

ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE INGENIERÍA

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL MANEJO DE INFORMACIÓN SOCIO EDUCATIVA, PSICOLÓGICA Y DELICTIVA PARA UN CENTRO DE ORIENTACIÓN JUVENIL

**PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO DE
SISTEMAS INFORMATICOS Y DE COMPUTACION**

MARCO VINICIO LUNA AIZAGA

DIRECTOR: ING. MARIA HALLO

Quito, Diciembre 2006

CAPITULO I.....	1
1 Marco Teórico.....	1
1.1 Centros de Orientación Juvenil en el Ecuador.....	1
1.1.1 Caracterización del adolescente.....	2
1.1.2 Descripción General del Proceso de Diagnóstico.....	3
1.1.3 Proceso Reeducativo.....	4
1.1.4 Tipos de Programas Reeducativos.....	5
1.2 Data Warehouse.....	6
1.2.1 Definición.....	6
1.2.2 Características.....	8
1.2.3 Necesidades.....	9
1.2.3.1 Perspectivas de requerimientos.....	10
1.2.4 Arquitectura.....	11
1.2.4.1 Bloques de la arquitectura.....	13
1.2.4.1.1 Bloque de fuente de datos.....	13
1.2.4.1.2 Bloque de construcción del Data Warehouse.....	15
1.2.4.1.3 Bloque de acceso y uso.....	17
1.2.4.2 Capas de la arquitectura.....	18
1.2.4.2.1 Capa de administración de datos.....	18
1.2.4.2.2 Capa de administración de metadatos.....	19
1.2.4.2.3 Capa de transporte.....	20
1.2.4.2.4 Capa de infraestructura.....	20
1.2.5 Modelo multidimensional.....	21
1.2.5.1 Hechos.....	22
1.2.5.2 Dimensiones.....	22
1.2.5.2.1 Esquema en estrella.....	23
1.2.5.2.2 Esquema copo de nieve.....	24
1.3 Data Mart.....	25
1.3.1 Definición.....	25
1.3.2 Ventajas y desventajas.....	25
1.3.3 Arquitectura del DM.....	26
1.4 Herramientas para el análisis.....	28
1.4.1 Cubos OLAP.....	29
1.4.1.1 Medidas.....	29
1.4.1.2 Dimensiones.....	30
1.4.2 Operaciones con los cubos OLAP.....	31
1.4.3 Esquemas de Almacenamiento.....	32
1.4.3.1 ROLAP.....	32
1.4.3.2 MOLAP.....	32
1.4.3.3 HOLAP.....	32
1.5 Data Mining.....	33
1.6 Metodología Kimball par el desarrollo de DW O DM.....	34
1.6.1 Introducción.....	34
1.6.2 Etapas de la metodología Kimball.....	35
1.6.2.1 Planificación del Proyecto.....	37
1.6.2.1.1 Formación de equipos de trabajo.....	37
1.6.2.1.2 Reunión Inicial.....	38
1.6.2.1.3 Elaboración del Plan Inicial del Proyecto.....	38
1.6.2.1.4 Documentación inicial del proyecto.....	39
1.6.2.2 Definición de Requerimientos.....	39

1.6.2.2.1	Elaboración de la agenda de citas.....	40
1.6.2.2.2	Pre Entrevistas	40
1.6.2.2.3	Entrevista	40
1.6.2.2.4	Resumen de los requerimientos.....	40
1.6.2.3	Análisis	41
1.6.2.3.1	Generalidades	42
1.6.2.3.2	Arquitectura técnica.....	42
1.6.2.3.2.1	Fuentes de datos	43
1.6.2.3.2.2	Reglas de extracción, transformación y carga.....	43
1.6.2.3.2.3	Políticas de uso	43
1.6.2.3.3	Modelo lógico.....	43
1.6.2.4	Diseño.....	44
1.6.2.4.1	Selección de la herramienta.....	44
1.6.2.4.2	Diseño de Físico del DM.....	45
1.6.2.4.2.1	Creación del Modelo de datos físico	45
1.6.2.4.2.2	Determinación de la Arquitectura.....	45
1.6.2.4.2.3	Especificaciones de procesos.....	45
1.6.2.4.2.4	Modelo de extracción y transformación.....	46
1.6.2.4.2.5	Esquema de almacenamiento.....	46
1.6.2.4.3	Especificación de usuario final.....	46
1.6.2.5	Implementación	47
1.6.2.5.1	Construcción del Data Warehouse o Data Mart	47
1.6.2.5.1.1	Creación de la Base de Datos	47
1.6.2.5.1.2	Creación de paquetes de Extracción, Transformación y Carga	47
1.6.2.5.1.3	Aplicación de las reglas de limpieza de datos	48
1.6.2.5.1.4	Creación de Cubos.....	48
1.6.2.5.2	Desarrollo de la Aplicación de Usuario Final	48
1.6.2.6	Pruebas	48
1.6.2.7	Despliegue	49
CAPITULO II.....		50
2	Análisis y Diseño.....	50
2.1	Planificación	50
2.1.1	Descripción del Problema.....	50
2.1.2	Elección Estrategia de Implementación	51
2.1.3	Elección de la Metodología de Desarrollo	51
2.1.4	Justificación de la Metodología.....	54
2.1.5	Formación del Equipo de Trabajo	55
2.1.6	Plan Inicial del Proyecto.....	55
2.1.6.1	Establecimiento del Contexto Inicial y Futuro	55
2.1.6.1.1	Mercado de Destino.....	55
2.1.6.1.2	Plataformas de uso actual y futuras	56
2.1.6.2	Estimación del tiempo de inicio de trabajo para el DM	57
2.1.7	Calendario y Presupuesto	57
2.1.8	Catalogo de Metadatos	58
2.2	Requerimientos.....	59
2.2.1	Agenda de citas con el usuario	59
2.2.2	Resumen de los requerimientos.....	60
2.2.2.1	Requerimientos del Propietario	60
2.2.2.2	Requerimientos del Desarrollador	61

2.2.2.2.1	Herramientas.....	61
2.2.2.2.2	Distribución	62
2.2.2.2.3	Disponibilidad	63
2.2.2.3	Requerimientos del Usuario Final	63
2.2.3	Requerimientos de Arquitectura.....	64
2.2.3.1	Arquitectura de Datos.....	64
2.2.3.2	Arquitectura de la Aplicación.....	64
2.3	Análisis	65
2.3.1	Generalidades	65
2.3.1.1	Directivas del Negocio	65
2.3.1.2	Objetivo del Negocio.....	66
2.3.1.3	Roles y Responsabilidades	66
2.3.2	Arquitectura técnica.....	67
2.3.2.1	Fuentes de Datos.....	67
2.3.2.2	Reglas de extracción, transformación y carga de datos.....	69
2.3.2.3	Políticas de Uso	70
2.3.2.3.1	Como se utilizara el DM.....	70
2.3.2.3.2	Calendarización de carga del DM	71
2.3.2.3.3	Reglas de Seguridad	71
2.3.2.3.4	Políticas de Respaldo y Recuperación.....	72
2.3.3	Modelo lógico.....	73
2.3.3.1	Diagrama de flujo de Datos	73
2.3.3.2	Indicadores.....	80
2.3.3.3	Modelo Conceptual de Datos del DM	87
2.4	Diseño.....	89
2.4.1	Selección de herramientas e instalación	89
2.4.2	Diseño Físico	89
2.4.2.1	Modelo de Datos Físico	89
2.4.2.2	Determinación de la Arquitectura.....	90
2.4.2.2.1	Bloque de la arquitectura.....	91
2.4.2.2.1.1	Bloque de Fuentes de Datos	91
2.4.2.2.1.2	Bloque del Data Mart	92
2.4.2.2.1.3	Bloque de acceso y uso.....	93
2.4.2.2.2	Capas de la arquitectura.....	93
2.4.2.2.2.1	Capa de Administración de datos	93
2.4.2.2.2.2	Capa de Infraestructura.....	94
2.4.2.2.2.3	Capa de Metadatos.....	94
2.4.2.3	Especificación de Procesos para el DM	95
2.4.2.3.1	Extracción, transformación y carga.....	96
2.4.2.3.2	Limpieza de datos.....	96
2.4.2.3.3	Transformación multidimensional	97
2.4.2.3.4	Explotación y Distribución.....	98
2.4.2.4	Modelo de Extracción y Transformación	99
2.4.2.5	Selección del Esquema de Almacenamiento.....	99
2.4.2.6	Especificaciones de Usuario Final.....	99
CAPITULO III	100
3	Implementación y Pruebas.....	100
3.1	Implementación	100
3.1.1	Construcción del Data Mart.....	100
3.1.1.1	Creación de la Base de Datos	100

3.1.1.2	Creación de los paquetes de extracción, transformación y carga de datos	102
3.1.1.2.1	Creación de ODBC's:	102
3.1.1.2.2	Creación de paquetes DTS	104
3.1.1.2.2.1	Conexiones	104
3.1.1.2.2.2	Tareas de ejecución SQL	109
3.1.1.2.2.3	Flujos de trabajo	110
3.1.1.2.2.4	Tareas de transformación de datos	111
3.1.1.3	Aplicación de las reglas de limpieza	118
3.1.1.4	Creación de Cubos	119
3.1.1.4.1	Origen de datos	119
3.1.1.4.2	Dimensiones compartidas	120
3.1.1.4.3	Cubo de datos	124
3.1.2	Desarrollo de Aplicación de Usuario final	128
3.2	Pruebas	132
3.2.1	Pruebas de Integración	133
3.2.2	Pruebas Funcionales	134
3.2.3	Pruebas de Desempeño	134
CAPITULO IV		136
4	Caso de Estudio.	136
4.1	Descripción del Ambiente	136
4.2	Personalización del sistema	136
4.3	Presentación de Resultados	137
4.4	Conclusión	137
CAPITULO V		139
5	Conclusiones y Recomendaciones.	139
5.1	Conclusiones	139
5.2	Recomendaciones	141
BIBILOGRAFIA		144

CAPITULO I

1 Marco Teórico.

1.1 Centros de Orientación Juvenil en el Ecuador

De acuerdo con el plan de operaciones del Ministerio de Bienestar Social, el Proyecto Socioeducativo para promover el desarrollo integral de los niños, niñas y jóvenes infractores de la ley en la República del Ecuador es una alternativa que pretende modificar determinados esquemas de represión, punición e internamiento aún existentes en la intervención con dicha población, a través de la implementación inmediata de medidas socioeducativas, psicológicas y de información delictiva no privativas de la libertad establecidas en la legislación

Dicha intención comprende diversos programas de atención tales como: Recepción, Ubicación Institucional, Libertad Asistida y Comunidad Terapéutica.

Este proyecto es elaborado de acuerdo a lo contemplado en el Código de la niñez y la adolescencia, en la Convención Internacional sobre Derechos del Niño y demás Documentos de las Naciones Unidas sobre Administración de Justicia de Niños, y busca permitir el desarrollo personal de los jóvenes para propiciar una adecuada ubicación en el medio socio-familiar y comunitario.

1.1.1 Caracterización del adolescente

Se define a la adolescencia como la etapa de transición entre la niñez y la edad adulta. En esta fase se producen modificaciones rápidas de orden biológico y psicológico que ocasionan modificaciones en la personalidad del niño. Por ejemplo: desarrollo físico acelerado mayor que el desarrollo social; modificación del rol que conlleva la renuncia de su dependencia infantil; cambio en las imágenes de referencia, que ocasionan comportamientos de exploración adaptiva; manifestación externa de su sexualidad que lo induce a la definición de su identidad; falta de seguridad en sí mismo, etc.

El adolescente infractor comparte las características anteriormente enunciadas, comunes a todos los niños(as) y adolescentes, pero factores individuales y de su entorno socio – familiar, hacen que genere conductas desadaptativa cuya evolución, por lo general culmina con infracciones de la Ley. Entre los factores de riesgo que pueden conducir a cometer tales infracciones sobresalen: Situaciones de extrema pobreza, marginación; familias conflictivas e inestables; La industrialización, el consumismo; la incidencia de los medios de comunicación; las pobres formas de identidad, los modelos de autoridad; la influencia ideológica; la escuela discriminatoria, el maltrato en sus diversas formas, la carencia de programas juveniles de apoyo.

Las actitudes que se presentan con mayor frecuencia en el adolescente son: rechazo a la autoridad, intolerancia, agresividad, astucia, rebeldía, manipulación, desconocimiento de sus potencialidades y dificultades; la no conciencia de su situación legal, las causas, motivos gravedad de la infracción,

demuestra carencia de afecto y satisfacción de necesidades básicas, niveles bajos de motivación, deserción y exclusión escolar, problemas de salud, etc.

1.1.2 Descripción General del Proceso de Diagnóstico

Para la realización del diagnóstico se toman en cuenta los siguientes parámetros:

- Una aproximación a la realidad tomando como referencia el modelo de atención y situación real
- Diagnóstico de la realidad jurídica en cuanto:
 - El menor
 - La familia
 - Los procesos de intervención
- Un diagnóstico de la realidad reeducadita:
 - Crecimiento personal
 - Programa escolar
 - Captación laboral
 - Salud mental y física
 - Participación familiar
 - Socialización
 - Recreación
 - Evaluación y seguimiento
- Un diagnóstico de la realidad de recursos
 - Humanos
 - Físicos

- Financieros
- Institucionales

1.1.3 Proceso Reeducativo

FASES	ETAPAS	AREAS DE INTERVENCIÓN	ESTRATEGIAS	EVALUACIÓN
Acogida	Inducción	<ul style="list-style-type: none"> ○ Salud ○ Lúdico – deportiva ○ Uso adecuado del tiempo libre ○ Jurídica ○ Familiar ○ Psicológica ○ Social 	<ul style="list-style-type: none"> ○ Internamiento provisional ○ Vida en grupo ○ Coloquio ○ Medidas socioeducativas ○ Ubicación e integración familiar ○ Apoyo interdisciplinario 	Diagnóstico
Tratamiento	Encauzamiento Afianzamiento Robustecimiento	<ul style="list-style-type: none"> ○ Salud ○ Lúdico – deportiva ○ Uso adecuado del tiempo libre ○ Jurídica ○ Familiar ○ Psicológica ○ Social 	<ul style="list-style-type: none"> ○ Internamiento provisional ○ Vida en grupo ○ Coloquio ○ Medidas socioeducativas ○ Ubicación e integración familiar ○ Apoyo interdisciplinario ○ Revisión de medidas ○ Permisos / salidas ○ Uso adecuado y progresivo de la libertad 	Clases: Auto evaluación Coevaluación Heteroevaluación Codificación Notas Vales Criterios Basado en un objetivo
Reinserción	Seguimiento Autonomía	<ul style="list-style-type: none"> ○ Familiar ○ Educativa ○ Laboral 	<ul style="list-style-type: none"> ○ Vistas ○ Coloquios ○ Encuentros 	Auto evaluación Heteroevaluación

		<ul style="list-style-type: none"> ○ Uso adecuado del tiempo libre ○ Jurídica ○ Psicológica ○ Social 		
--	--	--	--	--

Tabla 1.1: Proceso Reeducativo

1.1.4 Tipos de Programas Reeducativos

TIPO	CONCEPTO	MODALIDAD	FINALIDAD
RECEPCIÓN	Es un centro especial, primera instancia de abordaje por aprehensión del niño infractor	Internamiento provisional, breve y separado de adultos infractores	<ul style="list-style-type: none"> ○ Pre diagnóstico ○ Definición del grado de responsabilidad infraccional ○ Investigación que puede aproximarse hasta un mes
LIBERTAD ASISTIDA	Tiempo de apoyo y seguimiento (terapéutico)	<ul style="list-style-type: none"> ○ En su propia familia ○ Posibilidad de Servicios a la Comunidad y/o asistencia a otros programas 	<ul style="list-style-type: none"> ○ Acompañamiento y asesoría al niños y su familia ○ Normalizar situaciones problemáticas
CENTRO ABIERTO	Internamiento mitigado	<ul style="list-style-type: none"> ○ Mínimas medidas de seguridad ○ Pernoctan en el Centro ○ Posibilidad de Estudios y trabajar fuera 	<ul style="list-style-type: none"> ○ Atenuar privación de la libertad ○ Potenciar vida familiar ○ Facilitar integración comunitaria ○ Potenciar su autonomía
CENTRO SEMIABIERTO	Restrictivo parcial de la libertad, por infracciones continuadas, sin mayor gravedad e inestabilidad conductual	<ul style="list-style-type: none"> ○ Internamiento por breve periodo ○ Posibilidad de salidas ocasionales por méritos y 	<ul style="list-style-type: none"> ○ Garantizar presencia en el tratamiento

		logro de objetivos	
CENTRO CERRADO	<ul style="list-style-type: none"> ○ Privación de la libertad ○ Infracciones que revisten gravedad 	<ul style="list-style-type: none"> ○ Alta seguridad por el menor tiempo posible 	<ul style="list-style-type: none"> ○ Tratamiento intensivo ○ Protección y seguridad del niño y de la sociedad
COMUNIDAD TERAPÉUTICA	Centro especial para adolescentes con problemas de farmaco dependencia	<ul style="list-style-type: none"> ○ Residencial ○ Permanencia de acuerdo a logros alcanzados 	<ul style="list-style-type: none"> ○ Tratamiento especial ○ Superación de conflictos ○ Potenciación de elementos de resiliencia

Tabla 1.2: Programas Reeducativos

1.2 Data Warehouse.

1.2.1 Definición

Como en la mayoría de los casos, existen una gran cantidad de definiciones para una gran cantidad de conceptos, el Data Warehouse no es la excepción, de tal forma que se enunciará a continuación una de las definiciones mas acertadas: *“Un data warehouse es un conjunto de datos integrados orientados a una materia, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración¹”,* de esta manera se irá desglosando, en base a este concepto, el real significado de lo que es un data warehouse y principalmente su incidencia en el proceso de toma de decisiones de una empresa.

¹ GILL Harjinder, RAO Prakash, Data Warehousing La Integración De La Información Para La Mejor Toma De Decisiones, Ediciones Que, 1996

Se indica que un Data Warehouse es un conjunto de datos integrados y orientados a una materia, efectivamente el Data Warehouse es un almacén de datos con un arquitectura específica y una semántica consistente entre cada uno de sus componentes que fundamenta el proceso de toma de decisiones de una empresa extrayendo los datos de fuentes heterogéneas en un repositorio propio de tal forma que llena los requerimientos de acceso y reporte de datos.

Adicionalmente se dice que los datos de un Data Warehouse varían con el tiempo y que no son transitorios; se puede decir que este concepto va muy de la mano con el hecho de que la tecnología del Data Warehouse tiene entre una de sus dificultades el procurar en lo posible la integración de datos de varias fuentes heterogéneas, que puede ser desde archivos planos hasta sistemas externos de diferentes fabricantes; por lo tanto un Data Warehouse, también es, una tecnología de soporte a la toma de decisiones que debe contener dentro de su arquitectura un mecanismo que asegure la sincronización de dichas fuentes heterogéneas para asegurar la constante actualización del Data Warehouse conforme se crean nuevos datos en dichas fuentes, es decir juntar los datos operativos de una forma consistente de tal forma que sean la base para la exposición de datos estratégicos, en forma de reportes avanzados, para la compañía; orientándose siempre, como se menciona anteriormente, a soportar la toma de decisiones dentro del contexto del modelo de negocios de dicha compañía.

De esta manera se puede definir en forma práctica a un Data Warehouse como un *almacén*² que permite disponer de varias bases de datos que permitan extraer conocimiento de la información histórica almacenada en la organización que nos asegure los siguientes objetivos:

- Análisis de la organización
- Provisiones de evolución; y,
- Diseño de estrategias.

Adicionalmente se debe indicar que en el Data Warehouse, y como en toda la rama de tecnología, es necesario establecer una arquitectura estable, estándar y común, que se apoye en documentación formal y en una metodología específica. Es por esto que dentro del presente capítulo se hará un énfasis especial en dicha arquitectura.

1.2.2 Características

Dentro de los aspectos característicos de un Data Warehouse se puede enunciar los siguientes:

- Orientado hacia la información relevante de la organización: de tal forma que un Data Warehouse no se lo construye sin antes conocer específicamente que información importante de la empresa va a contener y en que contexto de la organización va a dar soporte para adoptar las soluciones mas adecuadas. Es decir que un Data Warehouse, se diseña para consultar eficientemente información relativa a las actividades básicas de la organización (ventas, compras,

² GILL Harjinder, RAO Prakash, Data Warehousing La Integración De La Información Para La Mejor Toma De Decisiones, Ediciones Que, 1996.

producción, etc.) y no para soportar los procesos que se realizan en ella (gestión de pedidos, facturación, etc.).

- Integrada: como se indico anteriormente la información que contiene un Data Warehouse se debe integrar desde varias fuentes heterogéneas internas o externas.
- Variable en el tiempo: esta característica se refiere que los datos del Data Warehouse son relativos a un periodo y deben ser incrementados periódicamente.
- No Volátil: es decir que los datos en el Data Warehouse deben ir incrementándose y no actualizándose directamente ya que si se actualiza alguna información dentro del Data Warehouse, debe ser porque se actualizo previamente la fuente de información y por consecuencia se actualizo el Data Warehouse; es por eso que el Data Warehouse solo soporta comandos de lectura.

1.2.3 Necesidades

Es importante definir, que para la apreciación de las necesidades, que el Data Warehouse debe cumplir, se debe definir y diferenciar las distintas maneras de ver, interpretar y/o formular los requerimientos³, por parte de los actores dentro de la arquitectura de un Data Warehouse, previos al desarrollo del mismo.

³GILL Harjinder, RAO Praskash, Data Warehousing La Integración De La Información Para La Mejor Toma De Decisiones, Ediciones Que, 1996.

1.2.3.1 Perspectivas de requerimientos

Una forma práctica de ver las necesidades empresariales que debe cumplir el Data Warehouse, es dividir las perspectivas de requerimientos desde tres visiones distintas que son:

- Desde el punto de vista del inversionista, es decir de la persona que proporciona el recurso financiero para la contratación, construcción, implantación y puesta en marcha el Data Warehouse.
- Desde el punto de vista del usuario empresaria, quien es la persona que conoce el modelo de negocio de buena manera y quien va a trabajar con el apoyo del Data Warehouse día a día.
- Desde el punto de vista del desarrollador, quien va a ser la persona que debe empaparse completamente de todo lo concerniente al negocio de tal forma que ese conocimiento lo refleje en la construcción, implantación y puesta en marcha del Data Warehouse a fin de que este cumpla con las expectativas de los usuarios antes mencionados.

De esta manera desglosaremos estas tres perspectivas de requerimientos de manera que se beneficie el rango de posibilidades para el desarrollo del Data Warehouse.

1.2.4 Arquitectura

La arquitectura de un DW se lo puede describir desde un punto de vista a alto nivel como:

- Un conjunto de datos extraídos desde repositorios de datos operacionales, internos y externos.
- Software que prepara los datos para ser accedidos por parte de los usuarios.
- Un conjunto de programas y herramientas que permiten explotar la información almacenada en el sistema de DW mediante consultas y análisis complejos.

Adicionalmente se puede subdividir esta representación arquitectónica, vista desde un nivel macro, en pequeños componentes que nos brindarán una visión mas detallada de lo que es un DW de tal manera que se pueda efectuar una implementación ordenada y correcta del DW.

Dentro de esta arquitectura se puede exponer los componentes necesarios para el Data Warehouse (DW) y el Data Mart (DM).

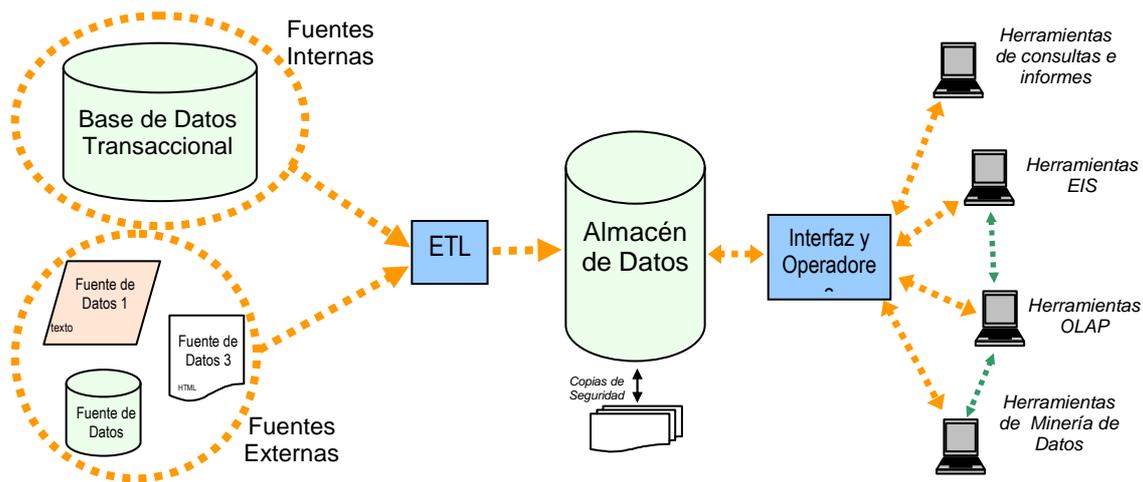


Figura 1.1: Arquitectura del DW a alto nivel

La arquitectura del DW a nivel de sus componentes específicos se divide en las capas de:

- Administración de datos
- Administración de metadatos
- Transporte
- Infraestructura
- Herramientas, tecnología y funciones

Donde las capas de administración de datos y metadatos aportan con las actividades de extracción, carga y actualización de la información para el Data Warehouse y el Data Mart. Así mismo, las capas de Transporte e Infraestructura proveen los servicios comunes necesarios para los bloques del Data Warehouse y el Data Mart que se describirán mas adelante. Finalmente la capa de herramientas, tecnología y funciones apoya a la metodología que se usa para la construcción y despliegue del Data Warehouse y el Data Mart.

Adicionalmente la arquitectura expone varios bloques que son los siguientes:

- Fuentes de datos
- Construcción del Data Warehouse
- Construcción del Data Mart
- Acceso y Uso

Los cuales aportan con la funcionalidad específica que se requiere tanto para el Data Warehouse como para el Data Mart

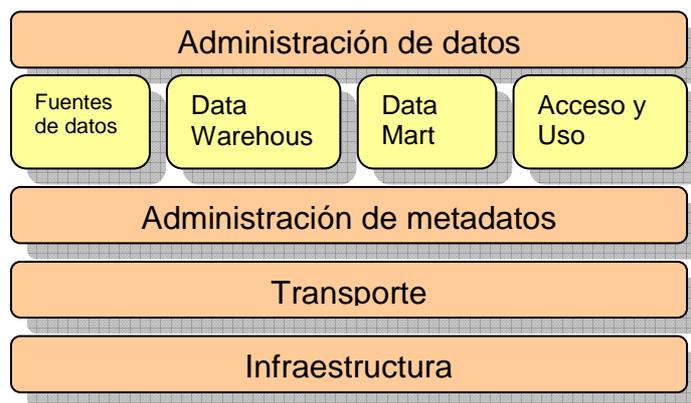


Figura 1.2: Arquitectura del DW a nivel de detalle

1.2.4.1 Bloques de la arquitectura

1.2.4.1.1 Bloque de fuente de datos

El bloque de fuentes de datos se compone de:

- Datos de producción: se refieren a la información almacenada en bases de datos operacionales alimentadas desde los sistemas operacionales los cuales pueden venir en varias tecnologías

- Datos de herencia: se refiere a información fuera de línea ya que están en archivos que no son necesarios para apoyar a los sistemas operacionales actuales pero que tiene una riqueza significativa especialmente para la minería de datos y en los cuales se debe especificar bien las fechas respectivas.
- Sistemas internos de oficina: los cuales aportan con información que no se encuentra almacenada en una base de datos operacional como tal; dichos datos pueden ser no estructurados (formas no electrónicas), estructurados (formas electrónicas como hojas de cálculo, etc.), semiestructurados (formas electrónicas en formato de impresión). La información que aportan estos datos pueden utilizarse dentro de un contexto departamental.
- Sistemas externos: se refiere a la información que la empresa no administra y que se mas bien es una información adquirida y que se encuentra en repositorios externos (como pueden ser los indicadores de las bolsas de valores, entre otros).
- Metadatos de las fuentes: es la información referente a los datos extraídos de las fuentes.

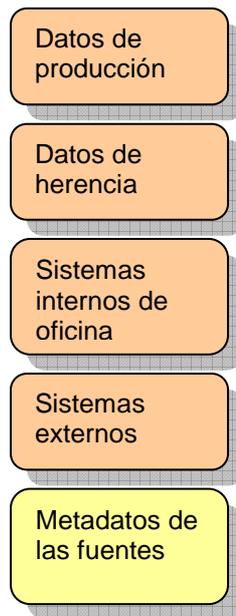


Figura 1.3: Bloque de fuentes de datos a detalle

1.2.4.1.2 Bloque de construcción del Data Warehouse

Este se constituye como uno de los más importantes bloques de la arquitectura y que a sus ves se subdivide en tres componentes principales:

- Refinamiento: que se encarga de estandarizar, limpiar, pulir, filtrar, confrontar, ajustar y verificar los datos que son extraídos de las distintas fuentes con el fin de que estos estén preparados para iniciar el proceso del Data Warehouse. Adicionalmente los metadatos que se colocan en este componente tienen que ver con la fecha en que estos datos son extraídos, la fuente y los campos adicionales que se hayan colocado para completar la información.
- Reingeniería: este componente difiere en concepto de lo que es la reingeniería para las aplicaciones ya que en este caso se trata de que los datos refinados son expuestos a procesos que conllevan el prepararlos para que sean congruentes con las necesidades de análisis

del usuario empresarial. Aquí se realizan tareas de integración de los datos de las distintas fuentes para luego irlos condensando y agregando también se realizan cálculos a fin de obtener mas datos que finalmente son traducidos y formateados para que se presenten de forma normalizada al usuario final. Finalmente se reubican dichos datos de las fuentes originales hacia el repositorio del Data Warehouse. Aquí los metadatos extraídos se refieren a los datos que han sido integrados, los datos de resumen y los datos calculados; las reglas de traducción y reubicación de los datos.

- Data Warehouse: este componente en sí refleja el modelamiento del Data Warehouse y el resumen de los grandes volúmenes de información que provienen de las distintas fuentes en porciones manejables. Los metadatos que se capturan en este componente tienen que ver con la estructura en sí del Data Warehouse, la descripción de las bases de datos, las rutas y reglas de navegación del Data Warehouse y la lista de consultas y reportes diseñados para el Data Warehouse.

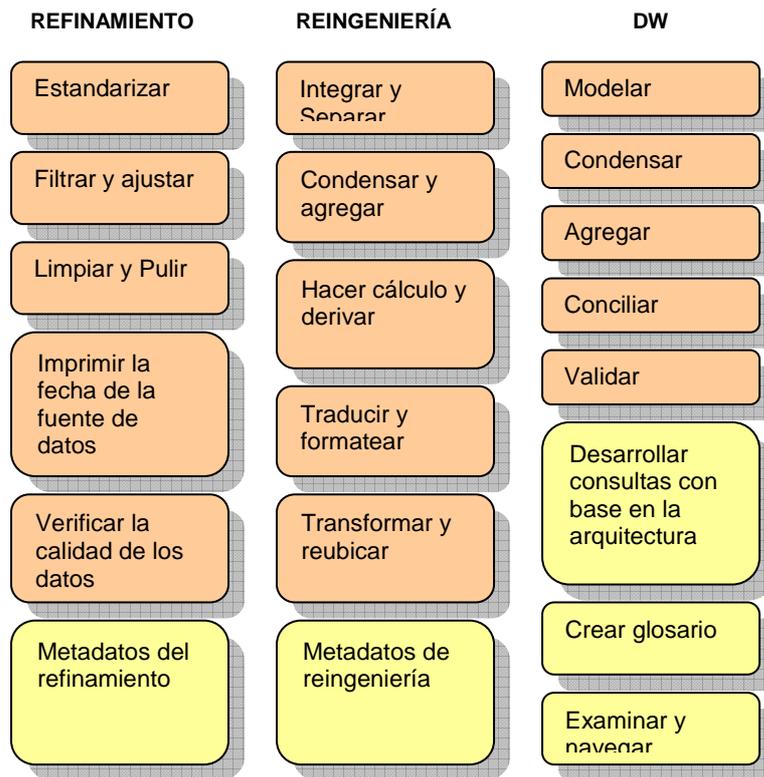


Figura 1.4: Bloque del Data Warehouse en detalle

1.2.4.1.3 *Bloque de acceso y uso*

Este bloque se encargará de proveer la retribución y valor agregado del Data Warehouse hacia la compañía que lo implementa. Para lo cual dicho bloque tiene dos componentes principales:

- **Acceso y Recuperación:** este componente proporciona el acceso al Data Warehouse o Data Mart, transforma los datos en vistas multidimensionales de acuerdo a los esquemas de almacenamiento que se verá mas adelante en un depósito local que permitirá el posterior análisis, minería y elaboración de reportes. En este caso los metadatos capturados nos servirán para entender el DW o DM, entender el contexto de los datos, determinar la ubicación de los datos, verificar la

confiabilidad de los mismos y encontrar la forma de acceder y usar los datos reales.

- **Análisis y Reporte:** este componente es responsable del conjunto de herramientas y aplicaciones necesarias para aprovechar las bondades del DW o DM; incluso las herramientas que nos ayudan a mantener y administrar los metadatos que se han ido capturando.

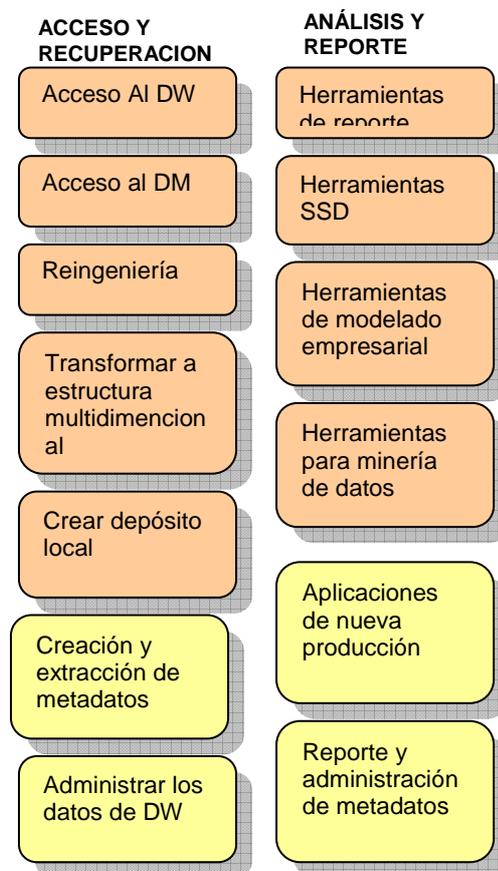


Figura 1.5: Bloque de Acceso y Uso en detalle

1.2.4.2 Capas de la arquitectura

1.2.4.2.1 Capa de administración de datos

Esta capa se encarga de los servicios de extracción, carga, actualización, seguridad, archivo y restauración del DW o el DM.

De esta manera los datos apropiados son extraídos de las fuentes de información seleccionadas para el refinamiento, reingeniería y posterior incorporación al DW o al DM.

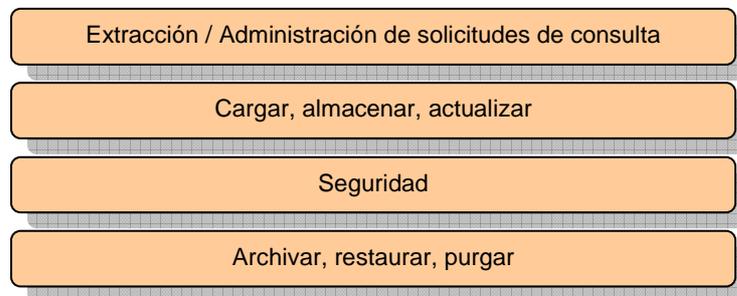


Figura 1.6: Capa de administración de datos en detalle

1.2.4.2.2 Capa de administración de metadatos

Esta capa es la responsable de manejar los metadatos que se requieren para el DW tales como los modelos y esquemas de datos físicos y lógicos del DW y el DM además del glosario técnico y empresarial entre otros. La capa de administración de metadatos es responsable de controlar:

- Definición estándar de los datos depositados en el DW
- Metadatos capturados en los bloques de refinamiento y reingeniería
- Los metadatos sobre granularidad, segmentación, etc.
- Los metadatos sobre consultas y reportes predefinidos y diseñados
- Sobre índices y perfiles para mejorar rendimiento
- Los metadatos que describen reglas para reparar y programar el ciclo de actualización y depuración.

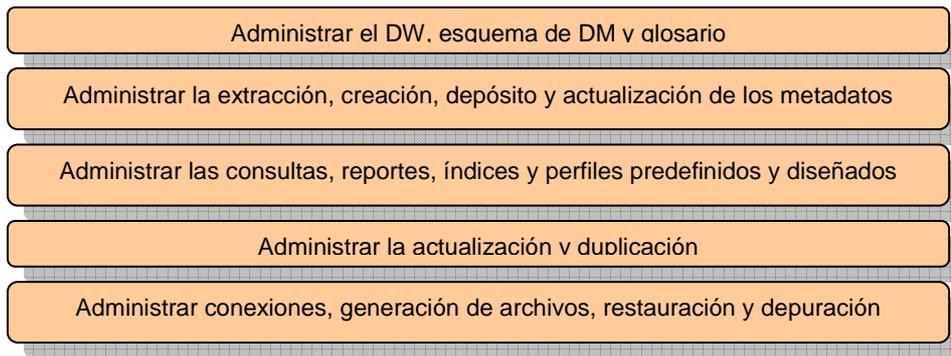


Figura 1.7: Capa de administración de metadatos en detalle

1.2.4.2.3 Capa de transporte

Esta capa se responsabiliza de retribuir el componente de transporte entre los distintos bloques de la arquitectura utilizando arquitectura de actualización y duplicación, red de transferencia y entrega de datos, además de componentes de middleware. Adicionalmente esta capa contribuye con el soporte de protocolos de red, sistemas operativos de red, tipos de redes, accesos a base de datos, sistemas de duplicación, sistemas de seguridad y autenticación entre otros los cuales contribuyen a la integración de la información necesaria al DW.

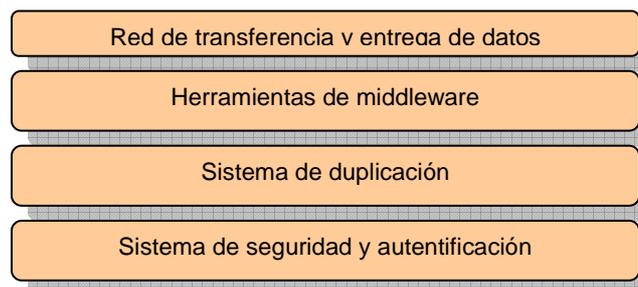


Figura 1.8: Capa de transporte en detalle

1.2.4.2.4 Capa de infraestructura

La capa de infraestructura se encarga de dar el soporte a los bloques de la arquitectura mediante la implementación de los siguientes componentes:

- Administración de sistemas; el cual contribuye el soporte necesario para el desarrollo y manejo de las aplicaciones que trabajen con el DW o el DM.
- Administración del flujo de trabajo; contribuye con el sustento al proceso de integración y administración para coordinar la ejecución ordenada y especificada de herramientas, aplicaciones y actividades.
- Sistemas de almacenamiento; el cual aporta con los servicios de administración de bases de datos tanto para las fuentes de datos como para el catálogo del DM y el almacenamiento multidimensional.
- Sistemas de procesamiento; contribuye con los ambientes operativos implícitos en los bloques de la arquitectura: fuente de datos, DW, DM, herramientas de acceso y uso, etc.

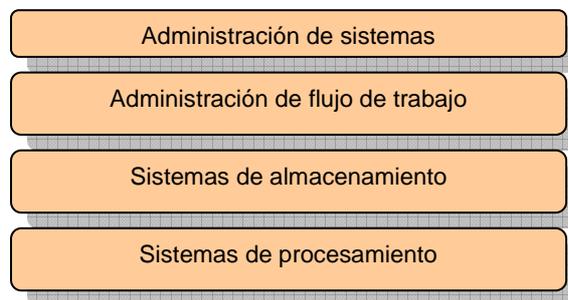


Figura 1.9: Capa de infraestructura en detalle

1.2.5 Modelo multidimensional

Para facilitar el proceso de visualización gráfica de un DW o DM, las herramientas de análisis, que serán producto de un estudio posterior, presentan al usuario un modelo multidimensional en el cual se muestra una actividad que es objeto de análisis y las dimensiones cuantificables que facilitarán dicho proceso de análisis.

1.2.5.1 Hechos

Hecho es la actividad que es objeto del análisis cuya información relevante se representa mediante un conjunto de indicadores que los llamaremos medidas o atributos del hecho; por ejemplo: “En un almacén X una venta es registrada de la siguiente manera: se vendió en el almacén #1 el día 03/02/2005, 5 unidades del producto Y por un valor de 400 USD”. El hecho o actividad que será objeto de un análisis el las ventas en dólares, que se intentará visualizar por almacén, por tiempo y por producto.

1.2.5.2 Dimensiones

Una dimensión será las características medibles de la actividad que es objeto del análisis o los puntos de vista desde los cuales se puede medir una actividad; es decir, que en el caso del ejemplo anterior las dimensiones que caracterizan la actividad o hecho serán las ventas (hecho) por almacén, las ventas en determinado tiempo o período y las ventas por producto de tal forma que si representáramos los hechos y sus dimensiones en un gráfico, quedaría de la siguiente manera:

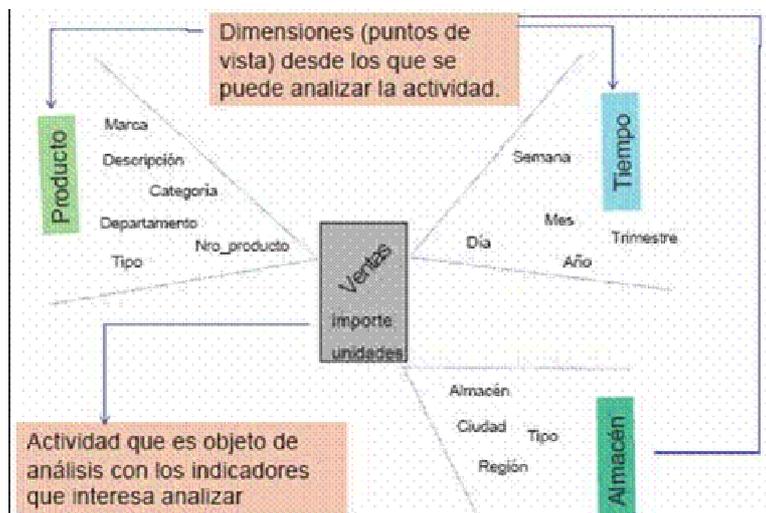


Figura 1.10: Hechos y Dimensiones de un modelo multidimensional

Este esquema dimensional se puede representar mediante jerarquías de los atributos de dichas dimensiones; por ejemplo, en el caso de la dimensión del tiempo los atributos de dicha dimensión se pueden representar ya sea por días, meses, semanas, trimestres o años la jerarquía que se muestra en el gráfico siguiente viene dada porque luego de medir en días, los meses o las semanas podrían ser cualquiera de los dos la jerarquía inmediatamente superior, luego de esto la medida en trimestres, etc. son la jerarquía superior de meses y mas no de las semanas ya que los trimestres están en otro contexto que de las semanas y la jerarquía inmediatamente superior a las semanas y a los trimestres sería los años.

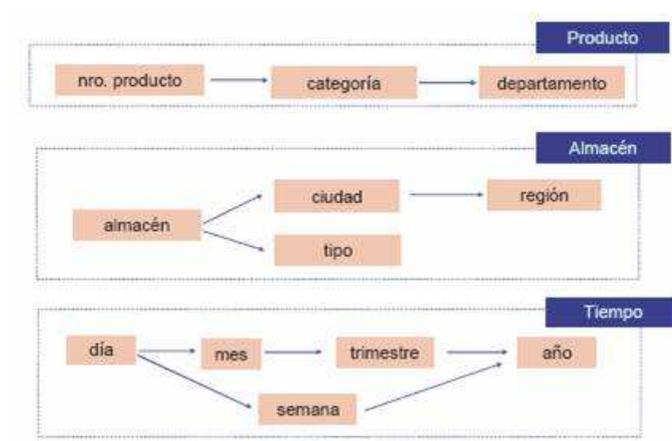


Figura 1.11: Jerarquías de los atributos de una dimensión

1.2.5.2.1 Esquema en estrella

Este esquema se representa cuando la jerarquía de los atributos de una dimensión es lineal; por ejemplo, si queremos analizar una actividad o hecho relacionado con el personal de una compañía usaríamos dimensiones como el tiempo (solo en días, semanas y años), proyecto y equipo; puestas así las dimensiones, sus atributos nos muestran una jerarquía lineal, es decir, sin

bifurcaciones o jerarquía en árbol por lo que el gráfico que representaría este esquema sería:

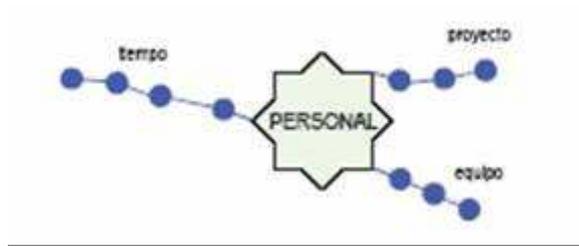


Figura 1.12: Esquema en estrella

1.2.5.2.2 Esquema copo de nieve

Este esquema se presenta cuando la jerarquía de los atributos de las dimensiones que miden una actividad, no es lineal; por ejemplo, en el caso del hecho de medir las ventas de un almacén determinado, la dimensión de tiempo y almacén nos presentaban una jerarquía de atributos no lineal, por lo que el gráfico que representaría este esquema sería:

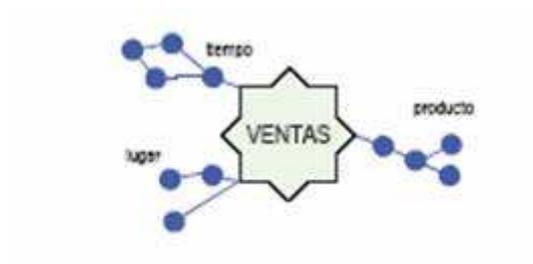


Figura 1.13: Esquema copo de nieve

1.3 Data Mart.

1.3.1 Definición

Dentro de la arquitectura de un DW, se puede constatar que existe la posibilidad de incluir un bloque adicional, el cual se lo nombrará como el bloque de construcción del Data Mart (DM). El cual acota el ámbito de un DW a límites departamentales o seccionales y que además enfoca la información de detalle departamental, con una implementación sencilla, con consultas rápidas y más información agregada que la que tendría un DW.

Adicionalmente un DM es representado por un único esquema en estrella o copo de nieve y el conjunto de estos esquemas formarán la estructura de un DW; por lo que, se puede definir que un DW se puede originar desde varios DM; es decir, que la primera iniciativa en una organización fue el crear varios DM departamentales, que han funcionando regularmente por algún tiempo, y que en base de estos se construye un DW, lo que se conoce como una arquitectura BOTTOM - UP. O por el contrario la organización inicia con el reto de construir un DW y luego lo subdivide en varios DM específicos lo que se conoce como arquitectura TOP - DOWN.

1.3.2 Ventajas y desventajas

VENTAJAS:

- Más simple y rápido de implementar
- Primero se logran las necesidades departamentales
- Las consultas son más rápidas

- Contienen datos que van más allá de los que mantiene normalmente un DW
- Menores requerimientos de software

DESVENTAJAS:

- Se puede inadvertidamente construir DM incompatibles entre si
- Si se construye un DW primero, se requiere hardware adicional para soportar a los DM individuales
- Se tienen datos descentralizados por lo que el usuario puede necesitar hacer búsquedas en más de un lugar

1.3.3 Arquitectura del DM

El DM tiene componentes similares a los expuestos dentro del bloque de construcción del DW en la arquitectura propuesta anteriormente, la principal diferencia entre estos dos es el enfoque al usuario final. Los componentes del mercado de datos se aplican a diferentes conjuntos de pasos de refinamiento y reingeniería para los objetivos del usuario final. Debido a la complejidad y el ámbito reducidos de las tareas de refinamiento y reingeniería del DM, estas actividades se combinan en un solo sub bloque de refinamiento y reingeniería. De esta manera y según la necesidad de la organización se puede remplazar el bloque de DW por el bloque de creación del DM.

En los casos en que la organización decide tener varios DM y ningún DW, el ámbito y la complejidad de las actividades de refinamiento y reingeniería están en la misma escala que las requeridas para un DW; al contrario, en los casos

en que los DM de la organización se alimentan por medio del DW corporativo, se realiza varias actividades de refinamiento y reingeniería.

El bloque de creación del mercado de datos posee los mismos componentes de los del bloque del DW, descrito antes. La diferencia radica en el enfoque de cada uno de los componentes. Por ejemplo, el bloque de modelado en el DM se concentra en las necesidades del departamento individual que posee DM, mientras que el bloque de modelado corporativo se basa en los diagramas de tiempo y granularidad de un departamento y no de todos los departamentos de la organización.

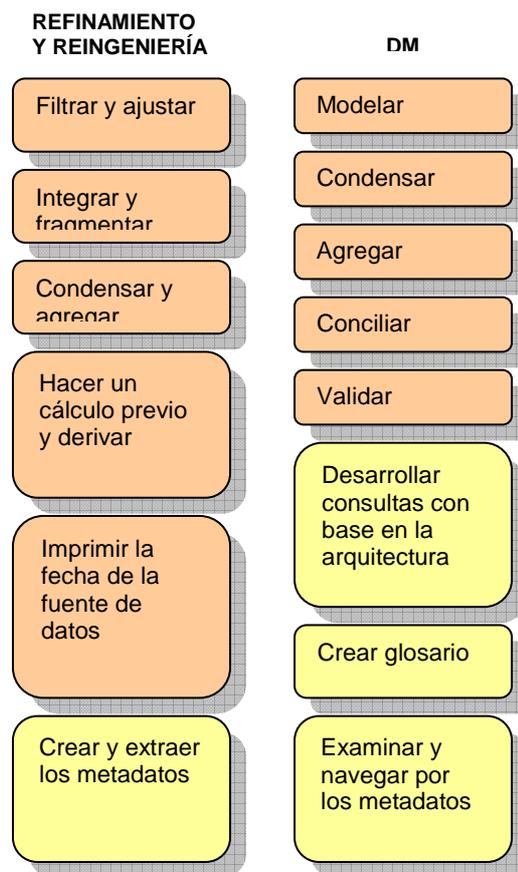


Figura 1.14: Bloque de creación del DM en detalle.

Cuando no existe un DW, se pueden aplicar los componentes de refinamiento y reingeniería descritos en el bloque del DW para la creación del DM. En el contexto de la arquitectura de referencia, descrita anteriormente, el componente del DM es esencialmente idéntico al componente del DW que se describió antes en el bloque del DW.

1.4 Herramientas para el análisis

Las herramientas de análisis que se presentan dentro de la tecnología del DW son las herramientas OLAP quienes presentan al usuario una visión multidimensional de los datos (esquema multidimensional) para cada actividad que es objeto de análisis.

El usuario formula consultas a la herramienta OLAP seleccionando atributos de este esquema multidimensional sin conocer la estructura interna (esquema físico) del almacén de datos.

La herramienta OLAP genera la correspondiente consulta y la envía al gestor de consultas del sistema.

Una consulta a un DW consiste generalmente en la obtención de medidas sobre los hechos parametrizados por atributos de las dimensiones y restringidas por condiciones impuestas sobre las dimensiones:



Restricciones: productos del departamento Bebidas, ventas durante este año

Parámetros de la consulta: por categoría de producto y por trimestre

Figura 1.15: Descripción de las herramientas OLAP

1.4.1 Cubos OLAP

El objeto principal dentro de la tecnología OLAP es el cubo. Un cubo es una representación multidimensional de un conjunto de datos, conteniendo detalle y datos de resumen. Una base de datos OLAP puede tener tantos cubos como sean necesarios y cada uno de estos conforma el DW. Un cubo se construye usando dos componentes principales las medidas y las dimensiones. Una medida es un valor numérico sobre un hecho específico dentro del contexto del DW, como por ejemplo, el precio de un computador o la cantidad de un producto determinado; mientras que una dimensión categoriza las medidas; como por ejemplo, cómo una medida cambia en el tiempo.

1.4.1.1 Medidas

Una medida es la unidad fundamental de análisis dentro de un cubo. Una medida puede contener una sola columna de una tabla hecho o puede ser computada mediante una expresión que envuelva una o más columnas de la tabla hecho.

Una medida puede ser aditiva o no aditiva. Aditiva implica que los valores de las medidas puedan ser totalizados. Un buen ejemplo de esto es el precio de venta final (suma el precio de venta al público más los impuestos). No Aditiva implica que las medidas no pueden ser sumadas pero se pueden contar y el resultado de este conteo si podría ser sumado de alguna manera.

1.4.1.2 Dimensiones

Una dimensión es una manera de categorizar una medida, como por ejemplo, cuántas computadoras fueron vendidas en este último trimestre. Dentro de una dimensión se pueden definir múltiples niveles. Un nivel representa un grado de agregación.

Consideremos el ejemplo de que cuántas computadoras fueron vendidas en el último trimestre. El número de computadoras vendidas es la medida, el tiempo es la dimensión. Usando la dimensión del tiempo se puede especificar un nivel de agregación en este caso es un trimestre, se podría medir por días, por años o por semestres.

Los niveles son ordenados jerárquicamente, el nivel mayor es el de mayor agregación y el nivel mas bajo es el último en la agregación. El único requisito para estos niveles de agregación es que el número de miembros en cada nivel debe ser menor o igual al número de miembros del nivel anterior. El ejemplo de cuántas computadoras se vendieron es una buena oportunidad para experimentar con los niveles de agregación en donde la cabeza de la

agregación será “Todo el tiempo”, seguido de “En un año”, seguido de “En un trimestre” y “Por mes”, etc.

1.4.2 Operaciones con los cubos OLAP

Lo interesante no es poder realizar consultas que, en cierto modo, se pueden hacer con selecciones, proyecciones, concatenaciones y agrupamientos tradicionales. Lo realmente interesante de las herramientas OLAP son sus operadores de refinamiento o manipulación de consultas.

- PROCESSING
- DRILL
- ROLL
- SLICE & DICE
- PIVOT

El carácter agregado de las consultas en el Análisis de Datos, aconseja la definición de nuevos operadores que faciliten la agregación (consolidación) y la disgregación (división) de los datos:

- Agregación (**roll**): permite eliminar un criterio de agrupación en el análisis, agregando los grupos actuales.
- Disgregación (**drill**): permite introducir un nuevo criterio de agrupación en el análisis, disgregando los grupos actuales.

1.4.3 Esquemas de Almacenamiento

El DW y en general las herramientas que se utilizan para el análisis, como son las herramientas OLAP, se basan en tres opciones de almacenamiento físico: ROLAP, MOLAP y HOLAP.

1.4.3.1 ROLAP

Este esquema de almacenamiento físico se implementa sobre tecnología relacional, pero disponen de algunas facilidades para mejorar el rendimiento (índices de mapas de bits, índices de JOIN).

1.4.3.2 MOLAP

Una forma común de almacenar los datos es en una base de datos multidimensional, no almacenando la información como en una base de datos relacional con tablas con filas y columnas; sino, mediante arreglos multidimensionales. De esta manera las dimensiones serán de fácil acceso y se pueden ejecutar consultas con un mejor rendimiento.

1.4.3.3 HOLAP

Quizás la solución más sensible al problema de almacenamiento sea el usar sistemas híbridos como son los sistemas HOLAP donde los datos se mantienen en un sistema ROLAP y los datos sumarios y de resumen en un sistema MOLAP y los servicios OLAP de los sistemas HOLAP son capaces de

usar los datos MOLAP y ROLAP simultáneamente para resolver una consulta multidimensional.

1.5 Data Mining

Podemos definir al Data Mining como “la extracción de información oculta y predecible de grandes bases de datos”⁴, es una poderosa tecnología nueva con gran potencial para ayudar a las organizaciones a concentrarse en la información más importante de sus DW. Las herramientas de Data Mining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información (knowledge-driven). Los análisis prospectivos automatizados ofrecidos por un producto así van más allá de los eventos pasados provistos por herramientas retrospectivas típicas de sistemas de soporte de decisión. Las herramientas de Data Mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas y a los cuales los usuarios de esta información casi no están dispuestos a aceptar

Vale la pena aclarar que solo se hace una referencia conceptual de lo que es el Data Mining, como parte del contexto general del Data Warehousing, ya que no se desarrollará ningún módulo de Data Mining en el presente proyecto.

⁴ VARIOS AUTORES. Data Mining Concepts.
http://www.pcc.qub.cc.uk/tec/courses/datamining/ohp/dm_OHP_final_1.html. 2001

1.6 Metodología Kimball par el desarrollo de DW O DM⁵

El Data Warehouse (DW) incluido el Data Mart (DM) siguen los mismos pasos que para el desarrollo de software convencional, con el mismo ciclo de desarrollo, metodología, entregables, administrables y tiempos de entrega planificados. Sin embargo existen metodologías específicas desarrolladas para la construcción de DW o DM, una de ellas es la Metodología Kimball la cual se resume en los siguientes puntos.

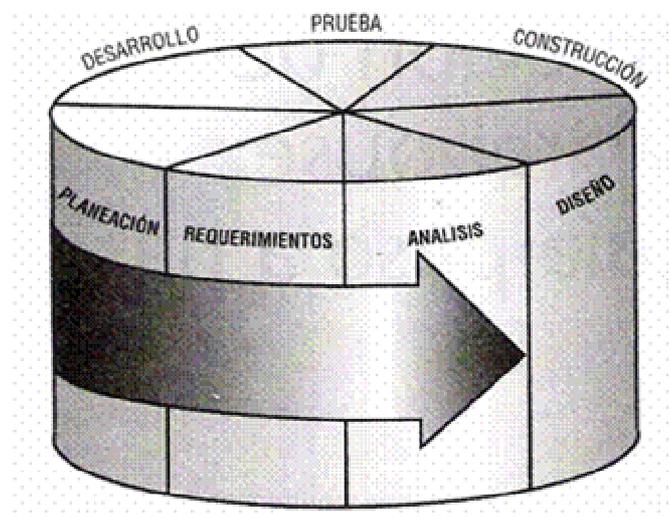


Figura 1.16: Ciclo de desarrollo de software convencional.

1.6.1 Introducción

Esta metodología se orienta principalmente al diseño de la Base de Datos la cual almacenará la información que servirá como apoyo a la toma de decisiones. No obstante toma en cuenta todos los elementos necesarios para la construcción de un DW o DM para lo cual deja abierto el campo para la utilización de artefactos, documentación, modelos de otras metodologías o disciplinas de desarrollo de software que aseguren la calidad del producto final.

⁵ KIMBALL Ralph, The Data Warehouse Lifecycle Toolkit, John Wiley & Sons Inc, 1998

Vale la pena enfatizar que la metodología Kimball maneja como principal interés el desarrollo rápido y consistente del DM.

1.6.2 Etapas de la metodología Kimball

El ciclo de vida de la metodología Kimball es en espiral cuyas etapas se resumen en el siguiente gráfico y se describen a continuación.

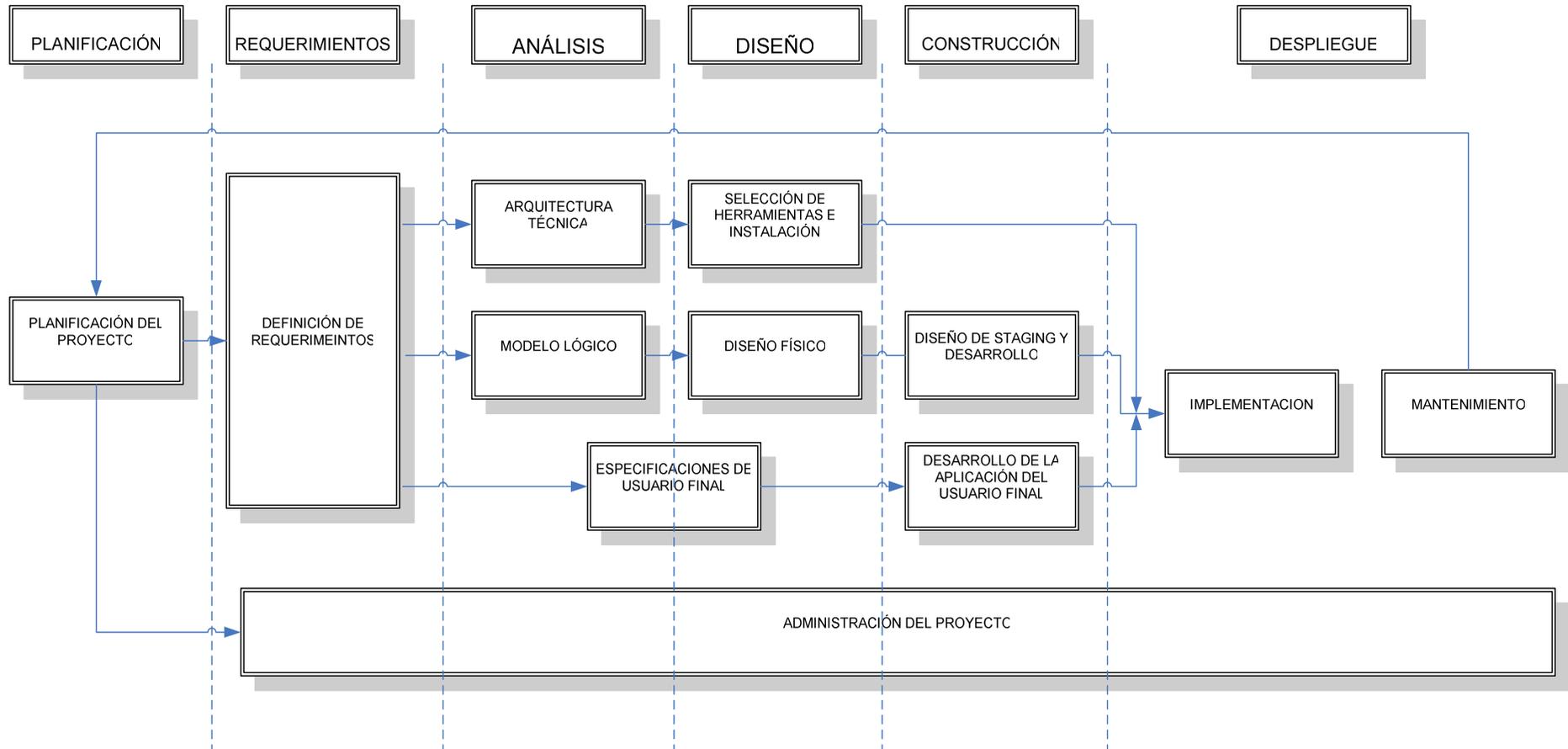


Figura 1.17: Etapas de la Metodología Kimball.

1.6.2.1 Planificación del Proyecto

La planificación es la etapa que tiene como hilo conductor la visión del dominio, naturaleza y estrategia de solución de un problema en particular lo cual nos lleva a determinar una serie de aspectos particulares e iniciales; mismos que son:

1.6.2.1.1 Formación de equipos de trabajo

Esta actividad se refiere a la determinación del equipo de trabajo que estará conformado de acuerdo a los siguientes roles:

ROL	DESCRIPCIÓN
ADMINISTRADOR DEL PROYECTO	Es la persona que se encarga del manejo de los tiempos, fases, recursos, monitoreando constantemente el avance y la productividad del equipo de trabajo y de los avances en el desarrollo del proyecto
ANALISTAS	Es el grupo de personas quienes se encargan de recopilar y analizar la información obtenida del cliente la cual es necesaria para el diseño posterior de la base de datos del DM. Adicionalmente determinan las políticas de uso posterior que se le dará al DM en común acuerdo con el cliente; vale la pena aclarar que las políticas de uso se refieren a la seguridad, calendario de carga, políticas de respaldo y recuperación; y otras actividades que se relacionen con el adecuado y eficaz uso del DM
DISEÑADORES	Es el grupo de personas quienes se encargan de diseñar la base de datos del DM y de los paquetes de extracción, transformación y carga en base a los requerimientos que han sido recopilados y analizados anteriormente. Adicionalmente también se encargan del diseño de las

	herramientas de usuario final para la explotación del DM.
IMPLEMENTADORES	Es el grupo de personas quienes traducen las especificaciones de diseño en: tablas de hechos y dimensiones, paquetes de extracción, transformación y carga, cubos y reportes multidimensionales físicamente existentes las cuales serán de uso práctico por parte del cliente.
PERSONAL DE PRUEBAS	Es el grupo de personas conformados por un coordinador, quien es parte del equipo de desarrollo técnico; y, personal de la organización quienes llevarán a cabo el plan de pruebas diseñado para el proyecto con el fin comprobar el cumplimiento de las especificaciones funcionales determinadas para la creación del DW O DM
CONTRAPARTE	Es la persona o grupo de personas encargadas de la coordinación general del proyecto por parte de la organización, quienes priorizan la satisfacción de los requerimientos asegurándose de que estos sean los que dirijan el desarrollo del proyecto, defendiendo siempre los intereses de dicha organización.

Tabla 1.3: Descripción de Roles.

1.6.2.1.2 Reunión Inicial

Como su nombre lo indica es la primera reunión formal en la cual se dan a conocer al personal, los roles y responsabilidades; así como también, se determina el plan inicial del proyecto. Finalmente se pacta la fecha de la siguiente reunión

1.6.2.1.3 Elaboración del Plan Inicial del Proyecto

El Plan Inicial del Proyecto debe constar de las actividades, fechas, hitos de revisión y entrega de la primera versión del DW o DM. Como parte de la creación de este plan se puede utilizar metodologías y herramientas automatizadas para planificación y monitoreo de proyectos. Vale la pena aclarar que como toda planificación, el Plan Inicial del Proyecto, esta sujeto a variaciones y ajustes según las necesidades de la Contraparte o del Equipo de Desarrollo representados por el Administrador del Proyecto.

1.6.2.1.4 Documentación inicial del proyecto

Está conformada por los primeros documentos del proyecto los cuales pueden incluir solo el acta de la reunión inicial y el plan inicial del proyecto o documentos con información adicional de ayuda al robustecimiento de la planificación y organización formal del proyecto; estos documentos adicionales pueden ser especificaciones metodológicas, especificaciones de estándares de documentación y modelamiento, entre otros los mismos que dependerán de la habilidad y conocimiento del equipo de trabajo

1.6.2.2 Definición de Requerimientos

La definición de requerimientos se refiere a la detección de las necesidades del negocio mediante la realización de entrevistas a los usuarios finales. Estas entrevistas serán registradas en una serie de documentos los cuales permitirán la comprensión exacta de los requerimientos de la empresa. Se registran las siguientes actividades.

1.6.2.2.1 Elaboración de la agenda de citas

La agenda de citas con el usuario es el documento en el cual se colocan las fechas de citas con las personas asignadas, según la contraparte, para el levantamiento de la información. Vale la pena aclarar que esta agenda se la debe realizar en común acuerdo con la contraparte con el fin de comprometer a los miembros de la organización para el cumplimiento de los tiempos colocados en dicha agenda.

1.6.2.2.2 Pre Entrevistas

Las cuales se realizan para establecer recursos que se necesitarán para la entrevista definitiva, como documentos informativos, formatos, equipos como radio grabadoras; también se debe tomar en cuenta las prohibiciones dictadas por la organización que influyen en la entrevista, como el uso de filmadoras, la extracción de documentos, acceso a ciertas áreas de la compañía, etc. Así mismo, esta actividad le sirve al equipo de trabajo para diseñar cuestionarios de apoyo para la ejecución de las entrevistas

1.6.2.2.3 Entrevista

Esta actividad se refiere a la ejecución de la entrevista en sí, utilizando el cuestionario de apoyo y material necesario para el desarrollo de la misma según las condiciones expuestas en las pre entrevistas.

1.6.2.2.4 Resumen de los requerimientos

Esta actividad se refiere al registro de la información receptada en las entrevistas, misma que puede ser documentada siguiendo alguna metodología específica siempre y cuando sirva para la formalización del levantamiento de los requerimientos según los conocimientos, habilidades y experiencia del personal del proyecto. Existen muchas buenas prácticas al respecto como el uso de tarjetas CRC, modelos de casos de uso, entre otros cualquiera puede servir siempre que el equipo de trabajo este de acuerdo en su utilización y sirvan para el seguimiento de los requerimientos a lo largo del proceso de desarrollo. Así mismo es importante el adecuado archivo de los documentos y formatos que se vayan recopilando como parte de las entrevistas. Vale la pena aclarar que como resultado de la recopilación de requerimientos deberá estar, si es posible, el conjunto de preguntas que el DW o DM debe responder, de otra manera deberá constar como parte del resumen de requerimientos las medidas o indicadores que el DW o DM deberá mostrar

1.6.2.3 Análisis

En la fase de análisis el equipo de trabajo convierte los requerimientos recolectados en un conjunto de especificaciones que puedan apoyar el diseño. En lo abstracto, hay tres especificaciones principales de entrada para el DW o DM:

- Especificaciones de requerimientos de enfoque empresarial que delinear las fronteras de la información que debe comprender el DW o

DM. El enfoque empresarial determinará también la audiencia y sus requerimientos de información.

- Especificaciones de requerimientos de fuentes de datos que delimitan las fronteras de información disponible en las fuentes de datos actuales.
- Especificaciones de requerimientos de usuario final y acceso, las cuales definen cómo se utilizará la información del DW o DM. Junto con éstas se encuentra la especificación de los tipos de herramientas y técnicas de exhibición que usan.

En lo concreto, la etapa de análisis consiste en derivar modelos lógicos de datos para el DW o DM y definir los procesos necesarios para conectar las fuentes de datos, DW o DM y las herramientas de acceso del usuario final. Este propósito se lo lleva a cabo mediante las siguientes actividades

1.6.2.3.1 Generalidades

En esta fase se enuncian aspectos de carácter general relacionados con la organización como objetivos, misión y visión. Así mismo se determinan los roles y responsabilidades del personal una vez que el DW o DM se ponga en producción.

1.6.2.3.2 Arquitectura técnica

En esta fase determina la siguiente información:

1.6.2.3.2.1 Fuentes de datos

Las fuentes de datos son los repositorios de los cuales se extraerá la información, como parte de la información de las fuentes se debe especificar su ubicación, naturaleza y accionar, así como también, la estructura en sí de cada fuente. Además de la plataforma tecnológica comercial a la cual pertenecen.

1.6.2.3.2.2 Reglas de extracción, transformación y carga

Estas reglas determinan la especificación de paquetes de extracción, transformación y carga que permiten la interacción entre las fuentes de datos y el almacén de datos del DW o DM.

1.6.2.3.2.3 Políticas de uso

Las políticas de uso constan de información como, calendarios de carga, reglas de seguridad, limpieza de datos, políticas de respaldo y recuperación y en general información reglamentaria de cómo se utilizará el DW o DM.

1.6.2.3.3 Modelo lógico

El modelo lógico se constituye de dos etapas, la primera que se refiere a la especificación de los procesos de la organización y la segunda a la especificación del modelo conceptual de la base de datos del DW o DM; para estas actividades se puede utilizar técnicas de modelamiento específicas, como se indicó anteriormente, siempre que contribuyan al mejor entendimiento del contexto del proyecto y al buen desarrollo y comunicación del mismo.

Vale la pena aclarar que todas estas actividades se deben documentar de manera formal ya sea en formatos diseñados previamente por el equipo de desarrollo en conjunto con el administrador del proyecto o utilizando plantillas determinadas por metodologías adicionales utilizadas para robustecer el proyecto.

1.6.2.4 Diseño

En la fase de diseño, los modelos lógicos desarrollados en la fase de análisis se convierten en modelos físicos. Los procesos identificados en la fase de análisis para conectar las fuentes de datos con el DW o DM y estos con las herramientas de estación de trabajo del usuario final, se convierten en diseños para programas que realizarán las tareas requeridas por los procesos. También se identifican y detallan los procesos que requiere, de manera interna, cada bloque de la arquitectura del DW o DM. El diseño se lo realiza mediante las siguientes actividades:

1.6.2.4.1 Selección de la herramienta

En esta actividad se elige una herramienta o plataforma de software y hardware adecuada a las necesidades expuestas en el levantamiento de requerimientos. Vale la pena aclarar que hasta antes de este punto no se debería hacer ninguna referencia hacia una plataforma específica de tal manera que se asegure un desarrollo independiente de dicha plataforma.

1.6.2.4.2 Diseño de Físico del DM

El diseño físico del DW o DM se refiere a la creación de modelos físicos para el DW o DM, una vez que se crea un modelo lógico en la etapa de análisis; el modelo físico se divide en los siguientes puntos:

1.6.2.4.2.1 Creación del Modelo de datos físico

En este punto se crea un modelo Entidad Relación, con especificaciones físicas, es decir, con especificaciones de tipos de datos y tamaño de los mismos, para la base de datos del DW o DM que contenga todas las tablas de hechos y dimensiones en base al modelo lógico que se estableció en la etapa de análisis.

1.6.2.4.2.2 Determinación de la Arquitectura

En este punto se determina el diseño arquitectónico que tendrá el DW o DM en base a la arquitectura de referencia que se explicará mas adelante. Identificando las capas y bloques que son adecuados para la construcción del DW o DM, mismos que deberán estar contenidos en los distintos módulos de la aplicación.

1.6.2.4.2.3 Especificaciones de procesos

En este punto se diseña un modelo de procesos para el DW o DM que especifique la interacción y el flujo de información. Es importante aclarar que

dicho modelo se lo puede construir en base a técnicas específicas siempre que el equipo de trabajo este de acuerdo en la utilización de las mismas.

1.6.2.4.2.4 Modelo de extracción y transformación

En este punto se crea un modelo e extracción de datos desde las fuentes, las transformaciones que se realizan sobre dichos datos y la carga final en las tablas respectivas de la base de datos del DW o DM. Al igual que los anteriores modelos, este se puede también desarrollar en base a técnicas específicas o a formatos adaptados por el equipo de trabajo.

1.6.2.4.2.5 Esquema de almacenamiento

En este punto se especifica que tipo de esquema de almacenamiento se utilizará para la implementación del DW o DM. Esto se refiere a que si el DW o DM utilizará un esquema de tipo MOLAP, HOLAP o ROLAP. Cualquiera de estos esquemas debe contribuir al adecuado desempeño del DW o DM en base a las especificaciones hechas en la etapa de análisis.

1.6.2.4.3 Especificación de usuario final

En este punto se describe las características de la información que, el usuario final, desea consultar mediante: formatos de reportes, filtros para sus consultas, encabezados, precisión de cantidades numéricas y fechas y la aplicación del cuestionario, si se lo obtuvo, de donde se obtendrán los reportes

que el DW o DM debe cubrir. Esta información ya fue recopilada en la etapa de levantamiento de requerimientos, pero que se especifican de manera formal en este punto.

1.6.2.5 Implementación

1.6.2.5.1 Construcción del Data Warehouse o Data Mart

La fase de construcción es responsable de implementar físicamente los diseños desarrollados durante la fase de diseño. Por medio de decisiones bien pensadas, es posible integrar una solución de DW o DM bastante rápido.

La construcción del DW o DM es similar a la construcción de un sistema de base de datos relacional grande. La construcción consta de las siguientes actividades:

1.6.2.5.1.1 Creación de la Base de Datos

En este punto se crean las tablas de la base de datos del DW O DM de acuerdo al modelo Entidad Relación que se diseñó, tomando en cuenta las especificaciones físicas determinadas en la etapa de diseño.

1.6.2.5.1.2 Creación de paquetes de Extracción, Transformación y Carga

En este punto se crean los paquetes que extraerán la información de las fuentes, transformarán los datos extraídos y los cargarán en las respectivas

tablas de la base de datos del DW o DM de acuerdo al modelo de extracción y transformación establecido en la etapa de diseño

1.6.2.5.1.3 Aplicación de las reglas de limpieza de datos

En este punto se aplican sobre los datos cargados sobre la base de datos del DW o DM, las reglas de limpieza de los datos, las cuales fueron definidas en la etapa de análisis.

1.6.2.5.1.4 Creación de Cubos

En este punto se crean los cubos multidimensionales que son necesarios para la obtención de los reportes determinados en la etapa de diseño. Con las respectivas dimensiones y medidas

1.6.2.5.2 Desarrollo de la Aplicación de Usuario Final

Desarrollo de la aplicación que el usuario final utilizará para visualizar los reportes que arroja el DW o DM, aplicando las especificaciones establecidas en la etapa de diseño.

1.6.2.6 Pruebas

En esta etapa se pueden aplicar un gran número de pruebas según especificaciones de varios autores siempre que permitan demostrar la

funcionalidad y la comprobación de las características deseables del DW o DM por parte de la organización. Es importante aclarar que se pueden aplicar pruebas intermedias conforme se crean cada uno de los módulos de la aplicación de DW o DM

1.6.2.7 Despliegue

La etapa de despliegue tiene que ver principalmente con los retos de instalación, puesta en marcha y uso de la solución de DW o DM. Es probable que la organización ya tenga una experiencia con la implantación de soluciones informáticas lo cual acelerará esta etapa. Existen determinadas buenas prácticas para la etapa de despliegue que se pueden haber ido desarrollando en base a casos de éxito y que se pueden adoptar por parte del equipo de desarrollo para el aseguramiento del correcto despliegue del DW o DM. En este punto también se puede establecer planificaciones para futuras versiones.

CAPITULO II

2 Análisis y Diseño.

2.1 Planificación

2.1.1 Descripción del Problema

En los Centros de Orientación Juvenil que se visitaron se pudo comprobar que comparten una problemática y común necesidad de obtener información primordialmente estadística que les facilite el análisis en las áreas socioeducativa, psicológica y delictiva, primordialmente, para la gestión de proyectos de carácter estratégicos en bien de la niñez y la juventud. Dichos Centros mantienen su información en hojas electrónicas y solo uno en una bases de datos relacional, ya que posee un software para el almacenamiento de información, pero en general no poseen conocimientos sobre lo que es un DW o un DM y los reportes con información estratégica actualmente se los elabora en forma manual solicitando reportes específicos del sistema para posteriormente hacer los cálculos y finalmente elaborar reportes de carácter estratégico en hojas electrónicas. Éste es el caso del Centro de Orientación Juvenil “Virgilio Guerrero” el cuál mantiene una relativa estructura de sistemas, ya que posee el software para en ingreso de información de los jóvenes infractores. Se utilizará al Centro de Orientación Juvenil “Virgilio Guerrero” como caso de estudio.

2.1.2 Elección Estrategia de Implementación

Por lo expuesto en la explicación inicial sobre los Centros de Orientación Juvenil y en el caso particular del Centro de Orientación Juvenil “Virgilio Guerrero”; la tecnología del DM no está comprendida del todo ni ubicada dentro del nivel estratégico en el cual debería estar; por lo que se escoge una estrategia de implementación del tipo BOTTOM-UP¹¹ lo cual nos significa el mantener como hilo conductor al DM; es decir, que se desarrolle el DM en base a la problemática estadística genérica de tal manera que arroje reportes adecuados a dichas necesidades los cuales posteriormente permitan demostrar las bondades del DM como herramienta de carácter crítico; vale la pena aclarar que dentro del alcance de este trabajo esta solo la implementación del DM mas no cubre la etapa de despliegue. Adicionalmente hay que hacer énfasis en la necesidad de que la metodología que se use para desarrollar el DM deberá permitir la estrategia de implementación y debe contribuir a un desarrollo rápido.

2.1.3 Elección de la Metodología de Desarrollo

Debido al estado tecnológico de los Centros de Orientación Juvenil ejemplificados en el Centro de Orientación Juvenil “Virgilio Guerrero”, como se indicó anteriormente, y a la naturaleza de la estrategia de implementación escogida, es recomendable utilizar una metodología ágil y específica para el desarrollo del DM y que además utilice estándares y técnicas formales y disciplinadas que se adapten a la necesidad de una implementación rápida.

¹¹ GILL Harjinder, RAO Prakash, Data Warehousing La Integración De La Información Para La Mejor Toma De Decisiones, Ediciones Que, 1996

En base al análisis de las metodologías investigadas, mediante las cuales se puede construir un DM, se decide que la mejor opción para el desarrollo de este proyecto es la utilización específica de la metodología KIMBALL¹² la cual cumple con el ciclo de vida de desarrollo de software en esencia, esta desarrollada específicamente para construcción de DW y DM; y, se orienta en el diseño y principalmente en los datos que es lo que conviene para el desarrollo del presente proyecto y está debidamente documentada en varios libros que forman parte de la bibliografía y que se resume en el punto 1.6 del Marco Teórico; vale la pena indicar que esta metodología además se implementa independientemente de alguna herramienta o plataforma tecnológica en particular y describe las etapas y actividades que se deben ir cumpliendo para la creación del DM, deja libre la forma en que se va a documentar y modelar por lo que en este punto se aclara que se utilizará información bibliográfica adicional para cumplir con la metodología escogida¹³.

A continuación un breve resumen de las etapas de la metodología de Kimball que deben cumplirse en cada iteración, incluyendo en las que se requiere de la elaboración de modelos y como se va a modelar para el presente proyecto:

¹² KIMBALL Ralph, The Data Warehouse Lifecycle Toolkit, John Wiley & Sons Inc, 1998.

¹³ PRESSMAN Roger. "Ingeniería de software, un enfoque práctico", editorial McGraw-Hill, 2005

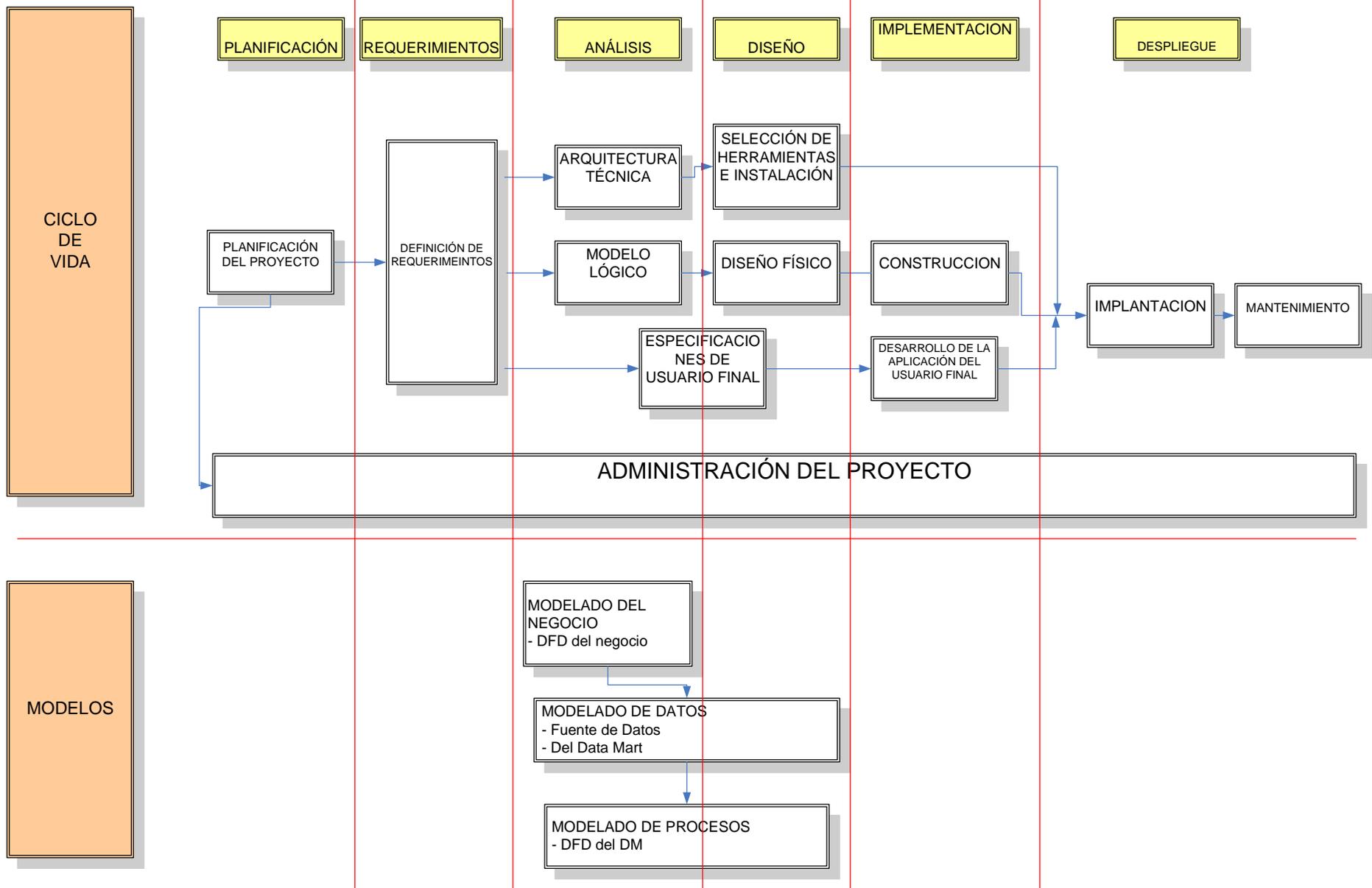


Figura 2.1: Etapas de la Metodología Kimball

2.1.4 Justificación de la Metodología

Es primordial definir el concepto de metodología a fin de justificar la elección mostrada anteriormente.

Metodología: Según Booch, es una colección de métodos aplicados a lo largo del ciclo de vida del Desarrollo de Software y unificados por algún enfoque filosófico o general.

Por lo expuesto se puede determinar que la Metodología Kimball¹⁷ cumple la definición de metodología ya que se unifica en torno a la orientación al diseño de la base de datos del DW o DM y a los datos que debe contener esta. Además define el conjunto de métodos especializados para la construcción de un DW o DM dejando abierta la posibilidad de uso de distintos modelos según las necesidades y características del proyecto por lo que abarca el conjunto de características que debe cumplir una metodología.

Cabe indicar que la metodología Kimball permite el uso de diagramas como el Diagrama de Flujo de Datos y el Entidad Relación, que son fáciles de entender para el personal no técnico del Centro de Orientación Juvenil “Virgilio Guerrero” y que apoyados con un Repositorio adecuado de Metadatos favorece las características del presente proyecto y posterior mantenimiento del mismo por parte del Centro. Por lo tanto se utilizarán dichos modelos en las etapas del ciclo de vida según se muestra en la Figura 2.1.

¹⁷ KIMBALL Ralph, The Data Warehouse Lifecycle Toolkit, John Wiley & Sons Inc, 1998

2.1.5 Formación del Equipo de Trabajo

El equipo de trabajo para el desarrollo del presente proyecto queda definido de la siguiente manera:

ROL	PERSONAL ASIGNADO
Administrador de proyecto	➤ Marco Luna
Desarrollador, Analista e Implementador	➤ Marco Luna
Contraparte y Personal de pruebas	➤ Director del Centro ➤ Psicólogo Asignado ➤ Educador Asignado ➤ Coordinador de Información del Centro

Tabla 2.1: Asignación de roles

2.1.6 Plan Inicial del Proyecto

Luego de la reunión inicial con la Contraparte se desarrollan los siguientes puntos como parte del Documento Inicial del Proyecto:

2.1.6.1 Establecimiento del Contexto Inicial y Futuro

2.1.6.1.1 Mercado de Destino

El DM deberá orientarse a una audiencia estándar que no ha comprobado aún los beneficios de un DM y que basa sus necesidades de información actual en bases de datos relacionales y reportes específicos y rígidos que vuelven el proceso analítico más laborioso.

2.1.6.1.2 Plataformas de uso actual y futuras

El DM deberá desarrollarse en base a la tecnología actual adquirida por el Centro la cual está establecida en base al siguiente esquema:

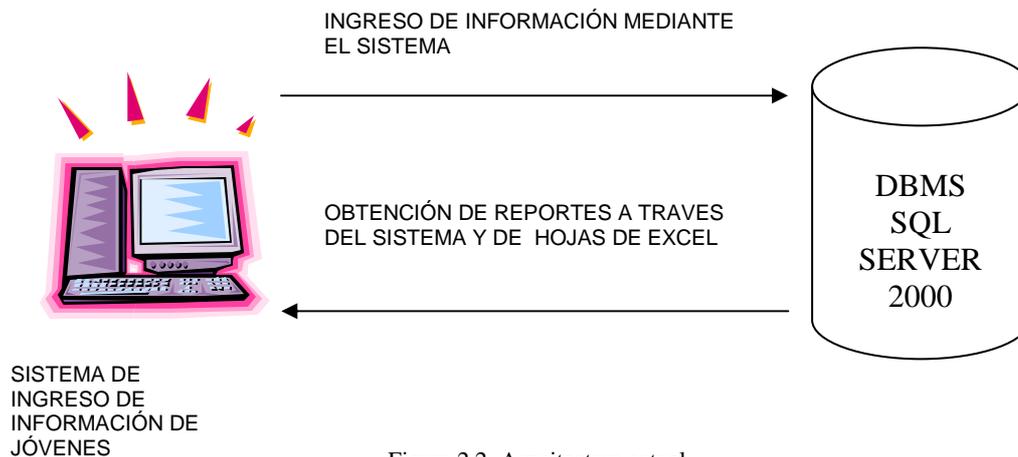


Figura 2.2: Arquitectura actual

Como se puede observar en el gráfico, la arquitectura actual se sustenta sobre plataformas Microsoft íntegramente y en forma bastante básica por lo que se utilizará esta misma plataforma sin adquisición alguna de ningún software específico adicional para que la implementación del DM no genere demasiado impacto económico y de conocimientos para las personas que utilizarán el DM. Se justifica por el hecho de que el DM será una herramienta nueva que deberá desarrollarse de tal manera que demuestre su necesidad de uso como herramienta crítica.

El DM luego de ser implementado deberá cumplir con la siguiente arquitectura:

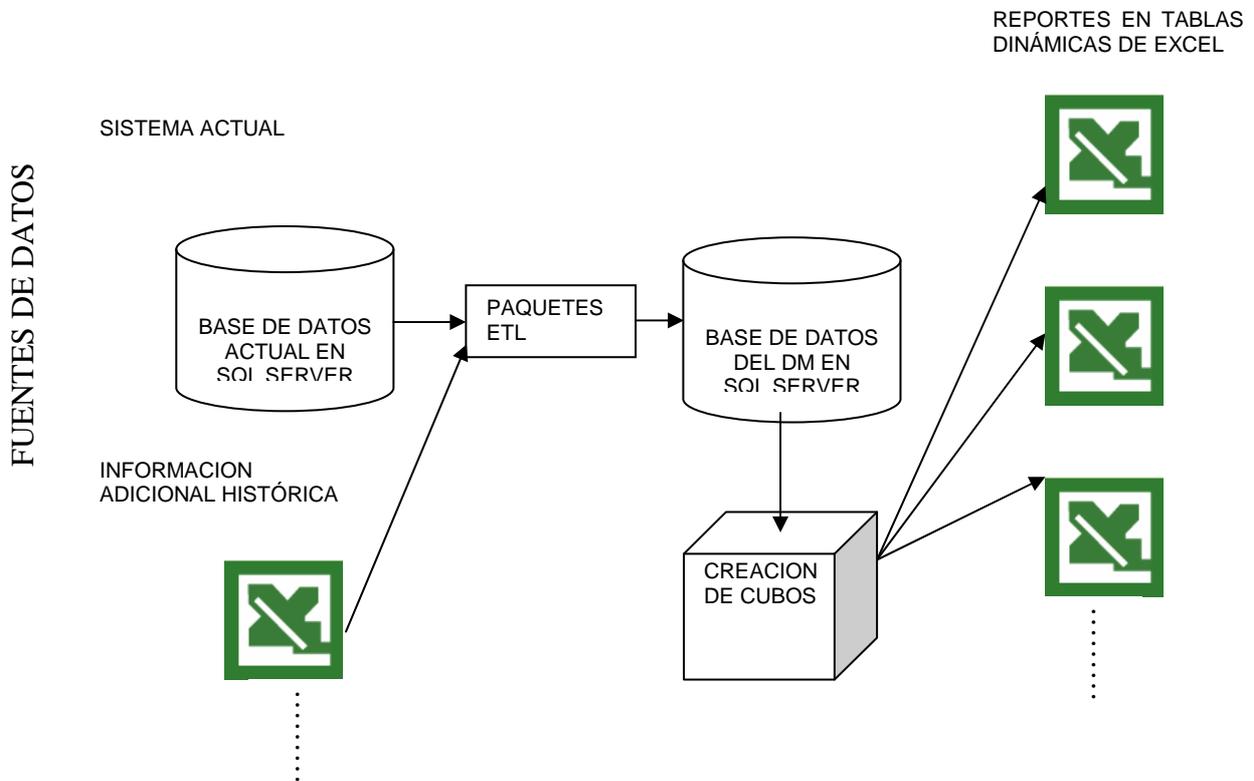


Figura 2.3: Arquitectura futura

2.1.6.2 Estimación del tiempo de inicio de trabajo para el DM

El DM es una herramienta que se va a implantar de forma que demuestre primero su utilidad y luego plantee su utilización definitiva o incluso un futuro crecimiento; por lo que la estimación del tiempo para el DM está fuera del alcance de este trabajo y no se planteará un tiempo exacto ni estimado.

2.1.7 Calendario y Presupuesto

El presente trabajo será desarrollado según el calendario que consta en el Temario de presentación del presente proyecto y con un presupuesto que se indica a continuación:

ITEM	VALOR (en USD)
SOFTWARE	0
HARDWARE	0
BIBLIOGRAFIA	300
MATERIAL DE OFICINA	700

Tabla 2.2: Presupuesto estimado

Vale la pena aclarar que el presupuesto no implica la adquisición de software o hardware adicionales o específicos para este trabajo ya que se utilizarán los ya adquiridos por el Centro.

2.1.8 Catalogo de Metadatos

Los Catálogos de Metadatos serán elaborados y administrados en base de una la misma herramienta que viene como parte de MS SQL Server y se utilizará solo para el modelamiento una herramienta CASE; Los Metadatos se mostrarán en forma impresa mediante el uso de los mismos utilitarios de oficina del Centro y que se especificarán en la etapa de requerimiento. Los metadatos se describirán en los siguientes anexos:

- ANEXO A: catalogo de la base de datos actual
- ANEXO B: catalogo de la base de datos del DM
- ANEXO G: Diagramas de Flujos de Datos de los Procesos Identificados
- ANEXO C: modelo de transformaciones

2.2 Requerimientos.

Los requerimientos que se describen a continuación se los ha levantado en base a entrevistas realizadas con los personeros del Centro de Orientación Juvenil “Virgilio Guerrero” y serán tomadas como punto de partida para el DM en forma estándar. Saliendo fuera del alcance de este trabajo el despliegue del DM.

2.2.1 Agenda de citas con el usuario

Se estableció la siguiente agenda en común acuerdo con la contraparte:

FECHA INICIO	FECHA FIN	PERSONA	INFORMACIÓN A SER RECOPIADA
02/01/2006	11/01/2006	Director del Centro	<ul style="list-style-type: none">➤ Información General➤ Requerimientos de propietario➤ Información Estratégica
12/01/2006	17/01/2006	Psicólogo asignado	<ul style="list-style-type: none">➤ Procesos que se desarrollan en el área➤ Información psicológica necesaria para el DM➤ Levantamiento de indicadores
18/01/2006	23/01/2006	Educador asignado	<ul style="list-style-type: none">➤ Procesos que se desarrollan en el área➤ Información socio educativa necesaria para el DM➤ Levantamiento de indicadores
24/01/2006	29/01/2006	Coordinador de	<ul style="list-style-type: none">➤ Procesos

		información	relacionados con el área de análisis delictivo ➤ Procesos de coordinación con las Entidades de Control ➤ Levantamiento de indicadores
30/01/2006	10/02/2006	Equipo de trabajo completo	➤ Elaboración del plan inicial ➤ Establecimiento de calendarios ➤ Establecimiento de indicadores finales a tomar en cuenta para el DM

Tabla 2.3: Agenda de citas con el usuario

Luego de efectuar las entrevistas expuestas en la agenda se obtuvieron los siguientes requerimientos¹⁸:

2.2.2 Resumen de los requerimientos

2.2.2.1 Requerimientos del Propietario

- Que el DM resuelva la necesidad de información estratégica para la gestión de proyectos a favor de la niñez y la adolescencia mejorando el proceso actual según los indicadores que se manejan en las áreas
 - Psicológica
 - Socio Educativa
 - Delictiva para gestión

¹⁸ GILL Harjinder, RAO Praskash, Data Warehousing La Integración De La Información Para La Mejor Toma De Decisiones, Ediciones Que, 1996.

- Disponer de una herramienta que agilice el proceso de análisis de información estadística relativa al trabajo que desarrolla el Centro de Orientación Juvenil “Virgilio Guerrero”
- Que el DM que se desarrolle permita observar las bondades de una nueva tecnología que sea beneficiosa y que se pueda ampliar o explotar de mejor manera en el futuro inclusive en otros Centros de Orientación amigos.
- Que el tiempo de desarrollo del DM sea el más corto posible y que se pueda experimentar y sacar conclusiones sobre su uso en poco tiempo.
- Que las herramientas en las que se desarrolle el DM sean las mismas de las cuales dispone el Centro.

2.2.2.2 Requerimientos del Desarrollador

Los requerimientos del desarrollador hacen consideraciones a la utilización de las herramientas ya existentes en el Centro.

2.2.2.2.1 Herramientas

- SQL SERVER 2000: DBMS ya adquirido por el Centro y que se utilizará par administrar tanto la BDD actual fuente como la BDD del DM; además de la administración de los metadatos
- MS Excel 2003: Utilitario de oficina que se utiliza actualmente en el Centro para distintos fines y que se lo empleará como herramienta de usuario final para emisión de reportes tanto textuales como gráficos

- Data Transformation Service: utilitario que forma parte del SQL SERVER 2000 y que se utilizará para la creación de los paquetes de extracción, transformación y carga.
- MS Analysis Services: componente de MS SQL SERVER que se utilizará para la creación de los cubos de información y de administración de los metadatos correspondientes
- SQL SERVER AGENT: Servicio de SQL SERVER 2000 que se utilizará para calendarizar la ejecución de las actividades del DM
- Enterprise Manager: Utilitario del SQL SERVER 2000 que se utilizará para la administración general del DM.
- MS Word 2003: Utilitario de oficina que se utiliza en el Centro y que será empleado para el levantamiento de la documentación formal del presente trabajo.
- SYBASE POWER DESIGNER: utilitario que se utilizará para el reporte impreso de los metadatos.

2.2.2.2 Distribución

Debido a que el alcance de este trabajo solo abarca la creación del DM en base a requerimientos genéricos, la etapa de distribución del DM queda fuera de dicho alcance y no se definirá algún método en particular; pero vale la pena resaltar que el Centro no cuenta con una red de datos por lo que lo mas probable sea que el DM se los ejecute en un solo computador que saque reportes y estos sean distribuidos en forma impresa a los interesados.

2.2.2.2.3 Disponibilidad

El DM debe estar disponible todo el tiempo, con información actualizada y real. Por lo que se deberá calendarizar de forma automática dicha actividad de manera que ésta no esté enclavada a alguna persona en particular. Para lo cual se utilizará una herramienta de automatización de este proceso luego de un proceso previo de control de calidad de la información que será desarrollada durante el proceso de desarrollo del DM y que se lo establecerá como un proceso periódico.

2.2.2.3 Requerimientos del Usuario Final

- Disponer de una herramienta que permita el análisis ágil y dinámico de reportes estadísticos sin tener que depender de herramientas o reportes específicos y rígidos.
- Que el DM sea una herramienta que favorezca el proceso de obtención de reportes para la entrega de información a las entidades de control del Centro y las contrapartes.
- Que la información que arroje el DM pueda mejorar la visualización de los indicadores psicológicos, socio educativos y delictivos.
- Que todo el entorno del DM se lo haga con herramientas fáciles de utilizar y de comprender o aprender del ser el caso.

2.2.3 Requerimientos de Arquitectura¹⁹

2.2.3.1 Arquitectura de Datos

La arquitectura de datos necesaria, será establecida luego del análisis de los indicadores adecuados para la toma de decisiones en el Centro y está planificado definirlo en detalle en el ANEXO B:

2.2.3.2 Arquitectura de la Aplicación

La arquitectura de aplicaciones deberá estar establecida de la siguiente manera:

- Modulo de Administración de datos: que administrará las actividades de extracción, transformación, carga y limpieza de los datos
- Modulo de Fuentes de Datos: el cual administrará la base de datos que alimentará el DM
- Modulo de DM: el cual administrará la base de datos del DM
- Modulo de Acceso y Uso: el cual contemplará la administración de las herramientas de explotación del DM por parte del usuario final
- Módulo de Infraestructura: el cual controlará la correcta interacción entre las herramientas que conforman el contexto del DM.
- Módulo de Administración de Metadatos: el cual administrará todos los metadatos involucrados en el DM

Estos módulos interactuarán de acuerdo a la arquitectura de referencia que se mencionó en el punto 1.2.4. Vale la pena aclarar que no se contempla el

¹⁹ GILL Harjinder, RAO Praskash, Data Warehousing La Integración De La Información Para La Mejor Toma De Decisiones, Ediciones Que, 1996

módulo de transporte debido a que el Centro no posee una red de datos y el único mecanismo pensado para dar a conocer los resultados del DM será la impresión de los reportes o la entrega de los mismos en medio magnético lo cual no constituye un verdadero módulo de la aplicación.

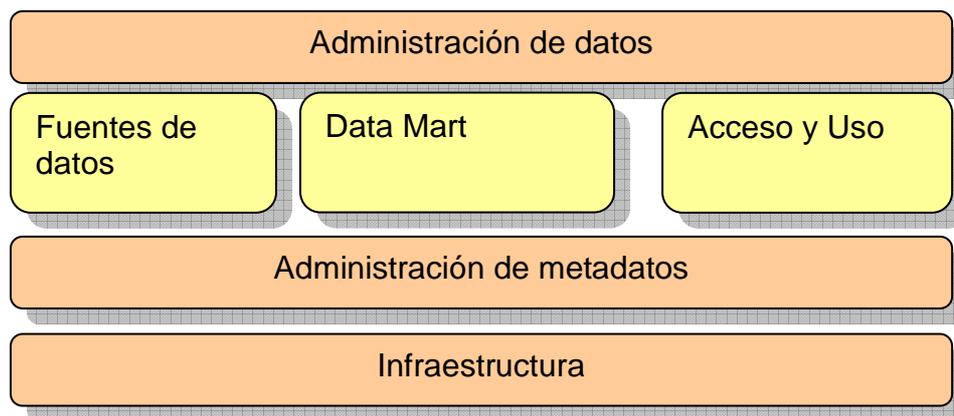


Figura 2.4: Arquitectura de módulos requerida

En la etapa de diseño se describirá como se implementará e integrará cada uno de los componentes de esta arquitectura

2.3 Análisis

2.3.1 Generalidades

2.3.1.1 Directivas del Negocio

- Entidades Controladoras
- Jóvenes Infractores
- Situación socio económica del país
- Políticas de Estado

2.3.1.2 Objetivo del Negocio

Se debe considerar como objetivo global el siguiente:

- “Construir una propuesta educativa, para niños, niñas y jóvenes infractores que garantice el ejercicio y defensa de los derechos a la vida, desarrollo, protección, y participación, a través de la implementación de estrategias pedagógicas que favorezcan la formación del niño(a) como persona responsable, autónoma, y comprometida consigo misma, con su pueblo y con el mundo en que vive”²⁰.

2.3.1.3 Roles y Responsabilidades

A continuación se enlistan los roles y responsabilidades del personal involucrado con el que iniciará el DM:

ROL	RESPONSABILIDAD
DIGITADOR	➤ Ingreso de información en los sistemas actuales
OPERADOR DEL DATA MART	➤ Monitoreo de la integridad y consistencia de la información ingresada ➤ Ejecución de la carga de información de las fuentes a la base de datos del DM ➤ Elaboración y comunicación de la información
USUARIO FINAL	➤ Recepción y uso de los reportes arrojados por el DM
ADMINISTRADOR	➤ Control y manejo de todos los componentes de la arquitectura final del DM

Tabla 2.4: Roles y responsabilidades involucradas

²⁰ Objetivo sacado de la publicación: REENCUENTRO del Centro de Orientación “Virgilio Guerrero”

2.3.2 Arquitectura técnica

2.3.2.1 Fuentes de Datos

Las fuentes de datos del Sistema de DM está definida en detalle en el ANEXO A, a continuación se muestra el Diagrama Entidad Relación de dicha fuente:

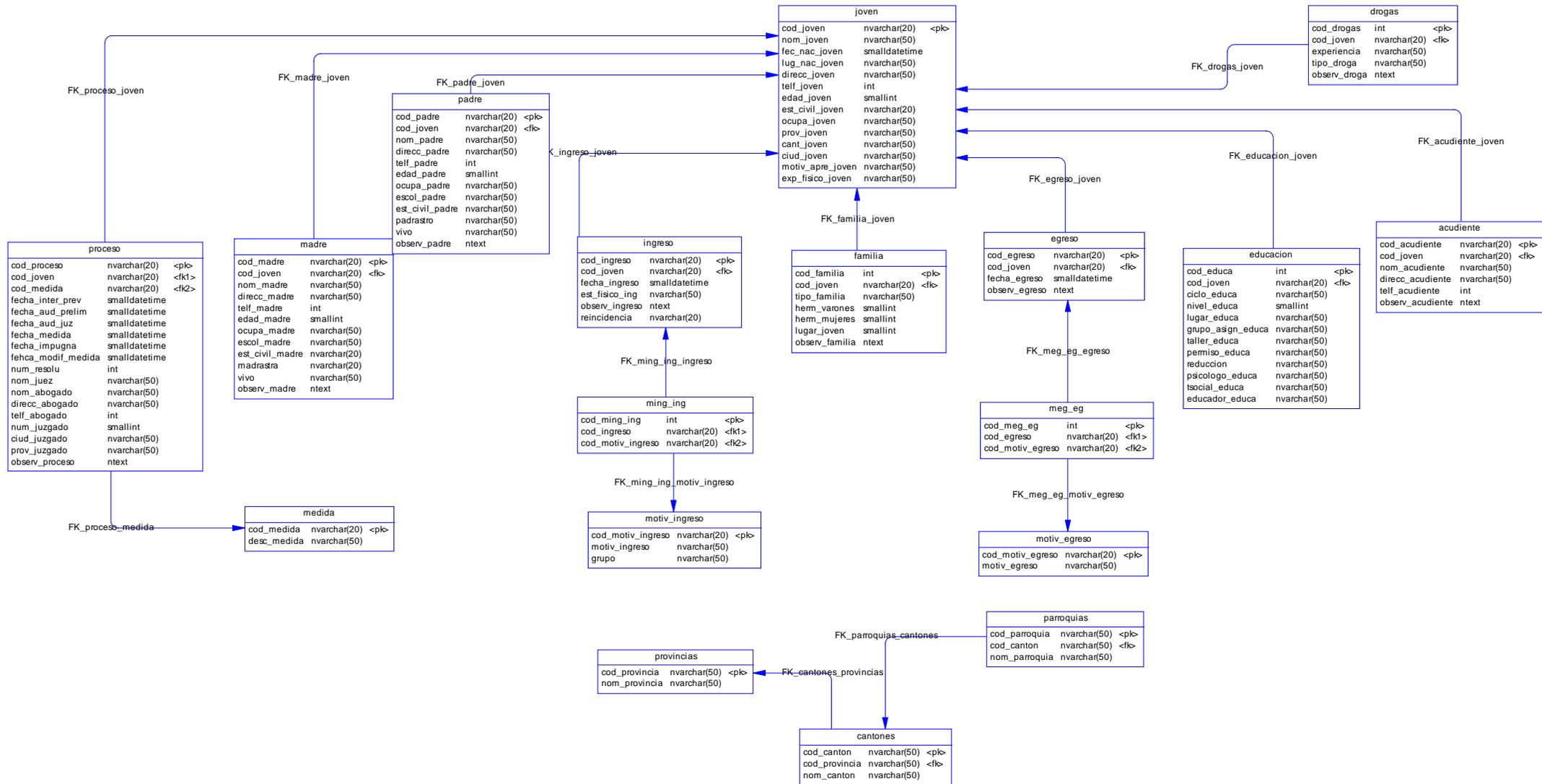


Figura 2.5: Arquitectura de la Fuente de Datos

2.3.2.2 Reglas de extracción, transformación y carga de datos

Reglas de Extracción:

- Se debe crear un objeto de acceso a datos (ODBC) para acceder a cada fuente de información
- Se debe crear un objeto de acceso a datos (ODBC) para acceder a la base de datos del DM
- Los objetos de acceso a datos (ODBC) deben contener la información de seguridad y conexión necesarios para acceder a las fuentes para las cuales han sido creados

Reglas de transformación y carga:

- Se deben crear paquetes de transformación y carga (DTS) para cada una de las tablas de dimensión y hechos definidas en la arquitectura de datos del DM luego del análisis y diseño respectivo.

Vale la pena indicar que estas reglas de transformación y carga se definen específicamente en el ANEXO C donde se planifica colocar el modelo de transformaciones para este proyecto y de acuerdo a la implementación del módulo respectivo según la arquitectura de la aplicación especificada en el punto 2.4.2.2.

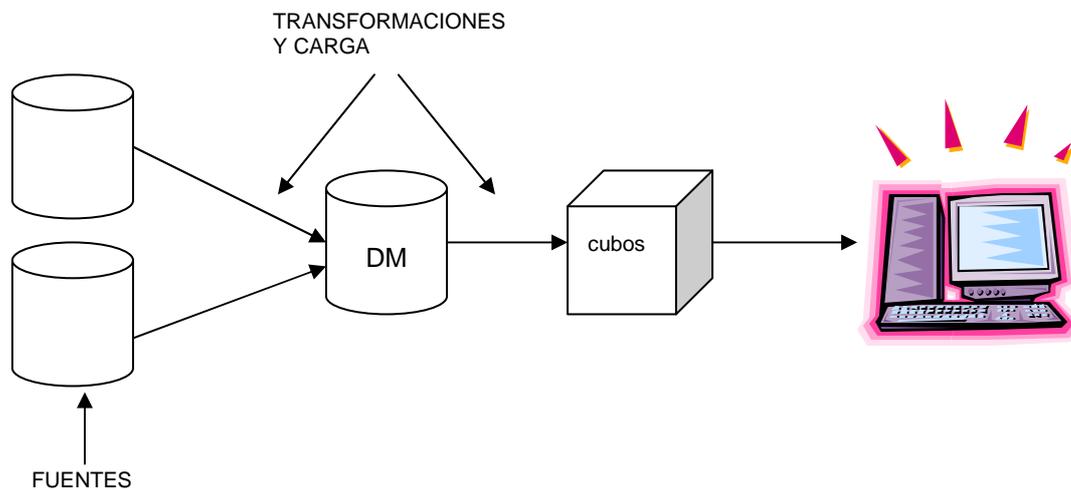


Figura 2.6: Esquema de fuentes y transformaciones

2.3.2.3 Políticas de Uso

2.3.2.3.1 Como se utilizara el DM

- El DM será de uso exclusivo del personal asignado con los roles descritos en el apartado 2.3.1.3 y la seguridad del DM será de responsabilidad de dichos roles.
- El DM deberá estar actualizado con una frecuencia diaria con responsabilidad única y exclusiva del personal asignado con el rol de OPERADOR DEL DATA MART.
- El USUARIO FINAL solo tendrá acceso a la información del DM a través de las herramientas de explotación autorizadas y mediante las opciones que dichas herramientas les brindan, cualquier necesidad adicional de información deberá informársele al SUPERVISOR.
- La información será divulgada a personas no asignadas con ningún rol bajo publicación autorizada por el ADMINISTRADOR del DM.

2.3.2.3.2 Calendarización de carga del DM

ACTIVIDAD	PERIODISIDAD	HORARIO
Cierre de ingreso de información a la base de datos operativa	Todos los días	19:00
Corrida de los paquetes DTS	Todos los días	19:30
Comprobación de consistencia	Todos los días	08:00 del siguiente día

Tabla 2.4: Roles Calendarización de la Carga

2.3.2.3.3 Reglas de Seguridad

- El ingreso al DM será única y exclusivamente a través de un computador central que almacenará tanto la información táctica como la estrategia
- El computador será de uso exclusivo del rol SUPERVISOR y ADMINISTRADOR
- El computador en que se encuentra residente el DM deberá autenticar a los usuarios autorizados bajo un esquema de nombre de usuario y clave de acceso a nivel de base de datos.
- El USUARIO FINAL solo tendrá acceso a los reportes generados y publicados por el SUPERVISOR, en medio magnético o en un repositorio central si se implementare una ambiente de red.

2.3.2.3.4 Políticas de Respaldo y Recuperación

- El respaldo de la información se lo hará a las 12:00 PM todos los días almacenando los archivos de respaldo respectivo en un directorio específico dentro del computador que aloja al DM y con su transferencia correspondiente a medio magnético también en forma diaria; dicha actividad se la realizará de manera automática a través de algún tipo de agente que deberá contener la herramienta en la que se desarrollará el DM.
- Los archivos de respaldo que se alojan en el directorio respectivo dentro del computador tendrán un período de residencia de un mes luego del cual serán borrados; no así su correspondencia en medio magnético la cual no será borrada y será almacenada en un lugar fuera del sitio de residencia del computador central.
- En general el período de vigencia de la información en el medio magnético, será establecida en forma mancomunada con las contrapartes y entidades reguladoras de la actividad del Centro.
- En caso de caídas del sistema será responsabilidad del OPERADOR DEL DATA MART el establecer la necesidad de restauración total o parcial de la BDD del DM mediante la especificación de la fecha del archivo de respaldo mas adecuado a dicha recuperación tomando como primera prioridad los archivos que se encuentran residentes en el computador central.

Vale la pena aclarar que esta estimación de reglas y políticas se las ha hecho enmarcándose en el contexto tecnológico del Centro por lo que no se

especifica la obligatoriedad de una red ni de un servidor central con capacidades de hardware y software específicas; todos esto sin dejar de recomendar que estas estructuras deberían ser las adoptadas por el centro para asegurar la calidad del servicio al cliente²⁴ interno y externo.

2.3.3 Modelo lógico

En este punto se plantea la necesidad de establecer los procesos a fin de conocer el modelo del negocio y posteriormente determinar el modelo conceptual de la base de datos del DM, por lo que se plantea la utilización de Diagramas de Flujo de Datos para la especificación de procesos y diagrama Entidad Relación para especificar el modelo conceptual, no sin antes indicar que como resultado del análisis de los procesos, se establecerá una tabla de indicadores, lo cuales se levantaron en la etapa de recopilación de requerimientos, ubicándolos en el contexto de cada uno de los proceso del negocio, manteniendo de esta manera la consistencia entre los requerimientos y la implementación. Este mecanismo nos permite reemplazar las preguntas del cuestionario que el DM debe responder por los indicadores.

2.3.3.1 Diagrama de flujo de Datos

Los Diagramas de Flujo de Datos, que muestran el modelo de negocio, están definidos de forma detallada en el ANEXO G de este proyecto tanto de los procesos empresariales como de los propios del DM. El DFD del negocio explotado de nivel 0 y de nivel 1 se especifican a continuación:

²⁴ Se considera cliente a la persona que se sirve de la información y bondades del DM, mas no del concepto comercial de un cliente regular.

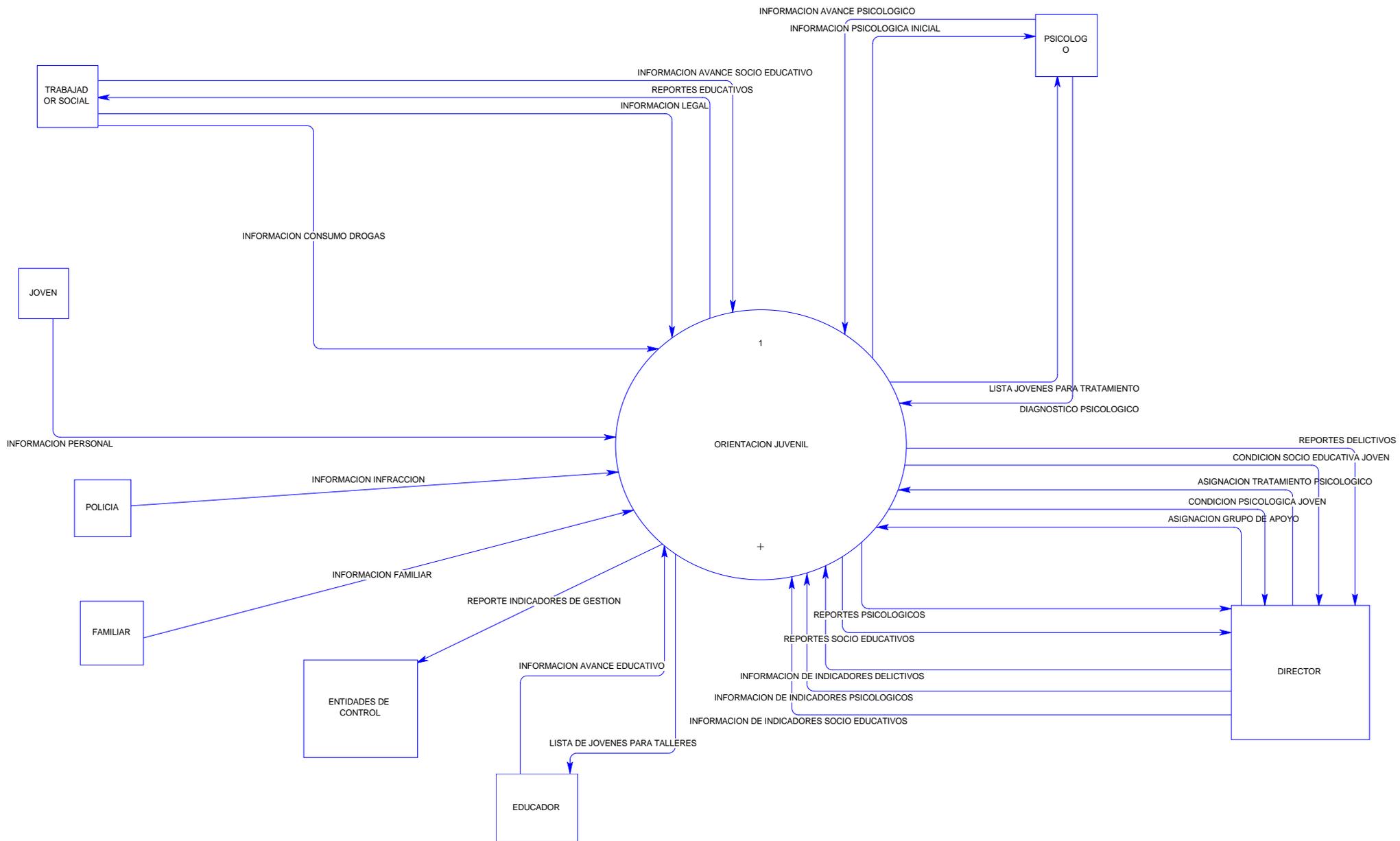


Figura 2.6: Diagrama de contexto del Negocio (DFD nivel 0)

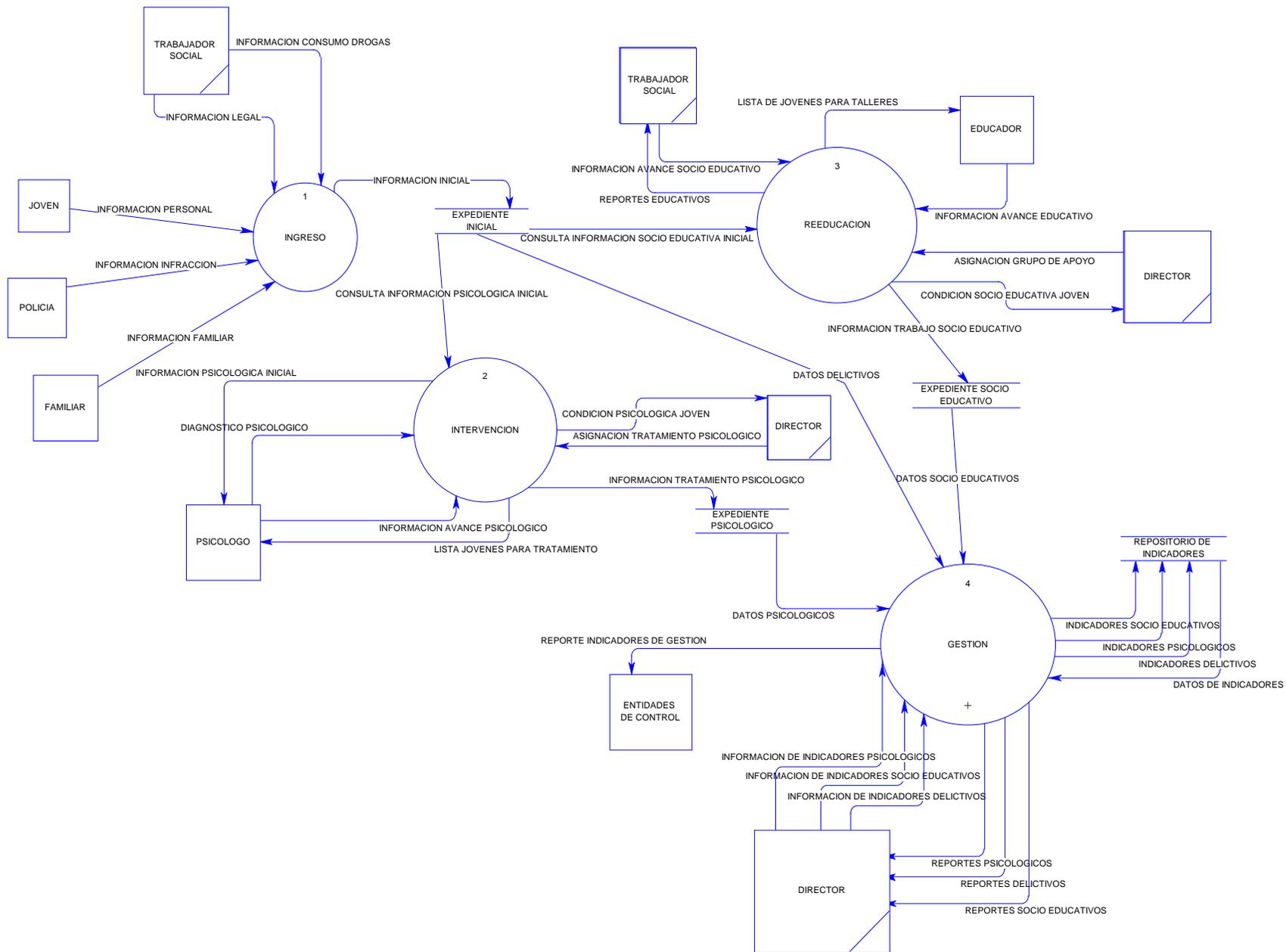


Figura 2.7: Diagrama de flujo de Datos nivel 1

De tal forma que se visualice mejor el DFD anterior lo ordenaremos en sus principales procesos aclarando que no es un refinamiento a un siguiente nivel de descomposición:

Proceso de Ingreso:

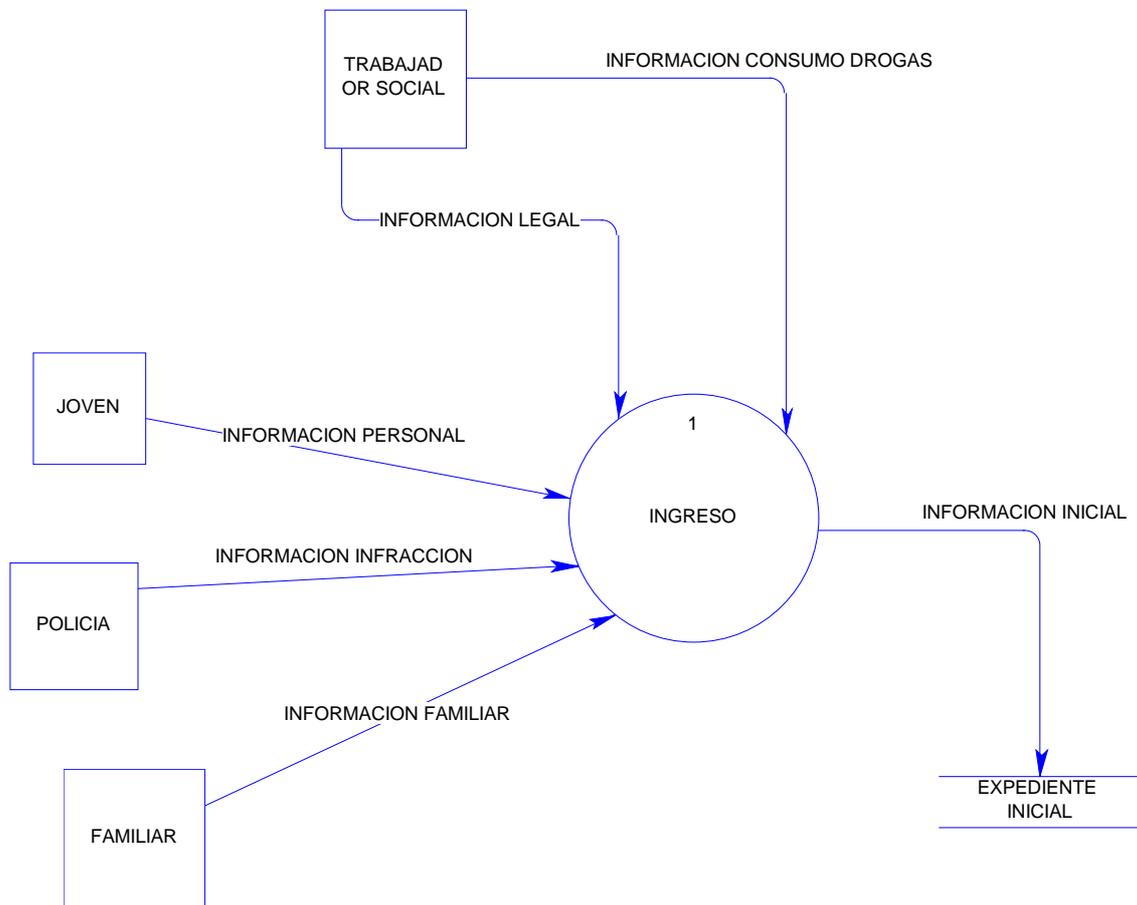


Figura 2.8: Diagrama de flujo del proceso de Ingreso

Proceso de Intervención:

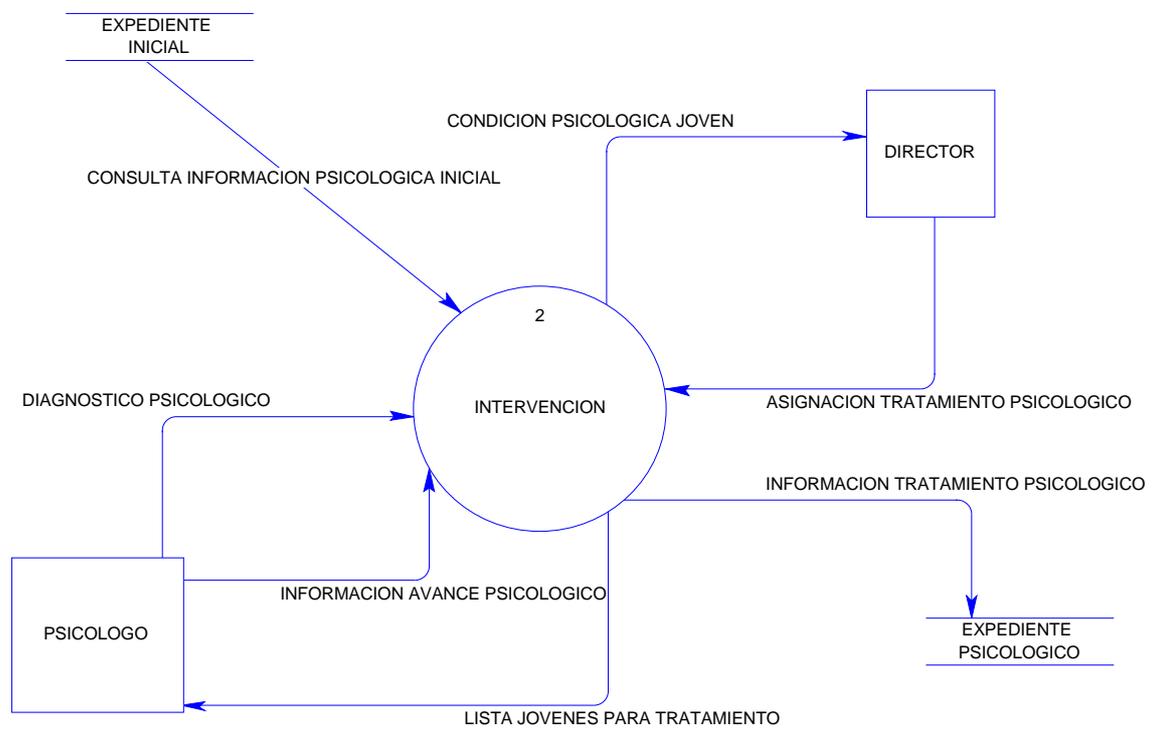


Figura 2.9: Diagrama de flujo del proceso de Intervención

Proceso de Reeducación:

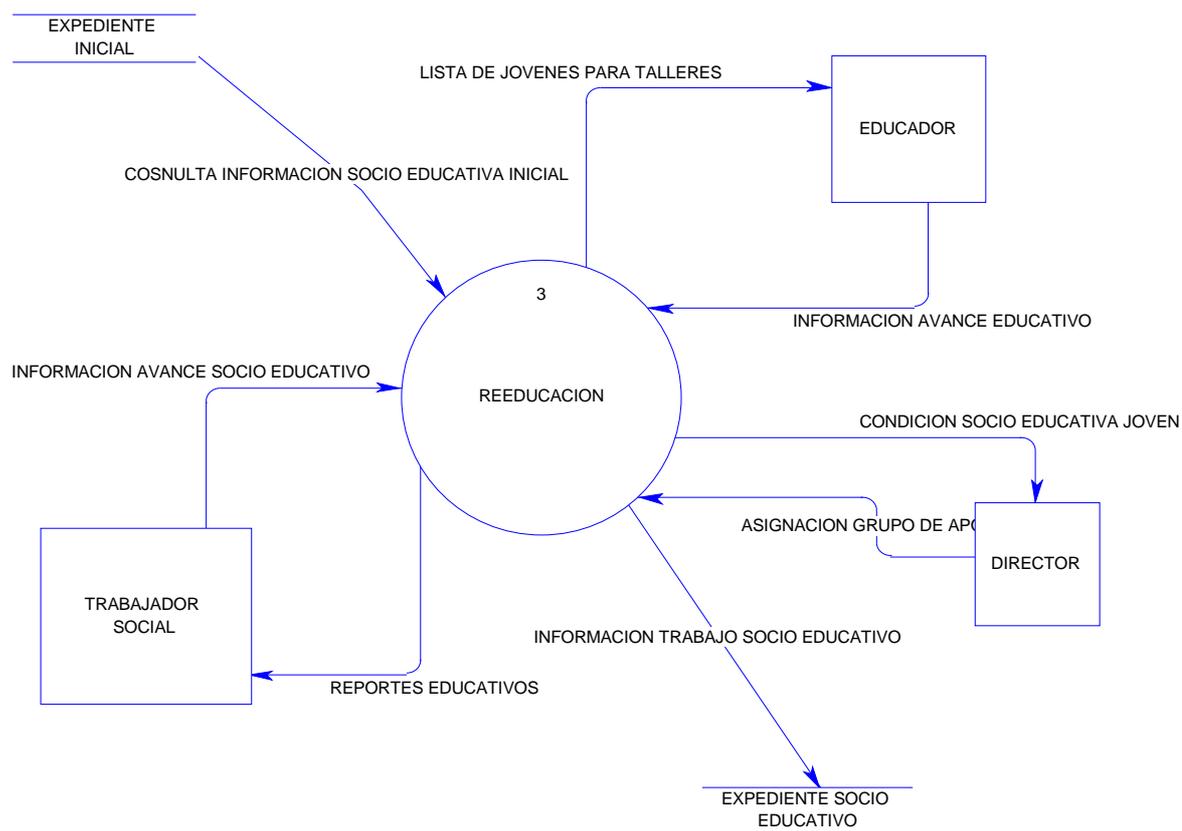


Figura 2.10: Diagrama de flujo del proceso de Reeduccion

En este punto se debe indicar que, si bien es cierto se puede seguir refinando los diagramas de los procesos de Intervención y Reeduación, a fin de conocer como se desarrollan, estos son de carácter operativo por lo que se explotará la burbuja del proceso de Gestión, el cual es de carácter estratégico y contribuye con los indicadores que se solicitaron en la etapa de levantamiento de requerimientos y cuyo DFD se describe a continuación:

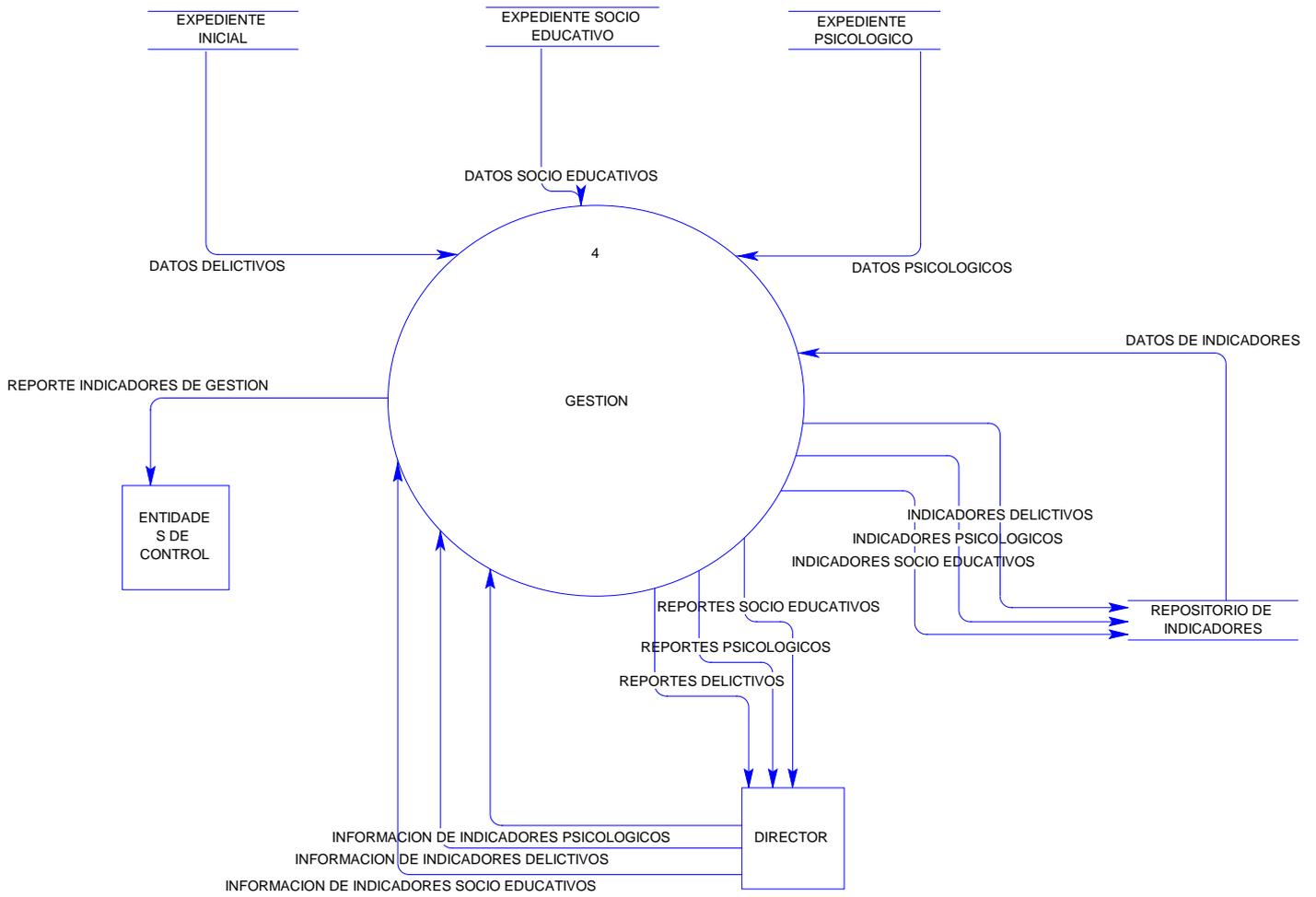


Figura 2.11: Diagrama de flujo del proceso de Gestión

El refinamiento de este diagrama a un nivel más de especificación se describe a continuación:

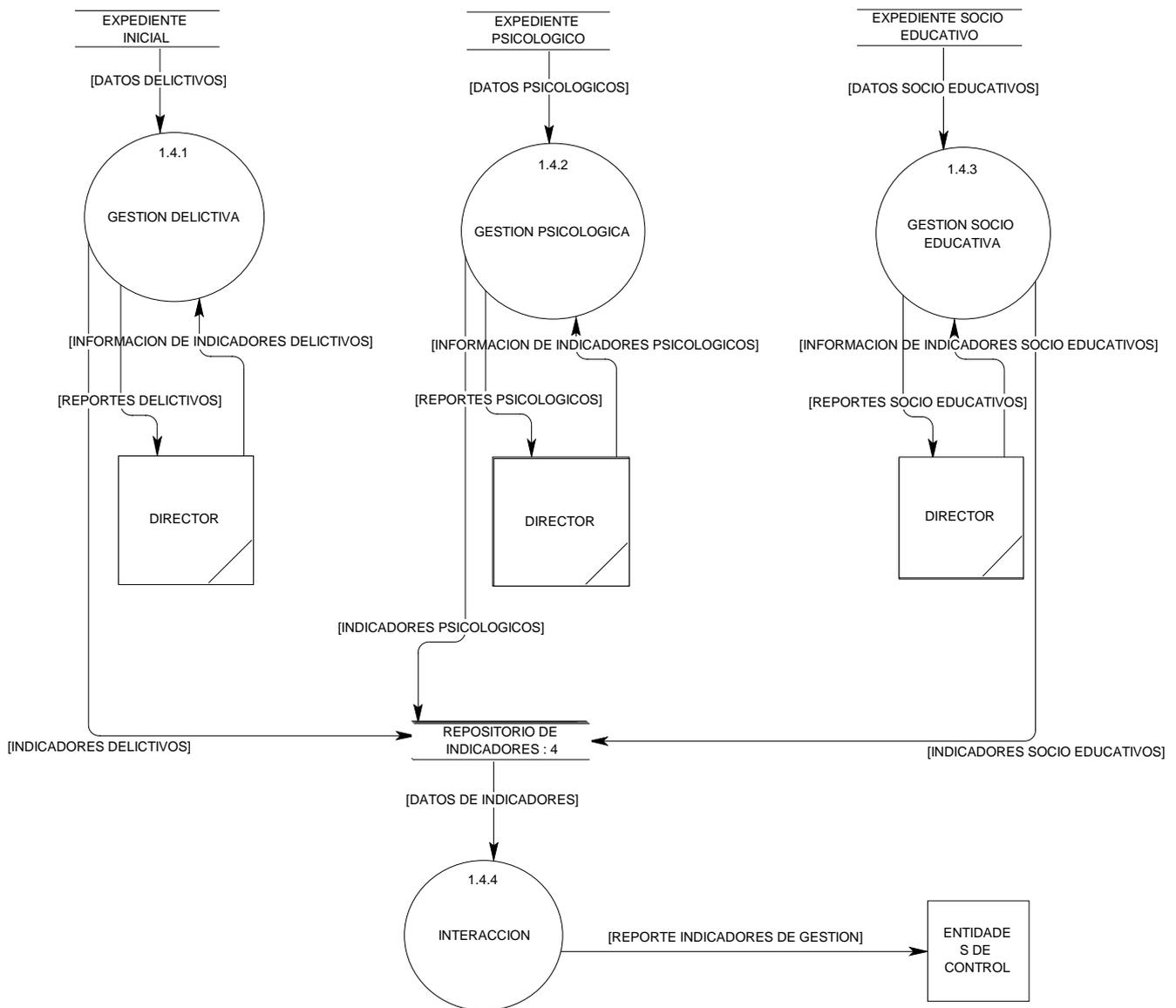


Figura 2.12: Refinamiento del diagrama de flujo del proceso de Gestión

Vale la pena repetir que el Catálogo de Metadatos de estos DFD está detallado en el ANEXO G

2.3.3.2 Indicadores

Luego del análisis del proceso de Gestión y de los requerimientos se define un listado de indicadores que faciliten la de toma de decisiones en el Centro de Orientación Juvenil “Virgilio Guerrero”, por lo que se describe a continuación la lista de dichos indicadores que han sido recopilados según entrevistas

mantenidas y asignándolos al respectivo subproceso mediante el respectivo código:

SUBPROCESO	CODIGO
Gestión Delictiva	PRC006
Gestión Psicológica	PRC007
Gestión Socio Educativa	PRC008

Tabla 2.5: Códigos de los subprocesos de Gestión

CODIGO	NOMBRE	DESCRIPCION	PROCESO
IN001	Número de jóvenes infractores por edades y por grupo de infracción	Este indicador sirve para el análisis de carácter psicológico a fin de determinar a que edad ocurre la mayor incidencia de una infracción determinada y actuar en campañas de prevención temprana	PRC006
IN002	Número de jóvenes infractores por edades y por grupo de infracción a través del tiempo	Este indicador permite identificar como se ha dado la incidencia de un grupo de infracción en determinadas edades a través del tiempo a fin de tener un soporte histórico para la toma de decisiones en cuento a campañas de prevención	PRC006
IN003	Reincidencia de infracción de los jóvenes por edades a través del tiempo	Este indicador permitirá conocer de manera histórica cual ha sido las edades en las cuales los jóvenes reinciden en una determinada infracción a fin de mitigar dichos hechos con procesos de apoyo psicológico ya sea a nivel de colegios, escuela o campañas en la comunidad	PRC006
IN004	Numero de jóvenes infractores catalogados por edad y medida impuesta	Este indicador permite conocer la tendencia de los tribunales a aplicar una medida especifica según la edad a fin de evaluar dichas sentencias y proponer mejoramiento en las leyes de ser el caso	PRC006

IN005	Edades en los que hay una mayor incidencia de jóvenes con delitos en contra del pudor	Este indicador permite conocer a que edades los jóvenes infractores cometen este tipo de delitos lo cual puede ayudar a la identificación de sectores de la sociedad vulnerables e iniciar una investigación de sus causas	PRC006
IN006	Numero de jóvenes infractores que son casados agrupados por edades	Este indicador permite determinar como influye el matrimonio a temprana edad en la consecución de delitos en los jóvenes	PRC006
IN007	Numero de jóvenes infractores que han sido sentenciados a internamiento por delitos en contra de la dignidad sexual	Este indicador permite conocer el grado de agravamiento de delitos de naturaleza sexual en los jóvenes a fin de elaborar campañas de prevención	PRC006
IN008	Numero de jóvenes infractores que han ingresado en malas condiciones físicas o psicológicas por infracciones contra la propiedad	Este indicador permite conocer si existe una tendencia al maltrato del joven antes del ingreso al centro como parte de la represión por el delito cometido	PRC007
IN009	Jóvenes que han sido aprehendidos por delitos contra la propiedad privada y ha reincidido en la misma falta	Este indicador permite afianzar tratamientos psicológicos que corrijan o atenúen este tipo de conductas en los jóvenes	PRC007

IN010	Numero de jóvenes infractores solteros frente a casados que han cometido delitos en contra el pudor	Este indicador permite conocer si influye el estado civil del joven infractor en la ocurrencia de delitos en contra del pudor a fin de determinar si los jóvenes que ya han tenido relaciones sexuales, no cometen este tipo de delitos	PRC007
IN011	Numero de jóvenes huérfanos de padre o madre que han cometido delitos en contra del pudor	Este indicador es útil a fin de determinar si existe una relación entre el abandono o la falta de una figura ya sea paterna o materna y el apareamiento de delitos en contra del pudor	PRC007
IN012	Numero de jóvenes con una figura familiar desorganizada en los distintos grupos de infracción	Este indicador permite ir descubriendo la influencia del modelo familiar en la ocurrencia de cierto tipo de delitos	PRC007
IN013	Numero de jóvenes con una figura familiar organizada en los distintos grupos de infracción	Este indicador complementa el descubrimiento de la influencia familiar en la ocurrencia de cierto tipo de delitos	PRC007
IN014	Numero de jóvenes de formación primaria que han cometido algún delito	Este indicador sirve para referenciar a que nivel de educación ocurren ciertos delitos	PRC007
IN015	Numero de jóvenes de formación secundaria que han cometido algún delito	Este indicador complementa el anterior en el sentido de evidenciar si a medida que el nivel de educación avanza, el numero de delitos cometidos disminuye y que se puede hacer en coordinación con los centros educativos	PRC007

IN016	Numero de jóvenes que conviven en unión libre agrupados por edad y grupo de infracción	Este indicador permite determinar la relación entre la convivencia de pareja no formal de los jóvenes y las edades en las que se cometen cierto tipo de delitos	PRC007
IN017	Número de jóvenes infractores en los diferentes grupos de apoyo socio educativo a través del tiempo	Este indicador sirve como aporte de carácter socio educativo para conocer cuáles son los grupos de apoyo que mas han beneficiado a los jóvenes infractores con el fin de tomar decisiones en cuanto a reforzar determinados grupos o impulsar nuevos	PRC008
IN018	Jóvenes que ante un delito en contra el pudor han sido sentenciado y en que grupo socio educativo han recalado	Este indicador permite conocer que grupo de apoyo socio educativo conviene mas para jóvenes que han cometido este tipo de delitos	PRC008
IN019	Incidencia de determinado grupo infractor por edad a través del tiempo	Este indicador sirve para determinar la variación de edades en los diferentes grupos infractores de forma histórica	PRC008
IN020	Edades de los jóvenes que ingresan a talleres con mayor proyección laboral	Este indicador permite conocer si jóvenes de edades cercanas a la mayoría de edad tienden a tratar de rehabilitarse obteniendo una posibilidad de reinserción laboral en la sociedad	PRC008
IN021	Numero de jóvenes que están en el programa de Libertad Asistida que han cometido una reincidencia	Este indicador permite medir el grado de efectividad de este programa	PRC008
IN022	Numero de jóvenes infractores que han tenido experiencia con drogas agrupados por edad	Este indicador permite conocer la tendencia de los jóvenes que han cometido una infracción a consumir drogas en determinadas edades	PRC008

IN023	Numero de jóvenes que han consumido drogas ilegales y su relación con determinado grupo de infracción	Este indicador permite determinar la influencia del consumo de drogas ilegales en la ocurrencia de un determinado delito	PRC008
-------	---	--	--------

Tabla 2.6: Indicadores

De estos indicadores se derivan los datos necesarios para las tablas de dimensiones y hecho que se especifican en el Modelo de Transformaciones detallado en el ANEXO C.

2.3.3.3 Modelo Conceptual de Datos del DM

En el apartado anterior al enlistar los indicadores, se define tácitamente el cuestionario que el DM debe responder, dicho cuestionario que se exige como un paso dentro de la metodología escogida para el Desarrollo del presente proyecto. Vale la pena aclarar que de dicha tabla de indicadores se seleccionará solamente los que contribuyan o se enmarquen en el alcance del presente proyecto; esto es solo los indicadores que tengan que ver con las áreas psicológica, socio educativa y delictiva; de esta manera se elabora el modelo conceptual del DM planteando las siguientes dimensiones y tabla de hechos a fin de que el DM responda a dicho cuestionario:

Dimensiones

- Dim_Educacion
- Dim_Ingreso
- Dim_Joven
- Dim_Medida
- Dim_Tiempo

Hechos

- Hecho_Infraccion

La arquitectura de estos objetos, mantiene un esquema en estrella y queda definida de la siguiente manera:

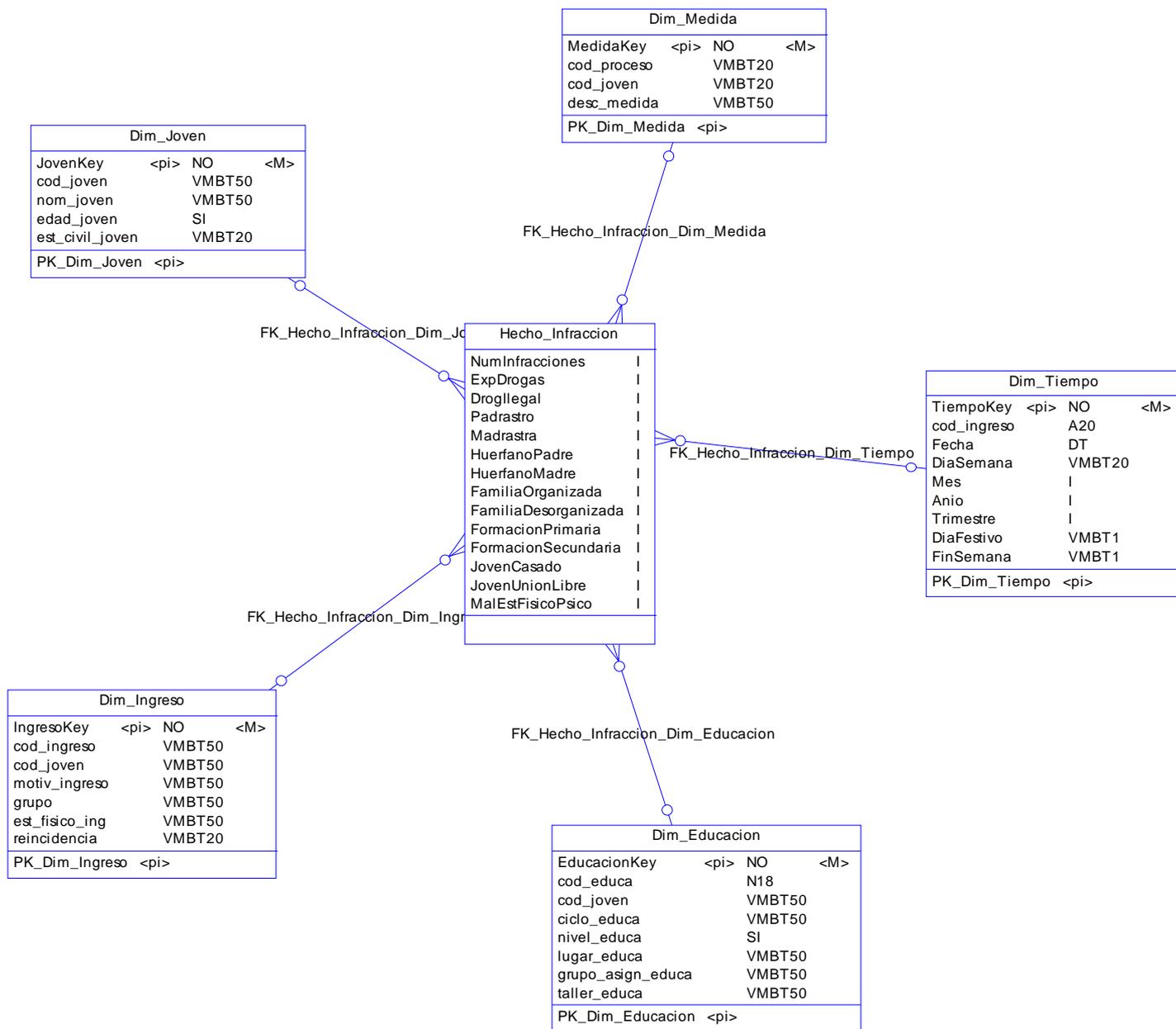


Figura 2.14: Modelo Conceptual del DM

2.4 Diseño

2.4.1 Selección de herramientas e instalación

Como se indicó en el apartado 2.2.2 se definieron las herramientas con las cuales se desarrollará el DM. La instalación y afinamiento de las mismas no se colocará como punto o anexo de este trabajo debido a que este proceso puede ser definido de forma total y mejorada en bibliografía específica creada para dicho efecto.

2.4.2 Diseño Físico

2.4.2.1 Modelo de Datos Físico

El Modelo de datos físico está definido en detalle en el ANEXO B al igual que sus metadatos de tal forma que cumpla con lo estipulado en el modelo conceptual y los indicadores. A continuación el diagrama Entidad Relación:

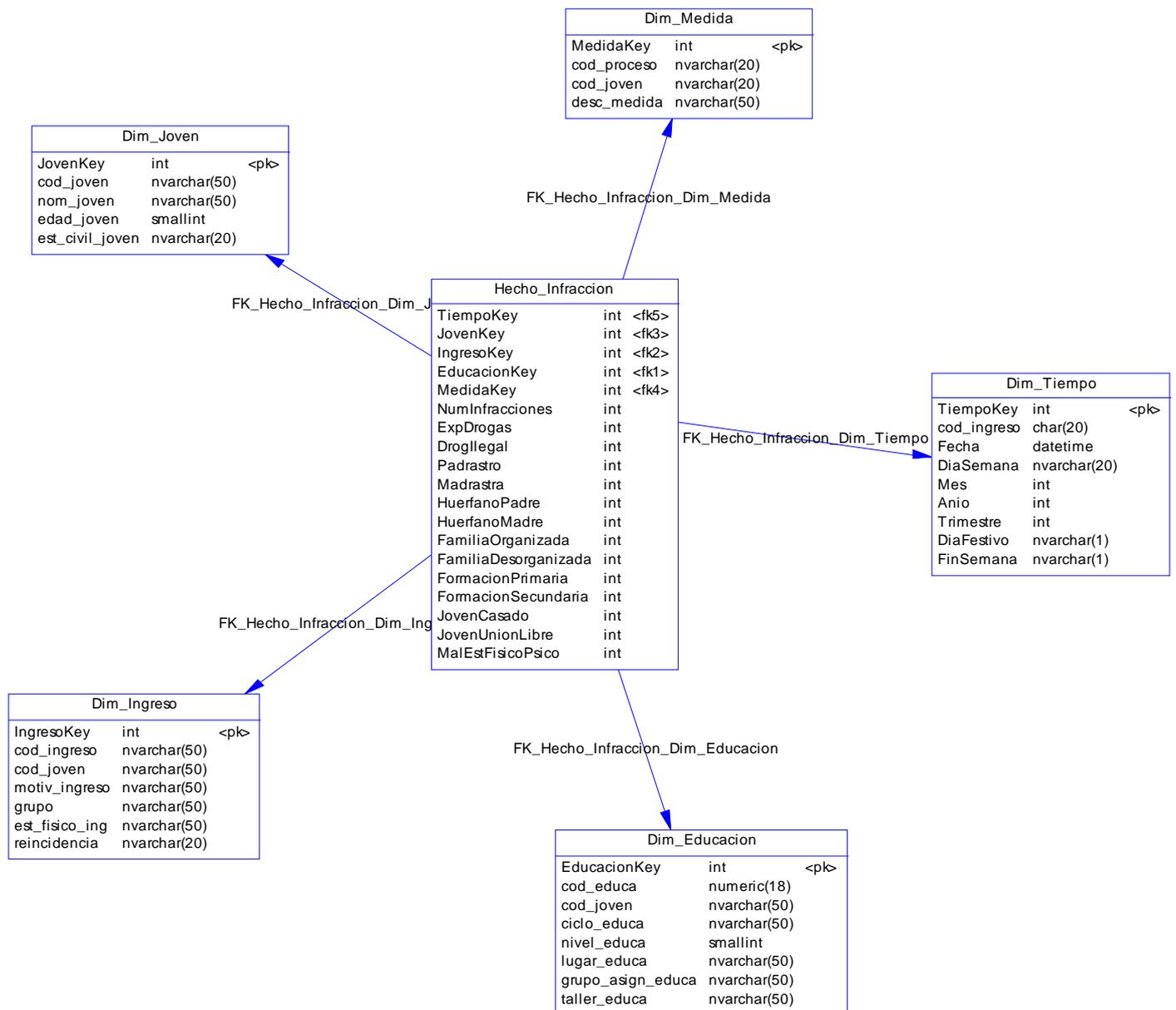


Figura 2.15: Diseño físico del DM

2.4.2.2 Determinación de la Arquitectura

La arquitectura de la aplicación está diseñada para que cumpla con el requerimiento respectivo y de acuerdo al siguiente esquema:

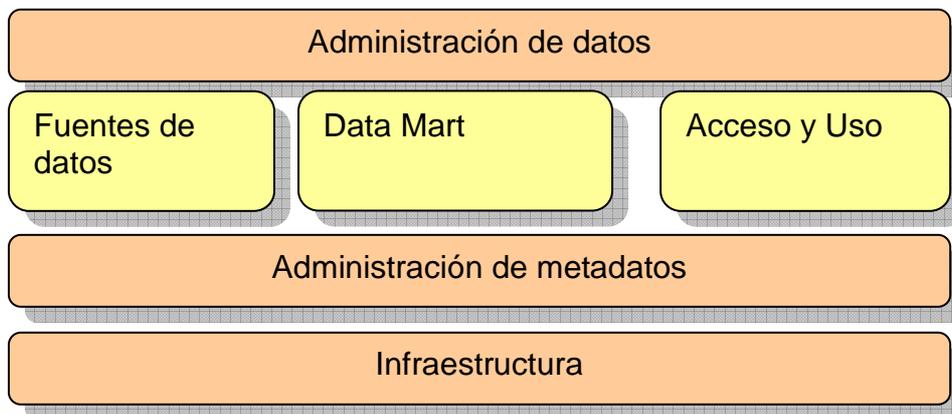


Figura 2.16: Arquitectura del DM

La descripción de los bloques y capas de la arquitectura se definen a continuación:

2.4.2.2.1 Bloque de la arquitectura

2.4.2.2.1.1 Bloque de Fuentes de Datos

El Bloque de Fuentes de Datos se compone de cinco sub bloques, que están definidos en detalle en el apartado 1.2.4.1.1 y que se resume de la siguiente manera:

- Datos de producción
- Datos de herencia
- Sistemas internos de oficina
- Fuentes externas
- Metadatos de las fuentes de datos

De estos sub bloques, el bloque de fuentes de datos del presente proyecto, solo se usará el de datos de producción ya que solo existe la base de datos del

sistema de información del Centro de Orientación Juvenil “Virgilio Guerrero” de donde extraer información; no existen datos de herencia ya que en administraciones anteriores a la actual no se mantenía un archivo formal y se ha ido perdiendo casi la totalidad de la información histórica pero que sin embargo en un futuro trabajo se podría hacer una búsqueda mas exhaustiva de los mismos. Los sistemas internos de oficina no se estructuran en torno a la información que el DM necesita. Las fuentes externas antes que alimentar al DM, se alimentan de él por lo que tampoco existe este sub bloque dentro de la arquitectura del DM. Los Metadatos de las fuentes de datos están detallados en el ANEXO A.

2.4.2.2.1.2 Bloque del Data Mart

El bloque del DM esta compuesto por dos sub bloques que se resumen a continuación:

- Refinamiento y reingeniería
- Data Mart

El sub bloque de refinamiento y reingeniería está compuesto por los paquetes DTS que soportan el modelo de transformaciones que se expone en el ANEXO C del presente proyecto y que a su vez forman parte del catálogo de metadatos general del DM.

El sub bloque del DM se compone, de manera global, de la base de datos según el diagrama Entidad Relación que se especifica en la figura 2.15, el cubo

OLAP que permite el análisis de la información solicitada en los indicadores de la tabla 2.6 y la calendarización automática según lo especificado en la tabla 2.4

2.4.2.2.1.3 Bloque de acceso y uso

El bloque de acceso y uso esta compuesto por dos sub bloque que son:

- Acceso y recuperación
- Análisis y reporte

El sub bloque de acceso y recuperación se compone de los componentes de acceso y recuperación de la información contenida en el cubo OLAP, componentes que son propios de la herramienta de explotación, en este caso del MS Excel 2000.

El sub bloque de Análisis y reporte está compuesto por las tablas dinámicas que muestran los indicadores solicitados en la tabla 2.6.

2.4.2.2.2 Capas de la arquitectura

2.4.2.2.2.1 Capa de Administración de datos

Esta capa está compuesta por los módulos de las herramientas requeridas para la construcción del DM que soporten la administración de los bloques de la arquitectura y que se detallan a continuación:

- La fuente de datos será administrada mediante el módulo Enterprise Manager del MS SQL Server 2000.
- El bloque del DM estará administrado mediante los módulos Enterprise Manager para la base de datos del DM, los paquetes DTS y la automatización de la carga, Analysis Services para la administración del cubo OLAP, ambos módulos forman parte del MS SQL Server 2000
- El bloque de acceso y uso estará administrado a través de los componentes del MS Excel 2000, que permiten la generación de reportes a través de tablas dinámicas que se generan en base a las dimensiones y medidas de los cubos OLAP

2.4.2.2.2 Capa de Infraestructura

La capa de Infraestructura esta compuesta por las plataformas tecnológicas que permiten el uso de las herramientas especificadas en los requerimientos de desarrollador en el apartado 2.2.2.2.

2.4.2.2.3 Capa de Metadatos

La capa de metadatos está conformada por el catálogo de metadatos especificado en el apartado 2.1.8 y automatizado mediante el uso del componente respectivo del Enterprise Manager del MS SQL Server 2000

2.4.2.3 Especificación de Procesos para el DM

El DM que es objeto de estudio en el presente proyecto tiene la finalidad de intervenir en el proceso de Gestión en el Centro de Orientación Juvenil “Virgilio Guerrero”, mismo proceso que fue definido en el punto 2.3.3.1. De esta manera se rediseña el proceso de gestión mediante el siguiente diagrama de flujo de datos para el DM mismo que se detalla en el ANEXO G:

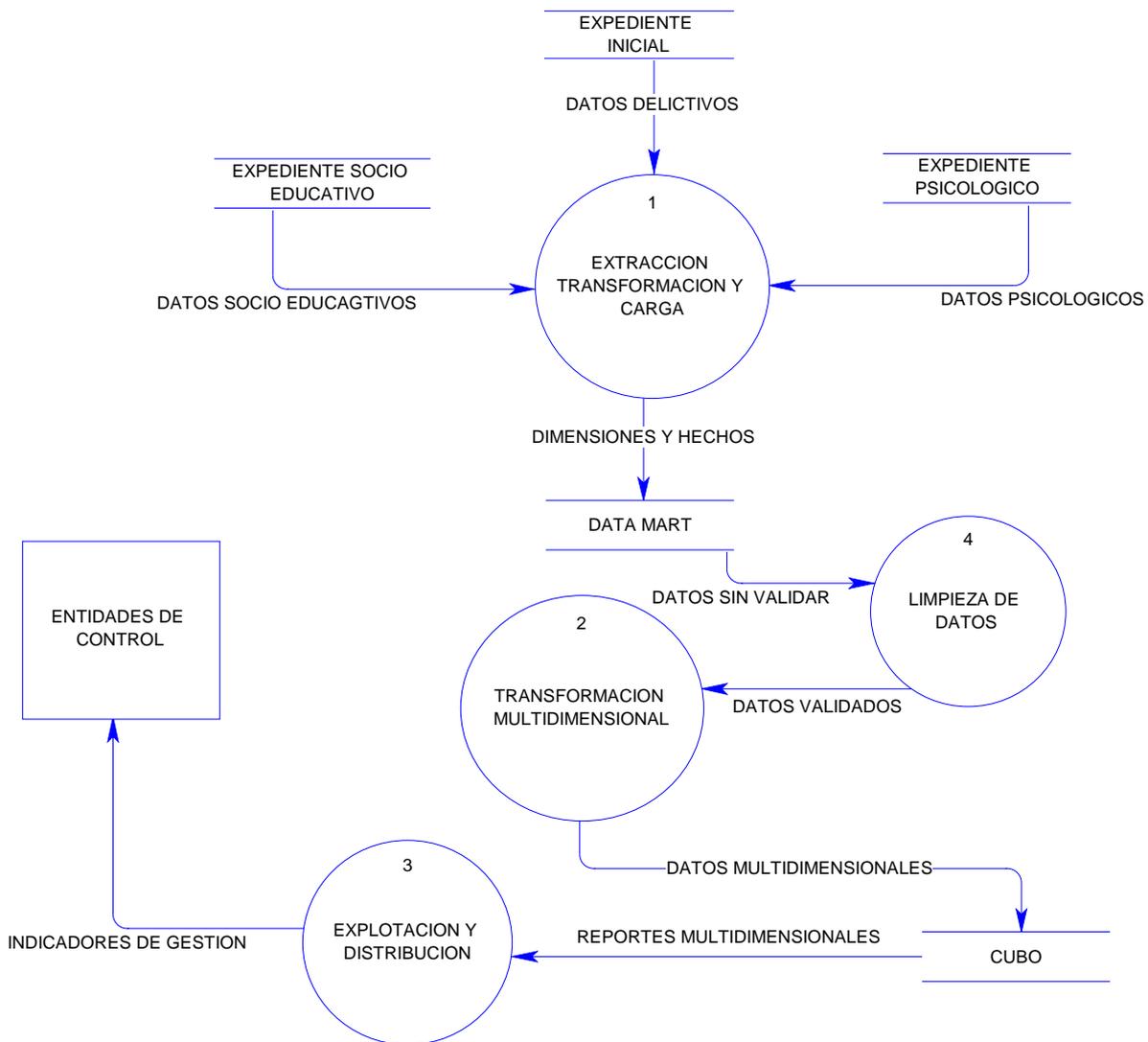


Figura 2.17: Diagrama de flujo de datos del DM

2.4.2.3.1 Extracción, transformación y carga

Este proceso es el encargado de llevar los datos desde la base de datos fuente hacia la base de datos del DM y se realizan mediante la ejecución de las siguientes actividades:

- Lectura de datos de las tablas de origen
- Transformación y filtro de los datos
- Escritura de los datos en las tablas de destino

2.4.2.3.2 Limpieza de datos

Por la estructura tecnológica que se tiene en el Centro, no se han definido políticas ni lineamientos específicos sobre la calidad de la información mas que la de tener los datos coherentes a la realidad práctica del Centro. Por esto se definen a continuación ciertas pautas para definir la limpieza y consistencia de la información:

- La información debe considerar a todos los jóvenes que pasan por el centro de forma temporal o definitiva como resultado de los procesos de investigación, juzgamiento y cumplimiento de penas
- La correspondencia de la información referente a un joven determinado debe ser de forma única; es decir que para un joven dado se debe mostrar la información estadística histórica considerando al joven como un único ente dentro de la base de datos independientemente de cuantas detenciones tenga
- Se debe tomar en cuenta que el joven será identificado por un solo código único y no por el nombre

- La información referente a padre y madre del joven deberá considerarse en forma única ya sea que estos no sean los progenitores del mismo, es decir, que si el joven vive con el padrastro o madrastra estos serán considerados como padre o madre almacenándose solo un registro de padre y madre por joven y solo se utilizará la regencia que existe en dichos registros para reverenciarlos como padrastro o madrastra.
- La información sobre acudientes se la considerará de carácter volátil por lo que se podrá aceptar que para un joven dado se podrá tener más de un acudiente independientemente que sea algún familiar o incluso el padre o la madre.

2.4.2.3.3 Transformación multidimensional

Este proceso contempla la conversión de los datos almacenados en el DM a estructuras multidimensionales como lo son los cubos; de acuerdo al siguiente diagrama:

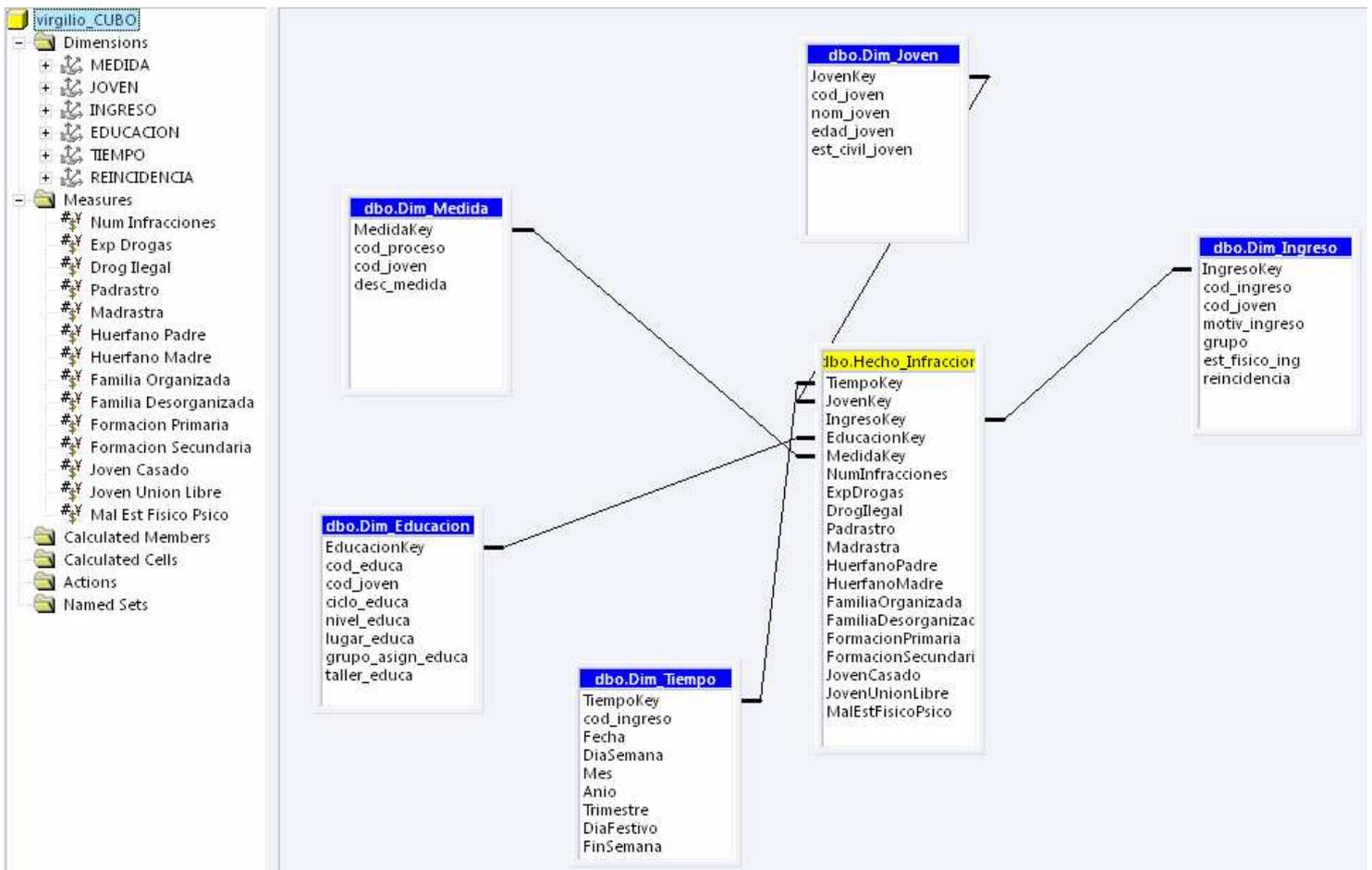


Figura 2.18: Diagrama multidimensional

2.4.2.3.4 Explotación y Distribución

Este proceso define la manera como se utilizará el DM. Por esta razón, este proceso, se encargará de generar los reportes multidimensionales necesarios para responder a los indicadores especificados en la tabla 2.6.

El proceso continua con la distribución de los reportes que arroje el DM, en este caso solo contempla los siguientes pasos:

- Actualización de los reportes
- Impresión de los reportes
- Entrega de reportes impresos al personal interesado

2.4.2.4 Modelo de Extracción y Transformación

Este modelo está descrito de forma detallada en el ANEXO C y contempla el esquema de paquetes DTS necesarios para realizar las extracciones y transformaciones respectivas sobre los datos fuentes y la carga en la base de datos del DM.

2.4.2.5 Selección del Esquema de Almacenamiento

Se optará por la implementación de un solo Cubo OLAP que utilice un modelo de almacenamiento HOLAP ya que las herramientas en las que va a desarrollarse el DM lo permiten. HOLAP es un modelo de almacenamiento que mantiene los datos de detalle en estructuras ROLAP y los datos de resumen los mantiene en estructuras MOLAP y se vale de los servicios OLAP para que hagan la integración entre las estructuras MOLAP y ROLAP

2.4.2.6 Especificaciones de Usuario Final

Por las razones ya descritas en apartados anteriores, el Centro no maneja una estructura formal de sistemas por lo que no existen formatos establecidos para reportes o documentación. De esta manera existe la apertura del Centro a adoptar los formatos de reportes que se propongan en el presente trabajo y que estarán sujetos, en lo futuro, a la habilidad de uso de la herramienta de explotación del DM

CAPITULO III

3 Implementación y Pruebas.

3.1 Implementación

3.1.1 Construcción del Data Mart

3.1.1.1 Creación de la Base de Datos

La base de datos del DM se construyo con base a los requerimientos descritos anteriormente y a los indicadores de tal manera que se mantenga la consistencia con el diseño

Las especificaciones para creación, en cuanto al tamaño, de la base de datos del DM son:

OBJETO	TAMAÑO
Archivo de datos	20 GB
Archivo de log	30 GB

Tabla 3.1: Especificaciones físicas de la BDD del DM

Las especificaciones físicos mostradas en el cuadro anterior son estimadas ya que se proporcionó solo información de prueba correspondiente al primer y segundo trimestre del año 2003 la cual arrojó un espacio de carga inicial de 8,69 MB para el archivo de datos y de 10,5 MB para el log de transacciones por lo que las dimensiones mostradas en dicho cuadro son suficientes para la

carga inicial que incluya además los datos de los años posteriores con una tasa de crecimiento del 200% anual

En el ANEXO F se describe el script de creación de la base de datos del DM la cual se muestra ya creada a continuación:

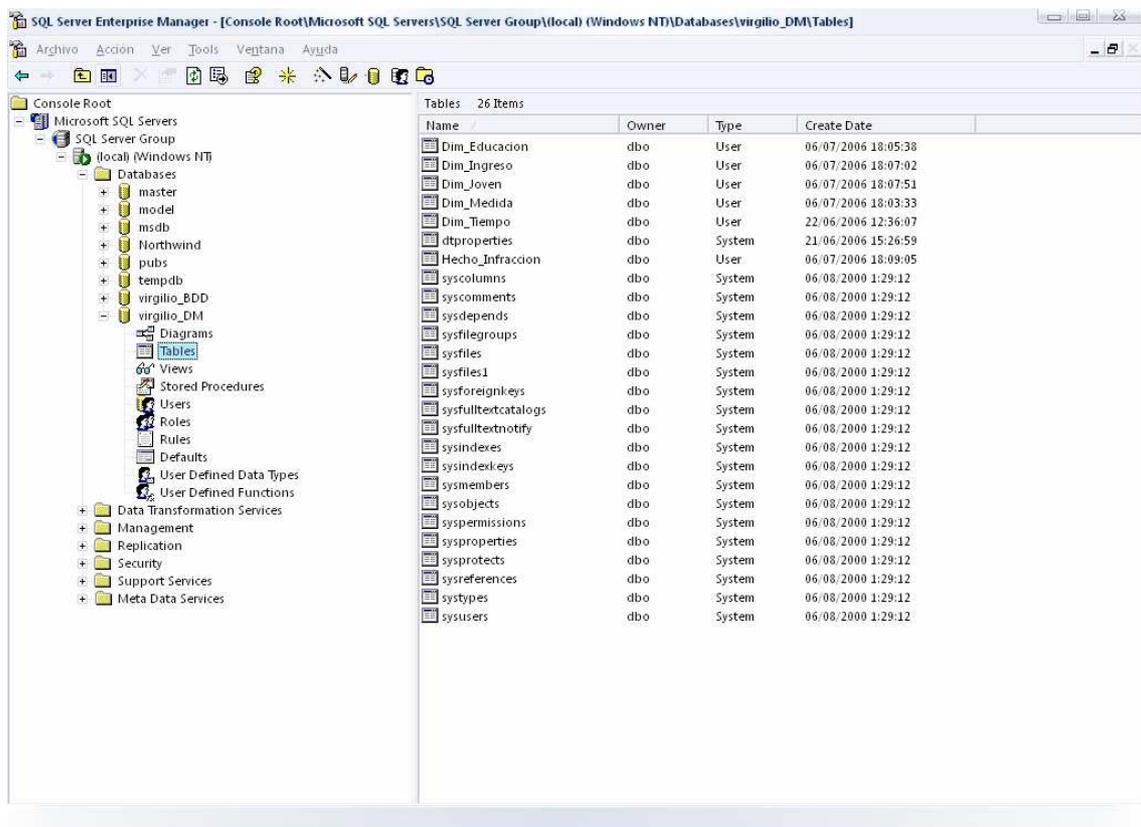


Figura 3.1: Creación de la BDD del DM

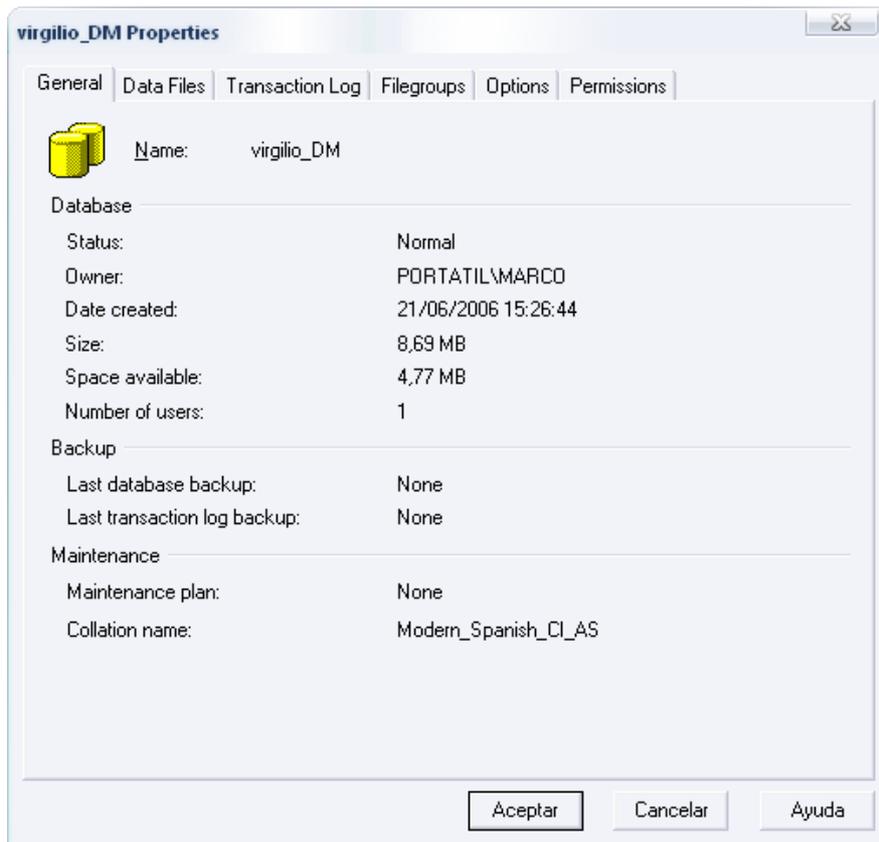


Figura 3.2: Información física de la BDD del DM

3.1.1.2 Creación de los paquetes de extracción, transformación y carga de datos

Según lo descrito en el modelo de transformaciones se procede a crear los siguientes objetos:

3.1.1.2.1 Creación de ODBC's:

Se crearon dos conexiones del tipo ODBC una para la base de datos fuente y otra para la base de datos del DM, estos ODBC fueron nombrados de la siguiente manera:

- virgilio_bdd; que conecta a la base de datos fuente
- virgilio_dm; que conecta a la base de datos del DM



Figura 3.3: Creación del ODBC de la BDD fuente



Figura 3.4: Creación del ODBC de la BDD del DM

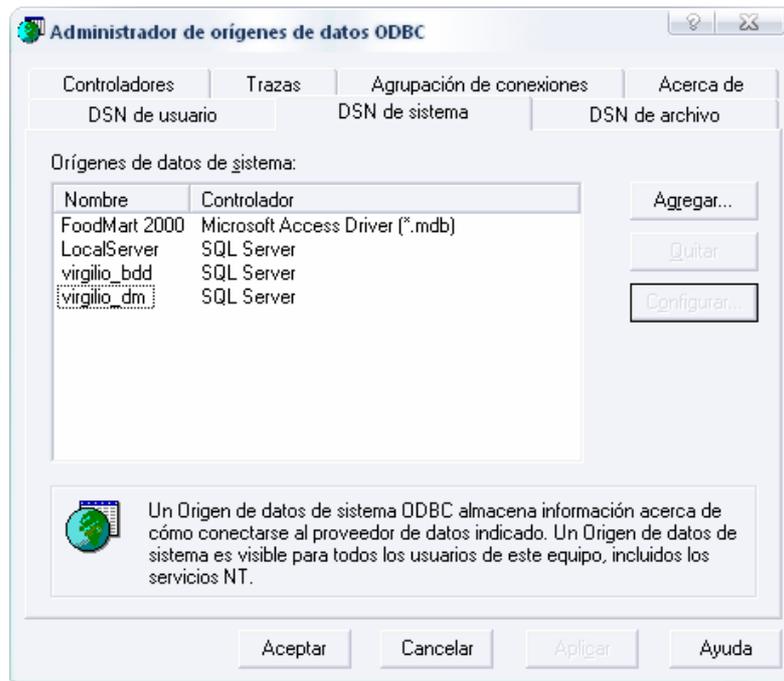


Figura 3.5: Creación de los ODBC's para las BDD fuente y del DM

3.1.1.2.2 Creación de paquetes DTS

El proceso de transformar los datos fuente consiste en realizar operaciones de base de datos sobre las columnas de las tablas que alimentan el DM de tal manera que se cumpla con el requerimiento de información que se solicita para el DM es por esto que se creó un paquete DTS en el MS SQL SERVER, que consta de los siguientes componentes:

3.1.1.2.2.1 Conexiones

- Virgilio_DMOLAP: esta conexión apunta hacia la base de datos del DM y mediante la cual se poblarán las tablas de las dimensiones y de la tabla de hechos

- tiempo_BDD: esta conexión apunta hacia la base de datos fuente y servirá para extraer la información y cargarla en la tabla Dim_Tiempo
- ingreso_BDD: esta conexión apunta hacia la base de datos fuente y servirá para extraer la información y cargarla en la tabla Dim_Ingreso.
- joven_BDD: esta conexión apunta hacia la base de datos fuente y servirá para extraer la información y cargarla en la tabla Dim_Joven
- educación_BDD: esta conexión apunta hacia la base de datos fuente y servirá para extraer la información y cargarla en la tabla Dim_Educacion
- medida_BDD; esta conexión apunta hacia la base de datos fuente y servirá para extraer la información y cargarla en la tabla Dim_Medida
- hecho_BDD; esta conexión apunta hacia la base de datos fuente y servirá para extraer la información y cargarla en la tabla Hecho_Infraccion

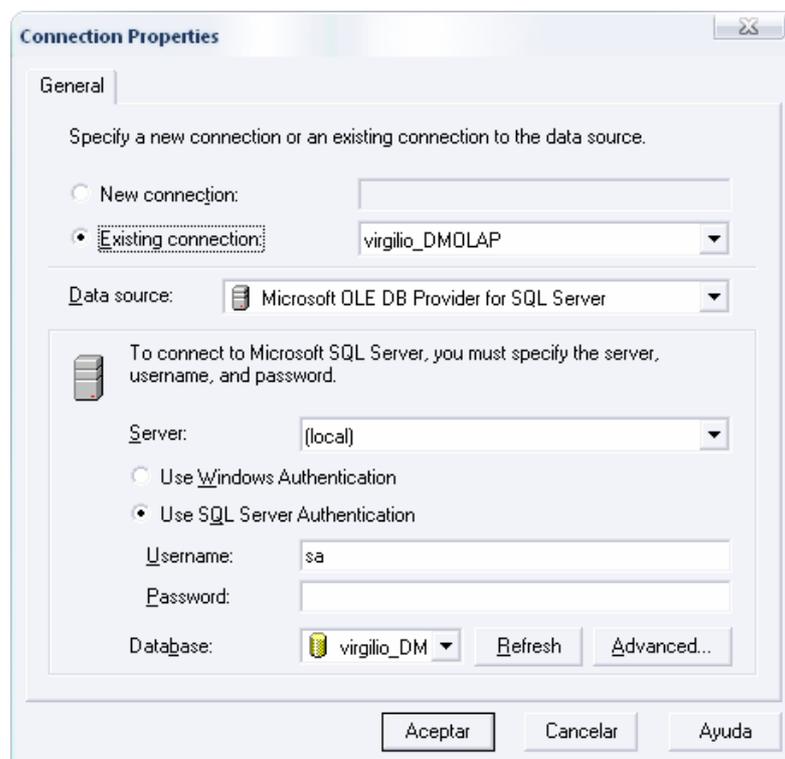


Figura 3.6: Conexión a la base de daos del DM

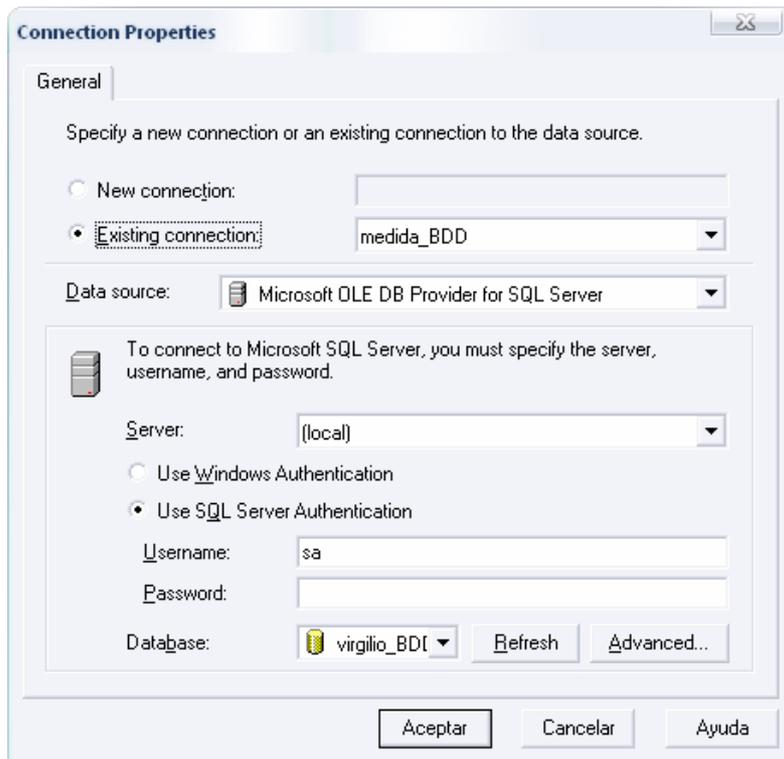


Figura 3.7: Conexión a la base de datos fuente para la tabla Dim_Medida

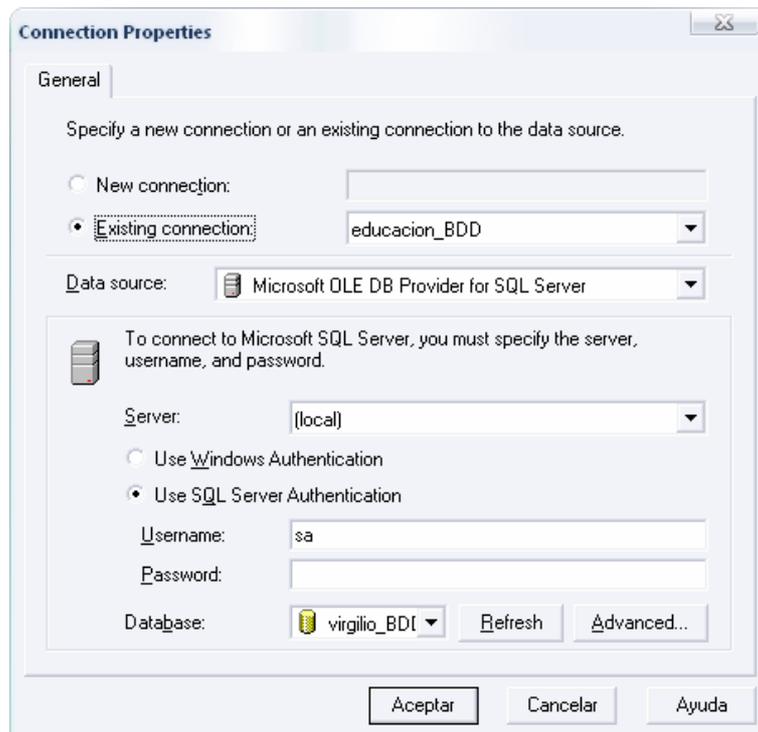


Figura 3.8: Conexión a la base de datos fuente para la tabla Dim_Educacion

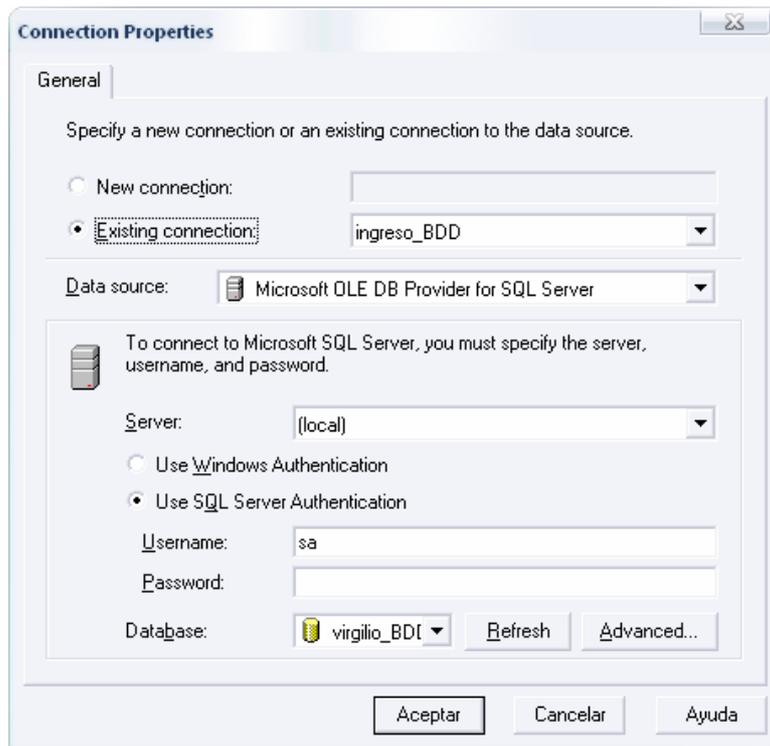


Figura 3.9: Conexión a la base de datos fuente para la tabla Dim_Ingreso

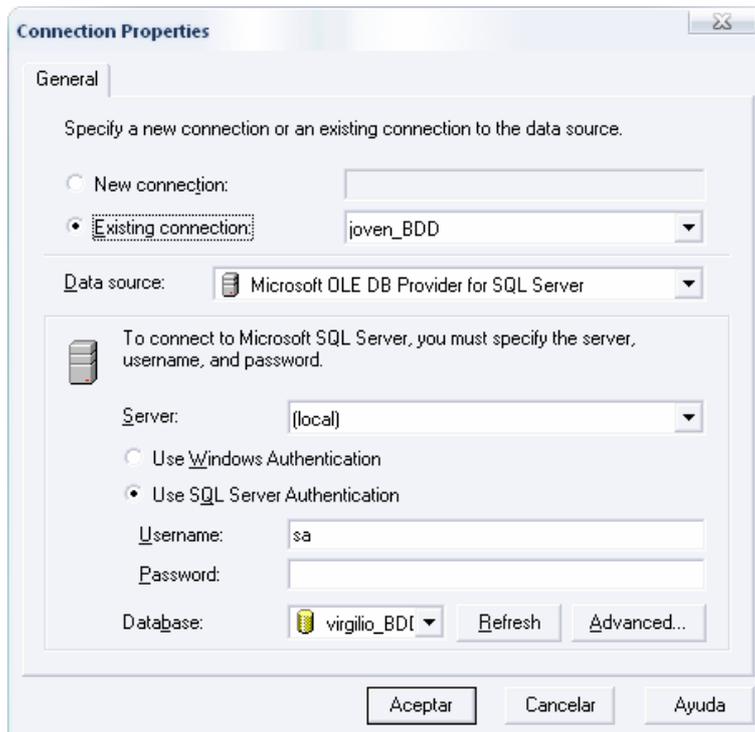


Figura 3.10: Conexión a la base de datos fuente para la tabla Dim_Joven

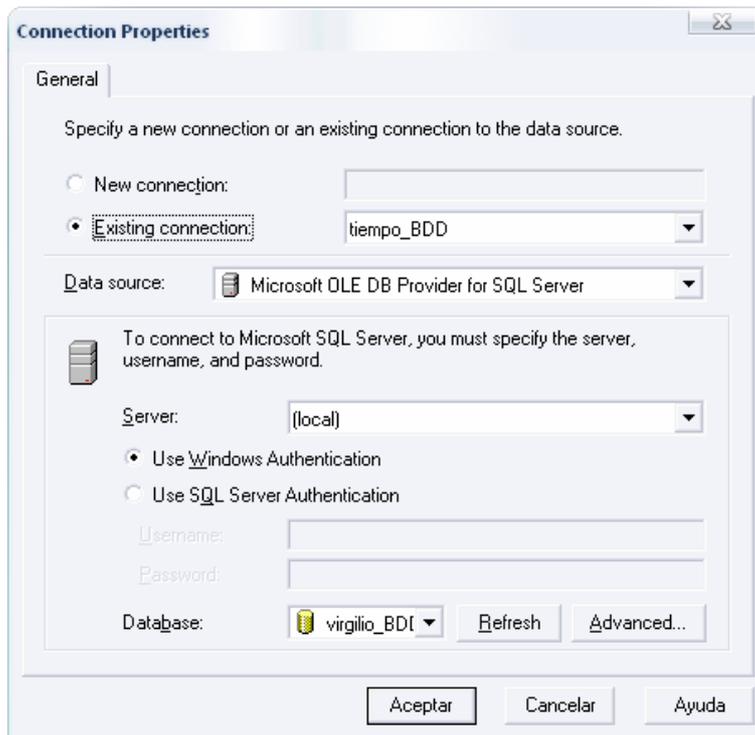


Figura 3.11: Conexión a la base de datos fuente para la tabla Dim_Tiempo

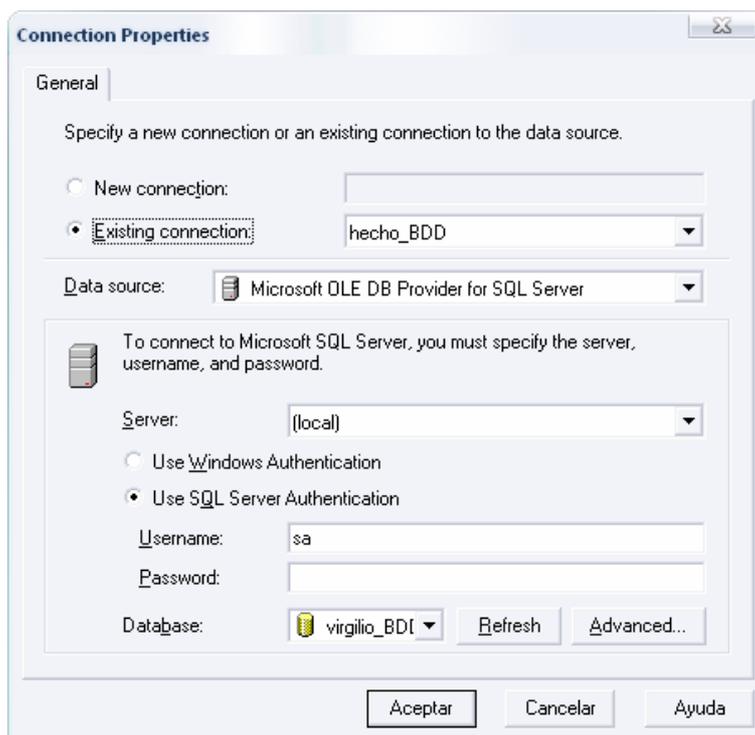


Figura 3.12: Conexión a la base de datos fuente para la tabla Hecho_Infraccion

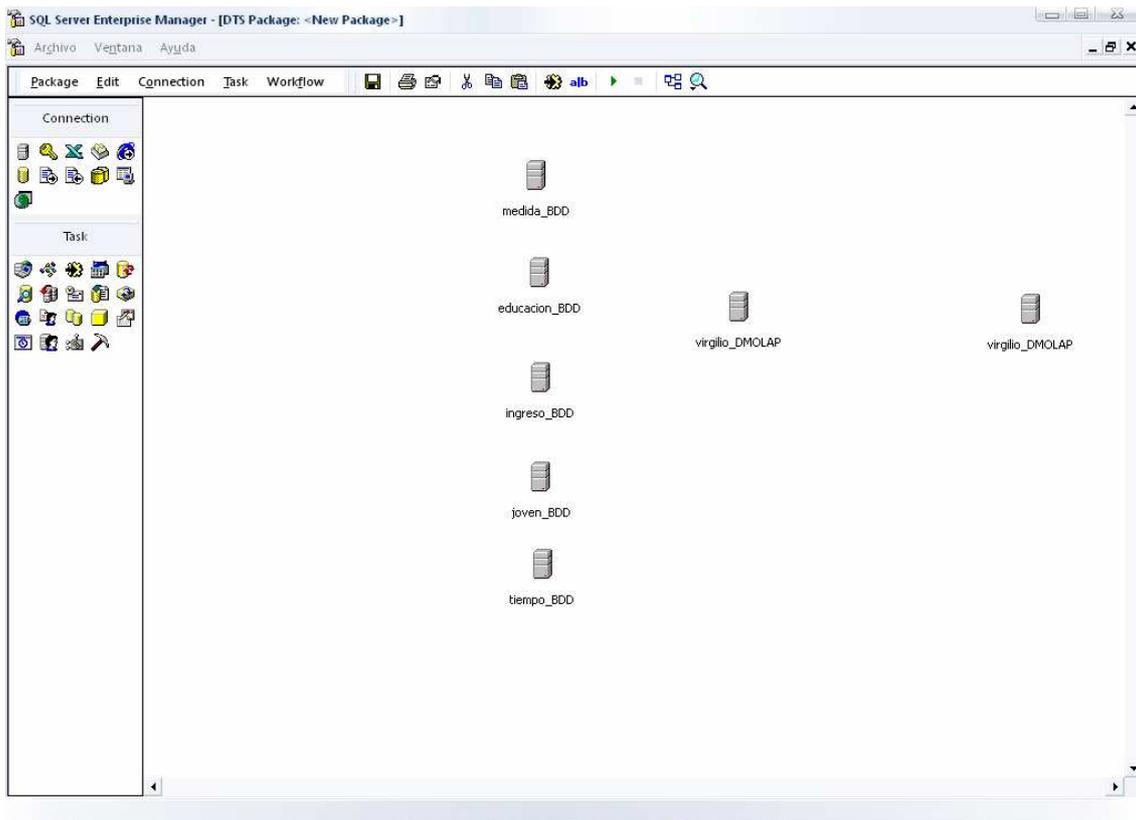


Figura 3.13: Conexiones

3.1.1.2.2 Tareas de ejecución SQL

- Limpiar Tablas: esta tarea se ejecuta al inicio y procede a borrar la información existente en las tablas

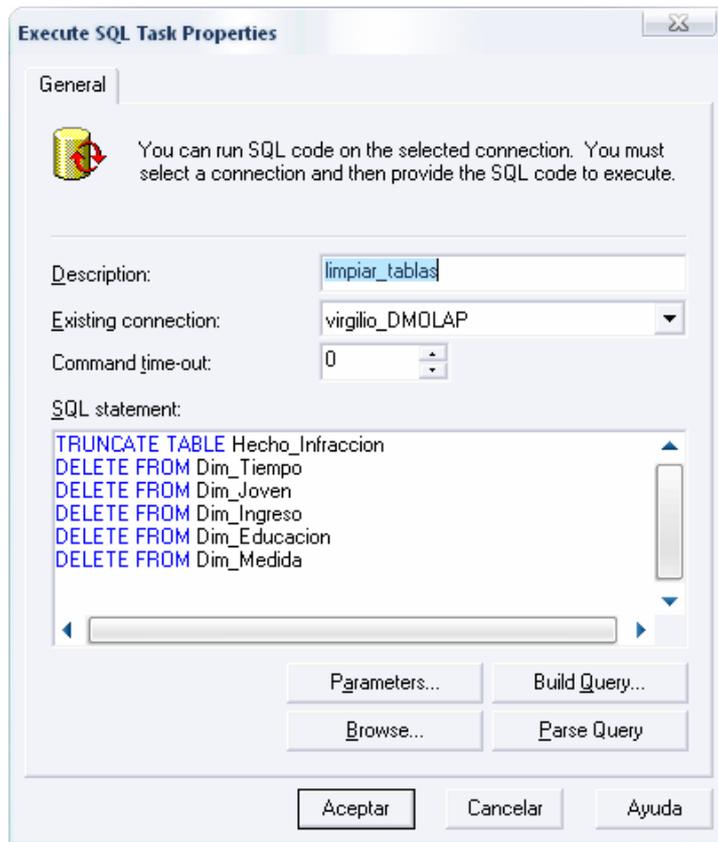


Figura 3.14: Tarea de ejecución SQL

3.1.1.2.2.3 Flujos de trabajo

- Se debe crear un flujo de trabajo entre la tarea de ejecución SQL y las conexiones a la base de datos fuente para las tablas de dimensiones a fin de que se condicione a que dichas las operaciones que se realicen sobre dichas conexiones se ejecuten si la limpieza de las tablas del DM ha sido ejecutada con éxito
- Se debe crear un flujo de trabajo ente la conexión a la base de datos del DM y la conexión a la base de datos fuente para la tabla de hechos a fin de que se condicione a que la carga en la tabla de hechos se ejecutará siempre que se hayan cargado con éxito las tablas de las dimensiones

3.1.1.2.2.4 Tareas de transformación de datos

Las tareas de transformación de datos se crean en base al modelo de transformaciones que se especifica en el ANEXO C, de la siguiente manera

- Carga Dimensión Medida: esta tarea permite la carga y transformación de los datos extraídos de las tablas de proceso y medidas a la tabla Dim_Medida.

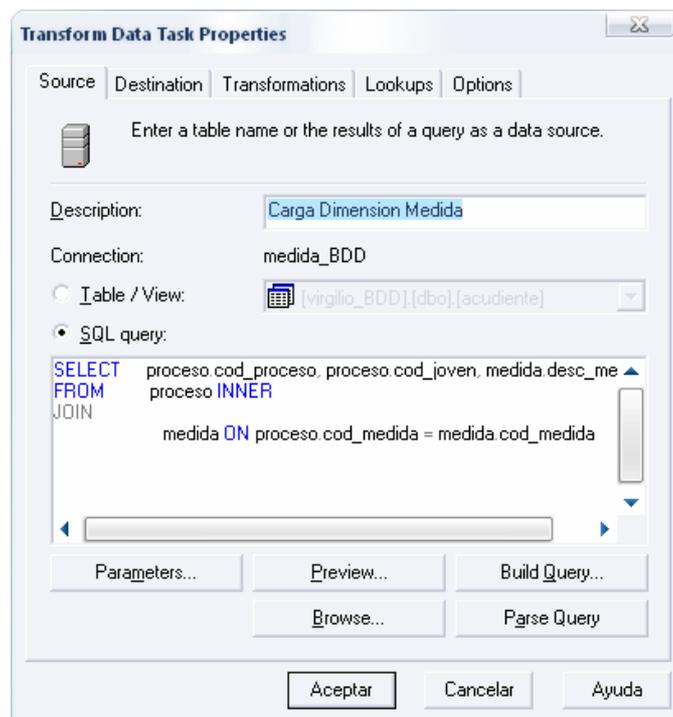


Figura 3.15: Carga de datos de la Dimensión Medida

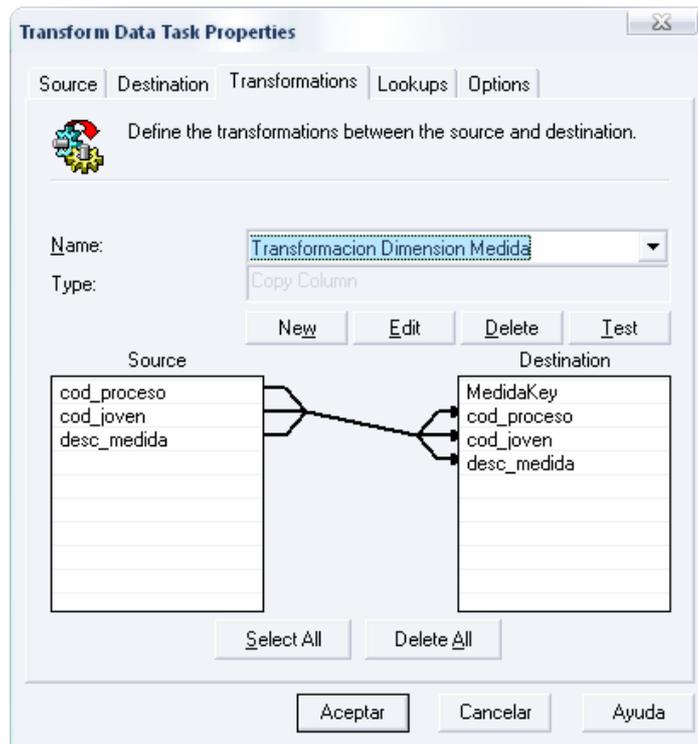


Figura 3.16: Transformación de datos de la Dimensión Medida

- Carga Dimensión Educación: esta tarea permite la carga y transformación de los datos extraídos de la tabla de educación a la tabla Dim_Educacion.

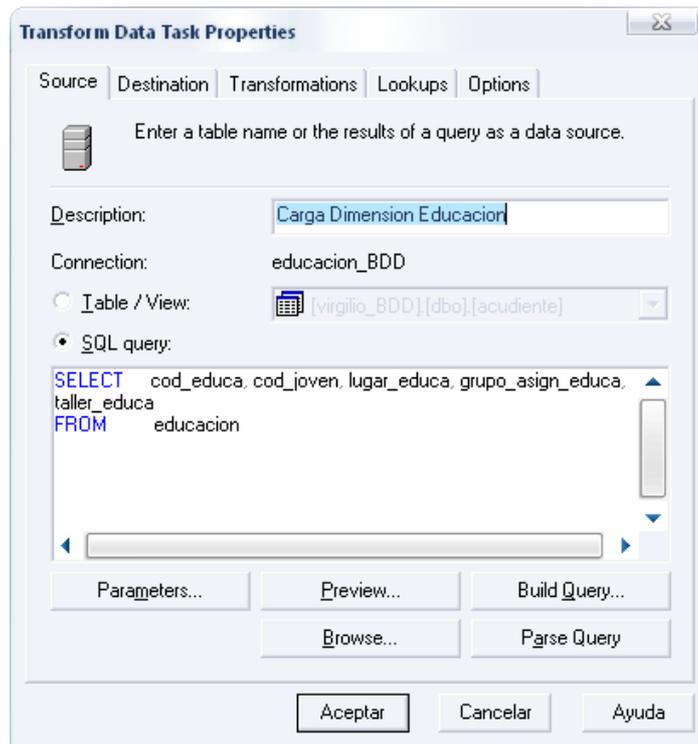


Figura 3.17: Carga de datos de la Dimensión Medida

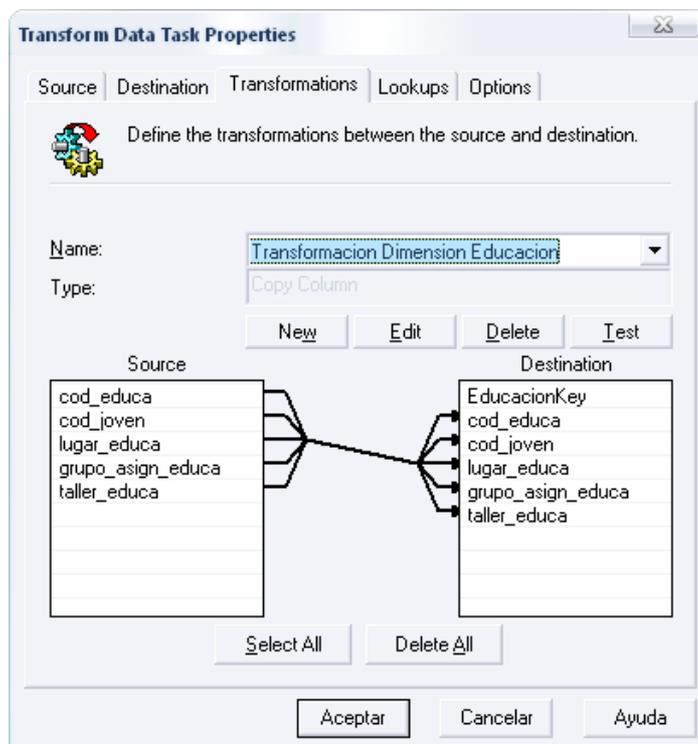


Figura 3.18: Transformación de datos de la Dimensión Medida

- Carga Dimensión Ingreso: esta tarea permite la carga y transformación de los datos extraídos de la tabla de ingresos a la tabla Dim_Educacion.

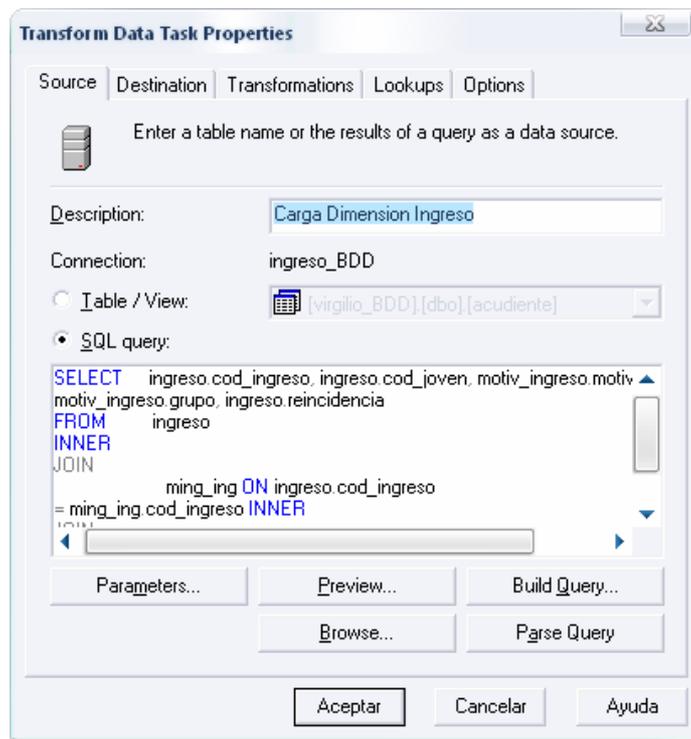


Figura 3.19: Carga de datos de la Dimensión Ingreso

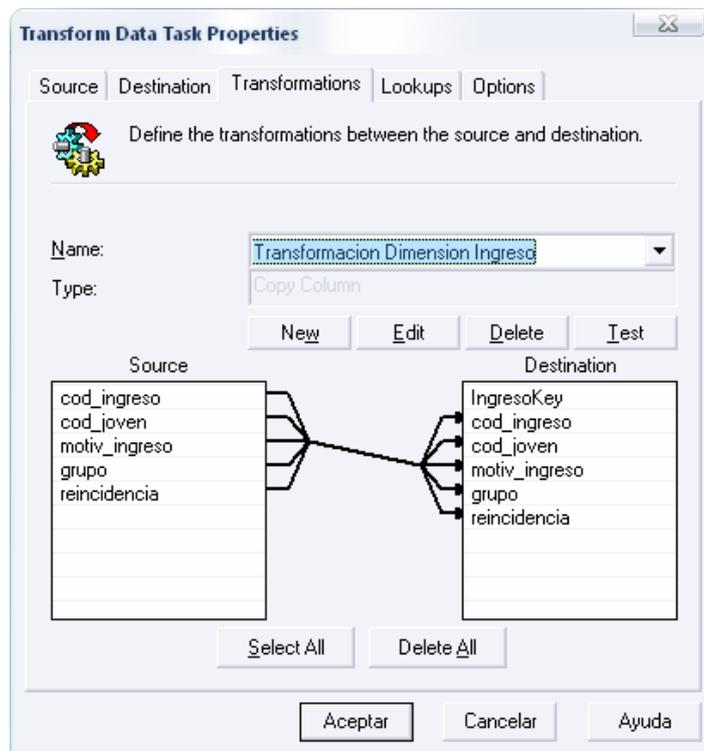


Figura 3.20: Transformación de datos de la Dimensión Ingreso

- Carga Dimensión Joven: esta tarea permite la carga y transformación de los datos extraídos de la tabla de jóvenes a la tabla Dim_Joven.

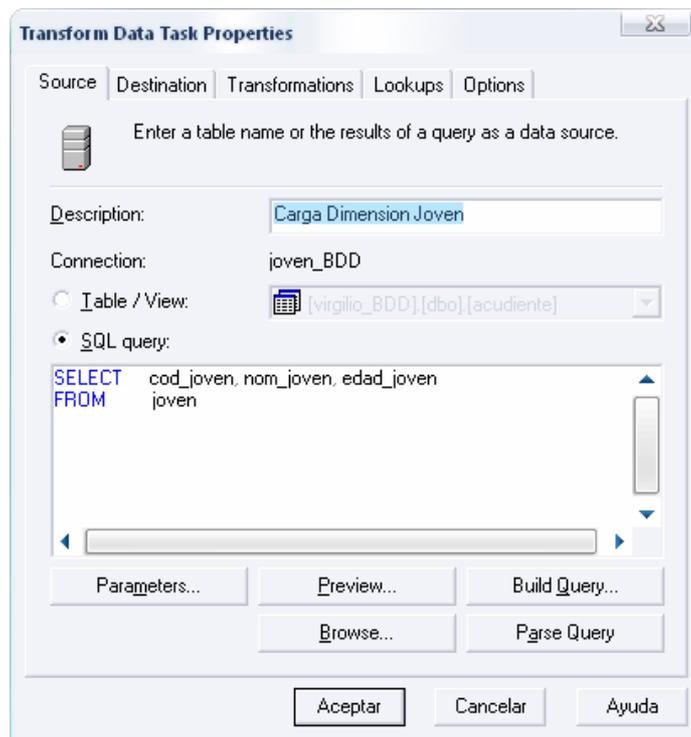


Figura 3.21: Carga de datos de la Dimensión Joven

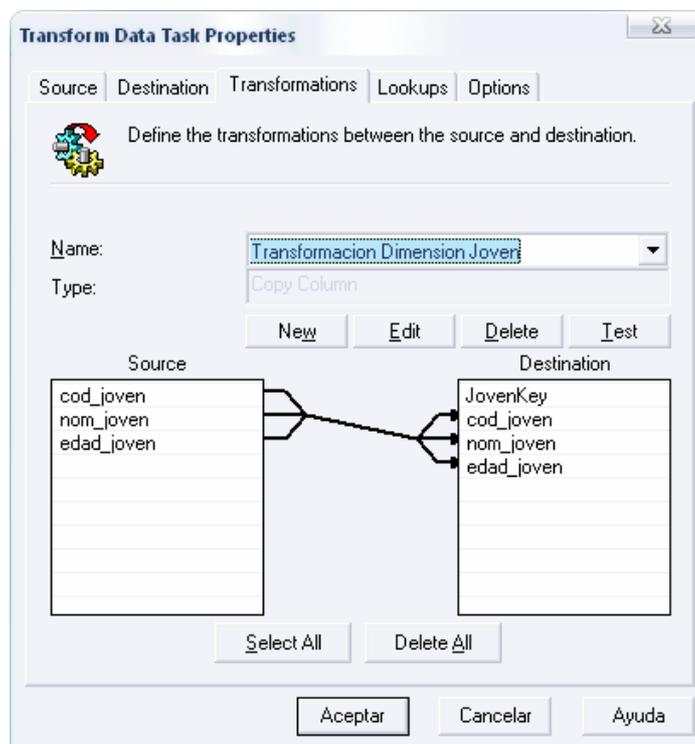


Figura 3.22: Transformación de datos de la Dimensión Joven

- Carga Dimensión Tiempo: esta tarea permite la carga y transformación de los datos de tiempo extraídos de la tabla de ingresos a la tabla Dim_Tiempo.

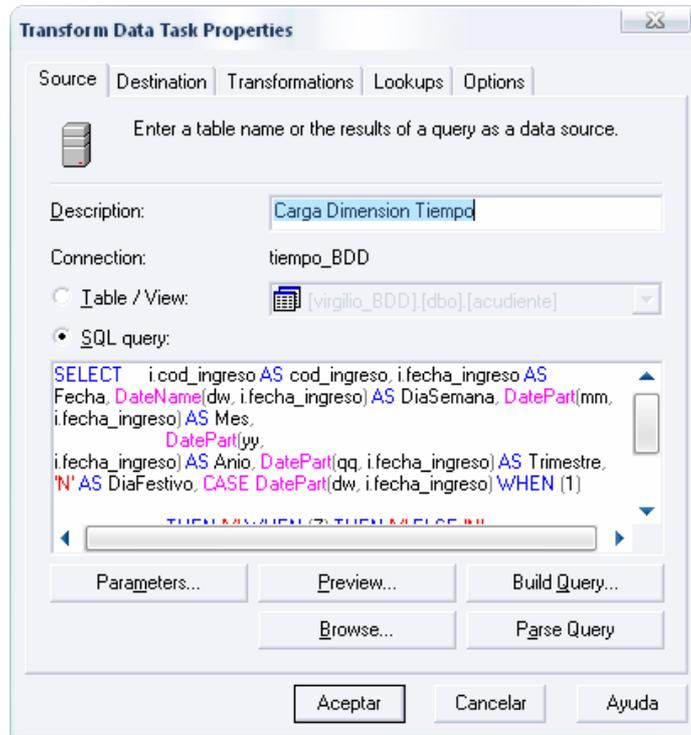


Figura 3.22: Carga de datos de la Dimensión Tiempo

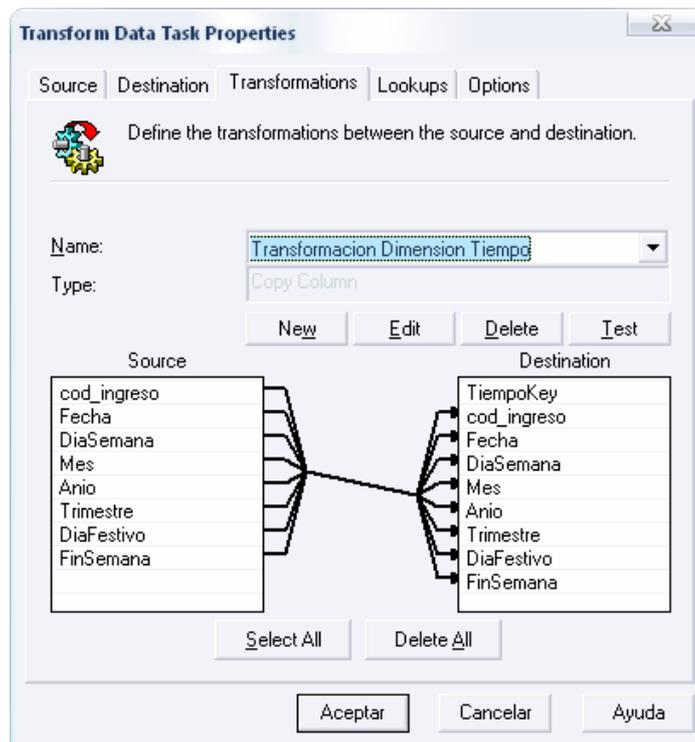


Figura 3.23: Transformación de datos de la Dimensión Tiempo

- Carga Hechos: esta tarea permite la carga y transformación de los datos extraídos de las tablas de ingresos, jóvenes, educación, familia, padres, madres, experiencia con drogas y de las dimensiones a la tabla Hecho_Infracción.

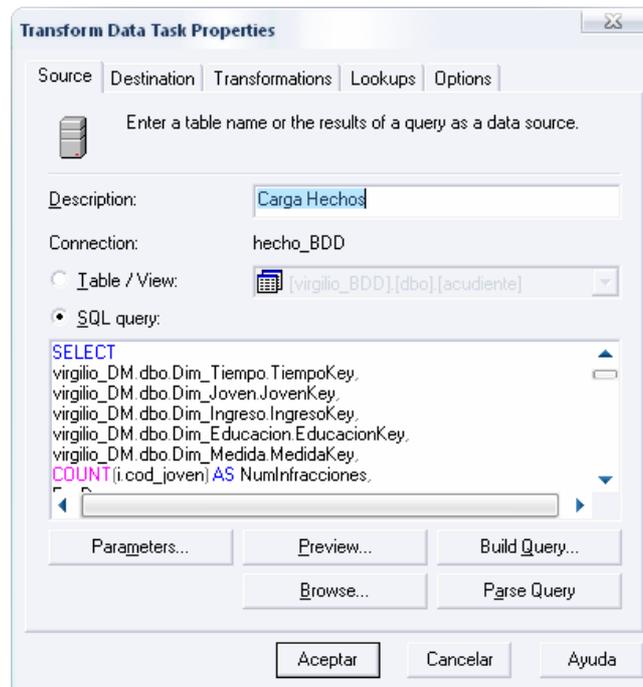


Figura 3.24: Carga de datos a la tabla de Hechos

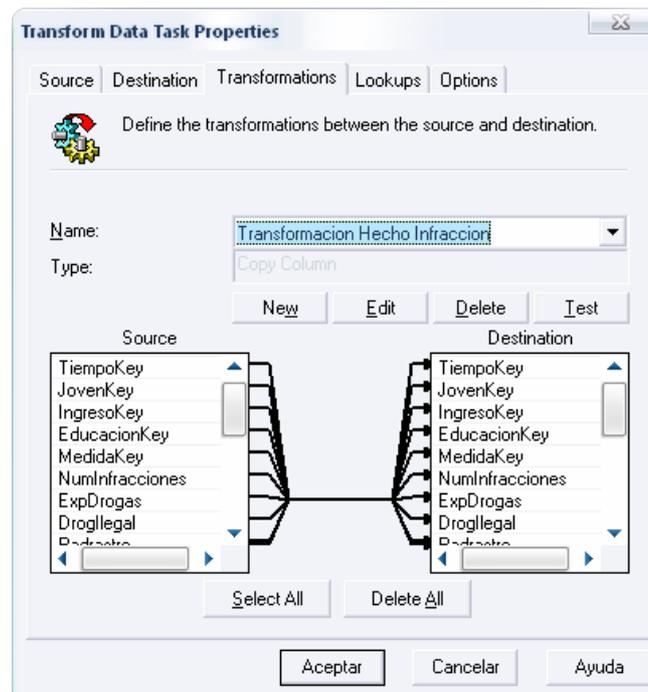


Figura 3.25: Transformación de datos para la tabla de Hechos

Todos estos componentes se han colocado de forma ordenada con el fin de que se ejecute la extracción, carga y transformación de datos de manera correcta y coherente; en un solo paquete DTS con el nombre de “virgilio_DTS”. Luego se creará un paquete DTS que ejecute el paquete Virgilio_DTS y el procesamiento y actualización del cubo multidimensional.

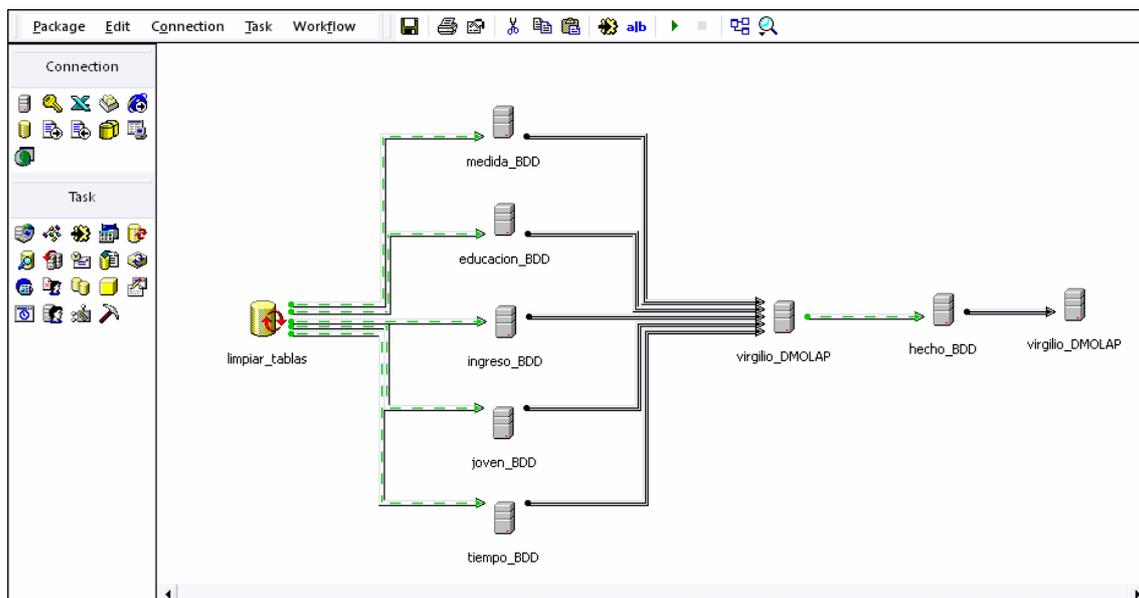


Figura 3.26: Diseño del paquete virgilio_DTS

3.1.1.3 Aplicación de las reglas de limpieza

Las reglas de limpieza de datos fueron implementadas dentro de las tareas del paquete DTS creado en el apartado anterior, según lo especificado en el punto 2.4.2.3.2.

3.1.1.4 Creación de Cubos

En este punto se crean las estructuras multidimensionales o CUBOS OLAP los cuales son necesarios a fin de que la información se almacene según el esquema definido en el apartado 2.4.2.5 y con el fin de proporcionar la visión y características multidimensionales del DM al usuario final

Por lo tanto se crea un CUBO OLAP de nombre “virgilioDMOLAP” con los siguientes componentes:

3.1.1.4.1 Origen de datos

Se crea un origen de datos que conecte a la base de datos del DM (Virgilio_DM) a través de un proveedor de conexión de SQL SERVER.

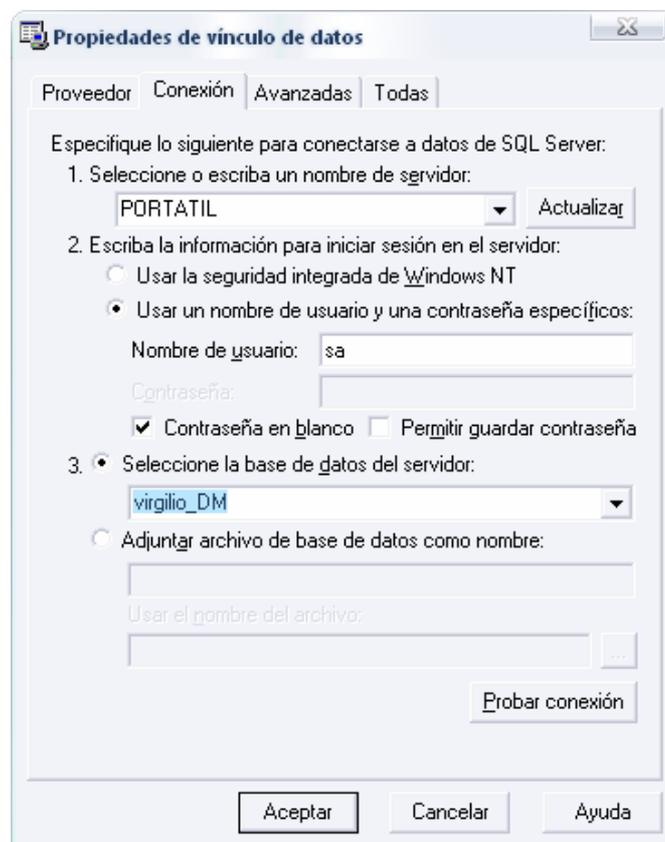


Figura 3.27: Creación del origen de datos para el DM

3.1.1.4.2 Dimensiones compartidas

Se crean cinco dimensiones compartidas, una para cada tabla de dimensión de la base de datos del DM, además de una dimensión adicional que permita analizar los datos del punto de vista de la reincidencia de penas. La dimensión adicional llamada “Reincidencia” se crea en base de la información contenida en la tabla Dim_Ingreso. Las dimensiones se crean con un esquema de agregaciones de columna de la siguiente manera:

DIMENSIÓN	AGREGACION
EDUCACION	Grupo_Asign_Educa └─ Taller_Educa
INGRESO	Grupo └─ Motiv_Ingreso
JOVEN	Edad_Joven └─ Nom_Joven
TIEMPO	Anjo └─ Trimestre └─ Mes
MEDIDA	Desc Medida
REINCIDENCIA	Reincidencia

Tabla 3.2: Dimensiones del CUBO OLAP y agregaciones de columna

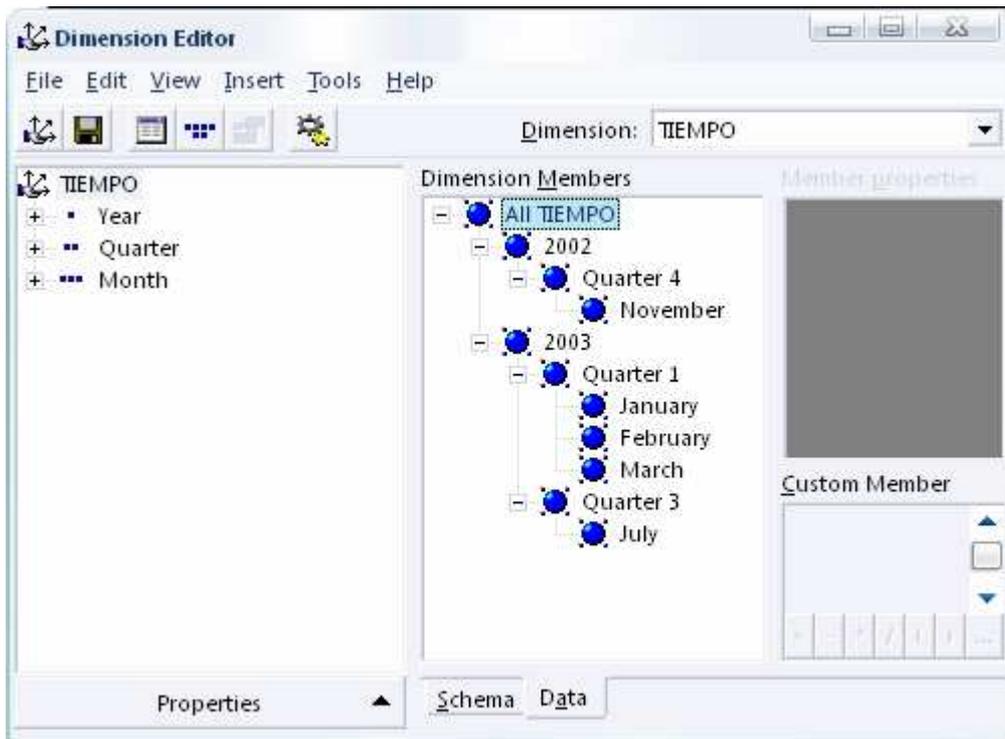


Figura 3.28: Creación Dimensión TIEMPO

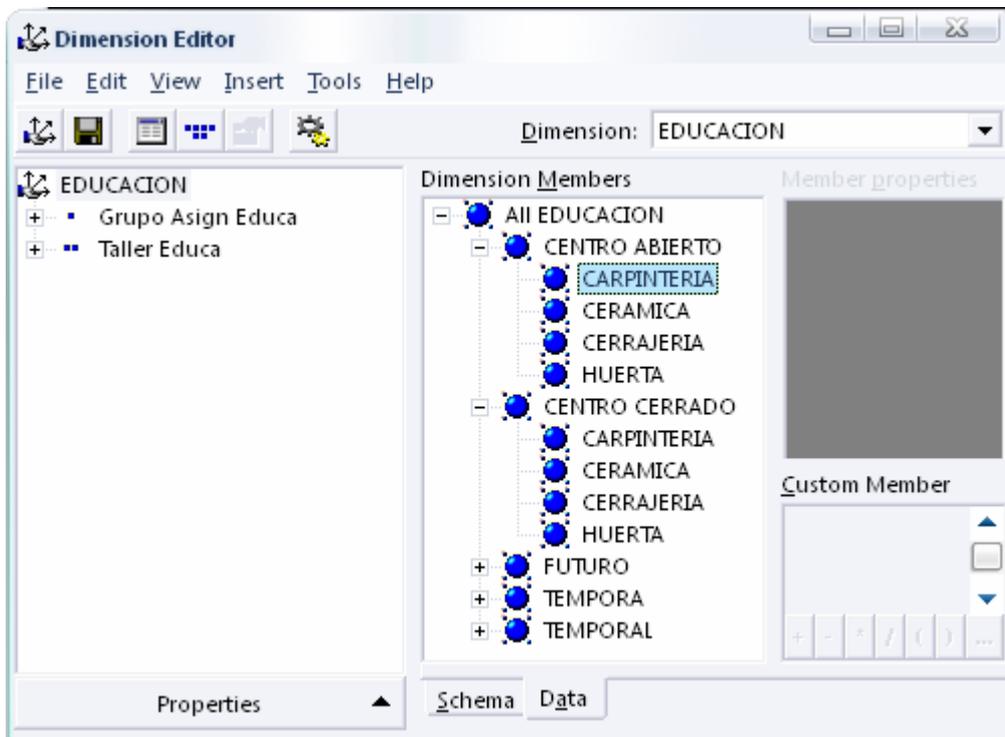


Figura 3.29: Creación Dimensión EDUCACION

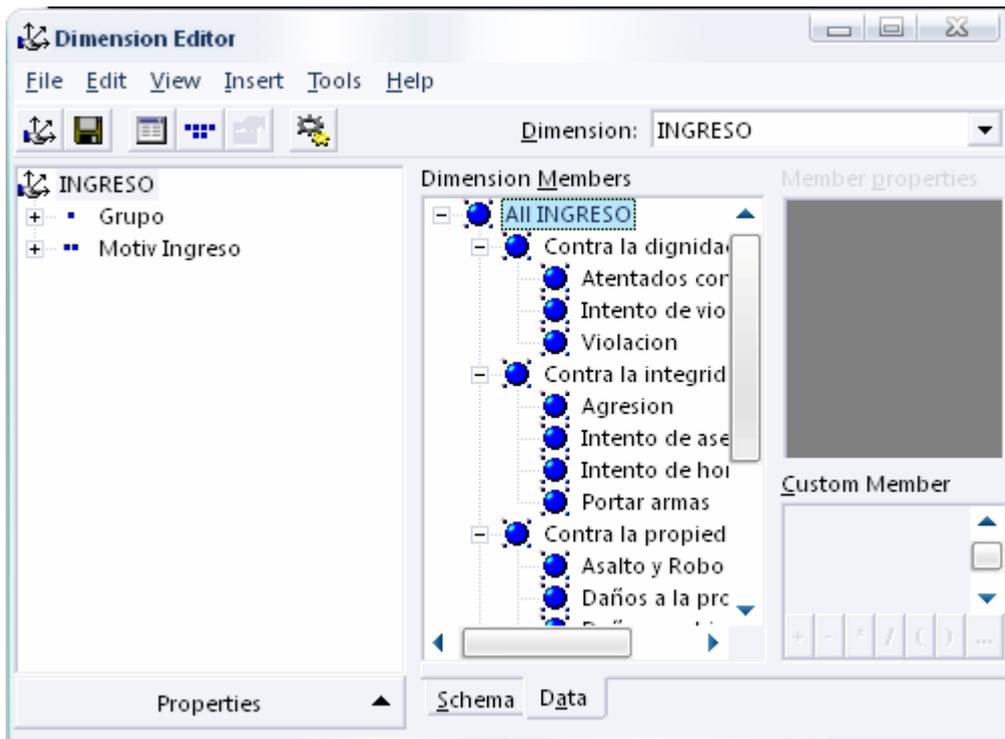


Figura 3.30: Creación Dimensión INGRESO

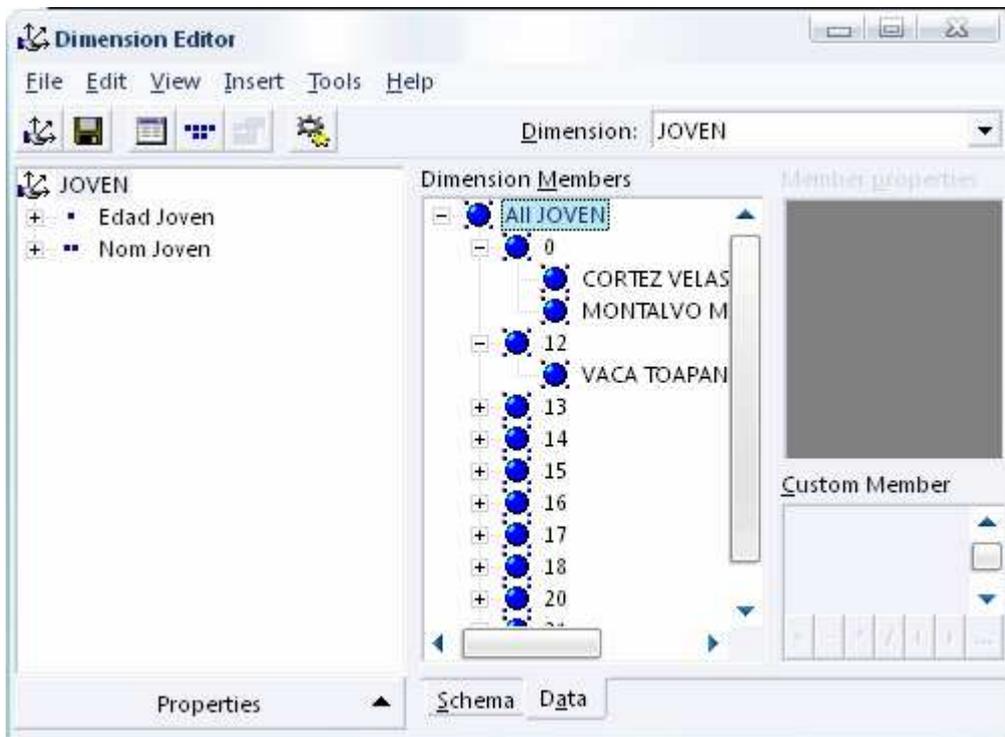


Figura 3.31: Creación Dimensión JOVEN

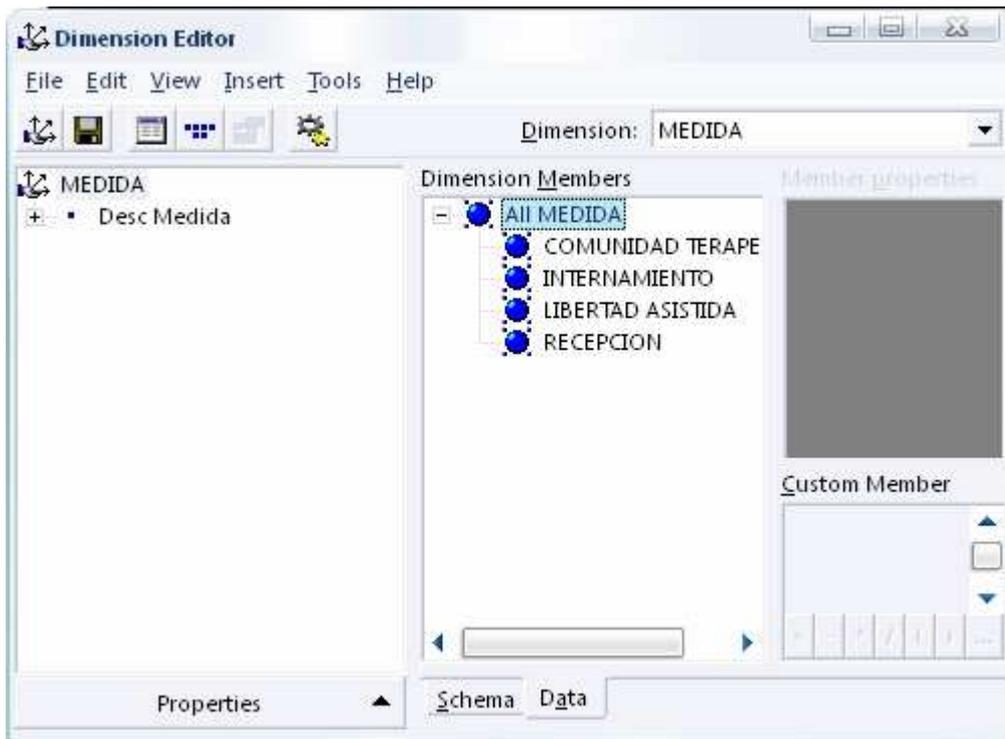


Figura 3.32: Creación Dimensión MEDIDA

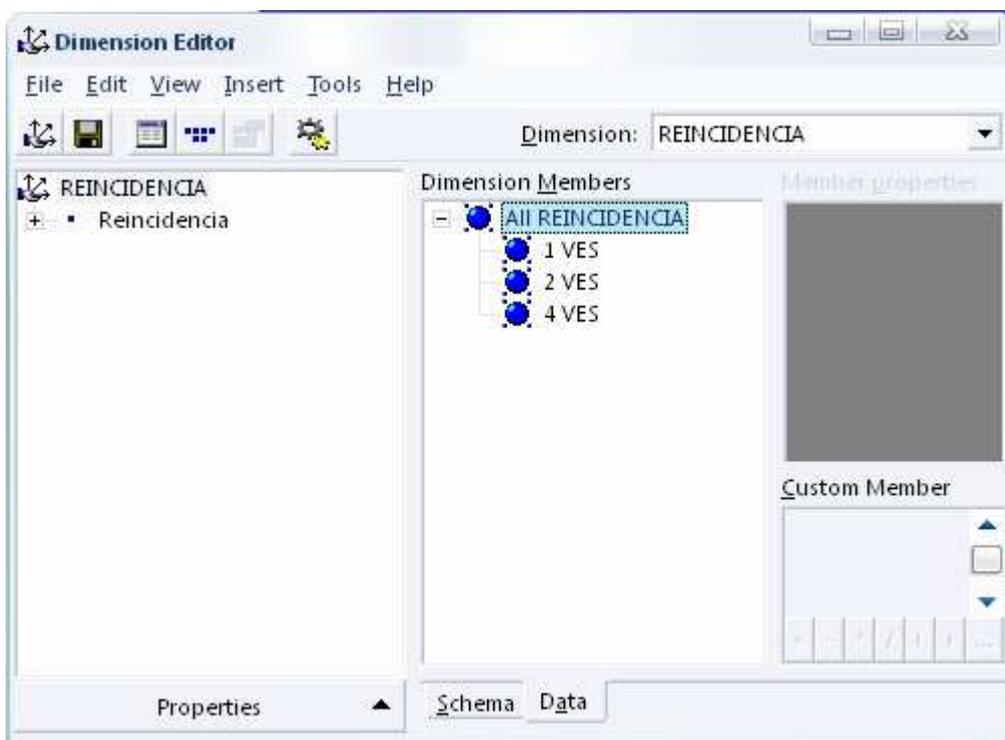


Figura 3.33: Creación Dimensión REINCIDENCIA

3.1.1.4.3 Cubo de datos

Se crea un CUBO OLAP de nombre “virgilio_CUBO” que utilice la conexión creada y las dimensiones respectivas para mostrar la información multidimensional que el DM requiere.

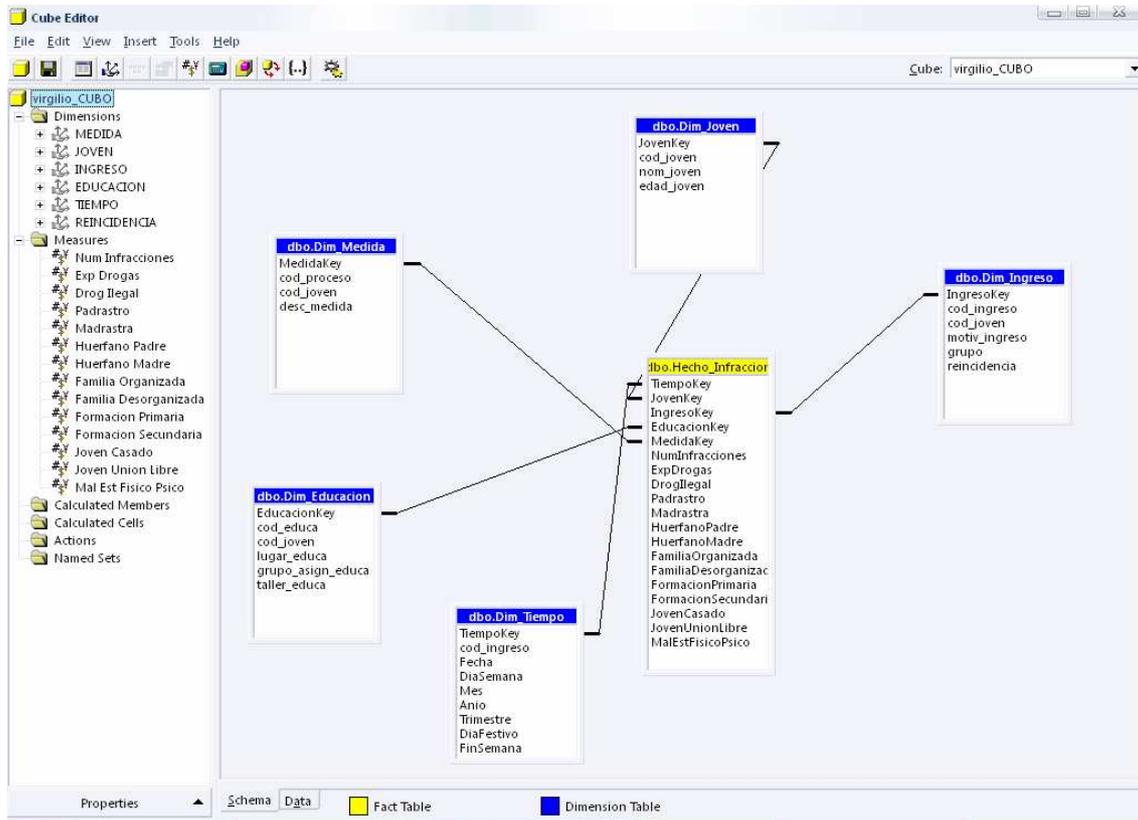


Figura 3.34: Creación del CUBO OLAP

De esta manera se obtiene una visión multidimensional de la información almacenada en el DM lo cual facilita el análisis y presentación de la información

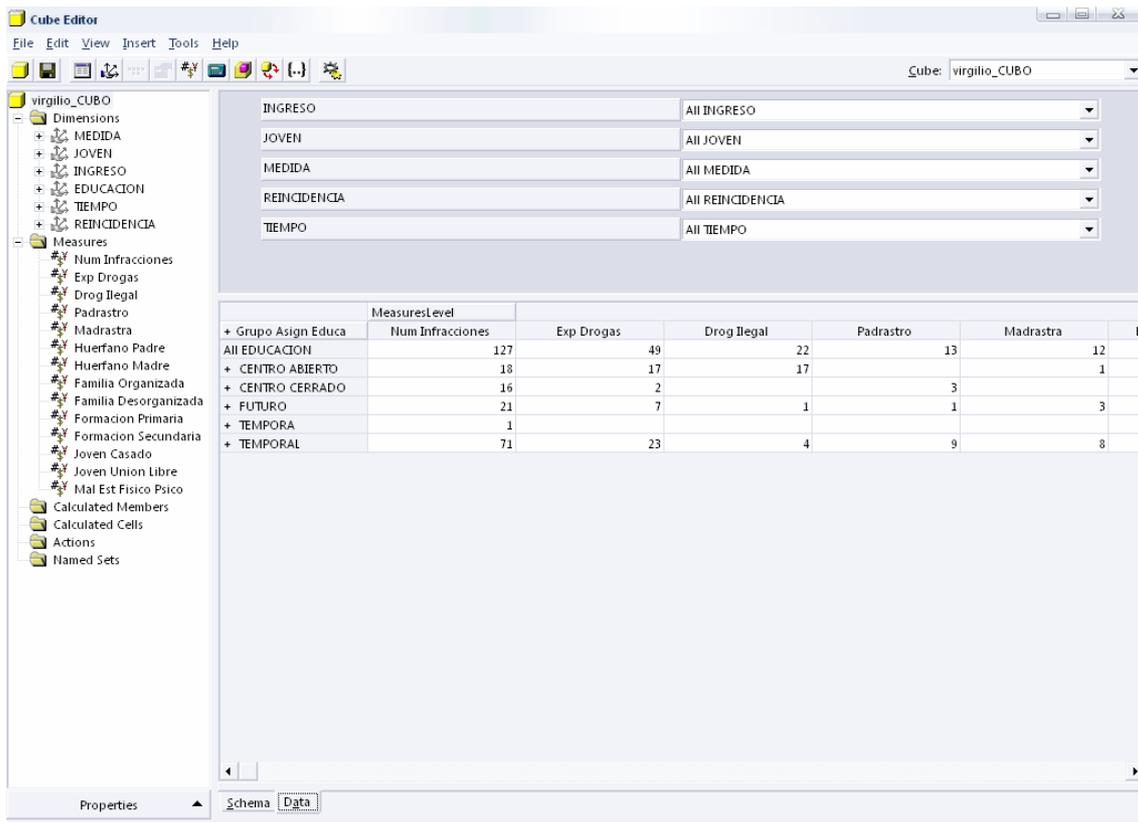


Figura 3.35: Visión multidimensional del CUBO OLAP

A fin de cumplir con el esquema de almacenamiento escogido para este DM se procede de la siguiente manera:

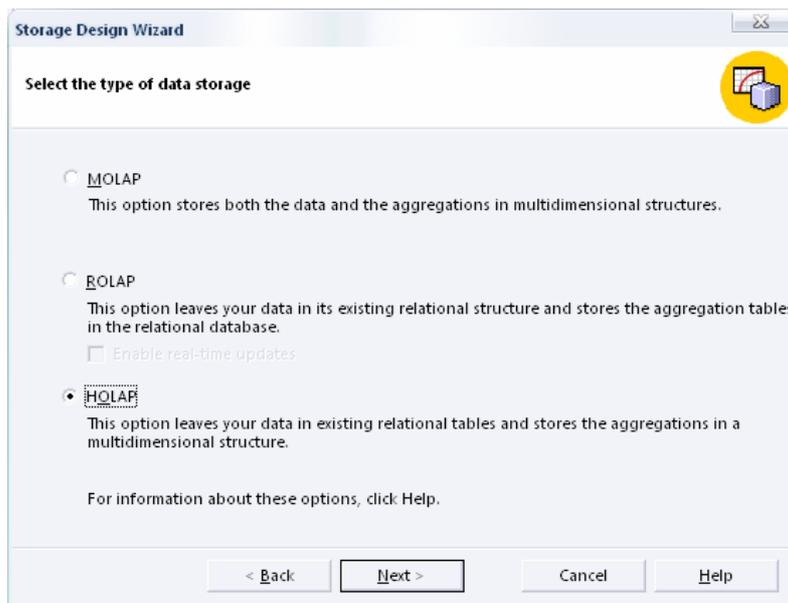


Figura 3.36: Esquema de almacenamiento.

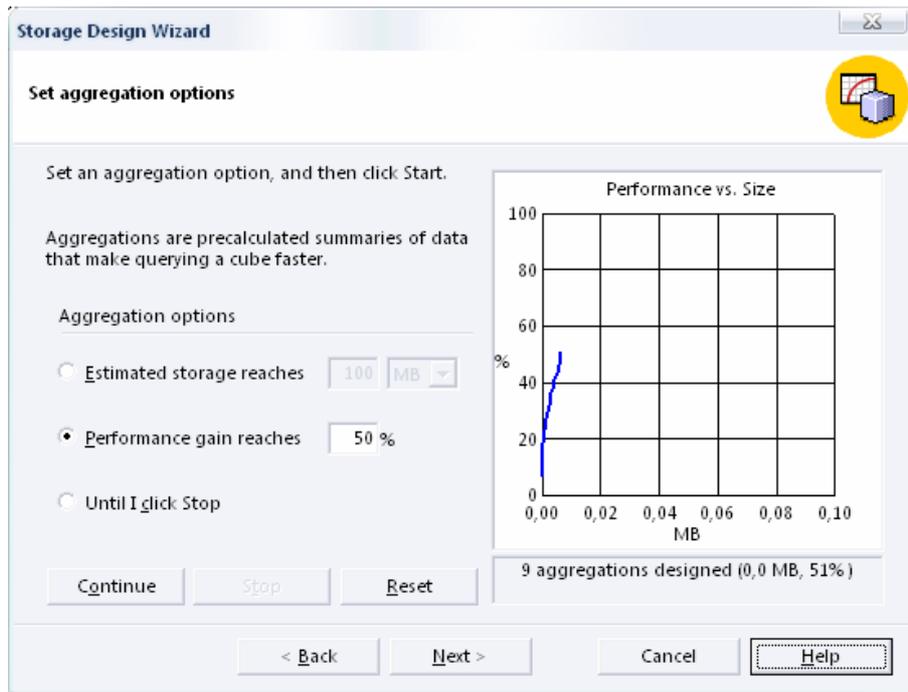


Figura 3.37: Aplicación del esquema para las agregaciones.

Una vez hecho esto se procede a diseñar un último paquete DTS que ejecute el paquete virgilio_DTS y que concluida dicha tarea con éxito procese el CUBO OLAP para su posterior explotación. Este paquete será nombrado como Virgilio_TOTAL

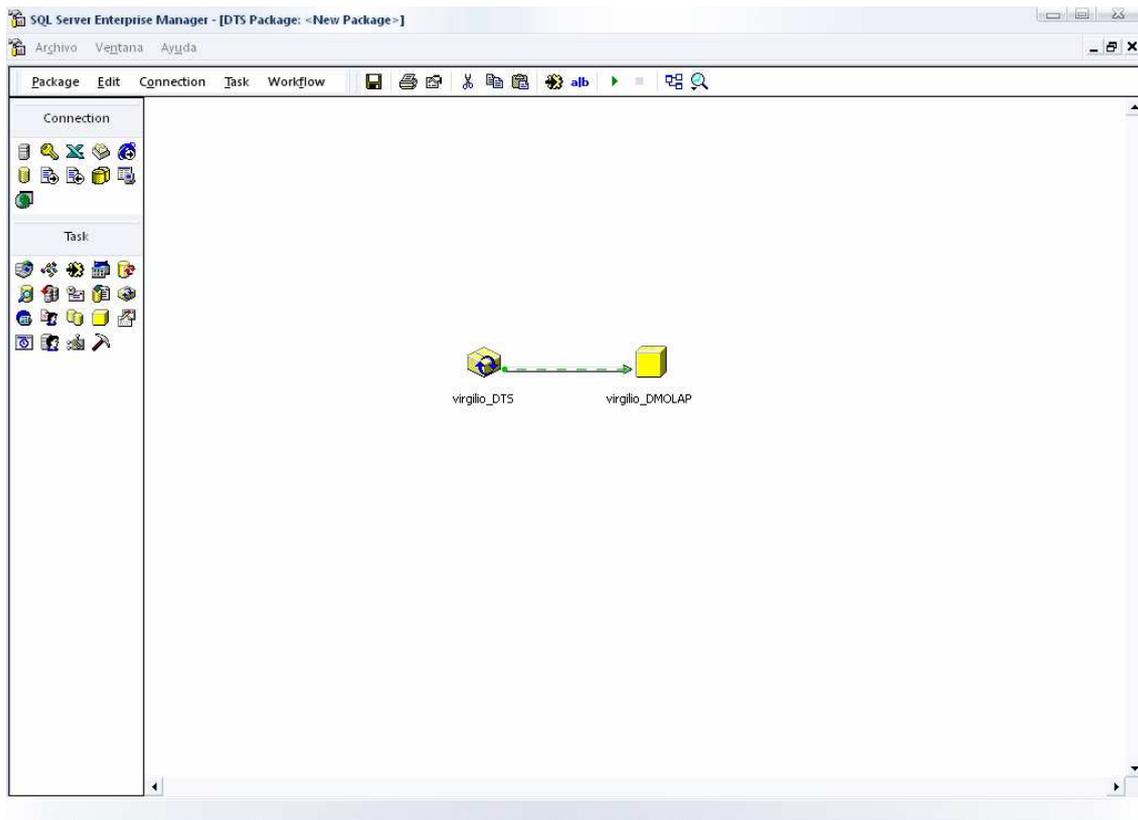


Figura 3.38: Creación del paquete de carga TOTAL.

De esta manera el paquete de carga creado puede calendarizarse a fin de que se refresquen los datos y se pueda tener información diaria.

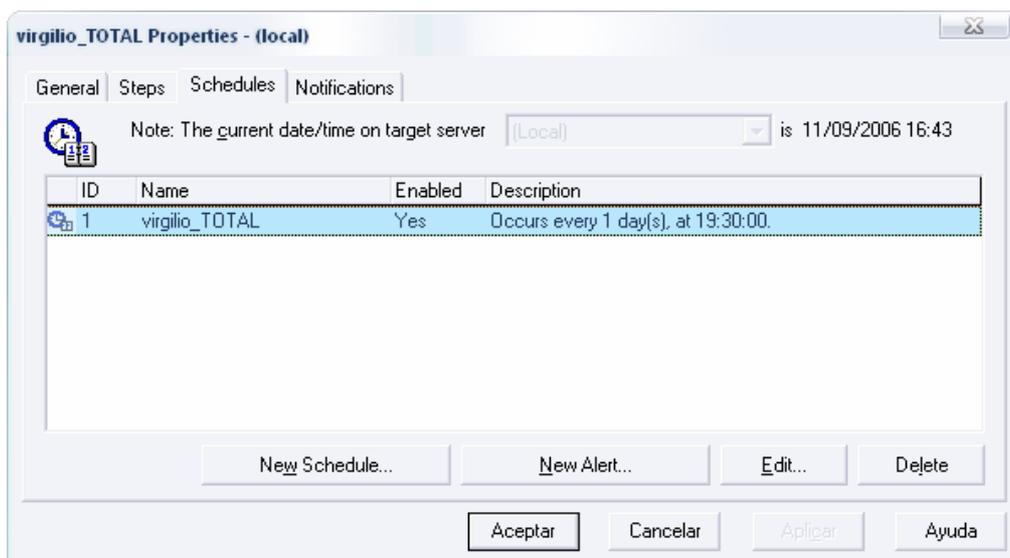


Figura 3.39: Calendarización del paquete de carga TOTAL.

3.1.2 Desarrollo de Aplicación de Usuario final

Para este propósito se procede a crear informes en base a tablas dinámicas en MS Excel 2000; según lo estipulado en los indicadores recolectados en el punto 2.3.3.2.

Inicialmente se procede a crear una conexión hacia un Cubo de OLAP y en este caso hacia el cubo Virgilio_DMOLAP:

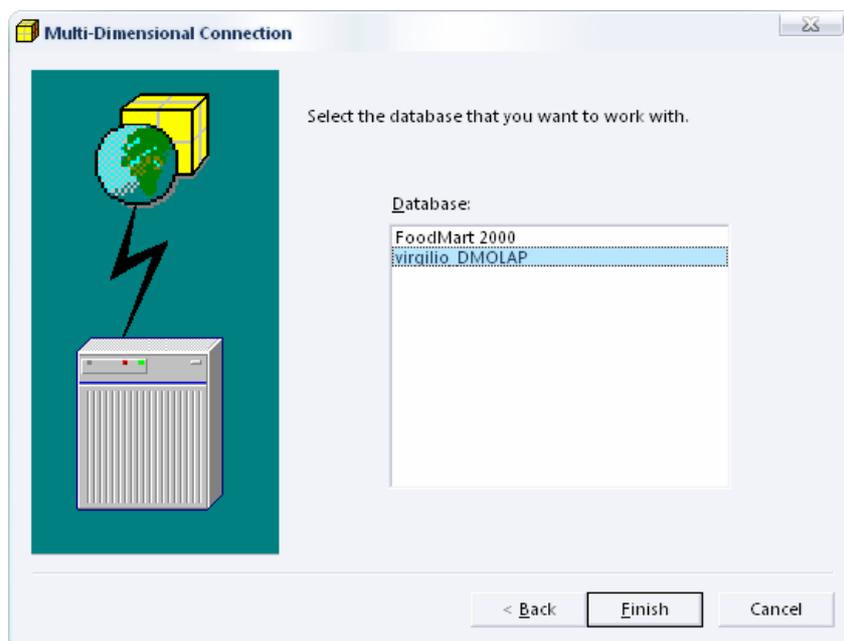


Figura 3.40: Conexión al servidor del CUBO OLAP.

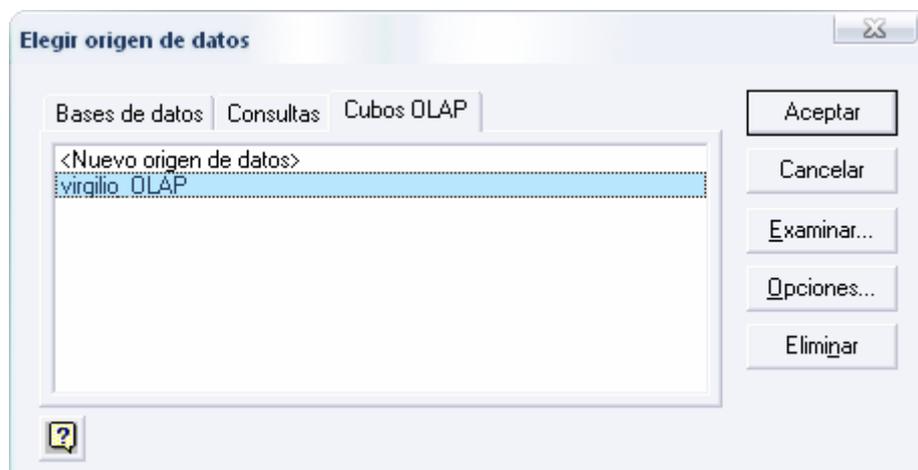


Figura 3.41: Creación de una conexión a un origen de datos externo a MS Excel 2000.

Luego se crea una serie de tablas dinámicas que accedan al cubo y lo muestren en tablas multidimensionales según la información requerida en los indicadores:



Figura 3.42: Diseño de la tabla dinámica.

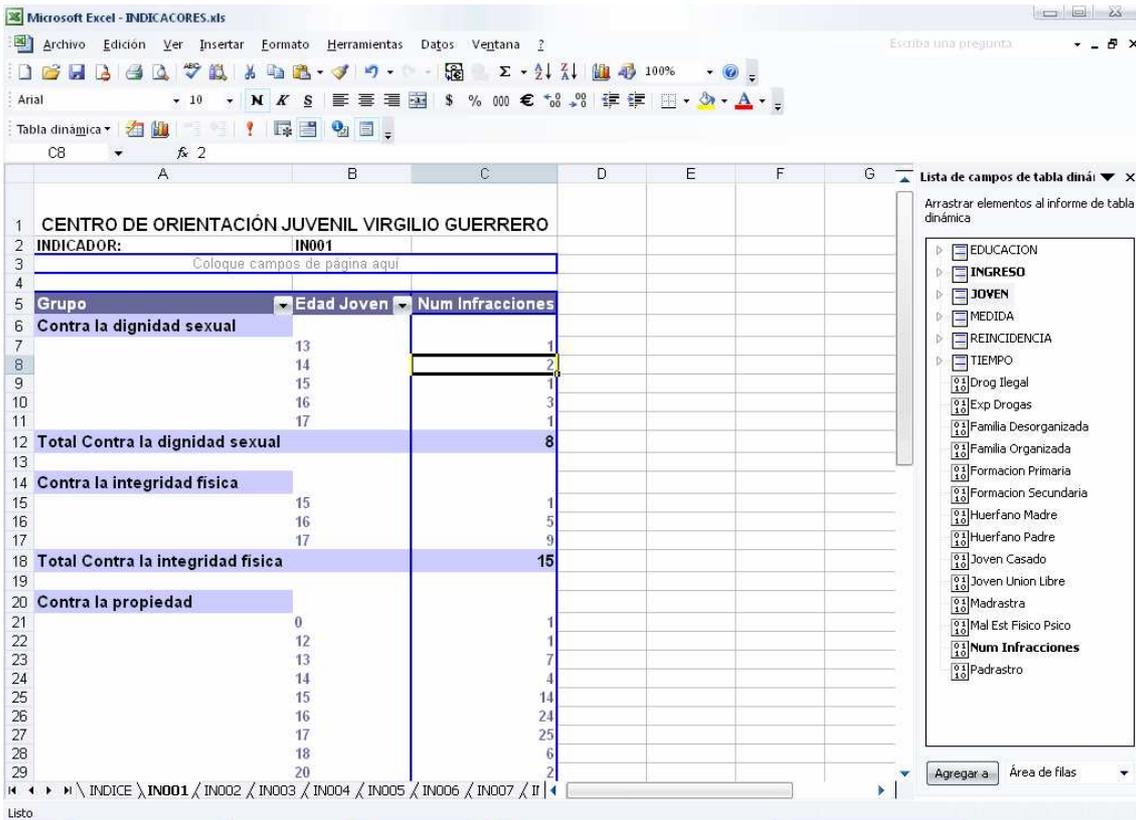


Figura 3.43: Ejemplo de reporte para el indicador IN001

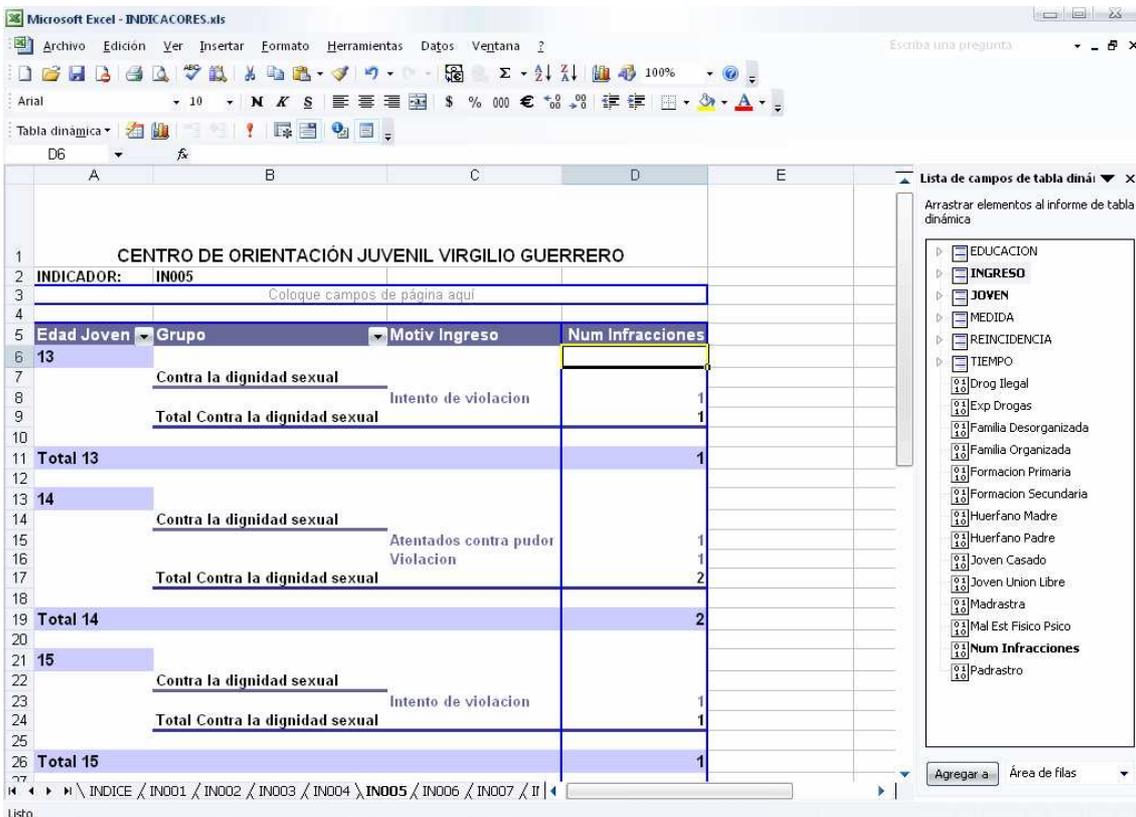


Figura 3.47: Ejemplo de reporte para el indicador IN005

Para mejorar la navegabilidad de los indicadores se crea una hoja de Excel con un índice cuyas entradas sean hipervínculos hacia las hojas donde están los reportes de los indicadores e hipervínculos desde los reportes hacia el índice:

CENTRO DE ORIENTACIÓN JUVENIL VIRGILIO GUERRERO				
INDICADORES DE GESTIÓN				
CODIGO	NOMBRE	DESCRIPCION	PROCESO	
IN001	Número de jóvenes infractores por edades y por grupo de infracción	Este indicador sirve para el análisis de carácter psicológico a fin de determinar a que edad ocurre la mayor incidencia de una infracción determinada y actuar en campañas de prevención temprana	PRC006	
IN002	Número de jóvenes infractores por edades y por grupo de infracción a través del tiempo	Este indicador permite identificar como se ha dado la incidencia de un grupo de infracción en determinadas edades a través del tiempo a fin de tener un soporte histórico para la toma de decisiones en cuento a campañas de prevención	PRC006	
IN003	Reincidencia de infracción de los jóvenes por edades a través del tiempo	Este indicador permitirá conocer de manera histórica cual ha sido las edades en las cuales los jóvenes reinciden en una determinada infracción a fin de mitigar dichos hechos con procesos de apoyo psicológico ya sea a nivel de colegios, escuela o campañas en la comunidad	PRC006	
IN004	Numero de jóvenes infractores catalogados por edad y medida impuesta	Este indicador permite conocer la tendencia de los tribunales a aplicar una medida específica según la edad a fin de evaluar dichas sentencias y proponer mejoramiento en las leyes de ser el caso	PRC006	
IN005	Edades en los que hay una mayor incidencia de	Este indicador permite conocer a que edades los jovenes infractores cometen este tipo de delitos lo cual puede ayudar a la identificación		

Figura 3.48: Índice de reportes de indicadores

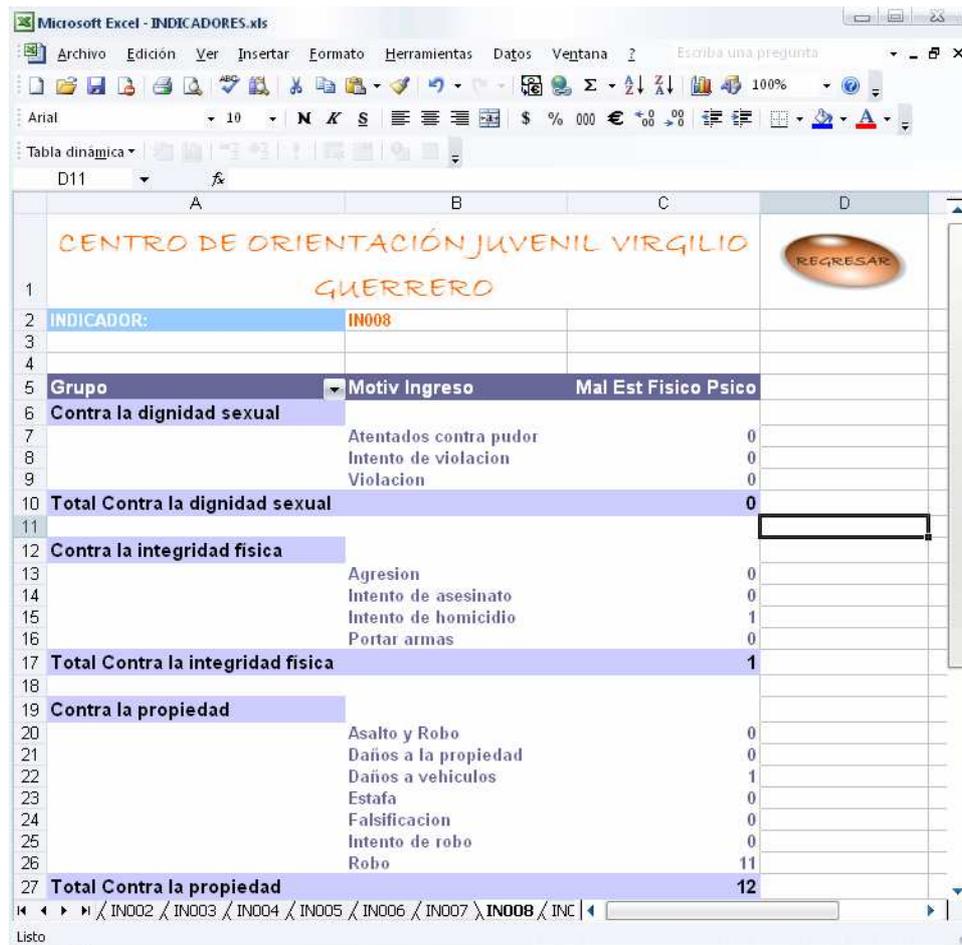


Figura 3.49: Ejemplo de reportes con botón de navegación

3.2 Pruebas

La metodología escogida para el desarrollo del presente proyecto no especifica una documentación ni un conjunto de pruebas específicas para esta etapa de desarrollo del DM por lo que se procede a realizar un conjunto de pruebas aplicables a un desarrollo de software convencional y se escoge un conjunto de pruebas sugerido por YOURDON; ya que se consideran las mas adecuadas para este tipo de proyectos y que aseguran la calidad del mismo, por lo que se opta por realizar pruebas funcionales, de integración y de desempeño. Vale la pena indicar que no se plantean pruebas de recuperación ya que estas se

realizarían siempre en una fase de producción plena del DM la misma que sale fuera del alcance del presente proyecto.

3.2.1 Pruebas de Integración

Estas pruebas se refieren a la correcta interacción y acoplamiento entre las diferentes plataformas tecnológicas en las que se ha implementado los distintos módulos del DM; el escenario esperado para dichas pruebas es el siguiente:

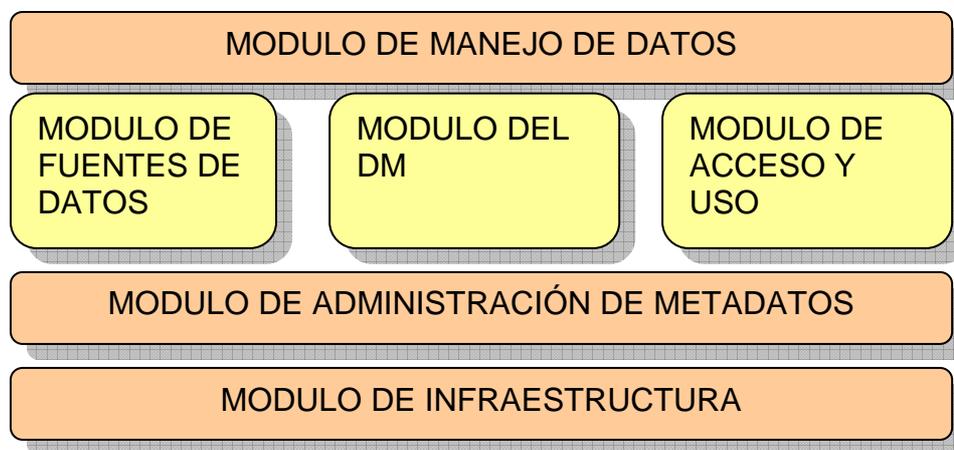


Figura 3.30: Escenario para pruebas de integración.

Debido a que los módulos de:

- Manejo de Datos
- Fuente de Datos
- Del Data Mart

Se implementaron utilizando las herramientas de MS SQL Server 2000; esto nos asegura la integración entre los mismos ya que pertenecen a la misma plataforma.

La integración manifiesta entre el modulo de infraestructura y los tres módulos mencionados, además del módulo de acceso y uso queda asegurada debido a

que se maneja una misma plataforma tecnológica, ya que el módulo de acceso y uso que es MS Excel 2000 tiene una infraestructura diseñada específicamente para interactuar con el MS SQL Server 2000.

El módulo de administración de metadatos también pasa las pruebas de integración debido a que se opta por la utilización del repositorio propio del MS SQL Server 2000.

3.2.2 Pruebas Funcionales

Estas pruebas con sus respectivos reportes de resultados se especifican en el ANEXO D; donde se abarca toda la funcionalidad solicitada para el DM a través de los indicadores y que en definitiva determina los criterios de aceptación del mismo.

3.2.3 Pruebas de Desempeño

Según lo estipulado en el apartado 2.1.6 se procedió a configurar el esquema de almacenamiento y agregación para que el porcentaje de ganancia de rendimiento sea del 50%; es decir que del tiempo de respuesta al procesar del CUBO se reduciría a la mitad con esta configuración, tomando en cuenta que este tipo de esquema requiere un mayor espacio de almacenamiento. En este sentido se procedió a ejecutar una corrida del sistema completo teniendo los siguientes resultados:

- Se procesaron 121 registros de jóvenes con sus respectivos datos en las tablas conexas y con 5 agregaciones configuradas para el DM. Se

obtuvo un tiempo de respuesta de 43 segundos. Adicionalmente se indica que los datos estadísticos del Centro de Orientación Juvenil “Virgilio Guerrero”, estipulan que el crecimiento diario en la población del Centro es de 7 casos por día como máximo.

- Con esta información se estima que, para una tasa de crecimiento de 7 jóvenes por día, a un año de funcionamiento del DM el tiempo de respuesta sería de 15,13 minutos
- Ahora considerando un porcentaje de error del +/- 50%; el tiempo de respuesta del DM estaría entre 7,57 a 22,70 minutos.

CAPITULO IV

4 Caso de Estudio.

El presente proyecto fue desarrollado en su totalidad tomando en cuenta el ambiente, información y características particulares del Centro de Orientación “Virgilio Guerrero”; generalizando dichas características, además de la naturaleza misma de este tipo de institución, para los demás centros existentes en el país.

4.1 Descripción del Ambiente

El ambiente en el cual funcionará el DM, quedó definido en los apartados: 2.1, 2.1.1, 2.1.6.1.1, 2.1.6.1.2, 2.1.6.1.3, el cual nos aclara que no existe una estructura de tecnología de información muy grande y mas bien se opta por un desarrollo oportunista que permita mostrar al usuario final las bondades del uso de un DM para el mejor análisis de la información estadística del Centro.

4.2 Personalización del sistema

Como se indicó en los apartados citados anteriormente, el Centro de Orientación Juvenil “Virgilio Guerrero”, es la institución que mejor estructura informática mantiene de entre los centros de este tipo y es por eso que se escogió desarrollar el DM en base a la estructura de información que mantiene el mismo por lo que el DM está personalizado ya para trabajar en el Centro de Orientación Juvenil “Virgilio Guerrero” y en vista que los demás Centros del

país no mantienen una estructura informática formal, estos podría adaptar su estructura al manejo informático que se tiene en el “Virgilio Guerrero”.

4.3 Presentación de Resultados

Para el desarrollo del presente proyecto no se contempla la etapa de Implantación y mantenimiento formal según lo estipulado en la Metodología elegida ya que esto correspondería al tiempo en que el DM estaría en producción y el alcance de éste proyecto no abarca el monitoreo de dicha instancia por lo que los resultados que se pueden presentar son solo los que arrojan las pruebas diseñadas para este trabajo principalmente las de funcionalidad que son las que establecen los criterios de aceptación que se tendrían de acuerdo a los requerimientos recolectados de los personeros del Centro de Orientación Juvenil “Virgilio Guerrero”. Quedando como potestad del Centro la adecuada utilización del DM

4.4 Conclusión

El presente proyecto tiene la finalidad de mostrar a los personeros de primer mando del Centro de Orientación Juvenil “Virgilio Guerrero” las bondades del uso de una herramienta como el DM para el análisis de información estadística misma que ha sido mantenida en forma manual hasta la presente fecha y ha constituido uno de los principales ejes de acción del Centro mostrando indicadores ya sea a las contrapartes externas como a los personeros internos de la institución. Por lo que se concluye que el DM contribuirá en gran medida al trabajo del Centro constituyéndose en una herramienta de uso crítico, como

es su naturaleza, en muy poco tiempo y que incluso abrirá la puerta para el desarrollo de posteriores trabajos dentro del área de las Tecnologías de la Información y que obligará a un replanteamiento de la significancia de éste dentro del quehacer del Centro.

CAPITULO V

5 Conclusiones y Recomendaciones.

5.1 Conclusiones

- Las herramientas utilizadas, si bien es cierto no son las últimas versiones de sus respectivas líneas de producto, sirven para la implementación de proyectos actuales sin problemas de integración y adaptabilidad a las plataformas nuevas.
- Se puede utilizar versiones anteriores para lograr un producto de interés actual debido a que el concepto de Data Warehousing ya tiene varios años en el ámbito de las Tecnologías de la Información y lo importante es el comprender el concepto del Data Warehousing independiente de las plataformas tecnológicas para su implementación
- Para un proyecto de creación de un Data Mart lo importante es el Diseño y la orientación hacia la arquitectura de datos que se requieren para su construcción; y se deben buscar procesos de desarrollo adecuados al nivel de complejidad en la creación de dicho Data Mart
- Las herramientas que se emplearon son de fácil utilización debido a su orientación intrínseca hacia el usuario final, lo cual facilitó la tarea de

integración de plataformas y facilitará la capacitación de dichos usuarios finales.

- El presente proyecto fue desarrollado utilizando como eje transversal la naturaleza y estructura informática del Centro caso de estudio por lo que se limitó su complejidad.
- La metodología utilizada es bastante formal, disciplinada y bibliográficamente justificada; ya que permitió cubrir todas las etapas convencionales para la construcción de software, desde los aspectos generales hasta los más específicos con un modelamiento y documentación técnica adecuadas y adaptadas a la naturaleza del presente proyecto.
- La metodología empleada no propone la construcción del Data Mart o Data Warehouse en base a una plataforma tecnológica específica sino que profundiza en la comprensión de los conceptos principales para la construcción de los mismos
- De la información que arroja el Cubo implementado en este proyecto, se puede concluir que la información es de importancia estratégica no solo para tomar decisiones operativas y logísticas dentro del centro sino también para la toma de decisiones estratégicas a nivel de Ministerios, Entidades de Control y Gobierno Nacional en lo referente a la atención a la niñez y la adolescencia.

- La información que arroja el Data Mart, implementado en el presente proyecto, permitirá acelerar los procesos de análisis estadístico para la creación de proyectos con financiamiento nacional o extranjero en bien de la niñez y la adolescencia
- El presente proyecto bien puede replicarse como una herramienta de uso genérico en instituciones de la misma naturaleza que el Centro caso de estudio.
- Es importante concluir que el presente proyecto permite la exposición de las habilidades científicas, técnicas y conceptuales que se necesitan para justificar la utilización de personal de nivel de ingeniería de sistemas quienes bien pueden incluirlo en proyectos más integrales y en ambientes complejos.

5.2 Recomendaciones

- Se debería estandarizar procesos propios de desarrollo de Data Warehouse o Data Marts debido a su difusión actual, posición estratégica actual en el Campo de las Tecnologías de la Información y la utilización de conceptos propios. Ya que la adaptación de procesos de desarrollo de software convencionales pueden retrasar la culminación de este tipo de proyectos.

- En los casos en los que las empresas que requieren una herramienta de análisis estratégico de la información sin una experiencia previa en la utilización de éstas, se debería partir por la construcción de uno o varios Data Marts dependiendo del caso y no por la creación de un Data Warehouse completo ya que con esto se ganaría en adaptación de los usuarios finales al uso de este tipo de herramientas y facilitaría una futura integración de dichos Data Marts en un Data Warehouse futuro.
- Para asegurar el éxito en la implementación de un Data Mart se debe procurar la utilización de plataformas integradas a fin de minimizar los híbridos que se producen por el uso de diferentes herramientas y el posterior costo de mantenimiento.
- Es importante que el usuario final utilice herramientas de explotación sencillas y fáciles de utilizar y comprender, con una interfaz amigable y una capacitación adecuada que potencie sus habilidades y destrezas en el uso de dichas herramientas.
- La utilización de una metodología específica, para la construcción de Data Marts o Data Warehouse, siempre se lo debe hacer orientándose a la independencia de la plataforma tecnológica para su implementación.
- Se deben incluir en el cuestionario que debe responder el Data Mart para su construcción, preguntas de carácter estratégico y frecuente; ya

que las preguntas furtivas bien pueden ser respondidas ya sea con uso de reportes planos convencionales o incluso con el uso de la herramienta del Data Mart pero con la consecuente consideración por parte del usuario final y administrador, del tiempo de respuesta a dicha pregunta por parte de esta herramienta.

BIBLIOGRAFIA

- COREY Michael, ABBEY Michael. Oracle Data Warehousing. Oracle Press. 1997.
- SIMON Alan. 90 DAYS TO THE DATA MART. John Wiley Press. 1998.
- MICROSOFT. Microsoft SQL Server 7.0 Data Warehousing Training. Microsoft Press. 1999.
- Centro de Orientación Juvenil Virgilio Guerrero. El Amigo – Diario Anual del Centro. Imprenta del Centro de Orientación Juvenil Virgilio Guerrero.2003.
- PRESSMAN Roger. "Ingeniería de software, un enfoque práctico", editorial McGraw-Hill, 2005.
- JACOBSON, BOOCH, RUMBAUGH, El Proceso Unificado de Desarrollo de Software, Addison-Wesley, 1999.
- JACOBSON, BOOCH, El Lenguaje Unificado de Modelado. Manual de Referencia, Addison-Wesley, 1999.
- KIMBALL Ralph, The Data Warehouse Lifecycle Toolkit, John Wiley & Sons Inc, 1998
- MICROSFOT, Unloocking OLAP With Microsoft SQL Server and Excel 2000, John Wiley & Sons Inc, 2000
- BARQUIN Ramon, EDELSTEIN Herb, Building and managing the data warehouse, Prentice Hall Travel, 1997

- GILL Harjinder, RAO Praskash, Data Warehousing La Integración De La Información Para La Mejor Toma De Decisiones, Ediciones Que, 1996
- BURCH David G. Methodology And Architecture For Data Mart Consolidation. <http://www.tcoug.org/Archive/folio/burch0798.ppt>. 2001.
- E-PLUS CONSULTING. Concepts and DW Methodology. [http://www.e-plusconsulting.com/DW_Concepts/Book I Concepts and DW Methodology/book i concepts and dw metho.html](http://www.e-plusconsulting.com/DW_Concepts/Book_I_Concepts_and_DW_Methodology/book_i_concepts_and_dw_methodology.html). 2000.
- VARIOS AUTORES. Data Mining Concepts. http://www.pcc.qub.ac.uk/tec/courses/datamining/ohp/dm_OHP_final_1.html. 2001
- VARIOS AUTORES. Data Warehousing Training and OLAP Tools. <http://www.microsoft.com/data/oledb/olap/excel.htm>. 1999.
- VARIOS AUTORES, Practical Techniques, Proven Results, <http://www.kimballgroup.com/html/articles.html>. 2006