desafío para un proyecto de este tipo ya que se debe tomar en cuenta aspectos como: estudio y análisis de los datos, integración de los mismos con el objetivo de resolver problemas de calidad, cargas automáticas; por lo que se considera fundamental al Proceso ETL en el marco de desarrollo de un proyecto de BI/DWH.

Se plantean pasos para el diseño y desarrollo de un proceso ETL, estos son:

Desarrollo del Plan ETL

2.3.7.1. Paso 1 – Trazar un plan de alto nivel.

Se diseña un mapa mediante cuadros que especifican la actividad a realizar, con el fin de mostrar el flujo de datos, desde el origen hacia el destino.

2.3.7.2. Paso 2 – Seleccionar la herramienta ETL.

Se selecciona la herramienta ETL adecuada, tomando en cuenta las funcionalidades que se van a utilizar.

2.3.7.3. Paso 3 – Desarrollar estrategias por defecto.

Teniendo la idea de cómo va a funcionar el proceso ETL, se desarrollará estrategias con el fin de garantizar y aumentar el rendimiento del mismo, como por ejemplo:

- Administración de cambios en los atributos de las Dimensiones.
- Carga de Datos.
- Frecuencia de carga.
- Organización del Área Temporal ETL.
- Secuencia de los Procesos de Transformaciones.
- Registro y Manejo de Errores.

2.3.7.4. Paso 4 – Profundizar la tabla objetivo.

Se diseña la jerarquía de las tablas objetivo mediante gráficos, es decir, se le da un orden a los niveles de las Dimensiones, con lo que se puede apreciar la estructura desde un nivel macro hasta el más detallado.

Desarrollo de Carga Inicial

2.3.7.5. Paso 5 – Poblar las tablas de Dimensiones con los datos históricos.

Las Dimensiones (tablas objetivo) serán pobladas por primera vez, para lo cual se implementarán actividades como:

- Extracción de Datos
- Transformaciones
- Carga de Datos.

2.3.7.6. Paso 6 – Realizar la carga histórica de la Tabla de Hechos.

Al igual que las Dimensiones, se realizará el poblado de los Hechos por primera instancia, tomando en cuenta los datos de las Dimensiones, y asignando adecuadamente los registros de las mismas a la de sus respectivos hechos

De igual manera se implementará las actividades de extracción, transformación según la necesidad, y de carga de datos.

Desarrollo de carga Incremental.

2.3.7.7. Paso 7 – Procesamiento incremental de la Dimensiones.

Las Dimensiones serán pobladas de manera incremental, es decir a partir de la segunda carga de datos. Al igual que el poblado inicial, se realizará las transformaciones, y carga de datos, con la diferencia ahora, que la extracción se realizará exclusivamente con datos nuevos o que hayan cambiando en el origen.

2.3.7.8. Paso 8 – Procesamiento incremental de los Hechos.

Así mismo el poblado incremental se realizará a partir de la segunda carga a la tabla de hecho, y tomando en cuenta únicamente de los datos nuevos o que hayan sufrido cambios; y al igual que la carga inicial se implementa las demás actividades como transformaciones.

2.3.7.9. Paso 9 – Operación y Automatización del Sistema ETL.

Con el fin de garantizar la ejecución periódica del proceso ETL, es necesario automatizar los trabajos y transformaciones.

Según la herramienta ETL que se haya seleccionado, esta implementará pasos para configurar tareas programadas, permitiendo que por cierto tiempo como por ejemplo cada día se proceda automáticamente con el poblado de datos hacia el Data Warehouse.

Esto es de gran ayuda, en el sentido que no se necesitará la intervención humana para que se realice el proceso ETL.

Como entregable en la etapa de Diseño y Desarrollo ETL es el documento con las especificaciones de acuerdo a cada paso de la misma sobre el Proyecto, lo que muestra cómo funcionará el proceso ETL.

2.3.8. DISEÑO Y DESARROLLO DE APLICACIÓN DE BI

Lo fundamental aquí, es especificar y proporcionar la interfaz que será mostrada al usuario final, este tipo de aplicaciones permitirá visualizar los datos del Data Warehouse de manera interactiva.

Las aplicaciones de BI son el primer medio de entrega de los valores de un sistema de DataWarehouse dentro de una organización.

Para poder determinar la Aplicación se determino lo siguiente:

2.3.8.1. Tipo de aplicación de BI

Permitirá establecer el tipo de aplicación de BI que se pretende desarrollar para el usuario final.

Las aplicaciones por ejemplo pueden ser:

- Herramientas de reporte y consulta de acceso directo.
- Minería de datos (Data Mining)
- Aplicaciones analíticas.
- Paneles y Cuadros de Mando.

2.3.8.2. Preparar el desarrollo de la aplicación.

En este punto se prepara la herramienta de BI, para que pueda ejecutar sin problemas las necesidades de los usuarios.

Para el desarrollo de esta actividad se tomará en cuenta los siguientes puntos:

- Instalación de la herramienta.
- Poblar los metadatos de la herramienta de BI.
- Crear los metadatos del Negocio.
- Testear la herramienta de BI.

2.3.8.3. Crear las aplicaciones

Primero se captura las especificaciones para la aplicación de BI después de completar la definición de los requerimientos. Posterior a esto se realiza el desarrollo de la aplicación de BI tomando en cuenta el tipo de aplicación a implementar.

En esta etapa, el entregable es la herramienta de BI con las configuraciones realizadas para poder así pasar al despliegue.

2.3.9. DESPLIEGUE

Aquí se unen todas las piezas, como la tecnología a utilizar para los datos, las aplicaciones de BI, la arquitectura, estas convergen en el despliegue.

En este instante, se realizará lo que se conoce en término inglés como *deploy* de las aplicaciones para dar como resultado el producto.

En esta etapa se realiza las siguientes actividades:

2.3.9.1. Despliegue ETL

Se realiza el deploy de todo el sistema ETL, como son las transformaciones y trabajos.

2.3.9.2. Despliegue de BI

Se realiza el despliegue de la herramienta de BI conjuntamente con las aplicaciones desarrolladas para la visualización de los datos.

El entregable en la etapa de despliegue son las aplicaciones ya disponibles para la ejecución, tanto del sistema ETL, como también de la herramienta y aplicaciones de BI.

2.3.10. MANTENIMIENTO Y CRECIMIENTO

En el mantenimiento se verifica que esté funcionando de manera correcta en cada etapa del desarrollo del DataWarehouse.

Se realiza el Diagnostico de los resultados, las operaciones, en sí, es una pequeña administración para que se encuentre funcionando sin problemas.

En esta etapa no se presentan entregables debido a que en el caso de mantenimiento, se lo debe realizar constante con el fin de verificar el correcto desempeño del Data Warehouse, mientras que con el crecimiento en caso de suceder se debe realizar nuevamente todo el ciclo para adecuar los cambios que se pretenda realizar al Almacén y a todo el proceso BI.

2.4. HERRAMIENTAS PARA EL DESARROLLO DEL DATAWAREHOUSE

Las herramientas que se especificarán a continuación serán utilizadas para el desarrollo del proyecto. Se dividió en tres tipos de herramientas las cuales

serán objeto de uso de acuerdo a cada etapa en que se encuentre el desarrollo del mismo.

2.4.1. MOTOR DE BASE DE DATOS.

Como se apreció hasta el momento un DataWarehouse es un almacén, y desde el punto de vista técnico es una base de datos con cualidades y características específicas de rendimiento. Es por ello la importancia de una Base de Datos para una aplicación de este tipo, ya que es la base, y el núcleo en el cual reposarán los datos que previamente han sido tratados.

Por lo que para la creación de la base de datos se debe tomar en cuenta ciertas funcionalidades como:

- Alto Rendimiento y Gestión de datos.
- Alta escalabilidad y desempeño para el crecimiento exponencial de datos, lo cual depende mucho de la cantidad de datos a tratar y ser alojados en el almacén.

2.4.2. HERRAMIENTA ETL.

Hoy en día existe en el mercado varias herramientas ETL de diferente proveedor ó casa comercial, aparte de esto se debe tomar en cuenta funcionalidades que permitan el buen desarrollo del Proceso ETL para lo cual se formuló características básicas que debe tener un herramienta de este tipo, lo que permite garantizar el trato de los datos.

- Extracción de datos desde archivos planos, hojas de cálculo, bases de datos.
- Carga de datos de forma masiva en bases de datos.
- Limpieza de datos.

- Interfaz gráfica amigable.
- Unificar Datos
- Seguridad de la información transmitida.
- Retención temporal de información hasta conseguir sincronización.
- Conectividad con diferentes Motores de Base de datos.
- Soporte de la herramienta para problemas técnicos.
- Interfaz para implementación de código SQL.

2.4.3. HERRAMIENTA DE PRESENTACIÓN.

Como parte de la explotación del DataWarehouse es importante contar con una herramienta que permita brindar una interfaz de usuario final, con facilidades y funcionalidades amigables que permitan una mayor y mejor interpretación de los datos que se encuentran en el Almacén.

Por lo que es recomendable que para la selección de una herramienta de este tipo, ella cuente con funcionalidades como:

- Visualización de gráficos para análisis.
- Navegabilidad dimensional.
- Soporte para agregaciones y sumalizaciones.
- Interfaz amigable para el usuario final.
- Opciones de filtrado o condicionales.
- Soporte a operaciones OLAP como Dril Down, Roll Up, Slice, Dice.

Con las funcionalidades antes mencionadas, facilitará el análisis de datos por parte de los usuarios finales.

2.5. LINEA DE NEGOCIO

En el proyecto que se va a realizar para el Instituto Geográfico Militar se cuenta con información de los negocios o procesos: Inventarios y Punto de Ventas.

Dentro de la Institución se manejan con Proyectos y Presupuestos a los que son asignados de acuerdo a la planificación anual.

En el Instituto se encuentran estructurando una Organización por proceso o Gestiones. Por lo que están conformados Centros de Gestión, los mismos que están encargados de Departamentos. Existen varias Bodegas en las cuales se lleva el control de los Artículos que el Instituto adquiere, así como también de los Productos que vende.

Para poder entender los negocios de mejor manera se los tratará por separado.

2.5.1. INVENTARIOS

El proceso de Inventarios en el Instituto, se efectúa de acuerdo a pedidos y movimientos que se realizan en el Instituto en lo referente a materiales., lo cual involucra las Bodegas.

En el proceso se realizan pedidos de materiales (artículos) o servicios. Estos pedidos son ejecutados dependiendo de las necesidades del personal de la Institución. Por ejemplo si un Militar necesita de botas, el mismo realiza un pedido con el fin de que sea entregado, ó a su vez si necesitan materiales de oficina u otros Artículos. En caso de no tener existencia en stock del Artículo del cual se hizo el pedido, se procede a realizar una orden de compra, que deberá ser autorizada para su adquisición por compra.

A continuación se listará los pasos que se siguen en el proceso de Inventarios:

- Se realiza el Pedido.

- Se procede con la Autorización correspondiente.
- Se verifica la existencia o no del Artículo en Bodega.
- Se realiza la Orden de Compra en caso que no haya existencia del artículo.
- Se realiza el Ingreso o Salida de Bodega.
- Se realiza la transferencia en bodega o se despacha el Artículo al personal.

Para tener más claro el proceso se puede observar en el Anexo A, que es el Diagrama de Flujo del Proceso de Inventarios de Materiales y Servicios facilitado por el Instituto Geográfico Militar.

2.5.2. PUNTO DE VENTAS

El Instituto Geográfico Militar cuenta con dos Puntos de Venta, la principal se encuentra en Quito y la segunda es la Regional ubicada en Guayaquil.

El proceso de Punto de Ventas esta dado por la venta de productos (Ítems) que produce el Instituto, como por ejemplo: Mapas, Cartas Topográficas, Derechos o papeles especiales, títulos, entre otros, los cuales son objeto de venta.

El proceso empieza por el pedido del cliente, al cual se le plantea una proforma con los productos (ítems) a comprar. La proforma puede ser de productos q se encuentran inventariados, para lo cual se debe realizar una orden de producción en caso de no existir en ítem en stock, lo que permitirá que entre en producción el ítem solicitado por el cliente y que deberá constar en proforma.

Terminado con esto se procederá con la facturación y el pago pertinente por parte del cliente.

CAPÍTULO III: DESARROLLO E IMPLEMENTACIÓN DEL DATAWAREHOUSE PARA EL IGM

Para el desarrollo del DataWarehouse, se utilizará como guía la Metodología de Ralph Kimball tratada en el Capítulo II. Esto permitirá realizar el proyecto en base a pasos que garantizan la calidad del mismo. Metodología que será adaptada a las necesidades del Proyecto de Tesis.

3.1. PLANEACIÓN Y ADMINISTRACIÓN DEL PROYECTO.

3.1.1. DEFINICIÓN DEL PROYECTO.

Para el desarrollo del Proyecto se identificó el alto interés por parte de las áreas de Negocio y del área de Gestión Tecnológica.

Las Áreas involucradas manifiestan su interés por obtener información que les permita tomar sus decisiones en función al monitoreo del Negocio.

Por lo que se definió el proyecto, dirigiéndole a las Áreas interesadas, como proyecto inicial de este tipo, cuyo nombre fue establecido para la Tesis.

3.1.2. ALCANCE Y JUSTIFICACIÓN DEL PROYECTO

En cuanto a estos puntos, ya fueron definidos en las especificaciones del Proyecto en el Capítulo I. Pero vale recalcar de manera puntual, el alcance de proyecto, que es el desarrollo un DataWarehouse, el cual mantendrá información de análisis, tanto de Inventarios como de Punto de Ventas. Su justificación, es de contar con información útil para el soporte de decisiones en función al análisis de la información

Para la Planeación del Proyecto se definió los roles a cumplir en el desarrollo de proyecto, estos son:

- Desarrollador del Proyecto.

Persona única la cual estará encargada de:

- Desarrollar.
- Monitorear el Desarrollo.
- Diseñar y Modelar.
- Analizar Negocio.

- Jefes Departamentales.

Personas con alto rango en el negocio a las cuales se les va a ir notificando periódicamente los avances del Proyecto, los jefes de:

- Gestión Tecnológica.
- Gestión de Servicios Institucionales.
- Gestión de Mercadotecnia.

- Personal involucrado en el Negocio.

Empleados que participan activamente en los negocios de Inventarios y Punto de Ventas.

3.2. DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO

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

- Realizar conjuntamente con el Jefe de la Gestión Tecnológica la petición a las áreas controladoras de negocio (Inventarios y punto de Ventas), acerca de qué tipo de información necesitan y tratan diariamente.
- Reunión con los jefes de negocio.

- Realizar el estudio de los reportes implementados en los Sistemas actuales del Instituto, Sistemas relacionados a los procesos de Negocio. En este punto vale recalcar que se dio más énfasis a los reportes estadísticos, ya que a través de ellos se puede recopilar información relevante para el negocio a evaluar.
- Estudiar los negocios con el fin de entender las necesidades del los Usuarios.

Tomando en cuenta los puntos anteriores se obtuvo los requerimientos que se mencionan a continuación:

3.2.1. Proceso de Inventarios

La directiva Inventarios supo dar a conocer la información que le interesa y sobre la cual se soportan para la toma de decisiones, estas son:

- Ingresos en costo de los Artículos que el IGM compra, de acuerdo a su Grupo y Subgrupo.
- Adquisiciones por compra de Artículos.
- Kardex, que son las existencias de Artículos.
- Movimientos que se realizan en Inventarios.
- El consumo anual, es decir los egresos que se produce en costo.
- Pedidos de Artículos que se realizan en Inventarios de acuerdo a las Áreas.
- Análisis de la información en un horizonte de Tiempo, anual, cuatrimestral, mensual.
- Cantidad de Artículos pedidos de acuerdo al tipo y estado.

- Valor total y cantidad de Artículos con su unidad de medida de acuerdo al tipo de movimiento.
- Cantidad de Artículos comprados a proveedores con su Ruc.
- Valor de orden de compra con su estado y número de factura.
- Valor total de los pedidos que se realizan de acuerdo al proceso y proyecto.
- Partidas de los Artículos con su fecha y costo de ultima compra.

Indicadores

- Ingresos a bodega anual sobre los dos millones de dólares y mensualmente superior a los ochenta mil dólares.

Descripción: Este indicador representa total en costo, que es producto de las Adquisiciones que se realiza sobre los Artículos.

- Un millón de dólares en Consumo anual que se produce en la Bodega Central.

Descripción: Este indicador representa el valor total de los egresos que se produce en Bodega Central.

Mencionada la información que le interesa a la parte Directiva de Inventarios, los mismos manifestaron consultas a considerar para la presentación, los mismos que se especifican a continuación:

- Ingresos: por Proveedores, por Bodega, de importaciones.

Descripción: Permite visualizar y explorar el Valor Total, que es el Costo Unitario por la Cantidad de Artículos.

- Egresos a Bodega.

Descripción: Permite visualizar y explorar el Valor Total de los Egresos que se produce de acuerdo a cada Bodega, es decir el valor de salida de los Grupos de Artículos por Bodega.

- Existencia.

Descripción: Permite visualizar y explorar el Saldo Total (cantidad) de Artículos en Bodega.

- Movimiento por Artículo.

Descripción: Permite visualizar y explorar el control de cantidad y existencia de los Artículos, que da lugar un Movimiento.

- Movimiento de ingreso.

Descripción: Permite visualizar y explorar el control de Cantidad, Costo Unitario y Valor Total de los Artículos que son transferidos en un Movimiento, cuya transacción es de Ingreso a Bodega.

- Compras por Proveedor.

Descripción: Permite visualizar y explorar el Valor total de las Compras que se realizaron de acuerdo a cada Proveedor.

- Control de Pedidos (Pedidos por Área, Artículos y proyectos)

Descripción: Permite visualizar y explorar la Cantidad Pedida, Costo Unitario y Valor Total de los Pedidos de acuerdo al Proyecto y Área de donde se dio origen al mismo.

3.2.2. Proceso de Punto de Ventas

Para el negocio de Punto de Ventas no se conto con el análisis de reportes estadísticos, debido a que no existe una implementación de este tipo en el Instituto, por lo que solo se realizó la entrevista y petición de necesidades al jefe de Gestión de Mercadotecnia (gestión encargada de Punto de ventas).

La directiva supo dar a conocer la información que le interesa y sobre la cual se soportan para la toma de decisiones, estas son:

- Total de ventas en precio.
- Ítems más vendidos, con su unidad de medida y si el ítem es o no inventariado.
- Ventas realizadas en las Sucursales.
- Valor total de la venta al cliente.
- Análisis de la información en un horizonte de Tiempo, anual, cuatrimestral, mensual, diario.
- Valor total de la venta con su número de factura y estado.
- Cantidad de Ítems que se conformaron en proforma, con su tipo y estado.
- Total de ventas que se realizan de acuerdo a los Proyectos.
- Valor de las proformas a las cuales se realizo un anticipo.
- Empleados con más ventas, con su número de cedula.

- Valor total de las ventas que se realizan por los centros de gestión y empresa.
- Total de las ventas que se realizan de acuerdo al grupo de ítem.

Indicadores

- Número de ventas a clientes anualmente superior a las cien, y mensualmente de veinte.

Descripción: Este indicador representa la cantidad de ventas que se realizó al cliente en el año.

- Numero de ventas anuales de Empleados superiores a cien, y mensualmente de veinte.

Descripción: Este indicador representa la cantidad de ventas que ejecutaron los Empleados durante el año.

Indicadores Clave de Rendimiento (KPI)

- Nivel de Cumplimiento del Trabajo Realizado

Descripción: Este indicador representa un valor de evaluación para los Empleados, en función de su cumplimiento, lo que permitirá medir anualmente su responsabilidad con las ventas.

Valor objetivo Anual: 200

$$\% \text{ Cumplimiento} = \frac{NUM_VENTAS * 100}{VALOR_OBJETIVO}$$

Tabla 3.1: Indicador de Cumplimiento

% de Cumplimiento	Nivel de Cumplimiento (KPI)
$\geq 90\%$	5
$\geq 80\%$ y $< 90\%$	4
$\geq 70\%$ y $< 80\%$	3
$\geq 60\%$ y $< 70\%$	2
$< 60\%$	1

Mencionada la información que le interesa a la parte directiva de Punto de ventas, manifestaron consultas para la visualización que se tomará en cuenta, los mismos que se especifican a continuación:

- Ventas y cantidad vendida de Ítems por :
 - Cliente.
 - Proyecto.
 - Familia de Ítem.
 - Sucursal.
 - Centro de Gestión.
 - Factura.
 - Proforma.
 - Empleado.
 - Empresa.

Descripción: Permite visualizar y explorar el Precio Total de Venta y la cantidad de Ítems vendidos que se realizaron tomando en cuenta cada una de las entidades antes mencionadas como Cliente, Ítems.

- Numero de ventas por Empleado, por Sucursal.

Descripción: Permite visualizar y explorar la Cantidad de Ventas realizadas anualmente por cada Empleado.

A estos requerimientos, tanto de Inventarios como de Punto de Ventas, se especificó un horizonte de Tiempo, que según la toma de requerimientos sus niveles fueron:

- Anual (Periodo).
- Cuatrimestral.
- Mensual.
- Diario.

Adicional a esto se especificó por parte de los directivos que necesitan los datos actualizados y no mantener el histórico de los mismos, empezando en un horizonte de tiempo desde el Año 2004 para Inventarios y desde finales del 2008 para Punto de Ventas.

No se definieron Indicadores Claves de Rendimiento para inventarios, debido a que los Directivos y Gerentes de los Negocios en cuestión, no cuentan con Sistemas Automatizados y Datos capaces de soportarlos.

Por lo que las necesidades según el Directivo de Inventarios, se dirigieron exclusivamente en visualizar las métricas establecidas, ya que en base al análisis de estas toman sus decisiones, así como también de los indicadores antes mencionados.

También debido a que para el negocio, no está establecido este tipo de evaluación que denota los Indicadores de Rendimiento, y en cuanto a Punto de

Ventas en las reuniones establecidas con los Directivos manifestaron que están desarrollando KPI's enfocados en su Negocio a cargo, como parte del plan para el año 2010.

En función a esto, para el Diseño y Desarrollo del Data Warehouse se contará con los requerimientos tomados por cada Negocio.

3.3. DISEÑO TÉCNICO DE LA ARQUITECTURA

En el Diseño de la Arquitectura se definió dos ambientes, el primero es el Back Room y el segundo que es Front Room, los mismos que serán detallados a continuación.

3.3.1. BACK ROOM

En este ambiente se especificará el ambiente del proceso ETL, así como los orígenes de datos.

Como se puede apreciar en la Figura 3.2, se detalla el Diseño que realizo para la implementación del Back Room, que consta desde el Origen de los Datos hasta el almacenamiento en el DataWarehouse.

Origen de datos.

La base de datos Origen se encuentra sobre la plataforma de Oracle 10g. Como se puede observar en la Figura 3.2 se establecen 2 orígenes de datos, IGM y GEO debido a que la primera se utilizará cuando el proyecto se encuentre en producción, mientras que la Base de datos GEO se la utilizará como base de datos origen para las pruebas y desarrollo. Vale recalcar que los

metadatos (esquemas, tablas, disparadores, funciones, etc.) de ambas bases son similares, ya que dentro del Instituto se utiliza la base GEO para la fase de desarrollo de cualquier proyecto, y cuando entre a producción se utilizar la base IGM.

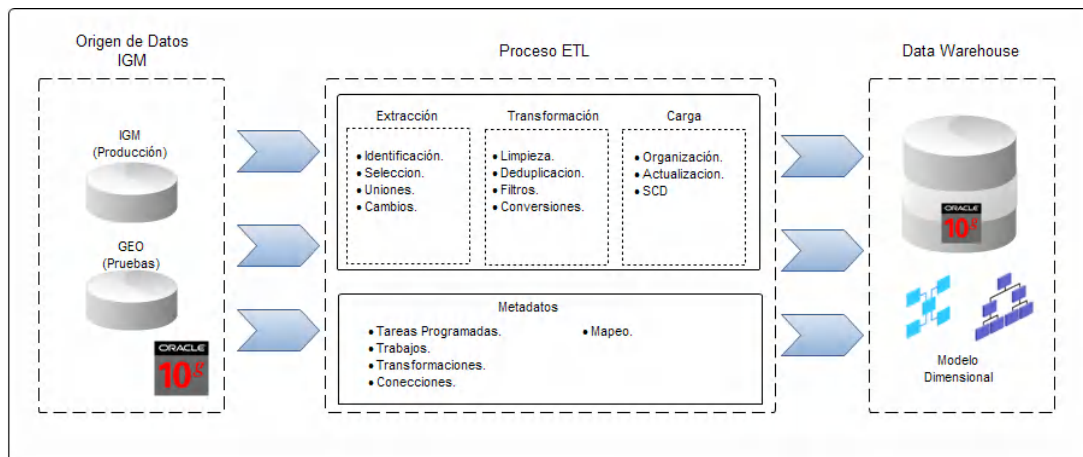


Figura 3.1: Back Room - Proyecto

ETL y Metadatos

En cuanto al proceso ETL se realizará el Mapeo de los datos Origen, que serán especificados posteriormente en el Modelamiento Dimensional, después se desarrollará el proceso de extracción, seguido de la transformación de los datos Origen, dependiendo de qué cambios se les debe adecuar, y finalmente se procederá con la carga hacia el Data Warehouse.

Los metadatos son las especificaciones de transformaciones, mapeos, tareas programadas, que se almacenarán en el Repositorio del Data Warehouse como soporte al proceso ETL.

Se especificará con más detalle el Proceso ETL y los Metadatos del Proyecto en la sección de Diseño y Desarrollo ETL.

Destino de datos (DataWarehouse)

El destino que es el Almacén de datos, mantendrá el modelo dimensional que será especificado en la sección Diseño y Modelamiento Dimensional.

La base de datos destino estará separado físicamente de las bases de datos Origen, por motivo de rendimiento y para poder separar el ambiente operacional del ambiente del DataWarehouse.

3.3.2. FRONT ROOM

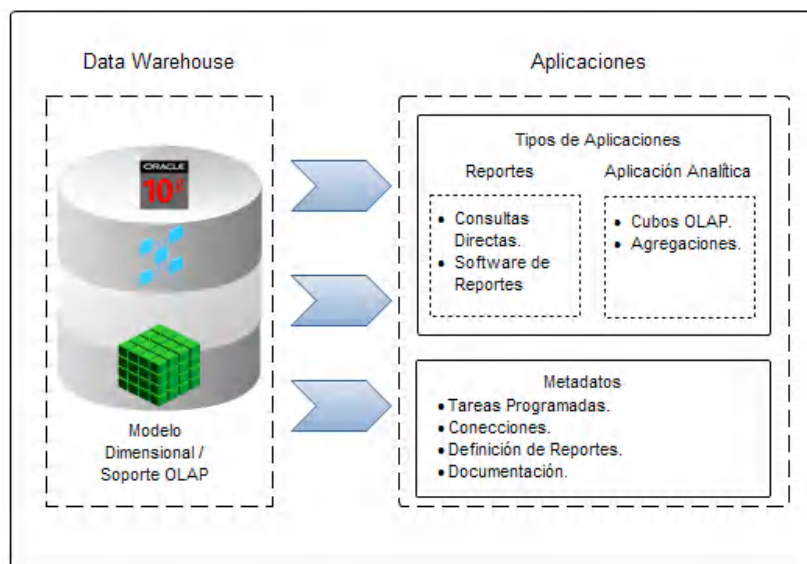


Figura 3.2: Front Room - Proyecto

En esta sección se especifica la Arquitectura de presentación del DataWarehouse, mediante aplicaciones dirigidas al usuario final.

Como se puede apreciar en la Figura 3.3, se parte del DataWarehouse en donde se encuentra los datos que ya han sido tratados y almacenados, posteriormente se procede a la presentación de acuerdo al tipo de aplicación de usuario final que se utilice. Para el desarrollo del presente proyecto se implementará una arquitectura de ambiente analítico para poder diseñar los cubos OLAP, así como también el DataWarehouse servirá como soporte para

lo que son consultas directas mediante Sistemas que Desarrollan en el Instituto.

El Desarrollo de los Cubos OLAP será detallado más adelante en la Sección Diseño y Desarrollo de Aplicación de BI.

3.4. SELECCIÓN DEL PRODUCTO

Para la selección de los productos a utilizar se tomo en cuenta las características planteadas anteriormente en el Capítulo II - Sección 4 sobre las Herramientas para el Desarrollo de un Data Warehouse, con lo cual se planteo las siguientes herramientas a ser utilizadas en el Proyecto:

3.4.1. Plataforma DBMS.

La plataforma a Utilizar es el Motor de base de datos Oracle 10g Release 2 (10.2.0.1.0) ya que el Instituto Geográfico Militar cuenta con la licencia y su estructura de base de datos sobre esta plataforma.

Vale recalcar que este motor tiene incorporado la funcionalidad de la tecnología OLAP, por lo que aparte de ser el motor de base de datos, será también un servidor o base de datos OLAP.

3.4.2. Herramienta ETL.

En cuanto a la herramienta ETL se planteo 2 opciones:

- Pentaho Data Integration (Kettle).
- Oracle Warehouse Builder (OWB).

La primera opción es una herramienta open source, parte de la suite de BI de Pentaho y la segunda es una herramienta privativa de Oracle.

Para poder seleccionar el producto adecuado se analizo 4 aspectos fundamentales para este proyecto:

Tabla 3.2: Comparación OWB vs PDI

Característica Herramienta	Licencia	Facilidad de Manejo (% Entendimiento)	Funcionalidad	Estabilidad
Oracle Warehouse Builder (OWB)	Privativo con costo de licencia	70%	100%	100%
Pentaho Data Integration (Kettle)	Open source sin costos de licencia	80%	100%	100%

- Licencia.- La licencia de Pentaho Data Integration no tiene costo alguno debido a que es una herramienta Open Source, mientras que Warehouse Builder de Oracle tiene un precio por costo de licencia al no contar con licencia de la base de datos Oracle 10g Enterprise Edition, por lo que para versiones Standard y Express de la base de datos Oracle se tiene que adquirir la licencia.
- Facilidad de Manejo.- Este parámetro dependió mucho del entendimiento del desarrollador sobre la herramienta ya que ambas son herramientas visuales. La diferencia marco en que la herramienta de Pentaho se presento más entendible para el desarrollador con respecto a la herramienta de Oracle.
- Funcionalidad.- La funcionalidad de las 2 herramientas cumplieron con las características necesarias para la integración de datos, así como para transformaciones, extracción y carga. Existe un funcionalidad que difiere de gran manera a las 2 herramientas y es que la herramienta de Oracle solo puede interactuar con un Motor de Base de datos exclusivamente de su

misma firma es Oracle, en cambio la herramienta de Pentaho puede actuar con varias Bases de Datos.

- Estabilidad.- En cuanto a la estabilidad la herramienta de Oracle en teoría se puede considerar estable debido a que es una versión Empresarial con soporte por parte de los técnicos de Oracle, mientras que la herramienta open source, no cuenta con soporte al menos que se costee el mismo. Se realizo pruebas con ambas herramientas y ninguna manifestó inestabilidad.

Tomando en cuenta estos aspectos y el del objetivo se concluyó en utilizar la herramienta Open Source, Pentaho Data Integration 3.2 (Kettle) la misma que fue considerada ya que en el Instituto Geográfico Militar están cambiando las políticas de uso de herramientas privativas a las de libre distribución.

Herramienta de Pentaho: Pentaho Data Integration

La Arquitectura de PDI se encuentra representada de la Siguiete manera:

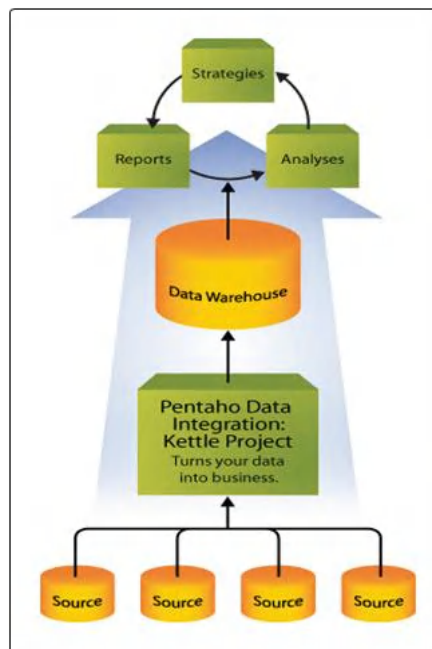


Figura 3.3: Arquitectura - Pentaho Data Integration⁹

⁹ Tomado de <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>

A parte de ser open source y sin costes de licencia, las características básicas de esta herramienta son:

- Entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma: Windows, Macintosh, Linux.
- Basado en dos tipos de objetos: *Transformaciones* (colección de pasos en un proceso ETL) y *trabajos* (colección de transformaciones).
- Incluye cuatro herramientas:
 - **Spoon:** para diseñar transformaciones o trabajos ETL usando el entorno gráfico.
 - **PAN:** para ejecutar transformaciones diseñadas con spoon mediante líneas de comando.
 - **CHEF:** para crear trabajos mediante líneas de comando.
 - **Kitchen:** para ejecutar trabajos mediante líneas de comando.

3.4.3. Herramienta de BI.

La herramienta de BI que se seleccionó para lo que la presentación de la información mediante cubos OLAP es el Discoverer OLAP Plus debido a que dentro de la Institución se cuenta con la licencia de este producto de Oracle, y debido a que los usuarios de la Institución se encuentran familiarizados con la herramienta, en este caso con el Discoverer Viewer que es la aplicación de visualización parte del Discoverer Plus. Por lo tanto se va a utilizar la herramienta mencionada.

Arquitectura de Discoverer

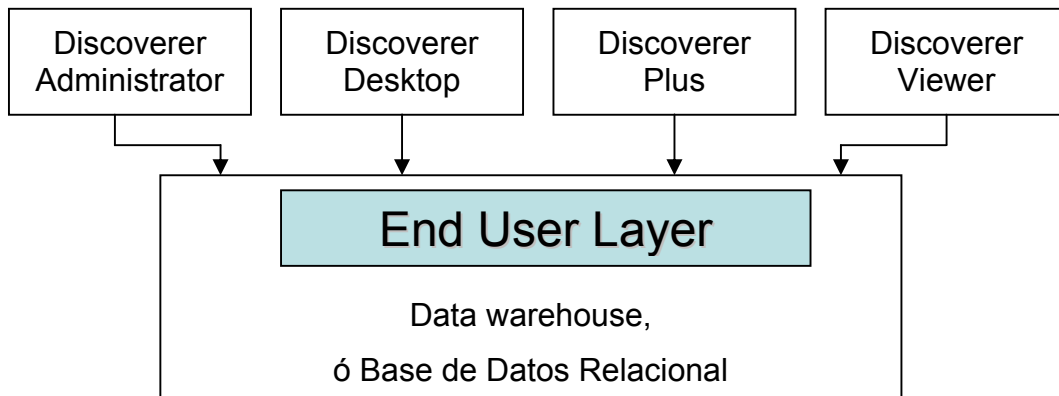


Figura 3.4: Arquitectura Discoverer

- **Discoverer Administrator:** Este producto sirve para crear, mantener y administrar datos en la capa de usuario final (EUL) y para definir como lo usuarios interactúan con los datos.
- **Discoverer Desktop:** Como usuario final, se usa este componente para ejecutar consultas AD HOC, y generar reportes, mediante un interfaz de Escritorio.
- **Discoverer Plus:** Es una interface Web y puede ser acceder usando el browser preferido. Como usuario final, se puede usar este componente para crear, visualizar y editar los libros de trabajo sobre el internet o intranet.
- **Discoverer Viewer:** Este componente es útil para visualizar los libros y hojas de trabajo, usando un browser. Los libros de trabajo creados en Discoverer Desktop y Discoverer Plus pueden ser abiertos, pero no pueden ser editados usando el Viewer. Discoverer Viewer hace que la información disponible en el Internet o intranet, a través de firewalls, y permite la información para ser transparentemente embebida en páginas Web y se pueda acceder desde portales corporativos.

- **End User Layer:** Este componente es una capa de metadatos basada en el servidor, lo que permite ocultar la complejidad de la base de datos, permitiendo que los usuarios puedan interactuar con ella sin tener que reconocer comandos propios de la Base. Esto permite a los usuarios usar el lenguaje de negocios con el que está más familiarizado.

3.5. ESTANDARES DE DESARROLLO PARA EL PROYECTO

Como parte del desarrollo se aumento un tema a tratar dentro de la metodología para adecuarla al proyecto, previo al Modelamiento Dimensional, esto son los estándares que se tomarán en cuenta al momento de desarrollar y que servirán para describir los objetos de manera abreviada.

3.5.1. Estándares de Tablas.

Los Nombres de la Tablas para su creación mantendrán el siguiente estándar::

<TIPO>_<NOMBRE>_[NEGOCIO]

<TIPO>: El tipo describe si la tabla es Dimensión, Hecho u otro.

<NOMBRE>: Describe el nombre de la tabla.

[NEGOCIO]: Especifica si fuese necesario el nombre del negocio al que pertenece la tabla.

3.5.2. Estándares de Campos.

Para los campos no se realizará ningún tipo de especificación para los nombres debido a la variedad que pueden tener pero básicamente mantendrá un estándar como se muestra a continuación:

[DESCRIPCION]_<NOMBRE_CAMPO>_[DESCRIPCION_2]

[DESCRIPCION] – [DESCRIPCION_2]: Especifica si fuese necesario una descripción abreviada corta.

<NOMBRE_CAMPO>: Especifica el Nombre del Campo.

3.5.3. Estándares de Índices.

Para los índices se incluyen dentro de este las claves primaria y foráneas especificaciones de nombre de la siguiente manera:

Clave Primaria

<NOMBRE_TABLA>_<TIPO>

<NOMBRE_TABLA>: Especifica el Nombre de la tabla a la que pertenece.

<TIPO>: Especifica el tipo de clave en este caso primaria.

Clave Foránea

<DESCRIPCION_CORTA_TABLA_PRINCIPAL>_<DESCRIPCIÓN_CORTA_TABLA_FORANEA>_<TIPO>

<DESCRIPCION_CORTA_TABLA_PRINCIPAL>: Descripción abreviada de la tabla Principal.

<DESCRIPCIÓN_CORTA_TABLA_FORANEA>: Descripción abreviada de la tabla Foránea.

Índices adicionales

Los índices adicionales como por ejemplo los índices que son creados en el desarrollo ETL por medio de la herramienta ETL serán dejados por defecto,

como son los índices lookup utilizados para las actualizaciones de datos por la Herramienta ETL.

3.5.4. Estándares de Disparadores (Triggers)

Para la Creación de disparadores se mantendrá el siguiente estándar:

```
CREATE OR REPLACE TRIGGER
<ESQUEMA>.<TIPO>_<TABLA_DE_ASIGNACION>
AFTER INSERT OR UPDATE
ON <ESQUEMA>.<TABLA>
REFERENCING NEW AS NEW OLD AS OLD
FOR EACH ROW
BEGIN
<CUERPO DEL DISPARADOR>
END <TIPO>_<TABLA_DE_ASIGNACION>;
```

3.5.5. Estándares ETL

En cuanto a los estándares para el Proceso ETL se los especificará en la *Sección: DISEÑO Y DESARROLLO DEL PROCESO ETL*, topando más puntos a tomar en cuenta para su elaboración.

3.6. MODELAMIENTO DIMENSIONAL

Para poder realizar el Modelamiento del almacén de datos primero hay que analizar el modelo relacional de los datos origen tanto de Inventarios como de punto de ventas.

Como ya se especifico anteriormente el origen de datos se encuentra bajo la plataforma Oracle 10g. El Instituto Geográfico Militar cuenta actualmente con los modelos Entidad Relación en sus sistemas transaccionales descrito en el Anexo B: Inventarios, el Anexo C: Punto de Ventas, y el Anexo D: tablas utilizadas de Financiero, Proyectos y Recursos Humanos.

Dada la definición de requerimientos así como también tomando en cuenta los modelos relacionales del Instituto que serán utilizados como origen. Se planteará los modelos dimensionales capaces de soportar dichos requerimientos para los negocios de Inventarios y Punto de Ventas.

Los nombres de los esquemas en la base de datos Origen son especificados de la siguiente manera:

Tabla 3.3: Esquemas – Base de datos Origen

Nombre de Esquema	Nombre de Esquema en la Base de Datos Origen.
Inventarios	INVENTARIOS
Punto de Ventas	PTV
Financiero	FINANCIERO
Proyectos	PROYECTOS
Recursos Humanos	RH

3.6.1. PASO 1: SELECCIÓN DEL PROCESO DE NEGOCIO.

Como se definió en el alcance y los requerimientos del presente proyecto los negocios enfocados son:

- Inventarios.
- Punto de Ventas.

3.6.2. PASO 2: DEFINICIÓN DE LA GRANURALIDAD.

En cuanto al nivel de la granularidad se la definió lo más baja o granular posible, para que el modelo dimensional pueda soportar datos y/o ciertas consultas que se encuentran a un Nivel Organizacional Operativo lo que beneficiará y aumentará las agregaciones de acuerdo a cada nivel, permitiendo un mayor análisis de los datos.

3.6.2.1. Inventarios

La granularidad del negocio de inventarios corresponde a un registro de pedido y/o movimiento para cada Dimensión, es decir un registro con un Artículo, un Área, una Bodega, una Orden de Compra, un Proyecto, en un día.

3.6.2.2. Punto de Ventas

La granularidad de Punto de Ventas está dada por un registro de venta de un Ítem, un cliente, una sucursal, un Centro de Gestión, una factura, un Proyecto, una Proforma, un empleado, en un día.

3.6.3. PASO 3: IDENTIFICAR LAS DIMENSIONES Y MAPEAR LOS DATOS ORIGEN.

Las Dimensiones a definir soportarán los requerimientos establecidos cumpliendo con la granularidad especificada para cada negocio. El mapeo de los datos será especificado en cada Dimensión.

Para el negocio de Inventarios y Punto de Ventas, se definieron las dimensiones que se detallan a continuación

3.6.3.1. Dimensión Tiempo

Contiene la Información acerca del horizonte de tiempo de las actividades del Instituto.

Nombre de Tabla: Dim_Tiempo

Tipo de Tabla: Dimensión

Esquema Origen: Ninguno

Origen: Sistema ETL

Destino			
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave
Key_Tiempo	Clave Primaria Subrogada	Integer	Primaria
Fecha	Fecha del día	Date	
Anio	Año de actual de la Fecha	Integer	
Cuatrimestre	Cuatrimestre del Año	Char(2)	
Mes_Nombre	Nombre del Mes	Varchar2(30 Byte)	
Dia_Nombre	Nombre del día de Semana	Varchar2(30 Byte)	

3.6.3.2. Dimensión Área.

Contiene la Información acerca de las Áreas que maneja el Instituto.

Nombre de Tabla: DIM_AREA
Tipo de Tabla: Dimensión
Esquema Origen: INVENTARIOS, FINANCIERO
Tablas Origen: T_AREAS, T_EMPRESA

Nombre - Columna	Destino				Origen		
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Area	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Area	Número del Area	Integer			INVENTARIOS - T_AREAS	NO_CD	NUMBER(4)
Nombre_Area	Nombre de la Area	Varchar2(50 Byte)		1	INVENTARIOS - T_AREAS	DESCRIP	VARCHAR2(200 BYTE)

3.6.3.3. Dimensión Artículo.

Contiene la Información acerca de los Artículos que maneja el Instituto, los mismos que pertenecen a un Sub Grupo y estos a su vez a un Grupo.

Nombre de Tabla: DIM_ARTICULO
Tipo de Tabla: Dimensión
Esquema Origen: INVENTARIOS
Tablas Origen: T_ARTICULO, T_GRUPOS_ARTICULOS, T_UND_MED

Destino			Origen				
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Tiempo	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Articulo	Número de Artículo	Integer			INVENTARIOS - T_ARTICULO	NO_ARTICULO	Number(9)
Nombre_Articulo	Descripción del Artículo	Varchar2(100 Byte)		1	INVENTARIOS - T_ARTICULO	DESCRIP	Varchar2(100 Byte)
No_Grupo	Numero de Grupo de Artículo	Integer			INVENTARIOS - T_GRUPOS_ARTICULOS	GRUPO	Number(2)
Nombre_Grupo	Descripción de Grupo de Artículo	Varchar2(80 Byte)		1	INVENTARIOS - T_GRUPOS_ARTICULOS	DESC_GRUPO	Varchar2(80 Byte)
No_SubGrupo	Numero de Sub Grupo de Artículo	Integer			INVENTARIOS - T_GRUPOS_ARTICULOS	SUBGRUPO	Number(2)
Nombre_SubGrupo	Descripción de Sub Grupo de Artículo	Varchar2(100 Byte)		1	INVENTARIOS - T_GRUPOS_ARTICULOS	DESC_SUBGRUPO	Varchar2(40 Byte)
Partida	Numero de Partida Presupuestaria del Grupo de Artículo	Varchar2(30 Byte)		1	INVENTARIOS - T_GRUPOS_ARTICULOS	C_PARTIDA_PRESUP	Varchar2(30 Byte)
Valor_Ultima_Compra	Valor de la Compra del Artículo	Number(15,7)		1	INVENTARIOS - T_ARTICULO	VALOR_UNIT_ULT_COMPRA	Number(15,7)
Fecha_Ultima_Compra	Ultima Fecha de Compra del Artículo	Date		1	INVENTARIOS - T_ARTICULO	F_ULTIMA_COMPRA	Date
Costo	Costo en Dolares del Artículo	Number(15,7)		1	INVENTARIOS - T_ARTICULO	V_COSTO_PROMEDIO	Number(15,7)
Unidad_Medida	Unidad de Medida del Artículo	Varchar2(30 Byte)		1	INVENTARIOS - T_UND_MED	DESCRIP	Varchar2(30 Byte)

3.6.3.4. Dimensión Bodega.

Contiene la Información acerca de las Bodegas que maneja el Instituto.

Nombre de Tabla: DIM_BODEGA
Tipo de Tabla: Dimensión
Esquema Origen: INVENTARIOS
Tabla Origen: T_BODEGA

Nombre - Columna	Destino				Origen		
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Bodega	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Bodega	Número de la Bodega	Integer			INVENTARIOS - T_BODEGA	NO_BODEGA	Number(3)
Nombre_Bodega	Nombre de la Bodega	Varchar2(30 Byte)		1	INVENTARIOS - T_BODEGA	DESCRIP	Varchar2(30 BYTE)

3.6.3.5. Dimensión Centro de Gestión

Contiene la Información de los Centros de Gestión de la Institución.

Nombre de Tabla: DIM_CENTRO_GESTION
Tipo de Tabla: Dimensión
Esquema Origen: FIANCIERO
Tablas Origen: T_CENTRO_GESTION, T_EMPRESA

Destino		Origen					
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Centro_Gestion	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Empresa	Número de la Empresa	Integer			FIANCIERO - T_EMPRESA	NO_EMPRESA	NUMBER(3)
Nombre_Empresa	Nombre de la Empresa	Varchar2(100 Byte)		1	FIANCIERO - T_EMPRESA	DESCRIP	VARCHAR2(50 BYTE)
No_Centro_Gestion	Número del Centro de Gestion	Integer			FIANCIERO - T_CENTRO_GESTION	CODIGO_CENTRO_GESTION	NUMBER(3)
Nombre_Centro_Gestion	Nombre del Centro de Gestion	Varchar2(100 Byte)		1	FIANCIERO - T_CENTRO_GESTION	DESCRIPCION	VARCHAR2(50 BYTE)

3.6.3.6. Dimensión Cliente.

Contiene la información acerca del cliente que realiza la compra de Ítems que ofrece el Instituto.

Nombre de Tabla: DIM_CLIENTE
Tipo de Tabla: Dimensión
Esquema Origen: PTV
Tablas Origen: TENTIDAD

Nombre - Columna	Destino			Origen			
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Cliente	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
Cedula_Ruc_Cliente	Cedula ó RUC del Cliente	Varchar2(15 Byte)			PTV - TENTIDAD	CIU	Varchar2(15 Byte)
Nombre_Cliente	Nombre y Apellido del Cliente	Varchar2(150 Byte)		1	PTV - TENTIDAD	NOMBRES , APELLIDOS	Varchar2(100 Byte) Varchar2(50 Byte)
Direccion_Cliente	Dirección de residencia del Cliente	Varchar2(100 Byte)		1	PTV - TENTIDAD	DIRECCION	Varchar2(100 Byte)

3.6.3.7. Dimensión Empleado.

Contiene la información acerca del empleado del Instituto.

Nombre de Tabla: DIM_EMPLEADO
Tipo de Tabla: Dimensión
Esquema Origen: RH
Tablas Origen: T_DAT_EMPLEADO

Nombre - Columna	Destino				Origen		
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Empleado	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Empleado	Número del Empleado en la Institucion	Integer			RH - T_DAT_EMPLEADO	NO_PERSONA	NUMBER(6)
Cedula_Empleado	Cedula del Empleado	Varchar2(10 Byte)			RH - T_DAT_EMPLEADO	NO_CEDULA	Varchar2(10 Byte)
Nombre_Empleado	Nombre del Empleado	Varchar2(100 Byte)		1	RH - T_DAT_EMPLEADO	PRI_APELLIDO, SEG_APELLIDO	VARCHAR2(20 BYTE), VARCHAR2(20 BYTE)

3.6.3.8. Dimensión Factura

Contiene la información acerca de la factura que se emite para la venta por parte del Instituto.

Nombre de Tabla: DIM_FACTURA

Tipo de Tabla: Dimensión

Esquema Origen: PTV

Tablas Origen: TFACTURA

Destino				Origen			
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Factura	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Factura	Número de la Factura	Integer			PTV - TFACTURA	ID_FACTURA	Number(11)
Estado_Factura	Estado de la Factura.	Varchar2(15 Byte)		1	PTV - TFACTURA	ESTADO	Varchar2(1 Byte)
Periodo_Factura	Numero de Periodo en el q se realizo la factura de compra.	Integer			PTV - TFACTURA	ID_PERIODO	Number(3)

3.6.3.9. Dimensión Ítem.

Contiene la Información de los Ítems (Productos) que el Instituto vende al público en general.

Nombre de Tabla: DIM_ITEM
 Tipo de Tabla: Dimensión
 Esquema Origen: PTV
 Tablas Origen: TFAMILIA, TITEM, TUNIDADMEDIDA

Nombre - Columna	Destino				Origen		
	Descripción	Tipo de Dato/Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Item	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
Id_Familia	Número único de Familia	Integer			PTV - TFAMILIA	ID_FAMILIA	Number(4)
Nombre_Familia	Descripción de la Familia de Item	Varchar(100 Byte)		1	PTV - TFAMILIA	DESC_FAMILIA	Varchar2(100 Byte)
Id_Item	Número único de Item	Integer			PTV - TITEM	ID_ITEM	Number(7)
Nombre_Item	Descripción del Item	Varchar(120 Byte)		1	PTV - TITEM	DESC_ITEM	Varchar2(120 Byte)
Iva_Item	Si el Item tiene o no IVA	Varchar(2 Byte)		1	PTV - TITEM	L_IVA	Number
Estado_Item	Estado del Item	Varchar(15 Byte)		1	PTV - TITEM	ESTADO	Varchar2(2 Byte)
Item_Inventariado	Si el Item esta o no en Inventario	Varchar(2 Byte)		1	PTV - TITEM	L_INVENTARIADO	Number
Unidad_Medida	Unidad de Medida del Item	Varchar(15 Byte)		1	PTV - UNIDADMEDIDA	DESC_UNIDAD	Varchar2(100 Byte)

3.6.3.10. Dimensión Movimiento.

Contiene la información de los Movimientos que se produce en el negocio de inventarios.

Nombre de Tabla: DIM_MOVIMIENTO
Tipo de Tabla: Dimensión
Esquema Origen: INVENTARIOS
Tablas Origen: T_TIPO_MOV_INV, T_MOV, T_DET_MOV

Destino		Origen					
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Movimiento	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Movimiento	Número de Movimiento	Integer			INVENTARIOS - T_MOV	NO_MOV	NUMBER(9)
Tipo_Movimiento	Tipo de Movimiento	Varchar2(30 Byte)		1	INVENTARIOS - T_TIPO_MOV_INV	DESCRIP	Varchar2(30 Byte)
Periodo_Movimiento	Periodo del Pedido	Integer			INVENTARIOS - T_MOV	PERIODO_CODIGO	NUMBER(3)

3.6.3.11. Dimensión Orden de Compra.

Contiene la Información acerca de las Órdenes de Compra que se producen en función de la existencia de Artículos en la Institución.

Nombre de Tabla: DIM_ORDEN_COMPRA
Tipo de Tabla: Dimensión
Esquema Origen: INVENTARIOS
Tablas Origen: T_ORDEN_COMPRA, T_DET_ORDEN_COMPRA

Destino			Origen				
Nombre - Columna	Descripción	Tipo de Dato/Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Orden_Comp	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Orden	Número de Orden de Compra	Integer			INVENTARIOS - T_ORDEN_COMPRA	NO_ORDEN	Number(9)
Estado_Orden	Estado de la Orden de Compra	Varchar2(15 Byte)		1	INVENTARIOS - T_ORDEN_COMPRA	ESTADO	Varchar2(1 Byte)
No_Factura	Numero de Factura de la Orden de Compra	Integer			INVENTARIOS - T_DET_ORDEN_COMPRA	NO_FACTURA	Number(9)
Periodo_Orden	Periodo de la realización de la Orden	Integer			INVENTARIOS - T_ORDEN_COMPRA	PERIODO_CODIGO	Number(3)

3.6.3.12. Dimensión Pedido.

Contiene Información acerca de los Pedidos que se realizan en Inventarios.

Nombre de Tabla: DIM_PEDIDO
Tipo de Tabla: Dimensión
Esquema Origen: INVENTARIOS
Tablas Origen: T_PEDIDO, T_DET_PEDIDO

Destino			Origen				
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Pedido	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Pedido	Numero de Pedido	Integer			INVENTARIOS - T_PEDIDO	NO_PEDIDO	NUMBER(9)
Tipo_Pedido	Tipo de Pedido	Varchar2(30 Byte)		1	INVENTARIOS - T_PEDIDO	TIPO_PEDIDO	VARCHAR2(1 BYTE)
Estado_Pedido	Estado del Pedido	Varchar2(30 Byte)		1	INVENTARIOS - T_PEDIDO	ESTADO	VARCHAR2(1 BYTE)
Importacion_Pedido	Si es o no un Pedido de Importacion	Varchar2(2 Byte)		1	INVENTARIOS - T_PEDIDO	FUE_IMPORTACION	NUMBER(1)
Periodo_Pedido	Periodo en que se realizo el Pedido	Integer			INVENTARIOS - T_PEDIDO	PERIODO_CODIGO	NUMBER(3)

3.6.3.13. Dimensión Proforma

Contiene información acerca de las proformas que se realizan para la venta de Ítems en la Institución.

Nombre de Tabla: DIM_PROFORMA
Tipo de Tabla: Dimensión
Esquema Origen: PTV
Tablas Origen: TPROFORMA, TDETPROFORMA

Destino				Origen			
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Proforma	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Proforma	Numero de la proforma	Integer			PTV - TPROFORMA	ID_PROFORMA	NUMBER(10)
Tipo_Proforma	Tipo de Proforma	Varchar2(20 Byte)		1	PTV - TPROFORMA	TIPO_PROFORMA	Varchar2(2 Byte)
Estado_Proforma	Estado de la Proforma	Varchar2(10 Byte)		1	PTV - TPROFORMA	ESTADO	Varchar2(15 Byte)
Anticipo_Proforma	Proforma de Anticipo	Varchar2(2 Byte)		1	Sistema ETL		
Periodo_Proforma	Periodo de la Proforma	Integer			PTV - TPROFORMA	ID_PERIODO	NUMBER(3)

3.6.3.14. Dimensión Proveedor

Contiene información acerca de los proveedores a los cuales el Instituto realiza la compra de artículos.

Nombre de Tabla: DIM_PROVEEDOR
Tipo de Tabla: Dimensión
Esquema Origen: FINANCIERO
Tablas Origen: T_PROVEEDOR

Nombre - Columna	Destino				Origen		
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Proveedor	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Proveedor	Número del Proveedor	Integer			FINANCIERO - T_PROVEEDOR	NO_PROVEEDOR	NUMBER(10)
Nombre_Proveedor	Nombre del Proveedor	Varchar2(100 Byte)		1	FINANCIERO - T_PROVEEDOR	NOMBRE	Varchar2(100 Byte)
Cedula_Ruc_Proveedor	Cedula o Ruc del Proveedor	Varchar2(25 Byte)		1	FINANCIERO - T_PROVEEDOR	NO_CEDULA, RUC	Varchar2(10 Byte), Varchar2(25 Byte)

3.6.3.15. Dimensión Proyecto.

Contiene la Información acerca de los Proyectos con el Proceso que maneja el Instituto.

Nombre de Tabla: DIM_PROYECTO
Tipo de Tabla: Dimensión
Esquema Origen: PROYECTOS
Tablas Origen: T_PRY_PROYECTO, T_PRY_MACROPROCESO

Nombre - Columna	Destino				Origen		
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Proyecto	Clave Primaria Subrogada	Integer	Primaria		Sistema ETL		
No_Proyecto	Número de Proyecto	Integer			PROYECTOS - T_PRY_PROYECTO	COD_PROYECTO	Number(9)
Nombre_Proyecto	Nombre del Proyecto	Varchar2(100 Byte)		1	PROYECTOS - T_PRY_PROYECTO	NOMBRE_PROYECTO	Varchar2(100 Byte)
No_Proceso	Número del Proceso de la Institución	Integer			PROYECTOS - T_PRY_MACROPROCESO	CODIGO_MACRO	Number(6)
Nombre_Proceso	Nombre del Proceso de la Institución	Varchar2(100 Byte)		1	PROYECTOS - T_PRY_MACROPROCESO	DESCRIPCION	Varchar2(100 Byte)

3.6.3.16. Dimensión Sucursal.

Contiene Información acerca de las Sucursales que maneja el Instituto.

Nombre de Tabla: DIM_SUCURSAL
 Tipo de Tabla: Dimensión
 Esquema Origen: PTV
 Tablas Origen: TSUCURSAL

Destino			Origen				
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Sucursal	Clave Primaria Origen	Integer	Primaria		PTV - TSUCURSAL	ID_SUCURSAL	Number(3)
Nombre_Sucursal	Descripción de la sucursal	Varchar2(100 Byte)		1	PTV - TSUCURSAL	DESC_SUCURSAL	Varchar2(100 Byte)
Dirección_Sucursal	Dirección de la sucursal	Varchar2(100 Byte)		1	PTV - TSUCURSAL	DIRECCION	Varchar2(100 Byte)

3.6.4. PASO 4: IDENTIFICAR LOS HECHOS Y MAPEAR LOS DATOS ORIGEN.

Los Hechos a definir soportarán los requerimientos establecidos cumpliendo con la granularidad especificada para cada negocio. El mapeo de los datos será especificado en cada Hecho.

En el desarrollo del DataWarehouse se definirán hechos referentes a las transacciones que se realicen de acuerdo al negocio. Los hechos tendrán Dimensiones en común así como también Dimensiones específicas para el negocio.

Las medidas que se implementan en los hechos cumplen con el detalle de granularidad que se especificaron anteriormente.

3.6.4.1. Pedidos.

Contiene los indicadores (medidas) de la transacción de Pedidos que se realiza en el Negocio de Inventarios, y las claves foráneas que permitirán relacionar con las Dimensiones.

Nombre de Tabla: Hec_Pedidos_Inv

Tipo de Tabla: Hecho

Esquema Origen: INVENTARIOS

Tablas Origen: T_DET_PEDIDOS, T_DET_ORDEN_COMPRA, DIM_TIEMPO, DIM_ARTICULO, DIM_BODEGA, DIM_PROYECTO, DIM_AREA, DIM_ORDEN_COMPRA, DIM_PEDIDO

Destino			Origen				
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Tiempo	Clave Foranea de la Dimension Tiempo	Integer	Primaria - Foranea		DIM_TIEMPO	KEY_TIEMPO	Integer
Key_Articulo	Clave Foranea de la Dimension Articulo	Integer	Primaria - Foranea		DIM_ARTICULO	KEY_ARTICULO	Integer
Key_Bodega	Clave Foranea de la Dimension Bodega	Integer	Primaria - Foranea		DIM_BODEGA	KEY_BODEGA	Integer
Key_Proyecto	Clave Foranea de la Dimension Proyecto	Integer	Primaria - Foranea		DIM_PROYECTO	KEY_PROYECTO	Integer
Key_Area	Clave Foranea de la Dimension Area	Integer	Primaria - Foranea		DIM_AREA	KEY_AREA	Integer
Key_Orden_Compra	Clave Foranea de la Dimension Orden de Compra	Integer	Primaria - Foranea		DIM_ORDEN_COMPRA	KEY_ORDEN_COMPRA	Integer
Key_Proveedor	Clave Foranea de la Dimension Proveedor	Integer	Primaria - Foranea		DIM_PROVEEDOR	KEY_PROVEEDOR	Integer
Key_Pedido	Clave Foranea de la Dimension Pedido	Integer	Primaria - Foranea		DIM_PEDIDO	KEY_PEDIDO	Integer
Cantidad_Pedida	Cantidad pedida de articulos.	Number(14,5)		1	INVENTARIOS - T_DET_PEDIDO	CANT_PEDIDA	Number(14,5)
Valor_Total	Valor total del pedido.	Number(15,7)		1	Sistema ETL		

3.6.4.2. Movimientos.

Contiene los indicadores (medidas) de la transacción de Movimientos que se realiza en el Negocio de Inventarios, y las claves foráneas que permitirán relacionar con las Dimensiones.

Nombre de Tabla: Hec_Movimientos_Inv
Tipo de Tabla: Hecho
Esquema Origen: INVENTARIOS
Tablas Origen: T_DET_MOV, DIM_TIEMPO, DIM_ARTICULO, DIM_BODEGA, DIM_PROYECTO, DIM_AREA, DIM_ORDEN_COMPRA, DIM_PEDIDO_INV, DIM_MOVIMIENTO_INV

Nombre - Columna	Destino			Origen			
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Tiempo	Clave Foránea de la Dimension Tiempo	Integer	Primaria - Foránea		DIM_TIEMPO	KEY_TIEMPO	Integer
Key_Articulo	Clave Foránea de la Dimension Articulo	Integer	Primaria - Foránea		DIM_ARTICULO	KEY_ARTICULO	Integer
Key_Bodega	Clave Foránea de la Dimension Bodega	Integer	Primaria - Foránea		DIM_BODEGA	KEY_BODEGA	Integer
Key_Proyecto	Clave Foránea de la Dimension Proyecto	Integer	Primaria - Foránea		DIM_PROYECTO	KEY_PROYECTO	Integer
Key_Area	Clave Foránea de la Dimension Area	Integer	Primaria - Foránea		DIM_AREA	KEY_AREA	Integer
Key_Orden_Compra	Clave Foránea de la Dimension Orden de Compra	Integer	Primaria - Foránea		DIM_ORDEN_COMPRA	KEY_ORDEN_COMPRA	Integer
Key_Proveedor	Clave Foránea de la Dimension Proveedor	Integer	Primaria - Foránea		DIM_PROVEEDOR	KEY_PROVEEDOR	Integer
Key_Pedido	Clave Foránea de la Dimension Pedido	Integer	Primaria - Foránea		DIM_PEDIDO_INV	KEY_PEDIDO	Integer
Key_Movimiento	Clave Foránea de la Dimension Movimiento	Integer	Primaria - Foránea		DIM_MOVIMIENTO_INV	KEY_MOVIMIENTO	Integer
Cantidad	Cantidad de articulos que se trasladaron	Number(14,5)		1	INVENTARIOS - T_DET_MOV	CANTIDAD	Number(14,5)
Valor_Total	Valor Total - Cantidad por el Costo Unitario	Number(15,7)		1	Sistema ETL		

3.6.4.3. Ventas

Contiene los indicadores (ó métricas) de la transacción de Venta que se realiza en el Negocio de Punto de Ventas, y las claves foráneas que permitirán relacionar con las Dimensiones.

Nombre de Tabla: HEC_VENTAS_PTV
 Tipo de Tabla: Hecho
 Esquema Origen: PTV
 Tablas Origen: TDETPROFORMA, TPROFORMA, TANTICIPO, TFACTURA, DIM_TIEMPO, DIM_CLIENTE, DIM_SUCURSAL, DIM_CENTRO_GESTION, DIM_FACTURA, DIM_ITEM, DIM_PROYECTO, DIM_PROFORMA, DIM_EMPLEADO

Destino			Origen				
Nombre - Columna	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Tiempo	Clave Foranea de la Dimension Tiempo	Integer	Primaria - Foranea		DIM_TIEMPO	KEY_TIEMPO	Integer
Key_Cliente	Clave Foranea de la Dimension Cliente	Integer	Primaria - Foranea		DIM_CLIENTE	KEY_CLIENTE	Integer
Key_Sucursal	Clave Foranea de la Dimension Sucursal	Integer	Primaria - Foranea		DIM_SUCURSAL	KEY_SUCURSAL	Integer
Key_Centro_Gestion	Clave Foranea de la Dimension Centro de Gestion	Integer	Primaria - Foranea		DIM_CENTRO_GESTION	KEY_CENTRO_GESTION	Integer
Key_Factura	Clave Foranea de la Dimension Factura	Integer	Primaria - Foranea		DIM_FACTURA	KEY_FACTURA	Integer
Key_Item	Clave Foranea de la Dimension Item	Integer	Primaria - Foranea		DIM_ITEM	KEY_ITEM	Integer
Key_Proyecto	Clave Foranea de la Dimension Proyecto	Integer	Primaria - Foranea		DIM_PROYECTO	KEY_PROYECTO	Integer
Key_Proforma	Clave Foranea de la Dimension Proforma	Integer	Primaria - Foranea		DIM_PROFORMA	KEY_PROFORMA	Integer
Key_Empleado	Clave Foranea de la Dimension Empleado	Integer	Primaria - Foranea		DIM_EMPLEADO	KEY_EMPLEADO	Integer
Cantidad	Cantidad de items vendidos	Number(15,7)		1	PTV - TDETPROFORMA	CANTIDAD	Number(15,7)
Valor_Total	Valor total de Venta Cantidad por Precio Unitario	Number(17,8)		1	Sistema ETL		Number(17,8)

3.6.4.4. Número de Ventas

Contiene el indicador (medida) que refleja el Número(os) de Ventas que se realiza en el Negocio de Punto de Ventas, y las claves foráneas que permitirán relacionar con las Dimensiones.

Nombre de Tabla: HEC_NUM_VENTAS_PTV
Tipo de Tabla: Hecho
Esquema Origen: PTV
Tablas Origen: TDETPROFORMA, TPROFORMA, TANTICIPO, TFACTURA, DIM_TIEMPO, DIM_CLIENTE, DIM_SUCURSAL, DIM_CENTRO_GESTION, DIM_FACTURA, DIM_ITEM, DIM_PROYECTO, DIM_PROFORMA, DIM_EMPLEADO

Nombre - Columna	Destino			Origen			
	Descripción	Tipo de Dato / Tamaño	Clave	SCD	Sistema / Esquema - Tabla	Campo Origen	Tipo de Dato Origen
Key_Tiempo	Clave Foranea de la Dimension Tiempo	Integer	Primaria - Foranea		DIM_TIEMPO	KEY_TIEMPO	Integer
Key_Cliente	Clave Foranea de la Dimension Cliente	Integer	Primaria - Foranea		DIM_CLIENTE	KEY_CLIENTE	Integer
Key_Sucursal	Clave Foranea de la Dimension Sucursal	Integer	Primaria - Foranea		DIM_SUCURSAL	KEY_SUCURSAL	Integer
Key_Factura	Clave Foranea de la Dimension Factura	Integer	Primaria - Foranea		DIM_FACTURA	KEY_FACTURA	Integer
Key_Empleado	Clave Foranea de la Dimension Empleado	Integer	Primaria - Foranea		DIM_EMPLEADO	KEY_EMPLEADO	Integer
Numero_Ventas	Numero de Ventas realizadas	Number		1	PTV - TFACTURA	ID_FACTURA	Number(11)

3.6.5. DISEÑO DEL MODELO DIMENSIONAL INVENTARIOS

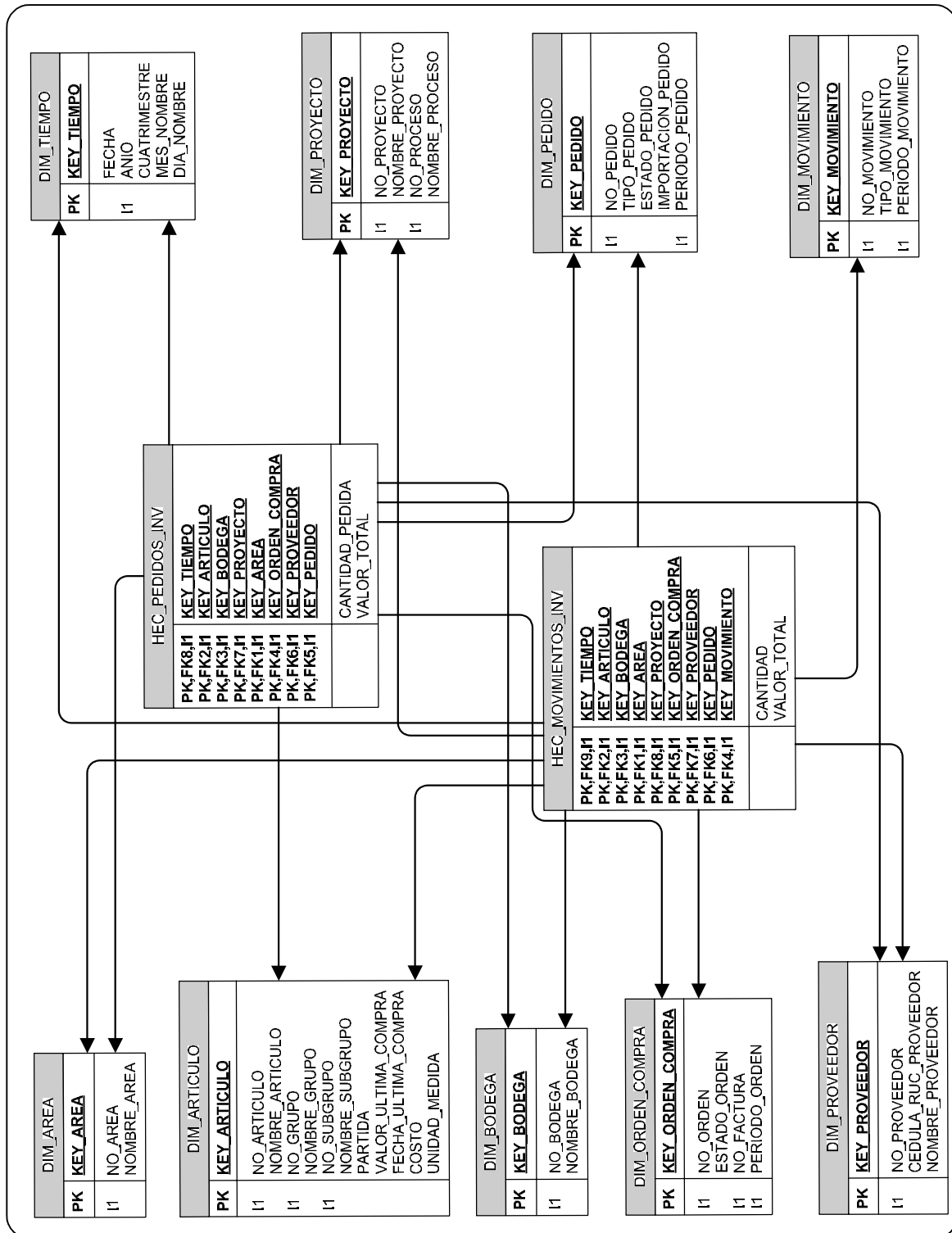


Figura 3.5: Modelo Dimensional - Inventarios

3.6.6. DISEÑO DEL MODELO DIMENSIONAL PUNTO DE VENTAS

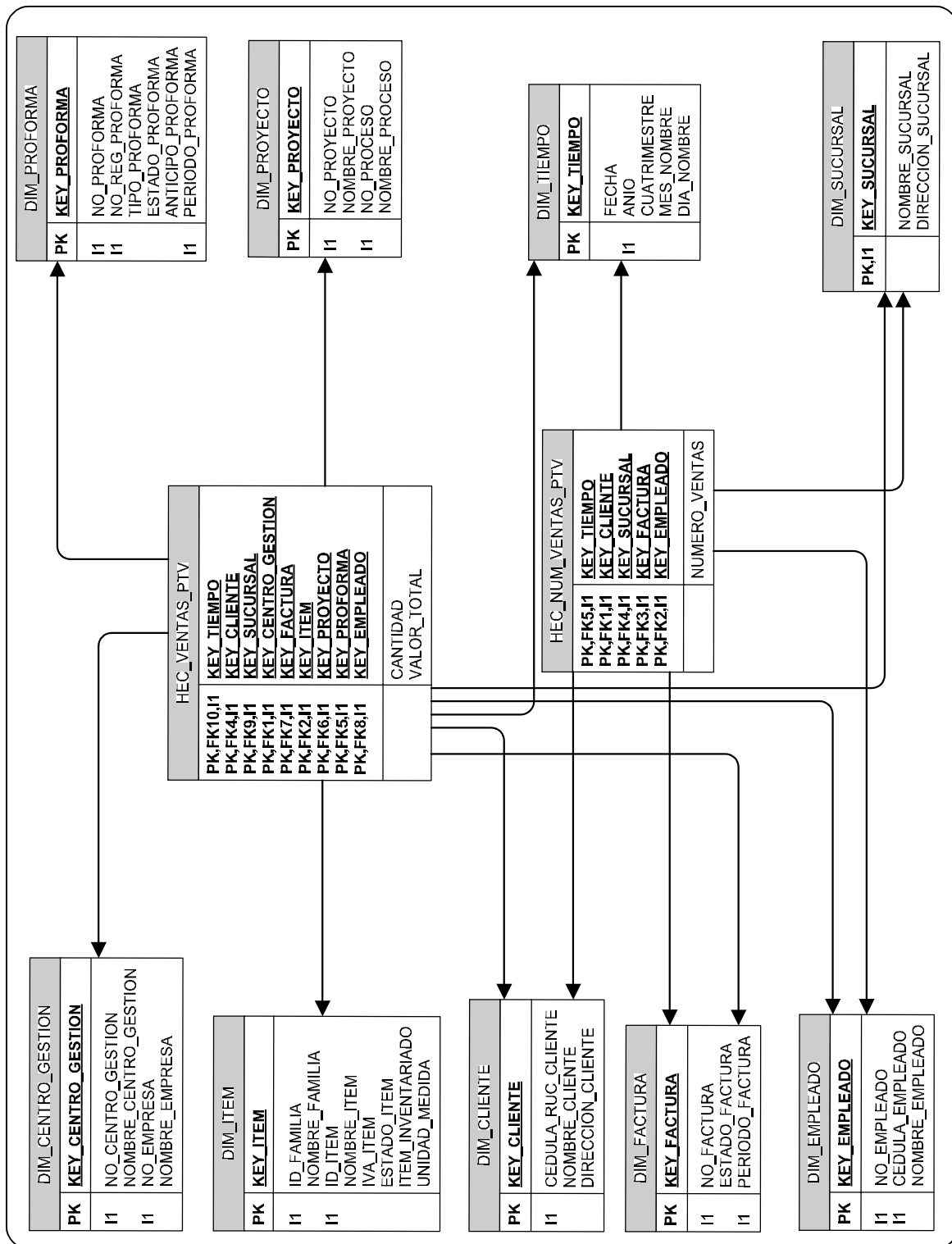


Figura 3.6: Modelo Dimensional - Punto de Ventas

3.6.7. DISEÑO DEL MODELO DIMENSIONAL TOTAL

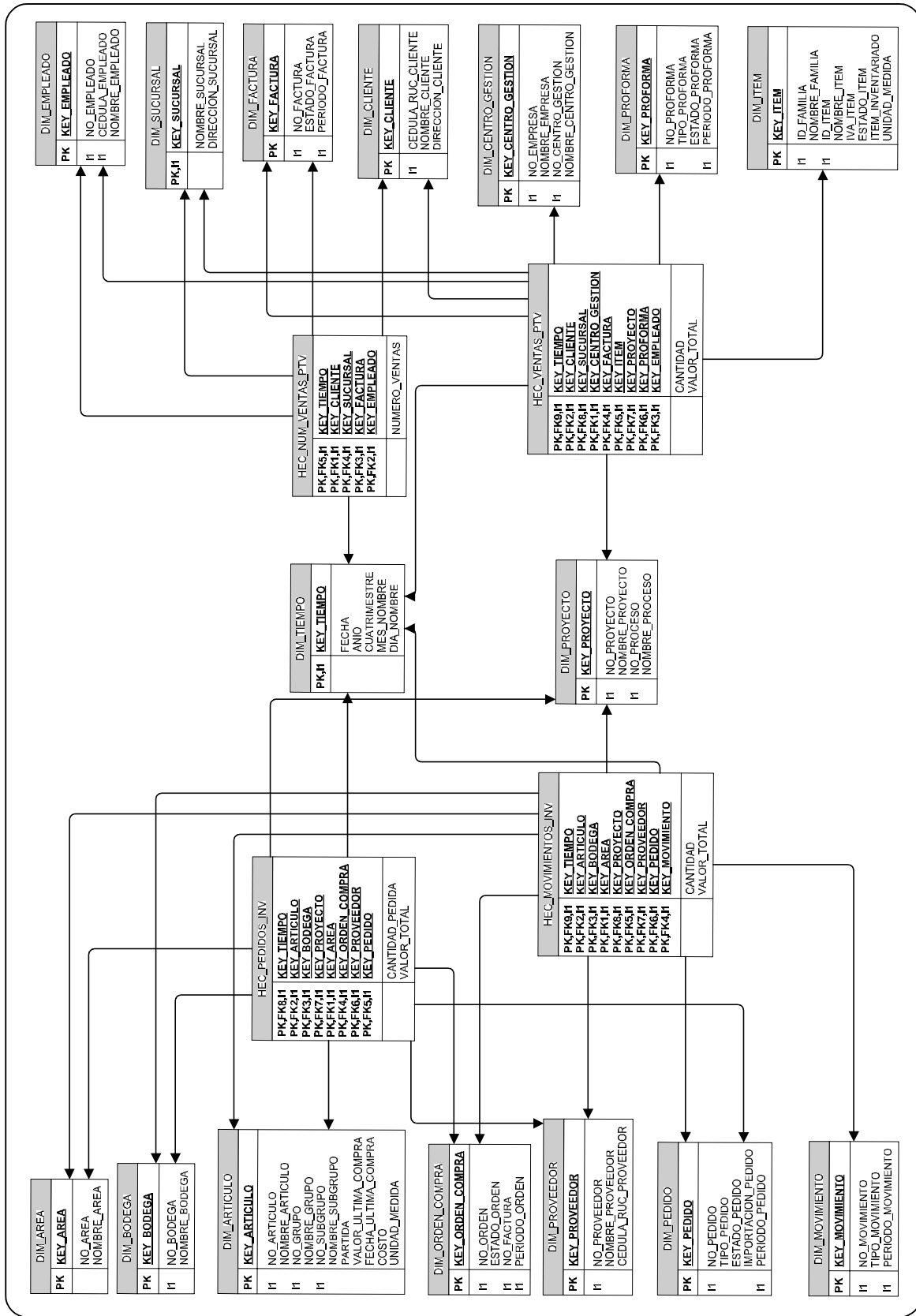


Figura 3.7: Modelo Dimensional - DataWarehouse

3.6.8. DATA WAREHOUSE BUS MATRIX FINAL

Proceso de Negocio	Tablas de Hecho	Granularidad	Dimensiones Comunes														
			Area	Bodega	Centro de Gestion	Cliente	Factura	Item	Movimiento	Orden de Compra	Pedido	Proforma	Proveedor	Sucursal	Empleado	Tiempo	
Inventarios	Pedidos	1 registro por cada transaccion de Pedido	X	X	X					X	X	X				X	
	Movimientos	1 registro por cada transaccion de Movimiento	X	X			X	X	X		X	X				X	
Punto de Ventas	Ventas	1 registro por cada transaccion de Venta de Item							X							X	X
	Numero Ventas	1 registro por cada transaccion de Venta Total							X							X	X

Data Warehouse - Bus Matrix Detallado

3.7. DISEÑO FÍSICO

El Diseño Físico se realizó en función a la Base de Datos ya mencionada, Oracle 10g, como son Esquemas, Espacios de Tabla, etc. Los Scripts de creación para la Base de datos Origen se encuentra en el Anexo E y para los scripts del DataWarehouse en el Anexo F.

3.7.1. BASE DE DATOS ORIGEN

Primero se empezará especificando los Scripts de creación para la base de datos Origen, lo que permitirá llevar a cabo la técnica de Extracción CDC (Change Data Capture) la misma que será explicada en el Diseño ETL para su mayor entendimiento así como de la funcionalidad que aportará.

3.7.1.1. ESPACIO DE TABLA (TABLESPACE)

Se creará un Tablespace para almacenar los datos que han cambiado en la base Origen. El tamaño asignado será de 10 MegaByte, con nombre:

- ADMINDATAWH

3.7.1.2. ESQUEMA

Se creará un Esquema o Usuario en la Base de Datos Origen que permita administrar los recursos que serán agregados en ella, así como también a través del cual permita leer los datos para la extracción en el proceso ETL. Los privilegios o permisos asignados para la Base son:

- Lectura de Datos.
- Creación de Tablas (Para su Esquema).
- Creación de cualquier Disparador (Trigger).

3.7.1.3. TABLAS

Se utilizará una única Tabla de Auditoria CDC, para la captura de datos que hayan cambiado o que sean nuevos en la Base Origen. El script de creación se encuentra en el Anexo E.

3.7.1.4. ÍNDICES

Se asignará un índice único para la clave primaria de la tabla de Auditoria.

3.7.1.5. RESTRICCIONES (CONSTRAINTS)

La tabla de Auditoria utilizará una única restricción que es la clave primaria.

3.7.1.6. SECUENCIA

Se utilizará una secuencia incremental para el campo de la clave primaria de la Tabla de Auditoría.

3.7.1.7. DISPARADORES (TRIGGERS)

Disparador para la tabla de Auditoría:

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

Disparadores para las Tablas de Negocio:

Se crearán disparadores para las diferentes tablas que serán utilizadas en el proceso ETL. Mediante estos disparadores se guardarán las claves de los registros nuevos o q han cambiando en la tabla de auditoría.

3.7.2. BASE DE DATOS DESTINO (DATAWAREHOUSE)

Primero de empezará con las especificaciones de la Base Destino, es decir el DataWarehouse.

3.7.2.1. ESPACIO DE TABLA (TABLESPACE)

Se crearán 2 Tablespace que se mencionan a continuación:

- REPOSITORIO_WAREHOUSE

En este espacio de tablas se alojarán todos los Metadatos utilizados por la herramienta ETL para todo el Proceso de integración de datos.

- DATA_WAREHOUSE

En este espacio de tablas se alojarán los datos propios del almacén, es decir todas las dimensiones, hechos, métricas, índices, etc.

3.7.2.2. TABLAS

Se crearán las Tablas del DataWarehouse, tanto lo que serán las Dimensiones así como también los Hechos. Además se creará una Tabla para monitorear en caso de que existan errores en la ejecución ETL, lo que permitirá llevar un control de errores.

3.7.2.3. ÍNDICES

Se definirán los índices para cada tabla de Dimensión, Hecho y Log de Errores.

3.7.2.4. RESTRICCIONES (CONSTRAINTS)

Se creará las restrincciones, lo que son las claves primarias y foráneas, tanto para las Tablas de Dimensiones como para los Hechos y para el Log de Errores.

3.7.2.5. SECUENCIA

Se creará una secuencia para el incremento de registros de la tabla Log de Error.

3.7.2.6. DISPARADORES (TRIGGERS)

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

3.8. DISEÑO Y DESARROLLO DEL PROCESO ETL

DESARROLLO DEL PLAN ETL

A continuación se laborará una Mapa gráfico, el cual permitirá tener una mayor idea de cómo se ejecutará el Proceso ETL.

3.8.1. PASO 1: TRAZAR UN PLAN DE ALTO NIVEL

ÁREA

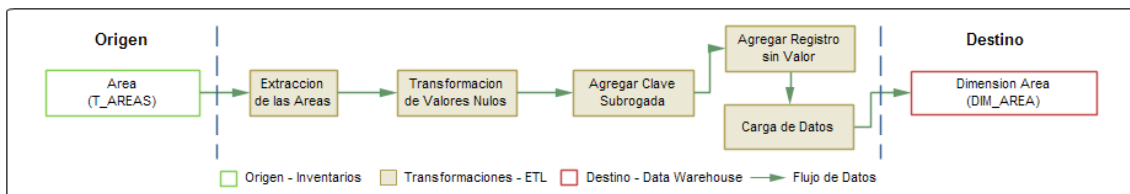


Figura 3.8: Plan ETL – Área

ARTÍCULO

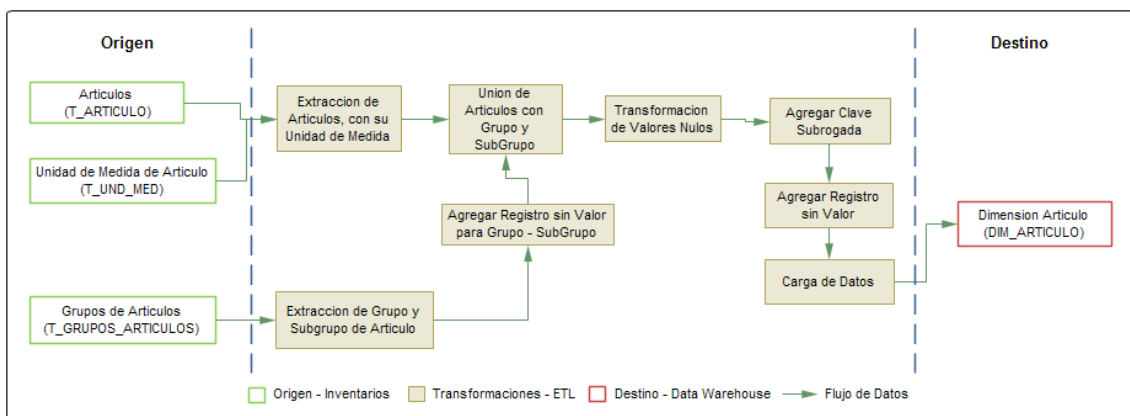


Figura 3.9: Plan ETL – Artículo

BODEGA

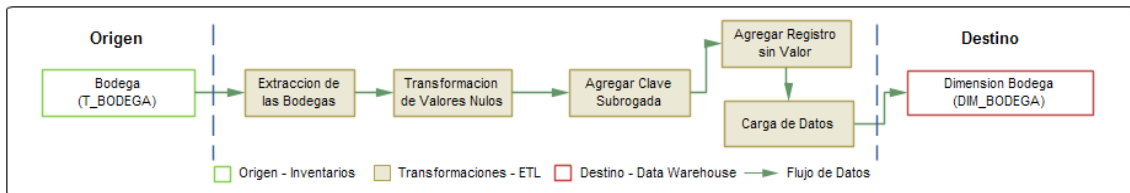


Figura 3.10: Plan ETL – Bodega

CENTRO DE GESTIÓN

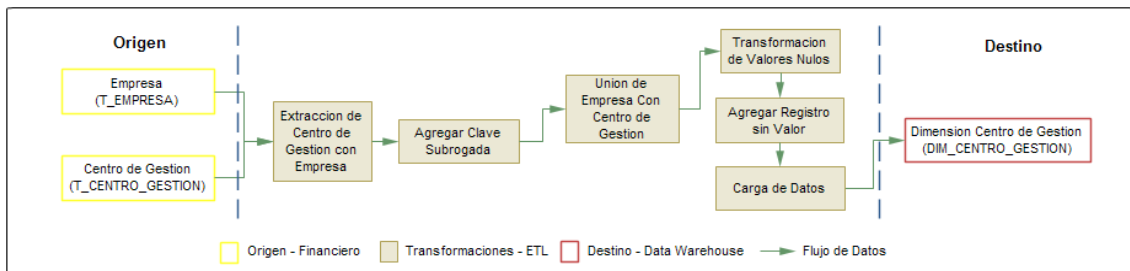


Figura 3.11: Plan ETL – Centro de Gestión

CLIENTE



Figura 3.12: Plan ETL – Cliente

EMPLEADO



Figura 3.13: Plan ETL – Empleado

FACTURA

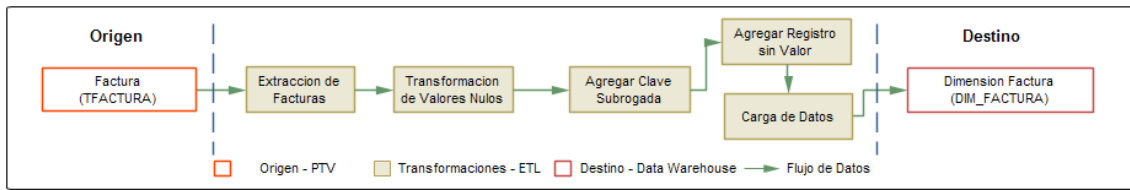


Figura 3.14: Plan ETL – Factura

ÍTEM

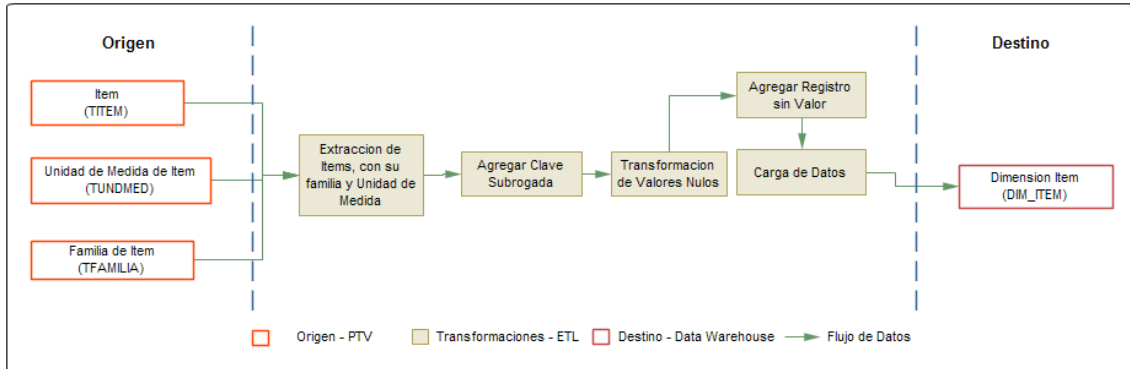


Figura 3.15: Plan ETL – Ítem

MOVIMIENTO

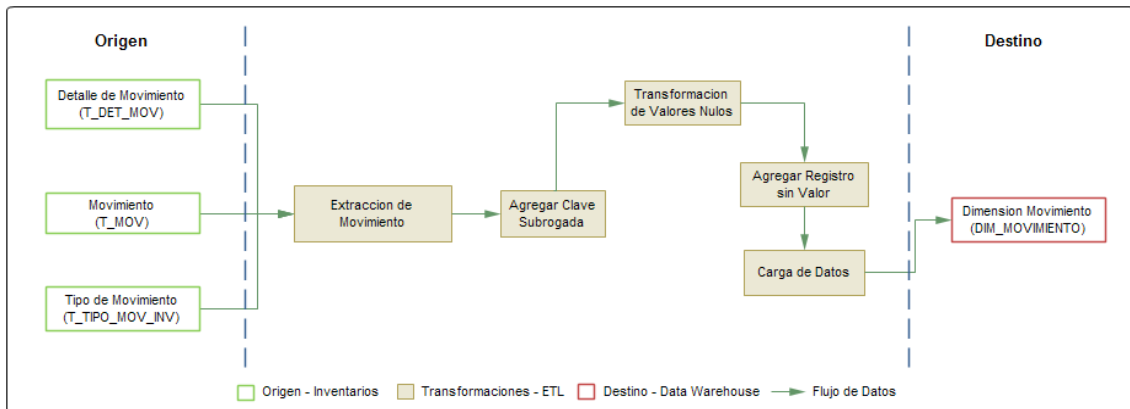


Figura 3.16: Plan ETL – Movimiento

ORDEN DE COMPRA

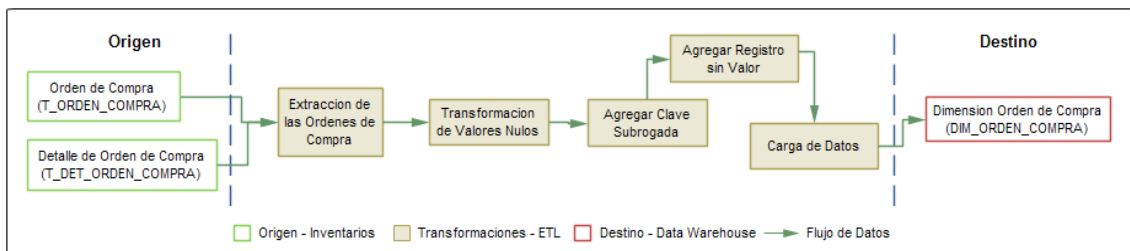


Figura 3.17: Plan ETL – Orden de Compra

PEDIDO

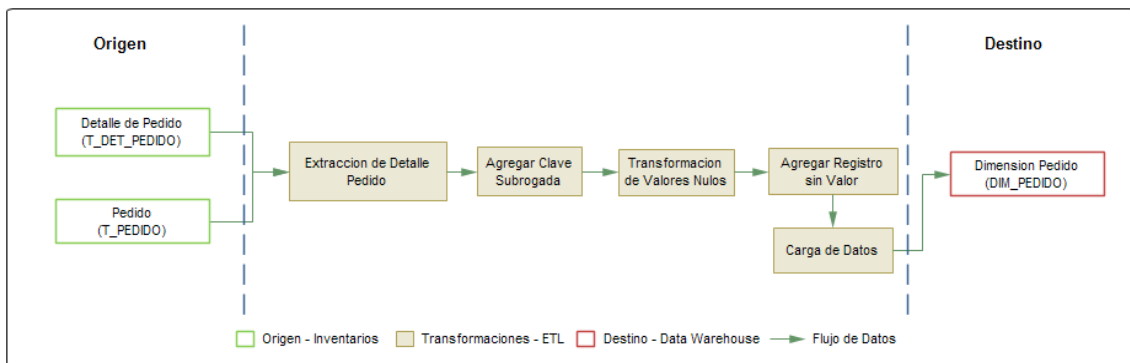


Figura 3.18: Plan ETL – Pedido

PROFORMA

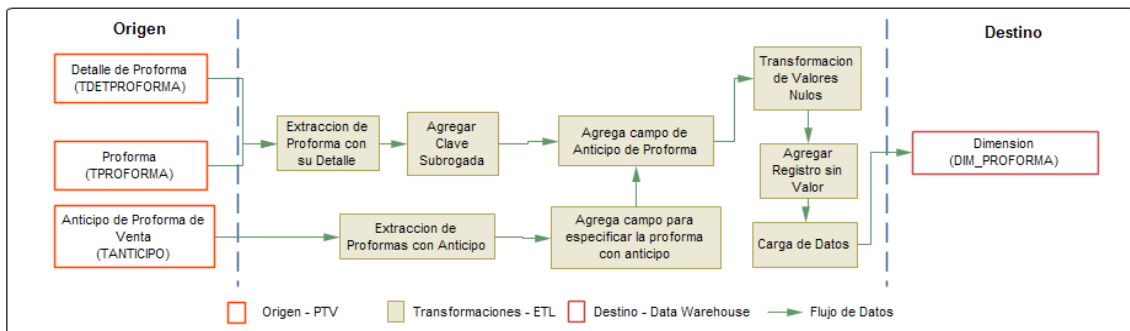


Figura 3.19: Plan ETL – Proforma

PROYECTO

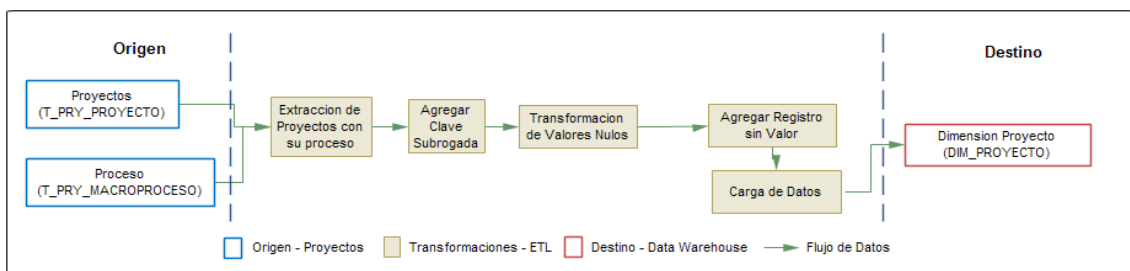


Figura 3.20: Plan ETL – Proyecto

PROVEEDOR

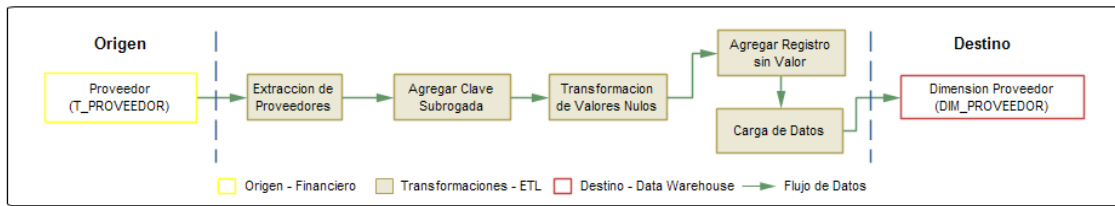


Figura 3.21: Plan ETL – Proveedor

SUCURSAL

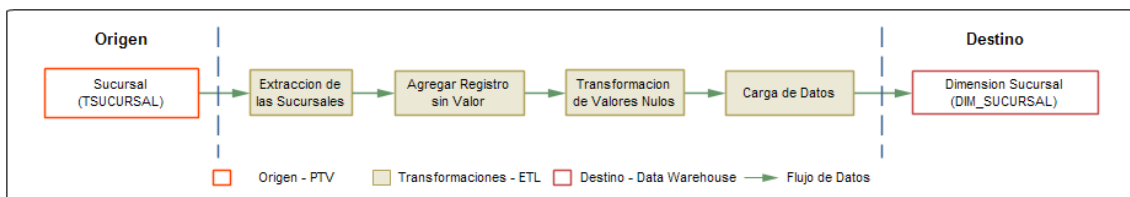


Figura 3.22: Plan ETL – Sucursal

TIEMPO

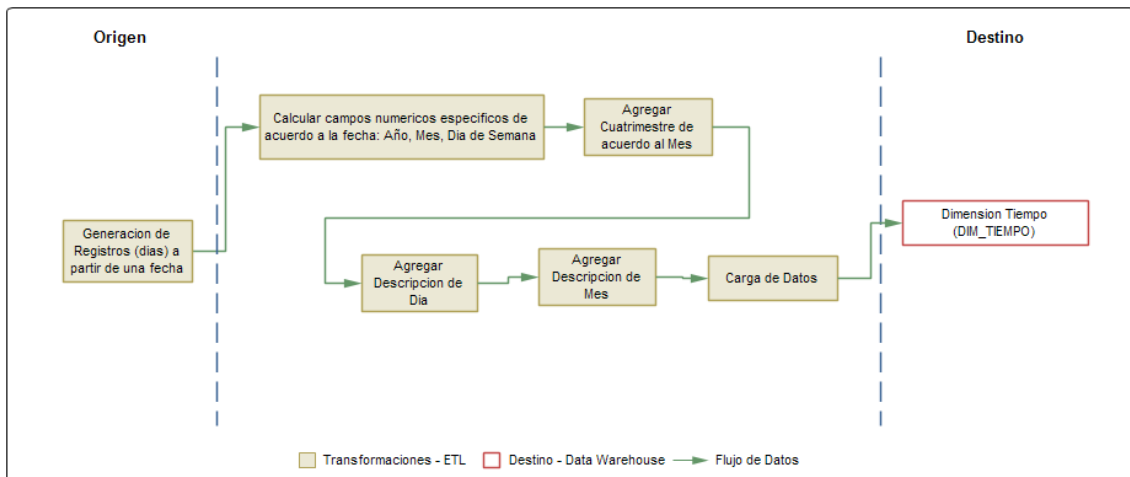


Figura 3.23: Plan ETL – Tiempo

PEDIDOS

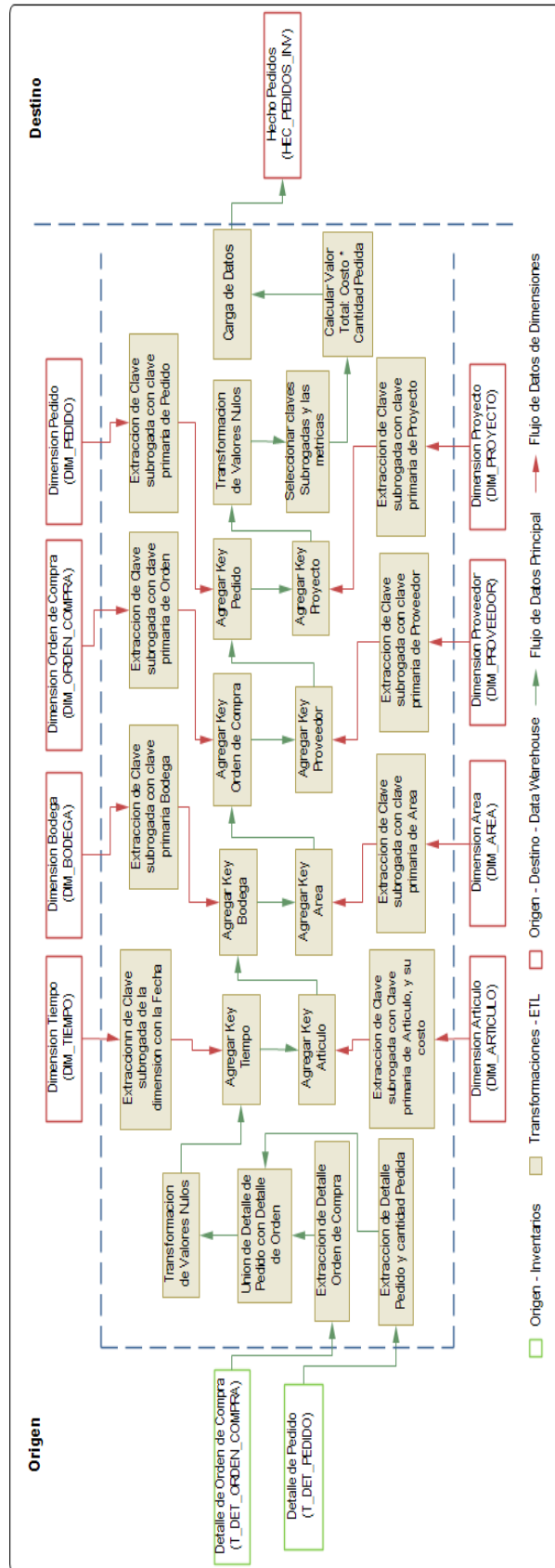


Figura 3.24: Plan ETL – Pedidos

MOVIMIENTOS

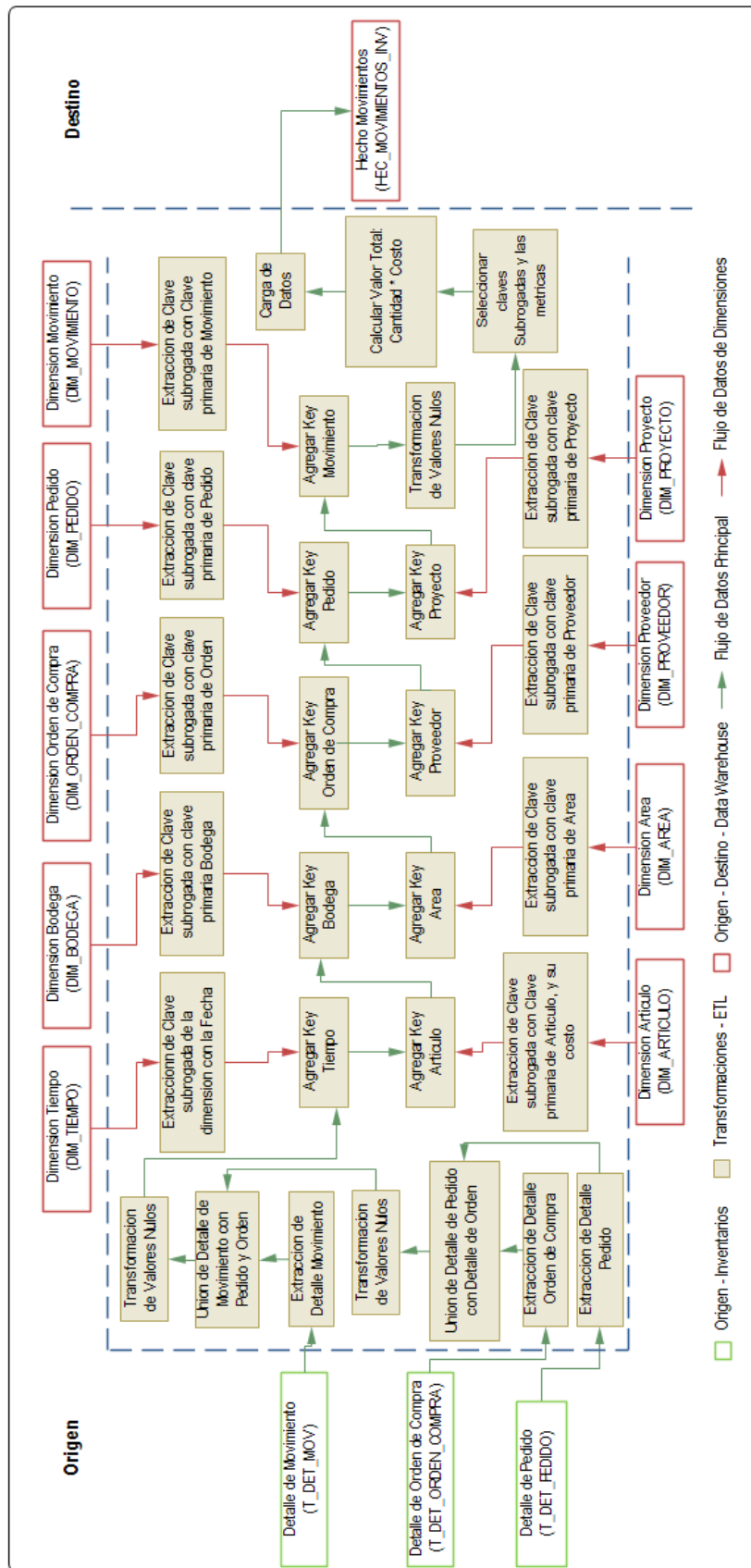


Figura 3.25: Plan ETL – Movimientos

VENTAS

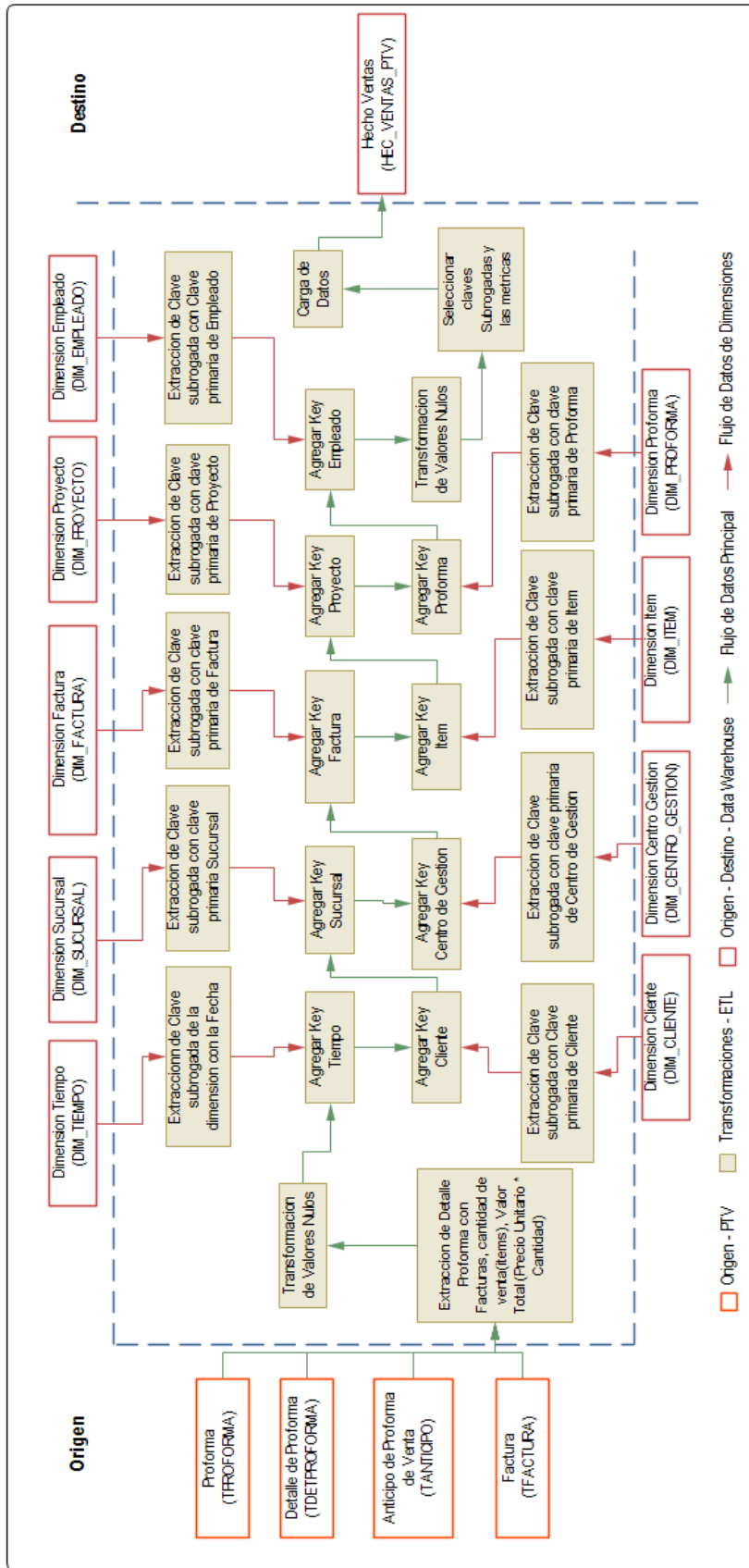


Figura 3.26: Plan ETL – Ventas

NÚMERO VENTAS

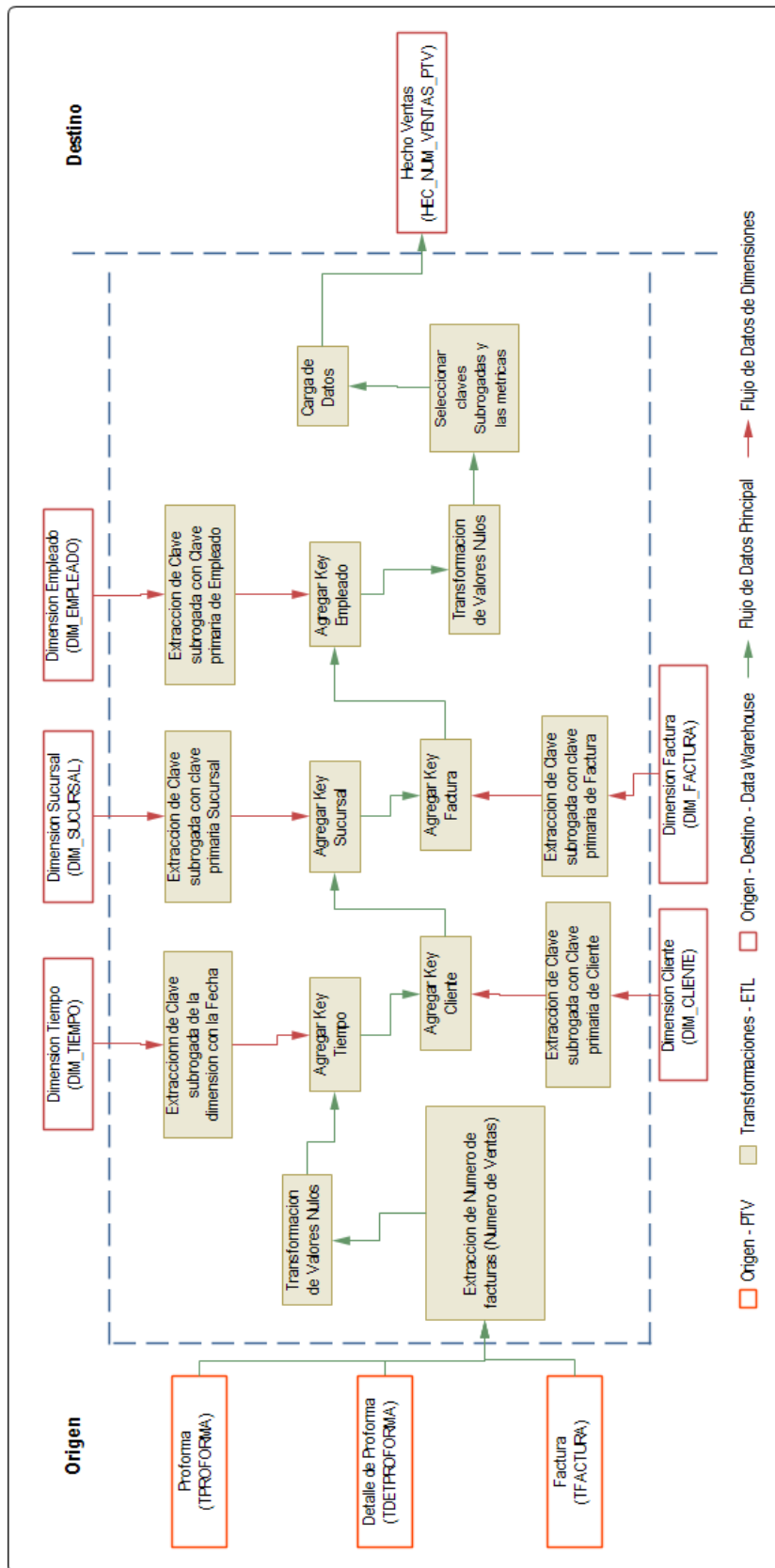


Figura 3.27: Plan ETL – Numero de Ventas

3.8.2. PASO 2: FUNCIONALIDAD DE LA HERRAMIENTA ETL

Se renombro este paso en función a que ya se especifico la Herramienta ETL a utilizar por lo que a continuación se detallará las funcionalidades a utilizar.

Para la elaboración de los Procesos ETL se utilizará dos herramientas del Software ETL de Pentaho. Estas son:

- Spoon, que va a permitir diseñar las transformaciones y trabajos de manera gráfica y amigable.
- Kitchen, pan, permiten ejecutar el Proceso ETL periódicamente mediante scripts batch, de acuerdo al Sistema Operativo.

3.8.3. PASO 3: DESARROLLAR ESTRATEGIAS POR DEFECTO

- Extracción de datos.

La Extracción de los Datos se lo relazará desde una sola Base de datos que se encuentra en producción con nombre: IGM, la misma que posee esquemas de acuerdo a los procesos y/o usuarios de la Institución.

Estos esquemas a utilizar ya se especificó anteriormente en el Modelamiento Dimensional.

- Captura de datos de cambio (CDC – Change Data Capture)

Como parte de la extracción de los datos se procedió a implementar una técnica la cual garantice que solo se extraiga los datos que hayan cambiado o sean nuevos en la Base de Datos Origen.

La manera como se implementará esta técnica se muestra en el grafico a continuación:

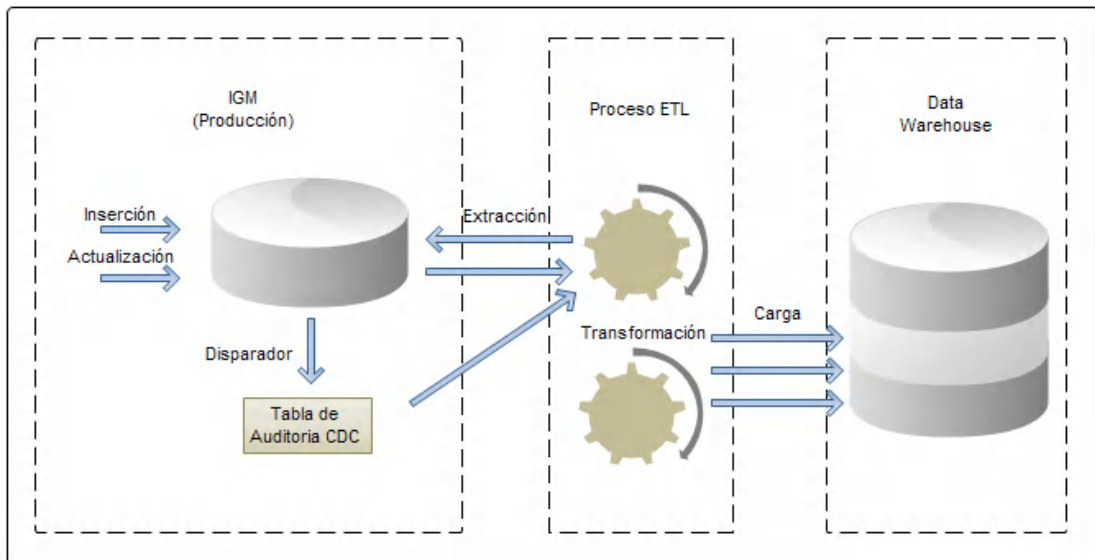


Figura 3.28: Arquitectura CDC (Change Data Capture)

Cualquier inserción de datos o actualización sobre las tablas origen, esto producirá que se lance un disparador, el cual guardará las llaves primaria de estas tablas en un tabla de Auditoría.

La tabla de auditoría estará en la Base Origen recolectado todos estos cambios, los mismos que serán extraídos por la Herramienta ETL para procesar los datos Origen que han sufrido cambios o que sean registros nuevos.

- Administración de cambios en los atributos de las Dimensiones.

Como se especifico en los requerimientos y en modelo dimensional los datos que se encuentran en el DataWarehouse serán actualizados, por lo que los atributos de las tablas destino serán sobrescritos con los datos actuales. De acuerdo a esto se especificara un SCD (Slowly Changing Dimension) de Tipo 1.

Para los atributos a los cuales se especifico este tipo de cambio, se lo puede apreciar en la sección Modelamiento Dimensional en la que se indica el SCD por atributo de acuerdo a cada tabla sea esta Dimensión o Hecho.

- Carga de Datos

La carga de los datos se enfatizo en dos tipos:

- Carga inicial.- Se la realizará cuando sea poblada por primera vez las tablas destino, por ser la primera carga durará más tiempo en poblar el DataWarehouse debido a que extraerá todos los datos de la base Origen.
- Carga Incremental.- Se realizará cuando las tablas destino sean pobladas con los datos que han sido modificados o que sean nuevos en la Base origen, garantizando mayor eficiencia del procesamiento de los datos. La carga incremental se ejecutará subsecuentemente a la primera carga es decir a partir de la segunda carga.

- Frecuencia de carga.

La frecuencia de carga se especifico para que se realice cada día, es decir sea una carga de datos hacia el DataWarehouse diaria.

- Secuencia de los Procesos de Transformaciones.

La secuencia de las transformaciones va a depender de las tablas destino y la dependencia (relaciones) que tenga con otras tablas. En

función a este punto la secuencia de las transformaciones será la siguiente:

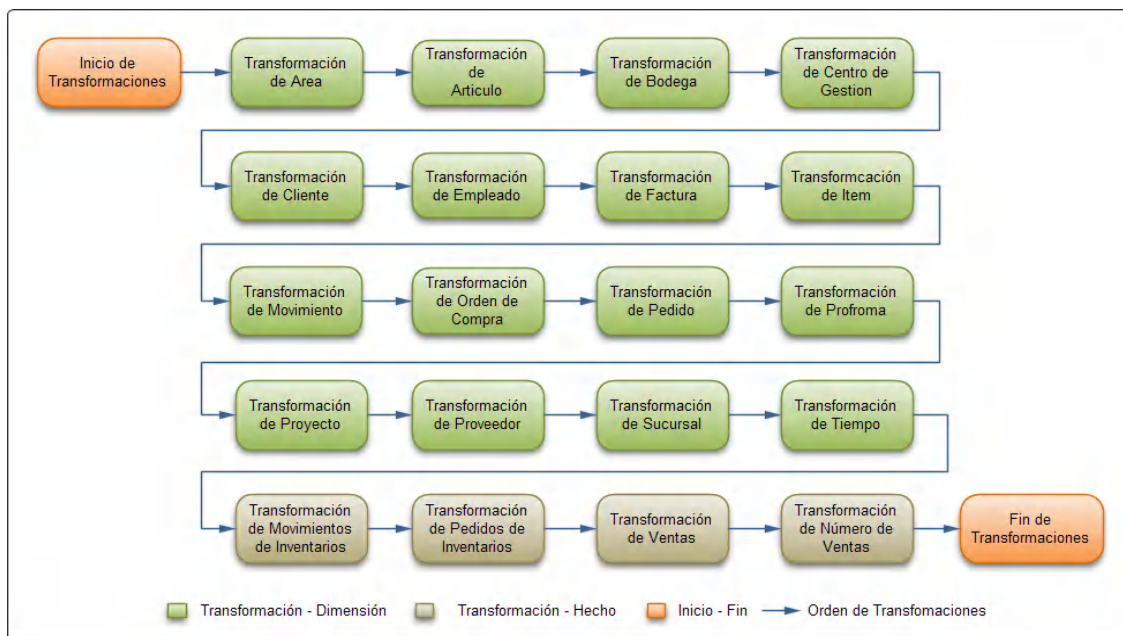


Figura 3.29: Secuencia de las Transformaciones

Primero se realizará las transformaciones para poblar las Dimensiones, a continuación se seguirá con los Hechos ya que como se pudo apreciar en el Modelo Dimensional los Hecho mantienen integridad referencial con las tablas de Dimensiones.

- Registro de Errores.

Como parte del proceso ETL se pretende realizar una Auditoria de errores para poder verificar y estar al tanto sobre errores que lleguen a producirse, y poder identificar el origen del error para tomar las medidas de corrección.

La manera como se pretende realizar el manejo de errores se lo puede interpretar de mejor manera en el siguiente grafico.

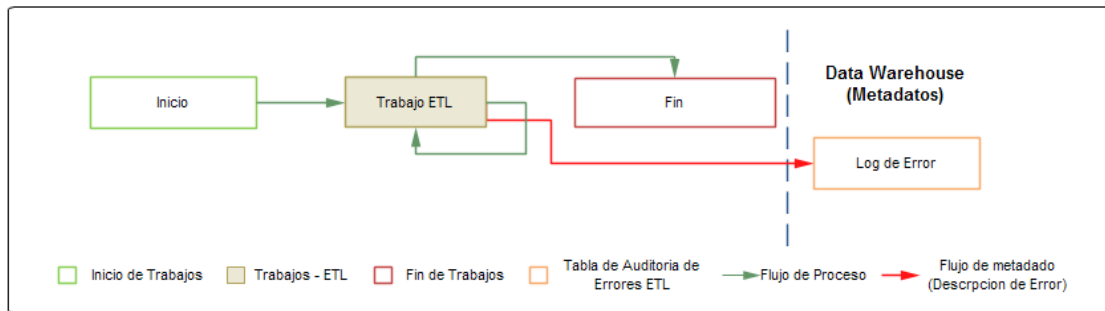
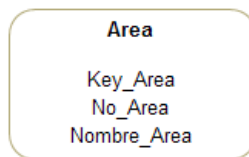


Figura 3.30: Auditoria – Error ETL

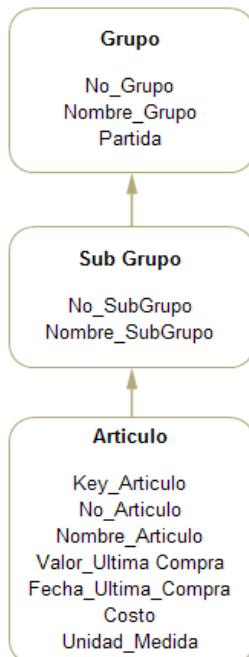
3.8.4. PASO 4: PROFUNDIZAR TABLA DESTINO

A continuación se especificará de manera detallada los niveles y jerarquía por cada dimensión.

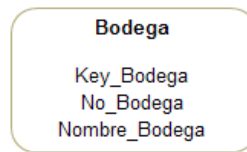
Área



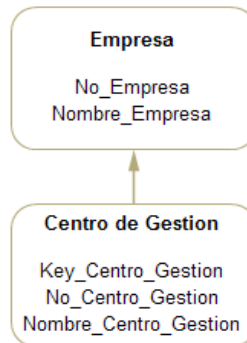
Artículo



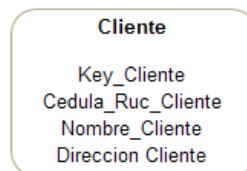
Bodega



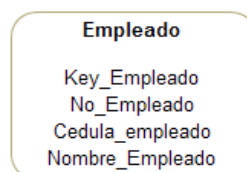
Centro de Gestión



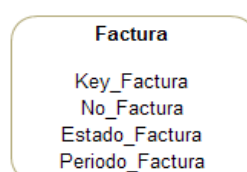
Cliente



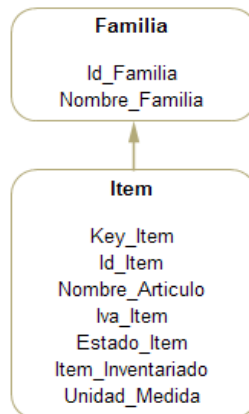
Empleado



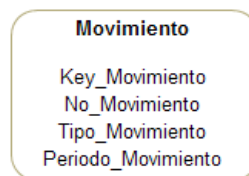
Factura



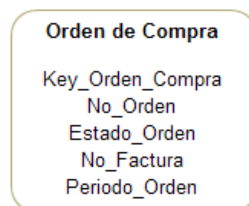
Ítem



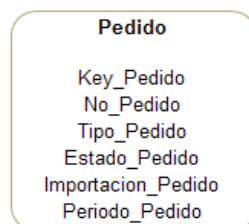
Movimiento



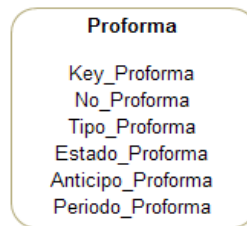
Orden de Compra



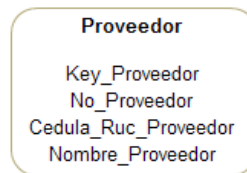
Pedido



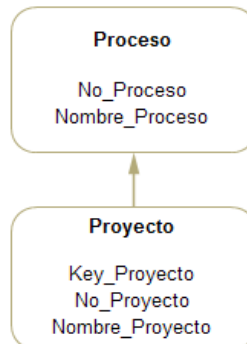
Proforma



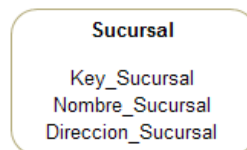
Proveedor



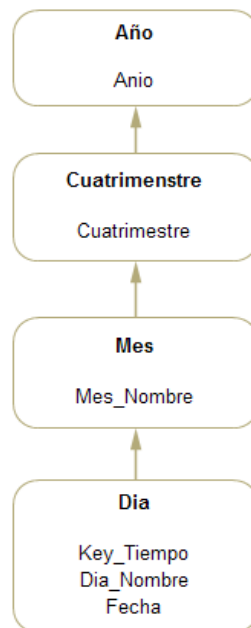
Proyecto



Sucursal



Tiempo



DESARROLLO DE CARGA INICIAL

Aquí se especificará la Carga de Datos que se da por primera instancia hacia el DataWarehouse.

3.8.5. PASO 5: POBLAR LAS TABLAS DE DIMENSIONES CON LOS DATOS HISTÓRICOS

3.8.5.1. Extracción de Datos

La extracción de los datos se lo realizará mediante sentencias SQL a las tablas de la Base Origen. Como resultado devolverá todos los registros de acuerdo a la tabla consultada debido a que es el Proceso de carga Inicial.

Las consultas desarrolladas para poblar las Dimensiones se encuentran en el Anexo G.

3.8.5.2. Transformaciones

Los datos antes de ser cargados en el almacén serán transformados según lo requieran, para lo cual se especifico ciertas transformaciones que se mencionan a continuación de acuerdo a cada Dimensión.

3.8.5.2.1. Valores Nulos

Los valores Nulos serán remplazados por valores que tengan un significado, y los cuales fueron especificados por el personal involucrado del Instituto. Estos valores tomados son de acuerdo a los valores origen que se extraerán.

Área

Tabla 3.4: Valores Nulos – Dimensión Área

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_AREAS	NO_CD	0	NO_AREA
	DESCRIP	NO ASIGNADO	NOMBRE_AREA

Artículo

Tabla 3.5: Valores Nulos – Dimensión Artículo

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_GRUPOS_ARTICULOS	GRUPO	0	NO_GRUPO
	DESC_GRUPO	NO ASIGNADO	NOMBRE_GRUPO
	SUBGRUPO	0	NO_SUBGRUPO
	DESC_SUBGRUPO	NO ASIGNADO	NOMBRE_SUBGRUPO
	C_PARTIDA_PRESUP	0	PARTIDA
T_ARTICULO	NO_ARTICULO	0	NO_ARTICULO
	DESCRIP	NO ASIGNADO	NOMBRE_ARTICULO
	VALOR_UNIT_ULT_COMPRA	---	VALOR_ULTIMA_COMPRA
	F_ULTIMA_COMPRA	---	FECHA_ULTIMA_COMPRA
T_UND_MED	V_COSTO_PROMEDIO	---	COSTO
	DESCRIP	NO ASIGNADO	UNIDAD_MEDIDA

Como se puede identificar en el cuadro, no se asignaron datos para remplazar en caso de que existe valor nulo en los campos origen, de VALOR_UNIT_ULT_COMPRA, F_ULTIMA_COMPRA y V_COSTO_PROMEDIO debido a que por parte de los empleados involucrados

del negocio en la Institución, sugirieron que no se realice cambios a los valores nulos con respecto a estos tres campos.

Se recomienda la transformación para todos los datos, pero en función de la sugerencia, y tomando en cuenta que a la final es una necesidad o requerimiento de usuario, se procedió a no realizar transformación para estos campos.

Bodega

Tabla 3.6: Valores Nulos – Dimensión Bodega

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_BODEGA	NO_BODEGA	0	NO_BODEGA
	DESCRIP	NO ASIGNADO	NOMBRE_BODEGA

Centro de Gestión

Tabla 3.7: Valores Nulos – Dimensión Centro de Gestión

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_CENTRO_GESTION	CODIGO_CENTRO_GESTION	0	NO_CENTRO_GESTION
	DESCRIPCION	NO ASIGNADO	NOMBRE_CENTRO_GESTION
T_EMPRESA	NO_EMPRESA	0	NO_EMPRESA
	DESCRIP	NO ASIGNADO	NOMBRE_EMPRESA

Cliente

Tabla 3.8: Valores Nulos – Dimensión Cliente

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
TENTIDAD	CIU	00000000000000	CEDULA_RUC_CLIENTE
	NOMBRES. APELLIDOS	NO ASIGNADO	NOMBRE_CLIENTE
	DIRECCION	NO ASIGNADO	DIRECCION_CLIENTE

Empleado

Tabla 3.9: Valores Nulos – Dimensión Empleado

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_DAT_EMPLEADO	NO_PERSONA	0	NO_EMPLEADO
	NO_CEDULA	0000000000	CEDULA_EMPLEADO
	NOMBRES, PRI_APELLIDO, SEG_APELLIDO	NO ASIGNADO	NOMBRE_EMPLEADO

Factura

Tabla 3.10: Valores Nulos – Dimensión Factura

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
TFACTURA	ID_FACTURA	0	NO_FACTURA
	ESTADO	NO ASIGNADO	ESTADO_FACTURA
	ID_PERIODO	0	PERIODO_FACTURA

Ítem

Tabla 3.11: Valores Nulos – Dimensión Ítem

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
TFAMILIA	ID_FAMILIA	0	ID_FAMILIA
	NOMBRE_FAMILIA	NO ASIGNADO	NOMBRE_FAMILIA
TITEM	ID_ITEM	0	NO_ITEM
	DESC_ITEM	NO ASIGNADO	NOMBRE_ITEM
	L_IVA	NO	IVA_ITEM
	ESTADO	NO ASIGNADO	ESTADO_ITEM
	L_INVENTARIADO	NO	ITEM_INVENTARIADO
TUNIDADMEDIDA	DESC_UNIDAD	NO ASIGNADO	UNIDAD_MEDIDA

Movimiento

Tabla 3.12: Valores Nulos – Dimensión Movimiento

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_MOV	NO_MOV	0	NO_MOVIMIENTO
	PERIODO_CODIGO	0	PERIODO_MOVIMINETO
T_TIPO_MOV_INV	DESCRIP	NO ASIGNADO	TIPO_MOVIMIENTO

Orden de Compra

Tabla 3.13: Valores Nulos – Dimensión Orden de Compra

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_ORDEN_COMPRA	NO_ORDEN	0	NO_ORDEN
	ESTADO	NO ASIGNADO	ESTADO_ORDEN
	PERIODO_CODIGO	0	PERIODO_ORDEN
T_DET_ORDEN_COMPRA	NO_FACTURA	0	NO_FACTURA

Pedido

Tabla 3.14: Valores Nulos – Dimensión Pedido

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_PEDIDO	NO_PEDIDO	0	NO_PEDIDO
	TIPO_PEDIDO	NO ASIGNADO	TIPO_PEDIDO
	ESTADO	NO ASIGNADO	ESTADO_PEDIDO
	FUE_IMPORTACION	NO	IMPORTACION_PEDIDO
	PERIODO_CODIGO	0	PERIODO_PEDIDO

Proforma

Tabla 3.15: Valores Nulos – Dimensión Proforma

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
TPROFORMA	ID_PROFORMA	0	NO_PROFORMA
	TIPO_PROFORMA	NO ASIGNADO	TIPO_PROFORMA
	ESTADO	NO ASIGNADO	ESTADO_PROFORMA
	ID_PERIODO	0	PERIODO_PROFORMA
TANTICIPO	---	NO	ANTICIPO_PROFORMA

El campo ANTICPO_PROFORMA no tiene un campo origen debido a que fue generado en el Sistema ETL, extrae las proformas con anticipo de la tabla origen TANTICIPO.

Proveedor

Tabla 3.16: Valores Nulos – Dimensión Proveedor

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_PROVEEDOR	NO_PROVEEDOR	0	NO_PROVEEDOR
	NO_CEDULA_RUC	0000000000000	CEDULA_RUC_PROVEEDOR
	NOMBRE	NO ASIGNADO	NOMBRE_PROVEEDOR

Proyecto

Tabla 3.17: Valores Nulos – Dimensión Proyecto

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
T_PRY_PROYECTO	COD_PROYECTO	0	NO_PROYECTO
	NOMBRE_PROYECTO	NO ASIGNADO	NOMBRE_PROYECTO
T_PRY_MACROPROCESO	CODIGO_MACRO	0	NO_PROCESO
	DESCRIPCION	NO ASIGNADO	NOMBRE_PROCESO

Sucursal

Tabla 3.18: Valores Nulos – Dimensión Sucursal

Tabla Origen	Cambo Origen	Valor a Reemplazar por Valor Nulo	Campo Destino
TSUCURSAL	ID_SUCURSAL	0	KEY_SUCURSAL
	DESC_SUCURSAL	NO ASIGNADO	NOMBRE_SUCURSAL
	DIRECCION	NO ASIGNADO	DIRECCION_SUCURSAL

3.8.5.2.2. Agregar registro sin Valor a las Dimensiones.

Se agregará un registro a cada Dimensión, el cual tendrá un valor que permitirá identificar hechos que no tienen especificación con la dimensión, por ejemplo: Una Venta a la cual no se le haya especificado un Empleado (Vendedor), ó un Pedido de Inventario que no se le haya asignado o tenga vinculado por el momento una Bodega.

Estos registros fueron especificados con la ayuda de las personas involucradas en el Negocio, y se detallan a continuación de acuerdo a cada Dimensión:

Área

Tabla 3.19: Registro sin valor – Dimensión Área

Tabla de Dimension	Cambo Dimension	Valor de Campo
DIM_AREA	KEY_AREA	0
	NO_AREA	0
	NOMBRE_AREA	SIN AREA

Artículo

Tabla 3.20: Registro sin valor – Dimensión Artículo

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_ARTICULO	KEY_ARTICULO	0
	NO_GRUPO	0
	NOMBRE_GRUPO	SIN GRUPO
	NO_SUBGRUPO	0
	NOMBRE_SUBGRUPO	SIN SUBGRUPO
	PARTIDA	0
	NO_ARTICULO	0
	NOMBRE_ARTICULO	SIN ARTICULO
	VALOR_ULTIMA_COMPRA	0
	FECHA_ULTIMA_COMPRA	01/01/2004
COSTO	0	

Bodega

Tabla 3.21: Registro sin valor – Dimensión Bodega

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_BODEGA	KEY_BODEGA	0
	NO_BODEGA	0
	NOMBRE_BODEGA	SIN BODEGA

Centro de Gestión

Tabla 3.22: Registro sin valor – Dimensión Centro de Gestión

Tabla de Dimension	Campo Dimension	Valor de campo
DIM_CENTRO_GESTION	KEY_CENTRO_GESTION	0
	NO_CENTRO_GESTION	0
	NOMBRE_CESNTRO_GESTION	SIN CENTRO
	NO_EMPRESA	0
	NOMBRE_EMPRESA	SIN EMPRESA

Cliente

Tabla 3.23: Registro sin valor – Dimensión Cliente

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_CLIENTE	KEY_CLIENTE	0
	CEDULA_RUC_CLIENTE	SIN CEDULA/RUC
	NOMBRE_CLIENTE	SIN CLENTE
	DIRECCION_CLIENTE	SIN DIRECCION

Empleado

Tabla 3.24: Registro sin valor – Dimensión Empleado

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_EMPLEADO	KEY_EMPLEADO	0
	NO_EMPLEADO	0
	CEDULA_EMPLEADO	SIN CEDULA
	NOMBRE_EMPLEADO	SIN EMPLEADO

Factura

Tabla 3.25: Registro sin valor – Dimensión Factura

Tabla de Dimension	Campo Dimension	Valor de Campo
TFACTURA	KEY_FACTURA	0
	NO_FACTURA	0
	ESTADO_FACTURA	SIN ESTADO
	PERIODO_FACTURA	0

Ítem

Tabla 3.26: Registro sin valor – Dimensión Ítem

Tabla de Dimension	Campo Dimension	Valor a Reemplazar por Valor Nulo
DIM_ITEM	KEY_ITEM	0
	ID_FAMILIA	0
	NOMBRE_FAMILIA	SIN FAMILIA
	ID_ITEM	0
	NOMBRE_ITEM	SIN ITEM
	IVA_ITEM	NO
	ESTADO_ITEM	SIN ESTADO
	ITEM_INVENTARIADO	NO
UNIDAD_MEDIDA	SIN MEDIDA	

Movimiento

Tabla 3.27: Registro sin valor – Dimensión Movimiento

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_MOVIMIENTO	KEY_MOVIMIENTO	0
	NO_MOVIMIENTO	0
	PERIODO_MOVIMINETO	0
	TIPO_MOVIMIENTO	SIN TIPO

Orden de Compra

Tabla 3.28: Registro sin valor – Dimensión Orden de Compra

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_ORDEN_COMPRA	KEY_ORDEN_COMPRA	0
	NO_ORDEN	0
	ESTADO_ORDEN	SIN ESTADO
	PERIODO_ORDEN	0
	NO_FACTURA	0

Pedido

Tabla 3.29: Registro sin valor – Dimensión Pedido

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_PEDIDO	KEY_PEDIDO	0
	NO_PEDIDO	0
	TIPO_PEDIDO	SIN TIPO
	ESTADO_PEDIDO	SIN ESTADO
	IMPORTACION_PEDIDO	NO
	PERIODO_PEDIDO	0

Proforma

Tabla 3.30: Registro sin valor – Dimensión Proforma

Tabla de Dimension	Campo Dimension	Valor de campo
DIM_PROFORMA	KEY_PROFORMA	0
	NO_PROFORMA	0
	TIPO_PROFORMA	SIN TIPO
	ESTADO_PROFORMA	SIN ESTADO
	ANTICIPO_PROFORMA	NO
	PERIODO_PROFORMA	0

Proveedor

Tabla 3.31: Registro sin valor – Dimensión Proveedor

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_PROVEEDOR	KEY_PROVEEDOR	0
	NO_PROVEEDOR	0
	CEDULA_RUC_PROVEEDOR	SIN CEDULA/RUC
	NOMBRE_PROVEEDOR	SIN PROVEEDOR

Proyecto

Tabla 3.32: Registro sin valor – Dimensión Proyecto

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_PROYECTO	KEY_PROYECTO	0
	NO_PROYECTO	0
	NOMBRE_PROYECTO	SIN PROYECTO
	NO_PROCESO	0
	NOMBRE_PROCESO	SIN PROCESO

Sucursal

Tabla 3.33: Registro sin valor – Dimensión Sucursal

Tabla de Dimension	Campo Dimension	Valor de Campo
DIM_SUCURSAL	KEY_SUCURSAL	0
	NOMBRE_SUCURSAL	SIN SUCURSAL
	DIRECCION_SUCURSAL	SIN DIRECCION

3.8.5.2.3. Decodificación de datos Origen

Como parte de la transformación esta la decodificación de datos origen que se encuentran almacenados de manera codificada, para soportar los ambientes transaccionales. A continuación se especificarán las Dimensiones para las cuales se decodificaron los datos de Origen.

Factura

Tabla 3.34: Decodificación Factura

Cambo Origen	Valor Codificado - Origen	Campo Destino	Valor Decodificado - Destino
ESTADO	P	ESTADO_FACTURA	PENDIENTE
	C		CANCELADA
	A		ANULADA

Ítem

Tabla 3.35: Decodificación Ítem

Cambo Origen	Valor codificado - Origen	Campo Destino	Valor decodificado - Destino
L_IVA	0	IVA_ITEM	NO
	1		SI
ESTADO	V	ESTADO_ITEM	VIJENTE
	A		ANULADO
L_INVENTARIADO	0	ITEM_INVENTARIADO	NO
	1		SI

Orden de Compra

Tabla 3.36: Decodificación Orden de Compra

Cambo Origen	Valor codificado - Origen	Campo Destino	Valor decodificado - Destino
ESTADO	A	ESTADO_ORDEN	ANULAR
	T		TERMINAR
	C		COMPRAR
	P		PROFORMA

Pedido

Tabla 3.37: Decodificación Pedido

Cambo Origen	Valor codificado - Origen	Campo Destino	Valor decodificado -Destino
TIPO_PEDIDO	R	TIPO_PEDIDO	REQUISICION
	P		PEDIDO
	T		TRANSFERENCIA
	F		FONDO RENDIR CUENTAS
	S		STOCK
ESTADO	V	ESTADO_PEDIDO	VIJENTE
	T		TRAMITAR
	J		JUSTIFICAR
	A		ANULAR
FUE_IMPORTACION	-1	IMPORTACION_PEDIDO	SI
	0		NO

Proforma

Tabla 3.38: Decodificación Proforma

Cambo Origen	Valor codificado - Origen	Campo Destino	Valor decodificado -Destino
TIPO_PROFORMA	OP	TIPO_PROFORMA	ORDEN DE PRODUCCION
	OI		ORDEN INTERNA
	OV		ORDEN INVENTARIADA
	OS		ORDEN DE SERVICIO
ESTADO	P	ESTADO_PROFORMA	PENDIENTE
	A		ANULADO
	F		EJECUTADO
	E		FACTURADO

3.8.5.3. Carga de Datos

El cargado de los datos se lo realizara utilizando el SCD de tipo 1 ya que para el análisis de datos se necesita los datos actualizados según los requerimientos anteriormente especificados.

3.8.5.4. Poblar Dimensión Estática

La Dimensión estática es considera la de Tiempo ya que en su carga o poblado de datos no se realizará ningún tipo de Actualización o Modificación. El contenido de la tabla se generó en función de la decodificación de la fecha en forma corto: *DD/MM/YYYY*, lo que permitió obtener nombres de días, mes, entre otros, como se muestra a continuación.

Tabla 3.39: Decodificación y Formato Dimensión Tiempo

Formato de Fecha Seleccionado	Valor de Dato Generado	Campo Destino	Valor de Dato Decodificado - Destino
DD/MM/YYYY	01/01/2004 ... 31/12/2014	KEY_TIEMPO	0, 1, 2, 3, ... n
DD/MM/YYYY	01/01/2004 ... 31/12/2014	FECHA	01/01/2004 ... 31/12/2014
YYYY	2004 - 2014	ANIO	2004 - 2014
MM	1, 2, 3, ..., 12	CUATRIMESTRE	Q1, Q2, Q3
MM	1, 2, 3, ..., 12	MES_NOMBRE	ENERO, FEBRERO, ..., DICIEMBRE
DD/MM/YYYY	01/01/2004 ... 31/12/2014	DIA_NOMBRE	LUNES - DOMINGO

Vale recalcar que para esta Dimensión no se cuenta con un campo origen ya que no es extraído de alguna fuente de datos, por lo que es generado desde la herramienta ETL.

3.8.5.5. Desarrollo ETL Inicial de Dimensiones con la Herramienta de Pentaho

A continuación se mostrará mediante gráficos el desarrollo del proceso ETL de acuerdo a cada Dimensión.

Dimensión Área

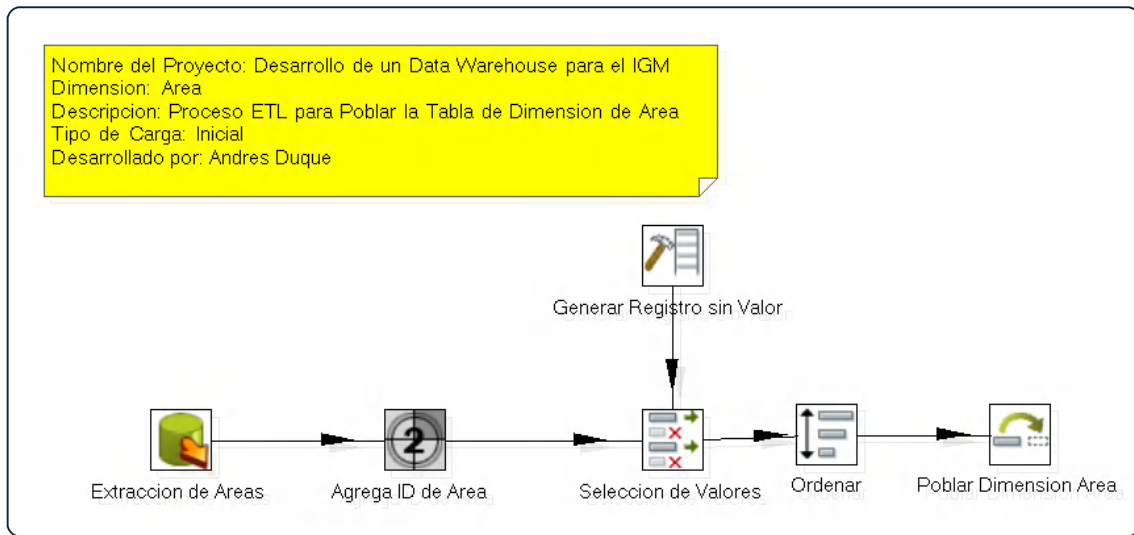


Figura 3.31: Transformación Dimensión Área

Dimensión Artículo

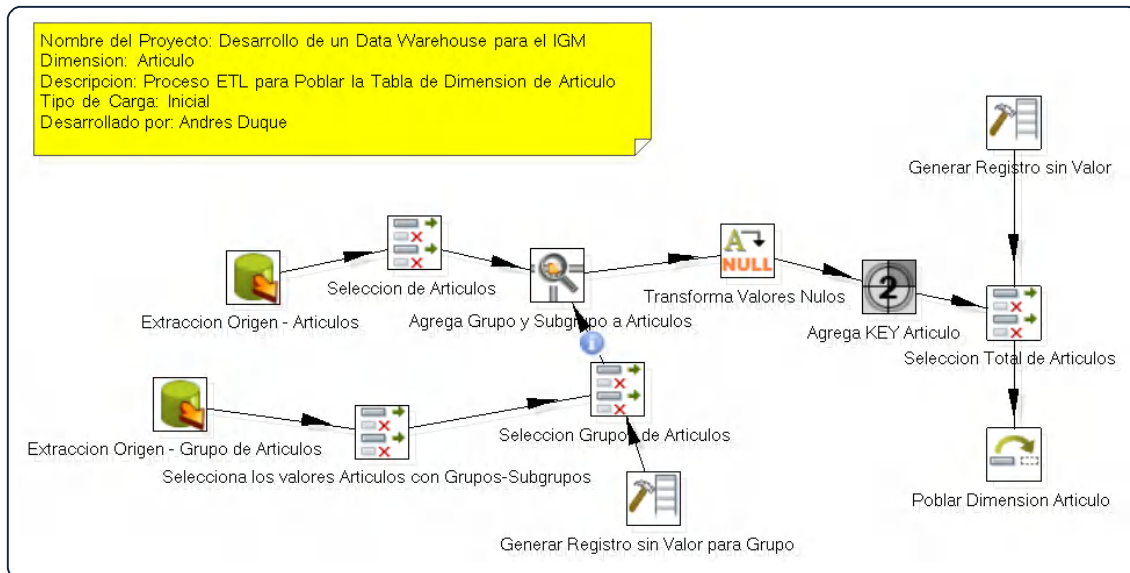


Figura 3.32: Transformación Dimensión Artículo

Dimensión Bodega

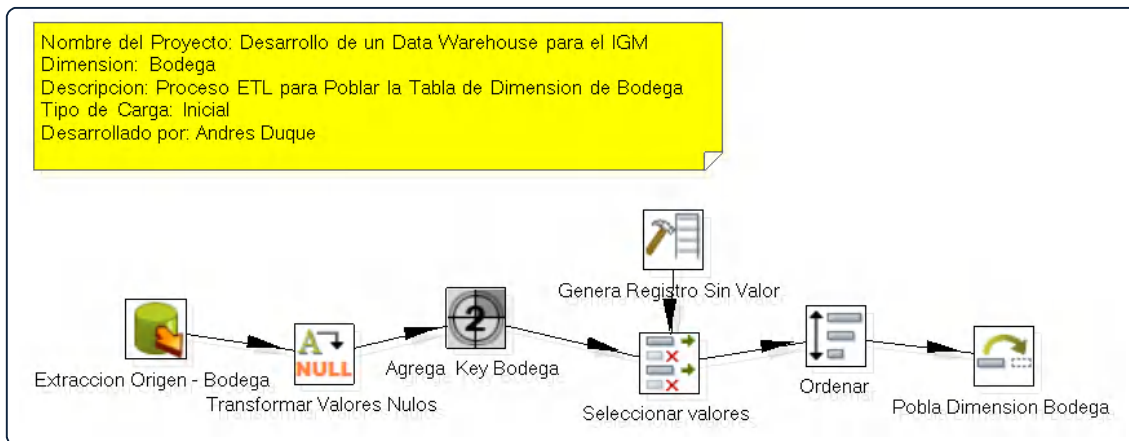


Figura 3.33: Transformación Dimensión Bodega

Dimensión Centro de Gestión

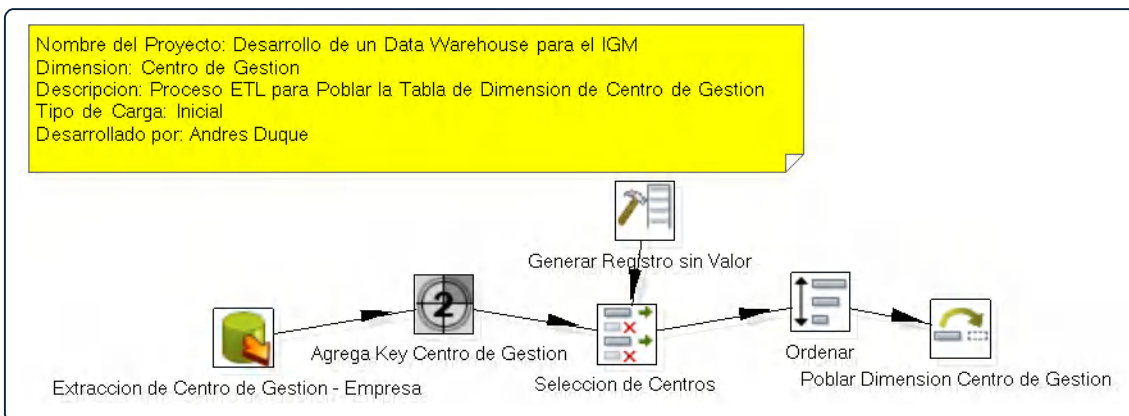


Figura 3.34: Transformación Dimensión Centro de Gestión

Dimensión Cliente

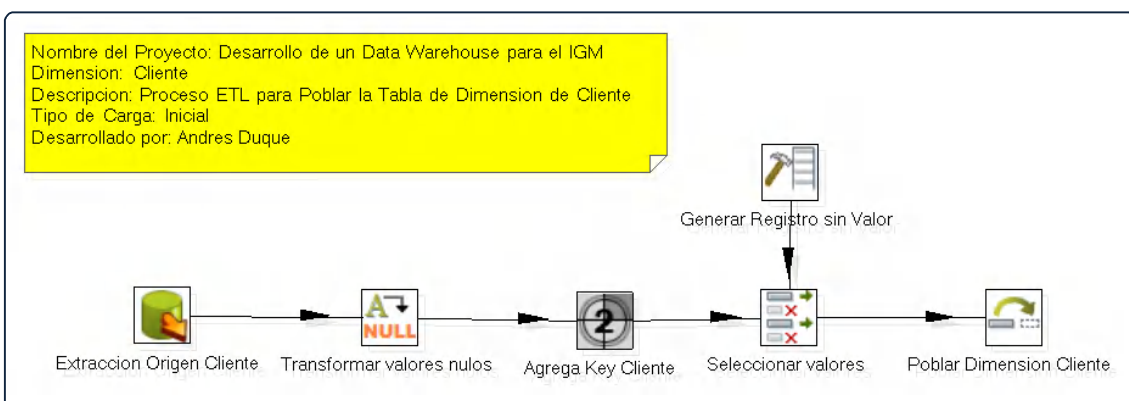


Figura 3.35: Transformación Dimensión Cliente

Dimensión Empleado

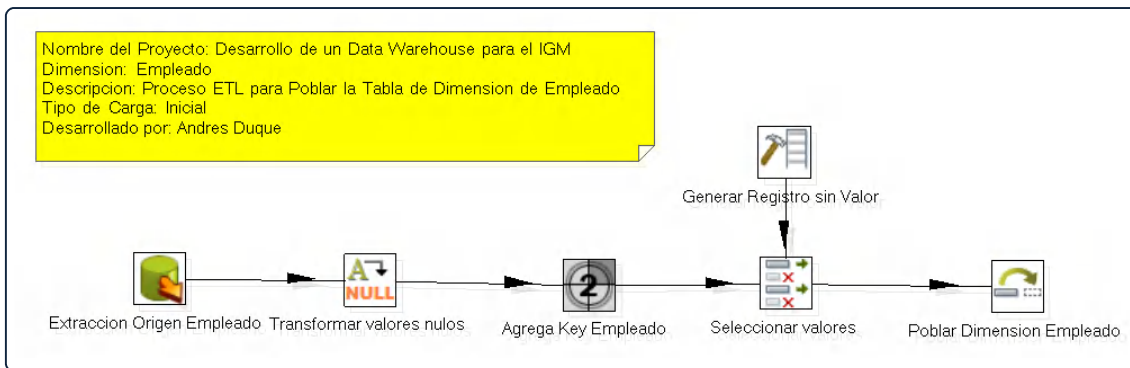


Figura 3.36: Transformación Dimensión Empleado

Dimensión Factura

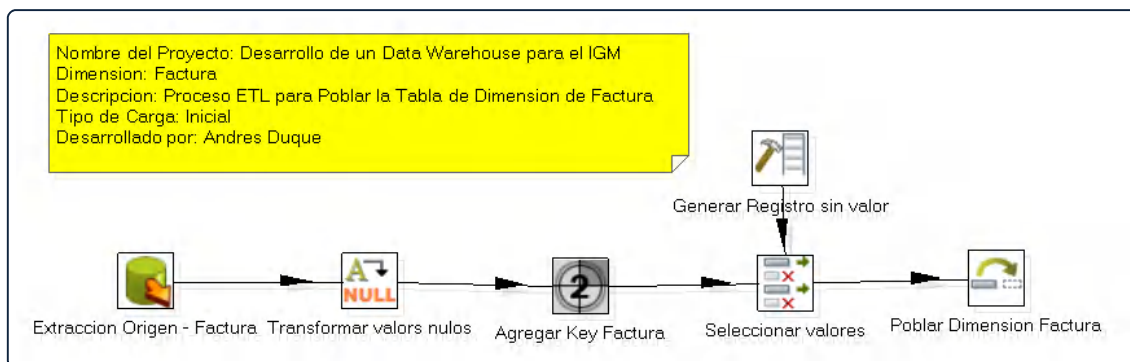


Figura 3.37: Transformación Dimensión Factura

Dimensiona Ítem

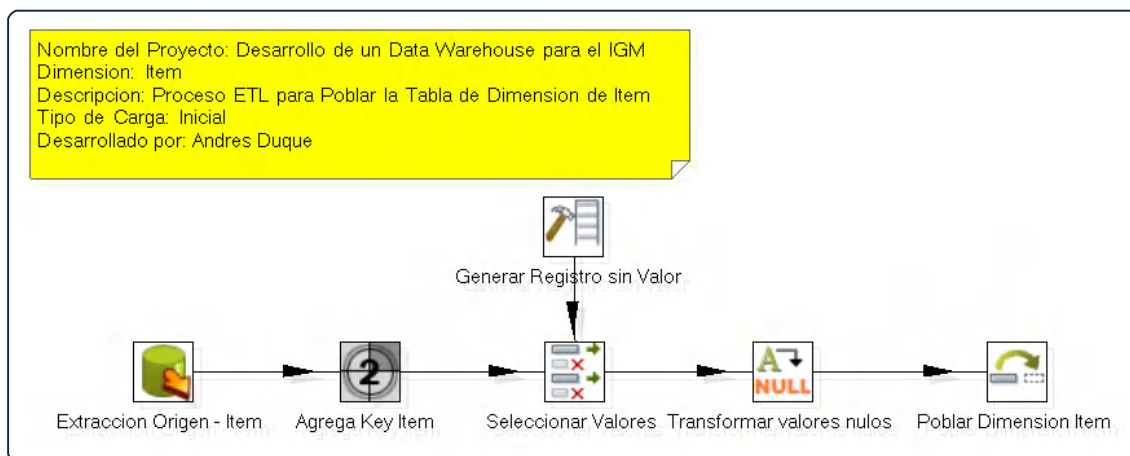


Figura 3.38: Transformación Dimensión Ítem

Dimensión Movimiento

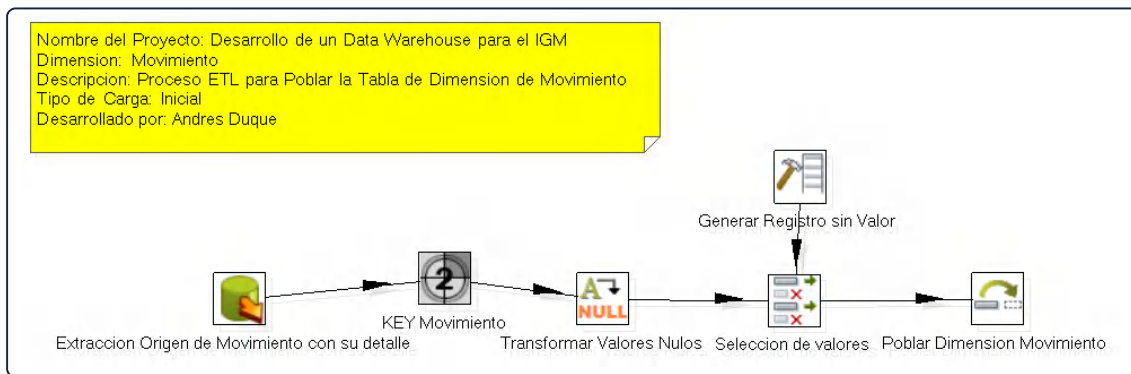


Figura 3.39: Transformación Dimensión Movimiento

Dimensión Orden de Compra

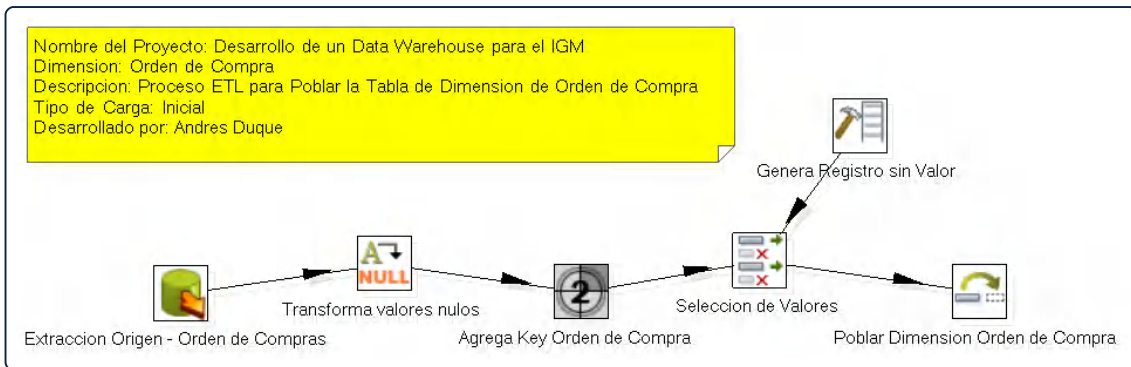


Figura 3.40: Transformación Dimensión Orden de Compra

Dimensión Pedido

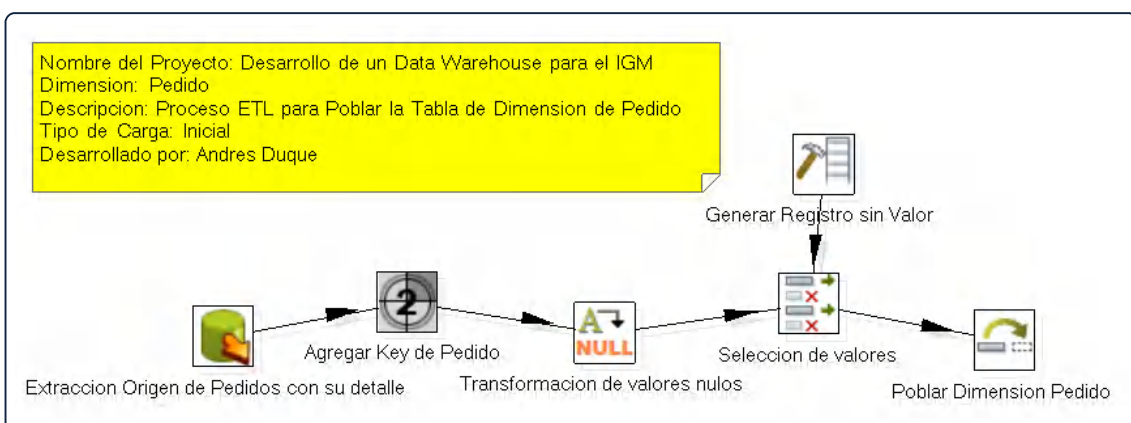


Figura 3.41: Transformación Dimensión Pedido

Dimensión Proforma

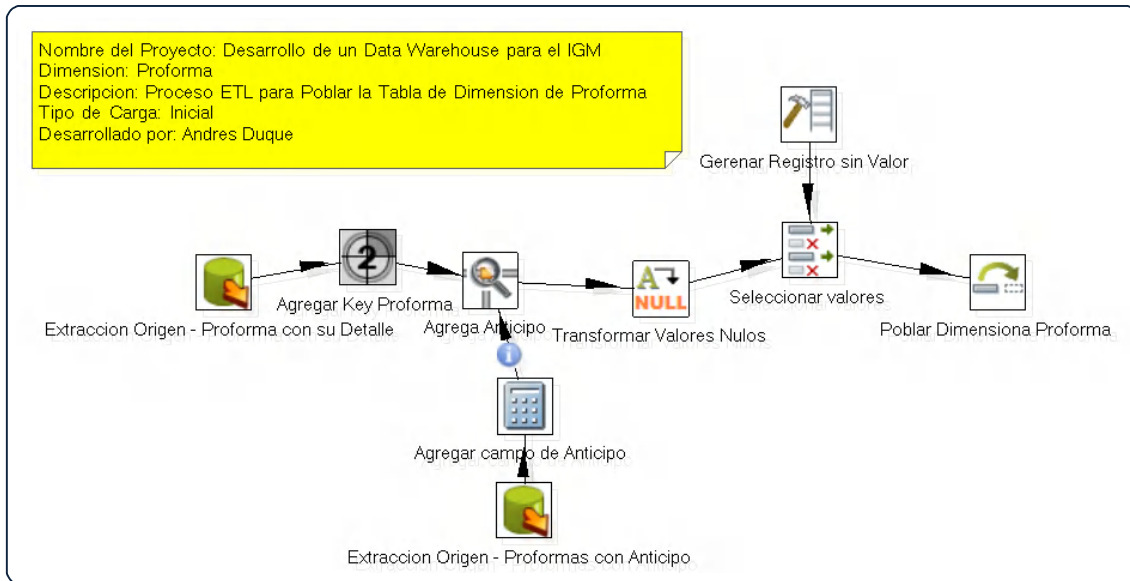


Figura 3.42: Transformación Dimensión Proforma

Dimensión Proveedor

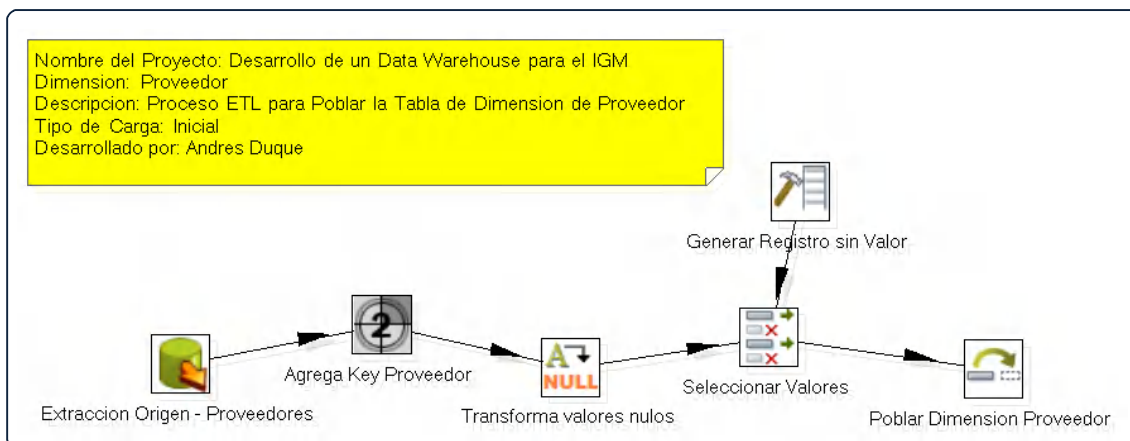


Figura 3.43: Transformación Dimensión Proveedor

Dimensión Proyecto

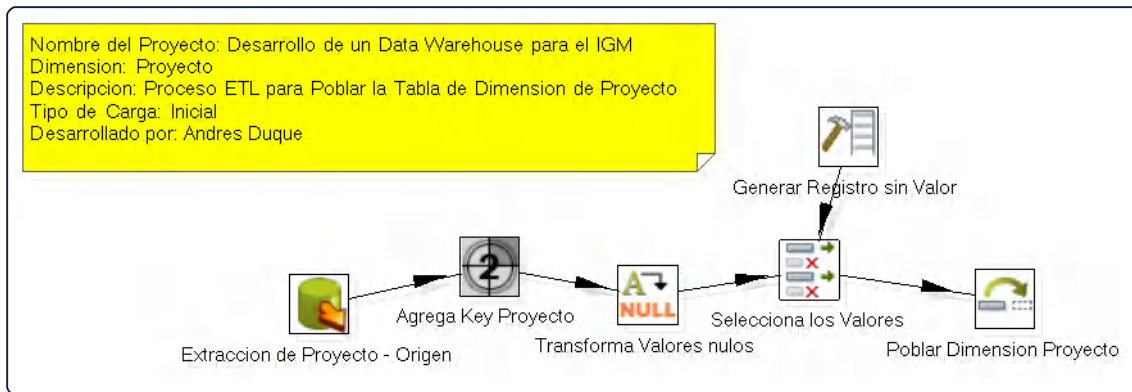


Figura 3.44: Transformación Dimensión Proyecto

Dimensión Sucursal

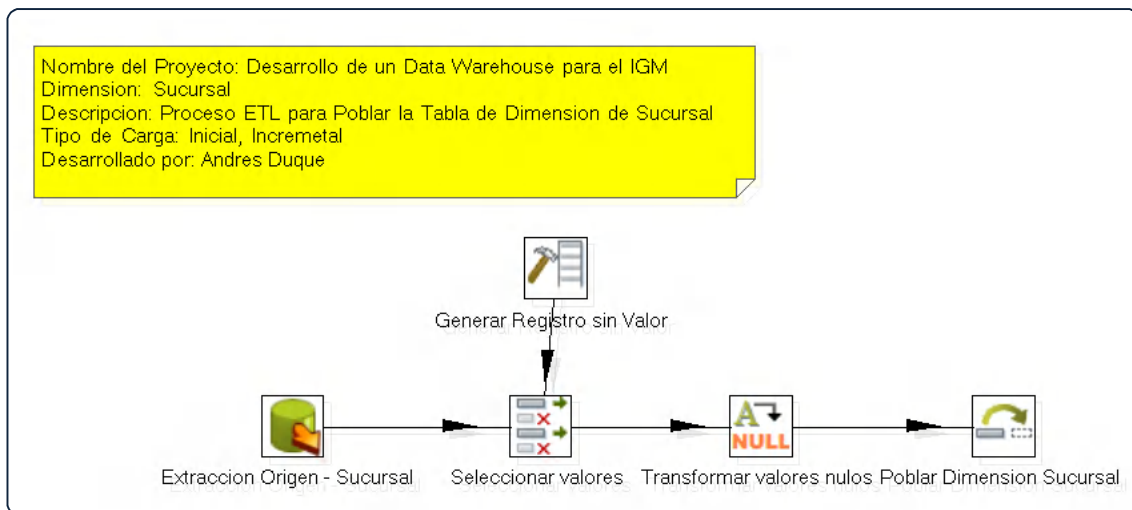


Figura 3.45: Transformación Dimensión Sucursal

Dimensión Tiempo

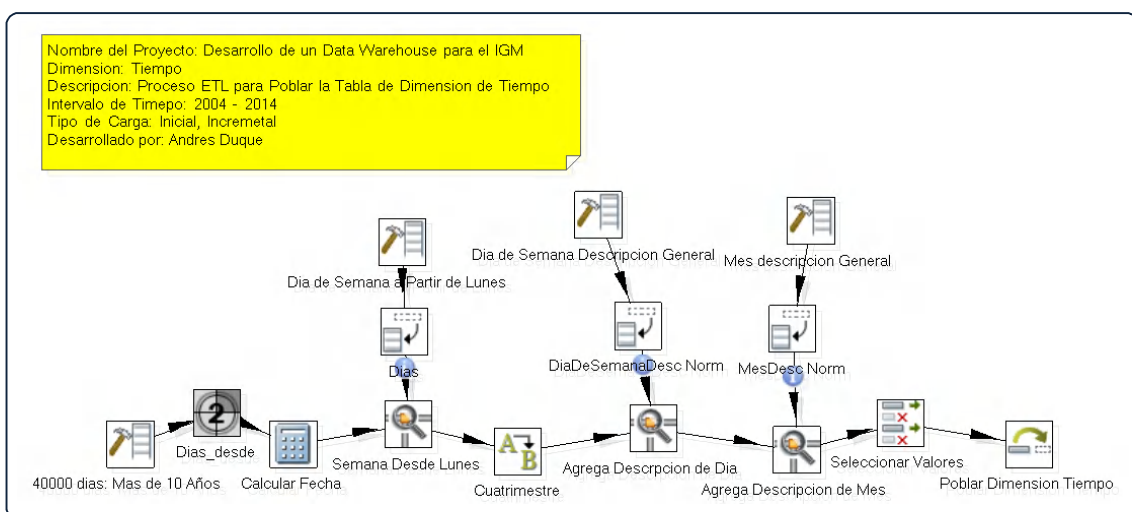


Figura 3.46: Transformación Dimensión Tiempo

3.8.6. PASO 6: REALIZAR LA CARGA HISTÓRICA DE LA TABLA DE HECHOS

La carga de Datos en las tablas de Hechos al igual que en las Dimensiones se lo realizará utilizando el SCD de tipo 1 en función de los requerimientos.

3.8.6.1. Extracción de Datos

Las consultas SQL desarrolladas, necesarias para poblar los hechos se especifican en el Anexo H. Se toman en cuenta tanto la extracción del Hecho desde el Sistema de base de datos origen, así como también el de las claves subrogas de las Dimensiones para asignar a los hechos.

3.8.6.2. Transformaciones

Los datos antes de ser cargados en el almacén serán transformados según lo requieran, para lo cual se especifican transformaciones que se mencionan a continuación de acuerdo a cada Hecho.

3.8.6.2.1. Valores Nulos

Los valores nulos para un hecho son tratados en función y tomando en cuenta los registros sin valor que se les dio a las Dimensiones, por lo que para un registro nulo en la tabla de Hecho, existirá un registro en su respectiva Dimensión, en la cual existe un registro con valor especificado, en caso de no existir entidad o Dimensión para ese Hecho.

Las Métricas no necesitan ser transformadas ya que por defecto desde la fuente de origen, los campos asociados para el cálculo son de característica no nula.

Pedidos

Tabla 3.40: Valores Nulos – Hecho Pedidos

Tabla Origen	Campo Origen	Valor a Reemplazar por Valor Nulo
T_PEDIDO	NO_PEDIDO	0
	PROYECTO_CODIGO	0
	PERIODO_CODIGO	0
	NO_CD	0
	NO_BODEGA	0
T_DET_PEDIDOS	NO_ARTICULO	0
	NO_REG_PEDIDO	0
T_DET_ORDEN_COMPRA	NO_ORDEN	0
	NO_PROVEEDOR	0

Movimientos

Tabla 3.41: Valores Nulos – Hecho Movimientos

Tabla Origen	Campo Origen	Valor a Reemplazar por Valor Nulo
T_MOV	NO_MOV	0
	NO_TIPO_MOV	0
	PERIODO_CODIGO	0
T_DET_MOV	NO_ORDINAL	0
	NO_PEDIDO	0
	NO_REG_PEDIDO	0
	NO_ARTICULO	0
	NO_BODEGA	0
T_DET_PEDIDOS	NO_CD	0
	PROYECTO_CODIGO	0
T_DET_ORDEN_COMPRA	NO_ORDEN	0
	NO_PROVEEDOR	0

Ventas

Tabla 3.42: Valores Nulos – Hecho Ventas

Tabla Origen	Campo Origen	Valor a Reemplazar por Valor Nulo
TPROFORMA	ID_PROFORMA	0
	CIU	0
	ID_CREADOR	0
TDETPROFORMA	NO_REG	0
	ID_PERIODO	0
	ID_SUCURSAL	0
	ID_ITEM	0
	ID_PROYECTO	0
	ID_CENTRO_GESTION	0
	ID_FACTURA	0

Número de Ventas

Tabla 3.43: Valores Nulos – Hecho Número de Ventas

Tabla Origen	Campo Origen	Valor a Reemplazar por Valor Nulo
TPROFORMA	CIU	0
	ID_CREADOR	0
TDETPROFORMA	ID_SUCURSAL	0
	ID_FACTURA	0

3.8.6.2.2. Búsqueda de Claves Subrogadas de Dimensiones

Como parte de las Transformaciones para poblar los hechos, un proceso fundamental es el reemplazo de las claves primarias de las Tablas Origen por las claves Subrogadas Primarias de Tablas Destino (Dimensiones), esto permitirá mantener la integridad referencial entre las tablas en el DataWarehouse, como se puede apreciar en el Plan de Alto Nivel que se realizó previamente.

Para poder apreciar de mejor manera se presenta un cuadro a continuación especificando la clave primaria Origen a reemplazar por la clave subrogada de la Dimensión.

Pedidos

Tabla 3.44: Búsqueda clave subrogada – Hecho Pedidos

Campos Origen	Campos Claves Origen	Campos Claves Subrogadas
NO_PEDIDO, PROYECTO_CODIGO, PERIODO_CODIGO, NO_CD, NO_BODEGA, NO_ARTICULO, NO_ORDEN, NO_PROVEEDOR, FECHA_PEDIDO	FECHA_PEDIDO	KEY_TIEMPO
	NO_ARTICULO	KEY_ARTICULO
	NO_BODEGA	KEY_BODEGA
	NO_CD	KEY_AREA
	NO_ORDEN	KEY_ORDEN_COMPRA
	PERIODO_CODIGO	
	NO_PROVEEDOR	KEY_PROVEEDOR
	NO_PEDIDO	KEY_PEDIDO
	PERIODO_CODIGO	
	PROYECTO_CODIGO	KEY_PROYECTO

Movimientos

Tabla 3.45: Búsqueda clave subrogada – Hecho Movimientos

Campos Origen	Campos Claves Origen	Campos Claves Subrogadas
NO_MOV, PERIODO_CODIGO, NO_ORDINAL, NO_PEDIDO, NO_REG_PEDIDO, NO_ARTICULO, NO_BODEGA, NO_CD, PROYECTO_CODIGO, NO_ORDEN, NO_PROVEEDOR, FECHA	FECHA	KEY_TIEMPO
	NO_ARTICULO	KEY_ARTICULO
	NO_BODEGA	KEY_BODEGA
	NO_CD	KEY_AREA
	NO_ORDEN	KEY_ORDEN
	PERIODO_CODIGO	
	PROYECTO_CODIGO	KEY_PROYECTO
	NO_PROVEEDOR	KEY_PROVEEDOR
	NO_PEDIDO	KEY_PEDIDO
	PERIODO_CODIGO	
	NO_MOV	KEY_MOVIMIENTO
	PERIODO_CODIGO	

Ventas

Tabla 3.46: Búsqueda clave subrogada – Hecho Ventas

Campos Origen	Campos Claves Origen	Campos Claves Subrogadas
FECHA, CIU, ID_SUCURSAL, ID_CENTRO_GESTION, ID_FACTURA, ID_PERIODO, ID_ITEM, ID_PROYECTO, ID_PROFORMA, NO_REG, ID_PERIODO, ID_CREADOR	FECHA	KEY_TIEMPO
	CIU	KEY_CLIENTE
	ID_SUCURSAL	KEY_SUCURSAL
	ID_CENTRO_GESTION	KEY_CENTRO_GESTION
	ID_FACTURA	KEY_FACTURA
	ID_PERIODO	
	ID_ITEM	KEY_ITEM
	ID_PROYECTO	KEY_PROYECTO
	ID_PROFORMA	KEY_PROFORMA
	ID_PERIODO	

Número de Ventas

Tabla 3.47: Búsqueda clave subrogada – Número de Ventas

Campos Origen	Campos Calaves Origen	Campos Calves Subrogadas
CIU, ID_CREADOR, ID_SUCURSAL, FECHA, ID_FACTURA, PERIODO_CODIGO	FECHA	KEY_TIEMPO
	CIU	KEY_CLIENTE
	ID_SUCURSAL	KEY_SUCURSAL
	ID_FACTURA	KEY_FACTURA
	ID_PERIODO	
	ID_CREADOR	KEY_EMPLEADO

3.8.6.3. Carga de Datos

La carga de los datos se lo realizara utilizando el SCD de tipo 1, de acuerdo a los requerimientos especificados, y al igual que la carga de las Dimensión, los datos serán actualizados.

3.8.6.4. Desarrollo ETL Inicial de Hechos con la Herramienta de Pentaho.

A continuación se mostrará mediante gráficos el desarrollo del proceso ETL de acuerdo a cada Hecho.

Hecho Pedidos

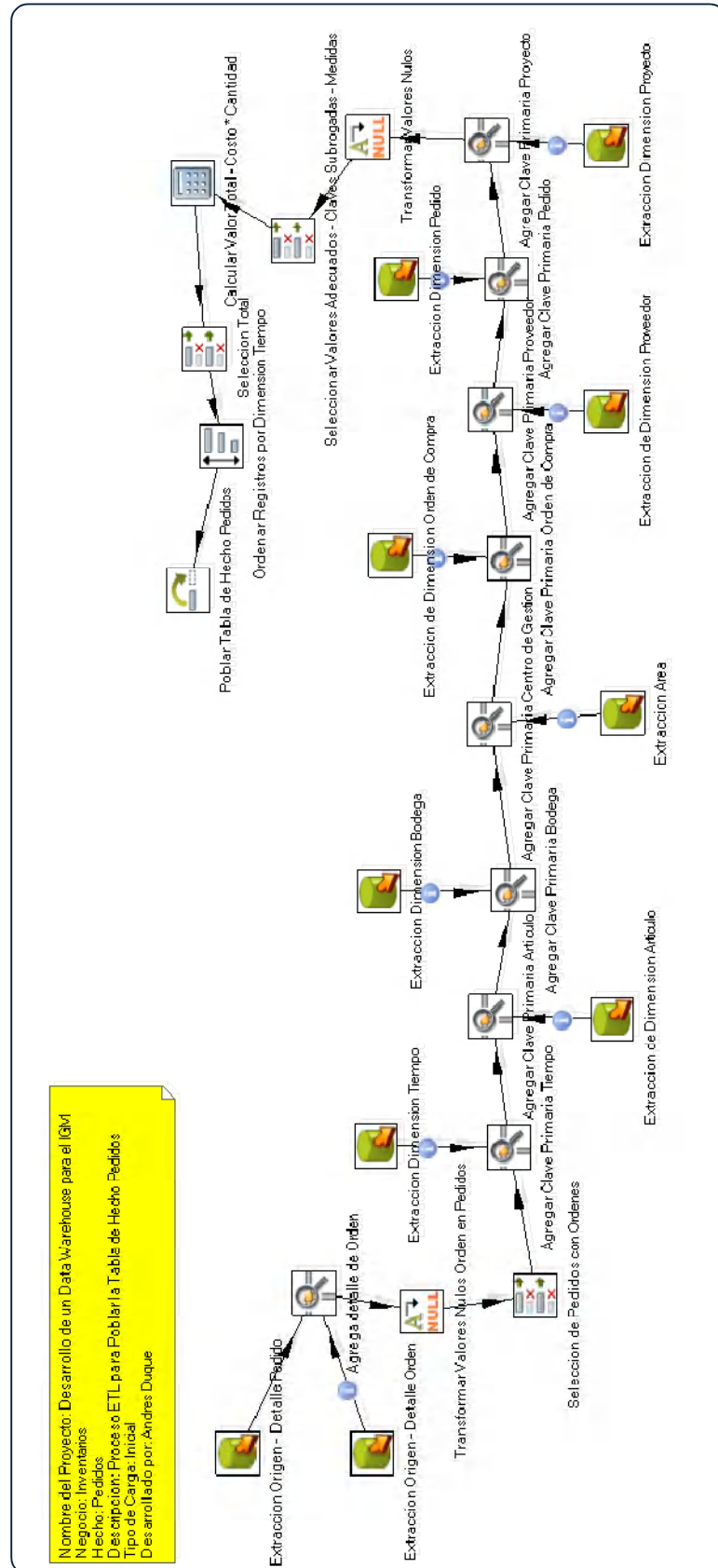


Figura 3.47: Transformación Hecho Pedidos

Hecho Movimientos

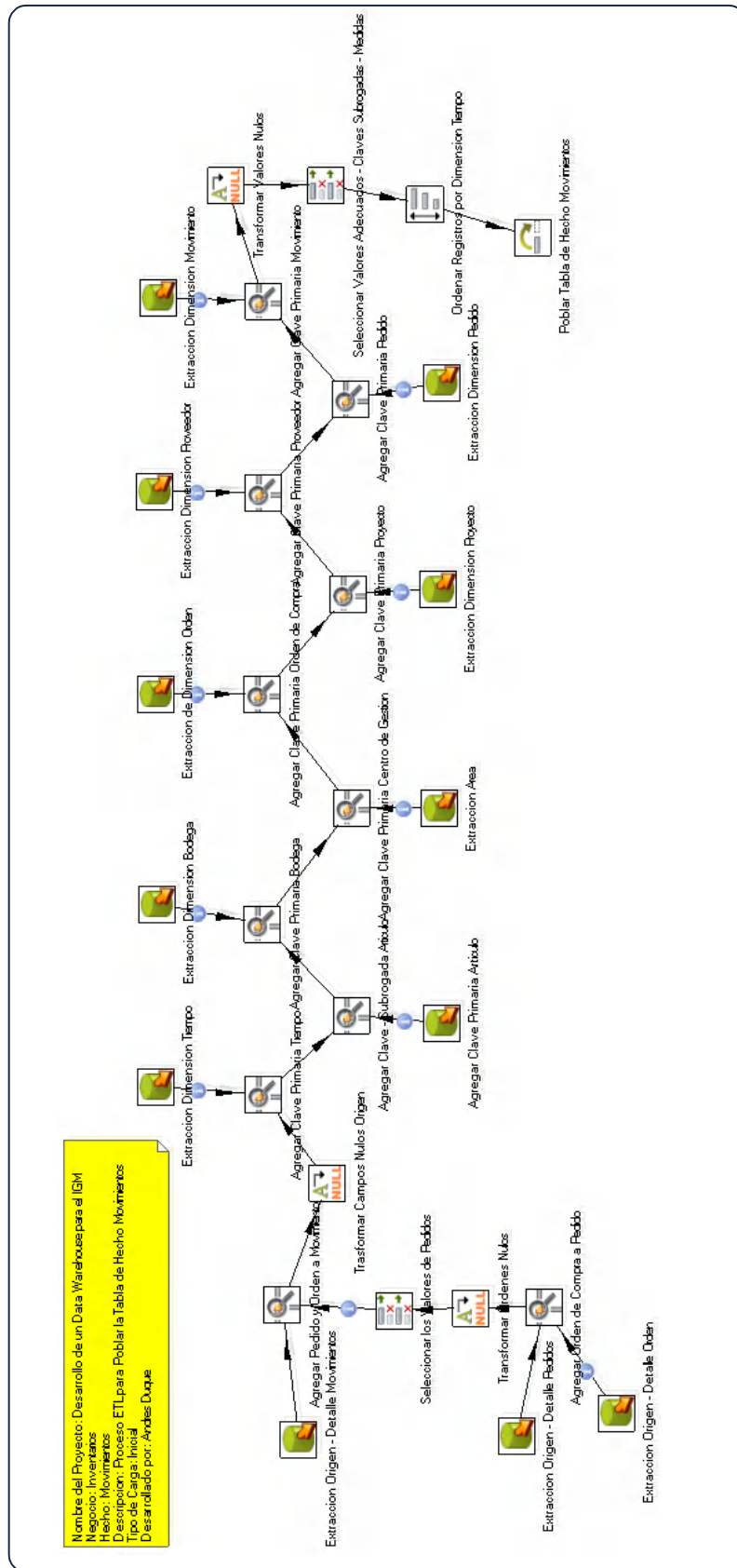


Figura 3.48: Transformación Hecho Movimientos

Hecho Ventas

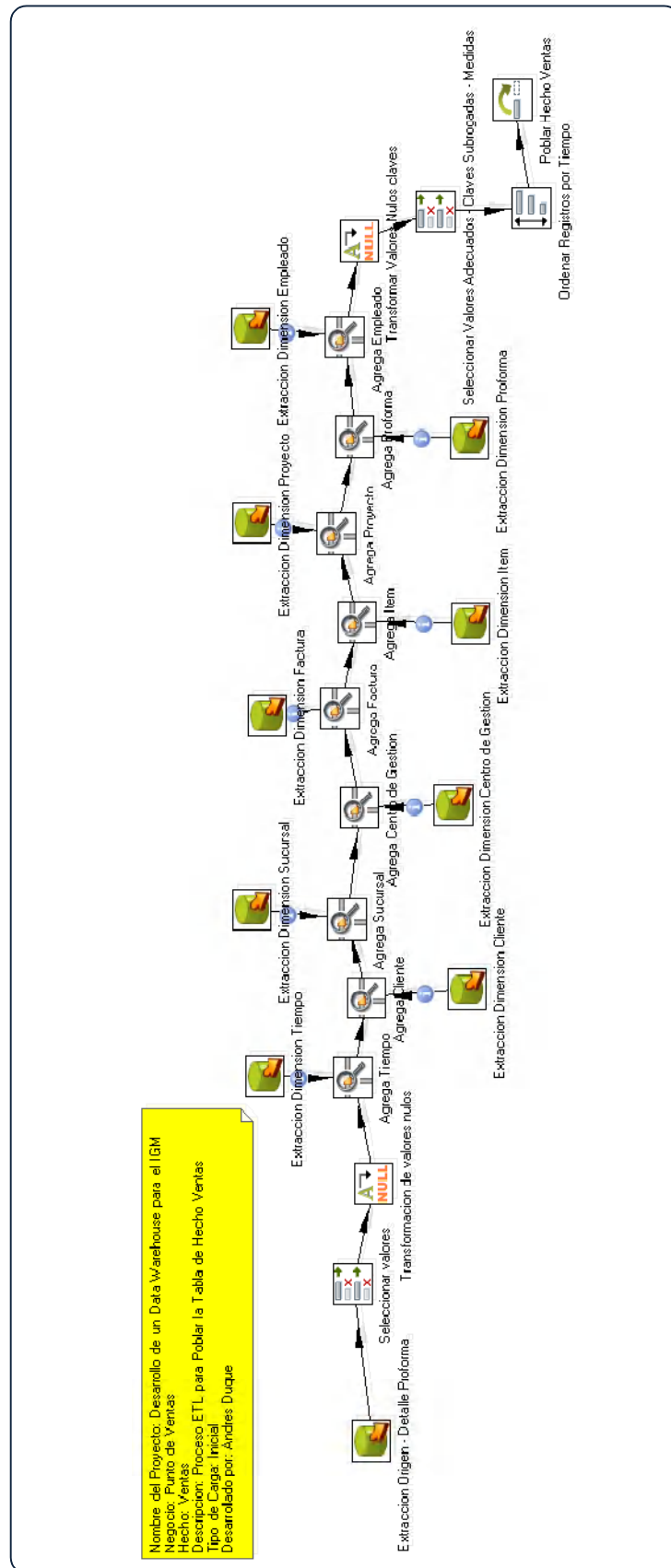


Figura 3.49: Transformación Hecho Ventas

Hecho Numero de Ventas

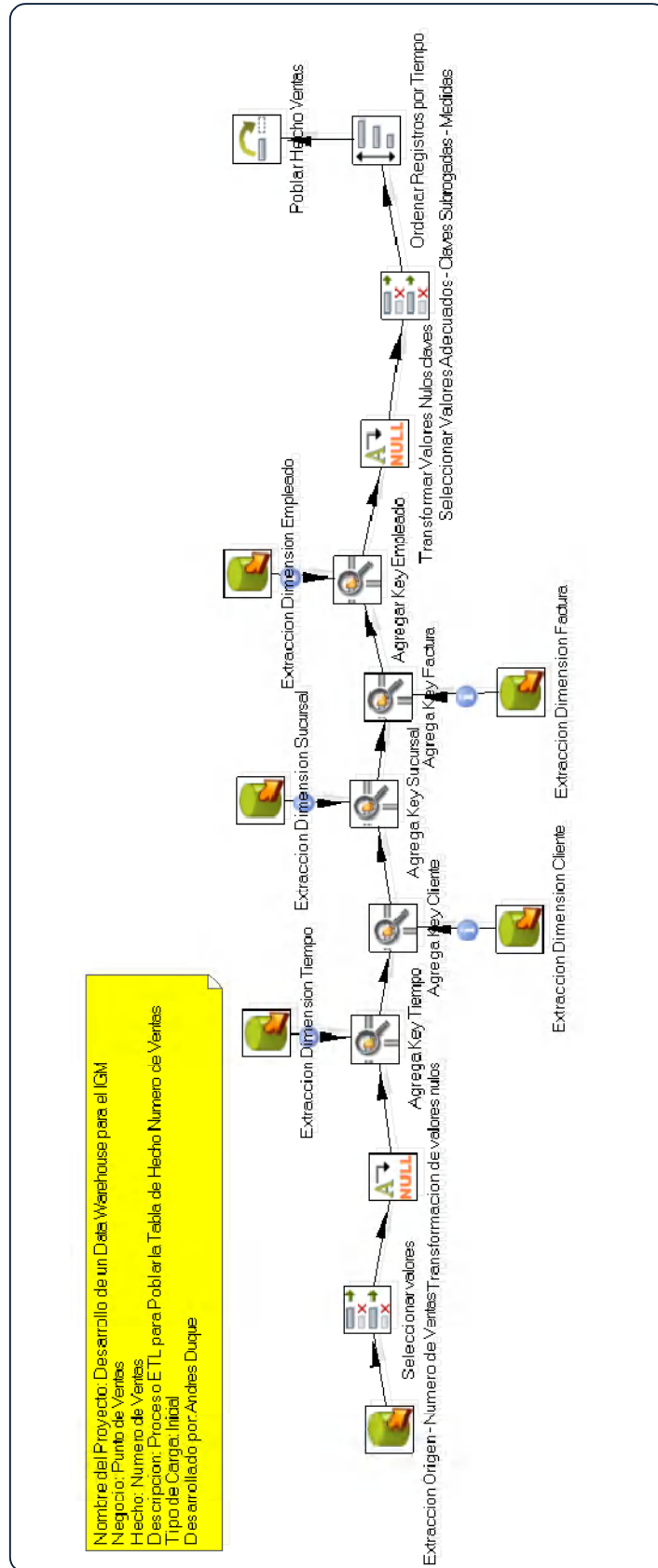


Figura 3.50: Transformación Hecho Número de Ventas

DESARROLLO DE CARGA INCREMENTAL.

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

Como parte de desarrollo del proceso de Carga Incremental es fundamental tomar en cuenta los datos que son nuevos o que han cambiado en la fuente de Origen. Esto permitirá que la extracción de ellos se realice únicamente sobre filas nuevas o sobre aquellas que hayan sufrido modificaciones.

Para poder monitorear estos campos en las fuentes Origen se implementó una pequeña arquitectura diseñada y que se menciona anteriormente en el Plan ETL la sección de *Desarrollar estrategias por defecto* acerca de CDC (Change Data Capture). La intención específica de esta Arquitectura es de capturar los datos que sean nuevos o q hayan sido modificados en la fuente o Base de Origen.

3.8.7. PASO 7: PROCESAMIENTO INCREMENTAL DE LA DIMENSIONES

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

3.8.7.1. Identificar Filas nuevas y con cambios.

Para implementar esta Arquitectura se realizó la extracción de las filas de la Base Origen, cuyas claves primarias se encuentran registradas en la tabla de Auditoria CDC. Para una mayor comprensión se muestran un gráfico a continuación que especifica cómo se realizará la identificación de cambios para las Dimensiones.

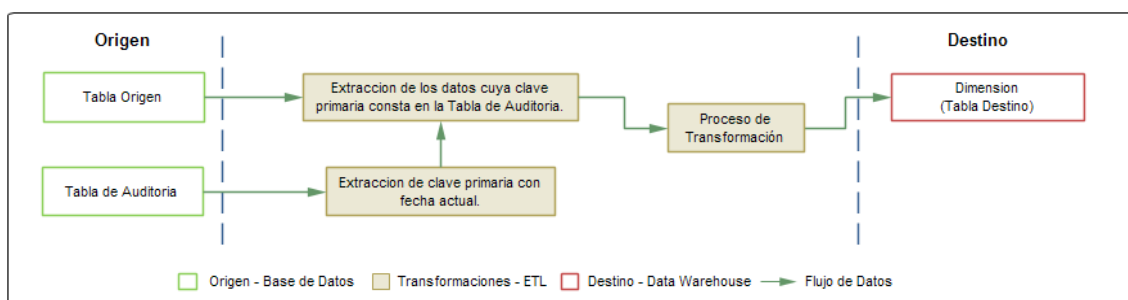


Figura 3.51: Funcionamiento CDC - Dimensiones

3.8.7.2. Extracción de Claves Subrogadas de Dimensiones

Este punto se agrego como parte de la metodología debido a la importancia técnica sobre la función que tendrá en el proceso incremental en las Dimensiones.

Como parte de la transformación se especifica la asignación de claves subrogadas para cada nivel en la dimensión. En función a esto y tomando en cuenta que para el proceso incremental solo se extraerán los datos nuevos o modificados, se diseño la manera de asignar las claves identificando el valor máximo de la clave subrogada del nivel actual que se encuentre en la Dimensión. Esto permitirá que en caso de ser datos nuevos, se designe la clave subrogada adecuada para la carga de estos datos en las Dimensiones evitando duplicidades de las claves o errores en la carga, y en caso de datos

modificados, no repercutirá ya que se actualizarán los registros de la Dimensión.

Para una mejor interpretación de lo dicho se especificará un grafico a continuación, en la cual se ilustra su funcionamiento.

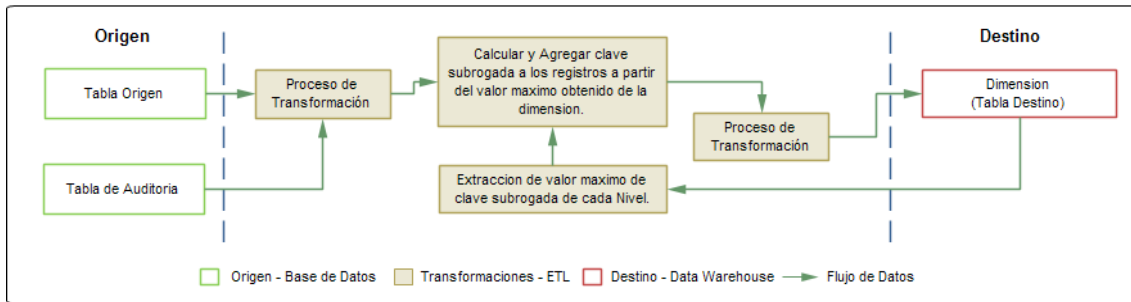


Figura 3.52: Agregar Clave Subrogada

3.8.7.3. Desarrollo ETL Incremental de Dimensiones con la Herramienta de Pentaho.

Dimensión Área

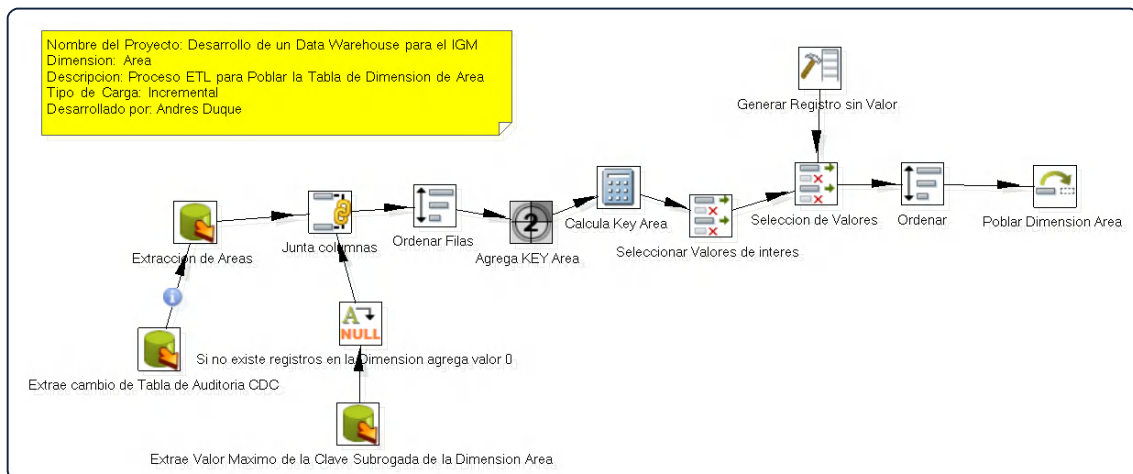


Figura 3.53: Transformación Dimensión Área

Dimensión Artículo

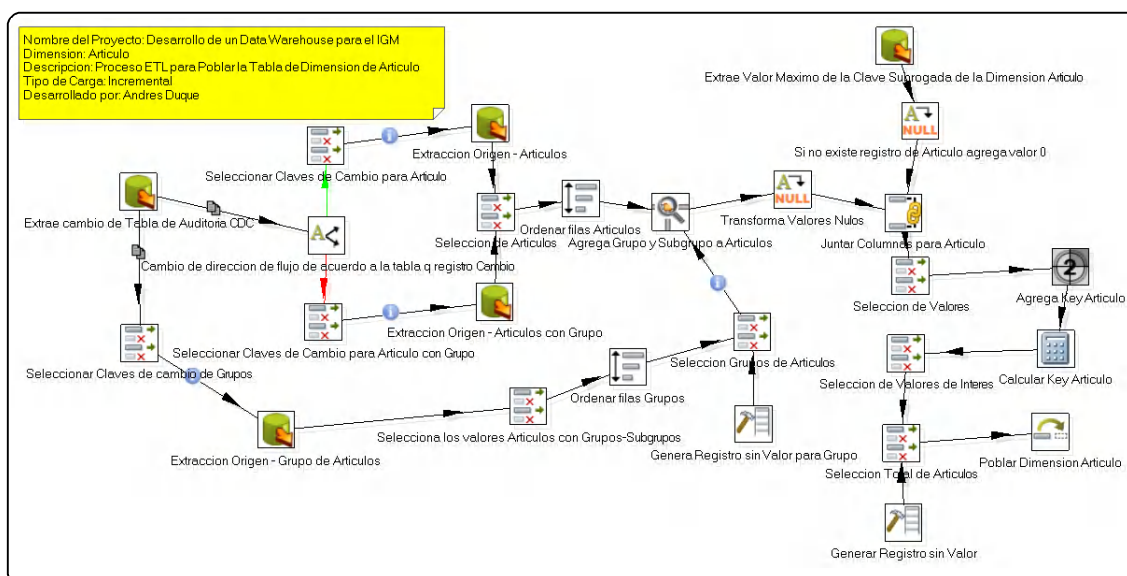


Figura 3.54: Transformación Dimensión Artículo

Dimensión Bodega

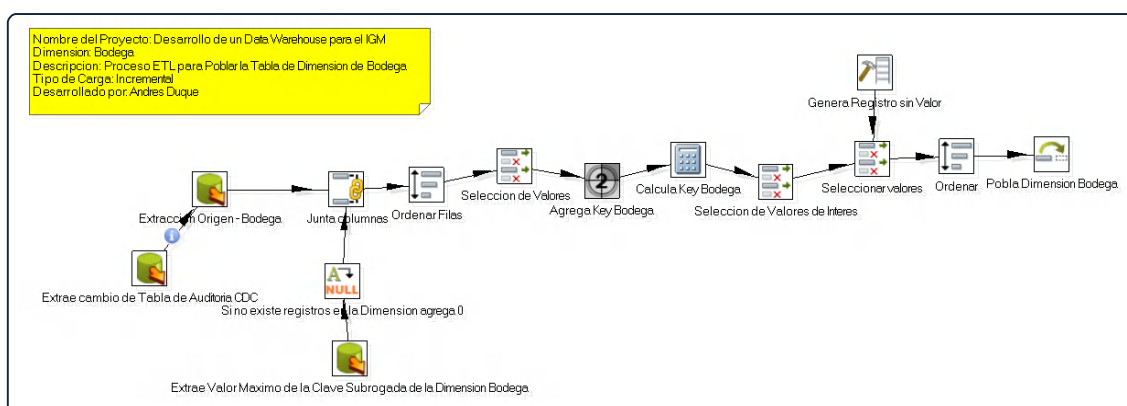


Figura 3.55: Transformación Dimensión Bodega

Dimensión Centro de Gestión

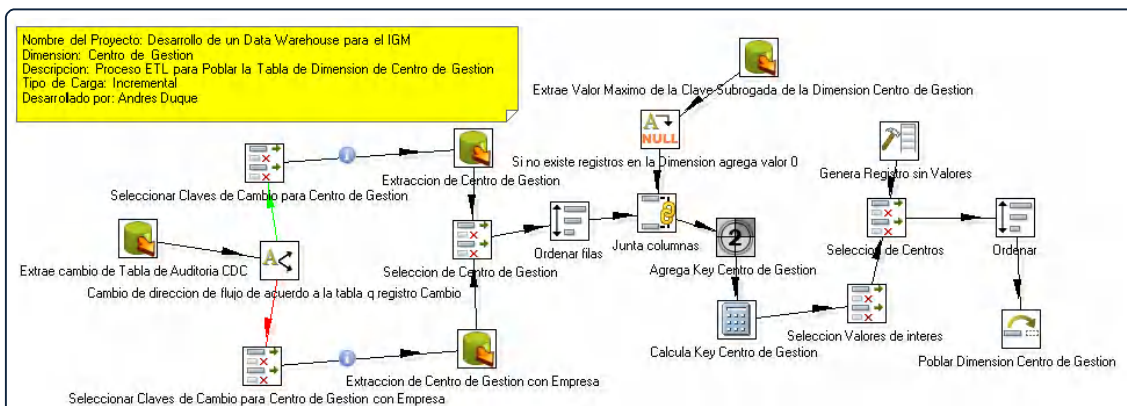


Figura 3.56: Transformación Dimensión Centro de Gestión

Dimensión Cliente

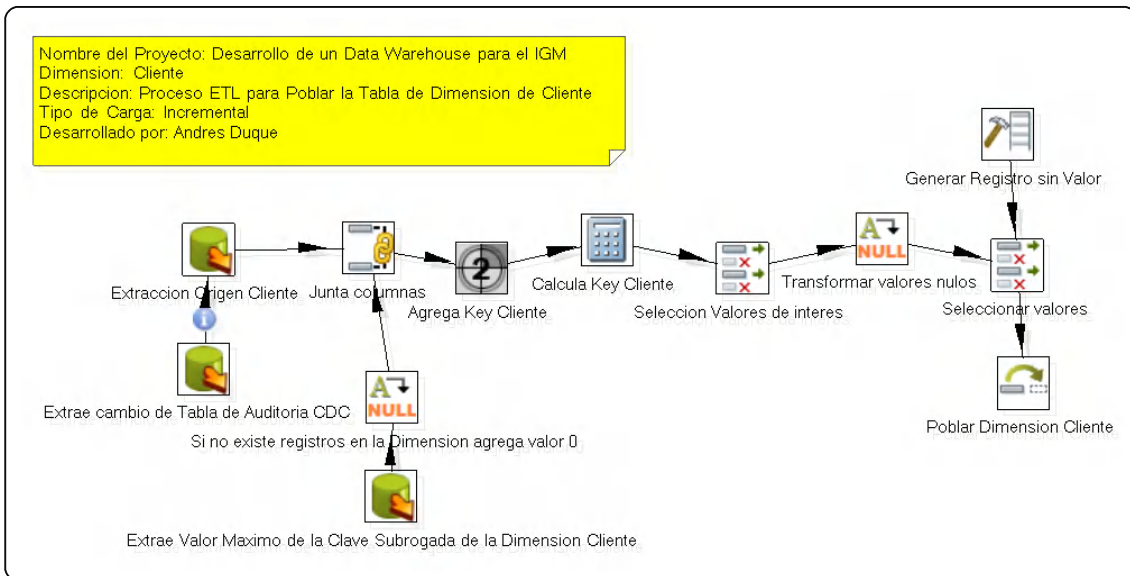


Figura 3.57: Transformación Dimensión Cliente

Dimensión Empleado

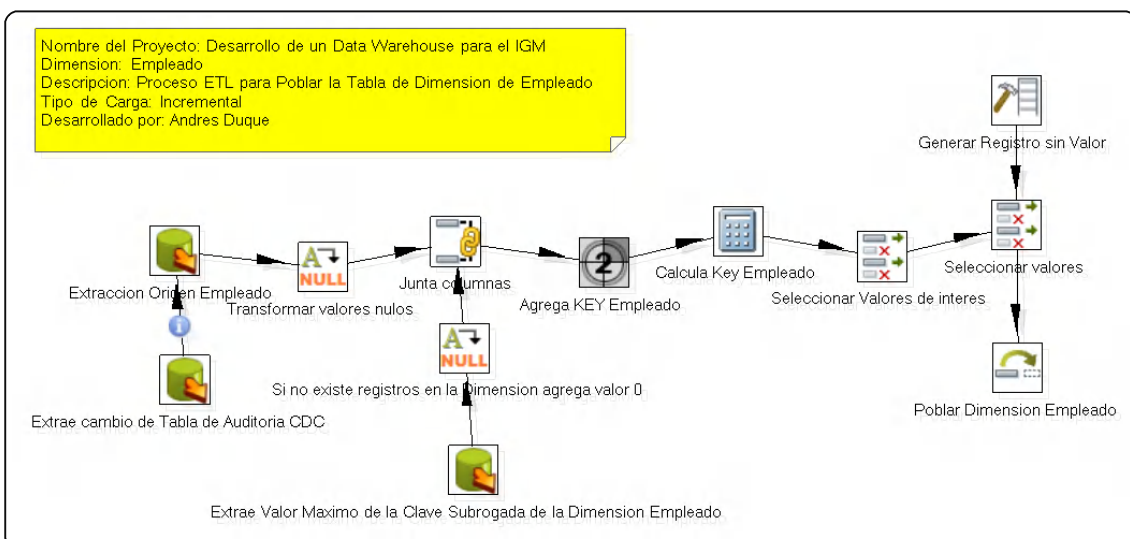


Figura 3.58: Transformación Dimensión Empleado

Dimensión Factura

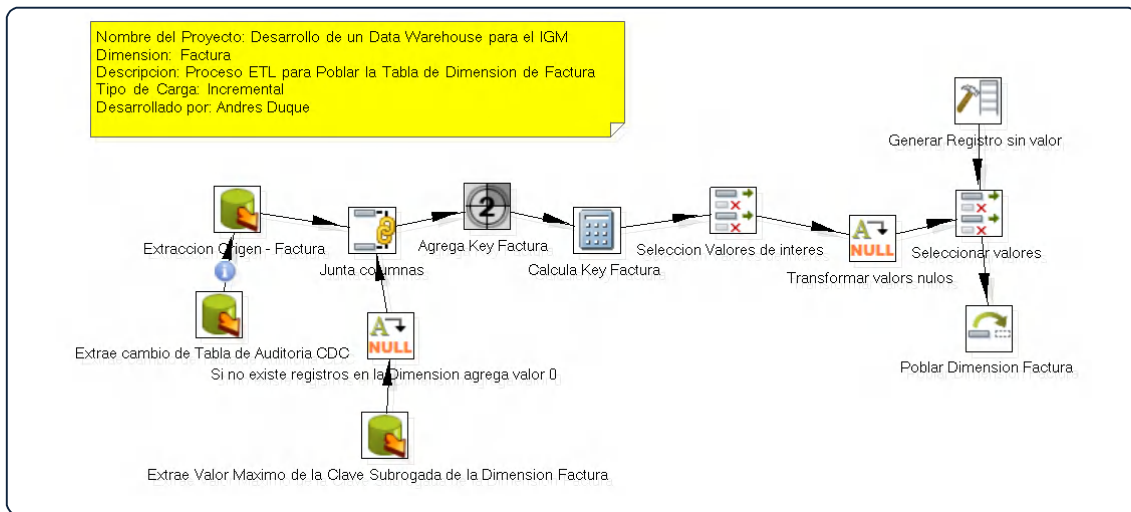


Figura 3.59: Transformación Dimensión Factura

Dimensión Ítem

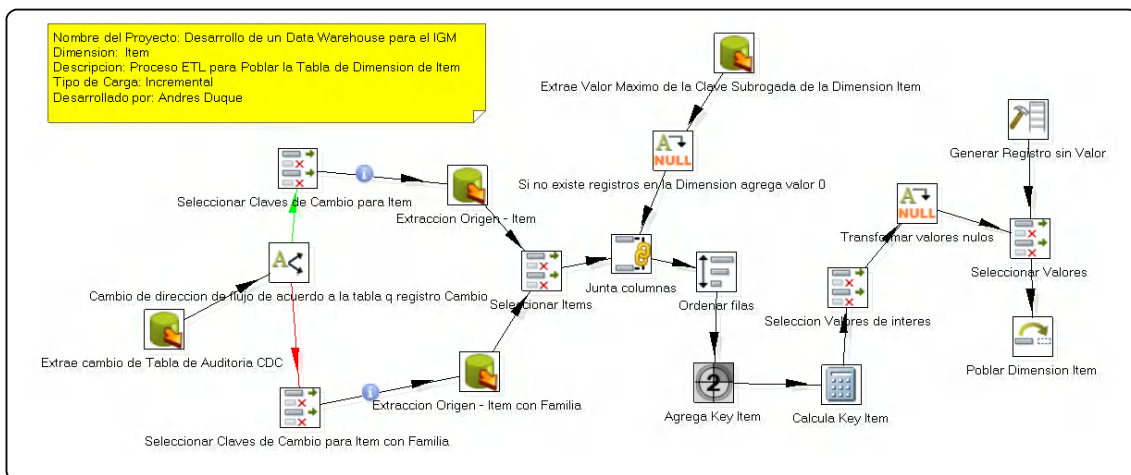


Figura 3.60: Transformación Dimensión Ítem

Dimensión Movimiento

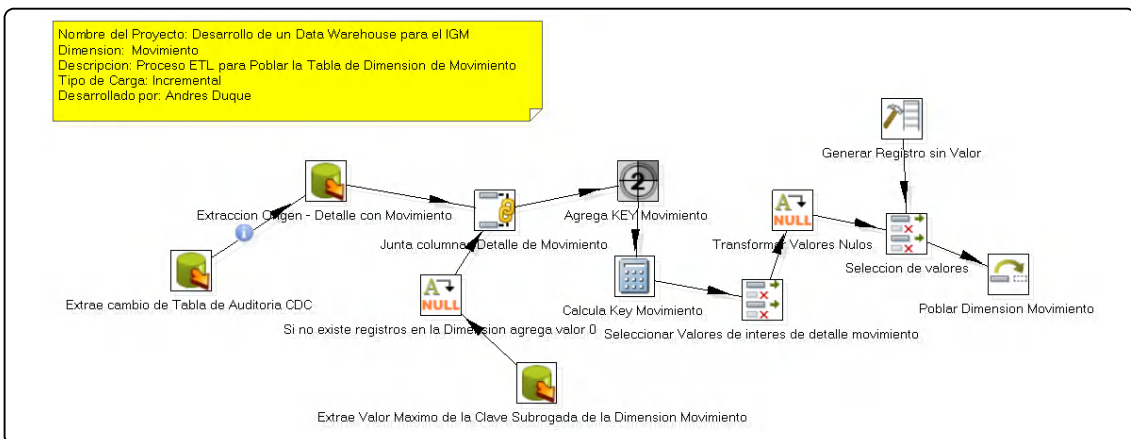


Figura 3.61: Transformación Dimensión Movimiento

Dimensión Orden de Compra

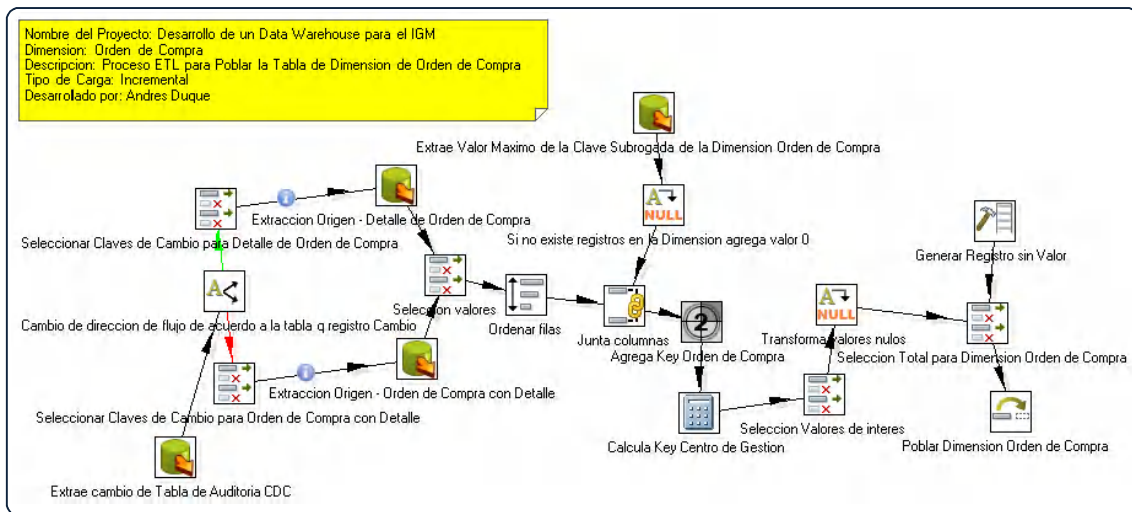


Figura 3.62: Transformación Dimensión Orden de Compra

Dimensión Pedido

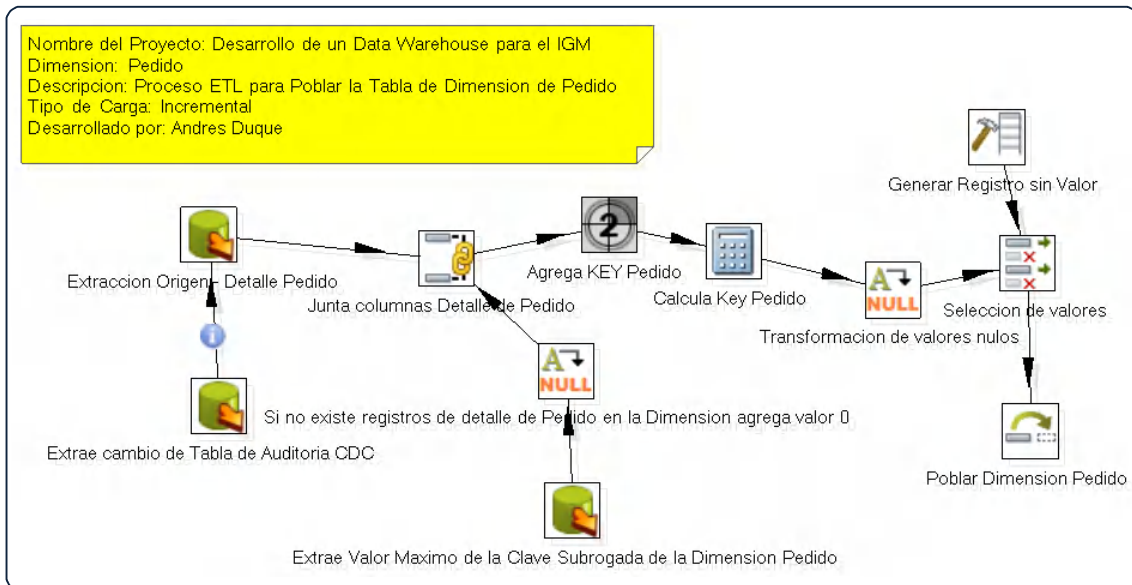


Figura 3.63: Transformación Dimensión Pedido

Dimensión Proforma

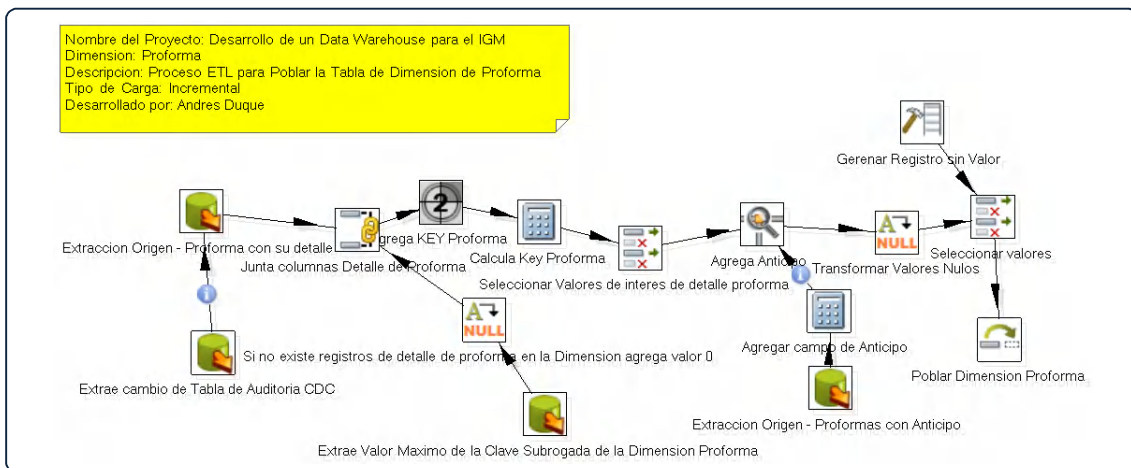


Figura 3.64: Transformación Dimensión Proforma

Dimensión Proveedor

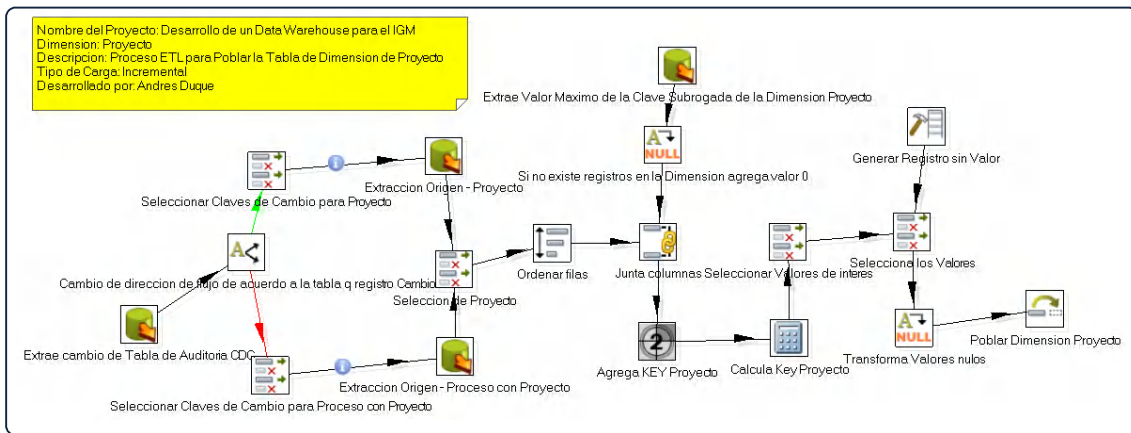


Figura 3.65: Transformación Dimensión Proveedor

Dimensión Proyecto

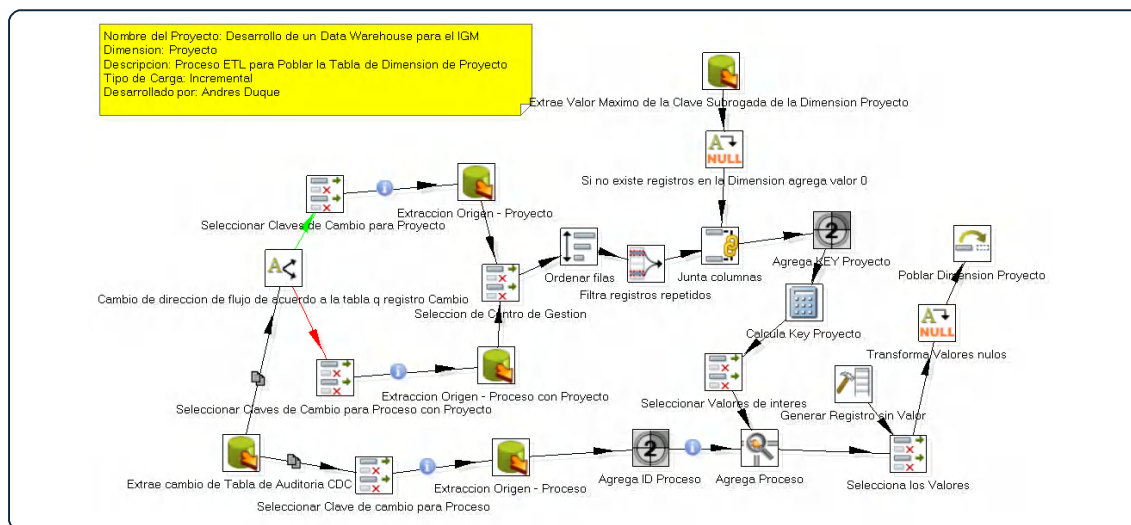


Figura 3.66: Transformación Dimensión Proyecto

Dimensión Sucursal y Tiempo

En cuanto a las Dimensiones de Tiempo y Sucursal se mantuvo el mismo diseño de la carga inicial debido a que la primera es considerada como dimensión estática y no tiene datos origen por su diseño, y mientras que los registros manipulados son bajos para Sucursales, tomando en cuenta que el Instituto tiene solo dos.

3.8.8. PASO 8: PROCESAMIENTO INCREMENTAL DE LOS HECHOS

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

3.8.8.1. Identificar Filas Nuevas y con Cambios

En la carga Incremental para los hechos en cuanto a identificar los registros se maneja la misma idea que las Dimensiones, como se muestra en el gráfico a continuación.

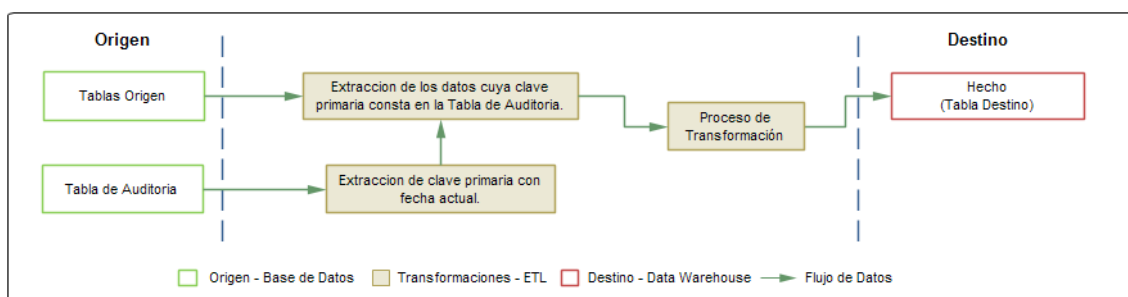


Figura 3.67: Funcionamiento CDC – Hechos

3.8.8.2. Extracción de Claves Subrogadas de Filas de cambio

Se agrego este punto por la importancia técnica y la cual merece ser destacada. En la carga Incremental después de haber extraído únicamente los datos nuevos o aquellos que han sufrido cambios se optimiza mucho el tiempo y el rendimiento en el Proceso ETL de los datos, pero al momento de agregarles la clave subrogada de acuerdo a cada Dimensión, se extraerá todos los datos de las mismas, haciendo que el rendimiento dependa de la cantidad de datos que exista en las Dimensiones. Por lo que para optimizar aun más el proceso ETL se diseño una manera en la cual solo se extraiga los registros pertenecientes o relacionados al nuevo registro de hecho.

Para entender de mejor manera esta técnica, se presenta un gráfico a continuación que muestra la idea de su funcionamiento.

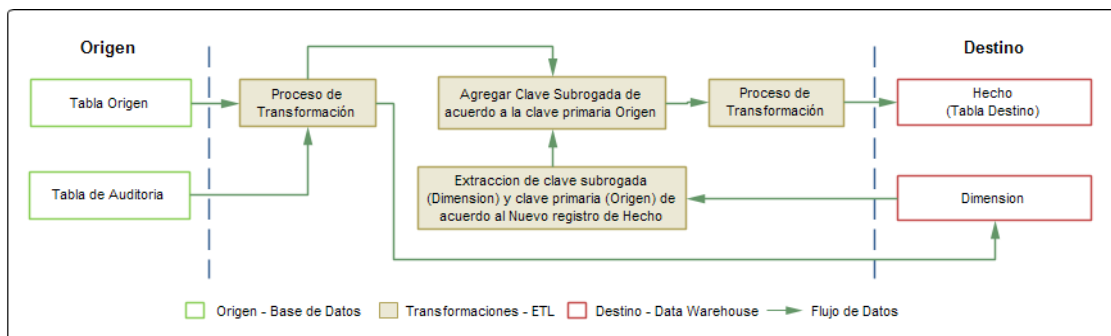


Figura 3.68: Extraer y Agregar Clave Subrogada

3.8.8.3. Desarrollo ETL Incremental de Hechos con la Herramienta de Pentaho

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

Hecho Movimientos

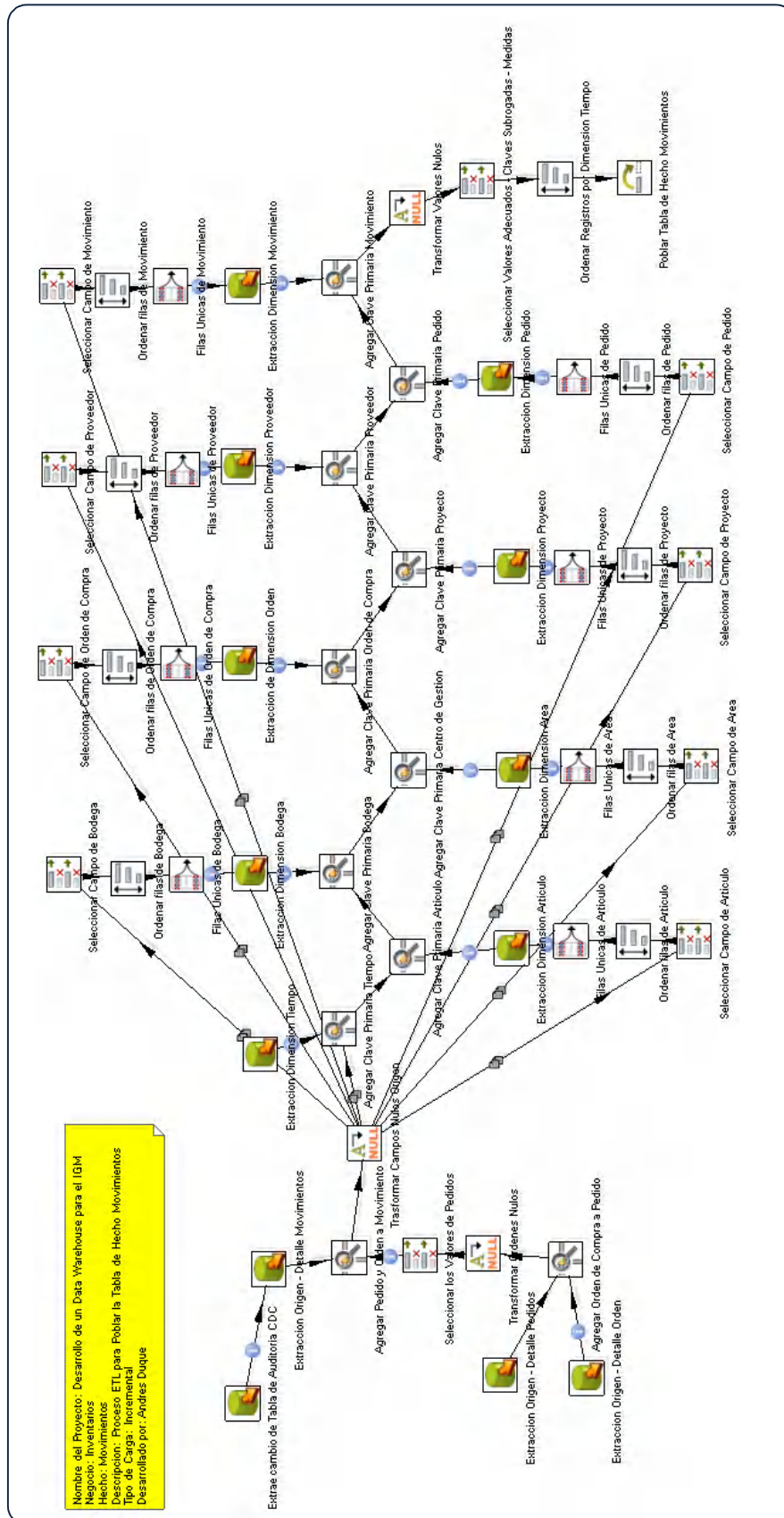


Figura 3.70: Transformación Hecho Movimientos

Hecho Ventas

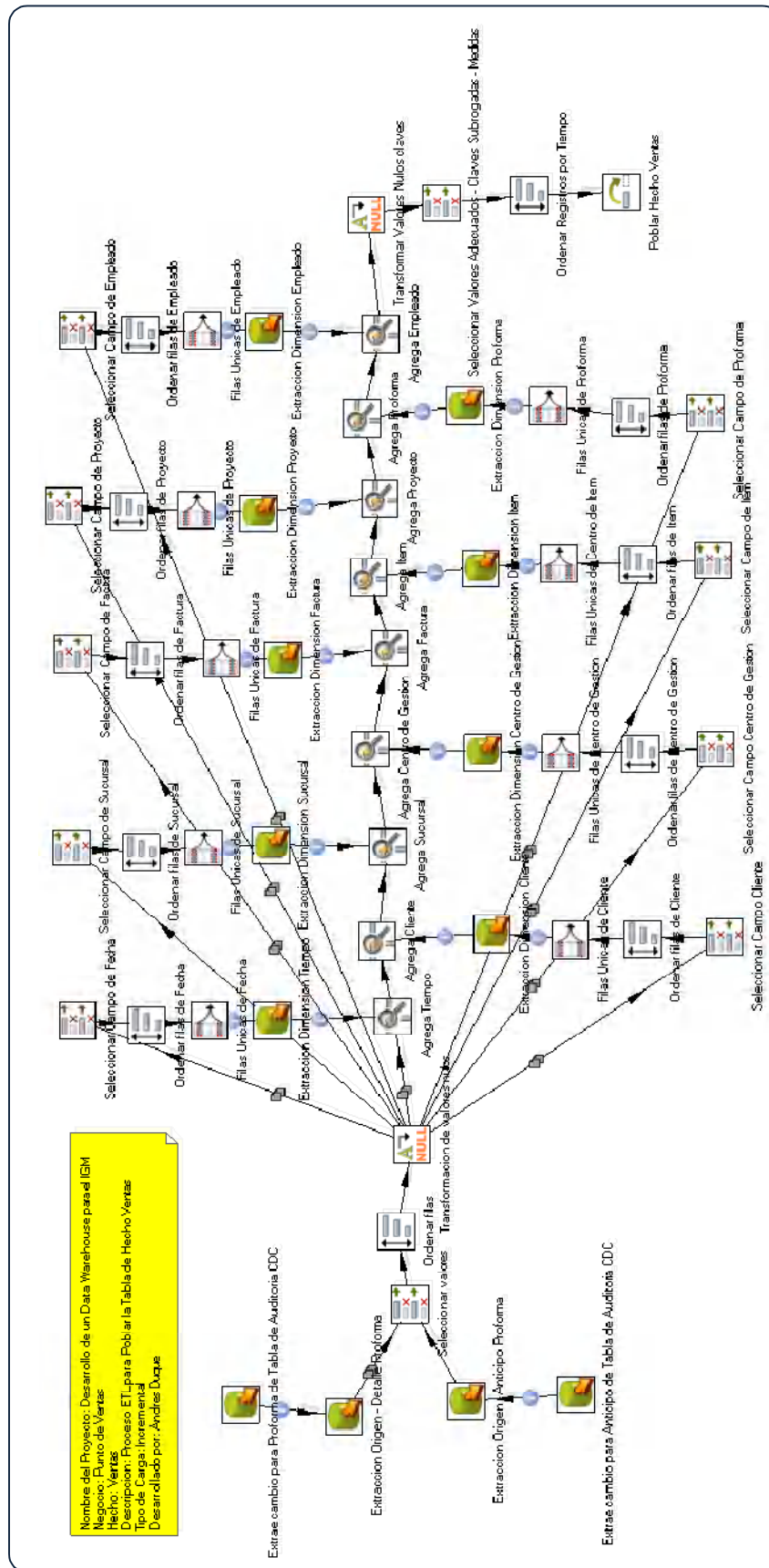


Figura 3.71: Transformación Hecho Ventas

Hecho Numero de Ventas

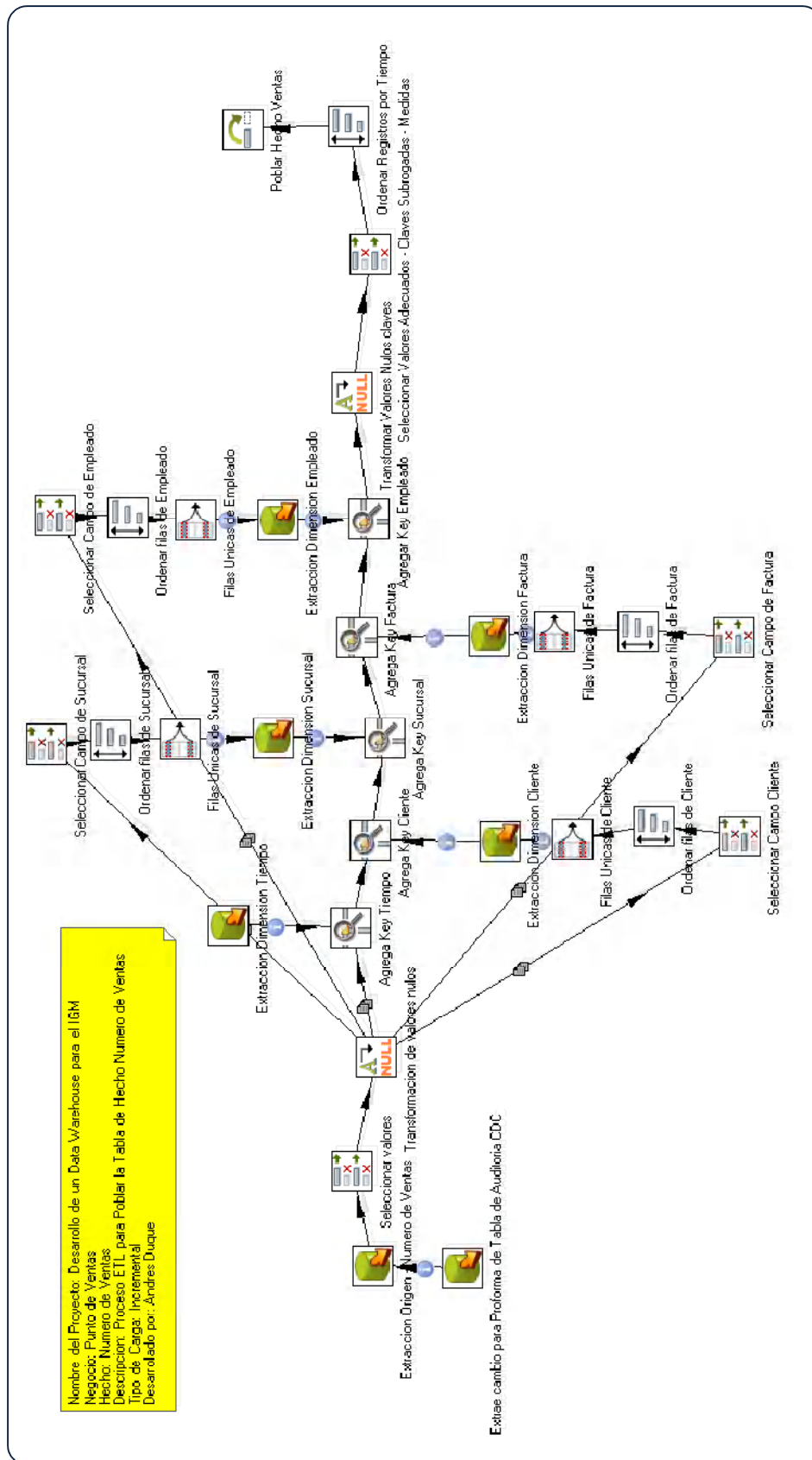


Figura 3.72: Transformación Hecho Numero de Ventas

3.8.9. PASO 9: OPERACIÓN Y AUTOMATIZACIÓN DEL SISTEMA ETL

Se realizó dos actividades principales, empleando trabajos para dar un orden y secuencia a las transformaciones, en conjunto con la programación de la tarea para que se ejecute periódicamente.

3.8.9.1. Trabajos ETL

Para los trabajos ETL se planteo los siguientes modelos, tanto para Dimensiones, como para hechos.

Verificar Dimensiones

Se empleará un trabajo el cual verifique si existe la tabla de Dimensión en el Data Warehouse como muestra el gráfico a continuación.

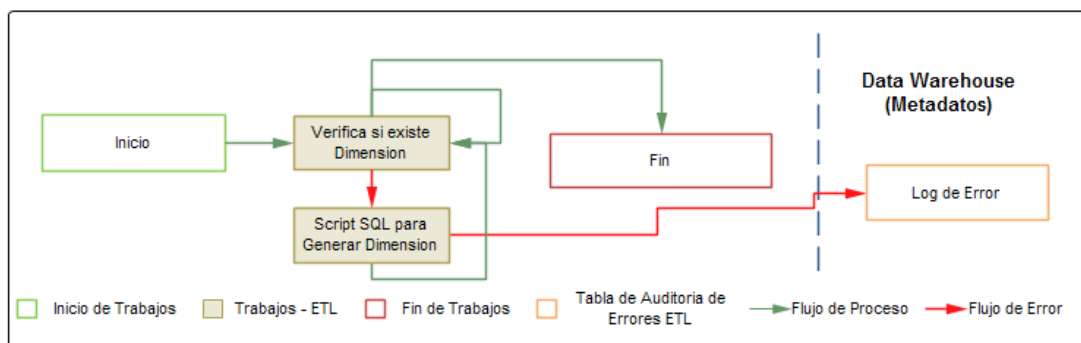


Figura 3.73: Verificar Dimensiones

Verificar Hechos

Al igual que las Dimensiones, para las Tablas de Hechos también se realizará una verificación de existencia o no, como muestra la figura a continuación.

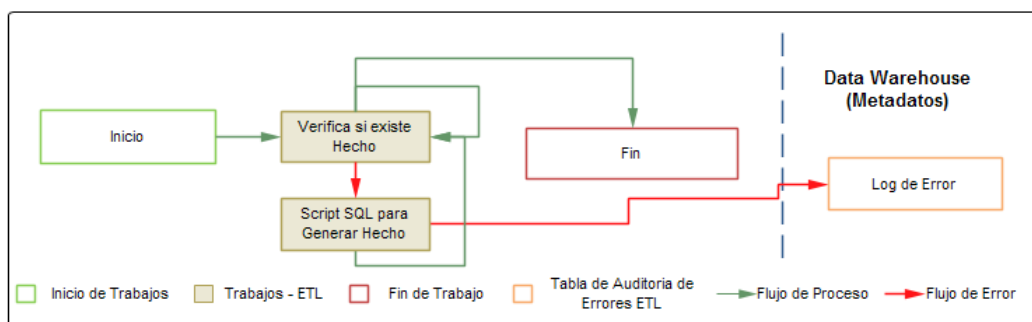


Figura 3.74: Verificar Hechos

Poblar Dimensiones

Para poblar las Dimensiones se realiza una condicional, en caso de haber registros en la Tabla se hará una carga Incremental de los datos, caso contrario se realizará un carga Inicial, al igual esto mantendrá una Auditoria de errores. Para una mejor comprensión se muestra el gráfico a continuación.

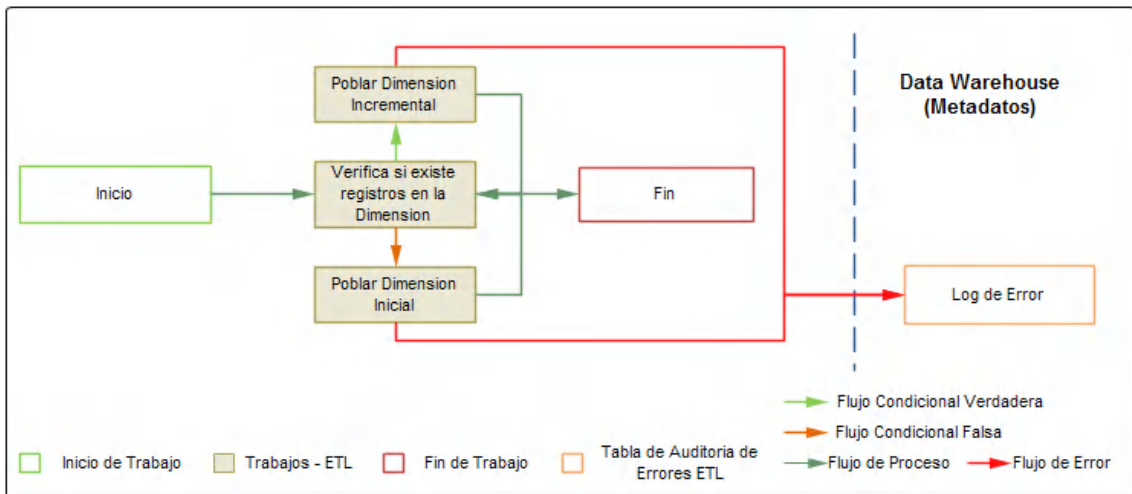


Figura 3.75: Poblar Dimensiones

Poblar Hechos

Al poblar las Tablas de Hechos de maneja la misma que las dimensiones, como se muestra el gráfico a continuación.

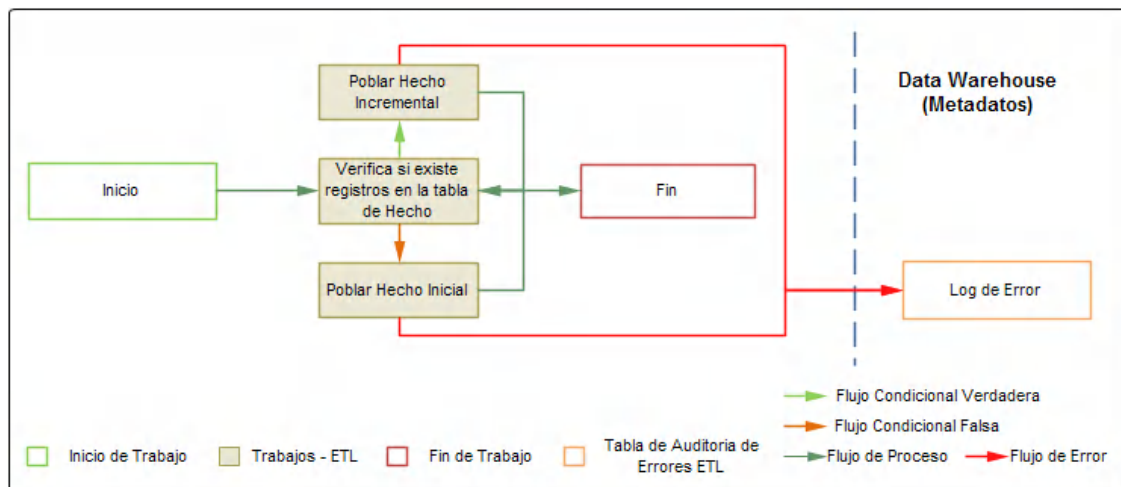


Figura 3.76: Poblar Hechos

Poblar DataWarehouse

Finalmente se realizará un trabajo el cual englobe los trabajos anteriores lo que permitirá seguir una secuencia para poblar el DataWarehouse. Para un mejor entendimiento se ilustra con un gráfico que se presenta a continuación.

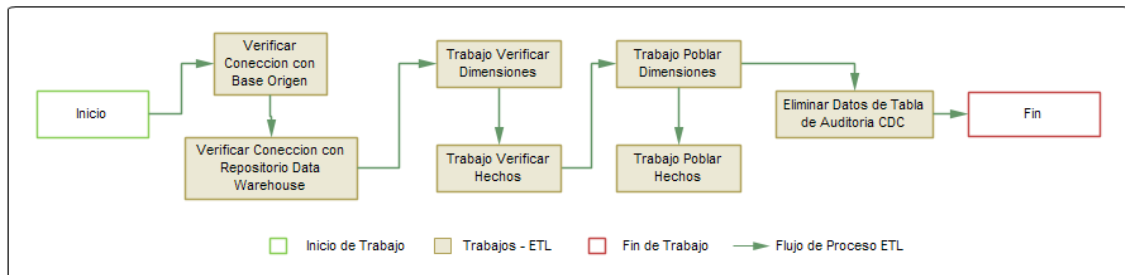


Figura 3.77: Poblar DataWarehouse

3.8.9.2. Programar Trabajos

El trabajo Principal que es el de Poblar el DataWarehouse es el que se establecerá un horario para que mantenga un periodicidad de ejecución. Al ser la herramienta ETL ajena a la Base de Datos en el sentido que no son de la misma casa fabricante, esto mantiene una debilidad al no poder establecer ejecuciones periódicas a estos trabajos directamente desde la Base de Datos como son los Jobs de Oracle.

Por lo que para poder programar el trabajo se lo realizará mediante comandos del Sistema, es decir mediante batch o script, especificando la frecuencia de ejecución del trabajo.

3.8.9.3. Desarrollo ETL de los Trabajos con la Herramienta de Pentaho

Verificar Dimensiones

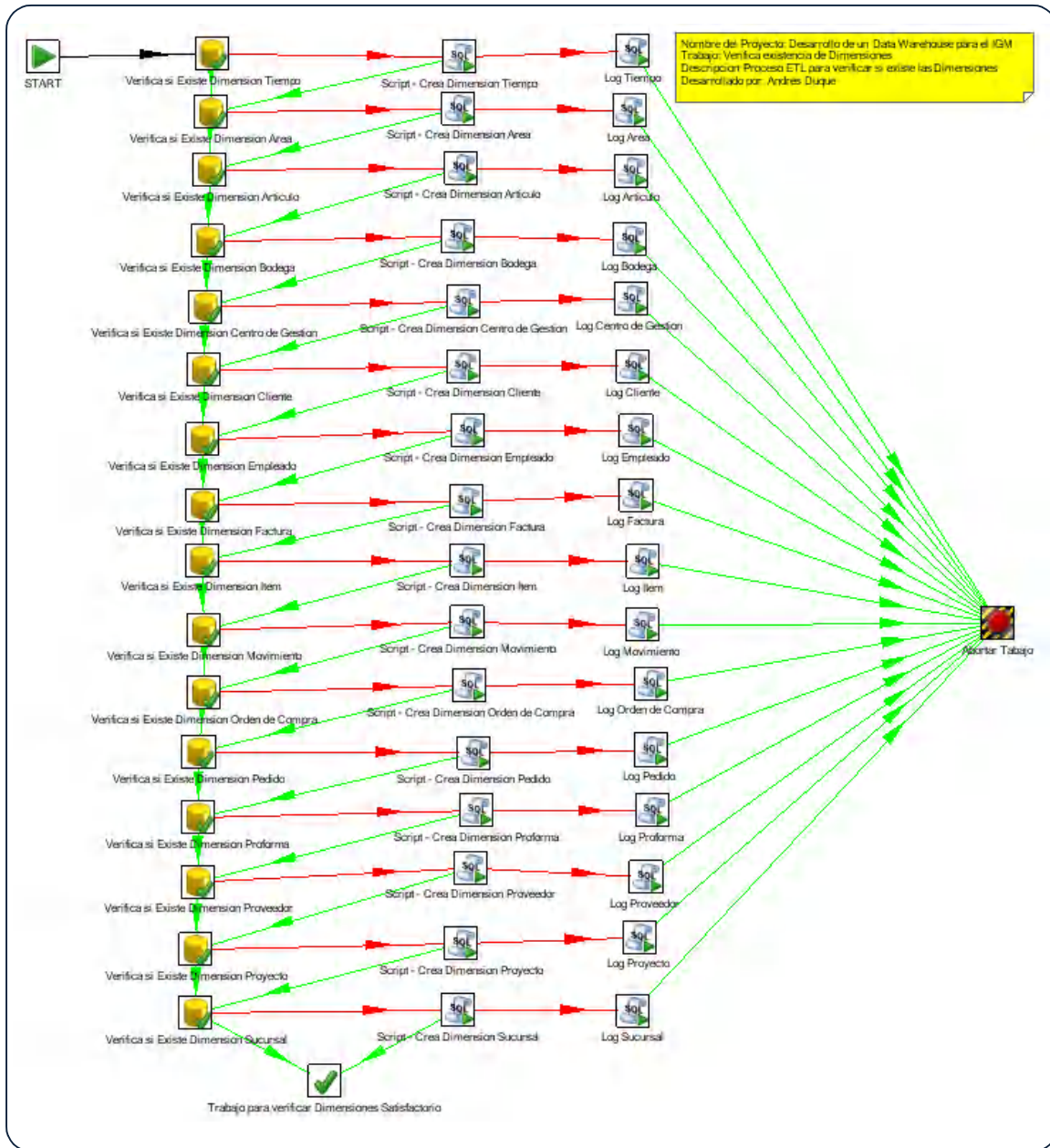


Figura 3.78: Trabajo Verificar Dimensiones

Verificar Hechos

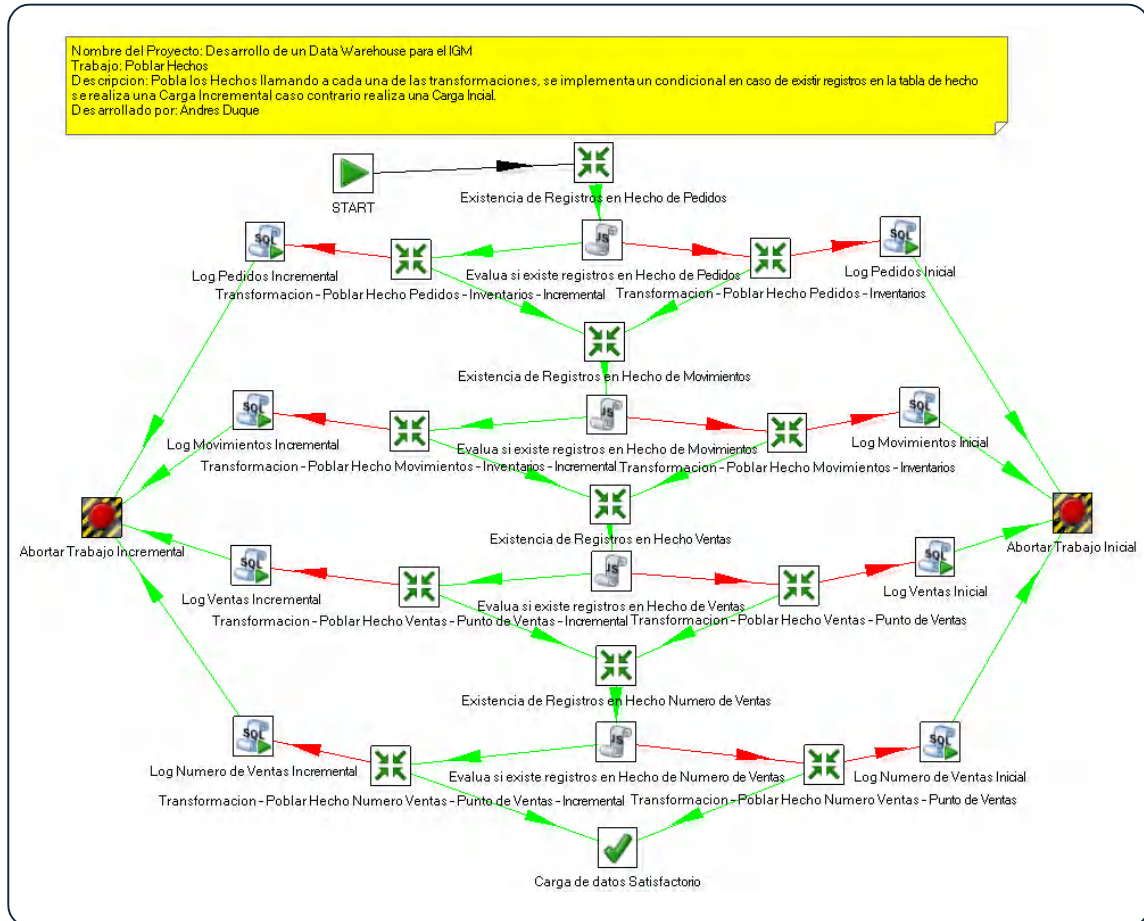


Figura 3.79: Trabajo Verificar Hechos

Poblar Hechos

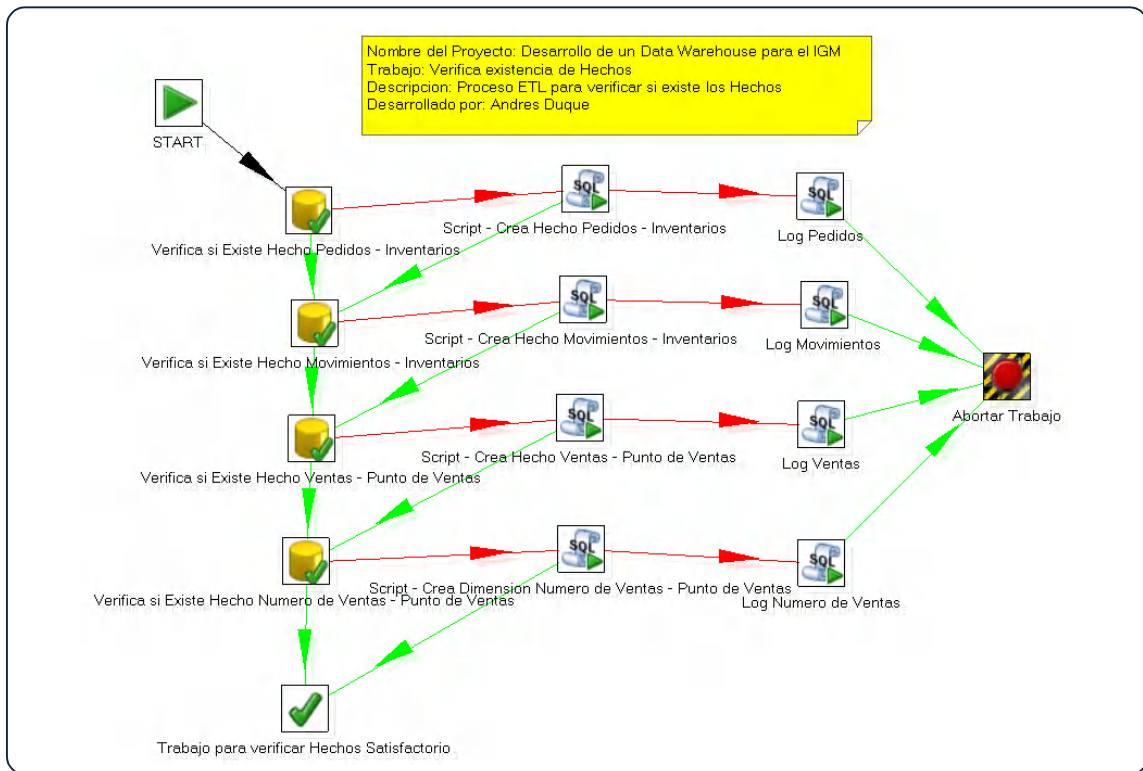


Figura 3.81: Trabajo Poblar Hechos

Poblar DataWarehouse

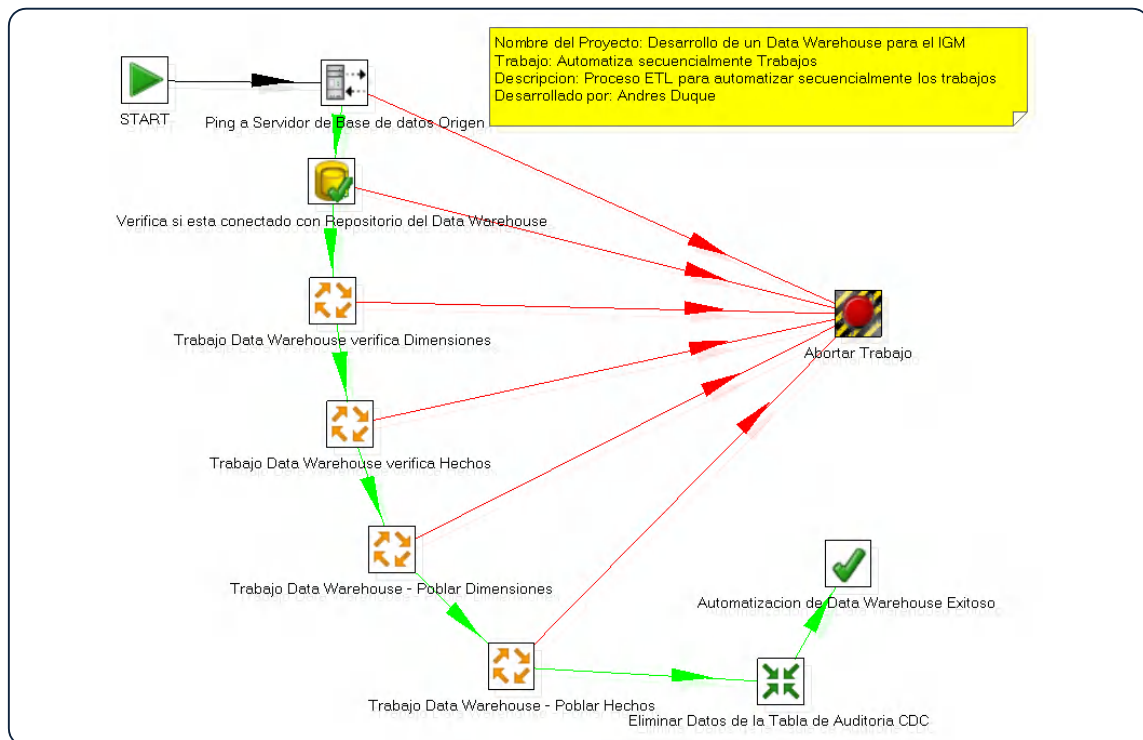


Figura 3.82: Trabajo Poblar DataWarehouse

3.9. DISEÑO Y DESARROLLO DE APLICACIÓN DE BI

3.9.1. TIPO DE APLICACIÓN DE BI

La aplicación de BI al cual está enfocado es:

- Reportes Estándar y Cubos de Información.

Este tipo de aplicación está orientado para presentar la información mediante reportes Diseñados, o los Cubos de Información para visualizar los datos (Métricas o Indicadores) de manera Multidimensional, permitiendo la navegabilidad entre Dimensiones.

Definido el Tipo de Aplicación de BI se procederá al desarrollo.

3.9.2. PREPARAR EL DESARROLLO DE LA APLICACIÓN

Se configurará la herramienta de BI para poder acceder a los datos del DataWarehouse, así como también los metadatos de ambos negocios, para que la herramienta presente los datos en una capa amigable para el usuario.

3.9.2.1. POBLAR LOS METADATOS DE LA HERRAMIENTA DE BI

Dentro del Instituto, se encuentra en funcionamiento la Herramienta de BI que dará el soporte al usuario final, esta herramienta es Oracle Discoverer, por lo cual no se necesitará poblar los metadatos de la herramienta.

3.9.2.2. CREAR LOS METADATOS DE NEGOCIO

Se crearán los metadatos de las Ares de Negocio tanto para Inventarios como para Punto de Ventas, ya que el Discoverer crea un Capa de Usuario Final (EUL), con el fin de generar una Interface para el Usuario lo más fácil de entender.

La creación de los metadatos se lo puede apreciar de manera grafica en el punto 3.9.4., en la que ya se muestra el desarrollo con la herramienta de Discoverer.

3.9.2.3. TESTEAR LA HERRAMIENTA DE BI

Se realiza pruebas con la herramienta a partir de las capas de usuario creadas, tanto de Inventarios como de Punto de Ventas, con el fin de verificar que todo ande bien, y de poder corregir errores en caso de que sucedan.

Se realiza pequeñas interfaces y consultas, para visualizar los datos del DataWarehouse, junto con gráficos, y así probar el correcto desempeño de la herramienta.

3.9.3. CREAR LAS APLICACIONES

Después de haber configurado el EUL del Discoverer, se procede a generar los reportes de análisis de acuerdo a cada negocio, esto se detalle de manera gráfica en el punto 3.9.4., todo el proceso de creación de las EUL.

3.9.4. DESARROLLO DE APLICACIÓN DE BI CON LA HERRAMIENTA DISCOVERER DE ORACLE

Se creará los Niveles de Usuario Final (EUL) que serán las Áreas de Negocio, tanto de Inventarios como Puntos de Venta.

Creación de las Áreas de Negocio (EUL)

Para la creación de los EUL se tomará en cuenta lo siguiente:

- Conexión a la Base de Datos con el Esquema USER_REPOSITORIO, debido a que en ella se guardarán los metadatos de los EUL.
- Referencia al esquema USER_DWH, con el fin de seleccionar las tablas, tanto de Inventarios como de Puntos de Venta, para conformar los EUL correspondientes.

Se procede a crear los EUL tomando las tablas pertenecientes a cada negocio, como muestran las dos figuras a continuación.

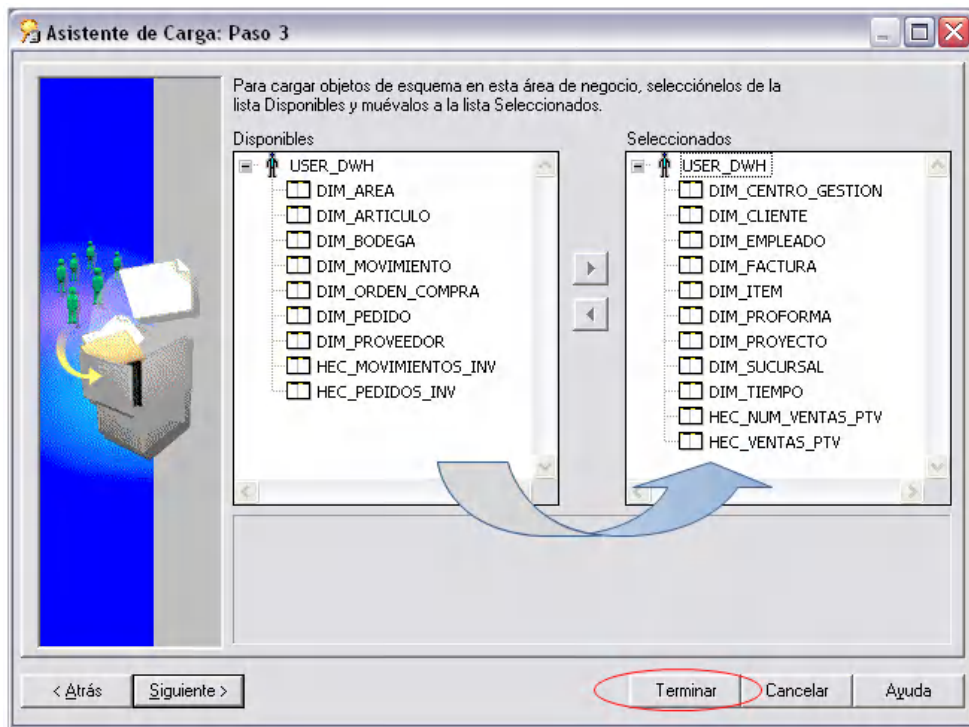


Figura 3.83: Selección de Dimensiones para EUL

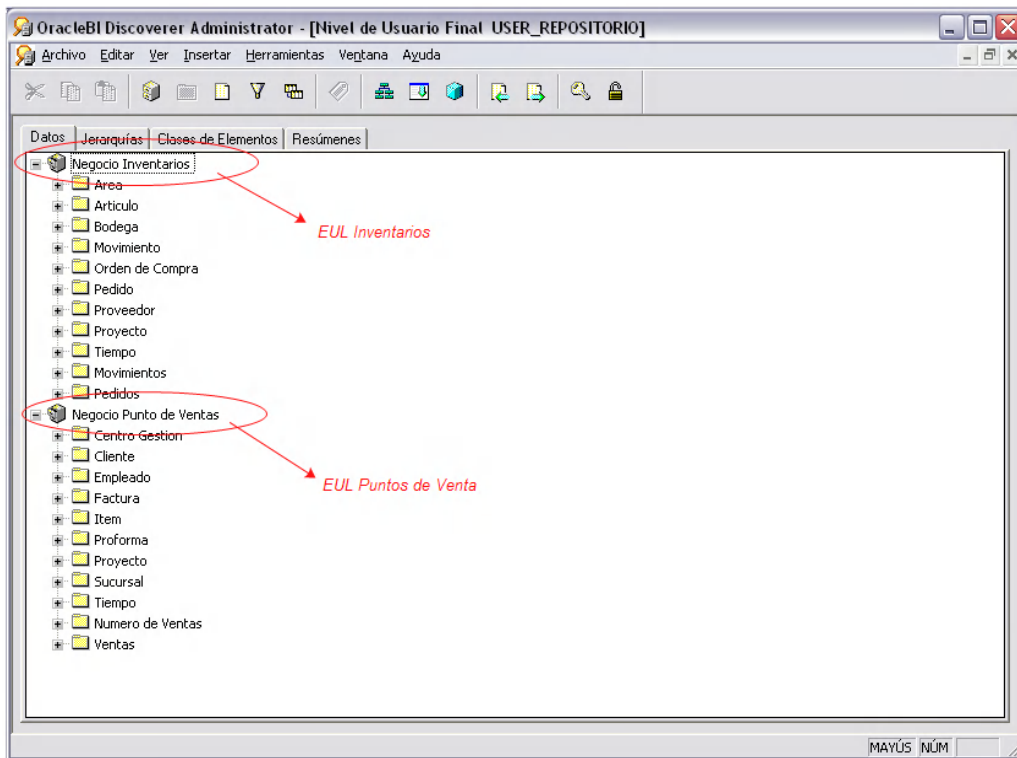


Figura 3.84: EUL's de Inventarios y Puntos de Venta

Modificación de las Áreas de Negocio

Se modifica las Áreas de Negocio con el fin de presentar información transparente y la necesaria al Usuario, por ejemplo el ocultar campos técnicos de las tablas como las claves primarias.

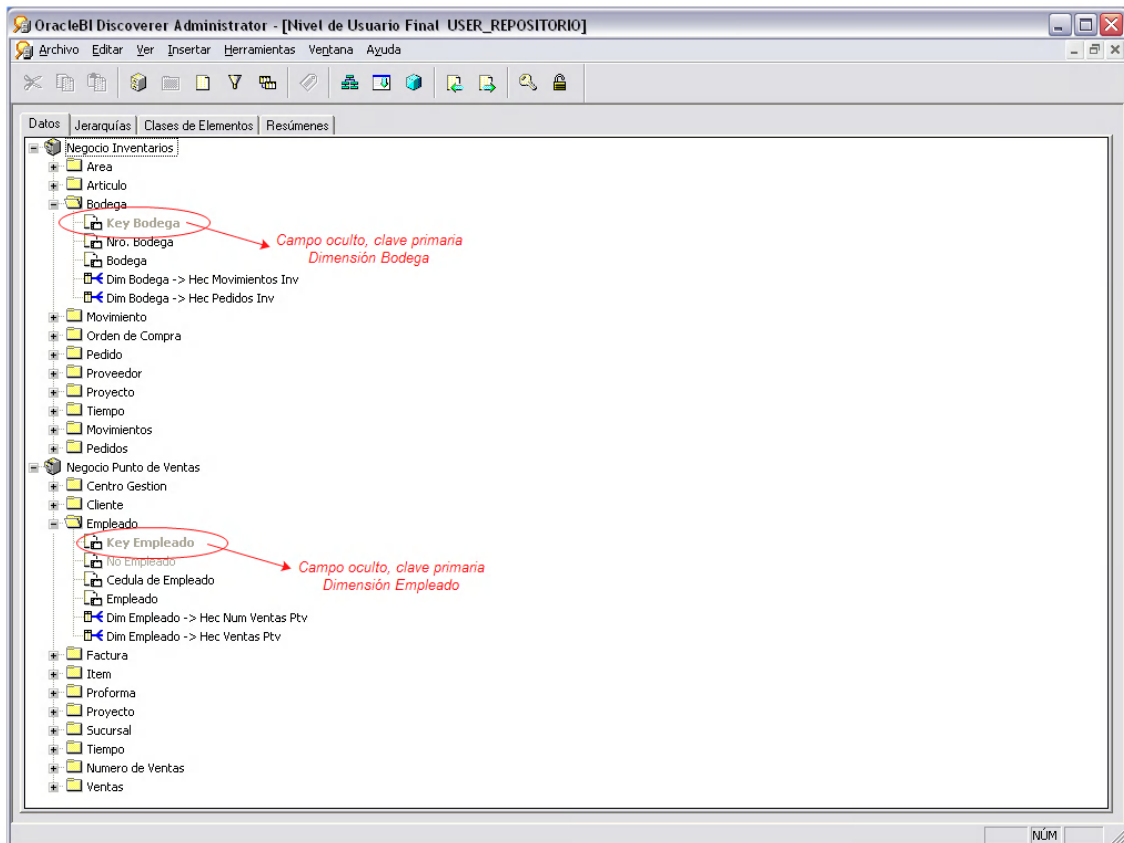


Figura 3.85: Campos ocultos de Dimensiones

Creación de Jerarquías en los EUL para Dimensiones

Se crea la jerarquía de la Dimensión de acuerdo a los niveles, dándoles un orden secuencial, como muestra la figura a continuación:

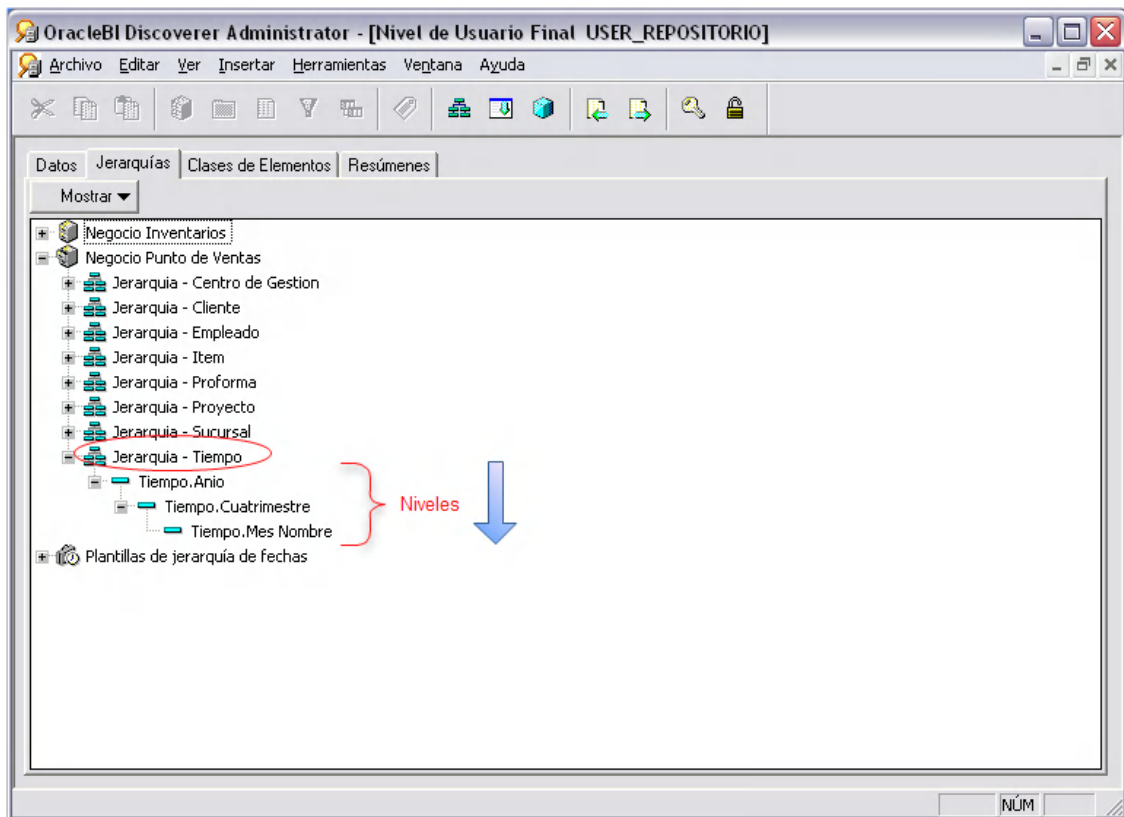


Figura 3.86: Niveles y Jerarquía Dimensión Tiempo

Concluido los puntos anteriores se tiene creado los Niveles de Usuario Final (EUL), los mismos que serán transparentes a los usuarios, manteniendo así campos únicamente con significado de negocio.

3.10. DESPLIEGUE

Antes de realizar el despliegue ETL y de BI se debe realizar las siguientes actividades, correspondiente a la parte física en la Base de Datos origen y el DataWarehouse:

- Creación de espacio de tablas y esquemas para la base de datos destino (DataWarehouse), para el repositorio donde se alojarán los metadatos, por medio de las representaciones del negocio, en Dimensiones y Hechos.

- Crear un espacio de tablas y un esquema en la base de datos origen, con el fin de alojar los datos de la tabla de Auditoría.
- Ejecutar el script de disparadores en la base de datos origen, para el sistema de Auditoría CDC.

Adicional y con el fin poner en producción el DataWarehouse se debe tomar en cuenta aspectos como:

- Recursos Humanos:
 - Personal técnico capacitado en el desarrollo y administración de sistemas de DataWarehouse.
 - Administrador de Base de Datos.
 - Personal capaz de recolectar requerimientos gerenciales.
 - Personal capacitado en Sistemas de Información.
 - Personal con un enfoque fuerte sobre el conocimiento del área de la empresa y de los procesos empresariales.
- Recursos de Hardware:
 - Servidor con prestaciones básicas para soportar las transacciones de datos, con almacenamiento superior al 1 Tera, debido a la cantidad de datos a contener en el Data Warehouse, además de escalable debido a que en futuros proyectos o por nuevos requerimientos se puede tener redimensionamiento en el almacenamiento.
- Recursos de Software:
 - Aplicación para Administrar la Base de Datos.
 - Aplicación o Herramienta de ETL para administrar todo el proceso de manipulación de datos Origen – Destino.

- Herramientas para la presentación de información, o reportes a usuarios finales.

3.10.1. DESPLIEGUE ETL

Creación del Repositorio

Se crea el repositorio y la estructura de almacenamiento de la herramienta ETL mediante tablas, en las cuales se registran los metadatos de las transformaciones y trabajos.

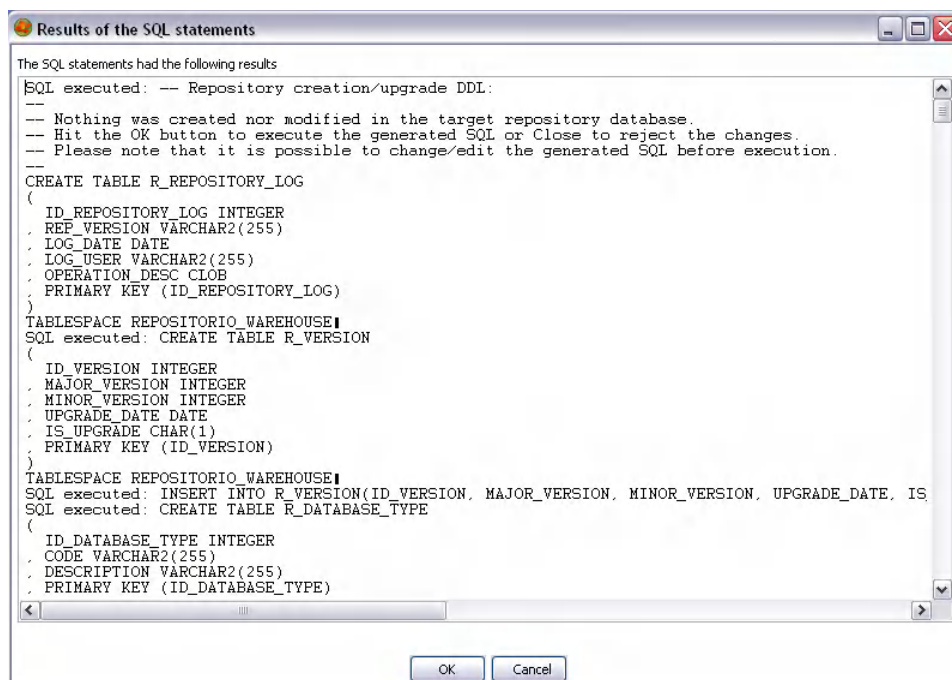


Figura 3.87: Resultado de Script de creación de repositorio

Guardar o Desplegar Transformaciones y Trabajos

Se guarda las transformaciones y trabajos en la base de Datos, como metadatos en el Repositorio.

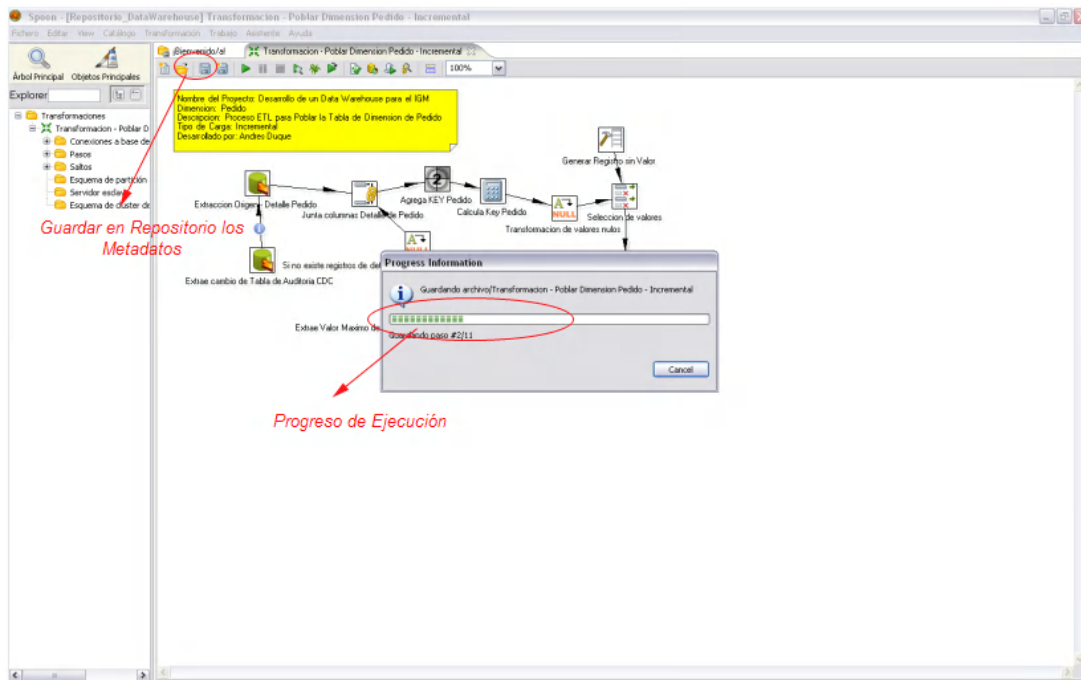


Figura 3.88: Progreso de metadatos (transformación) guardados

Verificación de Transformaciones y Trabajos Guardados

Se verifica los Metadatos de las transformaciones y los Hechos guardados en el Repositorio del DataWarehouse mediante la lista.

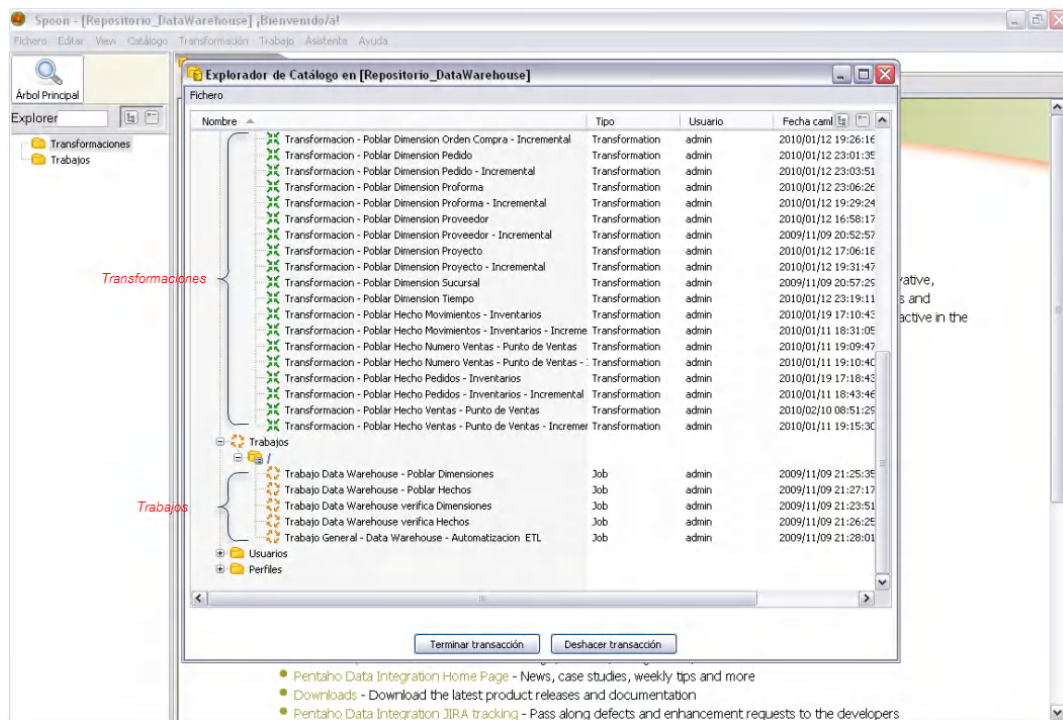


Figura 3.89: Catálogo de Transformaciones y Trabajos

Ejecución Inicial

Por medio del trabajo principal que automatiza al resto, se procede a ejecutar el proceso ETL, como muestra la figura a continuación, en la cual se puede ir observando el resultado de su ejecución.

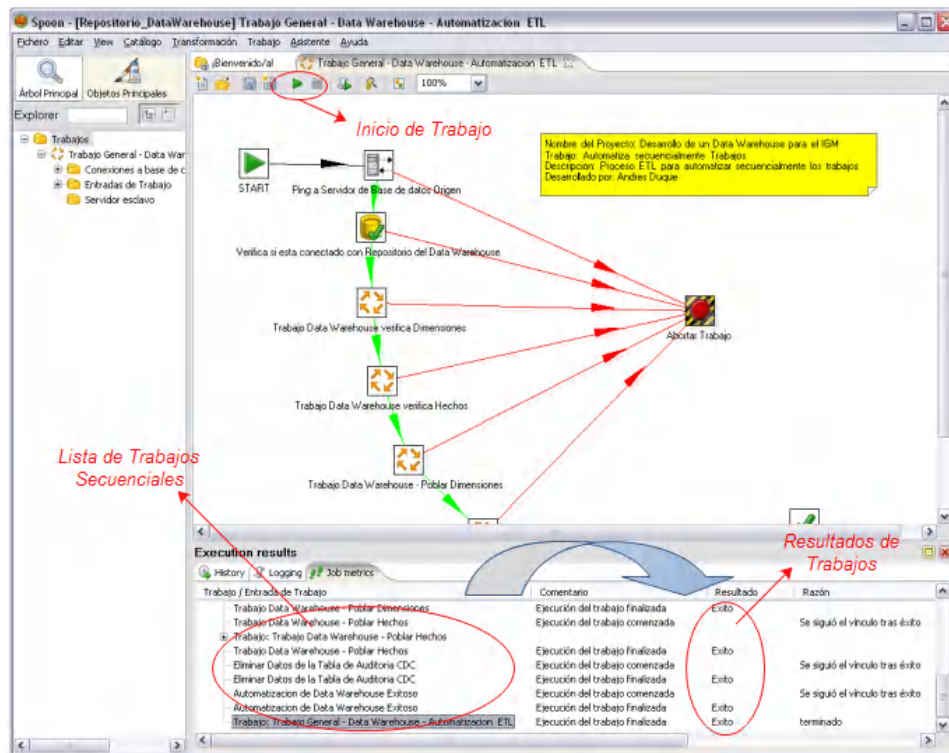


Figura 3.90: Ejecución Inicial

3.10.2. DESPLIEGUE BI

Tomando como base las Capas de Usuario Final (EUL), se procede a desarrollar los libros de trabajo junto con las hojas de datos de acuerdo a los requerimientos que se recolectaron para cada Negocio, Inventarios y Puntos de Venta.

3.10.2.1. Inventarios

Indicadores

- Ingresos a bodega anual sobre los dos millones de dólares y mensualmente superior a los ochenta mil dólares.

En los gráficos que se muestran a continuación se aprecia los Ingresos en dólares, tanto anual como mensual, identificados con colores, lo que indica: para el color verde, valores que están por encima de los 2 millones y ochenta mil respectivamente, mientras que de color amarillo: valores que se encuentran por debajo de las cifras antes mencionadas. Adicional se muestra los gráficos de barras para la evaluación anual y de tendencias para la mensual.

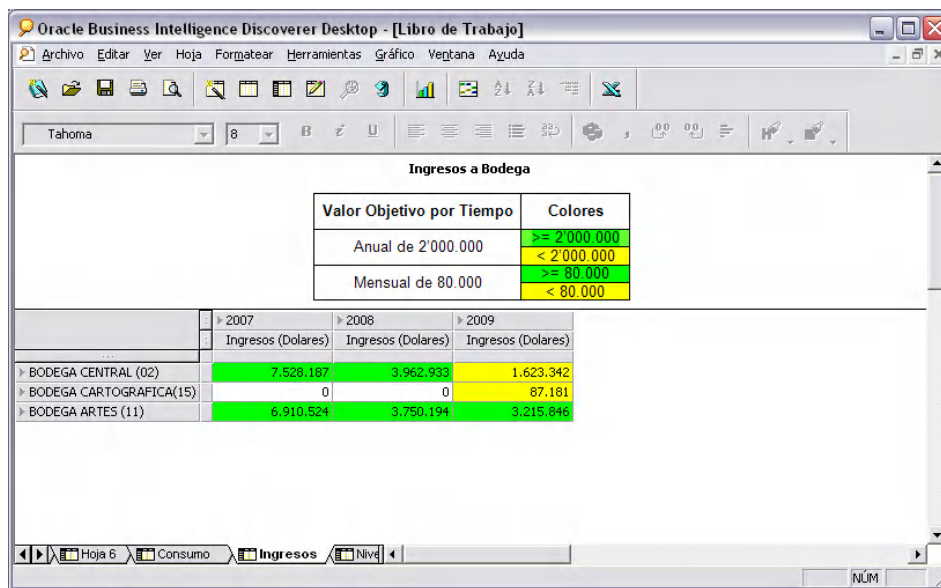


Figura 3.91: Ingresos Anuales

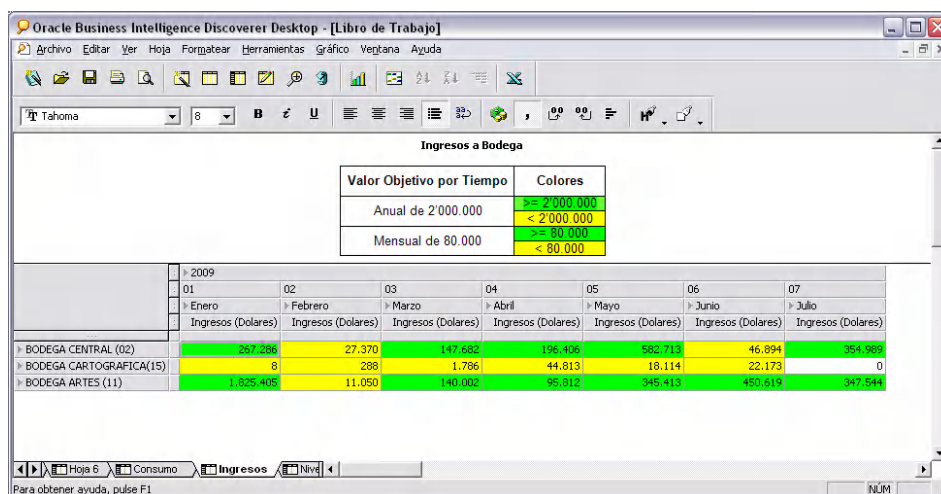


Figura 3.92: Ingresos Mensuales

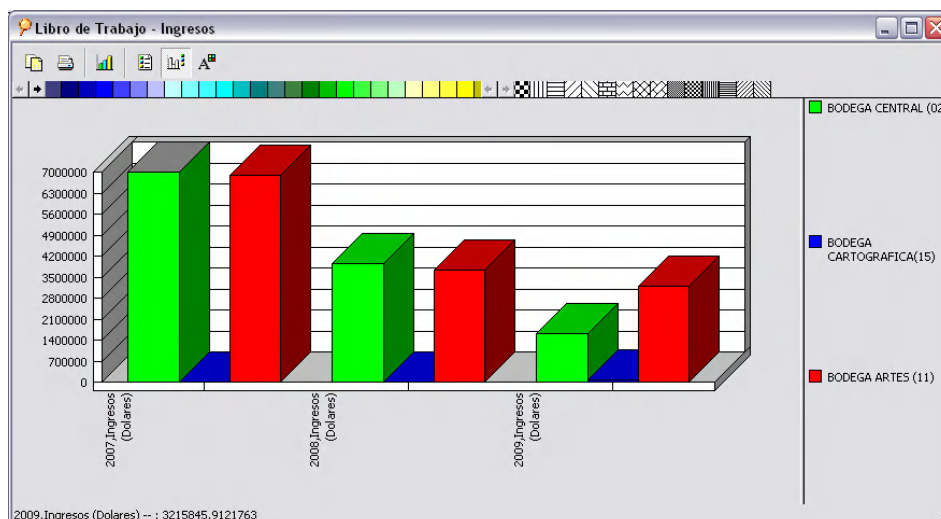


Figura 3.93: Gráfico de barras Ingresos Anuales

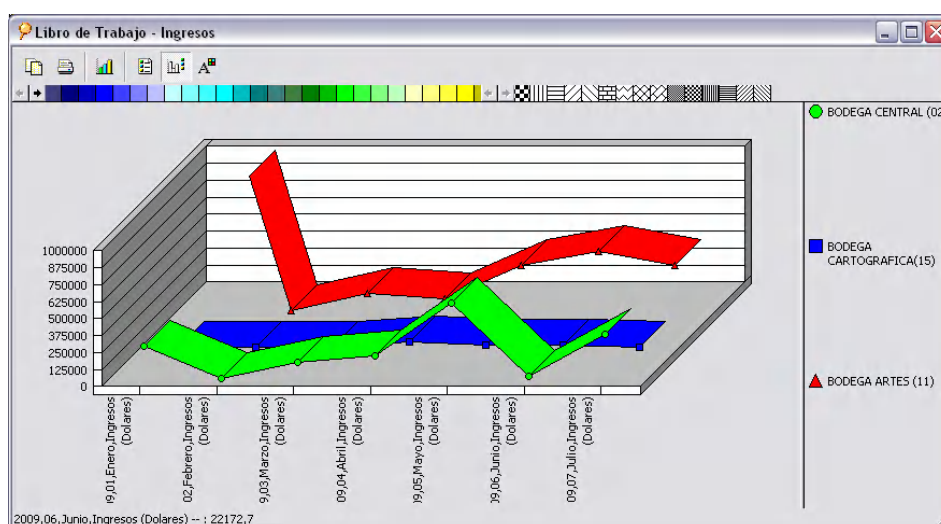


Figura 3.94: Gráficos de tendencia de Ingresos Mensuales

- Un millón de dólares en Consumo anual que se produce en la Bodega Central.

En los gráficos que se aprecia a continuación se identifica el consumo anual de la Bodega Central en al cual el color verde representa que se encuentra por encima del millón de dólares, mientras que el amarillo por debajo de esta cifra. Adicional se observa un gráfico de tendencias durante los años.

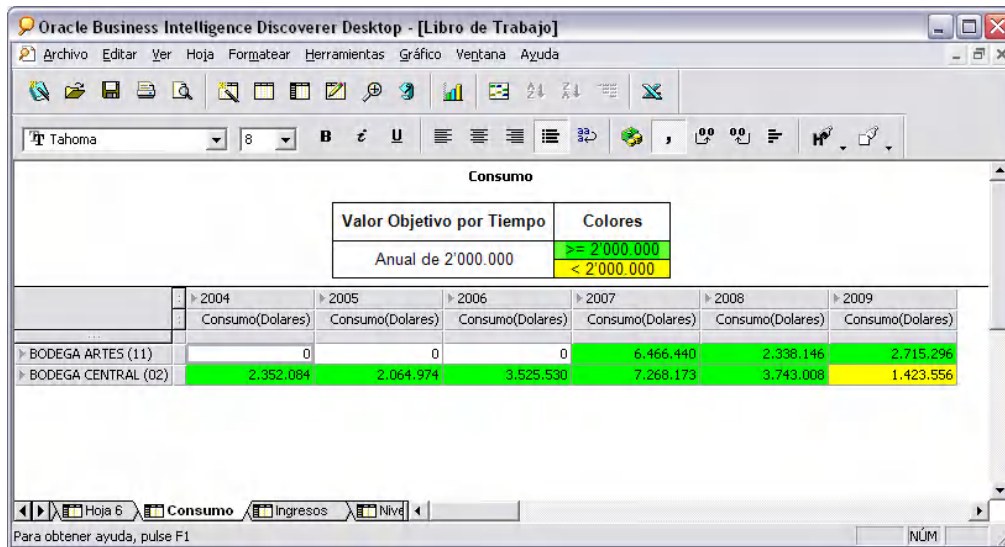


Figura 3.95: Consumo Anual

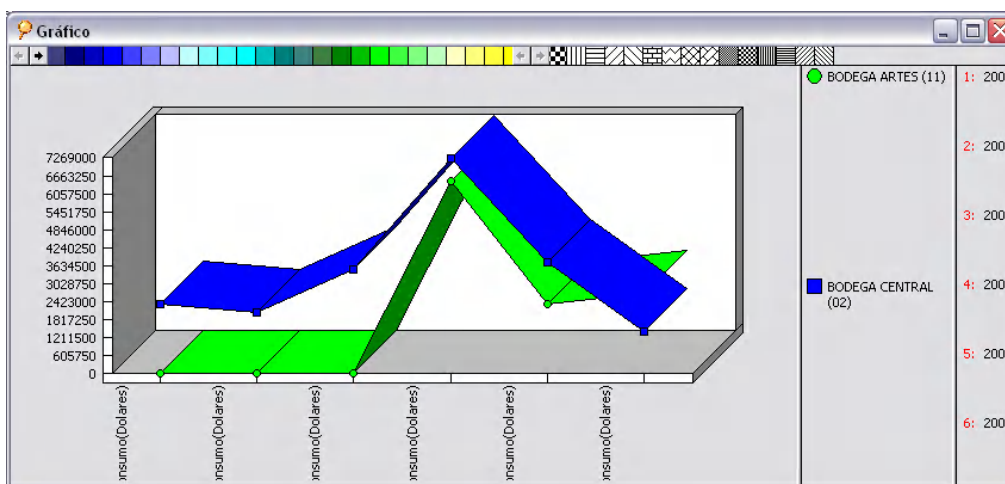


Figura 3.96: Gráfico de tendencia del Consumo Anual

Consultas

- Ingresos por proveedores.

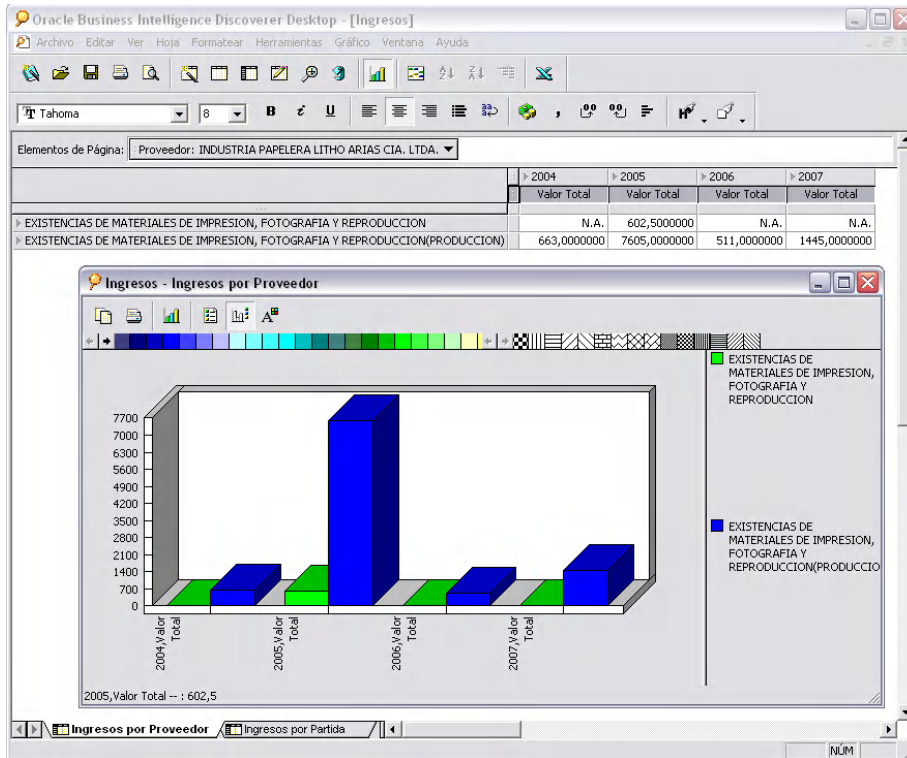


Figura 3.97: Ingresos por Proveedor con gráfico de barras

- Movimiento de ingreso.

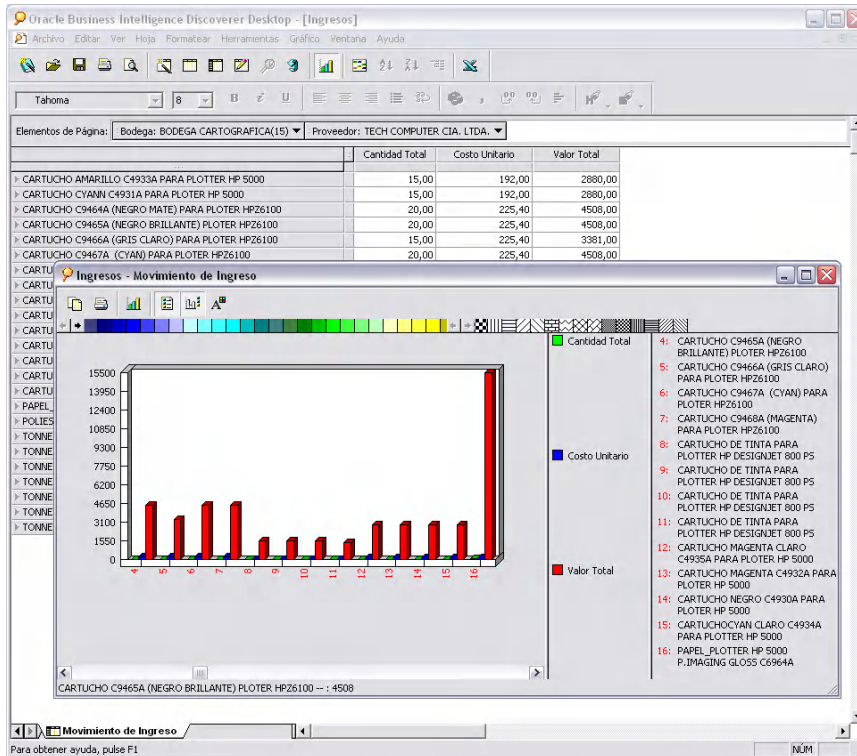


Figura 3.98: Movimiento de Ingreso con gráfico de barras

- Egresos a Bodega.

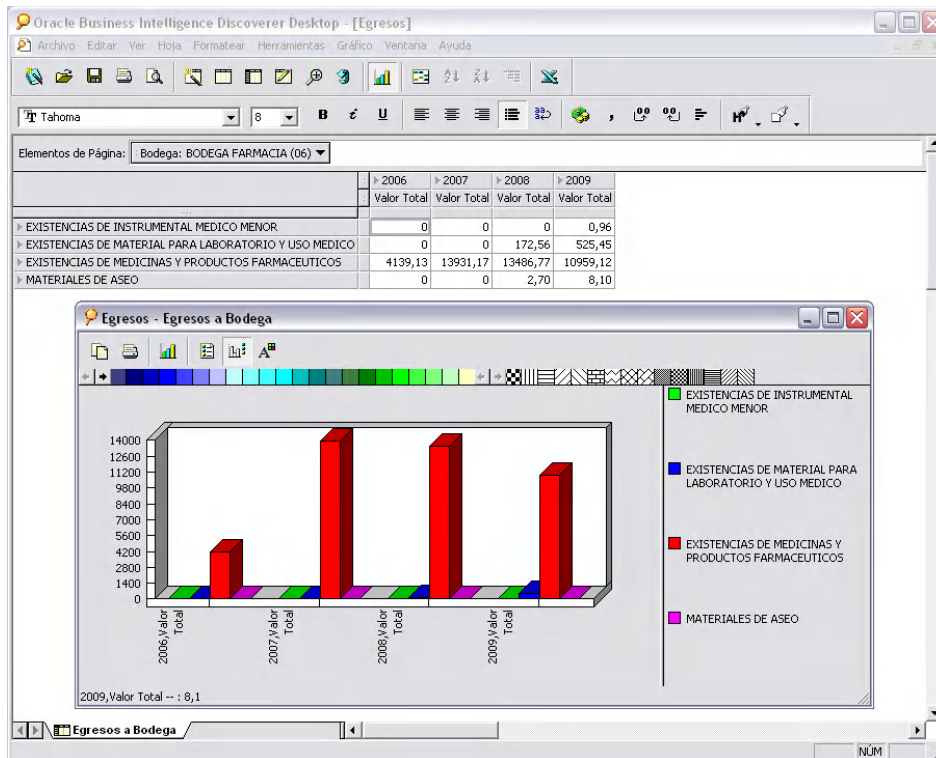


Figura 3.99: Egresos a Bodega con gráfico de barras

- Existencia

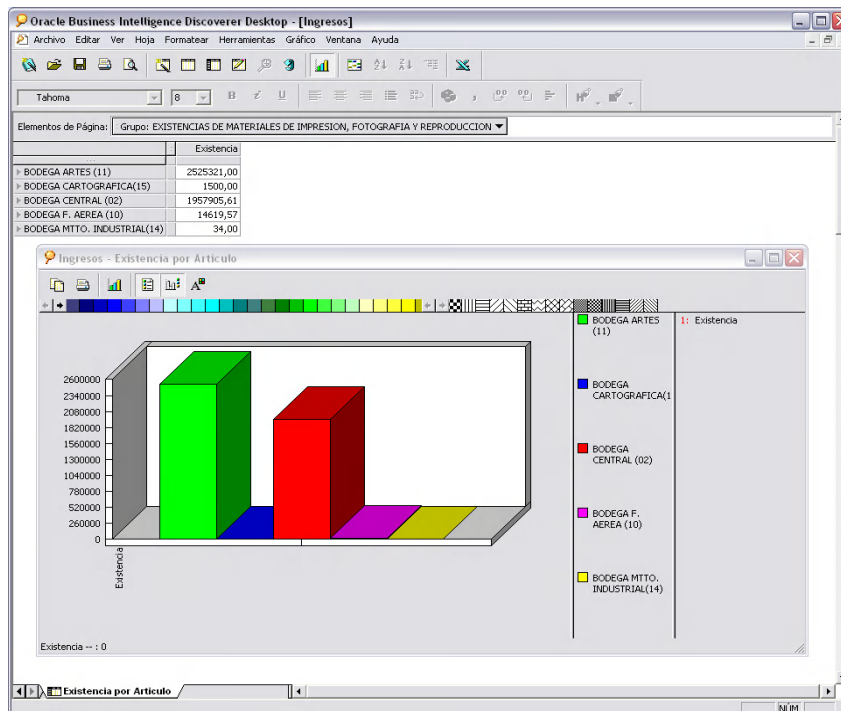


Figura 3.100: Existencia de Grupo de Artículo con gráfico de barras

- Movimiento por Artículo.

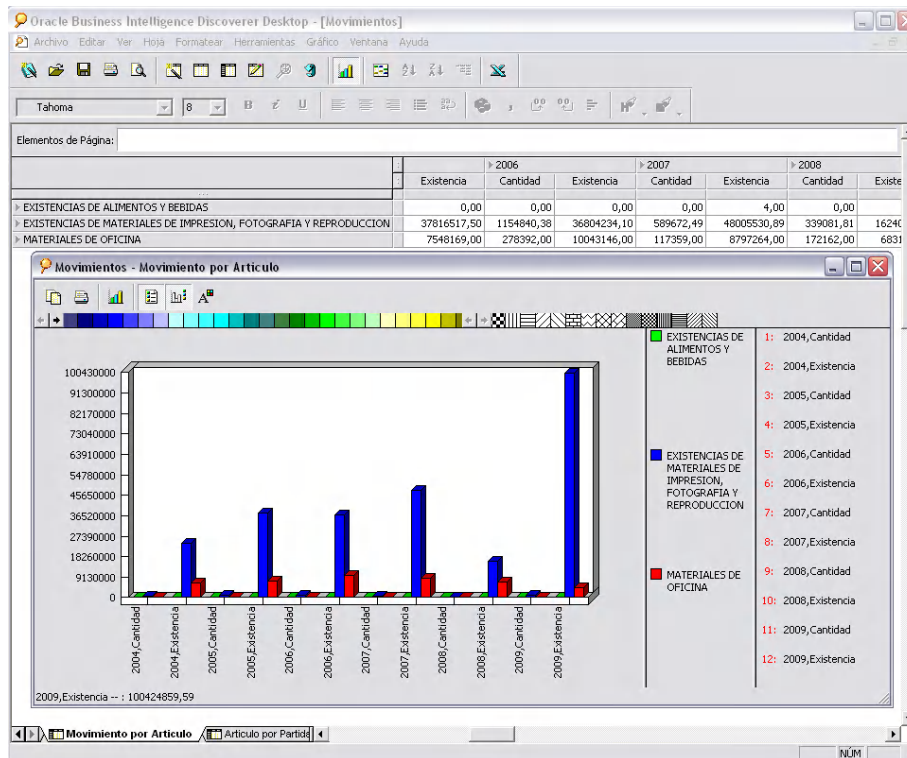


Figura 3.101: Movimientos por Grupo de Artículo con gráfico de barras

- Compras por Proveedor.

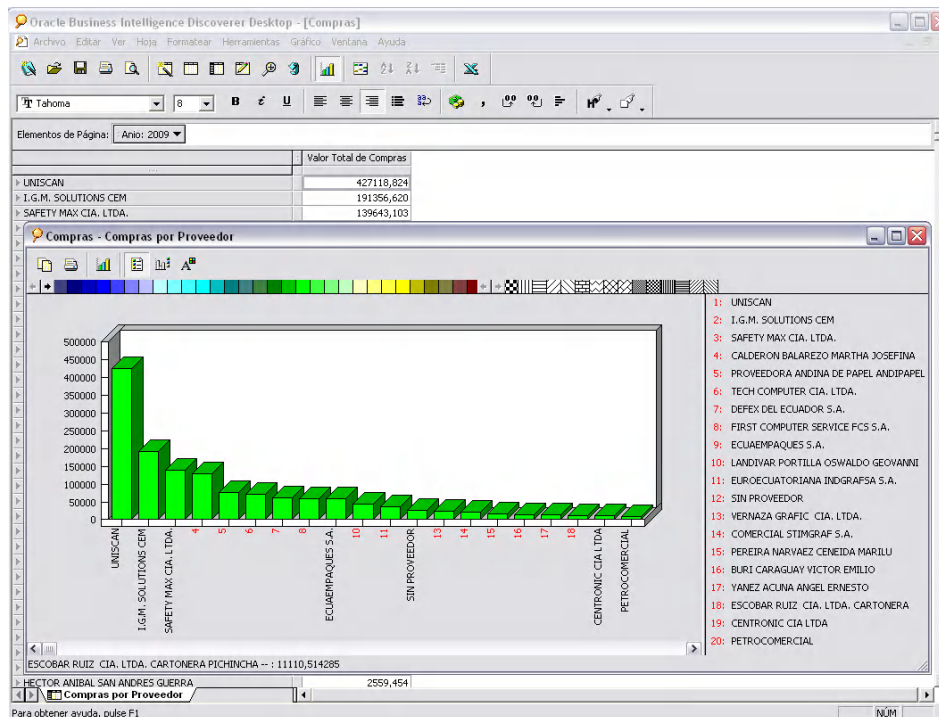


Figura 3.102: Compras por Proveedor en orden descendente con gráfico de barras

- Control de Pedidos (Pedidos por Área, Artículos y proyectos)

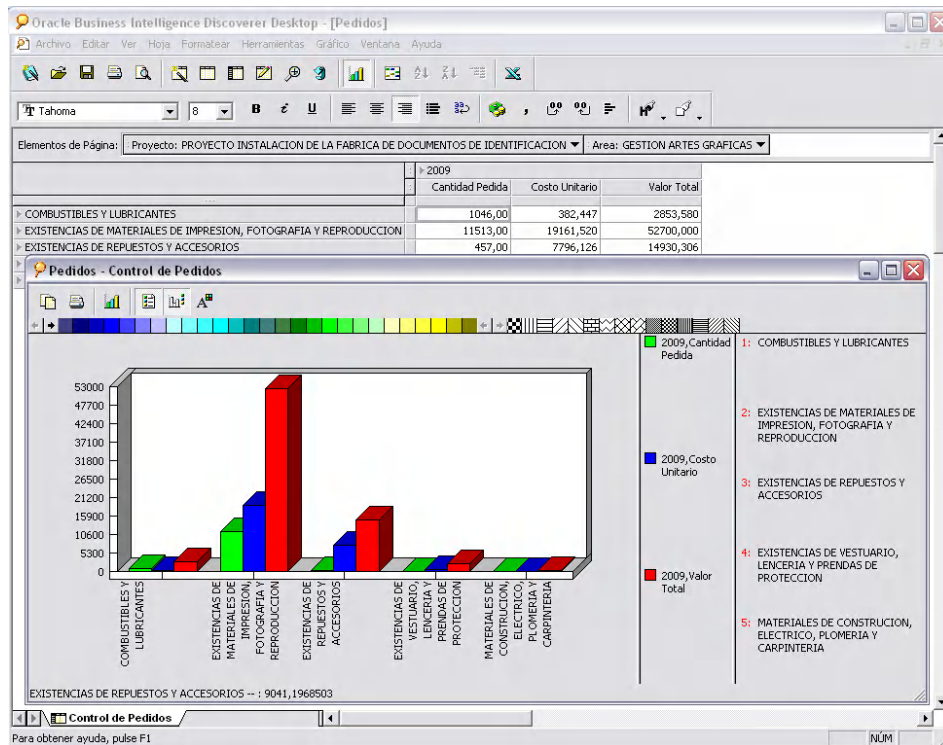


Figura 3.103: Pedidos por Grupo de Artículo, Área y Proyectos con gráfico de barras

3.10.2.2. Puntos de Venta

Indicadores

- Nivel de Cumplimiento

El Nivel de Cumplimiento está representado por la tabla de niveles a evaluar la misma que se encuentra organizado de la siguiente manera:

% de Cumplimiento	Nivel de Cumplimiento (KPI)
$\geq 90\%$	5
$\geq 80\%$ y $< 90\%$	4
$\geq 70\%$ y $< 80\%$	3
$\geq 60\%$ y $< 70\%$	2
$< 60\%$	1

Como se puede apreciar los colores representan el rango, con lo cual la consulta del Nivel de Cumplimiento quedará de la siguiente manera:

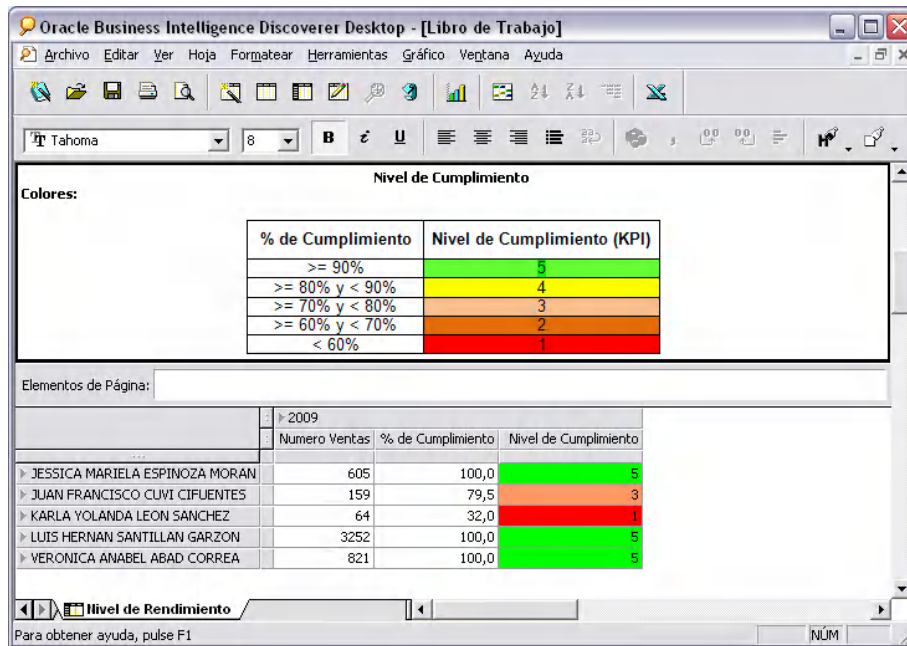


Figura 3.104: Nivel de Cumplimiento de Empleados

Consultas

Para las consultas se realizará el ejemplo de tres reportes, debido a que de acuerdo a las necesidades pueden agregar dimensiones que permitan ser evaluadas de otra manera y filtradas las medidas por otras dimensiones.

- Ventas por Sucursal.

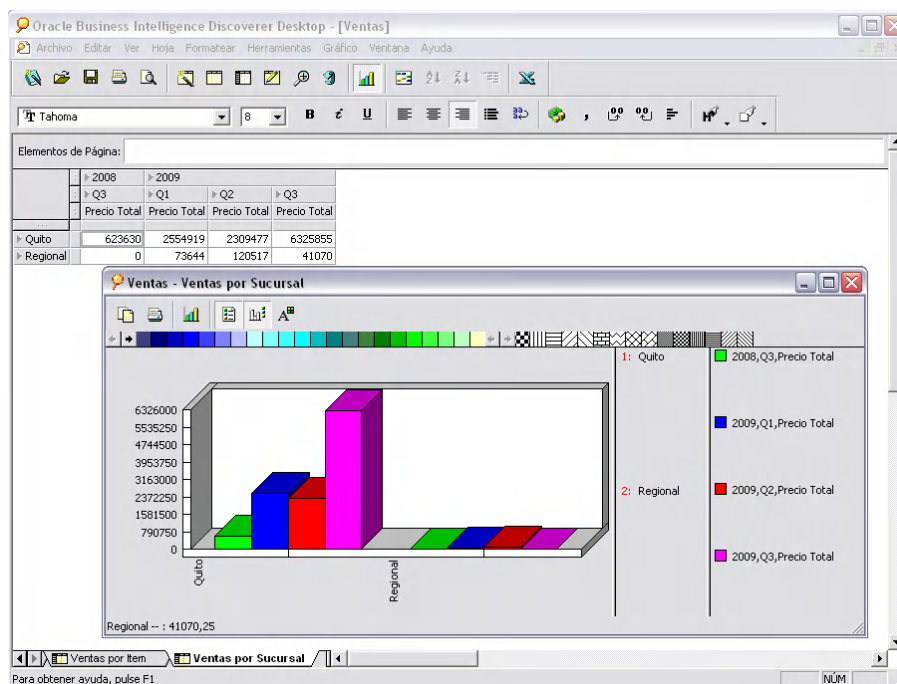


Figura 3.105: Total en Ventas por Sucursal con grafico de barras

- Cantidad vendida por cliente.

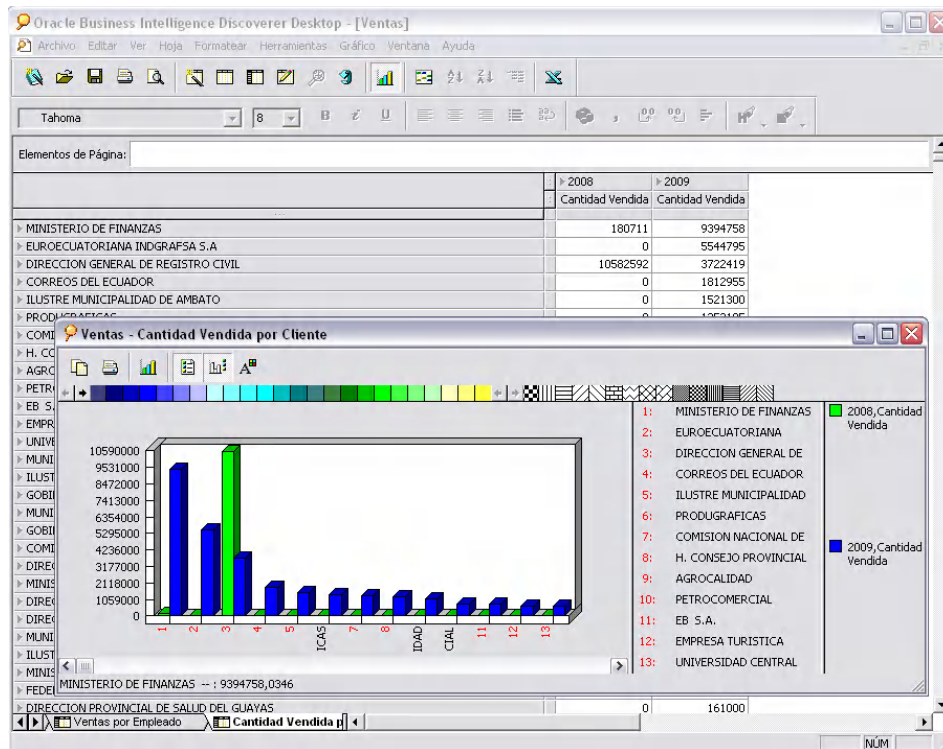


Figura 3.106: Cantidad de Ítems vendidos por Cliente con gráfico de barras

- Numero de Ventas por Sucursal.

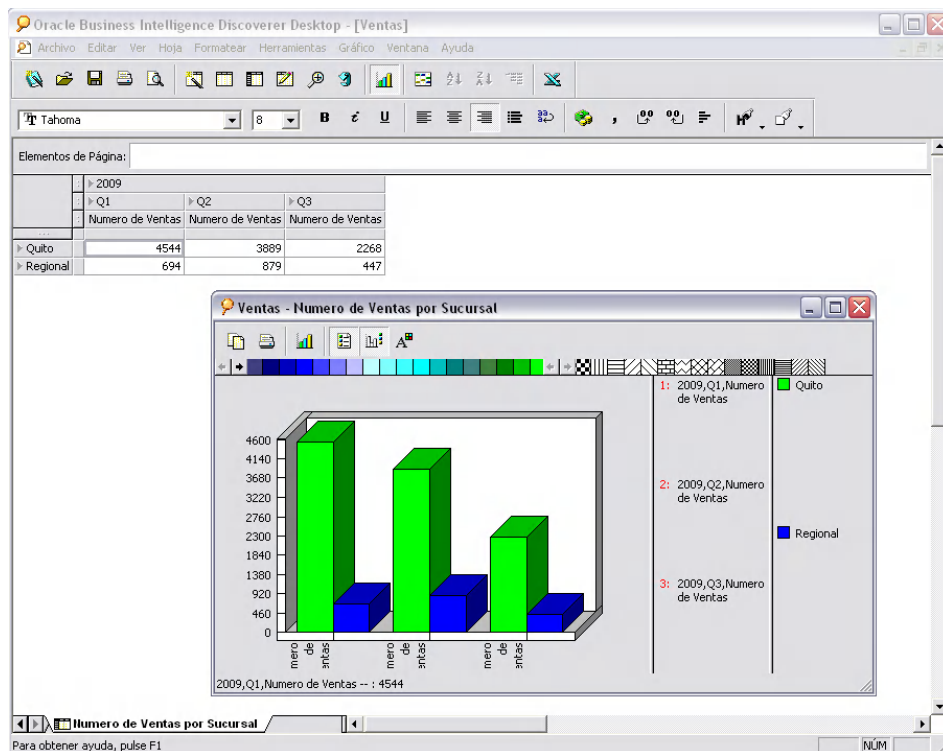


Figura 3.107: Número de Ventas por Sucursal con gráfico de barras

3.11. MANTENIMIENTO

En el Mantenimiento se enfoca en dos secciones:

- Front Room
- Back Room

3.11.1. FRONT ROOM

El mantenimiento del Back Room se refiere a la sección de presentación de los datos al usuario final.

3.11.1.1. Mantenimiento de la Aplicación

Se realiza el monitoreo semanalmente de los EUL y reportes implementados, con el fin de verificar el correcto funcionamiento en la presentación de los datos, esto se lo realiza de la siguiente manera:

- Discoverer Administrator, nos permite administrar lo referente a los EUL tanto de Inventarios como de Puntos de Venta, aspectos como por ejemplo:
 - Actualizaciones de las capas, contemplando creaciones de nuevos campos dentro de las tablas del DataWarehouse, lo que permite que se refleje dentro del EUL.
 - Creación de campos a partir del EUL, es decir independiente a los campos originales de las tablas del DataWarehouse, como la creación de un nuevo campo que especifique solamente el año, partiendo de un campo de fecha de una tabla del DataWarehouse.
 - Creación ó eliminación de niveles o jerarquías de las entidades para la navegabilidad en los datos, por ejemplo a la entidad

tiempo con jerarquía de estructura siguiente: Año -> Cuatrimestre
-> Mes – Día, se modifique a solo 2 niveles: Año -> Mes.

- Discoverer Plus ó Desktop, nos permite administrar los reportes creados, tomando en cuenta aspectos como:
 - Publicación de libros de trabajo, es decir compartir los libros creados a otros usuarios.
 - Modificación, eliminación ó creación de nuevas hojas o libros de trabajo, como sustituir en el reporte de consumo la entidad Sucursal por Bodega para analizar los datos desde otra perspectiva.

3.11.1.2. Manejo de Seguridad

Inicialmente los niveles de usuario final (EUL) están concebidos para que la parte directiva y personal de análisis pueda acceder, tanto para el negocio de Inventarios como de Puntos de Venta.

Según las necesidades, se deberá dar permisos a los empleados para que puedan tener acceso a los EUL creados.

Mediante el Discoverer Administrator, se realiza la asignación al usuario para que puedan hacer uso del EUL como muestra la Figura a continuación.

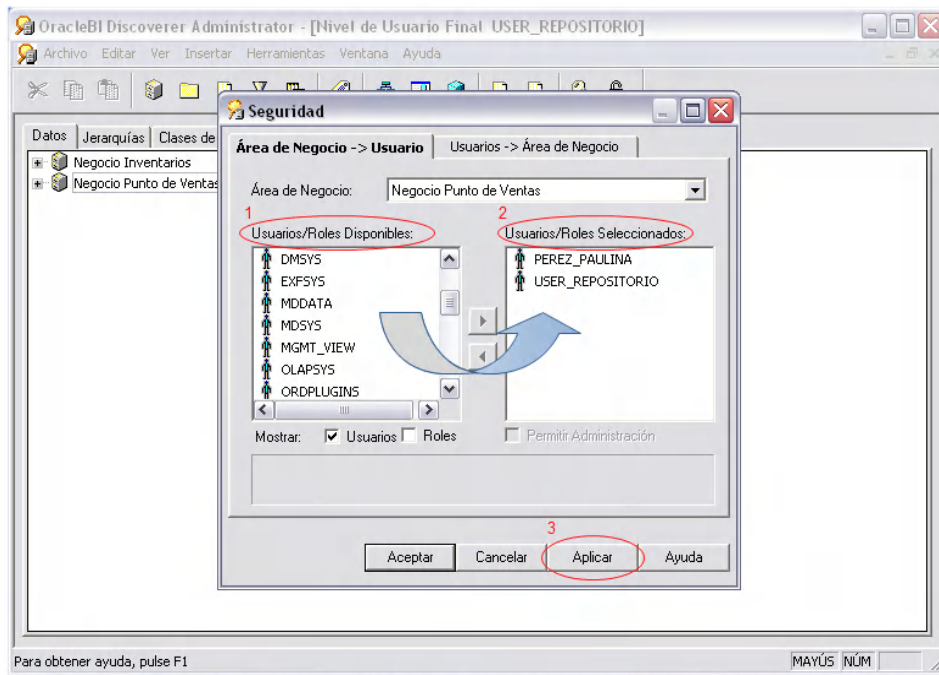


Figura 3.108: Permiso de Acceso por usuario al EUL

3.11.2. BACK ROOM

El mantenimiento del Back Room se refiere a la sección del DataWarehouse y las Fuentes Origen de los Datos.

3.11.2.1. Monitorear el Sistema ETL

Como parte del mantenimiento del Back Room, es el monitoreo del Sistema ETL con el fin de verificar como se encuentra su funcionamiento.

Se tomará en cuenta las siguientes actividades para monitorear el Sistema ETL:

- Verificación de errores mediante la Auditoria de Error que se planteo.
- Ejecución del Sistema ETL manualmente con el fin de observar el correcto funcionamiento.
- Verificación individual por cada transformación y trabajo.

- Chequeo para auditar la técnica CDC (Change Data Capture), para analizar que se esté registrando normalmente los cambios producidos en la Base de Datos Origen.

3.11.2.2. Administración de Espacio en Disco

Según la cantidad de datos que se cargan diariamente en la Base donde se encuentra el DataWarehouse, se tiene que monitorear semanalmente como está el espacio de tablas que maneja el Motor de Oracle. Por lo que de acuerdo a lo que está ocupando se debe aumentar la capacidad de almacenamiento.

3.11.2.3. Respaldo y Recuperación

Tomando en cuenta que el DataWarehouse no llevará un histórico de los datos del negocio, sino más bien datos actualizados, no es importante tomar medidas para respaldar la base de datos, ya que en alguna circunstancia en la que se pierda los datos, o se dañe el servidor, se deberá realizar una carga inicial de datos, lo cual garantiza que se recupere los datos sin ningún problema.

Un punto que se debe tomar en cuenta es el respaldo de los metadatos generados por la herramienta ETL Pentaho Data Integration, ya que la misma almacena todas las transformaciones y trabajos en la Base de Datos, lo cual es de suma importancia mantener los respaldos de estos metadatos. La manera más óptima de obtener estos paquetes es exportarlos como archivos con extensión KJB para los trabajos, y con extensión KTR para las transformaciones, como se muestra en el gráfico a continuación:

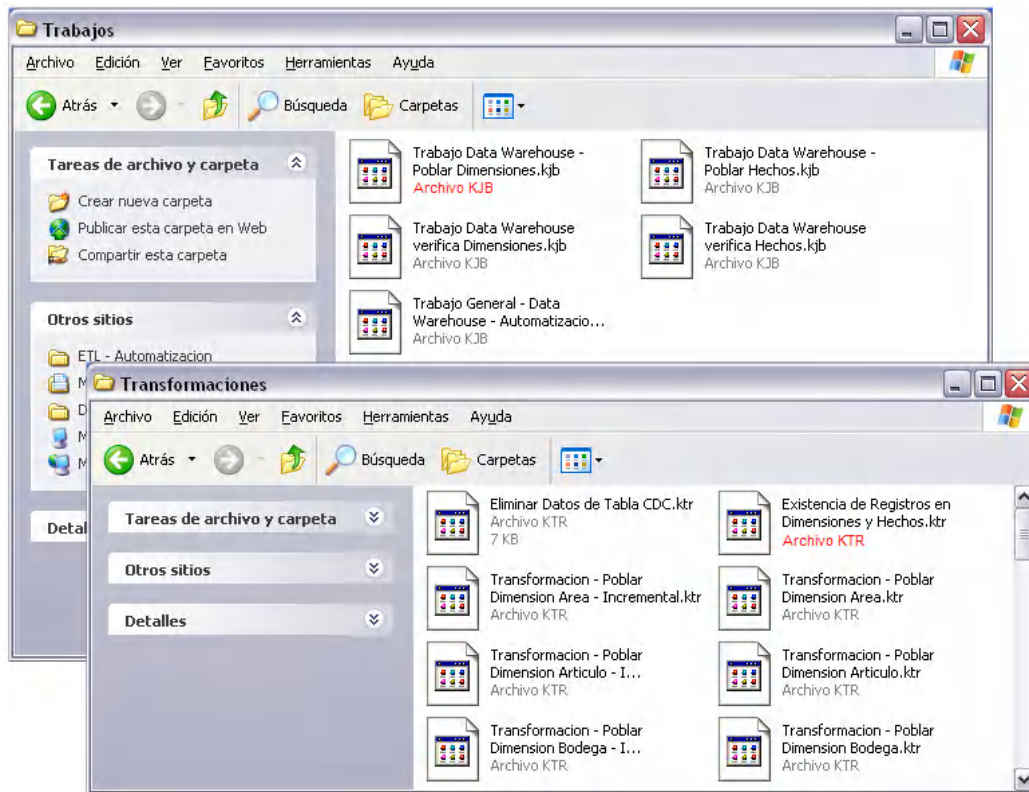


Figura 3.109: Respaldo de Transformaciones y Trabajos como Archivos *.ktr y *.kjb

3.12. CRECIMIENTO

La sección de crecimiento del DataWarehouse no se realizó para el proyecto debido a que el desarrollo del mismo es inicial, por lo cual no se implementó el crecimiento al Sistema. Pero se tratará esta sección con el fin de tener una idea clara para realizar aumentos o evoluciones futuras para el Proyecto.

Para llevar a cabo un aumento o crecimiento del DataWarehouse es importante tomar en cuenta puntos iniciales como:

- Comunicación constante con los usuarios de negocio.
- Evaluación de la satisfacción manifestada por los Usuarios de negocio.
- Supervisar el éxito y el servicio de medición.

Tomados en cuenta los puntos antes mencionados se puede empezar por preparar el crecimiento del DataWarehouse.

3.12.1. EVALUAR EL ENTORNO ACTUAL

Previo a al crecimiento o evolución, se debe primero evaluar el ambiente que se encuentra en ejecución con el fin de poder analizar si soporta aun o no las necesidades de los usuarios de Negocio.

Se evaluará puntos como:

- Datos consistentes y correctos almacenados en el DataWarehouse.
- Actualización de datos de manera oportuna.
- Fácil acceso por parte de los usuarios a Aplicaciones de BI.
- Datos organizados dimensionalmente.

3.12.2. PRIORIZAR OPORTUNIDADES PARA EL CRECIMIENTO

3.12.2.1. PRIORIDAD A PEQUEÑAS MEJORAS

Las pequeñas mejoras son orientadas sobre lo que se encuentra implementado, como por ejemplo el aumento de atributos a Dimensiones, o nuevos cálculos para los Hechos, ó adicionar aplicación de BI basados en la existencia actual de los datos en el almacén.

Esto significa que se debe dar importancia a las pequeñas modificaciones que se pretenda implementar sobre las Dimensiones ó Hechos ya sean estos de Inventarios o Punto de Ventas.

3.12.2.2. PRIORIDAD A MAYORES INICIATIVAS

Se enfoca en el cambio de datos del negocio que no se encuentran representados en el DataWarehouse, es decir, por ejemplo agregar nuevos hechos o Dimensiones representativos del negocio, o agregar al almacén nuevos Hechos y Dimensiones de Procesos que no consten en el

DataWarehouse, como por ejemplo incluir el análisis de ordenes de producción que se genera después de una compra para productos no inventariados.

3.12.3. GESTIONAR EL CRECIMIENTO ITERATIVO

Tomando en cuenta el tipo de cambio a implementar en el Sistema del DataWarehouse, ya sea este pequeños cambios, o cambios que impliquen el agregar análisis de nuevos Procesos, se debe realizar nuevamente el ciclo de desarrollo e implantación del DataWarehouse, con el fin de garantizar el correcto diseño y producción para poner en marcha las nuevas necesidades de negocio, junto con lo que se encuentre en funcionamiento.

CAPITULO IV: CONCLUSIONES Y RECOMENDACIONES

- La implementación de un DataWarehouse con alcance para Puntos de Venta e Inventarios, generó información analítica dentro de un nivel organizacional táctico – operativo, producto de los requerimientos de directivos de cada negocio. Se recomienda como alcance para los desarrollos futuros del DataWarehouse, trabajar en conjunto con las TIC y las áreas, con el fin de visionar y conformar indicadores representativos, que sean objeto de medición dentro de cada negocio.
- Se cumplió con el objetivo de utilizar la metodología propuesta por Ralph Kimball como guía para el proyecto, la misma que ayudó para planificar por etapas el desarrollo, permitiendo la administración y mantenimiento durante todo el ciclo de vida, empezando desde la planificación hasta el crecimiento futuro de nuevas mejoras o aumento del alcance del DataWarehouse. Por lo que es recomendable para el desarrollo de un proyecto de DataWarehouse y/o Bussiness Intelligence, guiarse por metodologías que colaboren durante fases de desarrollo, y que permitan llevar a cabo de manera organizada todas las etapas.
- Como parte del objetivo se cumplió con la utilización de herramientas Oracle adquiridas por el Instituto Geográfico Militar, el Motor de Base de Datos Oracle 10g para el almacenamiento de datos, en conjunto con Oracle Discoverer Plus 9 para la interfaz con el usuario, lo que sirvió para la integración en la presentación de los datos, pero con poca funcionalidad por parte del Discoverer, en el sentido de cómo ayuda a

visualizar la información al usuario, estando limitado en presentaciones gráficas, y de la ayuda que genera, más aun cuando se trata de análisis de indicadores ya sean estos por medio de medidores (dashboard). Por lo tanto, tomando en cuenta este aspecto a considerar, se recomienda implementar la presentación de indicadores con herramientas que visualicen gráficos que ayuden a interpretar el estado de los mismos para un mejor análisis.

- Se utilizó la herramienta Open Source Pentaho Data Integration, cumpliendo con el objetivo propuesto. El software libre mencionado fue intuitivo y accesible de comprender, así como su funcionamiento y estabilidad en la integración de datos. Sin embargo se tuvo un inconveniente en la actualización periódica entre las fuentes origen y el DataWarehouse, la misma que se solucionó con técnicas implementadas en el desarrollo del proyecto. Tomando en cuenta esto, se recomienda que para el desarrollo de un proyecto de DataWarehouse se evalúe la funcionalidad en conjunto del motor de Base de Datos con la herramienta ETL, con el fin de integrar el sistema ETL sin la necesidad de suplir falencias entre las herramientas, como la captura de datos con cambio para las cargas incrementales.
- El modelo Dimensional del DataWarehouse permite una mejor funcionalidad en las consultas, debido a que modela el negocio de importancia, almacenando datos relevantes para su análisis.

- Para el desarrollo del DataWarehouse fue indispensable contar con el apoyo tanto del área de informática, los cuales brindan el soporte tecnológico, así como del área referente al negocio con el fin de evaluar en conjunto las necesidades, y ser ente activo durante todas las etapas de desarrollo. Por lo que es recomendable contar con el apoyo técnico y de negocio, con el fin de evitar mal interpretaciones en las necesidades, y que las mismas sean reflejadas en el diseño del DataWarehouse.
- En el diseño del modelamiento dimensional, es importante identificar el nivel de detalle o granularidad hasta el que va llegar el análisis, esto permite realizar diseños que lo soporte, y que garantice la integridad de datos durante el proceso ETL. Por lo tanto se recomienda tomar en cuenta la granularidad con el fin de determinar técnicamente, como obtener los datos de las fuentes, integrarlos y sin que esto modifique la relación de los mismos cuando se encuentren en el ambiente de DataWarehouse.
- Es importante implementar técnicas de extracción de datos, esto ayuda al desempeño del proceso ETL y al ahorro de tiempo y recursos, por lo que se recomienda poner en práctica técnicas como es el caso de la captura de datos con cambio (CDC – Change Data Capture) dentro de la carga incremental, lo que permite extraer únicamente los datos nuevos y/o que hayan cambiado en las fuentes origen.

- Un DataWarehouse es más que el desarrollo de un producto, es un concepto, debido a que su impacto de alcance no es solamente generar información para la toma de decisiones, además que intangiblemente ayuda a las personas que se encuentran en el mundo de BI a tener una mayor visión sobre los procesos y negocio, desde una perspectiva gerencial, incluso ayuda a madurar los procesos con el fin de ser una base sólida para el desarrollo del DataWarehouse. Por lo que es aconsejable no subutilizar el desarrollo de un sistema de este tipo únicamente para presentar datos sin ningún fin específico de análisis, puesto que el desarrollo e implementación lleva tiempo e inversión.
- La importancia de entender cómo se evidencia el negocio en los sistemas transaccionales como son las Bases de Datos es fundamental, puesto que ayuda a tener mejor visión sobre los datos y permitir seleccionar los adecuados para el análisis de la información, después que hayan sido integrados por medio del proceso ETL. Por lo tanto es importante llevar a cabo un estudio del negocio y de las fuentes de datos que lo respalda, con el fin de sustentar un origen de calidad para la integración de datos.
- Para el desarrollo e implementación de una solución de Business Intelligence o DataWarehouse es recomendable contar con un grupo de personas por cada etapa de desarrollo, es decir un grupo encargado de recopilar requerimientos, otro sobre el Diseño y Modelamiento, el desarrollo ETL, y así las demás etapas. Esto garantiza que para cada

una de ellas se realice un estudio a profundidad, por lo que es importante identificar las etapas y las personas encargadas para intervenir en ellas.

- En el Instituto Geográfico Militar existe un grado de dificultad al implementar soluciones de DataWarehouse debido a que no cuentan con procesos maduros, por lo que los sistemas en general deben ajustarse a nuevas políticas que surgen dentro de la institución, esto es un gran problema y más aun si se trata de un sistema de DataWarehouse que se basa en el análisis de información, sobre datos mayormente recolectados de bases de datos transaccionales estables, y que se ven sujetas a cambios constantes en su diseño y estructura para soportar las nuevas políticas. Por lo tanto, tomando en cuenta este aspecto que es importante, es recomendable que se trabaje sobre procesos estables, o en su defecto madurar los procesos lo que garantice una estructura de negocio fuerte, sobre los que se puede sustentar un de sistemas de información. De igual manera se debe hacer partícipe de estos inconvenientes a la parte directiva, acerca de los problemas que conlleva el cambio de políticas con que se manejan los procesos.
- Dentro del Instituto no existe un ambiente gerencial por parte de la dirección de negocio, y más bien se encuentran a nivel operativo – táctico, por lo que es recomendable empezar un desarrollo de DWH/BI departamentales, y con un alcance como punto de partida medidas de negocio, y avanzar en su evolución con indicadores en función de

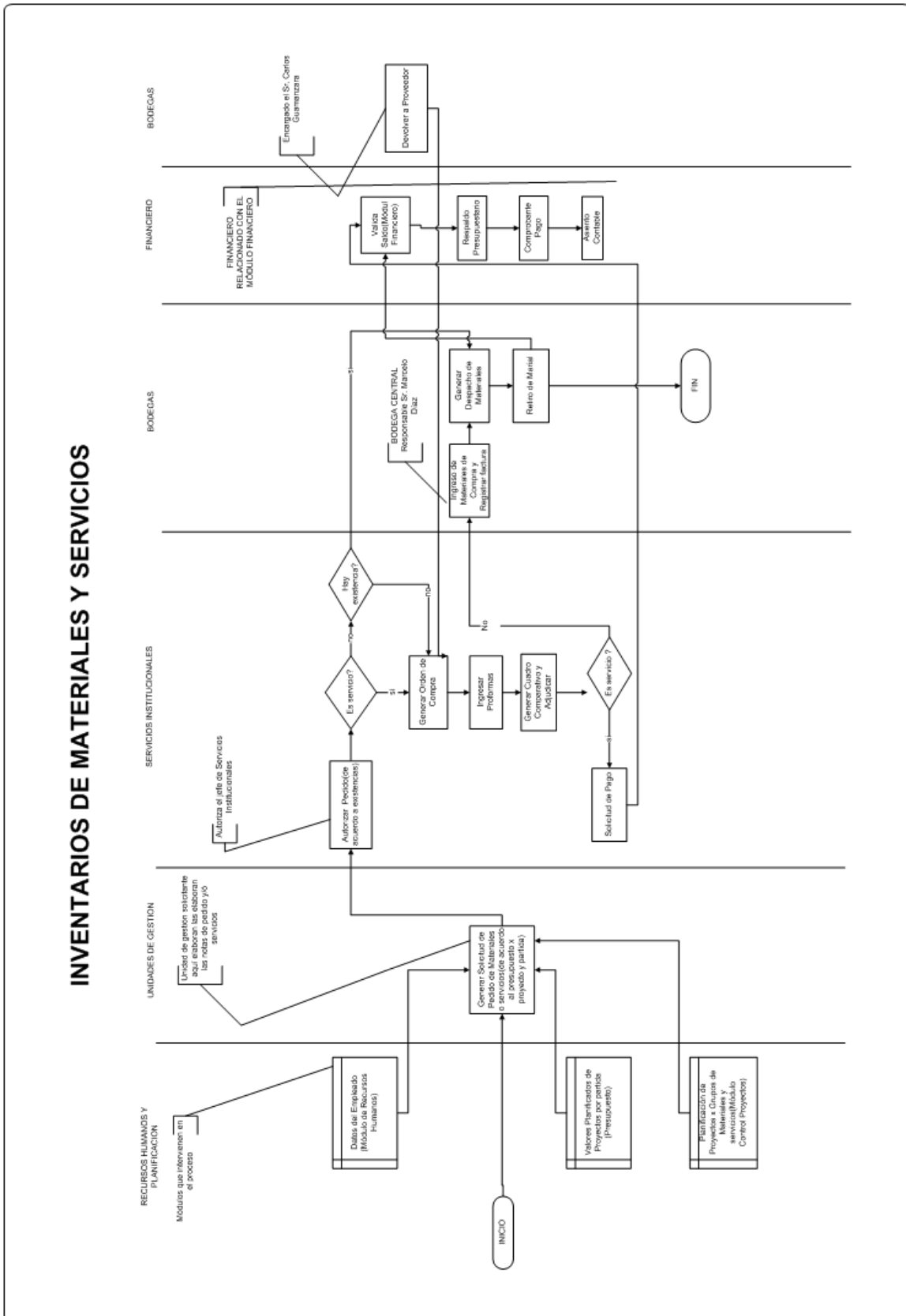
objetivos de la Institución, lo cual permita ir paralelamente la formación del empleado con visión gerencial, en conjunto con las aplicaciones informáticas que respalden esta información.

CAPÍTULO V: BIBLIOGRAFÍA Y FUENTES DE CONSULTA

- ✓ Kimball Ralph. "The Data Warehouse Lifecycle Toolkit", Second Edition.
- ✓ Bartle Phil PhD. Traducción de M^a Lourdes Sada. "Información para la Gestión y gestión de la información" [en línea]: [Fecha de consulta: 20 Febrero 2010], Disponible en: < <http://www.scn.org/mpfc/modules/mon-miss.htm> >.
- ✓ Wolff Carmen Gloria. "Modelamiento Multidimensional" [en línea]: [Fecha de consulta: 20 Septiembre 2009], Disponible en: < <http://www.inf.udec.cl/~revista/ediciones/edicion4/modmulti.PDF> >.
- ✓ MicroStrategy LATAM. "Scorecards y Tableros de Control" [en línea]: [Fecha de consulta: 22 Septiembre 2009], Disponible en: < http://latam.microstrategy.com/Solutions/5Styles/scorecards_dashboards.asp >.
- ✓ Sinnexus. "Bases de datos OLTP y OLAP" [en línea]: [Fecha de consulta: 25 Septiembre 2009], Disponible en: < http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx >.
- ✓ Cantos Henry, Cordero Diego, Larco Ciro, Méndez Alí, Echeverría Pedro. "Diseño E Implementación De Un Data Warehouse Para Las Empresas De Distribución Y Comercialización De Agua Potable" [en línea]: [Fecha de consulta: 21 Febrero 2010], Disponible en: <<http://www.dspace.espol.edu.ec/handle/123456789/6116>>
- ✓ MicroStrategy LATAM. "Scorecards y Tableros de Control" [en línea]: [Fecha de consulta: 22 Septiembre 2009], Disponible en: < http://latam.microstrategy.com/Solutions/5Styles/scorecards_dashboards.asp >.
- ✓ Gartner, Inc. "Glossary" [en línea]: [Fecha de consulta: 08 Octubre 2009], Disponible en: < http://www.gartner.com/6_help/glossary/Gartner_IT_Glossary.pdf>.
- ✓ Sinnexus. "Bases de datos OLTP y OLAP" [en línea]: [Fecha de consulta: 08 Octubre 2009], Disponible en: < http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx >.
- ✓ Pentaho Community, Pentaho Data Integration (Spanish), Doug Moran. "Documentación para Usuarios y Desarrolladores" [en línea]: [Fecha de consulta: 16 Agosto 2009], Disponible en: < <http://wiki.pentaho.com/display/EALes/Inicio> >.
- ✓ Dinesh Asanka. "CDC and Data Warehouse" [en línea]: [Fecha de consulta: 28 Octubre 2009], Disponible en: < http://www.sql-server-performance.com/articles/biz/CDC_and_Data_Warehouse_p1.aspx >.
- ✓ Talend. "Change Data Capture explained" [en línea]: [Fecha de consulta: 28 Octubre 2009], Disponible en: < <http://www.talend.com/blog/2008/06/16/change-data-capture-explained/> >.

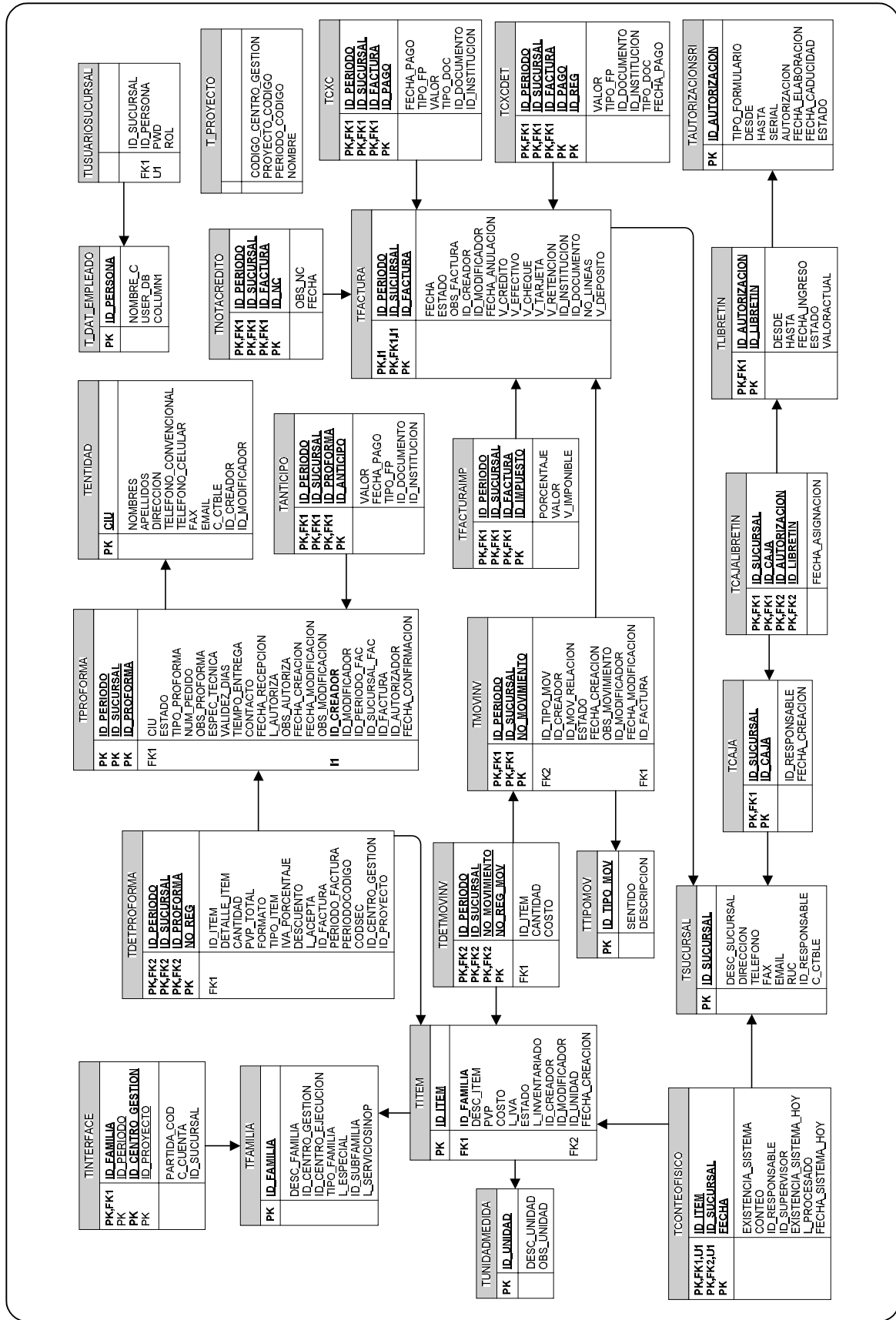
ANEXOS

ANEXO A – Flujo de Proceso de Inventarios



Flujo de Proceso de Inventarios
Referencia: Instituto Geográfico Militar

ANEXO C – Esquema de Puntos de Venta



Modelo Entidad Relación - Esquema Puntos de Venta - PTV

ANEXO D – Tablas Esquemas: Financiero, Proyectos, Recursos Humanos

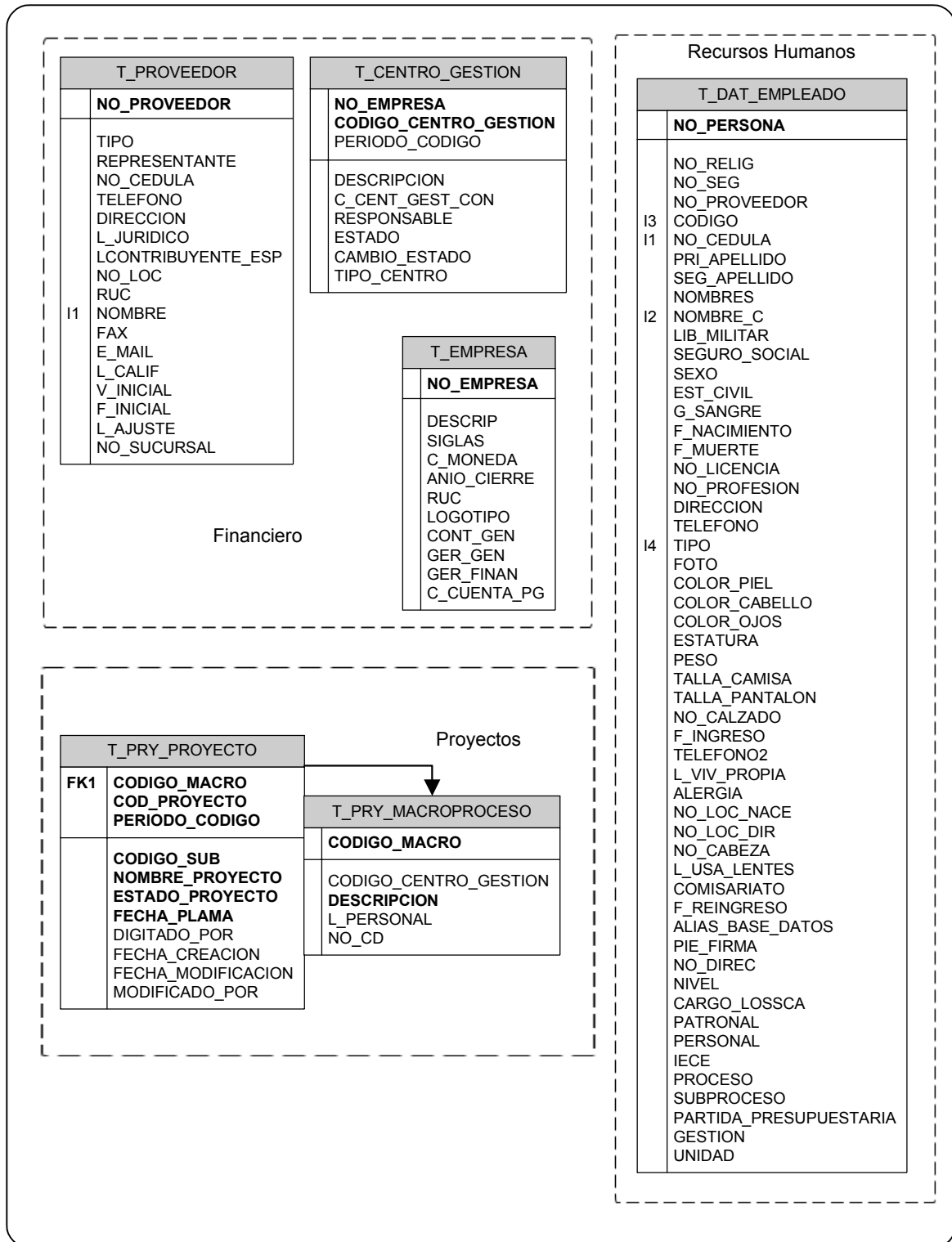


Diagrama Relacional de Tablas a utilizar de los Esquemas: Financiero – FINANCIERO,

Proyectos – PROYECTOS, Recursos Humanos - RH

ANEXO E - Scripts Base de Datos Origen

Se encuentra en el CD adjunto al Proyecto bajo la carpeta Anexos.

ANEXO F - Scripts – DataWarehouse

Se encuentra en el CD adjunto al Proyecto bajo la carpeta Anexos.

ANEXO G - Scripts – Extracción de Datos Origen

Se encuentra en el CD adjunto al Proyecto bajo la carpeta Anexos.

ANEXO H - Scripts – Extracción de Datos Destino para Carga Incremental

Se encuentra en el CD adjunto al Proyecto bajo la carpeta Anexos.

ANEXO I - Diccionario de Datos

Tabla de Auditoria CDC (Changed Data Captured)

Tabla: AUDIT_CDC

Descripción: Tabla que registra los cambios que hayan sufrido las tablas Origen.

Campo	Tipo de Dato	Descripción
ID	INTEGER	Clave Primaria de la Tabla
TABLA_CAMBIO	VARCHAR2(30)	Nombre de la tabla Origen q a cambiado.
ID_TABLA_CAMBIO	VARCHAR2(15)	1era. Clave primaria de la tabla Origen.
ID2_TABLA_CAMBIO	INTEGER	2da. Clave primaria de la tabla Origen.
ID3_TABLA_CAMBIO	INTEGER	3era. Clave primaria de la tabla Origen.
FECHA_CAMBIO	INTEGER	Fecha en la que se produjo el cambio.

Tabla de Auditoría para Errores – Log de Error

Tabla: LOG_ERROR

Descripción: Tabla que registra los errores que se pueden producir en el Proceso ETL.

Campo	Tipo de Dato	Descripción
ID	INTEGER	Clave Primaria de la Tabla
NOMBRE_TRABAJO	VARCHAR2(50)	Nombre del Trabajo en el que se produjo el Error.
DESCRIPCION_ERROR	VARCHAR2(100)	Descripción del Error, especificando la Tabla.
FECHA_ERROR	DATE	Fecha en la que se produjo el Error.

Tablas de Dimensiones y Hechos

Tabla: DIM_AREA

Descripción: Tabla de Dimensión que contiene las Áreas del Instituto.

Campo	Tipo de Dato	Descripción
KEY_AREA	INTEGER	Clave Subrogada Primaria de la Tabla
NO_AREA	INTEGER	Numero y Clave primaria Origen de la Área.
NOMBRE_AREA	VARCHAR2(100)	Nombre del Área.

Tabla: DIM_ARTICULO

Descripción: Tabla de Dimensión que contiene los Artículos.

Campo	Tipo de Dato	Descripción
KEY_ARTICULO	INTEGER	Clave Subrogada Primaria de la Tabla
NO_ARTICULO	INTEGER	Numero y Clave primaria Origen del Artículo. (Clave primaria compuesta - Origen)
NOMBRE_ARTICULO	VARCHAR2(100)	Nombre del Artículo.
NO_GRUPO	INTEGER	Numero del Grupo de Artículo.
NOMBRE_GRUPO	VARCHAR2(80)	Nombre del Grupo del Artículo.
NO_SUBGRUPO	INTEGER	Numero del Sub Grupo

		de Artículo. (Clave primaria compuesta - Origen)
NOMBRE_SUBGRUPO	VARCHAR2(80)	Nombre del Subgrupo de Artículo.
PARTIDA	VARCHAR2(30)	Numero de Partida del Grupo de Artículo.
VALOR_ULTIMA_COMPRA	NUMBER(15,7)	Valor de la Última Compra del Artículo.
FECHA_ULTIMA_COMPRA	DATE	Fecha en que se realizó la Última compra del Artículo.
COSTO	NUMBER(15,7)	Costo promedio del Artículo.
UNIDAD_MEDIDA	VARCHAR2(30)	Unidad de Medida del Artículo.

Tabla: DIM_BODEGA

Descripción: Tabla de Dimensión que contiene las Bodegas

Campo	Tipo de Dato	Descripción
KEY_BODEGA	INTEGER	Clave Subrogada Primaria de la Tabla
NO_BODEGA	INTEGER	Numero y Clave primaria Origen de la Bodega.
NOMBRE_BODEGA	VARCHAR2(100)	Nombre de la Bodega.

Tabla: DIM_CENTRO_GESTION

Descripción: Tabla de Dimensión que contiene Los Centros de Gestión.

Campo	Tipo de Dato	Descripción
KEY_CENTRO_GESTION	INTEGER	Clave Subrogada Primaria de la Tabla
NO_CENTRO_GESTION	INTEGER	Numero y Clave primaria Origen del Centro de Gestión.
NOMBRE_CENTRO_GESTION	VARCHAR2(100)	Nombre del Centro de Gestión.

NO_EMPRESA	INTEGER	Numero y Clave primaria Origen de la Empresa.
NOMBRE_EMPRESA	VARCHAR2(100)	Nombre de la Empresa.

Tabla: DIM_CLIENTE

Descripción: Tabla de Dimensión que contiene los Clientes.

Campo	Tipo de Dato	Descripción
KEY_CLIENTE	INTEGER	Clave Subrogada Primaria de la Tabla
CEDULA_RUC_CLIENTE	VARCHAR2(15)	Cedula o Ruc del Cliente.
NOMBRE_CLIENTE	VARCHAR2(150)	Nombre del Cliente.
DIRECCION_CLIENTE	VARCHAR2(100)	Dirección del Cliente.

Tabla: DIM_EMPLEADO

Descripción: Tabla de Dimensión que contiene los Empleados

Campo	Tipo de Dato	Descripción
KEY_CLIENTE	INTEGER	Clave Subrogada Primaria de la Tabla
NO_EMPLEADO	INTEGER	Número y Clave Primaria Origen del Empleado.
CEDULA_EMPLEADO	VARCHAR2(10)	Cedula del Empleado.
NOMBRE_EMPLEADO	VARCHAR2(100)	Nombre del Empleado.

Tabla: DIM_FACTURA

Descripción: Tabla de Dimensión que contiene las Facturas de la Venta.

Campo	Tipo de Dato	Descripción
KEY_FACTURA	INTEGER	Clave Subrogada Primaria de la Tabla
NO_FACTURA	INTEGER	Número y Clave

		Primaria compuesta Origen de la Factura.
ESTADO_FACTURA	VARCHAR2(15)	Estado de la Factura.
PERIODO_FACTURA	INTEGER	Periodo en que se produjo la factura, y clave primaria compuesta Origen.

Tabla: DIM_ITEM

Descripción: Tabla de Dimensión que contiene los Ítems de Venta.

Campo	Tipo de Dato	Descripción
KEY_ITEM	INTEGER	Clave Subrogada Primaria de la Tabla
ID_FAMILIA	INTEGER	Clave Subrogada de la Familia de Ítem, y Clave Primaria de la Familia del Ítem.
NOMBRE_FAMILIA	VARCHAR2(100)	Nombre de la Familia del Ítem.
ID_ITEM	INTEGER	Número y Clave Primaria Origen del Ítem.
NOMBRE_ITEM	VARCHAR2(120)	Nombre del Ítem.
IVA_ITEM	VARCHAR2(2)	Especifica si el Ítem tiene o no IVA.
ESTADO_ITEM	VARCHAR2(15)	Estado del Ítem.
ITEM_INVENTARIADO	VARCHAR2(2)	Especifica si el Ítem está o no Inventariado.
UNIDAD_MEDIDA	VARCHAR2(15)	Unidad de Medida del Ítem.

Tabla: DIM_MOVIMIENTO

Descripción: Tabla de Dimensión que contiene los Movimientos de Bodega.

Campo	Tipo de Dato	Descripción
KEY_MOVIMIENTO	INTEGER	Clave Subrogada Primaria de la Tabla.
NO_MOVIMIENTO	INTEGER	Número y Clave Primaria compuesta

		Origen del Movimiento.
TIPO_MOVIMIENTO	VARCHAR2(30)	Tipo de Movimiento.
NO_ORDINAL	INTEGER	Numero de Ordinal por cada Ítem, registro por ítem.
PERIODO_MOVIMIENTO	INTEGER	Periodo del Movimiento, y clave primaria compuesta origen.

Tabla: DIM_ORDEN_COMPRA

Descripción: Tabla de Dimensión que contiene las Órdenes de Compra.

Campo	Tipo de Dato	Descripción
KEY_ORDEN_COMPRA	INTEGER	Clave Subrogada Primaria de la Tabla
NO_ORDEN	INTEGER	Número y Clave Primaria compuesta Origen de la Orden de Compra.
ESTADO_ORDEN	VARCHAR2(15)	Estado de la Orden de Compra.
NO_FACTURA	INTEGER	Numero de la Factura de Compra.
PERIODO_ORDEN	INTEGER	Periodo en que se emitió la Orden de Copra, y clave primaria compuesta Origen.

Tabla: DIM_PEDIDO

Descripción: Tabla de Dimensión que contiene los Pedidos

Campo	Tipo de Dato	Descripción
KEY_PEDIDO	INTEGER	Clave Subrogada Primaria de la Tabla.
NO_PEDIDO	INTEGER	Número y Clave Primaria compuesta Origen del Pedido.
TIPO_PEDIDO	VARCHAR2(30)	Tipo de Pedido.
ESTADO_PEDIDO	VARCHAR2(30)	Estado en el que se

		Encuentra el Pedido.
IMPORTACION_PEDIDO	VARCHAR2(2)	Especifica si el Pedido que se realizo es o no de Importación.
PERIODO_PEDIDO	INTEGER	Periodo del Pedido, y clave primaria compuesta origen.

Tabla: DIM_PROFORMA

Descripción: Tabla de Dimensión que contiene las Proforma emitidas para la Venta.

Campo	Tipo de Dato	Descripción
KEY_PROFORMA	INTEGER	Clave Subrogada Primaria de la Tabla.
NO_PROFORMA	INTEGER	Número y Clave Primaria compuesta Origen de la Proforma.
TIPO_PROFORMA	VARCHAR2(20)	Tipo de Proforma.
ESTADO_PROFORMA	VARCHAR2(10)	Estado en el que se Encuentra la Proforma.
ANTICIPO_PROFORMA	VARCHAR2(2)	Especifica si se realizo un Anticipo de Pago de la Proforma.
PERIODO_PROFORMA	INTEGER	Periodo en el que se emitió la Proforma, y clave primaria compuesta origen.

Tabla: DIM_PROVEEDOR

Descripción: Tabla de Dimensión que contiene los Proveedores a los que se compra los Artículos.

Campo	Tipo de Dato	Descripción
KEY_PROVEEDOR	INTEGER	Clave Subrogada

		Primaria de la Tabla
NO_PROVEEDOR	INTEGER	Numero del Proveedor, y clave primaria Origen.
CEDULA_RUC_PROVEEDOR	VARCHAR2(25)	Cedula o Ruc del Proveedor.
NOMBRE_PROVEEDOR	VARCHAR2(150)	Nombre del Proveedor.

Tabla: DIM_PROYECTO

Descripción: Tabla de Dimensión que contiene el Proyecto y Proceso asignado para cada Negocio.

Campo	Tipo de Dato	Descripción
KEY_PROYECTO	INTEGER	Clave Subrogada Primaria de la Tabla
NO_PROYECTO	INTEGER	Numero y Clave primaria Origen del Proyecto.
NOMBRE_PROYECTO	VARCHAR2(100)	Nombre del Proyecto.
NO_PROCESO	INTEGER	Numero y Clave primaria Origen del Proceso.
NOMBRE_PROCESO	VARCHAR2(100)	Nombre del Proceso.

Tabla: DIM_SUCURSAL

Descripción: Tabla de Dimensión que contiene las Sucursales de Venta.

Campo	Tipo de Dato	Descripción
KEY_SUCURSAL	INTEGER	Clave Subrogada Primaria de la Tabla
NOMBRE_SUCURSAL	INTEGER	Descripción de la sucursal.
DIRECCION_SUCURSAL	VARCHAR2(100)	Dirección de la sucursal.

Tabla: DIM_TIEMPO

Descripción: Tabla de Dimensión que contiene el Horizonte de Tiempo.

Campo	Tipo de Dato	Descripción
KEY_TIEMPO	INTEGER	Clave Primaria Subrogada
FECHA	DATE	Fecha del día
ANIO	INTEGER	Año de actual de la Fecha
CUATRIMESTRE	CHAR(2)	Cuatrimestre del Año
MES_NOMBRE	VARCHAR2(30)	Nombre del Mes
DIA_NOMBRE	VARCHAR2(30)	Nombre del día de Semana

Tabla: HEC_PEDIDOS_INV

Descripción: Tabla de Hecho que contiene los Pedidos con su respectiva relación de Dimensión.

Campo	Tipo de Dato	Descripción
KEY_TIEMPO	INTEGER	Clave Foránea de la Dimensión Tiempo.
KEY_ARTICULO	INTEGER	Clave Foránea de la Dimensión Artículo.
KEY_BODEGA	INTEGER	Clave Foránea de la Dimensión Bodega.
KEY_PROYECTO	INTEGER	Clave Foránea de la Dimensión Proyecto.
KEY_AREA	INTEGER	Clave Foránea de la Dimensión Área.
KEY_ORDEN_COMPRA	INTEGER	Clave Foránea de la Dimensión Orden de Compra.
KEY_PROVEEDOR	INTEGER	Clave Foránea de la Dimensión Proveedor.
KEY_PEDIDO	INTEGER	Clave Foránea de la Dimensión Pedido.
CANTIDAD_PEDIDA	NUMBER(14,5)	Cantidad pedida de Artículos.
VALOR_TOTAL	NUMBER(15,7)	Valor del Pedido.

Tabla: HEC_MOVIMIENTOS_INV

Descripción: Tabla de Hecho que contiene los Movimientos con su respectiva relación de Dimensión.

Campo	Tipo de Dato	Descripción
KEY_TIEMPO	INTEGER	Clave Foránea de la Dimensión Tiempo.
KEY_ARTICULO	INTEGER	Clave Foránea de la Dimensión Artículo.
KEY_BODEGA	INTEGER	Clave Foránea de la Dimensión Bodega.
KEY_PROYECTO	INTEGER	Clave Foránea de la Dimensión Proyecto.
KEY_AREA	INTEGER	Clave Foránea de la Dimensión Área.
KEY_ORDEN_COMPRA	INTEGER	Clave Foránea de la Dimensión Orden de Compra.
KEY_PROVEEDOR	INTEGER	Clave Foránea de la Dimensión Proveedor.
KEY_PEDIDO	INTEGER	Clave Foránea de la Dimensión Pedido.
KEY_MOVIMIENTO	INTEGER	Clave Foránea de la Dimensión Movimiento.
CANTIDAD	NUMBER(14,5)	Cantidad de Artículos movidos o trasladados.
VALOR_TOTAL	NUMBER(15,7)	Valor del Movimiento.

Tabla: HEC_VENTAS_PTV

Descripción: Tabla de Hecho que contiene las Ventas con su respectiva relación de Dimensión.

Campo	Tipo de Dato	Descripción
KEY_TIEMPO	INTEGER	Clave Foránea de la Dimensión Tiempo.
KEY_CLIENTE	INTEGER	Clave Foránea de la Dimensión Cliente.
KEY_SUCURSAL	INTEGER	Clave Foránea de la Dimensión Sucursal.
KEY_CENTRO_GESTION	INTEGER	Clave Foránea de la Dimensión Centro de

		Gestión.
KEY_FACTURA	INTEGER	Clave Foránea de la Dimensión Factura.
KEY_ITEM	INTEGER	Clave Foránea de la Dimensión Ítem.
KEY_PROYECTO	INTEGER	Clave Foránea de la Dimensión Proyecto.
KEY_PROFORMA	INTEGER	Clave Foránea de la Dimensión Proforma.
KEY_EMPLEADO	INTEGER	Clave Foránea de la Dimensión Empleado.
CANTIDAD	NUMBER(14,5)	Cantidad de Ítems vendidos.
VALOR_TOTAL	NUMBER(15,7)	Precio Total de Venta, Cantidad por Precio Unitario.

Tabla: HEC_NUM_VENTAS_PTV

Descripción: Tabla de Hecho que contiene el Número de Ventas con su respectiva relación de Dimensión.

Campo	Tipo de Dato	Descripción
KEY_TIEMPO	INTEGER	Clave Foránea de la Dimensión Tiempo.
KEY_CLIENTE	INTEGER	Clave Foránea de la Dimensión Cliente.
KEY_SUCURSAL	INTEGER	Clave Foránea de la Dimensión Sucursal.
KEY_FACTURA	INTEGER	Clave Foránea de la Dimensión Factura.
KEY_EMPLEADO	INTEGER	Clave Foránea de la Dimensión Empleado.
NUMERO_VENTAS	NUMBER	Numero de Ventas realizadas.

HOJA DE LEGALIZACIÓN DE FIRMAS
ELABORADA(O) POR

Sr. Andrés Fabián Duque Gálvez

**COORDINADOR DE CARRERA DE INGENIERIA EN SISTEMAS E
INFORMÁTICA**

ING. DANILO MARTÍNEZ

Sr. Ing. Danilo Martínez

Lugar y fecha: _____