

UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS VILLAS

FACULTAD DE MATEMÁTICA, FÍSICA Y COMPUTACIÓN

Licenciatura en Ciencia de la Computación



**Mercado de Datos para el Departamento de Cuadros
de la UCLV**

Trabajo de Diploma

Autores: Yudith Ramos Díaz

Karel Alonso Quintanilla

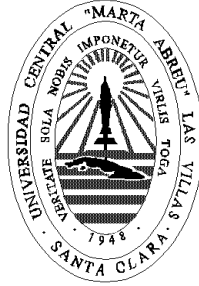
Tutores: Dra. Luisa M. González González

Ing. Liliaesther Delgado López

Santa Clara, Cuba

2014

DICTAMEN



Hacemos constar que el presente trabajo fue realizado en la Universidad Central Marta Abreu de Las Villas como parte de la culminación de los estudios de la especialidad de Ciencia de la Computación, autorizando a que el mismo sea utilizado por la institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos ni publicado sin la autorización de la Universidad.

Firma de los autores

Los abajo firmantes, certificamos que el presente trabajo ha sido realizado según acuerdos de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del tutor

Firma del jefe del
Laboratorio

FRASE:

*El verdadero signo de inteligencia no es el
conocimiento sino la imaginación.*

Albert Einstein.

AGRADECIMIENTOS

**A mis padres por su apoyo incondicional,
a mi tutora por su dedicación, a mi
hermana por estar ahí siempre que la he necesitado,
a Karel por enseñarme tantas cosas, a mis suegros, a Aleidita,
y a todos los que
de una a u otra forma han sido parte de este gran sueño.
Yudith.**

**A mi madre por el amor, el cariño y la paciencia que
siempre ha demostrado.
A mi padre por ser un guía ejemplar.
A mi hermano por estar siempre pendiente de mí.
A mi tutora por los consejos
y a Aleidita por la excelente atención en el Departamento de Cuadros.
Y a mi Yudith por su ternura.
Karel.**

RESUMEN

El Departamento de Cuadros de la UCLV es un órgano rector metodológico que dirige y controla los procesos que se corresponden con el sistema de trabajo de los cuadros y las reservas en dicho centro. Para el almacenamiento de la información se dispone de un sistema transaccional que no permite el tratamiento histórico necesario en el análisis sistemático de tendencias y comportamiento de los datos, útiles para apoyar el proceso de dirección.

Data Warehouse (DWH), es una tecnología para el manejo de la información, que soporta el procesamiento informático y provee una plataforma sólida que permite realizar análisis a partir de datos históricos y actuales. En el presente trabajo se diseñó e implementó un Mercado de Datos (MD) que tomó como fuente de datos el sistema operacional ya existente en el Departamento de Cuadros, contribuyendo a la accesibilidad, disponibilidad de la información y apoyo a la toma de decisiones.

Se utilizó la herramienta de inteligencia de negocio Pentaho 5.0.1 para la visualización y análisis de los datos. Para la validación de los resultados se utilizaron datos reales de dicho departamento.

ABSTRACT

The Department of Cuadros of the UCLV is a methodological rector organ that manages and controls the processes that correspond with the system of work of the cuadros and the reserves in this center. For the storage of the information they have a transactional system that doesn't allow the necessary historical treatment in the systematic analysis of tendencies and behavior of the data, useful to support the manage process.

Data Warehousing is a technology for the handling of the information that supports the computer processing and it provides a solid platform that allows achieving analysis starting from historical and current data.

In the present work a Data Mart, that took as data source the operational system already existent in the Department of Cuadros, was designed and implemented, contributing to the accessibility, readiness of the information and support to the decision making problems.

For the visualization and data analysis it was used Pentaho 5.0.1 a business intelligence tool. For the validation of the results it was used real data of this department.

TABLA DE CONTENIDO

INTRODUCCIÓN	1
CAPÍTULO 1. DESCRIPCIÓN DEL SISTEMA DE TRABAJO CON LOS CUADROS Y SUS RESERVAS EN LA UCLV	5
1.1 Designaciones y funciones del Departamento de Cuadros.	5
1.2 Conceptos propios del trabajo con los cuadros.	6
1.2.1 La reserva de los cargos de cuadros.	7
1.3 Procedimiento de selección y actualización de la Reserva.	8
1.4 Proceso de evaluación de los cuadros.	10
1.4.1 Indicadores del proceso de evaluación.	11
1.4.2 Conclusiones del proceso de evaluación.	14
1.5 Proceso de selección y tramitación de los movimientos de cuadros.	15
1.6 Procedimiento de planificación de demanda de cuadros.	17
1.7 Fuentes de datos en el Departamento de Cuadros.	19
1.8 Conclusiones del capítulo.	21
CAPÍTULO 2. ACERCA DE LOS MERCADOS DE DATOS.	23
2.1 La modelación dimensional.	23
2.2 Dimensiones lentamente cambiantes.	24
2.2.1 Dimensiones lentamente cambiantes tipo 2.	24
2.2.2 Dimensiones lentamente cambiantes: tipos 0, 4, 5.	25
2.3 Refinamiento del modelo.	27
2.3.1 Dimensiones junk.	27
2.3.2 Las tablas puentes (bridge): su utilización en jerarquías y relaciones muchos a muchos.	29
2.3.3 Esquemas copo de nieve o snowflakes.	33
2.3.4 Tratamiento de nulos en el esquema dimensional	34
2.3.5 Tratamiento de fechas.	36
2.4 Dimensiones que juegan rol.	38

2.5	Dimensiones conformadas.	39
2.6	El proceso de carga de los datos.	41
2.7	Análisis de datos a través de OLAP.	42
2.7.1	Categorías de sistemas OLAP.	43
2.7.2	Mondrian y OLAP.	44
2.8	Conclusiones del capítulo.	45
CAPÍTULO 3. DISEÑO E IMPLEMENTACIÓN DEL MERCADO DE DATOS PARA EL DEPARTAMENTO DE CUADROS DE LA UCLV.		46
3.1	Diseño lógico de la solución.	46
3.1.1	Selección de los procesos del negocio.	46
3.1.2	Declarando el grano.	47
3.1.3	Selección de las dimensiones.	47
3.1.4	Identificación de las tablas de hechos para el proceso de movimiento de cuadros 48	
3.1.5	Identificación de las tablas de hechos para el proceso actualización de la reserva 49	
3.1.6	Identificación de las tablas de hechos para el proceso evaluación de los cuadros 49	
3.1.7	Diseño del Mercado de Datos.	49
3.2	Implementación del MD para el Departamento de Cuadros.	53
3.3	Procesos de extracción, transformación y carga para poblar el Mercado de Datos 54	
3.3.1	Transformaciones y trabajos diseñados.	55
3.3.2	Automatización del proceso de carga.	62
3.4	Análisis de los datos a través de reportes y técnicas OLAP.	64
3.5	Conclusiones del capítulo.	67
CONCLUSIONES		68
RECOMENDACIONES		69
BIBLIOGRAFÍA		70

LISTA DE FIGURAS:

Figura 1-1 Diferencias entre los sistemas operacionales y los DWH	19
Figura 2-1 Tabla de hechos sin hechos para SCD tipo 2	25
Figura 2-2 SCD tipo 4 usando mini dimensión.	26
Figura 2-3 SCD tipo 5 usando mini dimensión y un Outtrigger de Tipo 1.....	27
Figura 2-4 Ejemplo de dimensión junk para el proceso de evaluación de cuadros.	29
Figura 2-5 Representación de jerarquías de profundidad conocida.	30
Figura 2-6 Jerarquía alternativa usando tabla puente.....	31
Figura 2-7 Utilización de una tabla puente para la relación muchos-a-muchos.	32
Figura 2-8 Ejemplo de tabla puente para el proceso de evaluación de cuadros.	32
Figura 2-9 Copo de nieve permisible.	33
Figura 2-10 Tratamiento de fechas con un outrigger.	38
Figura 2-11 Una tabla de hechos acumulativa donde la dimensión fecha juega distintos roles.....	39
Figura 2-12 Múltiples roles de la dimensión calendar.	39
Figura 2-13 El bus del DWH dentro de sus elementos básicos.	40
Figura 2-14 Ejemplo de una matriz de bus.	40
Figura 2-15 Ejemplo de dimensión conformada.....	41
Figura 2-16 Arquitecturas OLAP.....	43
Figura 2-17 DWH con Modrian.....	45
Figura 3-1 Relaciones de la tabla de hechos fact_eval_anual_cuadros.	51
Figura 3-2 Relaciones de la tabla de hechos fact_evaluacion_integral.	51
Figura 3-3 Relaciones de la tabla de hechos fact_mov_cuadros	52
Figura 3-4 Relaciones de la tabla de hechos fact_estado_reservas.....	53
Figura 3-5 Tablas de la BD en el phpMyAdmin.....	54
Figura 3-6 Transformación para la dimensión persona	55
Figura 3-7 Manejo de las SCD en la dimensión persona.	56
Figura 3-8 Transformación para tabla de hechos fact_mov_cuadros	57
Figura 3-9 Paso Calcular atributos.....	57

Figura 3-10 Paso Calcular Resto Atributos.	58
Figura 3-11 Configuración del paso Búsqueda en Base de Datos.	59
Figura 3-12 Transformación para la tabla de hechos fact_rel_cargo_cuadro.	60
Figura 3-13 Diseño del Trabajo mensual.	60
Figura 3-14 Transformación para la tabla de hechos fact_eval_anual.	61
Figura 3-15 Creación del id_grupo para la tabla puente.	61
Figura 3-16 Utilización del paso Add a checksum.	62
Figura 3-17 Parámetros genéricos para el Kitchen y el Pan.	63
Figura 3-18 Tarea del sistema para la carga mensual.	64
Figura 3-19 Reporte de estabilidad en el cargo.	66
Figura 3-20 Diseño de cubos en el Schema Workbench.	67

INTRODUCCIÓN

En el contexto de la sociedad actual cada vez más globalizada y competitiva es vital el rol que tienen recursos intangibles como los datos, la información y el conocimiento. La información es indispensable para decidir.

En un corto período de tiempo los volúmenes de datos existentes se han multiplicado exponencialmente haciendo muy difícil su manejo. Los sistemas operacionales u otros medios comúnmente utilizados no ofrecen a los usuarios un ambiente lo suficientemente integrado, centralizado, con grandes velocidades de respuesta, de fácil uso y que permitan el análisis eficaz y eficiente de la información y el apoyo a la toma de decisiones; surge de esta manera el concepto de DWH.

Un DWH es una colección de datos orientada a temas, variable en el tiempo, no volátil e integrado (Inmon, 2005). Según (Ponniah, 2003) una definición más amplia no se centra solo en los datos sino además en los medios para obtener y analizar esos datos, para extraerlos, transformarlos y cargarlos, así como las diferentes formas para realizar la gestión de los mismos.

En (Ralph Kimball, 2002a) un DWH se establece como un conjunto de MDs, caracterizados estos por la especificidad de los datos que contienen enfocados en un área determinada del negocio.

La toma de decisiones en cualquier contexto, en particular en un centro educacional, requiere de una base confiable y segura que ayude a los ejecutivos en su tarea de dirección. Los sistemas de soporte a la toma de decisiones tienen como respaldo informacional los DWH para hacer estimaciones, evaluaciones, comparación de alternativas, pronósticos, entre otras operaciones, sobre los datos almacenados; contribuyendo a mejorar las elecciones que los usuarios finales suelen hacer (Alter, 1980).

Debido a su orientación analítica, el DWH impone un procesamiento distinto al de los sistemas operacionales y requiere de un diseño cercano a la visión de los usuarios finales, para que sea más fácil la recuperación de información y la navegación (Trujillo, 2005).

Los cuadros son fundamentales para garantizar el funcionamiento correcto de las empresas y organizaciones del país, pues no son sólo importantes las tecnologías y otros medios materiales sino además el recurso humano del que se dispone. En el Departamento de Cuadros de la Universidad Central “Marta Abreu” de las Villas (UCLV) el análisis de la información recibida se realiza de forma manual y la elaboración de informes es costosa en esfuerzo y tiempo dificultándose la disponibilidad de la información estadística y la realización de análisis a través del tiempo que indiquen tendencias, características, el estado de determinados factores, etc.

Los datos en el Departamento de Cuadros de la UCLV se almacenan en un sistema transaccional cuyo soporte es una base de datos relacional sobre MySQL, la creación de un MD para esta dirección teniendo como base dicho sistema fuente apoyaría en gran medida el trabajo del personal administrativo.

Planteamiento del problema.

El Departamento de Cuadros de la UCLV es una dependencia estratégica para la dirección universitaria, tiene fuentes de datos históricas y no dispone de herramientas para el análisis que incluyan los períodos de tiempo como una variable y potencien la toma de decisiones.

El **objetivo general** de la presente investigación consiste en desarrollar un Mercado de Datos que contribuya a la accesibilidad, disponibilidad de la información y apoyo a la toma de decisiones en el Departamento de Cuadros de la Universidad Central “Marta Abreu” de las Villas usando la plataforma de inteligencia de negocio Pentaho 5.0.1.

Objetivos Específicos:

1. Caracterizar el caso de estudio del Departamento de Cuadros de la UCLV.

2. Realizar el diseño dimensional para definir la estructura del MD utilizando la herramienta ERStudio 8.0.
3. Definir e implementar los procesos ETL que permitan poblar el MD utilizando Pentaho Data Integration 4.0.1 (PDI).
4. Diseñar los reportes y los cubos que permitan el análisis de la información.
5. Ofrecer al usuario una interfaz de fácil uso y actualización a través de la Suite Pentaho de Inteligencia del Negocio.

Preguntas de investigación:

1. ¿Cuáles son las características presentes en las fuentes de datos?
2. ¿Cuáles son los hechos y dimensiones necesarios para diseñar el esquema dimensional?
3. ¿Qué transformaciones implementar en el proceso de integración de los datos que garanticen evaluar el estado, la calidad y estructura de los mismos?
4. ¿Cómo definir el manejo de las actualizaciones del mercado de datos valorando la frecuencia de tiempo y los posibles errores en las fuentes?
5. ¿Qué reportes el cliente necesita obtener como resultado de la aplicación?

Justificación de la investigación

Las necesidades identificadas en el Departamento de Cuadros de la UCLV referentes al almacenamiento y tratamiento de la información demuestran que con el presente trabajo se lograrán beneficios en la práctica ofreciendo un mayor nivel de gestión, control, calidad y seguimiento de la información, mayor eficacia y eficiencia en cuanto al procesamiento de datos. El almacenamiento de la información se realizará de forma homogénea y fiable. Se agilizará además la recuperación y elaboración de informes apoyando de esta forma el proceso de toma de decisiones.

Viabilidad de la investigación

Para el desarrollo del presente trabajo se cuenta con el apoyo del personal del Departamento de Cuadros de la UCLV que será el encargado de aportar los requisitos y fuentes de datos para el sistema, siendo viable además porque se cuenta tanto con los medios materiales como humanos necesarios para la realización del proyecto.

El presente trabajo se encuentra estructurado de la siguiente forma:

Capítulo 1: En este capítulo se realiza una descripción del caso de estudio para identificar los requerimientos y necesidades de los usuarios, así como el análisis del sistema fuente.

Capítulo 2: Durante este capítulo se tratan aspectos teóricos importantes a tener en cuenta durante el diseño de un MD y se analizan temas específicos que podrían aplicarse en la solución del problema planteado.

Capítulo 3: Este capítulo se dedica al diseño e implementación del MD para el Departamento de Cuadros de la UCLV, destacándose entre otros puntos el desarrollo de los procesos ETL y la creación de reportes.

CAPÍTULO 1. DESCRIPCIÓN DEL SISTEMA DE TRABAJO CON LOS CUADROS Y SUS RESERVAS EN LA UCLV

El presente capítulo tiene como objetivos identificar las principales características que definen el sistema de trabajo con los cuadros y sus reservas en el Departamento de Cuadros de la UCLV. Se han tomado como referencia documentos que describen los procedimientos realizados con relación a los cuadros y sus reservas, las funciones, atribuciones, objetivos y actividades del departamento así como las Instrucciones sobre el Perfeccionamiento del Sistema de Trabajo con los Cuadros del Estado y el Gobierno de julio del 2010.

1.1 Designaciones y funciones del Departamento de Cuadros.

El Departamento de Cuadros de la UCLV es un órgano rector metodológico subordinado directamente al Rector de dicha entidad, que esta designado para auxiliarlo en la aplicación de la Política de Cuadros aprobada por el Partido Comunista de Cuba y el Gobierno; implementar y controlar su cumplimiento en las diferentes nomenclaturas de cargos; así como ejercer la autoridad funcional hacia el resto de los órganos y especialistas de cuadros en los niveles de la estructura de base; cumplir las funciones y atribuciones comunes que le corresponden y entre las específicas se citan las siguientes:

1. Proponer y tramitar los movimientos y otras decisiones sobre los cuadros y sus reservas.
2. Proponer, de conjunto con los restantes órganos y especialistas el Plan de Preparación y Superación de los cuadros y sus reservas.
3. Participar en los estudios para el perfeccionamiento de la estructura de la Universidad.
4. Proponer la organización, planificación, realización y control del proceso de evaluación de los cuadros y una vez aprobado, controlar su cumplimiento.

5. Asegurar el funcionamiento sistemático de las comisiones de cuadros, tramitar con la calidad necesaria y en el tiempo establecido los temas que son objeto de análisis; registrar los acuerdos y controlar su cumplimiento; elaborar los resúmenes anuales de su labor y las proyecciones de trabajo para el próximo período.
6. Realizar análisis integrales y valoraciones sobre el estado del completamiento cuantitativo y cualitativo de los cuadros y sus reservas y de otros aspectos vinculados al cumplimiento de la Política de Cuadros.

La Comisión de Cuadros es un órgano asesor que tiene como objetivos analizar las decisiones que sobre el trabajo con los cuadros y sus reservas se toman en la Universidad y velar por el cumplimiento de la Política de Cuadros en la UCLV y el resto de las normativas legales.

Para continuar con el desarrollo de la investigación es importante tener en cuenta un conjunto de definiciones que sustentan el trabajo del Departamento de Cuadros.

1.2 Conceptos propios del trabajo con los cuadros.

La definición de cuadro según las Instrucciones sobre el Perfeccionamiento del Sistema de Trabajo con los Cuadros del Estado y el Gobierno de julio del 2010 expresa que: “Es el trabajador que posee la capacidad técnico-laboral requerida, una sólida preparación y habilidad de dirección, disciplina administrativa, constante exigencia, vela por el uso correcto de los recursos puestos a su disposición, cuenta con un conjunto de cualidades político-ideológicas y éticas, asume los principios consagrados en nuestra constitución y el Programa del Partido; es elegido o designado para ocupar cargos de dirección u otros que por la importancia de las funciones que cumple lo requieran, en los órganos superiores, demás órganos estatales, entidades presupuestadas y el sistema empresarial.”

Los **funcionarios** son trabajadores que pertenecen a la categoría ocupacional de técnicos, por la naturaleza de sus funciones, una parte de ellos, se encuentran próximos a las categorías de cuadros y constituyen la principal cantera para la selección de la reserva de los cuadros ejecutivos. Los funcionarios son designados, al igual que los

trabajadores de otras categorías ocupacionales, titulares de cargos que debido a sus características se cubren por designación.

La **categorización** de los cuadros constituye un elemento esencial a los fines organizativos o de otra índole para el trabajo con los cuadros (jerarquización, escalas salariales, estructuras y plantillas, facultades y estimulación, entre otras). Las categorías de los cuadros son las siguientes: Directivos superiores y Directivos y Ejecutivos.

La **nomenclatura** de cargos de cuadros es una de las vías con que cuentan el Partido Comunista de Cuba y el Gobierno para el control de los cargos fundamentales en cada nivel de decisión y se define como: “el conjunto de cargos de cuadros, cuyo nombramiento corresponde a un mismo nivel de decisión por el que responde su jefe. La relación de cargos se determina atendiendo a la categorización de los cuadros, sus jerarquías y funciones que cumplen dentro del sistema de la entidad y por regla general, en ella se combinan cargos de diferentes niveles de dirección”.

Los jefes al conformar las nomenclaturas, lo harán con un sentido racional y práctico, incluyendo cada cargo en una sola de ellas, mediante la realización de un estudio detallado de los niveles de dirección, la estructura, y la atención que requieren los cargos de cuadros de su entidad. Las nomenclaturas y la delegación de facultades se ponen en vigor una vez aprobadas en la Comisión de Cuadros correspondiente, mediante documentos legales firmados por el jefe máximo de la entidad.

1.2.1 La reserva de los cargos de cuadros.

La reserva de los cargos de cuadros es el conjunto de cuadros y de otros trabajadores de otras categorías ocupacionales que se seleccionan y aprueban siguiendo los mismos criterios y requisitos que, de forma general, se aplican a los titulares de los cargos de cuadros y que se les prepara para asumir cargos de un nivel superior o de mayor complejidad. Entre sus objetivos están:

1. Cubrir las vacantes de los cargos de cuadros conforme al plan previsto de manera que se puedan resolver, de manera eficaz y sin improvisaciones, las necesidades de las diferentes categorías.

2. Asegurar la preparación y superación de la reserva de los cargos de cuadros de manera que les permita asumir responsabilidades superiores, mediante un programa de actividades.
3. Garantizar el cumplimiento de la Política de Cuadros trazada por el Partido Comunista de Cuba en relación con la correcta combinación de los cuadros nuevos con los más experimentados, la promoción de compañeras a los cargos de dirección, el necesario equilibrio entre los de más edad con los de menos edad y una adecuada composición atendiendo al color de la piel en correspondencia con las características de nuestra población.

Como principios los jefes responden por la selección y aprobación de la reserva de los cargos que conforman su nomenclatura, previo análisis con las respectivas Comisiones de Cuadros. Los cuadros que se designan para desempeñar los cargos de segundos de los jefes, deben cumplir los requisitos para constituir la primera reserva de ellos. Por otra parte, constituyen una cantera importante para conformar la cantera de cuadros, los trabajadores conceptuados como funcionarios y otras categorías ocupacionales, cuyas características exigen de quienes las desempeñan poseer, entre los requisitos de idoneidad demostrada, el de confiabilidad reconocida, previo a su designación. Deben ser compañeros preparados, con iniciativas y perspectivas de desarrollo. Se debe determinar más de una reserva por cada cargo de cuadro, de manera que se garantice su utilización en el momento que se requiera. La cifra debe estar acorde a la realidad de cada entidad, ajeno a todo enfoque burocrático.

El proceso de selección y actualización de la reserva debe realizarse teniendo en cuenta un conjunto de pasos bien establecidos que garantizarán la efectividad del proceso; a continuación se explicará dicho procedimiento.

1.3 Procedimiento de selección y actualización de la Reserva.

El Departamento de Cuadros en correspondencia con un cronograma que forma parte del plan temático, orienta a los cuadros facultados a identificar el personal con potencialidades para desempeñarse como cuadro en los cargos que se encuentran en el

área de competencia de estos, por relaciones de subordinación lineales o por delegación del Rector.

Los cuadros facultados al conformar la propuesta de reserva de cuadros de cada cargo, tienen que determinar al menos dos reservas por cargo, priorizando las fuentes internas siempre que resulte viable. Un trabajador puede ser considerado como reserva de más de un cargo, siempre que se articule con una secuencia lógica de progreso en la estructura jerárquica de la unidad organizativa o Universidad.

Una vez definida o actualizada la reserva de cuadros, el cuadro facultado entrega en el formato establecido a la dirección del Departamento de Cuadros, la propuesta de reserva de cuadros de los cargos que le corresponde según su área de competencia.

La dirección del Departamento de Cuadros, de conjunto con el personal técnico y especialistas que lo integran, revisa la propuesta de las áreas, sobre la base entre otros criterios de los siguientes:

- Completamiento y actualización de la información según lo establecido en el formato establecido.
- Cumplimiento de los requisitos por alguno de los propuestos.

La Comisión de Cuadros analiza las propuestas en sesión de trabajo y de ser aprobada la lista de reservas propuesta en la Comisión de Cuadros, el cuadro facultado procede a informarla oficialmente al integrante de la reserva y posteriormente al núcleo del Partido del área. De no resultar aprobada la propuesta de reservas contenida en la lista o alguna de ellas, por la Comisión de Cuadros, el cuadro facultado debe proceder a reconsiderar su propuesta y volver a presentar en otra sesión, la cual es planificada por la dirección del Departamento de Cuadros.

Una vez concluido el proceso de aprobación de la lista de reserva por todos los implicados referidos en las actividades anteriores se procede por el cuadro facultado y los cuadros en funciones de la unidad organizativa a elaborar el plan de preparación de los integrantes de la lista de reserva de su cargo.

Otro proceso fundamental que deben realizar los jefes con sus subordinados como parte esencial del Sistema de Trabajo con los Cuadros es la evaluación, pues esta actividad

sistemática les permite valorar el comportamiento y resultados obtenidos por el cuadro en el cumplimiento de sus misiones y tareas, así como el desarrollo integral alcanzado por él; dichos resultados permiten valorar sus aciertos, deficiencias y sus causas, determinando, a su vez, cómo han sido aplicados los principios de la política de Cuadros y las medidas para el trabajo futuro

1.4 Proceso de evaluación de los cuadros.

La evaluación es un proceso de interacción cuadro-dirección de la entidad mediante el cual:

- a) El cuadro conoce como valoran los resultados de su trabajo, sus cualidades profesionales y personales, su correspondencia con las responsabilidades que desempeña, potencialidades de desarrollo, y qué se prevé para su futuro inmediato.
- b) La dirección de la entidad conoce como piensa el evaluado sobre sí mismo y su entorno, si se siente reconocido su trabajo, si son atendidas sus necesidades, su nivel de compromiso en erradicar las dificultades y señalamientos indicados, y al mismo tiempo tiene la oportunidad de transmitirles confianza y de exhortarlo a mejores desempeños.

La evaluación tiene como objetivos:

1. Analizar de forma crítica y objetiva los valores, cualidades y capacidades del cuadro para contribuir a estimularlas y desarrollarlas, así como, sus debilidades con el fin de rectificarlas, en interés de lograr mejores resultados y contribuir a su desarrollo.
2. Contribuir a la mejor selección, preparación y superación, ubicación, atención y estimulación de los cuadros y con ello asegurar superiores resultados.
3. Determinar si cada cuadro ocupa el cargo para el cual reúne las condiciones requeridas. Conocer cuales rebasan la exigencias del cargo que desempeñan y si están aptos para ser promovidos o si han dejado de responder a los requisitos del cargo que ocupan y deben ser demovidos o trasladados.

4. Arribar a criterios sobre el resultado cualitativo de la evaluación de los cuadros y conocer en qué medida estos responden a las necesidades actuales y perspectivas de cada unidad organizativa en particular y del organismo en su conjunto.
5. Aportar elementos para la aplicación correcta de las sanciones disciplinarias en los casos en que sean necesarios.

Para su mejor comprensión y utilización a los fines de la evaluación, el contenido general de la evaluación está comprendido en dos grupos de indicadores, y cada uno de ellos se subdivide, a su vez, en diferentes aspectos.

1.4.1 Indicadores del proceso de evaluación.

Los dos grupos con sus indicadores son:

(Para la evaluación integral)

I. Indicadores técnico-laborales

1.1 Ética y disciplina.

1.2 Actitud ante el trabajo.

1.3 Desempeño del cargo, resultados del trabajo y de la actividad de preparación.

1.4 Capacidad de dirección.

1.5 Trabajo con los subordinados.

II. Indicadores de personalidad

2.1 Características individuales

2.2 Cualidades intelectuales

Para la **evaluación parcial** se mantienen los mismos indicadores, pero agrupados de modo diferente de la siguiente forma:

I. Indicadores técnico-laborales

1.1 Ética, disciplina y actitud ante el trabajo.

1.2 Desempeño del cargo, resultados del trabajo y de la actividad de preparación.

1.3 Capacidad de dirección y trabajo con los subordinados.

II. Indicadores de personalidad

2.1 Características individuales y cualidades intelectuales

En los respectivos modelos de evaluación integral y parcial cada uno de los indicadores está dividido por aspectos (Anexo 2). Para calificar un aspecto se utiliza una escala de valores (puntuación del 0 al 10) en base a los criterios: Excelente (10); Muy Bien (9 y 8); Bien (7, 6 y 5); Regular (4 y 3) y Mal (0, 1, y 2), tal como se muestra en la escala siguiente:

Escala de valores (puntos)	Mal		Regular		Bien			Muy bien		Excelente	No evaluado
	M		R		B			MB		E	NE
	0	1	2	3	4	5	6	7	8	9	

Para calificar el indicador, se promedian las puntuaciones otorgadas a cada uno de los aspectos que le corresponden, si en la cifra resultante aparece un decimal de 0,5 o mayor, esta se redondea al número entero inmediato superior o al número entero inmediato inferior, si el decimal es menor de 0,5.

El desarrollo del cuadro evaluado ofrece como conclusión si: avanza, está estable, estancado o en retroceso. Para ello se analiza el desarrollo del cuadro en los últimos tres años incluido el que se evalúa. Se considera que un cuadro **avanza** si comparado

con el período anterior, mantiene un progreso constante de sus cualidades y sus resultados del trabajo mejoran.

Se considera que un cuadro está **estable** si ha consolidado los resultados obtenidos en el período anterior y aunque no se aprecia una mejoría constante en los resultados del trabajo, se valora que tiene tendencia a avanzar.

Se considera que un cuadro está **estancado** si comparado con el período anterior, no mantiene una tendencia al desarrollo, no ha seguido mejorando sus cualidades y los resultados de su trabajo no superan el nivel anterior. Ha llegado al límite de sus posibilidades de desarrollo.

Se considera que un cuadro **retrocede**, si en el período anterior estaba estancado, mantiene una tendencia a perder cualidades y los resultados de su trabajo evidencian un retroceso.

Este momento de la evaluación muestra los resultados del trabajo del cuadro, que pueden ser buenos o negativos, en cualquier caso, si tiene señalamientos relacionados con la idoneidad, violaciones de los principios éticos o de disciplina se debe valorar si es necesario producir algún movimiento.

Como norma, el proceso de evaluación se lleva a cabo de forma parcial anualmente y cada dos años de manera integral siguiendo una serie de actividades que permiten el desarrollo correcto de dicho proceso. Los cuadros que tienen la responsabilidad de realizar la actividad de evaluación, después de estudiar detalladamente las particularidades del proceso y realizar los esclarecimientos necesarios, proceden a recopilar información sobre el desempeño de los cuadros a evaluar, para lo cual entre otros aspectos tienen que:

- Revisar la evaluación anterior con énfasis en los señalamientos y recomendaciones.
- Estudiar el cumplimiento de los objetivos de las áreas y/o procesos que dirigen.
- Revisar el registro de seguimiento al desempeño del cuadro a evaluar.
- El cumplimiento del plan anual del área que dirige a partir de los controles al plan mensual.

- Resultado de supervisiones, auditorías u otros controles internos o externos realizados al área y/o procesos que dirige.
- Resultado de encuestas, entrevistas, observaciones u otras técnicas aplicadas.

1.4.2 Conclusiones del proceso de evaluación.

Concluido el proceso, el Departamento de Cuadros elabora el Resumen del Proceso de Evaluación para su análisis en la Comisión de Cuadros, el mismo debe contener:

- 1. Principales resultados y deficiencias presentadas durante la realización del proceso de evaluación.**
- 2. Calidad del proceso de evaluación.**
- 3. Resumen estadístico del proceso de evaluación.**
 - Total de cuadros con que cuenta la entidad, cuántos deben ser evaluados y cuántos lo fueron.
 - Total que por sus resultados: avanzan, estable, estancados o en retroceso y la cantidad de cuadros con problemas.
 - Cuadros que requieren superarse según las distintas modalidades.
 - Cantidad de cuadros que deben ser liberados.
 - Cantidad de candidatos para integrar la reserva de cuadros. Cuántos de los propuestos fueron ratificados y cuántos están listos para promover.
 - Cantidad de compañeros propuestos a mantener en los cargos o efectuar movimientos con ellos.
 - Cantidad de cuadros que están en desacuerdo con su evaluación, cuántas reclamaciones fueron realizadas y sus resultados.

El Departamento de Cuadros a partir del informe resumen anterior procede a identificar con detalles las insuficiencias de cada uno de los cuadros, para que sirva de fuente de información a la elaboración del Plan de Preparación y Superación de Cuadros; así como clasifica los cuadros según sus resultados a fin de facilitar el análisis en el proceso de estímulo y reconocimiento.

Del análisis de los resultados de las evaluaciones a cualquier nivel, se originan decisiones como la inclusión o exclusión en la reserva de los cuadros, en los listados de

candidatos a cursos, la elaboración del Plan de Desarrollo Perspectivo y la Proyección de los Cuadros, y otros, que lo convierten en un efectivo instrumento para la aplicación de la Política de Cuadros. Ello obliga a ser profundos y objetivos en las valoraciones que se realicen, pues un error puede determinar desde la ubicación o promoción de un cuadro hasta su cese en la entidad. De los resultados individuales del proceso de evaluación depende ante todo, la selección, movimientos, preparación y superación de los cuadros. A continuación se abordará el proceso relativo a los movimientos de cuadros.

1.5 Proceso de selección y tramitación de los movimientos de cuadros.

Se consideran movimientos de cuadros: la promoción, el traslado, la reubicación, la liberación, la democión, la separación y además la preparación y superación.

La aplicación de cualquiera de estos procedimientos implica una elevada responsabilidad que deben asumir los jefes facultados, los integrantes de las Comisiones y el Departamento de Cuadros, en la aplicación de la política en general y fundamentalmente en la selección y promoción a los cargos de dirección, en cada caso, a los compañeros que más condiciones tengan para ello y que mejor preparados estén para desempeñar esta actividad con eficiencia. Todas las propuestas de movimientos son analizadas por la Comisión de Cuadros constituida en la Universidad, previa indicación del jefe que la preside, para lo cual se auxilian del Departamento de Cuadros que prepara la documentación correspondiente. El análisis debe realizarse en un clima y ambiente propicio que estimule la participación activa de cada integrante y a que se expresen las opiniones y sugerencias de forma franca y espontánea.

El procedimiento en general cuenta con tres etapas:

Etapas de selección: Es responsabilidad de los jefes facultados y consiste en el análisis de los posibles candidatos, teniendo en cuenta el tipo de movimiento a realizar, los requisitos, funciones y atribuciones del cargo; las cualidades políticas-ideológicas y éticas exigibles de los cuadros; si está incluido en la reserva; el cumplimiento de sus objetivos, tareas y el plan de desarrollo individual; la valoración de su trayectoria y los

resultados de las evaluaciones, así como, otras informaciones adicionales que se obtengan del proceso de consultas.

Entre los indicadores que miden el desarrollo de este proceso de selección en la Universidad se encuentran:

- Por ciento de movimientos de cuadros en la sede central que tienen como fuente la reserva.
- Por ciento de movimientos de cuadros en la FUM que tienen como fuente la reserva.
- Cantidad de cargos con 2 reservas en la sede central.
- Cantidad de cargos con 2 reservas en la FUM.
- Por ciento de cuadros con estabilidad en el cargo por al menos 3 años.

Etapas de consulta: Se realiza previo a la reunión de la Comisión de Cuadros y tiene como objetivo posibilitar que el jefe facultado (Presidente de la Comisión) disponga de un conjunto de elementos obtenidos por diferentes vías que le permita analizar con la mayor profundidad posible los candidatos seleccionados, y poder adoptar la decisión de presentar la propuesta a la Comisión de cuadros, o desestimarla y buscar nuevos candidatos.

Etapas de aprobación: Cumplidas las etapas anteriores y seleccionados los candidatos, el Órgano de Cuadros, prepara los fundamentos de la propuesta apoyándose en el análisis de:

- a) El expediente de cuadro y laboral.
- b) La última evaluación integral realizada u otras que se consideren necesario;
- c) La planilla “Datos del Cuadro” actualizada; y
- d) Criterios que sobre el propuesto surjan del proceso de consulta.

Analizada la propuesta por la Comisión de cuadros y aprobada por el jefe facultado, el Órgano de Cuadros emite el documento (en forma de acuerdo o similar) contentivo de la decisión, archiva el original y envía una copia al área jurídica para que se dicte la

Resolución correspondiente, si fuese el caso, que le da forma legal a dicha decisión. Una copia certificada de la resolución se le entrega al Órgano de cuadros, otra se envía al área de Recursos Humanos y se archiva en el expediente laboral, y se le entrega una al interesado.

Concluido el proceso de selección y tramitación de los movimientos de cuadros y adoptada la aprobación final, se le debe informar de ello al propuesto, por el jefe facultado o con la participación de algunos miembros de la comisión de Cuadros, con el objetivo de escucharlo y obtener una valoración más directa de su proyección.

Los movimientos realizados implican el análisis de:

- Por ciento de incremento de cuadros negros y mestizos en funciones.
- Por ciento de incremento de negros y mestizos en la reserva.
- Por ciento de incremento de cargos ocupados por mujeres.
- Por ciento de incremento de cargos decisorios ocupados por mujeres.
- Por ciento de cuadros docentes que son profesores Titulares o Auxiliares.
- Por ciento de incremento de cuadros con categoría docente que son doctores.

Cuando ocurre un movimiento es preciso llevar a cabo un **proceso de entrega y recepción de cargos** donde:

Entrega: Es la actividad de traspaso de la autoridad de un cuadro saliente de un cargo, al cuadro que ha sido aprobado para ocuparlo, en la cual después de una supervisión de los procesos, recursos materiales, financieros y humanos; así como de la gestión, se presentan los resultados de forma documentada.

Recepción: actividad en que un cuadro que ha sido aprobado para ocupar un cargo, es investido de la autoridad sobre los recursos materiales, financieros y laborales, a partir del conocimiento de los resultados de la gestión de los mismos por el cuadro saliente.

1.6 Procedimiento de planificación de demanda de cuadros.

Este procedimiento tiene por objeto definir la forma de proceder para determinar las necesidades de cuadros en la Universidad.

Demanda de cuadros: Es la necesidad de cuadros de la Universidad en correspondencia con los puestos de trabajo de dirección que por sus características e importancia clasifican como tales.

Planificación de la demanda de cuadros: Es el proceso de identificación de los cargos de dirección que se categorizan como cuadros y están aprobados en la plantilla por el MES y se sustentan en una unidad organizativa o funciones vigentes en la Universidad.

La dirección del Departamento de Cuadros de conjunto con el Director de Recursos Humanos analizan en el mes de mayo de cada año la estructura de dirección de la Universidad e identifican los puestos de trabajo que constituyen cargos de dirección y cuáles de ellos clasifican como cuadro. Este análisis se realiza por personal de ambas áreas cada vez que se efectuó un cambio en la estructura organizativa del centro.

La dirección del Departamento de Cuadros de conjunto con los especialistas que atienden las áreas, procede a revisar los resultados de la actividad anterior, comparando con el inventario de cuadros del centro y lo establecido en los documentos normativos, de existir diferencias se procede a realizar por el Departamento de Cuadros un proceso de desnombramiento, si el cargo no está vigente o a nombrar a los que ocupan posiciones de dirección y no han sido nombrados de encontrarse vigente la unidad organizativa que dirigen o las funciones que cumplen.

La Dirección de Recursos Humanos, de no coincidir el inventario de cuadros con lo oficialmente aprobado procede a actualizar la plantilla eliminando las unidades organizativas o cargos no vigentes para su presentación al ministerio.

Una vez aprobada la plantilla se procede a realizar un nuevo cotejo entre el Departamento de Cuadros y la Dirección de Recursos Humanos.

La dirección del Departamento de Cuadros del centro concilia con la Dirección de Cuadros del MES la plantilla aprobada. Los resultados del análisis se presentan al presidente de la Comisión de Cuadros, quien después de revisar la plantilla de cuadros o inventario, procede a presentarla en la Comisión de Cuadros del Centro.

1.7 Fuentes de datos en el Departamento de Cuadros.

Los DWH posibilitan organizar y almacenar los datos que se necesitan para un procesamiento analítico abarcando una amplia perspectiva de tiempo. Su función esencial es ser la base de un sistema de información y brindar una visión integrada de dicha información, especialmente enfocada hacia la toma de decisiones por parte del personal de la organización.

El sistema fuente del Departamento de Cuadros es un sistema transaccional por ello sería interesante abordar algunos conceptos que se hace tratan en (Adamson, 2006) acerca de los sistemas operacionales y los almacenes de datos o DWH.

Los sistemas operacionales soportan directamente la ejecución de los procesos de negocio mientras que el DWH da soporte a la evaluación de dicho proceso. La figura siguiente muestra algunas de las principales diferencias.

	Operational System	Data Warehouse
Also Known as	Transaction System On Line Transaction Processing (OLTP) System Source system	Analytic system Data mart
Purpose	Execution of a business process	Measurement of a business process
Primary Interaction Style	Insert, Update, Query, Delete	Query
Scope of Interaction	Individual transaction	Aggregated transactions
Query Patterns	Predictable and stable	Unpredictable and changing
Temporal Focus	Current	Current and historic
Design Principle	Third normal form (3NF)	Dimensional design (star schema)

Figura 1-1 Diferencias entre los sistemas operacionales y los DWH

El análisis de estadísticas y tendencias es fundamental en los datos relacionados con los cuadros pues constituye una tarea estratégica debido a que hay que garantizar el

completamiento de los cuadros y la Política de Cuadros para un funcionamiento adecuado del sistema de dirección.

La información que se maneja en el Departamento de Cuadros de la UCLV se almacena en un sistema de base de datos transaccional. En este sistema encontramos:

Tabla persona: Describe los datos personales y otros de interés relativos a los cuadros y las reservas. Tiene 23 atributos y el carné de identidad es la llave primaria.

Tabla cargo: Recoge la información relativa a los cargos como el nombre, el lugar etc. Cada cargo se identifica por un id_cargo.

Tabla indicador: Almacena los indicadores actuales y su descripción, tanto de la evaluación parcial como integral.

Tabla causa_mov: Describe las causas bajo las cuales tiene lugar un movimiento.

Relación mov_cuadro: Define un movimiento de cuadro como la relación actual de una persona con un cargo anterior y un cargo actual por determinada causa. La llave primaria la constituyen las dos llaves foráneas a la tabla cargo, la llave foránea a la tabla causa_mov y la llave foránea a la tabla persona.

Relación reserva: Describe la relación entre un cargo y sus reservas de modo que una persona puede ser reserva de más de un cargo. La llave primaria está formada por las llaves foráneas a la tabla cargo y persona respectivamente.

Relación eval_cuadros: Asocia a cada persona con uno o más indicadores y la evaluación que le corresponde. La llave primaria son las llaves foráneas ci de la tabla persona y nombre_ind de la tabla indicador.

En la base de datos los nombres de las llaves primarias se mantienen invariables cuando se referencian desde otra tabla como llaves foráneas.

Los datos almacenados pertenecen no sólo a la sede central sino además a cada uno de los municipios de la provincia de Villa Clara donde existe un centro o filial universitaria.

El sistema almacena los datos de una forma bastante consistente y maneja varias condiciones para mantener la integridad de los datos y evitar errores en el momento en

que estos son introducidos. Algunos campos se introducen como espacios en blanco si se desconoce su valor. Los reportes que el sistema ofrece son muy sencillos y no responden a las necesidades de realizar análisis a través del tiempo que tiene el Departamento de Cuadros.

1.8 Conclusiones del capítulo.

En el análisis del Sistema de Trabajo con los Cuadros se evidenciaron procesos importantes que tienen lugar en el Departamento de Cuadros.

- Selección y actualización de la reserva:

La selección de la reserva es orientada a los cuadros facultados por la dirección del Departamento de Cuadros anualmente. Las propuestas para integrar la reserva deben ser aprobadas por la Comisión de Cuadros. Los cuadros facultados al conformar la propuesta de reserva de cuadros de cada cargo, tienen que determinar al menos dos reservas por cargo y un trabajador puede ser considerado como reserva de más de un cargo. Los cuadros pueden constituir reservas y viceversa. Mensualmente cuando tienen lugar los movimientos de cuadros pueden existir cambios en las reservas.

- Evaluación de cuadros:

El proceso de evaluación de los cuadros se realiza anualmente y tiene como base un conjunto de indicadores a cuyos aspectos específicos se les asigna una evaluación con un valor entre cero y diez, además del aspecto relacionado con el desarrollo del cuadro ya sea: avanza, estable, estancado o en retroceso. De este proceso solo se almacenan los resultados finales de cada indicador y el estado del desarrollo del cuadro. Se realiza una evaluación parcial anualmente e integral cada dos años. Los indicadores han presentado estabilidad en los últimos tres años pues se aplican a todos los niveles de cuadros en todo el país, sin embargo pueden estar sujetos a cambios si al establecerse una nueva política de cuadros así se designa.

- Movimientos de cuadros:

Otro proceso es el relacionado con los movimientos de cuadros que son analizados mensualmente por la comisión de cuadros teniendo en cuenta las propuestas o necesidades de cada área. Ante cada movimiento se expresan las causas y se redacta una resolución que le da validez.

La Universidad se divide en una estructura que de modo general contiene al rectorado, las direcciones, los departamentos independientes, los centros de estudios no adscritos a facultades, las facultades, el Colegio Universitario de Formación Básica y los centros y filiales universitaria municipales. Los cargos que existen en estas dependencias están sujetos a cambios ya sea por la variación de la estructura externa (facultad, departamento) o interna (eliminación o adición de nuevos cargos). Ante estos cambios se procede además a realizar los movimientos de cuadros correspondientes.

Además de estos procesos el Departamento de Cuadros debe generar según se le soliciten informes con resultados estadísticos sobre cualquier aspecto relativo al trabajo que ellos desempeñan con los cuadros y sus reservas por tanto, teniendo en cuenta las ventajas que ofrece un Mercado de Datos para el trabajo con datos históricos y el apoyo a la toma de decisiones; lo convierten en una alternativa para el tratamiento de la información en el Departamento de Cuadros de la Universidad. El capítulo siguiente abordará aspectos teóricos de interés que ayudarán al diseño e implementación del Mercado de Datos que responde a las necesidades del caso de estudio ya descrito.

CAPÍTULO 2. ACERCA DE LOS MERCADOS DE DATOS.

El presente capítulo tiene como objetivos explicar aspectos relevantes del diseño de un MD. Aborda aspectos importantes a tener en cuenta durante la fase de modelación y el uso de las técnicas OLAP para el análisis de los datos.

2.1 La modelación dimensional.

Un DWH no se puede comprar se tiene que construir, siendo la modelación un factor clave para ello. Posterior a la fase de definición de objetivos y necesidades de los usuarios se diseñará un modelo capaz de soportar estos requerimientos (Ralph Kimball, 2002a).

El modelo Entidad-Relación (ER) es una técnica poderosa para el diseño de sistemas transaccionales en el entorno de las bases de datos relacionales. Permite la normalización de la estructura de datos física, obteniéndose un diseño sin redundancias en los datos y ocupándose un menor espacio de almacenamiento. Sin embargo, una técnica mucho más poderosa para la consulta de los datos, es el modelo dimensional o multidimensional (Kimball, 1995). Esta técnica de diseño de bases de datos responde a las necesidades analíticas que impone la tecnología DWH, se logra un diseño más cercano a la visión del usuario final pues la información es más fácil de encuestar y navegar y se obtiene mayor velocidad de recuperación de las consultas en comparación con el modelo E/R.

Albert Einstein capturó la principal razón para usar el modelo dimensional cuando dijo: “Make everything as simple as possible, but not simpler” (Joy Mundy, 2006). El modelo dimensional contiene menos tablas que los típicos sistemas fuentes y la información está agrupada en categorías del negocio que son más intuitivas para los usuarios. Desde la perspectiva del modelo relacional el modelo multidimensional consiste en una tabla de hechos normalizada con tablas de dimensiones desnormalizadas. Crear tablas de hechos

y tablas de dimensiones y establecer las relaciones entre cada tabla de dimensión y la tabla de hechos dará como resultado para el modelo un esquema estrella (Ponniah, 2001b).

Generalmente la información del DWH se caracteriza por ser invariable, sin embargo lograr un buen diseño implica entre otras cuestiones manejar los posibles cambios que tienen lugar en los sistemas fuentes y que pueden ser de interés como información histórica.

2.2 Dimensiones lentamente cambiantes.

Las reglas para las dimensiones lentamente cambiantes (SCD) son identificadas durante el proceso de diseño del esquema dimensional (Adamson, 2006).

Para cada atributo en las tablas de dimensiones es necesario especificar una estrategia para manejar los cambios que puedan presentarse en los sistemas fuentes.

2.2.1 Dimensiones lentamente cambiantes tipo 2.

El objetivo principal del DWH es el mantenimiento de la historia. Con el cambio de tipo 2 se logra particionar la historia almacenando tanto los valores antiguos como los actuales. Se adiciona una nueva fila a la dimensión que refleje el nuevo valor del atributo cambiado, manteniendo el resto de los atributos invariables y generando una nueva llave subrogada. Alternativamente podría ser útil adicionar una columna que indique el más reciente. Otra técnica a utilizar es mantener una columna de fecha efectiva referente al momento a partir del cual el valor de atributo es válido o inválido en el caso de una columna que tuviera la fecha de expiración. Principalmente estas columnas son necesarias para el proceso de extracción y carga de los datos donde se precisa saber qué llaves subrogadas son válidas para realizar una entrada en la tabla de hechos. Es importante recalcar que estas fechas no guardan relación con las fechas que se almacenan en la dimensión fecha del DWH (Ralph Kimball, 2002b).

Un recurso que podría utilizarse en determinadas condiciones para complementar los cambios lentos de tipo 2 serían las tablas de hechos sin hechos pues según (Becker, 2003) esta técnica es factible cuando se tiene una dimensión que puede tener millones de filas y varios atributos que pueden cambiar potencialmente rápido, se decide entonces separar estos atributos en mini dimensiones e incluir, en las tablas de hechos correspondientes, las llaves foráneas a estas mini dimensiones y a la dimensión principal, estas llaves foráneas representan el estado existente cuando ocurrió el hecho. En el caso en que se necesita brindar soporte para que sea posible, por ejemplo, ver el perfil de un cliente en cualquier punto del tiempo, la solución es usar una *tabla de hechos sin hechos*. Esta *tabla de hechos sin hechos* captura la relación entre la dimensión principal y las mini dimensiones a través del tiempo (Figura 2-1). Así, cada vez que ocurra un cambio lento de tipo 2 en la dimensión base o un cambio en la relación entre la dimensión base y las mini dimensiones, se carga una nueva fila en la *tabla de hechos sin hechos*. Además de las llaves foráneas a la dimensión base y a las mini dimensiones, esta *tabla de hechos sin hechos* contiene la fecha de efectividad y de expiración para poder localizar un perfil en cualquier punto del tiempo. También se puede añadir una dimensión simple como bandera para indicar el perfil actual, y adicionalmente, una dimensión que indique la causa del cambio (Thorntwaite, junio 1, 2006).



Figura 2-1 Tabla de hechos sin hechos para SCD tipo 2

2.2.2 Dimensiones lentamente cambiantes: tipos 0, 4, 5.

Cuando se analiza cambiar un atributo de una dimensión, inmediatamente se valoran los tres enfoques fundamentales: dimensiones lentamente cambiantes de tipo 1 (sobrescribir), 2 (añadir una fila) y 3 (añadir una columna). Estas tres técnicas son

adecuadas para la mayoría de las situaciones. Sin embargo a veces se necesitan variaciones construidas sobre estas técnicas básicas para brindar servicio a usuarios más avanzados del DWH (Ross, 2005). Por esta razón han surgido técnicas híbridas más avanzadas a las cuales se les han asignado también números al igual que a las tres técnicas básicas para facilitar la comunicación entre los miembros del DWH. Algunas de estas nuevas técnicas descritas en (Ross, 2013) son las siguientes:

Tipo 0: Retener el Original

Con el tipo 0, el valor del atributo de la dimensión nunca cambia, así los hechos son siempre agrupados por este valor original. El tipo 0 es apropiado para cualquier atributo etiquetado como “original” como el puntaje de crédito original de cliente o cualquier identificador duradero. El tipo 0 también se aplica a la mayoría de los atributos de la dimensión fecha.

Tipo 4: Añadir una Mini Dimensión

Esta técnica es usada cuando un grupo de atributos dimensionales son separados para una mini dimensión. Este enfoque es útil cuando los valores de los atributos de las dimensiones son relativamente volátiles. En tablas de dimensión con millones de filas, los atributos frecuentemente usados, son también candidatos a ser separados para mini dimensiones, incluso aunque estos no cambien frecuentemente. Una llave subrogada es asignada a cada perfil único o combinación de valores de los atributos en la mini dimensión. Las llaves subrogadas de la dimensión base y de las mini dimensiones son capturadas como llaves foráneas en la tabla de hechos.

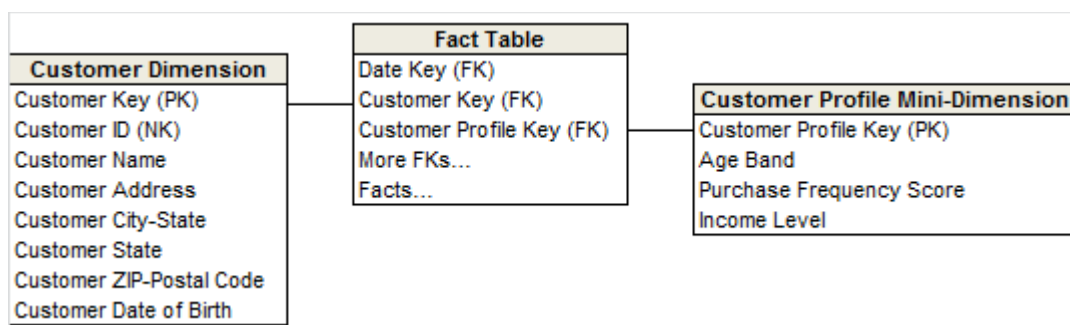


Figura 2-2 SCD tipo 4 usando mini dimensión.

Tipo 5: Añadir una Mini Dimensión y un Outtrigger de Tipo 1

Se basa en el tipo 4, embebiendo en la dimensión base, una llave “perfil actual” a la mini dimensión, esta llave es sobrescrita como un atributo de tipo 1. Este enfoque permite que los valores de los atributos de la mini dimensión asignados actualmente sean accedidos a través de la dimensión base en vez de vincularlos a través de la tabla de hechos. Lógicamente, la dimensión base y la mini dimensión outtrigger “perfil actual” son representadas como una sola tabla en la capa de presentación. Los atributos del outtrigger deben tener nombre distintos como “Nivel de renta actual” para diferenciarlos de los atributos de la mini dimensión enlazada a la tabla de hechos. El equipo de ETL debe sobrescribir la referencia de tipo 1 a la mini dimensión “perfil actual” cada vez que esta cambie a lo largo del tiempo.

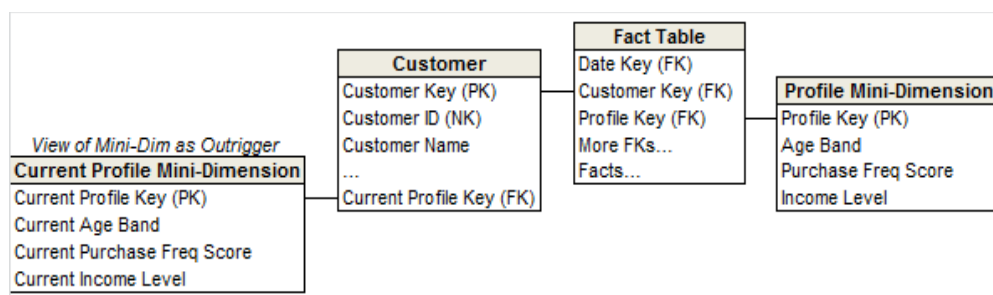


Figura 2-3 SCD tipo 5 usando mini dimensión y un Outtrigger de Tipo 1.

2.3 Refinamiento del modelo.

Cuando se analizan las fuentes de datos se identifican generalmente de forma preliminar cuales son las dimensiones, medidas etc., pero en fases posteriores del diseño pueden surgir variaciones del esquema estrella base, como son copos de nieve, tablas puentes, dimensiones degeneradas, dimensiones junk, entre otras.

2.3.1 Dimensiones junk.

Al desarrollar un modelo dimensional a menudo nos encontramos con indicadores y banderas que no tienen una pertenencia lógica a las tablas de dimensiones. Estos atributos independientes son usualmente demasiado valiosos para ignorarlos o

excluirlos. A veces se propone tratarlos como hechos, o en pequeñas tablas de dimensiones pero otra solución son las dimensiones junk.

Las dimensiones junk pueden ser usadas para almacenar todas las posibles combinaciones de valores discretos de un conjunto de atributos de manera que la tabla de hechos solo tendría una llave extranjera a esta tabla (Ralph Kimball, 2002b).

En (Ross, August 7, 2003) se exponen dos enfoques para crear una dimensión junk. Una primera opción puede ser crear la dimensión con una fila para cada posible combinación. Otro enfoque es crear nuevas filas durante el proceso de ETL, a medida que aparecen nuevas combinaciones se van agregando con su respectiva llave subrogada.

Si el número total de posibles filas en la tabla de dimensión junk es relativamente pequeño se recomienda usar el primer enfoque. Por otro lado, si el número de posibles filas es muy grande pudiera ser más ventajoso crear la dimensión junk a medida que se encuentran nuevas filas únicas.

El total de filas de la dimensión junk no debe exceder ni estar muy cerca del número de filas de la tabla de hechos que la referencia.

La dimensión junk incluye todas las combinaciones válidas de los atributos, por tanto no es necesario tener en cuenta las estrategias para los cambios lentos.

El proceso de evaluación en el Departamento de Cuadros ofrece ciertos requisitos que permiten valorar el uso de una dimensión junk para su modelación. Se usan un conjunto de indicadores para la evaluación cuyos valores oscilan entre cero y diez, por tanto, se podrían obtener todas las combinaciones posibles de evaluaciones. Como resultado parcial se obtendría el diseño mostrado en la siguiente figura.

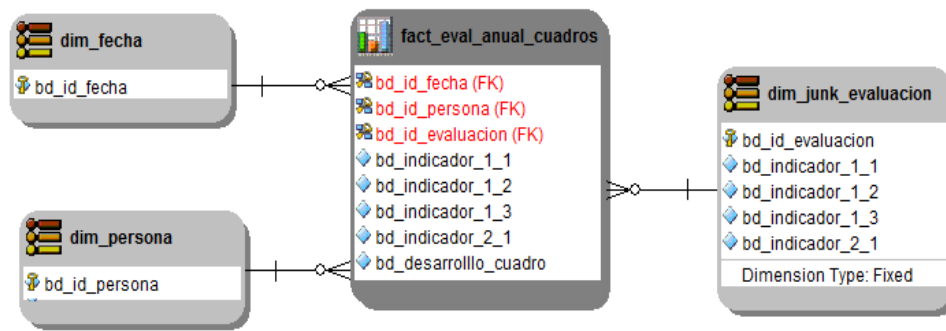


Figura 2-4 Ejemplo de dimensión junk para el proceso de evaluación de cuadros.

2.3.2 Las tablas puentes (bridge): su utilización en jerarquías y relaciones muchos a muchos.

El término drill-down (Claudia Imhoff, 2003) se refiere al acto de navegar a través de una jerarquía desde el nivel más alto de agregación hasta el nivel más bajo de detalle (o viceversa drilling up o roll up). Las jerarquías son parte integral de cualquier proceso de negocio y un aspecto fundamental del DWH.

Las jerarquías representadas en ocasiones como árboles tienen un tipo especial de relación padre-hijo. En este tipo de relación un padre puede tener uno o más hijos y un hijo debe pertenecer a un único padre. Estas jerarquías son simples y generalmente deseables pero también existe el caso en que un hijo puede pertenecer a múltiples padres por ejemplo cuando una compañía es parcialmente dirigida por otras dos compañías. La profundidad de una jerarquía se refiere al número de generaciones o niveles desde la raíz hasta las hojas existiendo entonces dos categorías: jerarquías de profundidad conocida y de profundidad desconocida o variable.

Las jerarquías identifican potenciales puntos de agregación y pueden ayudar a estimar el grado de sumariación. Es posible encontrar varias jerarquías en una dimensión (Adamson, 2006). Cuando la profundidad es conocida una opción que facilita el cálculo de agregados para recorrer la estructura jerárquica es representando en una misma dimensión la jerarquía deseada como se muestra en la figura siguiente (Ponniah, 2001a):

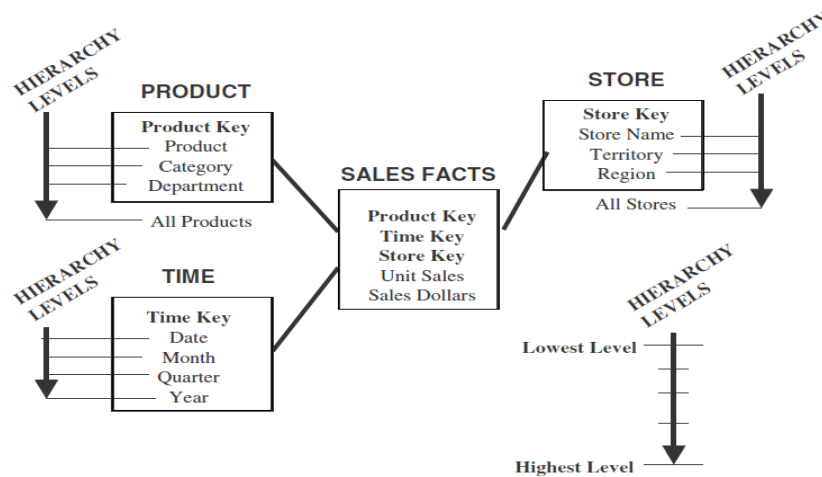


Figura 2-5 Representación de jerarquías de profundidad conocida.

Las tablas puentes o bridges pueden ser usadas para representar jerarquías de profundidad variable. En el artículo (Thorntwaite, diciembre 7,2004) se describe el caso de las jerarquías alternativas partiendo del hecho de que diferentes usuarios desean ver los datos agrupados de maneras distintas, en el caso más simple, un departamento quiere ver los clientes agrupados en una jerarquía y otro departamento quiere ver los clientes agrupados en una jerarquía alternativa. Una solución simple sería incluir ambas alternativas en la tabla de dimensión cliente y etiquetarlas apropiadamente. Desafortunadamente, no son muchas las jerarquías alternativas que se pueden incluir en una dimensión antes que esta se vuelva inusable.

La necesidad de acomodar las jerarquías alternativas de forma más flexible surge cuando varios departamentos quieren ver las formas a su propia manera. En este caso, generalmente se trabaja con el usuario para definir la forma más común en que los datos serán agrupados. Esta sería la jerarquía estándar o por defecto en la dimensión base. Cualquier otra jerarquía comúnmente usada es también incluida en la dimensión para mantener simplicidad para los usuarios.

La solución es proveer una tabla de jerarquía alternativa que permita a los usuarios agrupar los datos basados en la selección de una de las jerarquías disponibles. Cada jerarquía en la tabla de jerarquías alternativas tiene que incluir la jerarquía completa, desde su punto de partida donde se une a su dimensión asociada, hasta la cima. La

(Figura 2-6) muestra un ejemplo de tabla puente de jerarquía alternativa llamada *CustomerRegionHierarchy* para agrupamiento por regiones geográficas.

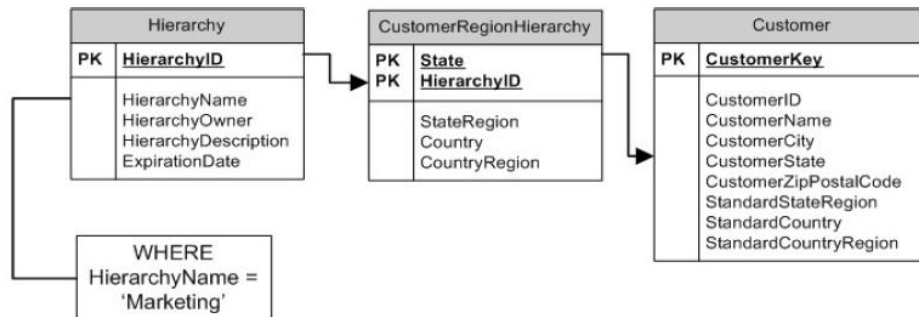


Figura 2-6 Jerarquía alternativa usando tabla puente

Para simplificar la confección de reportes y el análisis, la tabla puente incluye la definición de la jerarquía estándar. Esta elección se convierte entonces en la elección por defecto en todos los reportes estructurados, permitiendo a los usuarios intercambiar entre la jerarquía estándar y la alternativa.

Construir reportes y consultas a través de las relaciones muchos a muchos constituye por otra parte un trabajo difícil siendo útiles en este caso las tablas puentes entre las tablas de hechos y dimensiones para expresar esta relación. Un ejemplo comúnmente usado para mostrar las relaciones muchos a muchos son los múltiples diagnósticos que puede tener un paciente simultáneamente bajo los mismos síntomas. En (Ralph Kimball, 2002a) se propone un diseño para resolver este problema a través de una tabla puente entre la dimensión que nombran *Diagnosis* y la tabla de hechos (Figura 7). El diagnóstico original se generaliza en la tabla de hechos con una llave al *Diagnosis Group* (grupo de diagnósticos). La tabla de *Diagnosis Group*, es la tabla puente, y contiene un conjunto específico de entradas para cada paciente. Cada fila contiene la llave del grupo (*diagnosis_group_key*), una llave al diagnóstico individual (*diagnosis_key*), y un factor de peso (*weighting_factor*). Un paciente con tres diagnósticos tendría tres filas en su respectivo grupo (*Diagnosis Group*).

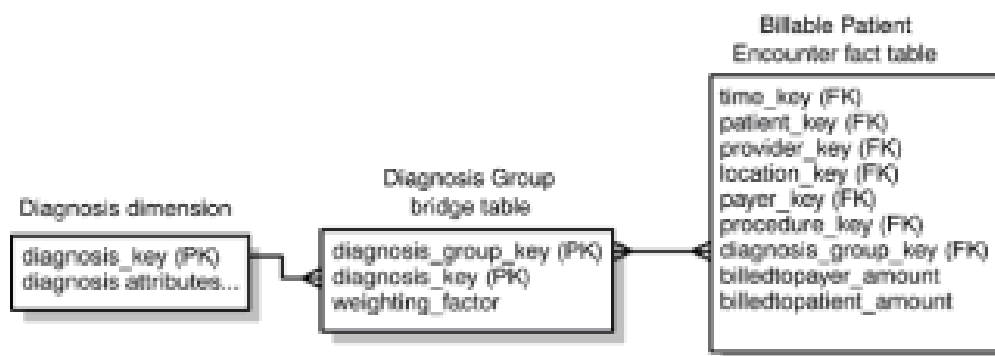


Figura 2-7 Utilización de una tabla puente para la relación muchos-a-muchos.

La llave de diagnóstico individual se une a la tabla *Diagnosis dimensión* que contiene la descripción total del diagnóstico.

En un período de tiempo un paciente podría tener diferentes grupos de diagnósticos. Una forma de embellecer el diseño sería añadiendo además de la llave paciente una fecha efectiva de comienzo y una fecha efectiva de final del grupo de diagnóstico.

Si se valora nuevamente el proceso de evaluación y se enfoca desde otra perspectiva se tiene que a cada cuadro se le asignan varios indicadores para su evaluación. El grupo de indicadores sería como el grupo de diagnósticos del ejemplo anterior, pudiéndose almacenar información adicional sobre cada indicador. La figura siguiente muestra un posible diseño.

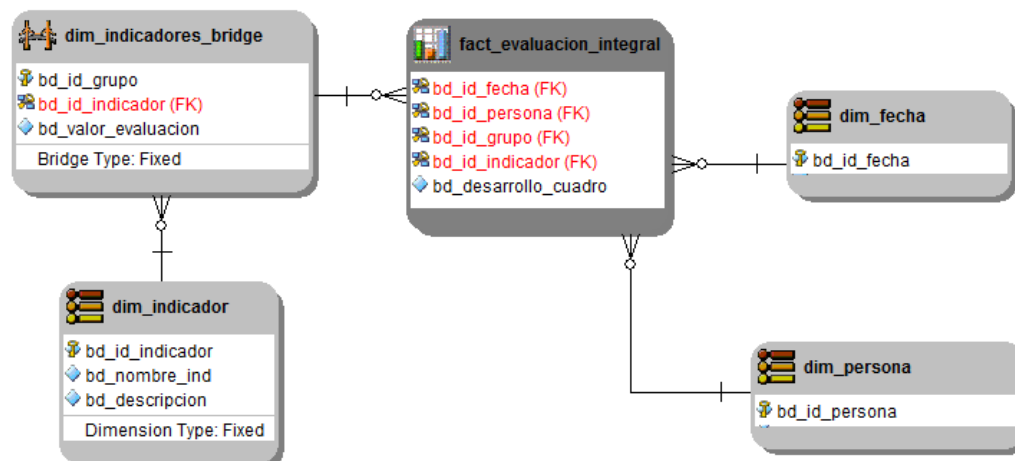


Figura 2-8 Ejemplo de tabla puente para el proceso de evaluación de cuadros.

2.3.3 Esquemas copo de nieve o snowflakes.

Un esquema copo de nieve es como el esquema estrella, un modelo dimensional de un proceso de negocio. Con los copos de nieve, las tablas de dimensiones son normalizadas lo que tiene como objetivo reducir la posibilidad de almacenar datos redundantes. En los sistemas transaccionales, donde sí se busca la normalización, el diseño más adecuado se consigue cuando se logra llevar los datos a la tercera forma normal.

Los copos de nieve para el diseño de agregados pueden ser muy útiles (Adamson, 2006) simplificando el proceso de construirlos y mantenerlos mejorando el tiempo de respuesta del sistema. Usualmente no se necesita construir dimensiones agregadas y añadir llaves subrogadas, sino que las tablas de hechos agregadas pueden ser cargadas incrementalmente usando una simple consulta ya que las llaves subrogadas necesarias están presentes en el esquema base.

Otra situación según (Ralph Kimball, 2002b) donde se tendría un copo de nieve permisible es cuando se requiere de la construcción de un outrigger, aunque siempre es necesario validar cuan beneficioso puede ser. Un outrigger podría considerarse cuando un conjunto de datos de la dimensión base se cargan y administran en momentos de tiempo diferentes al resto de los datos de la dimensión (Figura 2-9).

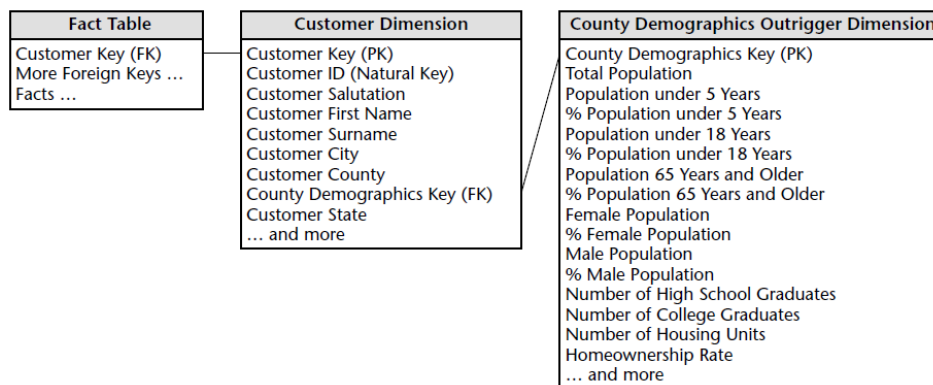


Figura 2-9 Copo de nieve permisible.

Por otra parte el manejo de las relaciones jerárquicas muchos a uno se recomiendan en una tabla de dimensión simple y no en un copo de nieve. Los copo de nieve pueden ser una carga para las ETL al tener que manejar las llaves enlazadas a las tablas

normalizadas las cuales pueden empezar a ser excesivamente complejas cuando el enlace entre las relaciones jerárquicas es objeto de cambios (Ross, Septiembre 3, 2008).

En (Ralph Kimball, 2002b) se encuentran algunas reflexiones interesantes sobre los copos de nieve aclarando lo válido de su uso pero también sus desventajas:

1. Los copos de nieve hacen el diseño más complejo y la simplicidad es el objetivo primario de un modelo dimensional desnormalizado.
2. Los numerosos enlaces entre las tablas se traducen en consultas con un bajo desempeño.
3. La habilidad de los usuarios para navegar a través de las dimensiones se ve afectada cuando están envueltos atributos de una o más dimensiones en presencia de restricciones, pues los usuarios necesitarían entender la relación entre los valores de los atributos de las dimensiones.

Como conclusión el uso de los copo de nieve debe estar bien justificado o se estará comprometiendo el desempeño final del DWH.

La existencia de nulos en los sistemas fuentes es otro aspecto que requiere atención especialmente durante el proceso de carga de los datos por lo que significan para las consultas que el usuario final realizará, sobre este tema se tratará en el siguiente epígrafe.

2.3.4 Tratamiento de nulos en el esquema dimensional

La mayoría de las bases de datos relacionales soportan el uso de valores nulos para representar la ausencia de datos. Los nulos pueden confundir tanto a los desarrolladores del DWH como a los usuarios porque la base de datos trata los nulos de forma distinta a los espacios en blanco o ceros, incluso aunque se vean igual. En (Thorntwaite, febrero 6, 2003) se aborda el caso de los valores nulos en los hechos, en las llaves foráneas de las tablas de hechos y en los atributos de las dimensiones.

En el caso de las llaves foráneas en las tablas de hechos existen situaciones en las que estas podrían tomar valor *null*, por ejemplo, si el valor de la llave foránea no es conocido

al momento de la extracción o si simplemente no es aplicable al hecho que se mide. Obviamente la integridad referencial es violada si una columna de la tabla de hechos declarada como llave foránea a una tabla dimensión toma valor *null*, ya que en una base de datos relacional *null* no es igual a él mismo. Para manejar esta situación se recomienda añadir una entrada especial en la tabla dimensión referenciada que indique la ausencia de valor, por ejemplo, “Dato no disponible todavía”.

Cuando el valor *null* pertenece a un hecho puede ser debido a que el valor no existe o el sistema de medición falló al capturar el valor, de cualquier manera generalmente se deja el valor como *null* pues la mayoría de los gestores de base de datos manejarán apropiadamente los *null* en las funciones de agregación como SUM, MAX, MIN, COUNT and AVG. Si se sustituye el valor *null* por cero, podrían distorsionarse estos cálculos de agregación.

En las tablas de dimensión generalmente se encuentran atributos nulos debido a, por ejemplo, quizás no todos los atributos han sido capturados todavía, y por tanto, se tienen algunos atributos desconocidos por un período de tiempo. También pueden haber ciertos atributos que solo aplican a un subconjunto de los miembros de la dimensión. En ambos casos se aplica la misma recomendación. Poner un nulo en estos campos puede resultar confuso para el usuario, ya que aparecerá como un espacio en blanco en reportes y menús desplegados, y requiere una sintaxis de consulta especial para encontrarlo. En lugar de esto se recomienda sustituir los valores nulos por una palabra descriptiva como las siguientes expuestas en (Becker, octubre 6, 2010):

Valor ausente: el atributo estaba ausente en la fuente de datos.

No ha ocurrido todavía: el atributo no está disponible todavía debido a asuntos relacionados con el tiempo en el sistema fuente.

Violación de dominio: el dato suministrado por el sistema fuente es inválido para este tipo de columna o está fuera de la lista de valores válidos.

No aplicable: El atributo no es válido para la fila de la dimensión en cuestión.

Los atributos textuales en las tablas de dimensión usualmente pueden contener los valores actuales que describen las condiciones nulas. Es recomendable evitar el uso de valores por defecto como espacios en blanco o cadenas de símbolos sin significado como @@@ ya que esto solo confunde a los usuarios del negocio. Los valores por defecto para cada atributo dimensional deben considerarse cuidadosamente y proveerlos del mayor significado posible.

Los atributos numéricos en las tablas de dimensión necesitarán tener un conjunto de valores especiales. El valor cero es, a menudo, la mejor opción porque es usualmente obvio para los usuarios que es artificial. Algunos atributos numéricos se presentaran como una elección difícil si el usuario combina estos valores en algún tipo de cálculo. Si en el valor numérico actual se usa cero para que actúe como nulo, este valor participará en los cálculos produciendo resultados engañosos. Un valor actual nulo causará un error en los cálculos, pero al menos no producirá resultados incorrectos. Una opción sería mostrar a los usuarios los valores nulos de los atributos numéricos de las dimensiones como *null*, de esta forma se pueden hacer reportes y cálculos sobre estos atributos sin preocuparse por resultados distorsionados.

2.3.5 Tratamiento de fechas.

En un esquema dimensional es usual que se identifiquen varias fechas diferentes, cada una con un significado específico para el negocio, que deben ser incluidas en el diseño dimensional. Según (Becker, 2004) hay que tener en cuenta que no todas las fechas son creadas ni manipuladas de la misma manera. Muchas fechas terminan como llaves foráneas a la dimensión fecha en las tablas de hechos. La mayoría de las restantes fechas se convierten en atributos de otras dimensiones. Finalmente, algunas fechas son incluidas en el diseño para facilitar el proceso de ETL y/o capacidades de auditoría.

Así, si diseñamos una tabla de hechos que integre los tres tipos de transacción de una compañía, cada fila de nuestra tabla de hechos incluiría una dimensión “tipo de transacción” que identifique el tipo de transacción que representa, así como una dimensión fecha de transacción. De esta forma el significado para el negocio de la fecha

está definido por la dimensión “tipo de transacción”. En este caso no se necesitan tres llaves de fecha ya que con una sola llave es suficiente para una fila dada.

En otras situaciones, una transacción representada por una fila de la tabla de hechos puede estar definida por múltiples fechas, como la fecha en que ocurre el evento y la fecha en que se entra al sistema. En este caso ambas fechas serán incluidas como llaves foráneas. Se podría usar una única dimensión fecha pero jugando distintos roles.

Muchas de las fechas significativas para el negocio serán incluidas como atributos en las tablas de dimensión. Por ejemplo la fecha de apertura de cuenta pudiera ser incluida en la dimensión cuenta.

Cuando las fechas son atributos de tablas de dimensión, es necesario considerar el uso que se les va a dar en los reportes y análisis. ¿Es suficiente saber la fecha en que fue abierta una cuenta o debemos incluir también atributos como año de apertura de cuenta y mes de apertura de cuenta? Estos atributos adicionales mejoran la habilidad de usuario de negocio de hacer preguntas analíticas interesantes agrupando las cuentas basado en el año o el mes de apertura de la cuenta.

Para soportar un análisis relativo a fechas más extensivo de estos atributos de las dimensiones, se puede incorporar una robusta dimensión fecha como un outrigger a la tabla de dimensión. En este caso se incluye la llave subrogada para la fecha que aplica en nuestra dimensión en vez de la fecha en sí (Figura 2-10). Esta técnica brinda acceso a la riqueza de atributos de la dimensión fecha principal para el análisis. Sin embargo, hay que recordar que el uso extensivo de dimensiones outrigger puede comprometer la usabilidad y el rendimiento. También hay que tener cuidado de que todas las fechas outrigger estén dentro del rango almacenado en la tabla dimensión estándar.

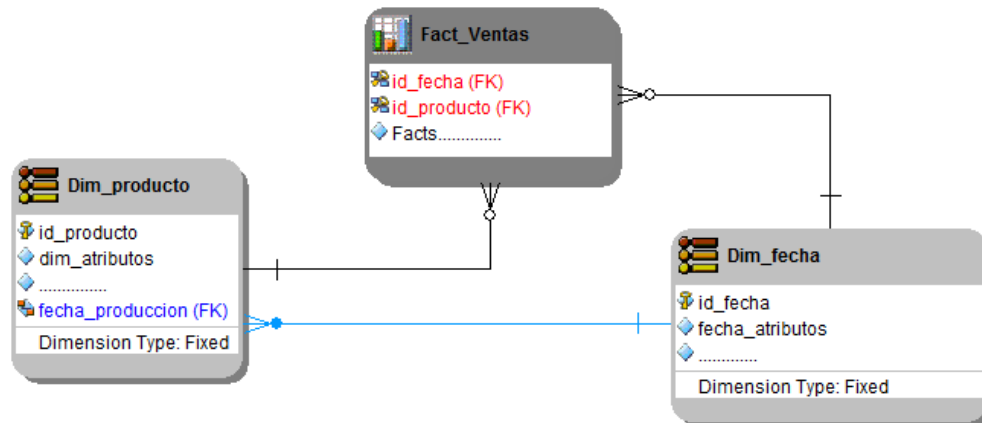


Figura 2-10 Tratamiento de fechas con un outrigger.

Hay fechas adicionales para ayudar al proceso de ETL y apoyar la habilidad de auditoría de los datos. Fechas como fecha de efectividad de fila, fecha de expiración de fila, fecha de carga de fila o fecha de última actualización de fila deben ser incluidas en cada tabla de dimensión. Aunque estas fechas pueden no ser accesibles para el usuario, ellas son muy útiles para el equipo del DWH.

2.4 Dimensiones que juegan rol.

En el epígrafe anterior se mencionaba la posibilidad de que la dimensión fecha jugara distintos roles. Efectivamente el juego de roles se evidencia cuando una simple tabla de dimensión aparece varias veces en una misma tabla de hechos (Ralph Kimball, 2002b). Probablemente la dimensión que comúnmente más roles juegue sea la dimensión fecha. En las tablas de hechos acumulativas generalmente encontramos varios usos de las fechas.

En la implementación de las dimensiones que juegan rol se recomienda primero construir una simple tabla de dimensión genérica y entonces implementar cada uno de los roles con una vista en esta tabla genérica (Ralph Kimball, 2004) como muestra la (Figura 2-12).

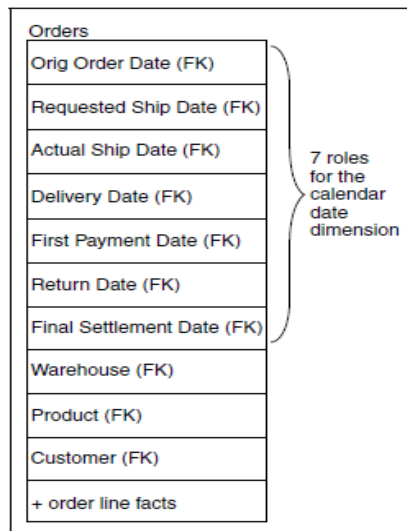


Figura 2-11 Una tabla de hechos acumulativa donde la dimensión fecha juega distintos roles.

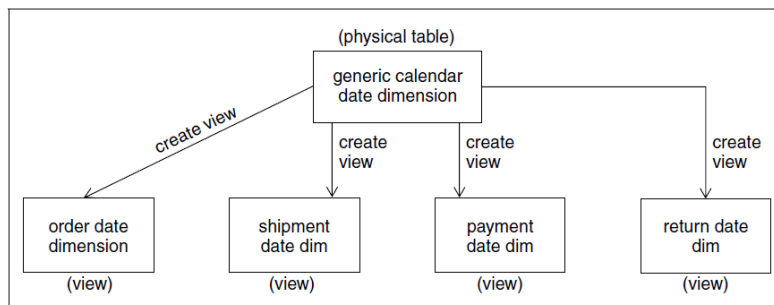


Figura 2-12 Múltiples roles de la dimensión calendar.

2.5 Dimensiones conformadas.

En (Adamson, 2006) se explica que para maximizar el valor y el éxito de un DWH resulta crítico asegurar la conformidad dimensional (Figura 2-13). Un conjunto común de dimensiones es planeado y multi referenciado con los distintos procesos del negocio que serán representados con tablas de hechos. Esta estructura dimensional común es conocida como el **bus del DWH**. Planear la conformidad dimensional con anticipación permite al DWH ser implementado en varios MD (área específica del negocio) a la vez, evitando potenciales incompatibilidades. Cada MD provee valor directo cuando es implementado y se integrará con los demás a medida que estos sean desarrollados.

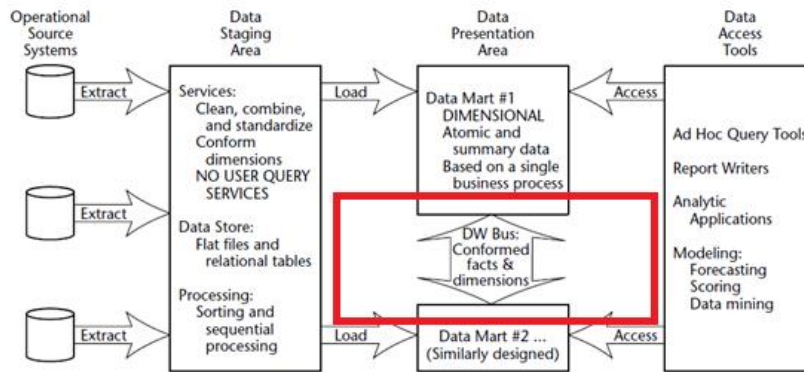


Figura 2-13 El bus del DWH dentro de sus elementos básicos.

Para crear, documentar y comunicar la arquitectura de bus se utiliza como herramienta la matriz de bus (Ralph Kimball, 2002b) (Figura 2-14). Las filas (rows) de esta matriz representan los procesos del negocio o MD y sus dimensiones asociadas. La matriz describe a los MD y no a las tablas de hechos individuales. Construir la matriz implica comenzar un útil ejercicio creativo de determinar si dada una dimensión esta podría estar relacionada con algún MD. Al observar la matriz pueden determinarse que dimensiones requieren especial atención por sus múltiples apariciones en los MD.

BUSINESS PROCESSES	COMMON DIMENSIONS							
	Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
Retail Sales	X	X	X	X				
Retail Inventory	X	X	X					
Retail Deliveries	X	X	X					
Warehouse Inventory	X	X			X	X		
Warehouse Deliveries	X	X			X	X		
Purchase Orders	X	X			X	X	X	X

Figura 2-14 Ejemplo de una matriz de bus.

Las dimensiones conformadas son entonces la implementación de esta matriz estableciendo el conjunto de atributos que serán útiles a cada uno de los MD. Deben definirse con el mayor nivel de granularidad posible. Obviamente requieren de una implementación coordinada y si se modifican los atributos existentes o se añaden otros nuevos se debe examinar con todo el equipo de trabajo. Las dimensiones cuando se han hecho conformadas a pesar de estar en diferentes MD tienen idénticos nombres de columnas, idénticos valores de los atributos y la misma combinación de valores de los

atributos. Sin embargo cuando se tiene una dimensión que es un subconjunto perfecto de otra también se dice que es conformada.

La (Figura 2-15) muestra un ejemplo de dimensión conformada (Adamson, 2006) donde:

- Los atributos de una tabla son un subconjunto de los atributos de las otras. La única excepción es el atributo llave.
- Los valores de los atributos son almacenados idénticamente ("January" no conforma con "JANUARY" o "Jan").
- Cada tabla contiene las mismas combinaciones distintas de valores para los atributos comunes (No hay una combinación month/year presente en una tabla que no esté en la otra).

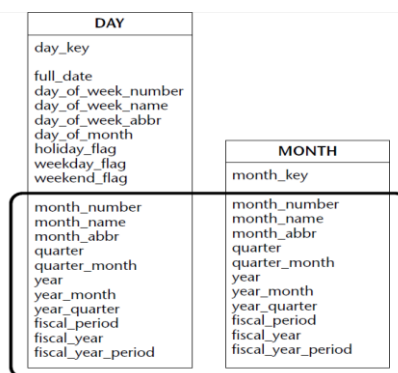


Figura 2-15 Ejemplo de dimensión conformada.

2.6 El proceso de carga de los datos.

Los sistemas de Extracción-Transformación-Carga (ETL por sus siglas en inglés) constituyen los cimientos del DWH. Un diseño apropiado de un sistema ETL debe extraer los datos de los sistemas fuentes, imponer calidad y consistencia a los datos y finalmente ofrecer un formato de presentación de los datos a los desarrolladores de modo que estos puedan construir aplicaciones y los usuarios finales puedan tomar decisiones (Ralph Kimball, 2004).

Cada paso en la construcción del DWH tiene sus dificultades específicas pero una vez que estas han sido superadas el mantenimiento del mismo es otro reto. Para ello existen varias técnicas entre ellas la carga incremental que es un proceso que se realiza

periódicamente para mantener sincronizado el DWH con los respectivos sistemas fuentes. Este proceso puede correr en un intervalo de tiempo o continuamente (en tiempo real) aunque lo primero se recomienda como una mejor práctica. Las rutinas ETL que cargan datos incrementalmente son usualmente el resultado de un proceso de carga inicial de los datos históricos hacia el DWH.

Es útil, por otra, parte (Ralph Kimball, 2002a), mantener en los metadatos la fecha de la última carga para facilitar el proceso y es mucho más eficiente hacer la carga incremental solo de los registros que han sido cambiados o añadidos desde la carga previa.

El proceso de carga incluye entre otros pasos, el manejo de las llaves subrogadas y de las tablas de dimensiones. Las tablas de hechos no pueden cargarse hasta que no haya finalizado exitosamente la carga de cada una de las tablas de dimensiones.

Una vez que se ha logrado diseñar e implementar un MD con la información necesaria, son imprescindibles herramientas de apoyo para realizar análisis, y para ello las técnicas OLAP (*On-Line Analytical Processing*) ofrecen una serie de ventajas que van desde minimizar el tiempo de respuesta, hasta ofrecer un modelo que facilita la selección, recorrido y exploración de los datos.

2.7 Análisis de datos a través de OLAP.

Los MD son diseñados para soportar variadas técnicas de análisis multidimensionales incluyendo las herramientas OLAP.

El término OLAP fue acuñado por E.F. Codd y se refiere a un tipo de software que permita el análisis interactivo de los datos a través de una interfaz hombre-computadora. Es comúnmente usado para etiquetar una categoría de tecnologías de software que posibilita a analistas, administradores y ejecutivos llevar a cabo el acceso y el análisis ad hoc de los datos basados en su dimensionalidad. La forma multidimensional provee rapidez, consistencia, acceso interactivo a una variedad de posibles vistas de la información. Sin embargo el término en sí mismo no implica el uso de análisis o estructuras multidimensionales (Claudia Imhoff, 2003).

Los cubos son la estructura de datos clave en OLAP, son subconjuntos de un DWH, organizados y resumizados dentro de una estructura multidimensional.

2.7.1 Categorías de sistemas OLAP.

La arquitectura ROLAP (OLAP relacional) almacena las medidas detalladas en una base de datos relacional que es la fuente del cubo. Es capaz de usar agregados precalculados si están disponibles, o de generar dinámicamente los datos originales.

Los usuarios finales ejecutan sus análisis multidimensionales a través de un motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Éstas son ejecutadas sobre la base de datos relacional y con los resultados obtenidos se genera un conjunto multidimensional para devolver a los usuarios. Este tipo de sistemas los datos detallados estarán tan actualizados como los del MD. La desventaja de ROLAP es que la recuperación de cualquier tipo de medida es más lenta que para otras arquitecturas (Bolado, 2012).

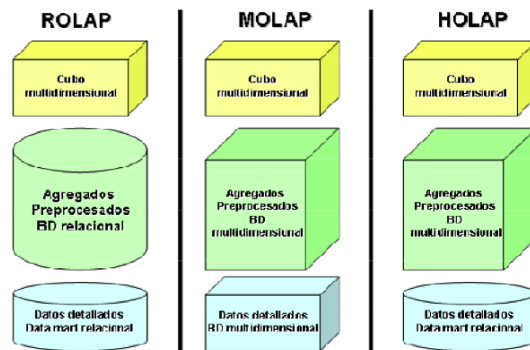


Figura 2-16 Arquitecturas OLAP.

En la arquitectura MOLAP (OLAP multidimensional) el formato original en el que los datos son almacenados es propiamente una base de datos multidimensional. Los agregados y detalles de los datos son almacenados dentro de la misma, y es la base de datos la encargada del manejo, acceso y obtención de los datos solicitados.

En HOLAP (OLAP Híbrido) los agregados y datos de la navegación se almacenan en una estructura MOLAP pero los detalles de los datos se mantienen en una base de datos

relacional (Dongen, 2009). HOLAP sacrifica el tiempo de respuesta de los datos detallados para conseguir una menor latencia y mejorar los tiempos de carga.

2.7.2 Mondrian y OLAP.

Mondrian es un motor OLAP de código abierto escrito en java. El núcleo de Mondrian es un JAR que actúa como "JDBC para OLAP". En él se ejecutan consultas en el lenguaje MDX¹, leyendo los datos desde una base de datos relacional y presentando los resultados en un formato multidimensional a través de un API escrito en java (Dongen, 2009). Un sistema que utilice Mondrian consta de cuatro capas fundamentales: la capa de presentación, cálculo, agregación y de almacenamiento.

La capa de presentación determina qué verá el usuario final y como puede éste interactuar con el sistema. Existen variadas formas de presentación pero todas tienen en común la gramática multidimensional de dimensiones y medidas. La capa de cálculo parsea, valida y ejecuta las consultas MDX. La capa de agregación es la responsable de mantener los agregados en la cache. Una agregación es un resumen de datos precalculados que mejora el tiempo de espera al tener preparadas las respuestas antes de que se planteen las consultas. La capa de almacenamiento es un RDBMS (Sistema de Gestión de Bases de Datos Relacionales), responsable de proveer las celdas y miembros desde las tablas de dimensiones.

Otra importante característica de Mondrian (Dongen, 2009) es la seguridad de su modelo dando soporte a los roles. Los roles pueden ser usados para restringir los datos que son accesibles por un usuario, limitando el número de vistas OLAP y reportes que necesitan ser desarrollados.

Modrian es además el motor OLAP de la suite de inteligencia de negocio Pentaho, este ofrece la herramienta visual Pentaho Schema Workbench (PSW) para construir y probar los esquemas multidimensionales sobre este motor. Un esquema es un mapeo entre el modelo relacional y el multidimensional que es usado para ayudar a traducir las consultas

¹ MDX es la abreviación de Multi Dimensional eXpressions, es un lenguaje diseñado específicamente para consultas OLAP a las bases de datos. Originalmente fue desarrollado por Microsoft.

MDX a consultas SQL, y luego transformar los resultados relacionales recibidos en respuesta a la consulta SQL en resultados multidimensionales. El esquema contiene un modelo lógico consistente de cubos, dimensiones, jerarquías, niveles y miembros que constituyen los constructores usados para escribir las consultas MDX, el modelo físico o esquema estrella es representado a través de este modelo lógico.



Figura 2-17 DWH con Modrian.

2.8 Conclusiones del capítulo

Para el diseño de un DWH el modelo dimensional ofrece múltiples ventajas al responder a las necesidades analíticas de esta tecnología. Durante el proceso de diseño se deben identificar las dimensiones lentamente cambiantes (SCD) y seleccionar el tratamiento más adecuado ya sean las técnicas de SCD tipo 1, 2 y 3 u otras como los tipos 0, 4 y 5.

En el refinamiento del diseño una opción es utilizar las dimensiones junk si se dispone de un conjunto de valores de los cuales podemos obtener todas sus posibles combinaciones.

Las tablas puentes ofrecen facilidades para representar jerarquías de profundidad variable y relaciones muchos a muchos.

Teniendo en cuenta las características del caso de estudio descrito se valora el uso de la tabla puente para diseñar el proceso de evaluación, el tratamiento de nulos en las tablas de dimensiones y de hechos y el uso de dimensiones que juegan rol.

En el capítulo siguiente se explica el proceso de diseño e implementación seguido para dar solución al problema planteado.

CAPÍTULO 3. DISEÑO E IMPLEMENTACIÓN DEL MERCADO DE DATOS PARA EL DEPARTAMENTO DE CUADROS DE LA UCLV.

En este capítulo se da una solución práctica al caso de estudio descrito a partir del diseño e implementación de un MD. Siguiendo la metodología propuesta por Kimball se siguen cuatro pasos fundamentales para el diseño. Se exponen además el uso de las herramientas de Pentaho para la implementación de los procesos ETL y el diseño de los reportes que permiten validar el funcionamiento de la aplicación final con datos reales del Departamento de Cuadros de la UCLV.

3.1 Diseño lógico de la solución.

Diseñar es, en muchas ocasiones, la tarea más compleja pero también la más necesaria pues garantiza minimizar los errores futuros tratando de modelar todos los requerimientos del usuario final antes de iniciar la construcción física del MD. Según (Ralph Kimball, 2002b) se definen cuatro pasos para guiar el proceso del modelado. Teniendo como referente la descripción del caso de estudio hecha en capítulos anteriores y los aspectos teóricos que han sido abordados se expone el resultado obtenido.

3.1.1 Selección de los procesos del negocio.

Un proceso es una actividad natural del negocio que se lleva a cabo y es soportada por una colección de datos fuentes (Ralph Kimball, 2002b).

Como resultado del análisis del caso de estudio del Departamento de Cuadros se determinó la existencia de tres procesos de negocio que trabajan con un conjunto de datos específicos y se actualizan en intervalos de tiempo diferentes.

Proceso 1: Movimiento de cuadros.

Proceso 2: Actualización mensual del estado de la reserva.

Proceso 3: Evaluación de los cuadros (Anual e integral cada dos años).

3.1.2 Declarando el grano.

Cuando se han identificado los procesos el siguiente paso es definir la granularidad para cada uno. El nivel de detalle contenido en una tabla de hechos es llamado grano (Joy Mundy, 2006). Se recomienda construir tablas de hechos con el mayor nivel de detalle posible que permitan los sistemas fuentes. El nivel atómico provee completa flexibilidad para operaciones roll up de los datos a cualquier nivel de sumarización necesario y a través de cualquier dimensión.

EL proceso de actualización mensual de la reserva interesa por reserva, por cargo, si funge como sustituto directo al cargo, y por mes.

La evaluación de los cuadros se realiza en dos momentos: anuales y cada dos años. En cada caso son importantes el grupo de indicadores utilizados para evaluar, el cuadro, el desarrollo del cuadro y la fecha.

El movimiento de cuadros por su parte interesa por cuadro, cargo anterior, cargo actual, causa del movimiento, por fecha de nombramiento, número de resolución, si viene de la reserva y si corresponde con la proyección.

3.1.3 Selección de las dimensiones.

La granularidad en sí misma determina un conjunto mínimo o primario de dimensiones (Ralph Kimball, 2002a). Por tanto como una versión inicial de las dimensiones necesarias se hayan: cuadro, reserva, cargo, fecha, causa del movimiento, e indicadores para la evaluación.

La granularidad del paso anterior ofrece estas dimensiones pero sin duda un análisis más detallado de cada uno de los procesos podría brindar otros puntos de vista.

Primeramente se encuentra que la actualización de la reserva hace uso de la dimensión reserva y el resto de los procesos utiliza específicamente a los cuadros. Si se observan los datos fuentes que se almacenan de cada uno respectivamente se encuentra una gran correspondencia con la exclusión de algunos datos que son específicos de los cuadros.

Otro aspecto es el hecho de que en un gran número de casos los cuadros constituyen reservas y viceversa. Para el sistema se recomienda entonces almacenar en una dimensión persona los datos de cuadros y reservas precisos para cada uno de los procesos descritos. El paso anterior a este hubiese sido diseñar una dimensión conformada útil a otros departamentos dentro de la UCLV, como el de Recursos Humanos, sin embargo al no existir un proceso común de desarrollo de un DWH que incluya estos requerimientos solo se puede trabajar con los datos del Departamento de Cuadros que se tienen a disposición.

La evaluación de los cuadros constituye otro proceso que se estuvo tomando como ejemplo cuando en el capítulo 2 se hacía referencia al uso de las dimensiones junk y las tablas puentes para modelar esta situación por las características del problema: un conjunto de indicadores que toman valor en un rango finito, con los que se evalúa a cada cuadro y que pueden ser objeto de cambio. La utilización de una dimensión junk sería muy recomendable por el mejor uso del espacio de almacenamiento que hace, pero no permitiría almacenar ninguna información extra sobre estos indicadores. Sin embargo se observó que con el uso de una dimensión puente se resolvía esta problemática.

3.1.4 Identificación de las tablas de hechos para el proceso de movimiento de cuadros

Las tablas de hechos sin hechos son útiles cuando la única medición de un evento es su ocurrencia, o cuando es útil llevar la cuenta de intersecciones significativas de valores dimensionales que no corresponden con transacciones.

Para el movimiento de cuadros se han seleccionado dos tablas de hechos sin hechos y periódicas, ya que este proceso se actualiza mensualmente: una tabla de hechos movimientos **fact_mov_cuadro** para almacenar cada mes los movimientos que tienen lugar y una tabla **fact_rel_cargo_cuadro** que tiene como objetivo recoger la relación mensual de cada cargo y la persona que lo ocupa de modo que se pueda obtener un estado de esta relación en cualquier momento que se especifique.

3.1.5 Identificación de las tablas de hechos para el proceso actualización de la reserva

La actualización de la reserva se modela con el objetivo de mantener el estado de la relación entre un cargo y sus reservas para lo cual se utiliza también una tabla de hechos sin hechos y periódica, **fact_estado_reserva**.

3.1.6 Identificación de las tablas de hechos para el proceso evaluación de los cuadros

El proceso de evaluación tiene lugar en dos períodos diferentes por ello se seleccionaron dos tablas de hechos sin hechos: **fact_eval_anual_cuadros** y **fact_eval_integral_cuadros** para modelar dicho proceso, con periodicidad anual y cada dos años respectivamente.

3.1.7 Diseño del Mercado de Datos.

La metodología que se ha seguido permite hacer un diseño inicial y posteriormente analizar los detalles que responden a características del problema que precisan de transformaciones del diseño primario. Como resultado se obtuvieron las siguientes especificaciones de las dimensiones y las tablas de hechos:

Dimensiones:

Dim_fecha: Se diseña con el objetivo de mantener las fechas que determinan el carácter histórico del MD. Los atributos usados responden a las características de los datos almacenados en los sistemas fuentes y el campo curso se calculará para su posible uso en análisis que se hagan posteriormente.

Dim_cargo: Esta dimensión describe los atributos de un cargo. Su construcción responde a la representación interna de la estructura jerárquica presente en la UCLV. Los cargos se asocian a direcciones, departamentos, facultades etc. Y estos forman una jerarquía de cuatro niveles.

Dim_causa_mov: La dimensión causa de movimiento de un cuadro se presenta como una dimensión causal con el objetivo de almacenar información sobre las condiciones bajo las que ocurre un movimiento.

Dim_persona: Es la encargada de almacenar los datos que describen a los cuadros y las reservas.

Dim_indicador: Describe los indicadores utilizados para la evaluación anual e integral auxiliado de una tabla puente que almacena la relación entre las tablas de hechos y esta dimensión junto al valor de cada indicador para cada evaluación individual realizada.

Tablas de hechos:

Las tablas de hechos periódicas cargan sus datos al final de períodos de tiempo definidos ya sea un día, un mes, un año. De cierto modo es como si resumieran las actividades transaccionales que ocurren en ese período (Ralph Kimball, 2004).

Para facilitar las operaciones de agregación sobre las tablas de hechos, se ha incluido en cada tabla de hechos identificada un hecho llamado *bd_cant* que toma siempre valor 1.

Las tablas de hechos identificadas corresponden a tablas de hechos sin hechos periódicas y son las siguientes:

fact_eval_anual_cuadros: La (Figura 3-1) muestra las relaciones con el resto del esquema. Esta tabla registra las evaluaciones que se realizan a los cuadros anualmente. Se ha decidido utilizar *bd_desarrollo_cuadro* como una dimensión degenerada ya que lógicamente no se puede considerar como atributo de ninguna otra dimensión y tampoco existen otros atributos que describan el desarrollo de un cuadro que se puedan incluir en una dimensión estándar. Aun así, el desarrollo del cuadro es imprescindible para describir la evaluación.

fact_eval_integral_cuadros: Registra las evaluaciones integrales que se realizan cada dos años. En la (Figura 3-2) se muestran las relaciones con otras tablas. Es similar a la anterior, solo que cambia el periodo de tiempo que representa.

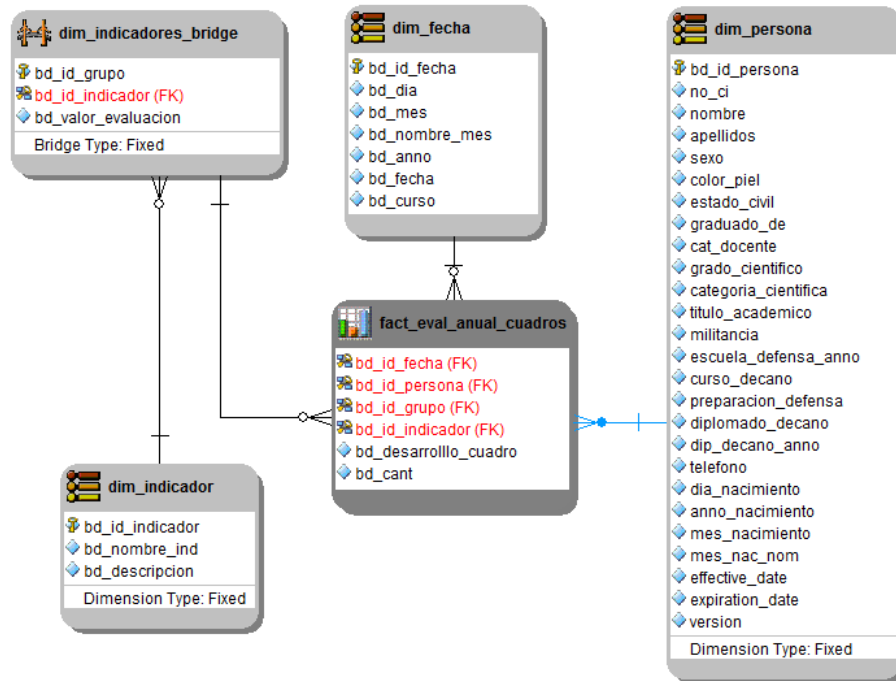


Figura 3-1 Relaciones de la tabla de hechos fact_eval_anual_cuadros.

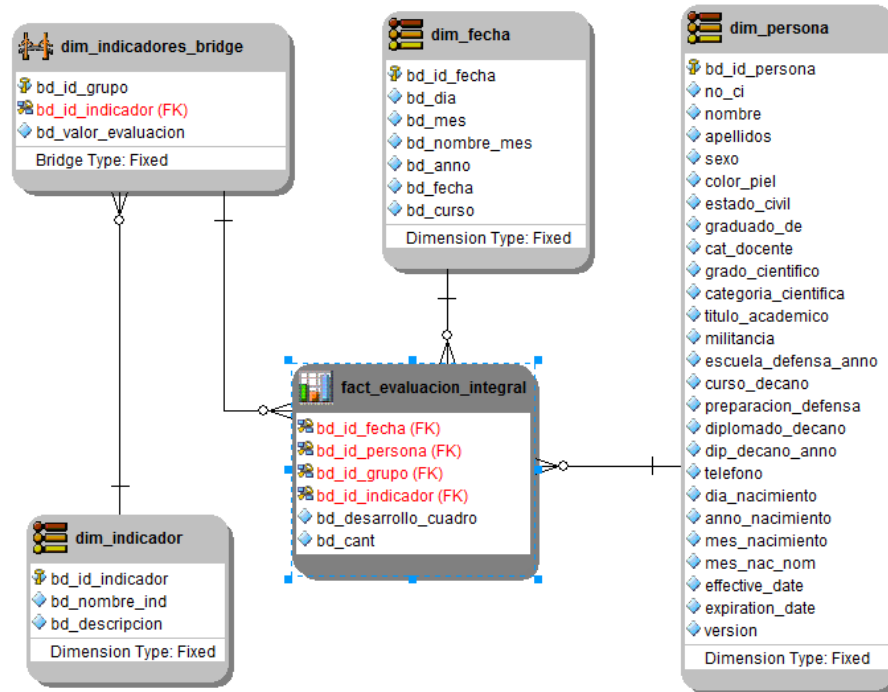


Figura 3-2 Relaciones de la tabla de hechos fact_evaluacion_integral.

fact_mov_cuadros: Esta tabla contiene tres dimensiones degeneradas: *bd_resolución* que indica la resolución a la que responde el movimiento del cuadro, *bd_desde_reserva* que indica si el cuadro proviene de la reserva y *bd_proyectado* para saber si el movimiento estaba proyectado. Los movimientos se registran mensualmente. La (Figura 3-3) muestra las relaciones. Para el diseño de esta tabla se tuvo en cuenta el tratamiento de nulos en las llaves foráneas de las tablas de hechos mencionado en el capítulo 2. En este caso cuando, por ejemplo, un cuadro, causa baja, y por tanto ya no ocupa ningún cargo, en vez de poner valor *null* en el campo *id_cargo_act*, la solución es asignarle a este campo la llave subrogada de una entrada especial en la dimensión **cargo** que se ha nombrado “Baja”.

fact_rel_cargo_cuadro: Se utiliza para registrar la relación completa de cada cargo y el cuadro que lo ocupa y como acompañante de la tabla anterior, que solo registra los movimientos ocurridos en un mes, este dúo es muy útil ya que hay respuestas que solo se pueden obtener consultando una de estas dos tablas y que no es posible encontrarla en la otra fácilmente, por ejemplo, el porciento de cuadros en funciones, en determinado mes, que son negros o mestizos, se obtiene con una simple consulta a esta tabla mientras que recuperar esta misma información en la tabla anterior sería demasiado complejo.

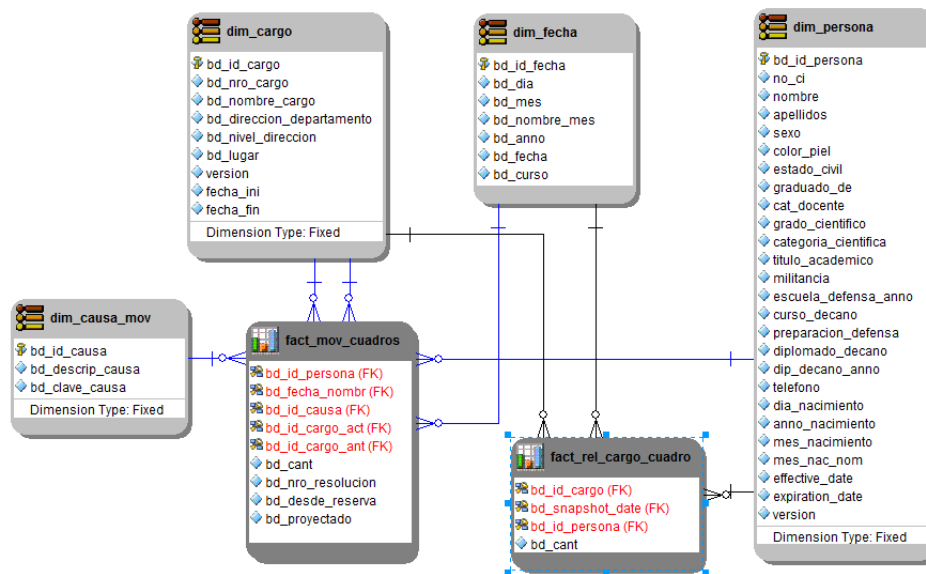


Figura 3-3 Relaciones de la tabla de hechos fact_mov_cuadros

fact_estado_reservas: La actualización de la reserva ocurre anualmente pero cuando se efectúan movimientos de cuadros generalmente ocurren cambios en la reserva. Para indicar si la reserva funge como sustituta directa al cargo se utiliza la dimensión degenerada *bd_funge_sustituto*. La siguiente muestra las relaciones que se establecen con las tablas de dimensiones.

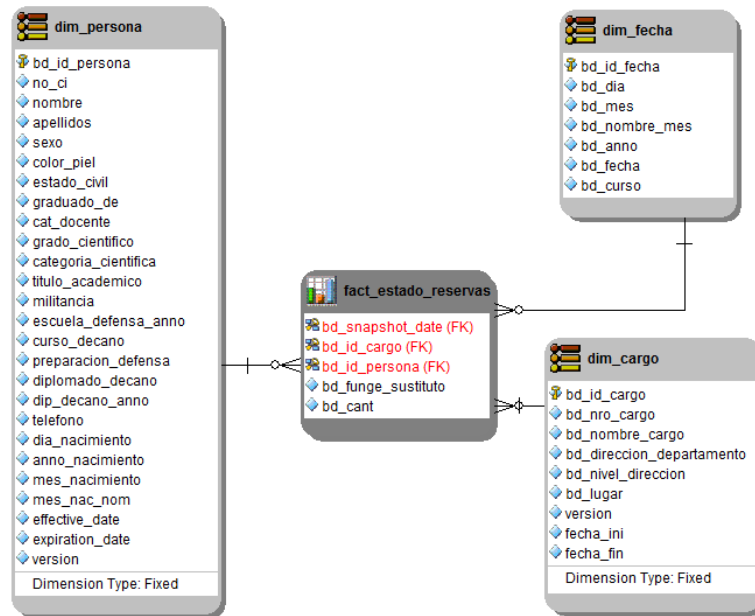


Figura 3-4 Relaciones de la tabla de hechos fact_estado_reservas

3.2 Implementación del MD para el Departamento de Cuadros.

Al concluir el diseño lógico del Mercado de Datos el diseño físico es el siguiente paso. La herramienta ERStudio ofrece las facilidades para la generación automática del modelo físico relacional especificando algunos parámetros como la plataforma de base de datos objetivo, que en este caso es MySQL. En el modelo físico generado se encuentran tablas y columnas en lugar de tablas de hechos, dimensiones y atributos clásicos del modelado dimensional. Las llaves foráneas de las tablas de hechos se transforman también en llaves primarias. A partir del modelo físico se genera un script con extensión `.sql` que contiene las sentencias necesarias para el SGBD que utiliza el MD.

Mediante la herramienta **phpMyAdmin** que es una interfaz web para la administración de MySQL se crea la base de datos **dep_cuadros** y se importa el archivo .sql creando las tablas y relaciones correspondientes en la BD, de la cual se muestra una vista en la siguiente figura.

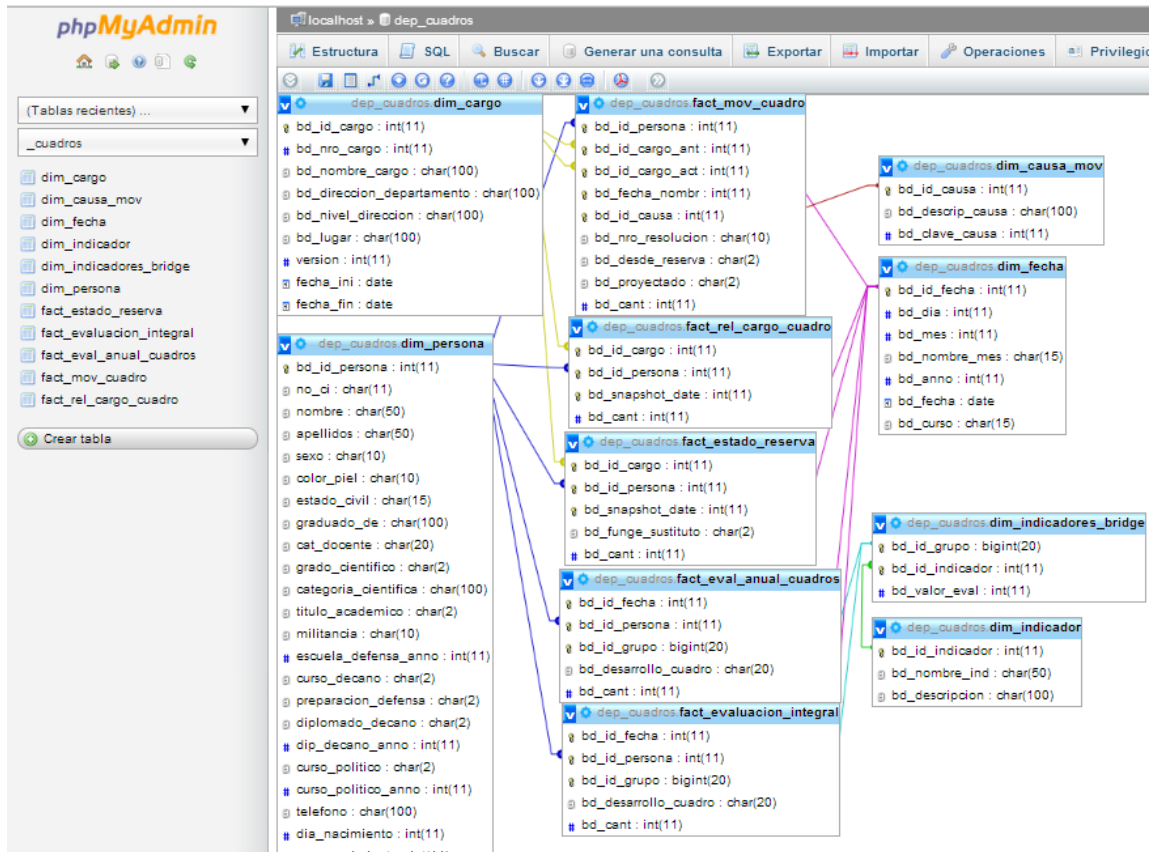


Figura 3-5 Tablas de la BD en el phpMyAdmin.

3.3 Procesos de extracción, transformación y carga para poblar el Mercado de Datos

Pentaho ofrece una colección de herramientas conocidas como Pentaho Data Integration que están diseñadas para dar soporte a los procesos ETL (Matt Casters, 2010). Está construido sobre la base de dos objetos fundamentales: las transformaciones y los trabajos, además de herramientas y utilidades para crearlos, manejarlos y ejecutarlos:

Spoon: IDE gráfico para la creación de transformaciones y trabajos.

Kitchen: Herramienta de línea de comandos para ejecutar los trabajos.

Pan: Herramienta de línea de comandos para ejecutar las transformaciones.

Carte: Un servidor ligero para ejecutar las transformaciones y trabajos desde un host remoto.

3.3.1 Transformaciones y trabajos diseñados.

El Spoon es la herramienta escogida para el diseño de las transformaciones y los trabajos por las facilidades gráficas que brinda.

Para obtener la dimensión cargo y la dimensión causa de movimiento se parte de las tablas cargo y causa_mov del sistema fuente. Los datos de tipo string se someten a algunas transformaciones para lograr homogeneidad en los mismos y en el momento de la carga se definen las dimensiones de cambio lento que se han determinado como tipo 2 y el tratamiento que recibirán, en este caso Insertar para el campo de **Tipo de actualización en dimensión** del paso **Búsqueda/Actualización en Dimensión**.

La siguiente transformación es diseñada para poblar la dimensión persona.

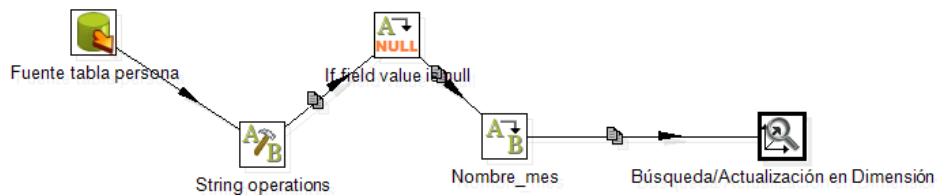


Figura 3-6 Transformación para la dimensión persona

El paso **If field value is null** es utilizado con el objetivo de manejar los nulos en la dimensión persona y ofrecer a los usuarios más información sobre los valores nulos: la no disponibilidad en el sistema fuente en el momento de la carga, la no presencia de determinada característica etc.

Los cambios de tipo lento identificados se manejan de la siguiente forma en el paso de **Búsqueda/Actualización en Dimensión**:

Campos Búsqueda/Actualización

#	Campo de Dimensión	Campo de flujo con el que comparar	Tipo de actualización de dimensión
1	nombre	nombre	Punch through
2	apellidos	apellidos	Punch through
3	sexo	sexo	Punch through
4	color_piel	color_piel	Punch through
5	estado_civil	estado_civil	Insertar
6	graduado_de	graduado_de	Actualizar
7	cat_docente	categoria_docente	Insertar
8	grado_cientifico	grado_cientifico	Insertar
9	categoria_cientifica	categoria_cientifica	Insertar
10	titulo_academico	titulo_academico	Insertar
11	militancia	militancia	Insertar
12	escuela_defensa_anno	escuela_defensa_anno	Insertar
13	curso_decano	curso_decano	Insertar
14	preparacion_defensa	preparacion_defensa	Insertar
15	diplomado_decano	diplomado_decano	Insertar
16	dip_decano_anno	dip_decano_anno	Insertar
17	curso_politico	curso_politico	Insertar
18	curso_politico_anno	curso_politico_anno	Insertar
19	telefono	telefono	Actualizar
20	dia_nacimiento	dia_nacimiento	Punch through
21	mes_nacimiento	mes_nacimiento	Punch through
22	anno_nacimiento	anno_nacimiento	Punch through
23	mes_nac_nom	nombre_mes	Punch through

Figura 3-7 Manejo de las SCD en la dimensión persona.

La opción Actualizar maneja el cambio de tipo 1 haciendo una corrección del dato escrito en el campo seleccionado para la instancia válida en el momento que se detecta el cambio.

La opción Punch through es utilizada para realizar una actualización de tipo 1 pero en cada una de las instancias que se han insertado con la misma llave natural por eso se seleccionan los campos nombre, apellidos, sexo, color de la piel, año, mes y día de nacimiento para este caso ya que estos datos deben mantenerse invariables en cualquier persona.

La opción Insertar es para el caso del tipo 2, o sea, se inserta una nueva fila con la misma llave natural pero una nueva llave subrogada. Los atributos que se encuentran en esa categoría han sido con el objetivo de conocer el estado de la persona en los distintos instantes de tiempo, por ejemplo como ha cambiado su categoría docente, su categoría científica, militancia, si ha cursado o no diplomado de decano, cursos políticos, etc.

La tabla de hechos fact_mov_cuadros puede llenarse una vez que se han cargado para el mercado de datos las dimensiones mencionadas anteriormente y se han generado cada una de las llaves subrogadas correspondientes. La tabla mov_cuadros del sistema fuente

es la que almacena los movimientos que tienen lugar cada mes por tanto es la fuente en este caso.

La dimensión fecha es la base del MD permitiendo el carácter histórico del mismo. Una opción a valorar para la creación de la dimensión fecha es generar un número determinado de fechas inicialmente sin embargo el sistema fuente con el que se trabaja generalmente ingresa pocas fechas de cada mes por lo que una generación múltiple abarcaría más espacio de almacenamiento y un gran número de fechas que nunca se utilizarían, surgiendo entonces la variante de adicionar a la dimensión fecha solo las fechas que son utilizadas en cada uno de los procesos o aquellas que definen los momentos de carga para las tablas de hechos periódicas en cuyo caso se obtendrán de los datos del sistema en el momento de la carga. La siguiente figura es la transformación realizada para llenar la tabla de hechos fact_mov_cuadros.

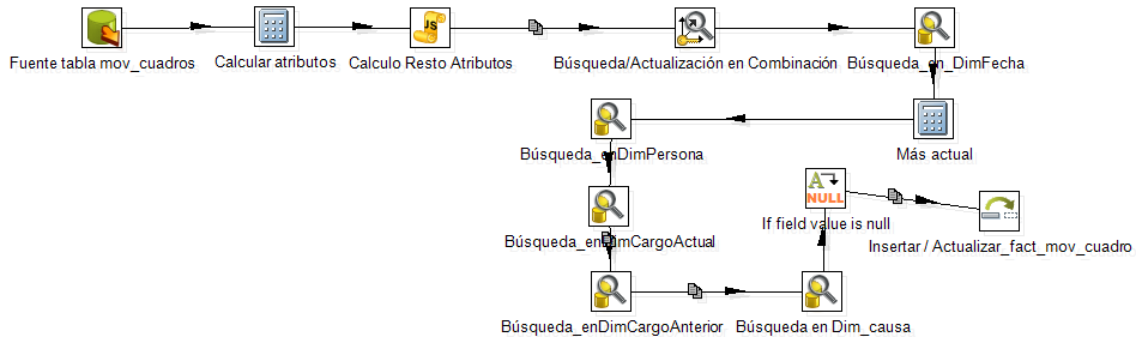


Figura 3-8 Tranformación para tabla de hechos fact_mov_cuadros .

El paso **Calcular atributos** se utiliza para obtener el día, mes y año de las fechas en las que se realizan los movimientos de cuadros obtenidas desde la tabla mov_cuadros del sistema fuente.

Calculadora										
Nombre paso <input type="text" value="Calcular atributos"/>										
Campos:										
#	Nuevo campo	Cálculo	Campo A	Campo B	Campo C	Tipo de valor	Longitud	Precisión	Eliminar	Conversion mask
1	fecha_id	Create a copy of field A	fecha_nombr			Date			N	yyyy-MM-dd
2	diames_id	Día del mes de fecha A	fecha_id			Integer			N	
3	mesn_id	Mes de fecha A	fecha_id			Integer			N	
4	anyo_id	Año de fecha A	fecha_id			Integer			N	

Figura 3-9 Paso Calcular atributos

El paso **Calcular Resto Atributos** utiliza dos funciones en java script para obtener el nombre del mes en la variable **mesn_desc** de cada fecha y el curso en la variable **curso** respectivamente.

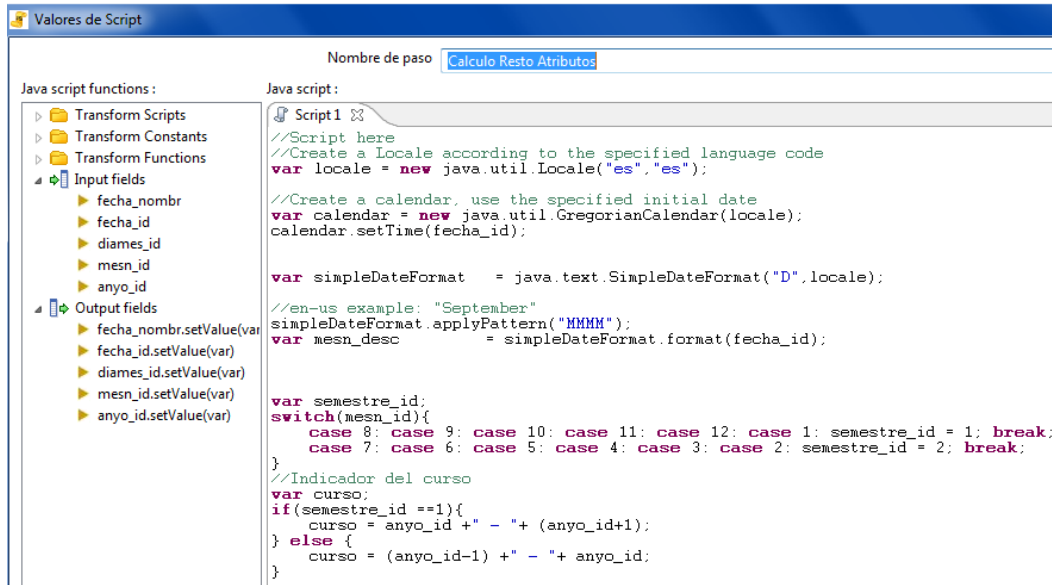


Figura 3-10 Paso Calcular Resto Atributos.

El paso **Más actual** es diseñado para crear un campo con la fecha máxima de expiración que puede tener una fila tanto en la dimensión persona como en la dimensión cargo, de esta manera utilizando como condición que la fecha de expiración sea igual a la fecha generada se obtiene la llave subrogada más actual.

El paso **Búsqueda en Base de Datos** es utilizado en múltiples ocasiones al realizar la carga de las tablas de hechos porque permite obtener las llaves subrogadas usadas en el MD que corresponden a las llaves naturales definidas y usadas en el sistema fuente. La configuración de este paso tomando como referencia el ejemplo específico de la dim_persona es la siguiente:

Conexión: Se especifica una conexión existente hacia el MD o se crea una nueva si así se requiere.

Tabla de búsqueda: Se selecciona la tabla de dimensión donde se realizará la búsqueda.

La tabla superior con el título: “La clave para realizar la búsqueda de valor” es usada para especificar como los campos que vienen en el flujo de datos se harán coincidir con las columnas de la tabla de búsqueda.

Campo de tabla: Corresponde con las columnas de la tabla de dimensión.

Campo1 y Campo 2: Son los nombres de las columnas en el flujo de datos con los que se harán las comparaciones.

En este caso además del ci como llave natural se compara usando el comparador (=) con una fecha de expiración como ya se había mencionado, para obtener el más actual.

En la tabla “Valores a devolver” se especifica el campo de la dimensión que se desea obtener de la búsqueda, siendo en este ejemplo el valor de la llave subrogada (bd_id_persona). El resto de las configuraciones aplicadas varían en las tablas objetivos y el nombre del campo final a devolver.

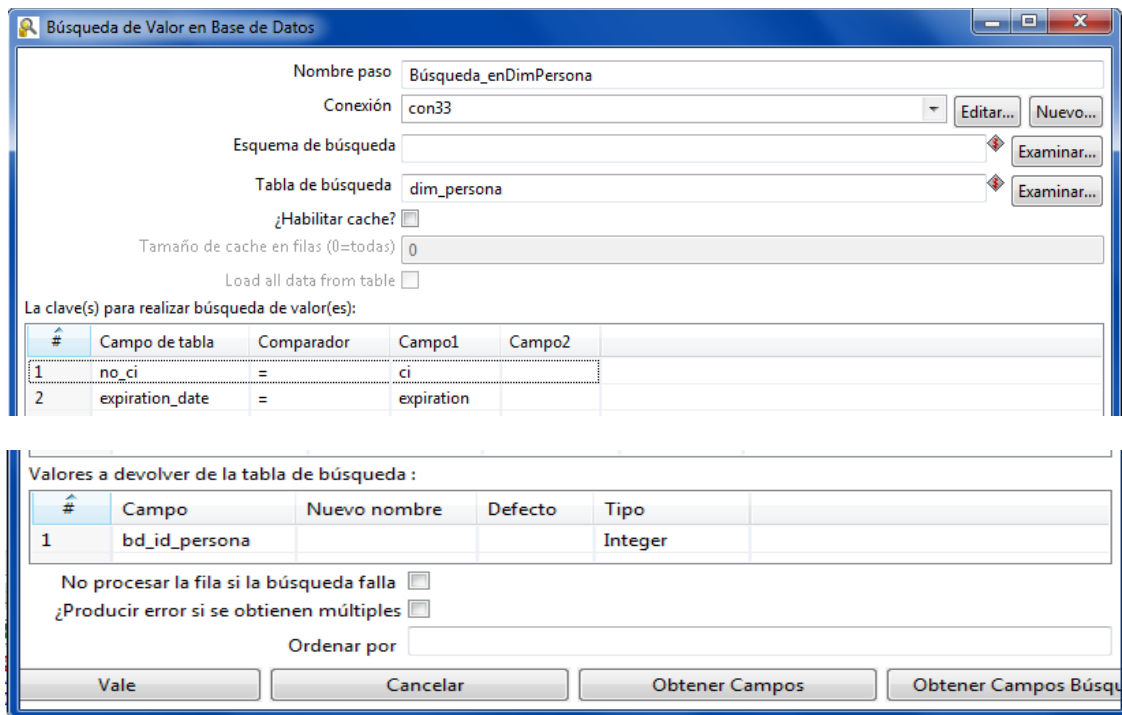


Figura 3-11 Configuración del paso Búsqueda en Base de Datos.

La transformación que llena la tabla de hechos fact_rel_cargo_cuadro (Figura 3-12) tiene como particularidad la necesidad de obtener la fecha en el momento de la carga

para lo cual se utiliza el paso **Información del Sistema** con la opción **fecha sistema (fijo)**. Esta fecha se carga hacia la dimensión fecha y a continuación se obtiene la llave subrogada con que fue insertada para su utilización en el llenado de la tabla de hechos.

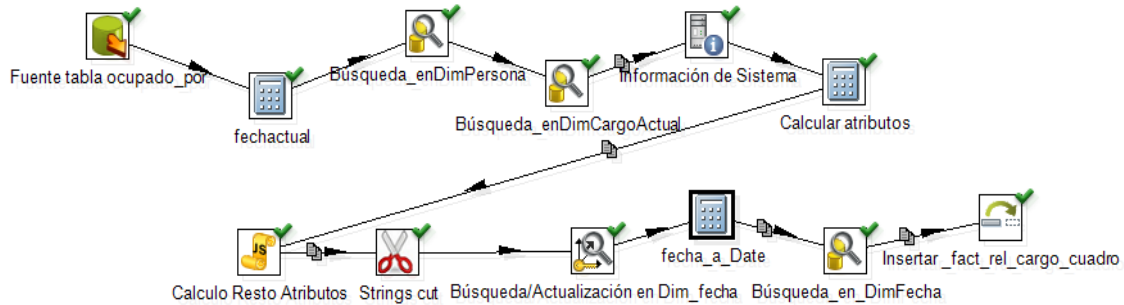


Figura 3-12 Transformación para la tabla de hechos fact_rel_cargo_cuadro.

Finalizado el diseño de las transformaciones para los procesos que se cargan mensualmente se orquestan en el siguiente trabajo:

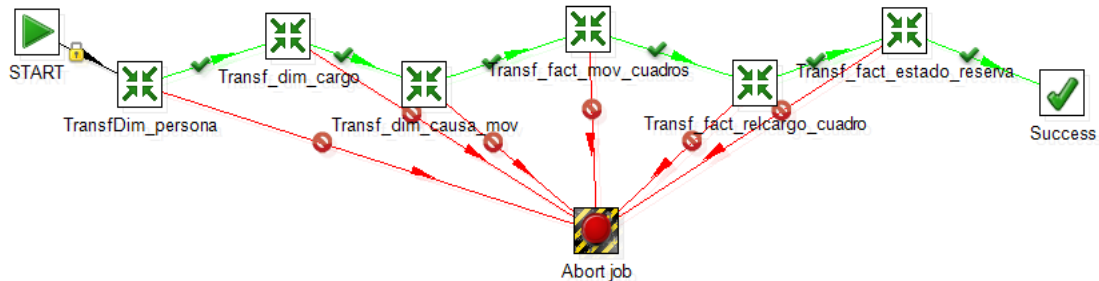


Figura 3-13 Diseño del Trabajo mensual.

El proceso de evaluación anual e integral fue diseñado con el uso de una tabla puente entre la dimensión indicador y cada tabla de hechos. Teniendo en cuenta la similitud de las tablas involucradas en estos procesos y siendo la periodicidad con que se cargan las tablas la principal diferencia se seleccionará el caso de la evaluación anual para explicar la transformación realizada.

El flujo general consiste en obtener la fecha en la que se cargan los datos para establecerla como fecha de evaluación, generar el identificador de grupo que se asociará a cada persona para insertarlo en la tabla puente y pasar al llenado de la tabla de hechos.

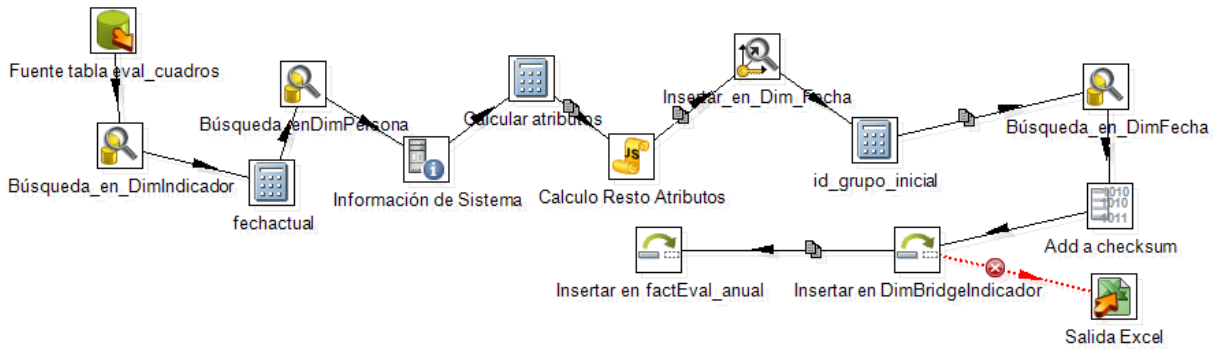


Figura 3-14 Transformación para la tabla de hechos fact_eval_anual.

La característica fundamental de esta transformación reside en que la tabla puente tiene un identificador de grupo que debe ser único para cada persona y para cada indicador de evaluación que le corresponda ser evaluado en un momento determinado, ante una nueva evaluación el grupo cambiará. Para garantizar un identificador de grupo único para cada persona se utilizó el paso **id_grupo_inicial** donde se crea una cadena que es la concatenación del año actual de la evaluación un valor constante igual a 1 si la evaluación es anual e igual a 2 si es integral y el carné de identidad de la persona.

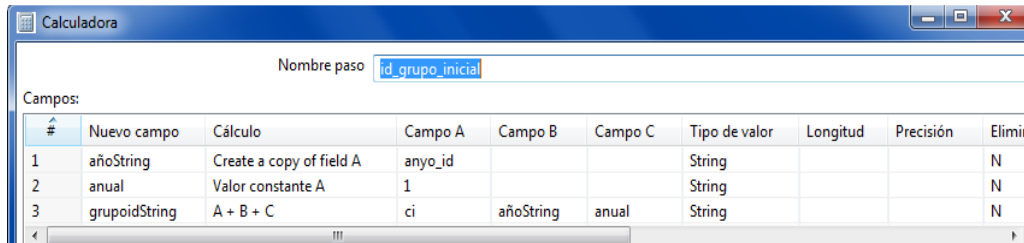


Figura 3-15 Creación del id_grupo para la tabla puente

Posteriormente con el paso **Add a checksum** se genera un valor numérico a partir de la cadena creada en el paso anterior garantizándose un valor numérico único para cada caso.

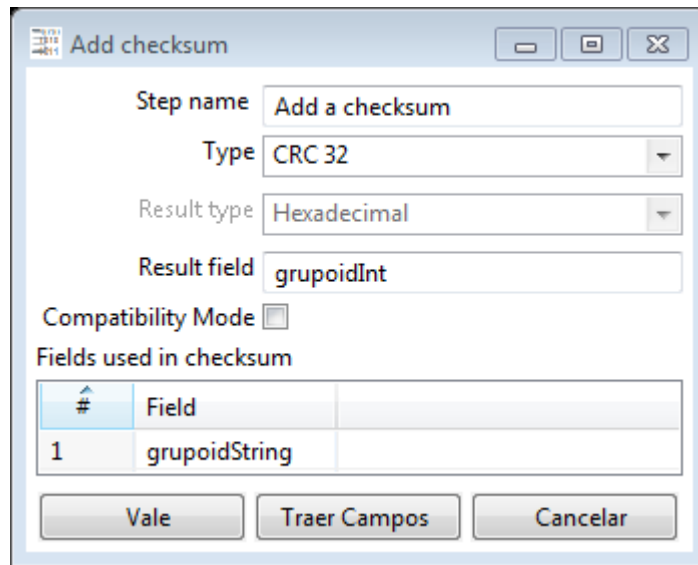


Figura 3-16 Utilización del paso Add a checksum.

3.3.2 Automatización del proceso de carga.

La herramienta Spoon es completamente dependiente de la interacción con el usuario y requiere de una interfaz gráfica, sin embargo lo deseable es que las transformaciones y trabajos se ejecuten de forma autónoma, con este propósito existen las herramientas de líneas de comando Kitchen y Pan (Matt Casters, 2010). Para el caso de estudio que nos ocupa se utilizará específicamente el Kitchen con el objetivo de ejecutar los trabajos que posibilitan el llenado de cada una de las tablas de hechos en los momentos que el usuario considere oportuno restringido por los períodos previamente definidos, ya sea mensual anual o cada dos años.

Para la iniciación del Kitchen primeramente ejecutamos el archivo Kitchen.bat. La sintaxis para especificar los parámetros consiste en un slash (/) o dash (-) seguido inmediatamente del nombre del parámetro **[/-] name [[:=] value]**. El valor del parámetro es especificado después de dos puntos (:) o el signo igual (=), puede ser encerrado en comillas simples o dobles opcionalmente pero si el valor del parámetro contiene espacios en blanco esto es obligatorio. Los parámetros genéricos pueden ser categorizados como aquellos que:

1. Especifican un trabajo o una transformación.

2. Control de logging.
3. Especifican un repositorio.
4. Listan los repositorios disponibles y su contenido.

NAME	VALUE	PURPOSE
Norep	Y	Don't connect to a repository. Useful to bypass automatic login.
Rep	Repository name	Connect to repository with the specified name.
User	Repository username	Connect to repository with the specified username.
Pass	Repository user password	Connect to repository with the specified password.
Listrep	Y	Show a list of available repositories.
Dir	Path	Specify the repository directory.
Listdir	Y	List the available repository job/repository directories.
File	Filename	Specify a job or transformation stored in a file.
Level	Error Nothing Basic Detailed Debug Rowlevel	Specify how much information should be logged.
Logfile	Filename for logging	Specify to which file you want to log. By default, the tools log to the standard output.
Version		Show the version, revision number, and build date of the tool.

Figura 3-17 Parámetros genéricos para el Kitchen y el Pan

Para automatizar el proceso de carga de los datos hacia el MD en el Departamento de Cuadros se creó una tarea básica en el programador de tareas de Windows 7 como se muestra en la figura siguiente:

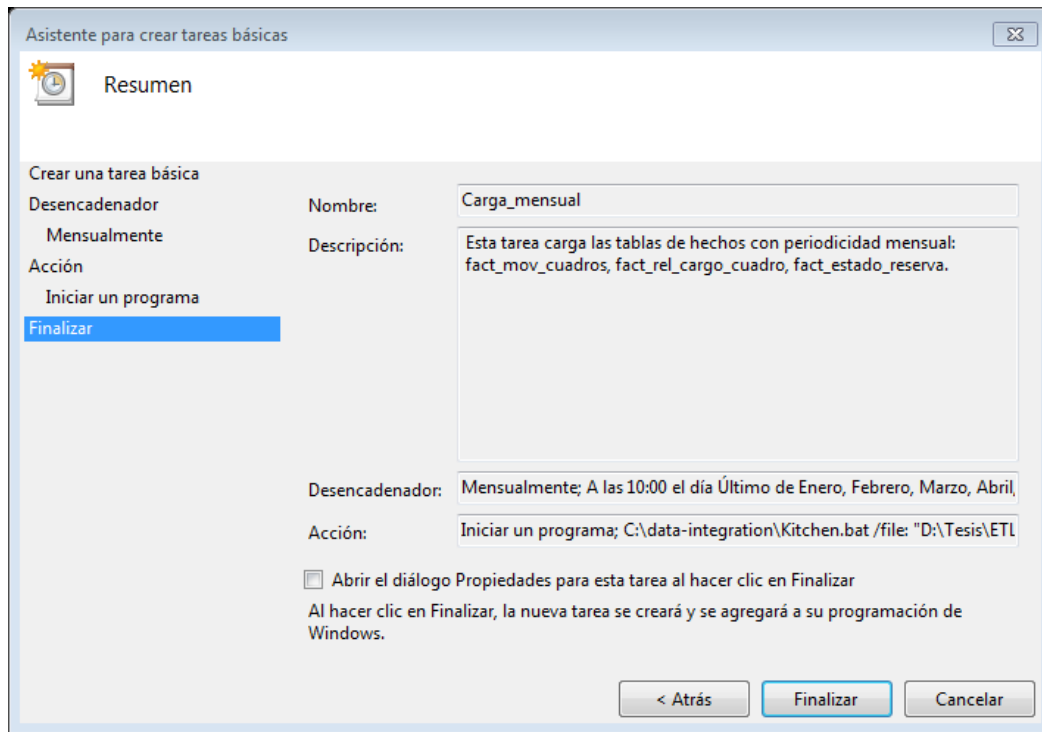


Figura 3-18 Tarea del sistema para la carga mensual.

Similarmente se diseñó una tarea que se desencadenará anualmente para la evaluación parcial e integral respectivamente.

3.4 Análisis de los datos a través de reportes y técnicas OLAP.

Los reportes ofrecen al usuario una panorámica abarcadora del proceso de negocio, responden a interrogantes y son indispensables para registrar y visualizar análisis y resultados.

Pentaho Report Designer es una herramienta gráfica que provee acceso a fuentes de datos relacionales, OLAP o basadas en XML, además de ofrecer varios formatos de salida como PDF, HTML, Excel o hasta texto plano. Después de diseñar los reportes estos se publican en la consola de usuario de Pentaho donde el usuario finalmente interactúa con ellos.

Para el Departamento de Cuadros se diseñaron reportes estáticos con parámetros que tienen como funcionalidad seleccionar o filtrar datos de campos para luego mostrar la información de acuerdo a esa selección.

Los reportes incluyen múltiples categorías tanto para cuadros como para las reservas entre las que se pueden citar:

- Sexo: Femenino, Masculino.
- Categoría Docente: Profesor titular, auxiliar, asistente, instructor.
- Categoría Científica: Investigador agregado, auxiliar, titular y aspirante a investigador.
- Título académico: Si es master o no.
- Grado científico: Si es doctor o no.
- Militancia: UJC, PCC.

En los reportes se representa:

- Tendencia histórica del comportamiento por categorías.
- Por ciento de incremento por categorías seleccionando las fechas de comparación.
- Por ciento de estabilidad en el cargo seleccionando la cantidad de años que se desea medir.
- Comportamiento por rango de edades.
- Representación por cuadro de evaluaciones y tendencia de desarrollo.

En la siguiente figura se muestra el resultado del reporte que mide la estabilidad en el cargo. El parámetro **tiempo** expresa la cantidad de años que ha sido estable. Los gráficos utilizados representan el por ciento de estabilidad por lugar y el por ciento con respecto al total de cuadros.

Dirección de cuadros MES Universidad Central "Marta Abreu" de Las Villas

Reporte de estabilidad en el cargo

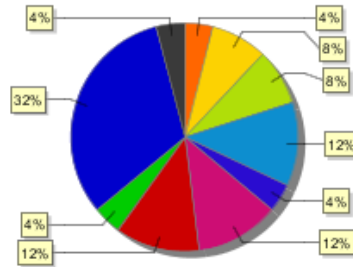
Fecha de confección: 2013-oct-25

Revisado por: _____

Cuadros con estabilidad en el cargo en los últimos 4 años

Nombre	Apellidos	Cargo	Lugar
Rafael Esteban	Bello Pérez	Director	Facultad De Matemática, Física Y Computación
Heriberto	Expósito Santana	Vicedecano Docente	Facultad De Construcciones
José Orestes	Guerra León	Jefe De Departamento	Facultad De Química Y Farmacia
Pedro Jesús	Iturria Quintero	Segundo Jefe De Departamento	Facultad De Química Y Farmacia
Luis Orlando	Maroto Martín	Jefe De Departamento	Facultad De Ciencias Agropecuarias
Jorge Luis	Barroso González	Vicedecano De Universalización	Facultad De Derecho

Porciento de estabilidad por lugar



Porciento de estabilidad en el cargo por 4 año(s)



Figura 3-19 Reporte de estabilidad en el cargo.

La versión 5.0.1 de Pentaho permite configurar los Datasources de manera que no es necesario publicar los cubos para el análisis OLAP como en versiones anteriores. A partir del esquema creado en Mondrian se puede crear un nuevo Datasource, solo se necesita el archivo XML asociado al esquema. Utilizando la herramienta Pentaho

Schema Workbench se diseñaron los cubos para el análisis de los datos, la figura siguiente muestra el diseño del cubo creado con la tabla de hechos fact_mov_cuadros.

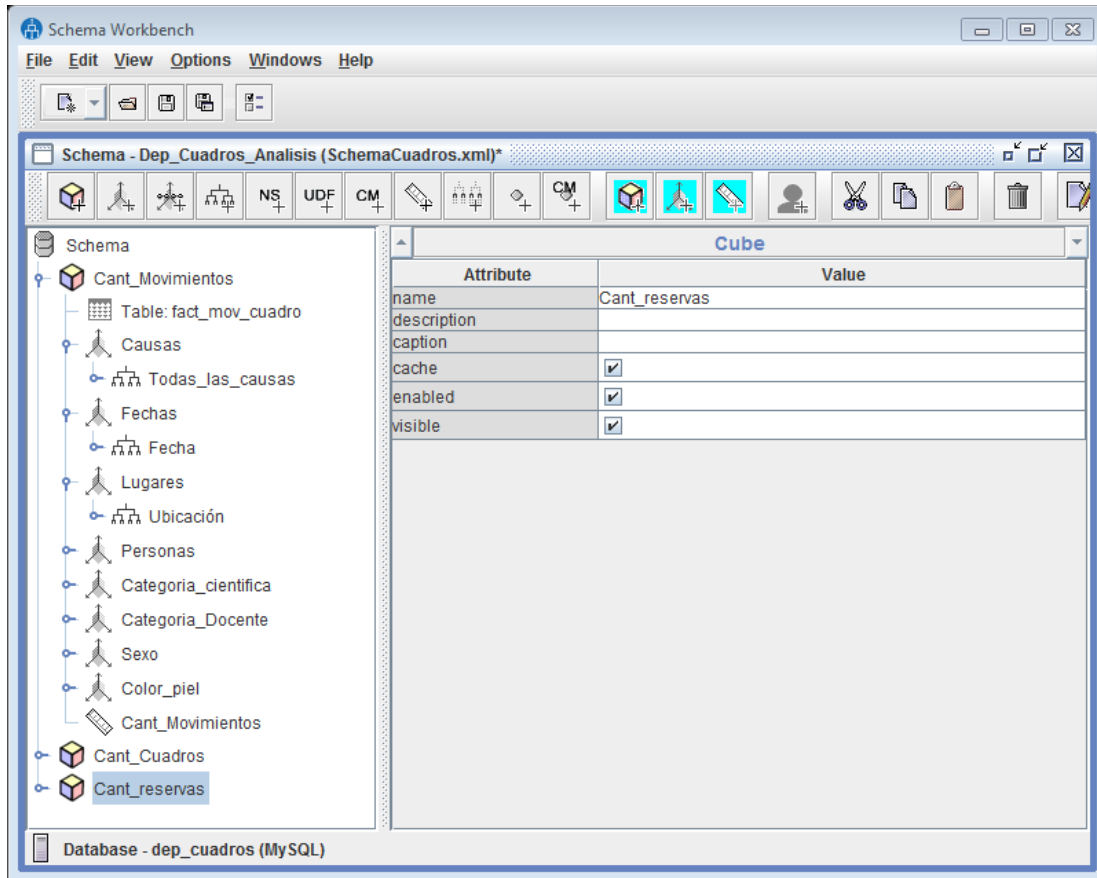


Figura 3-20 Diseño de cubos en el Schema Workbench

3.5 Conclusiones del capítulo

1. Como conclusión del capítulo se diseñó el MD para el Departamento de Cuadros de la UCLV usando la herramienta ERStudio 8.0 y se implementó sobre el SGBD MySQL.
2. Se implementaron los procesos ETL utilizando la herramienta Pentaho Data Integration teniendo en cuenta para su diseño el manejo de los cambios lentos en los atributos y la automatización del proceso de carga.
3. Se diseñaron reportes y cubos para permitir el análisis final de los datos a través de la consola de usuario de Pentaho 5.0.1.

CONCLUSIONES

Como resultado de este trabajo:

1. Se caracterizó el caso de estudio del Departamento de Cuadros de la UCLV identificando las fuentes de datos.
2. Se realizó un diseño dimensional que definió la estructura del MD y representó los requerimientos del negocio.
3. Utilizando Pentaho Data Integration 4.0.1 se definieron e implementaron los procesos ETL para una carga inicial y el posterior mantenimiento del MD.
4. Se diseñaron los reportes y los cubos para facilitar el análisis de la información usando la interfaz de Pentaho.

RECOMENDACIONES

Como recomendación se sugiere:

1. Diseñar los metadatos para la creación de reportes dinámicos.
2. Integrar el MD del Departamento de Cuadros con otros que se implementen en la UCLV con el objetivo de construir un DWH en la universidad.

REFERENCIAS BIBLIOGRÁFICAS

- ADAMSON, C. 2006. *Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance*, Wiley Publishing, Inc.
- ALTER, S. L. 1980. *Decision support systems: current practice and continuing challenges.*, Addison-Wesley
- BECKER, B. 2003. Design Tip #50: Factless Fact Tables? Sounds Like Jumbo Shrimp? Available: <http://www.kimballgroup.com/2003/10/16/design-tip-50-factless-fact-tables-sounds-like-jumbo-shrimp/>.
- BECKER, B. 2004. Design Tip #61: Handling All The Dates. Available: <http://www.kimballgroup.com/2004/10/28/design-tip-61-handling-all-the-dates/>.
- BECKER, B. octubre 6, 2010. Design Tip #128 Selecting Default Values for Nulls. *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2010/10/06/design-tip-128-selecting-default-values-for-nulls/>.
- BOLADO, S. R. 2012. *Desarrollo de un cliente web para la exploración y visualización de cubos OLAP*. Ingeniero en Informática, Universidad de Cantabria.
- CLAUDIA IMHOFF, N. G. Y. J. G. G. 2003. *Mastering Data Warehouse Design Relational and Dimensional Techniques*.
- DONGEN, R. B. J. V. 2009. *Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and MySQL.*, Wiley Publishing, Inc.
- INMON, W. H. 2005. *Building the Data Warehouse, Fourth Edition*.
- JOY MUNDY, W. T. Y. R. K. 2006. *The Microsoft Data Warehouse Toolkit : With SQL Server 2005 and the Microsoft Business Intelligence Toolset*, John Wiley & Sons
- KIMBALL, R. 1995. Is ER Modeling Hazardous to DSS.
- MATT CASTERS, R. B., JOS VAN DONGEN 2010. *Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration.*, Wiley Publishing, Inc.
- PONNIAH, P. 2001a. *DATA WAREHOUSING FUNDAMENTALS: A Comprehensive Guide for IT Professionals*
- PONNIAH, P. 2001b. *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals.*, John Wiley & Sons, Inc.

- PONNIAH, P. 2003. Database Design and Development an Essential Guide for IT Professionals.
- RALPH KIMBALL, J. C. 2004. *The Data Warehouse ETL Toolkit. Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*, Wiley Publishing, Inc.
- RALPH KIMBALL, L. R., MARGY ROSS Y WARREN THORNTHWAITE 2002a. *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data warehouses*.
- RALPH KIMBALL, M. R. 2002b. *The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling*, John Wiley & Sons, Inc.
- ROSS, M. 2005. Slowly Changing Dimensions are Not Always as Easy as 1, 2, 3. *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2005/03/10/slowly-changing-dimensions-are-not-always-as-easy-as-1-2-3/>.
- ROSS, M. 2013. Design Tip #152 Slowly Changing Dimension Types 0, 4, 5, 6 and 7. *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2013/02/05/design-tip-152-slowly-changing-dimension-types-0-4-5-6-7/>.
- ROSS, M. August 7, 2003. Design Tip #48: De-Clutter With Junk (Dimensions) *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2003/08/07/design-tip-48-de-clutter-with-junk-dimensions-by-margy/>.
- ROSS, M. Septiembre 3, 2008. Design Tip #105 Snowflakes, Outriggers, and Bridges. *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2008/09/03/design-tip-105-snowflakes-outriggers-and-bridges/>.
- THORNTHWAITE, W. diciembre 7, 2004. Design Tip #62: Alternate Hierarchies. *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2004/12/07/design-tip-62-alternate-hierarchies/>.
- THORNTHWAITE, W. febrero 6, 2003. Design Tip #43: Dealing With Nulls In The Dimensional Model. *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2003/02/06/design-tip-43-dealing-with-nulls-in-the-dimensional-model/>.
- THORNTHWAITE, W. junio 1, 2006. Design Tip #80 Adding a Row Change Reason Attribute. *Kimball Group* [Online]. Available: <http://www.kimballgroup.com/2006/06/01/design-tip-80-adding-a-row-change-reason-attribute/>.
- TRUJILLO, A. C. 2005. Modelo Multidimensional.