

Universidad de las Ciencias Informáticas

Facultad 6



**Título: Mercado de datos Comedores para la Sala Situacional de la
Universidad de las Ciencias Informáticas.**

*Trabajo de Diploma para optar por el título de Ingeniero en
Ciencias Informáticas*

Autores:

Ana Margarita Jorrín Pacheco.

Yuneidy Mercedes Alonso Duque.

Tutor: Lic. Lisdan Rodríguez Pérez.

Co-tutor: Ing. Leonel Pérez Nieblas

La Habana, Junio 2012

“Año 54 de la Revolución”



*"Lo que hace crecer el mundo no es el descubrir cómo está hecho,
sino el esfuerzo de cada uno por descubrirlo"*

José Martí

Declaración de autoría

Declaración de Autoría

Declaramos ser autoras del presente trabajo de diplomas y reconocemos a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste se firma la presente a los ____ días del mes de _____ del año _____.

Ana Margarita Jorrín Pacheco

Yuneidy Mercedes Alonso
Duque

Firma del Autor

Firma del Autor

Lic. Lisdan Rodríguez
Pérez

Firma del Tutor

Datos de contacto

Tutor: Lic. Lisdan Rodríguez Pérez.

Especialidad de graduación: Licenciatura en Ciencias de la Computación.

Categoría docente: Instructor.

Años de graduado: 4

Correo Electrónico: lisdanrp@uci.cu

Institución a la que pertenece: Universidad de las Ciencias Informáticas (UCI).

Dirección: Carretera San Antonio de los Baños, Torrens, Municipio Boyeros, Ciudad de La Habana, Cuba, Código postal 19370.

Co-tutor: Ing. Leonel Pérez Nieblas.

Especialidad de graduación: Ingeniero en Ciencias Informáticas.

Categoría docente: Recién graduado en Adiestramiento.

Años de graduado: 1

Correo Electrónico: lnieblas@uci.cu

Institución a la que pertenece: Universidad de las Ciencias Informáticas (UCI).

Dirección: Carretera San Antonio de los Baños, Torrens, Municipio Boyeros, Ciudad de La Habana, Cuba, Código postal 19370.

Primeramente agradecer a mi dúo de tesis, por su dedicación, por estar siempre a mi lado y batallar juntas en todo momento para lograr este sueño nuestro. A las personas más importantes de mi vida, a mi familia a ellas va dedicada esta tesis:

A mi mamá por ser mi mejor amiga, mi ejemplo de ser humano, por darme siempre tanto amor, tanto apoyo y tanta alegría. A mi papá por quererme tanto, por su comprensión y su apoyo incondicional.

A mi pasi por ser el mejor hermano del mundo, por apoyarme y complacerme en todos mis gustos, por cuidarme y quererme tanto. A milo que ya es parte de la familia. A mi tía Pilar y a Tony por su gran ayuda, su preocupación y cariño. A mima Elvira, Tata y pipo Benito por estar siempre pendientes de mí y haberme ayudado a convertirme en quién soy.

A mi novio Osmar por todo su amor, por ser mi apoyo aquí en la UCI por estar siempre ahí cuando lo necesité, por aguantar todas mis malcriadeces, por cuidarme y hacerme sentir que soy lo más importante del mundo.

A las mejores amigas de la historia a mis feas Tuta, Yuyu y la nerra a ustedes también va dedicada esta tesis por estar siempre conmigo en los buenos y malos momentos desde que éramos unas pulgas y nunca fallarme, por su amor y comprensión.

A mi mejor amigo Silvio mi gordito preferido, a ti moll muchas gracias por ser mi amigo y hermanito desde 1er año, gracias por cuidarme y por tantos momentos alegres que pasamos juntos. A mis otros dos mejores amigos Raul y Hector y no se pongan celosos de Silvio que yo también los quiero muchos ustedes lo saben, gracias por existir en mi vida.

A mis amistades de aquí de la UCI que ya se han convertido en parte de mi familia, a Marla, Juliet, Ana Rosa, Claudia, Nelly por tantos momentos lindos que pasamos juntas no las voy a olvidar nunca.

A los muchachos de mi grupo, a Yariel, Radame, Ale, Yan, Ramiro, Janier, Mailen, Cebo, Dalie, Ismel, Olivia, Karen.

A mi tutor por su ayuda y constancia. A Yoan por habernos sacado de una gran apuro y por nunca decirnos que no cuando lo necesitábamos a él muchas gracias. A Elio nuestro oponente por ayudarnos y aclararnos todas nuestras dudas.

A Leonel por estar siempre cuando lo necesitábamos, por su ayuda incondicional, por enseñarnos tanto y no ponerse bravo por tanta jodedera. A Esley por su ayuda incondicional, por ser nuestro profe y amigo. A todos muchas gracias.

Margarita

Quiero agradecer

A todas aquellas personas que de una forma u otra formaron parte de mi vida en estos cinco años, en especial a mis amigos Yulio, Marisbelis, Julito, Liset, Yusliel, Miguel, Ariel, Soto, Lianet, Yohana, Yeilen y Margarita mi dúo de tesis; ustedes se han convertido en parte de mi familia, gracias a todos por estar siempre presentes y compartir momentos de alegría que no olvidaré jamás.

A mis padres Milagro y Roberto. Mítica y Pitico no existen palabras en el mundo con las que pueda agradecer todo lo que han hecho para convertirme en la persona que soy, gracias por enseñarme el concepto de familia, a ser amable, respetuosa y a perseguir mis propias metas. Gracias por tener fe y confiar en mí.

A una de las personas más importantes en mi vida, mi hermano Yunier. Tati gracias por enseñarme y educarme; siempre has sido mi guía, mi ejemplo e inspiración.

A mi abuelita Gladys, tan quisquillosa y preocupándose siempre por mí. Abu agradezco que hayas contribuido en mi educación, sé que si estuvieras hoy aquí estuvieras muy contenta por mí.

A mis primos, a mis tíos, por estar siempre pendientes de mí, a mi tía Esther, gracias por ser como una madre, por apoyarme, aconsejarme y complacerme en todos mis antojos.

A Leonel, Maikel, Esley y Yusdenis, gracias a los cuatro por ser compañeros y amigos, por apoyarnos y estar dispuestos en todo momento a ayudarnos para obtener siempre los mejores resultados en el desarrollo de la presente investigación.

A Yoan por habernos ayudado siempre que lo necesitábamos.

A Elio nuestro oponente por enseñarnos y aclararnos todas nuestras dudas.

A mi tutor Lindan por su apoyo y confianza.

A todas aquellas personas que han contribuido en el éxito de esta investigación. A todos muchas gracias.

Dedico el presente trabajo a toda mi familia principalmente a mis padres, abuela y hermano por ser el motor inspirador de mis logros como persona y estudiante.

Yuneidy.

Resumen

La necesidad de tomar decisiones rápidas y certeras en un mundo que está en continua transformación, puede llegar a ser un proceso complejo por el gran cúmulo de información que se maneja en las empresas. Debido a esto se requiere de tecnologías modernas para el almacenamiento y procesamiento de la información. Los mercados de datos son una tecnología de almacenamiento de datos que tienen como objetivo fundamental brindar soporte a la toma de decisiones dentro de un área específica de la organización. La presente investigación tiene como objetivo general desarrollar un mercado de datos para el área Comedores de la Universidad de las Ciencias Informáticas que contribuya a la toma de decisiones. Esta área presenta deficiencias en la gestión diaria de sus recursos y en el procesamiento de la información, por lo que se hace necesario centralizar los indicadores que manejan los comedores para realizar análisis estadísticos que faciliten el proceso de toma de decisiones. Para ello se realizó el análisis, diseño, implementación y prueba al mercado de datos, utilizando varias herramientas entre ellas las que integran la Suite de Pentaho y la metodología de desarrollo "Propuesta de metodología para el desarrollo de almacenes de datos en DATEC". Una vez concluido el análisis, diseño e implementación se obtiene como resultado un mercado de datos poblado, así como la capa de visualización donde se muestran los reportes que incluyen el análisis de los principales indicadores del área antes mencionada.

Palabras claves: Comedores, mercado de datos, toma de decisiones.

Índice de Contenido

Introducción	1
Capítulo 1: Fundamento teórico de los almacenes de datos	4
1.1 Comedores	4
1.2 Aplicaciones existentes para el área de los Comedores	4
1.3 Tecnologías de almacenamiento de datos	5
1.3.1 Base de datos relacional	5
1.3.2 Almacén de datos	5
1.3.3 Mercado de datos	7
1.4 Metodologías para el desarrollo de almacenes de datos	7
1.5 Modelo de datos	9
1.6 Herramientas de modelado	11
1.7 Sistemas Gestores de Bases de datos	12
1.8 Modos de almacenamiento de datos	13
1.9 Herramientas para el proceso ETL	14
1.10 Herramientas para Inteligencia de Negocio (BI)	15
Capítulo 2: Análisis y Diseño del Mercado de datos Comedores	18
2.1 Análisis.....	18
2.1.1 Tema de análisis.....	18
2.1.2 Definición del negocio	18
2.1.3 Reglas del negocio	18
2.1.4 Necesidades de los usuarios	20
2.1.5 Requerimientos.....	20
2.1.6 Actores del sistema	23
2.1.7 Diagrama de casos de uso del sistema.....	24
2.2 Arquitectura del Mercado de datos	26
2.3 Diseño.....	28
2.3.1 Diseño del subsistema de almacenamiento	29
2.3.2 Diseño del subsistema de integración	35
2.3.3 Diseño del subsistema de visualización.....	37
2.4 Políticas de seguridad	38
2.4.1 Roles y permisos	38

2.4.2	Salva de la base de datos	39
Capítulo 3:	Implementación del Mercado de datos Comedores	40
3.1	Implementación del modelo de datos	40
3.2	Implementación de los procesos de Extracción, Transformación y Carga	40
3.3	Implementación del subsistema de visualización	45
Capítulo 4:	Pruebas al Mercado de datos Comedores	50
4.1	Pruebas de software.....	50
4.2	Herramientas para la aplicación de las pruebas	51
4.3	Resultados de las pruebas	55
	Conclusiones generales	58
	Recomendaciones.....	59
	Referencias Bibliográficas.....	60
	Bibliografía	62
	Glosario de términos	64
	Anexos	65

Índice de tablas

Tabla 1. Descripción de los actores del sistema.	23
Tabla 2. Descripción del caso de uso "Presentar información de los portadores energéticos".	25
Tabla 3. Dimensiones.	29
Tabla 4. Medidas.	29
Tabla 5. Hechos	32
Tabla 6. Reporte candidato " Consumo de electricidad por complejo comedor".	38
Tabla 7. Roles y permisos.	38
Tabla 8. Esquemas y tablas del mercado.	40
Tabla 9. Caso de prueba basado en el caso de uso Presentar información de los portadores energéticos.	51
Tabla 10. Lista de chequeo para el artefacto de ETL Diccionario de datos.	53

Índice de figuras

Figura 1. Diagrama de casos de uso del sistema.	24
Figura 2. Diseño de la arquitectura de un almacén de datos.	27
Figura 3. Diseño de los subsistemas.	28
Figura 4. Matriz BUS.	33
Figura 5. Modelo de datos.	34
Figura 6. Diseño de las transformaciones para la carga de las dimensiones estáticas.	35
Figura 7. Diseño de las transformaciones para la carga de las dimensiones lentamente cambiantes.	36
Figura 8. Diseño de las transformaciones para la carga de los hechos.	37
Figura 9. Diseño del mapa de navegación.	37
Figura 10. Transformación para la carga de la dimensión dim_complejo_comedor.	42
Figura 11. Transformación para la carga de la dimensión dim_evento.	43
Figura 12. Transformación para la carga del hecho hech_energia.	44
Figura 13. Transformación del trabajo principal.	44
Figura 14. Transformación del trabajo general.	45
Figura 15. Implementación de los cubos OLAP.	45
Figura 16. Arquitectura de información del Mercado de datos Comedores.	46
Figura 17. Consulta MDX del consumo de electricidad.	47
Figura 18. Reporte Consumo de electricidad por complejo comedor.	48
Figura 19. Gráfico dinámico del consumo de electricidad en Kw/h por año.	48
Figura 20. Modelo V.	50
Figura 21. Comportamiento de las no conformidades identificadas en las pruebas unitarias y de integración.	56
Figura 22. Comportamiento de los indicadores de la lista de chequeo.	57

Introducción

En la actualidad la información se ha convertido en uno de los recursos más valiosos y la materia prima indispensable para el desarrollo, equilibrio y adaptación de las empresas, sectores y países. Ha devenido como recurso estratégico determinante de la competitividad y un sustento a los procesos de toma de decisiones, siendo estos capaz de crear conocimientos. No basta con lograr la persistencia de los datos en el tiempo, pues se hace necesario su almacenamiento y procesamiento. En este sentido, se han dado avances significativos en los repositorios de información, los que han ido evolucionando desde un fichero de texto hasta grandes bases de datos.

En Cuba se han adquirido conocimientos informáticos que han permitido el desarrollo en diferentes sectores como el económico, el político y el social. En este campo la Universidad de las Ciencias Informáticas (por sus siglas en español UCI), desempeña un papel muy importante con la producción de *software* para la exportación e informatización de la sociedad cubana. La UCI cuenta con el Centro Tecnología de Gestión de Datos (por sus siglas en español DATEC), el cual tiene como objetivo proveer soluciones integrales para el análisis de la información. Sala Situacional UCI es un proyecto desarrollado por el centro DATEC con el fin de facilitar el análisis en las diferentes áreas de la universidad, dentro de ellas se encuentra Comedores, dicha área está concebida para la elaboración y distribución de los alimentos básicos hacia toda la comunidad universitaria.

El crecimiento de la comunidad universitaria con el pasar de los años, ha originado deficiencias en la gestión diaria de los recursos destinados a la producción de alimentos, causando que dicha producción en ocasiones esté por encima o por debajo de la demanda de los comensales en cada complejo comedor. Esta situación ocasiona afectaciones y molestias en el quehacer diario de los estudiantes, profesores y trabajadores, así como gastos innecesarios a la universidad y por ende al país. Estas deficiencias se deben en gran medida a las limitaciones existentes en cuanto a la planificación y estimación de los comensales. Además se hace difícil el proceso de toma de decisiones, pues los indicadores de interés para el área de los Comedores de la UCI se encuentran almacenados en diferentes fuentes.

Por lo antes planteado se identifica el siguiente **problema de la investigación**: ¿Cómo contribuir al proceso de toma de decisiones en el área Comedores para la Sala Situacional de la Universidad de las Ciencias Informáticas?

Definiendo como **objeto de estudio**: Los almacenes de datos, centrando su **campo de acción** en el Mercado de datos Comedores para la Sala Situacional de la Universidad de las Ciencias Informáticas.

Para dar respuesta al problema planteado se propone como **objetivo general**, Desarrollar el Mercado de datos Comedores para la Sala Situacional de la Universidad de las Ciencias Informáticas, que contribuya a la toma de decisiones.

Desglosándose en los siguientes **Objetivos Específicos**:

1. Realizar el marco teórico de la investigación
2. Realizar el análisis y diseño del Mercado de datos Comedores para la Sala Situacional de la Universidad de las Ciencias Informáticas.
3. Implementar el Mercado de datos Comedores para la Sala Situacional de la Universidad de las Ciencias Informáticas.
4. Probar el Mercado de datos Comedores para la Sala Situacional de la Universidad de las Ciencias Informáticas.

Para dar cumplimiento a los objetivos de la investigación se plantean las siguientes tareas:

1. Estudio de soluciones existentes en el área de los Comedores.
2. Caracterización y selección de la tecnología de almacenamiento de datos.
3. Caracterización y selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
4. Levantamiento de requisitos.
5. Descripción de los casos de uso del mercado de datos.
6. Definición de los hechos, las medidas y las dimensiones del mercado de datos.
7. Diseño del modelo de datos.
8. Definición de la arquitectura del mercado de datos.
9. Diseño del subsistema de integración.
10. Diseño del subsistema de visualización.
11. Implementación del subsistema de integración.
12. Implementación del subsistema de visualización.
13. Aplicación de las listas de chequeo.
14. Diseño de los casos de pruebas.
15. Aplicación de los casos de pruebas.

Como **resultado esperado** se obtendrá un mercado de datos poblado, así como la capa de visualización de los datos.

El presente Trabajo de Diploma se ha estructurado en introducción, cuatro capítulos, conclusiones generales, recomendaciones, referencias bibliográficas, bibliografía y anexos.

Capítulo 1: Fundamento teórico de los almacenes de datos

En el capítulo se definen conceptos fundamentales relacionados con las tecnologías de almacenamiento de datos, sus ventajas y desventajas, así como la selección de la metodología y herramientas que serán utilizadas para el desarrollo de la investigación.

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

En el capítulo se detallan las características del negocio y se identifican las necesidades de los usuarios, las reglas del negocio, los requisitos funcionales y no funcionales, así como los casos de uso del sistema. Se confecciona la matriz bus, el modelo de datos dimensional y el modelo físico. Se diseñan los tres subsistemas del Mercado de datos Comedores, además de que se establecen los roles y privilegios sobre la base de datos y la aplicación.

Capítulo 3: Implementación del Mercado de datos Comedores

En el capítulo se muestra todo lo referente a la implementación de los tres subsistemas del Mercado de datos Comedores con el objetivo de darle solución a las necesidades de los usuarios.

Capítulo 4: Pruebas al Mercado de datos Comedores

En el capítulo se le realizan pruebas unitarias, de integración y de sistema al Mercado de datos Comedores. Se muestran los resultados obtenidos en cada una de ellas luego de su aplicación. Dichas pruebas son realizadas para garantizar el cumplimiento de las exigencias del cliente.

Capítulo 1: Fundamento teórico de los almacenes de datos

El presente capítulo abordará el análisis de las tecnologías de almacenamiento de datos. Se hará un estudio del estado del arte tanto a nivel nacional como internacional de las aplicaciones existentes para el área de los Comedores. Se abordará sobre las metodologías y las principales herramientas para el desarrollo de los mercados de datos, así como la justificación de su uso.

1.1 Comedores

Los comedores a los que accede un gran número de clientes diariamente constituyen una de las áreas más importantes dentro de una entidad. Esta área debe contar con una infraestructura sólida que garantice su buen funcionamiento para satisfacer las necesidades de los clientes. En la UCI el área de los Comedores se encuentra ubicada dentro de la Dirección general de alimentos (Ver anexo 1), la cual da soporte a los servicios básicos de alimentación de toda la comunidad universitaria. Estos servicios traen implícitos una serie de procesos, entre ellos, la planificación y control del acceso de los comensales a los eventos del comedor (desayuno, almuerzo, comida y dobles respectivamente), así como la distribución de las raciones, la planificación del menú y el control económico de los ingresos que tiene esta área por la venta de tickets a trabajadores y profesores de cada uno de los tres complejos existentes.

1.2 Aplicaciones existentes para el área de los Comedores

En el mundo

La planificación y estimación de los comensales en el área de los Comedores, a la que acceden grandes cantidades de clientes, se hace cada vez más necesaria. En el mundo existen empresas que se dedican al desarrollo de aplicaciones que facilitan el trabajo y controlan el acceso de los clientes a esta área. Algunos ejemplos de estas soluciones son:

- **Sistema de Tratamiento y Análisis Nutricional** (por sus siglas en inglés SATN2007): está enfocado fundamentalmente al área de la salud. Este software cuenta con una comprensiva base de información y conocimiento de los alimentos. SATN2007 profesional proporciona análisis completos de dietas, recetas y menús. Su intuitiva interfaz de usuario y su funcionalidad ayudará en el trabajo y productividad, cualquiera que sea el área de la nutrición donde se mueva, medicina, servicio de comida o el campo educativo. [7]
- **EquiLibra Professional**, es un software que proporciona más de 1500 platos nacionales e internacionales con sus respectivos componentes nutricionales y la preparación culinaria. Dispone de un mecanismo de intercambios nutricionales utilizando todos los platos y alimentos del sistema,

lo que permite personalizar dietas al gusto y la disponibilidad de los alimentos. Tiene un acceso automático a búsquedas de alimentos y platos preparados de la base de datos para equilibrar en una forma fácil y automática su ingesta y de esta manera, poder suplementar el déficit a través de los alimentos adecuados. Además maneja y controla de forma histórica un número ilimitado de usuarios o grupos de usuarios, entre otras funcionalidades. [8]

En Cuba

En Cuba la Universidad de las Ciencias Informáticas cuenta con el *software* **Sistema control de acceso a los comedores en la Universidad de las Ciencias Informática**. Este sistema tiene como objetivo fundamental tener un control de los clientes que acceden a los servicios del comedor. Consiste en registrar el acceso a los servicios del comedor de cada comensal, comprobando que el usuario existe en la base de datos y tiene autorización para pasar por esa puerta. [9]

A pesar de que las soluciones mencionadas con anterioridad permiten analizar, controlar y planificar indicadores referentes al menú y accesos de clientes en el área de los Comedores, no posibilitan realizar reportes estadísticos de estos indicadores, ni análisis de datos almacenados. Por estas razones ninguno de los sistemas estudiados es factible para su empleo como solución a la situación existente. Por tanto se hace necesario desarrollar una solución que responda a todas las necesidades del área Comedores de la UCI, para ello se debe seleccionar tanto la tecnología, como las herramientas y una metodología que sirva de guía para el desarrollo del *software*.

1.3 Tecnologías de almacenamiento de datos

Las tecnologías de almacenamiento de datos se han convertido en herramientas fundamentales para el control y manejo de operaciones dentro de una empresa. Estas tecnologías facilitan la organización y centralización de grandes cúmulos de datos que pueden ser procesados para la obtención de información. A continuación se realizará un estudio de algunas de estas tecnologías.

1.3.1 Base de datos relacional

Las bases de datos relacionales de procesamiento de transacciones en línea (por sus siglas en inglés OLTP) involucran operaciones de inserción, modificación y borrado de datos. Presentan un modelo entidad-relación donde sus tablas se encuentran normalizadas, es decir en tercera forma normal. Estas permiten el acceso simultáneo de muchos usuarios que agregan y modifican datos con regularidad, no están diseñadas para almacenar información histórica por lo que suelen limitarse a los datos actuales o recientes.

1.3.2 Almacén de datos

Un almacén de datos es una gran colección de datos que recoge información de múltiples sistemas fuentes, con el propósito de responder preguntas de negocios y ayudar a la toma de decisiones. El mismo contiene datos históricos derivados de datos transaccionales.

Existen muchas definiciones para los almacenes de datos dadas por grandes personalidades, entre ellas **William Inmon y Ralph Kimball**, conocidas por sus grandes aportes al mundo de los almacenes de datos:

Según **William Inmon**: [1]

"Un almacén de datos es una colección de datos orientado a temas, integrados, no-volátiles y variantes en el tiempo, organizados para soportar necesidades empresariales".

- **Orientado a temas**: los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Variantes en el tiempo**: los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No volátil**: la información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de solo lectura y se mantiene para futuras consultas.
- **Integrado**: la base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

Según **Ralph Kimball**: [2]

"Un almacén de datos es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis, es la unión de todos los mercados de datos de una entidad"

Objetivos de un almacén de datos: [3]

- Hacer accesible la información de la organización: la información contenida en el almacén de datos debe ser navegable, fácilmente comprendida por los usuarios, y sobre todo de acceso rápido.
- Ser una fuente adaptable de información: el almacén de datos está diseñado para afrontar con éxito continuos cambios. Cuando surgen nuevas necesidades de información, nuevas preguntas o nuevos datos añadidos, las tecnologías y los datos existentes no deben verse afectadas.
- Ayudar a la toma de decisiones: los datos contenidos en el almacén de datos son adecuados para justificar decisiones estratégicas de la organización. Las decisiones se toman una vez que el almacén de datos ha aportado los datos que las justifican.

Ventajas del uso de un almacén de datos:

- Hacen más fácil el acceso a una gran variedad de datos.
- Se obtiene una base de datos clasificada por temas de análisis.
- Se obtiene una base de datos histórica.
- Integración de información procedente de múltiples sistemas externos.

Desventajas del uso de un almacén de datos: [4]

- Problemas con los sistemas de origen de los datos.

1.3.3 Mercado de datos

Un mercado de datos es un subconjunto de un almacén de datos, especializado en el almacenamiento de los datos de un área de negocio específica dentro de una organización, mientras que un almacén de datos analiza la información a nivel corporativo. El mercado se caracteriza por disponer de la estructura de datos necesaria para analizar la información al detalle desde todas las perspectivas que afecte a los procesos del departamento al cual está aplicado. [5]

Ventajas del uso de un mercado de datos: [6]

- Acelera las consultas al reducir la cantidad de datos a recorrer.
- Permite el acceso a los datos mediante un gran número de herramientas del mercado, logrando así la independencia de estas.
- Centraliza y transforma la información de manera homogénea.

Se selecciona como tecnología de almacenamiento un mercado de datos debido a que los sistemas transaccionales de almacenamiento presentan estructuras normalizadas, haciendo más lento el tiempo de respuesta de las consultas. Los sistemas transaccionales no almacenan información histórica, por tanto no son la mejor alternativa para el apoyo a la toma de decisiones. Además están concebidos para realizar operaciones de inserción, actualización y algunas selecciones de datos, pero es menos probable que este tipo de sistema se organice de forma tal que produzca reportes para datos resumidos con cierta jerarquía. Por tanto para dar solución a esto se podría utilizar un mercado de datos, debido a que están enfocados a un área de negocio en específico y enmarcado por departamentos. Esta sería la solución más factible para resolver los problemas que presenta el área de los Comedores, permitiendo profundizar en el análisis de la información y ver su variación en el tiempo.

1.4 Metodologías para el desarrollo de almacenes de datos

Una metodología es una guía que indica qué hacer y cómo actuar para realizar las acciones propias de una investigación. Para guiar todo el ciclo de vida de desarrollo de un almacén de datos existen

diferentes metodologías entre las que se destacan la metodología de Inmon y la metodología de Kimball, dos de las personalidades más influyentes en esta área.

Entre estas dos metodologías la principal diferencia está en el enfoque que cada uno le da. **Inmon** se basa en un enfoque descendente, propone construir primero el almacén de datos y a partir de este los mercados de datos, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente e histórica.

Muy diferente de Inmon, la propuesta de **Kimball** se basa en dividir el mundo de Inteligencia de Negocio en hechos y dimensiones, el mismo considera que el almacén de datos tiene que ser entendido fácilmente por los usuarios y ofrecer respuestas correctas en el menor tiempo posible. [2] Una de sus características principales es que su arquitectura es ascendente partiendo de los requisitos del negocio, plantea que se debe crear por cada departamento un conjunto de mercados de datos independientes orientados a los temas que estén relacionados con él.

Propuesta de metodología para el desarrollo de almacenes de datos en DATEC, es presentada por un conjunto de autores, entre ellos Leopoldo Zenaido Zepeda Sánchez, los que proponen como aspecto novedoso la incorporación de los casos de uso a la etapa de levantamiento de requisitos para guiar el proceso de desarrollo y una etapa de prueba a la metodología de Kimball. [10]

Después de realizar un estudio de las diferentes metodologías, se decide utilizar la Propuesta de metodología para el desarrollo de almacenes de datos en DATEC para guiar el proceso de desarrollo del Mercado de datos Comedores. Esta metodología se basa en el ciclo de vida Kimball, pues la misma crea los conceptos de hechos y dimensiones, lo que proporciona mayor agilidad en el proceso de toma de decisiones. Propone ir construyendo el almacén de datos a través de la construcción de los mercados de datos departamentales, lo que coincide con la división lógica de las empresas y entidades. Es una metodología madura, con abundante documentación sobre la misma y es reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles. Agrega los casos de uso a la etapa de levantamiento de requisitos y una etapa de prueba para estar en correspondencia con las características de la universidad y del centro. [10]

Los flujos de trabajo que existen dentro de esta metodología son: [10]

- **Estudio Preliminar o Planeación:** se realiza el estudio de la entidad cliente, así como la planeación, se definen los objetivos, el alcance y costos estimados.
- **Requerimientos:** se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar.

- **Arquitectura y Diseño:** participa el módulo de Extracción transformación y carga (por sus siglas en inglés ETL), el módulo de Almacenamiento de datos (por sus siglas en inglés AD) y el módulo de Inteligencia de Negocio (por sus siglas en inglés BI) realizando tareas como el diseño de las reglas de ETL, y en la definición de la arquitectura de información que regirá el desarrollo de la solución, participan los arquitectos de cada uno de los módulos anteriormente mencionados.
- **Implementación:** se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de BI.
- **Prueba:** se realizan las pruebas al sistema desde las pruebas de unidad llevadas a cabo por los desarrolladores de cada uno de los módulos (ETL, AD, BI) hasta las de aceptación con el cliente final.
- **Despliegue:** se demuestra que el sistema funciona a través de un despliegue piloto donde se configuran los servidores, se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona.
- **Soporte y Mantenimiento:** pueden brindarse los servicios de soporte en línea, vía telefónica, web u otras.
- **Gestión y Administración del Proyecto:** se realizan actividades relacionadas con la gestión de proyecto como son los gastos, las utilidades, los recursos y demás actividades.

Al ser seleccionada la metodología y las fases: requerimientos, arquitectura y diseño, implementación y prueba para el desarrollo del mercado de datos, es necesario definir el modelo de datos que se adapte a la tecnología de almacenamiento seleccionada.

1.5 Modelo de datos

Un modelo de datos es un conjunto de conceptos que pueden servir para describir la estructura de una base de datos, esto se refiere a tipos de datos, sus vínculos y las restricciones que deben cumplir los datos. [11]

Al ser seleccionado un mercado de datos como tecnología de almacenamiento se escoge el modelo de datos dimensional para representar la estructura que tendrán los datos en la base de datos, haciendo uso de los conceptos de hechos y dimensiones. Además este modelo permite que las tablas de la base de datos estén desnormalizadas, lo que agiliza el tiempo de respuesta de las consultas.

Modelo dimensional

A diferencia de los clásicos sistemas de bases de datos que presentan sus estructuras diseñadas mediante el modelo entidad-relación los mercados de datos se diseñan mediante un modelo dimensional.

El modelo dimensional divide el mundo de los datos en dos conjuntos: las tablas de hecho que almacenan las mediciones numéricas del negocio (medidas) y las tablas de dimensiones que almacenan las descripciones del entorno de estas medidas (dimensión). Las dimensiones son denominadas **dimensiones estáticas** que son aquellas cuya información no es propensa a cambios con el tiempo o **dimensiones lentamente cambiantes** (por sus siglas en inglés SCD). Estas últimas contienen datos que pueden variar con el tiempo y que, en dependencia de su tipo, permiten o no mantener un registro histórico de los valores asociados a un identificador del área temporal en el mercado [2].

Los tipos básicos de SCD son: [2]

Tipo 1. Sobrescribir: se sobrescriben los valores de la dimensión con los valores nuevos que vienen del área temporal. Este tipo es el más básico y no permite mantener una traza de los cambios ocurridos en los datos ya que siempre se sobrescriben los valores.

Tipo 2. Añadir fila: se añaden nuevas filas a la dimensión con los valores nuevos que vienen del área temporal. Este tipo permite mantener una traza de los cambios ocurridos en los datos ya que siempre se añaden nuevas filas con los valores nuevos.

Tipo 3. Añadir columna: este tipo requiere de una columna para el valor actual y otra para el valor anterior. Los valores existentes se almacenan en la columna valor_anterior y los valores nuevos provenientes del área temporal se almacenan en la columna valor_actual. Este tipo permite mantener una cantidad limitada de los valores anteriores a los valores actuales. En caso de que sea necesario guardar más de uno de los valores anteriores se debe crear una columna para cada uno de los valores anteriores que se desee mantener.

Tablas de Hechos: es la tabla primaria en el modelo dimensional, las mismas poseen una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas con este. [6]

Tablas de Dimensiones: las tablas de dimensiones definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Representan los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos. Cada tabla de dimensión posee un identificador único y al menos un campo de referencia

que describe los criterios de análisis relevantes para la organización. Estos son por lo general de tipo texto. [6]

Cubo multidimensional: un cubo es la representación de un hecho que agrega las medidas mediante los niveles y jerarquías de cada una de las dimensiones que se desee analizar.

Cubo virtual: es la unión de dos o más hechos para realizar cálculos con las medidas asociadas a estos.

Granularidad: representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando. [6]

Jerarquías: una jerarquía representa una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación “padre-hijo”. La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles. [6]

Para la materialización física del modelo dimensional es utilizada comúnmente la propuesta realizada por Ralph Kimball llamada “**esquema estrella**” que consta de una tabla de hecho central y de varias tablas de dimensiones que son las relacionadas con las tablas de hechos, las cuales deben estar totalmente desnormalizadas, es decir, que no pueden presentarse en tercera forma normal. Existen otras estructuras que surgen por modificaciones realizadas al esquema de estrella, ellas son:

Esquema en copo de nieves: representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones. Es el más cercano a un modelo de entidad relación, debido a que sus tablas de dimensiones están normalizadas.

Constelación de hechos: está compuesto por una serie de esquemas en estrella, formado por una tabla de hechos principal y una o más tablas de hechos auxiliares. Estas yacen en el centro del modelo y están relacionadas con una o más dimensiones.

Como esquema arquitectónico del modelo dimensional se escogió el esquema constelación de hechos, debido a que en la presente investigación se modelarán varios hechos que se relacionan con una o más dimensiones.

1.6 Herramientas de modelado

Las herramientas para el modelado de aplicaciones informáticas son usadas por muchas empresas para automatizar aspectos claves como son el diseño, construcción e implementación de todo el

proceso de desarrollo de un sistema. Además, son de gran ayuda a lo largo de todo el proceso de desarrollo de un *software*. A continuación se abordarán algunas de estas herramientas:

Rational Rose: es una herramienta de modelado para el desarrollo de aplicaciones informáticas y tiene como base el Lenguaje unificado de modelado (por sus siglas en inglés UML). Dispone de varios diagramas entre ellos diagramas de clases, de objetos y de casos de uso del negocio. Permite generar documentación automáticamente y posibilita realizar ingeniería inversa (crear modelos a partir de códigos). No posee un entorno gráfico amigable para el usuario, además de que no es un software libre. [12]

Visual Paradigm for UML: es una herramienta que soporta todos los diagramas UML necesarios para el desarrollo de diferentes metodologías, esquemas y diagramas de entidad-relación. Reduce la duración del ciclo de desarrollo de *software* brindando ayuda a los arquitectos, diseñadores y desarrolladores. Es una herramienta multiplataforma que permite la ingeniería inversa, generar códigos desde diagramas y documentación. [13]

Ventajas Visual Paradigm for UML: [13]

- Generación de bases de datos: transformación de diagramas de entidad-relación en tablas de base de datos.
- Ingeniería inversa y directa de bases de datos: desde Sistemas Gestores de Bases de Datos (por sus siglas en inglés DBMS) a diagramas de entidad-relación y viceversa.
- Posee un entorno gráfico amigable para el usuario.
- Permite el control de versiones.
- Se pueden asignar estereotipos por defecto, o crear estereotipos. [14]

Por sus características y facilidades ofrecidas se selecciona al **Visual Paradigm for UML** en su versión 8.0 como herramienta de modelado para el desarrollo del Mercado de datos Comedores. Además se posee un amplio conocimiento del trabajo con la herramienta y la universidad cuenta con la licencia para su uso.

1.7 Sistemas Gestores de Bases de datos

Un Sistema Gestor de Bases de datos es una colección de programas que permiten definir y manipular una base de datos, especificando tipos, estructuras y restricciones de datos. Facilitan el manejo de grandes volúmenes de información, controlan la redundancia de datos y restringen accesos no autorizados de usuarios que no tengan permiso o autorización.

Oracle: es uno de los gestores de base de datos más completos que existe en la actualidad, fue desarrollado por Oracle Corporation. La herramienta Oracle organiza eficientemente los datos de una manera relacional. [15] Es multiplataforma, soporta grandes bases de datos e implementa una buena seguridad. Es privativo y exige una gran cantidad de recursos de la máquina donde se encuentre instalado.

PostgreSQL: es un sistema que aproxima los datos a un modelo objeto-relacional, pues incluye características de la orientación a objetos, como puede ser la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. [16] Funciona en diferentes sistemas operativos tales como: Linux, UNIX, Mac Os y Windows y posee una buena interfaz de instalación y administración. Incorpora una estructura de datos *array* y soporta el uso de índices, reglas, vistas y procedimientos almacenados. Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos. Tiene soporte para lenguajes procedurales internos, y un lenguaje nativo llamado PL/pgSQL, el cual es comparable con el lenguaje procedural de Oracle, PL/SQL. [16] PostgreSQL permite que mientras un proceso escribe en una tabla, otros accedan a la misma tabla sin necesidad de bloqueos. [17]

Por las características antes mencionadas y por ser una herramienta de *software* libre se selecciona **PostgreSQL** en su versión 9.1 como gestor de base de datos. Además la universidad promueve el uso de herramientas libres para el desarrollo de productos informáticos.

Al ser seleccionado como gestor de base de datos PostgreSQL, se hace necesario tener un cliente para la administración de las bases de datos. En este caso, por sus características, se selecciona el **PgAdmin** en su versión 1.14. Esta herramienta diseña, mantiene y administra fácilmente las bases de datos almacenadas en el gestor PostgreSQL, es multiplataforma y está liberada bajo la licencia Open Source. [16]

1.8 Modos de almacenamiento de datos

El Procesamiento Analítico En Línea (por sus siglas en inglés OLAP) es una tecnología que facilita el análisis en un almacén de datos y proporciona respuestas rápidas a consultas analíticas complejas, a través de un modelo de datos intuitivo y natural. Existen tres modos de almacenamiento OLAP, estos son:

Procesamiento Analítico Relacional (ROLAP)

El modo de almacenamiento ROLAP, accede a los datos almacenados en un almacén para proporcionar los análisis OLAP. Los datos son almacenados en filas y columnas de forma relacional.

El sistema ROLAP utiliza una arquitectura de tres niveles, los mismos son:

- El nivel de base de datos: usa bases de datos relacionales para el manejo, el acceso y la obtención de datos.
- El nivel de aplicación: es el motor que ejecuta las consultas multidimensionales de los usuarios.
- El motor ROLAP: se integra con niveles de presentación mediante los cuales los usuarios realizan los análisis OLAP. [18]

Procesamiento Analítico Multidimensional (MOLAP)

MOLAP es un modo de almacenamiento que usa una base de datos propietaria multidimensional, en las que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis.

Utiliza una arquitectura de dos niveles: la base de datos multidimensional y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato; y el motor analítico es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP. [19]

Procesamiento Analítico Híbrido (HOLAP)

HOLAP combina los modos de almacenamiento ROLAP y MOLAP para brindar una solución con las mejores características de las dos tecnologías anteriores: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP independiente. [18]

Se escoge como modo de almacenamiento de datos al Procesamiento Analítico Relacional (ROLAP) debido a que el gestor de base de datos seleccionado para la solución modela las bases de datos de forma relacional y no permite hacerlo de manera multidimensional.

1.9 Herramientas para el proceso ETL

El flujo de datos desde el/los sistema/s fuente/s hacia el mercado de datos en este tipo de soluciones, se logra mediante el uso de herramientas de Extracción Transformación y Carga (ETL).

Extracción: la fase de extracción es el primer paso para la obtención de la información de las diferentes fuentes hacia el mercado de datos. Deja los datos en formato preparado para iniciar el proceso de transformación. [20]

Transformación: la fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que puedan ser cargados. Para ello este proceso se compone de limpieza y estandarización e integración de los datos. [20]

Carga: una vez que los datos han sido extraídos de las diferentes fuentes y transformados, se procede a realizar la carga histórica hacia el mercado de datos, después de la carga inicial se hacen cargas incrementales para mantener el mercado actualizado. [20]

Las herramientas para la integración de datos agilizan el proceso de ETL, garantizan la homogeneidad y calidad de los datos. Entre las más conocidas se encuentran:

Talend Data Integration, es una herramienta de ETL de código abierto, que transforma y verifica la integridad de los datos. Es una herramienta multiplataforma, poco intuitiva y difícil de entender. Además no cuenta con productos complementarios de calidad de datos y gestión de metadatos. [21]

Por las ventajas que ofrece **Pentaho Data Integration** (por sus siglas en inglés PDI) sobre Talend se decide utilizar esta para el proceso de ETL de la presente investigación en su versión 4.2. Entre las principales ventajas que llevan a seleccionar a PDI sobre Talend se encuentra la amplia experiencia en la utilización de la misma en el centro, es una herramienta de código abierto y multiplataforma que integra, limpia los datos y la información que se encuentra almacenada en aplicaciones y bases de datos separadas y la pone en manos del usuario. Además forma parte de una Suite que cuenta con una comunidad de desarrollo. [22] Incluye algunas herramientas para realizar consultas, generación de informes, reportes, análisis interactivos, tableros de mando, ETL/integración de datos, minería de datos y un servidor para la plataforma de Inteligencia de Negocio. Es extremadamente versátil, pues permite leer y escribir de cualquier base de datos, ficheros Excel, Access y de texto plano. Por otra parte también permite el mapeo de valores, búsquedas auxiliares en bases de datos y normalización de los datos de distintas filas en una sola y viceversa.

Antes de realizar el proceso de extracción, transformación y carga de los datos es necesario conocer la calidad de los datos provenientes de los sistemas fuentes, para ello se utiliza la herramienta **DataCleaner** en su versión 1.5.3, esta es una aplicación de código abierto y multiplataforma para el perfilado, validación y comparación de datos, garantizando que la información sea útil y aplicable a su situación de negocio. Permite perfilar datos que se encuentren almacenados en diferentes sistemas fuentes como son: bases de datos en Oracle, MySQL, PostgreSQL, ficheros Excel (. Xls y Xlsx.), XML, Access (. Mdb). [23]

1.10 Herramientas para Inteligencia de Negocio (BI)

Las herramientas de Inteligencia de Negocio son de gran ayuda para el proceso de toma de decisiones independientemente de la empresa que la utilice, facilitando un mejor entendimiento en el análisis y en la presentación de los datos. A continuación se mencionan algunas de las principales herramientas más utilizadas en el mundo de la Inteligencia de Negocio.

Oracle Business Intelligence Suite Enterprise Edition Plus: ofrece una serie de elementos independientes que tienen a Oracle BI Server como pieza central y están diseñados para proporcionar una extensa y completa visibilidad de los datos de negocio a amplias audiencias de usuarios, pudiendo acceder por vía web a aquellas informaciones relevantes que les interesan. Una de sus ventajas es su arquitectura multiplataforma para el desarrollo de sus aplicaciones. La Suite cuenta con diferentes componentes algunos de ellos son: Oracle Business Intelligence Server, Oracle Business Intelligence Dashboards, Oracle Business Intelligence Answers, Oracle Delivers. [15]

Pentaho BI Server: la plataforma Pentaho BI Server provee el soporte y la infraestructura necesarios para crear soluciones que ayuden a resolver problemas de negocios. La plataforma proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. También incluye un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos. La Suite incluye: [24]

Schema Workbench: es un entorno visual para el desarrollo y prueba de cubos OLAP. Con esta aplicación, se pueden elaborar esquemas lógicos de manera simple y efectiva. El mismo permite realizar el modelo lógico del cubo OLAP al cual se le realizarán las consultas Expresiones Multidimensionales (por sus siglas en inglés MDX). [25] Se conecta directamente con la base de datos para así poder diseñar los cubos OLAP que se requieren para que el usuario final pueda visualizar los indicadores de manera multidimensional.

Mondrian OLAP Server: es una de las aplicaciones más importantes de la plataforma Pentaho BI. Es un servidor OLAP de código abierto que gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente. Entre sus principales características se encuentra la facilidad para el análisis de grandes volúmenes de información almacenada en BD que soporten Java Database Connectivity (por sus siglas en inglés JDBC). Mondrian soporta el lenguaje Microsoft's Multidimensional Expressions (MDX), permite crear cubos de información, los cuales se componen de archivos XML (por su definición en español Lenguaje de Marcas Extensibles) donde se definen las dimensiones y las conexiones de los datos.

Se seleccionó como herramienta para la visualización de la solución la Suite de Pentaho BI Server en su versión 3.10. Para la creación de los cubos OLAP, se escogió Schema Workbench en su versión 3.2 y como motor OLAP Mondrian OLAP Server en su versión 3.0.4, pues a diferencia de Oracle esta

Suite está registrada bajo la licencia GPL (por sus siglas en inglés General Public License). Además se seleccionó como servidor web el Apache Tomcat en su versión 6.0 por ser un servidor robusto, que gestiona solicitudes y respuestas HTTP (por sus siglas en inglés Hyper Text Transfer Protocol). Es gratis, fácil de instalar y se puede ejecutar en máquinas con pocos recursos. Está implementado con tecnología Java bajo la licencia de Apache 2, por lo que funciona en cualquier sistema operativo que se encuentre la máquina virtual de Java.

Conclusiones del capítulo

En el presente capítulo se describieron los principales conceptos relacionados con las tecnologías de almacenamiento de datos, se analizaron los conceptos de almacenes de datos y mercados de datos así como las principales aplicaciones existentes para el área de los comedores. El estudio realizado permitió determinar que no existía ninguna aplicación tanto nacional como internacional que facilitara el análisis estadístico para la ayuda a la toma de decisiones en el área de los Comedores de la UCI, así como el uso del modelo dimensional para la representación del modelo de datos.

Se fundamentó la selección de la “Propuesta de metodología para el desarrollo de almacenes de datos en DATEC” como metodología guía para el desarrollo del Mercado de datos Comedores, aprovechando su ajuste a las condiciones y características de producción en el centro y en la Universidad.

Se estableció el uso del PostgreSQL en su versión 9.1 como gestor de base de datos, el cual se encarga del almacenamiento de los datos; y como cliente para la administración de la base de datos el PgAdmin3 en su versión 1.14. Se seleccionó el DataCleaner en su versión 1.5.3 para el perfilado de los datos, permitiendo arribar a conclusiones sobre el estado y calidad de la información almacenada en el sistema fuente y crear las reglas de transformación a partir de los resultados arrojados por la herramienta. Se escogió como herramienta de modelado al Visual Paradigm en su versión 8.0 para el modelado del diseño e implementación de la solución.

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

Introducción

En este capítulo se describe el negocio, donde se realiza el análisis y diseño del mercado de datos, se identifican los requisitos funcionales, no funcionales y los requisitos de información. Se especifican las necesidades de los usuarios y las reglas del negocio. Se identifican los actores y su relación con los casos de uso del sistema. En el diseño se obtiene el modelo de datos, después de haberse hecho un análisis de las relaciones entre las dimensiones y los hechos a través de la Matriz Bus.

2.1 Análisis

2.1.1 Tema de análisis

La identificación de los temas de análisis es de suma importancia para el desarrollo de un mercado de datos. Los mismos son áreas de una entidad o empresa, que manipulan información y que poseen objetivos o metas propias. Para la construcción de la propuesta se define como tema de análisis en el área de los Comedores de la Universidad de las Ciencias Informáticas: la planificación del menú, el consumo de materia prima, el consumo en los comedores, la distribución de raciones, los portadores energéticos y el control económico.

2.1.2 Definición del negocio

El área de los Comedores de la universidad cuenta con tres complejos, los cuales están conformados por varios comedores, encargados de brindar servicios a toda la comunidad universitaria. La elaboración de los alimentos se realiza de manera centralizada por cada uno de los complejos. Esta área cuenta con el sistema de control de acceso, que se ejecuta en las puertas de los comedores de cada complejo, el mismo guarda información de la cantidad de clientes que acceden a los servicios del comedor. Existen otros indicadores que son importantes analizar en esta área, ellos son: el control del consumo de electricidad, el consumo de materia prima, la planificación del menú y las raciones, así como el control de los ingresos por ventas de tickets. Algunos de estos indicadores se recogen de forma manual impidiendo que se puedan hacer reportes estadísticos que ayuden a la toma de decisiones en esta área.

2.1.3 Reglas del negocio

Las reglas del negocio (por sus siglas en español RN) describen políticas, normas, operaciones, definiciones y restricciones especiales que se deben cumplir en algunas actividades de los procesos diarios de la organización. [26] Son las transformaciones que se le deben realizar a ciertos datos para

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

obtener a partir de ellos otros datos. A continuación se muestran las reglas del negocio del Mercado de datos Comedores.

RN1: cant_total_consumo_gas = sumatoria del consumo de gas.

RN2: cant_total_consumo_electricidad = sumatoria del consumo de electricidad.

RN3: cant_total_raciones_elaboradas = sumatoria de las raciones elaboradas.

RN4: cant_total_raciones_consumidas = total_clientes.

RN5: cant_raciones_vendidas_trabajadores = sumatoria de las raciones que se le venden a los trabajadores del complejo.

RN6: cant_raciones_exentas_pago = sumatoria de las raciones que son exentas de pagos.

RN7: ingresos_ventas_tickets = sumatoria de los ingresos por la venta de los tickets.

RN8: cant_clientes_aplicacion = sumatoria de los clientes que se registran mediante la pc.

RN9: cant_clientes_torpedo = sumatoria de los clientes que se registran mediante el torpedo.

RN10: total_clientes = cant_clientes_aplicación + cant_clientes_torpedo.

RN11: cant_total_raciones_porcino = cant_total_raciones_elaboradas - cant_total_raciones_consumidas

RN12: porciento_raciones_entregadas_porcino = cant_total_raciones_porcino * 100 / raciones entregadas por complejo comedor.

RN13: porciento_raciones_consumidas_dobles = total_clientes_en_doble * 100 / cant_total_raciones_elaboradas.

RN14: porciento_raciones_consumidas_ventas = cant_raciones_vendidas_trabajadores * 100 / cant_total_raciones_elaboradas.

RN15: tickets_recoger_consumidos_en_trabajadores = cant_clientes_torpedo + cantidad de trabajadores externos (PC) - cant_raciones_exentas_pago.

RN16: tickets_recoger_consumo_dobles_y_ventas = (total_clientes_dobles (PC)/2) - (raciones_vendidas_trabajadores)

RN17: diferencia_tickets_e_ingresos_sin_dobles_ni_raciones_vendidas = ingresos_ventas_tickets - tickets_recoger_consumidos_en_trabajadores.

RN18: diferencia_tickets_e_ingresos_con_dobles_y_raciones_vendidas = ingresos_ventas_tickets - cant_raciones_exentas_pago - tickets_recoger_consumo_dobles_y_ventas.

RN19: $\text{cant_total_productos} = \text{sumatoria de los productos.}$

RN20: $\text{cant_total_platos_planificados} = \text{sumatoria de los platos planificados en la confección del menú.}$

RN21: $\text{cant_total_platos_reales} = \text{sumatoria de los platos reales del menú.}$

RN22: $\text{porciento_platos_ofertados} = \text{cant_total_platos_reales} * 100 / \text{cant_total_platos_planificados.}$

RN23: $\text{total_aporte_nutricional} = \text{sumatoria (aporte nutricional de los productos elaborados).}$

RN24: $\text{porciento_cumplimiento_nutricional_desayuno} = \text{total_aporte_nutricional} * 100 / 480.$

RN25: $\text{porciento_cumplimiento_nutricional_almuerzo_comida} = \text{total_aporte_nutricional} * 100 / 960.$

RN26: $\text{cant_frecuencia} = \text{sumatoria de la frecuencia del producto.}$

2.1.4 Necesidades de los usuarios

Para el desarrollo del análisis en el proceso del negocio es necesario conocer lo que necesitan los usuarios, así como la implicación de los mismos en el proyecto, aumentando así las probabilidades de satisfacer sus necesidades. Luego de un análisis del negocio que tiene lugar en el área de los Comedores de la Universidad de las Ciencias Informáticas se define como necesidad de los usuarios el análisis que estos deseen realizar sobre los datos, de ellas se derivan los requisitos de información.

2.1.5 Requerimientos

Requisitos de información

Son las principales informaciones que deben estar disponibles al realizar los análisis sobre los datos. Constituyen una entrada fundamental para el proceso de Inteligencia de Negocio y futuros reportes. Los requisitos de información identificados son los siguientes:

Requisitos de información del consumo en los comedores

1. Obtener la cantidad total de clientes que accedieron a los servicios del comedor por tipo de evento, categoría, complejo comedor, comedor, puerta y tiempo.
2. Obtener la cantidad total de clientes que accedieron a los servicios del comedor en un rango de tiempo.
3. Obtener la cantidad total de clientes por complejo comedor, comedor, puerta y tiempo.
4. Obtener la cantidad total de clientes que se registraron manualmente por cada punto de acceso (torpedo) por tipo de evento, categoría, complejo comedor, comedor, puerta y tiempo.

Requisitos de información del consumo de materia prima

5. Obtener la cantidad total de productos consumidos por tipo de evento y tiempo.

Requisitos de información del menú planificado y el cumplimiento de la norma nutricional.

6. Obtener la cantidad total de platos planificados, tipo de evento y tiempo.
7. Obtener la cantidad total de platos reales, tipo de evento y tiempo.
8. Obtener la cantidad de frecuencias de productos por tipo de evento y tiempo.
9. Obtener el porcentaje de platos ofertados por tipo de evento y tiempo.
10. Obtener el porcentaje de cumplimiento de la norma nutricional por tipo de evento y tiempo.

Requisitos de información del control de los ingresos y pagos por doble y raciones vendidas

11. Obtener el ingreso total por venta de tickets por tipo de evento, complejo comedor y tiempo.
12. Obtener la cantidad de raciones exentas de pago por tipo de evento, complejo comedor y tiempo.
13. Obtener la cantidad de tickets a recoger por tipo de evento, complejo comedor y tiempo.
14. Obtener la diferencia entre la cantidad de tickets e ingresos por tipo de evento, complejo comedor y tiempo.

Requisitos de información de las raciones elaboradas y su distribución

15. Obtener la cantidad total de raciones elaboradas por tipo de evento, complejo comedor y tiempo.
16. Obtener la cantidad total de raciones consumidas por tipo de evento, complejo comedor y tiempo.
17. Obtener la cantidad total de raciones entregadas al porcino por tipo de evento, complejo comedor y tiempo.
18. Obtener el porcentaje de las raciones elaboradas consumidas por tipo de evento, complejo comedor y tiempo.
19. Obtener el porcentaje de las raciones elaboradas entregadas al porcino por tipo de evento, complejo comedor y tiempo.
20. Obtener la cantidad total de raciones vendidas a los trabajadores.

Requisitos de información de los portadores energéticos

21. Obtener la cantidad total del consumo de gas por complejo comedor y tiempo.
22. Obtener la cantidad total de consumo de energía por complejo comedor y tiempo.

Requisitos funcionales

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

Los requisitos funcionales son capacidades o condiciones que el sistema debe cumplir [27] y su proceso de obtención es importante en el desarrollo del mercado de datos, de la calidad de los mismos depende el éxito del producto.

Para el desarrollo de los procesos de integración se identificaron los siguientes requisitos funcionales:

RF1- Validar usuario y contraseña.

RF2- Adicionar rol.

RF3- Modificar rol.

RF4- Eliminar rol.

RF5- Eliminar usuario.

RF6- Adicionar usuario.

RF7- Modificar usuario.

RF8- Modificar reporte.

RF9- Eliminar reporte.

RF10- Adicionar reporte.

RF11- Personalizar reporte.

RF12- Realizar transformación de los datos.

RF13- Extraer datos de la fuente.

RF14- Cargar los datos al Mercado de datos Comedores.

RF15- Exportar a otro formato.

Requisitos no funcionales

Los requisitos no funcionales definen propiedades o cualidades que el producto debe tener. [27] Para el desarrollo del mercado de datos se identificaron 15 requerimientos no funcionales del sistema, a continuación se muestran algunos ejemplos de requisitos no funcionales, de usabilidad, soporte, fiabilidad y restricciones del diseño. Para tener una visión más detallada de los requisitos no funcionales del mercado, se debe consultar en el Expediente de proyecto el artefacto "0113_Especificación de Requisitos de Software.doc".

Usabilidad

Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.

Los títulos de los componentes de la interfaz, los mensajes para interactuar con los usuarios y los mensajes de error, deben ser en idioma español y tener una apariencia uniforme en todo el sistema. Los mensajes de error deberán ser lo suficientemente informativos para dar a conocer la severidad del error.

Soporte

Lograr que los elementos definidos en el almacén tengan una nomenclatura homogénea.

Las estructuras del mercado de datos se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje. Se definen convenciones de nombrado con el objetivo de manejar un vocabulario común en el mercado de datos para un entendimiento claro y conciso de las estructuras por parte de los desarrolladores que interactúen con el mercado de datos.

Interfaz

Acceder al sistema

Se recomienda acceder a la aplicación a través del protocolo HTTP usando el navegador Firefox.

Software

Instalar en las estaciones de trabajo los requisitos mínimos de software

Las configuraciones de software de las máquinas clientes deben contar al menos el siguiente elemento:

Navegador web (preferentemente Firefox)

Java Virtual Machine

2.1.6 Actores del sistema

Los actores representan los usuarios del sistema u otras aplicaciones que interactúan con él, es decir, representan terceros fuera del sistema que inicializan los casos de uso (por sus siglas en español CU).

Descripción de los actores del sistema

Tabla 1. Descripción de los actores del sistema.

Actor	Descripción
Administrador	El administrador interactúa con el sistema para realizar todas las operaciones de administración de roles, usuarios y reportes.
Administrador ETL	El administrador ETL interactúa con la base de datos para monitorizar los procesos de extracción, transformación y carga de los datos.
Especialista	El especialista se autentica en el sistema y visualiza los reportes de la aplicación.

2.1.7 Diagrama de casos de uso del sistema

El diagrama de casos de uso del sistema (por sus siglas en español DCUS) es una representación de todos los actores del mismo, los casos de uso y las relaciones que existen entre ellos. Para la confección de éste, se agruparon los 15 requisitos funcionales del sistema en siete casos de uso funcionales (por sus siglas en español CUF) y los 22 requisitos de información en seis casos de uso informativos (por sus siglas en español CUI), estos últimos fueron agrupados a través del criterio de agrupación: tema de análisis. (Ver figura 1).

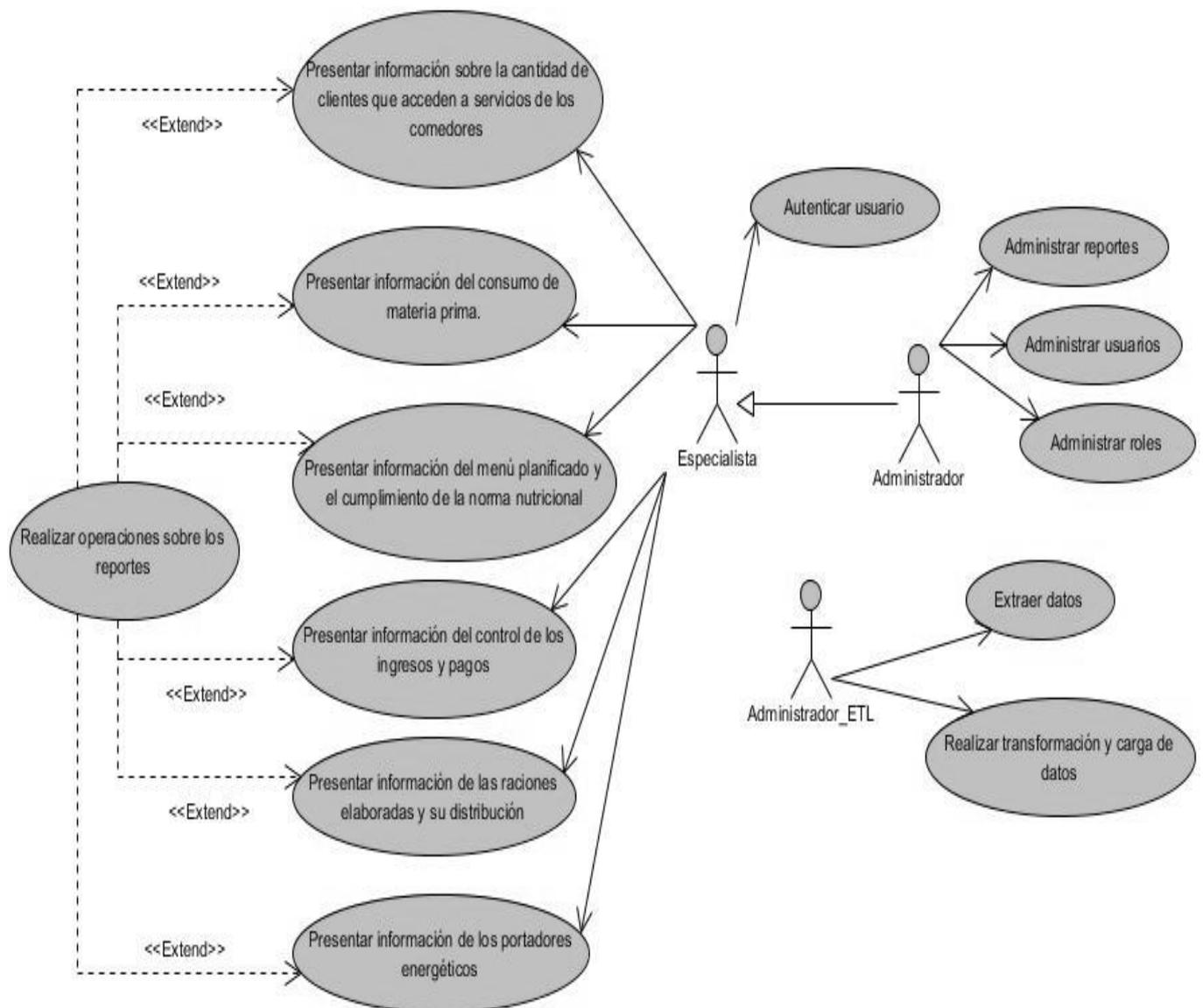


Figura 1. Diagrama de casos de uso del sistema.

Descripción textual de los casos de uso

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

Al describir textualmente los casos de uso, se puede tener una mejor visión y un nivel de detalle más específico de las necesidades del cliente para así poder comenzar el desarrollo del sistema. A continuación se describe el CUI “Presentar información de los portadores energéticos” (Ver tabla 2), Para ver todas las descripciones textuales de los CU se debe consultar el artefacto “0114_Especificación de casos de uso.doc”.

CUI: Presentar información de los portadores energéticos.

Tabla 2. Descripción del caso de uso "Presentar información de los portadores energéticos".

Objetivo	Presentar información del consumo de energía.	
Actores	Especialista	
Resumen	El caso de uso inicia cuando el especialista selecciona Presentar información de los portadores energéticos, luego se realiza la selección de la información a mostrar. Finaliza cuando se muestran los resultados.	
Complejidad	Alta	
Prioridad	Media	
Precondiciones	El especialista debe estar autenticado.	
Postcondiciones		
Flujo básico		
	Actor	Sistema
1.	Selecciona el Área de Análisis General (por sus siglas en español A.A.G) Sala_Situacional_UCI.	
2.		Muestra las Áreas de Análisis (por sus siglas en español A.A) de la Sala Situacional UCI.
3.	Selecciona el A.A Comedores.	
4.		Muestra los libros de trabajo del A.A Comedores.
5.	Selecciona el Libro de Trabajo (por sus siglas en español LT) Comedores.	

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

6.		Muestra los reportes existentes en el L.T del A.A de los Comedores.
7.	Selecciona el reporte que desea analizar.	
8.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Realizar operaciones sobre los reportes.
Opciones de reportes de <Presentar información de los portadores energéticos>		
Perspectivas de análisis		Posibles resultados
		Medidas
		Periodicidad
Variables de entrada relacionadas con el CU Presentar información de la cantidad de usuarios que acceden a los servicios de los comedores:		Variables de salida disponibles en el hecho Planificación de usuarios: Cant_total_consumo_gas Cant_total_consumo_electricidad
<ul style="list-style-type: none"> • Tiempo_mes • Complejo_comedor 		Rango de tiempo en que se solicitan las variables de salida: Mensual
Relaciones	CU Incluidos	No aplica
	CU Extendidos	CU Realizar operaciones sobre los reportes

2.2 Arquitectura del Mercado de datos

El proceso de desarrollo de *software* requiere de la definición previa de una arquitectura. Esta representa la guía para el diseño y construcción del mercado de datos, así como también sirve como marco de organización para apoyar la integración de las tecnologías. [5]

La arquitectura escogida para el desarrollo del Mercado de datos Comedores consta de cinco niveles:

Fuentes de datos: se refiere al origen de los datos.

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

Área temporal: es una base de datos donde se guardan los datos que son de interés para el negocio antes de ser cargados al mercado de datos.

Subsistema de integración: comprende un conjunto de herramientas para la limpieza, extracción, transformación y carga de los datos hacia el mercado de datos.

Subsistema de almacenamiento: se representa a través de una base de datos relacional que contiene las tablas de dimensiones y hechos cargadas a través de los procesos de ETL.

Subsistema de visualización: comprende las interfaces orientadas a usuarios que extraen información para apoyar el proceso de toma de decisiones. La figura muestra la arquitectura antes descrita (Ver figura 2).

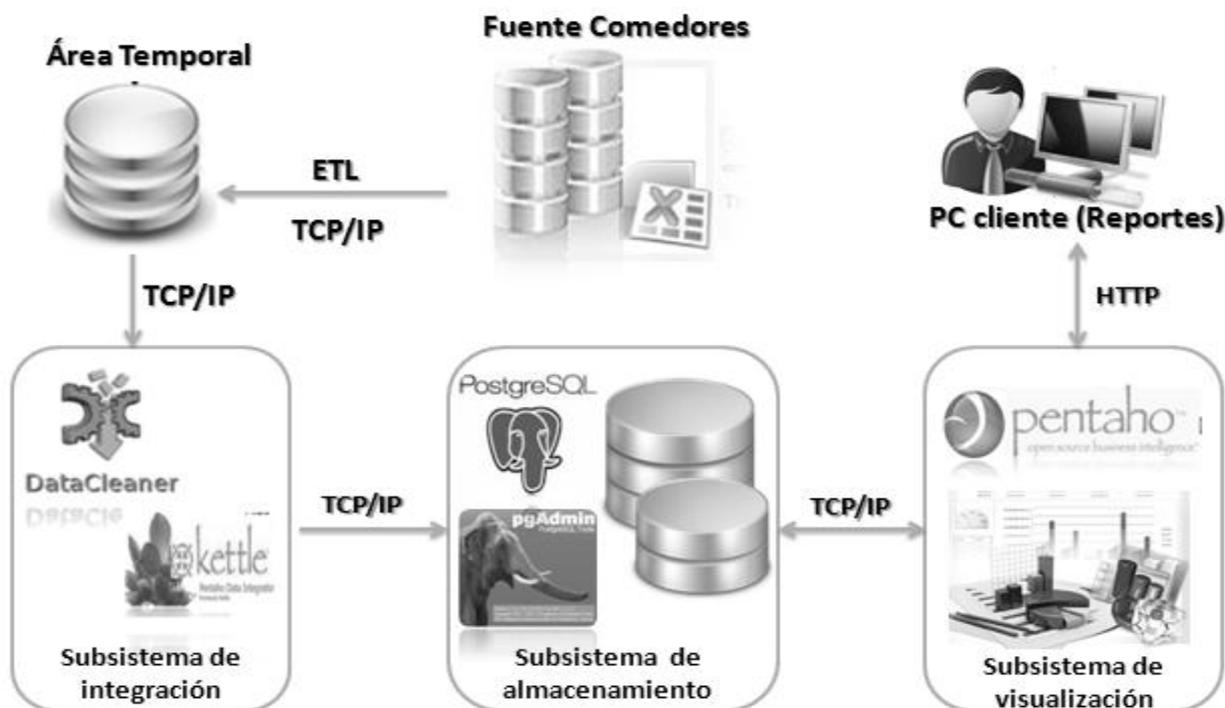


Figura 2. Diseño de la arquitectura de un almacén de datos.

En el primer nivel se encuentran las fuentes primarias del área Comedores, en esta área los indicadores relacionados con el acceso a los comedores se encuentran almacenados en una base de datos y los indicadores del menú, del control económico, del consumo de energía, del consumo de productos y de la distribución de las raciones se recogen de forma manual en ficheros Excel.

En un segundo nivel se encuentra el área temporal la cual recibirá información de los indicadores antes mencionados de los sistemas fuentes primarios, mediante un proceso de extracción, transformación y carga de los datos.

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

Una vez almacenados los indicadores en el área temporal son liberados de los sistemas de origen y pasan a ser limpiados y transformados por herramientas de integración de datos hasta que son cargados en el mercado de datos “mart_comedores”.

Por su parte, el nivel cuatro recibe la información procedente de la extracción y transformación, en una base de datos soportada por el gestor de base de datos PostgreSQL y administrada por los usuarios autorizados mediante la herramienta pgAdminIII.

Por último los usuarios finales pueden visualizar y explotar la información mostrada a través de los reportes que son implementados utilizando herramientas de Inteligencia de Negocio.

2.3 Diseño

En el proceso de diseño de un mercado de datos intervienen tres subsistemas:

Subsistema de Almacenamiento: en este subsistema se tienen los sistemas fuentes de donde son extraídos los datos. Del mismo dependen el subsistema de integración y el de visualización.

Subsistema de Integración: es el subsistema encargado de extraer los datos de los sistemas fuentes, los cuales pasan por un proceso de transformación y finalmente son entregados en un formato listo para la presentación, de forma tal que los desarrolladores puedan construir la aplicación y los usuarios finales puedan consultar los reportes.

Subsistema de Visualización: encargado de transformar la información de un gran conjunto de datos necesarios para las labores diarias, en datos que brindan conocimiento y apoyan el proceso de toma de decisiones. Las funcionalidades de dicho subsistema están encaminadas a generar reportes, gráficos en diferentes perspectivas y vistas de análisis, con el objetivo de arribar a conclusiones correctas con mayor rapidez (Ver Figura 3).

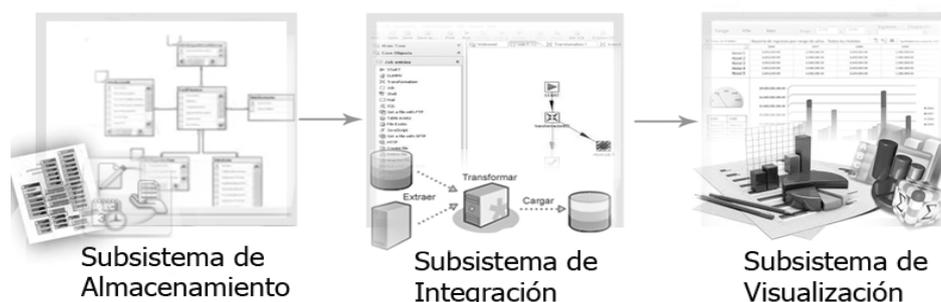


Figura 3. Diseño de los subsistemas.

2.3.1 Diseño del subsistema de almacenamiento

Como parte del subsistema de almacenamiento se identifican las dimensiones, los hechos y las medidas.

Las dimensiones definidas se muestran en la tabla 3.

Tabla 3. Dimensiones.

Dimensiones estáticas	Descripción
dim_tiempo_mes	Presenta la información asociada a un tiempo donde el nivel más bajo es el mes.
dim_tiempo_semana	Presenta la información asociada a un tiempo donde el nivel más bajo son las semanas.
dim_tiempo_dia	Presenta la información asociada a un tiempo donde el nivel más bajo son los días.
Dimensiones cambiantes	Descripción
dim_tipo_evento	Representa los diferentes eventos a los que pueden asistir los usuarios (desayuno, almuerzo, comida y los dobles respectivamente).
dim_complejo_comedor	Representa el complejo comedor con todos sus comedores y puertas.
dim_categoria	Representa las diferentes categorías de clientes que existen (estudiantes, trabajadores, profesores).
dim_tipo_producto_bruto	Representa los productos en bruto.
dim_tipo_productos_elaborados	Representa los productos elaborados que conforman el menú.
dim_norma_nutricional	Representa la norma que tiene cada evento.
dim_unidad_medida	Representa la unidad de medida de los productos en bruto.

Las medidas definidas se muestran en la tabla 4.

Tabla 4. Medidas.

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

Nombre	Descripción
cant_total_consumo_gas	Muestra la cantidad total de consumo de gas licuado en libras de los complejos comedores.
cant_total_consumo_electricidad	Muestra la cantidad total de consumo de electricidad en Kw/h de los complejos comedores.
cant_total_raciones_elaboradas	Muestra la cantidad total de raciones elaboradas en los complejos comedores.
cant_total_raciones_consumidas	Muestra la cantidad total de raciones consumidas en los complejos comedores.
cant_raciones_vendidas_trabajadores	Muestra la cantidad total de raciones que se les venden a los trabajadores después que terminan los servicios del comedor.
ingresos_ventas_tickets	Muestra la cantidad de ingresos que se tienen por las ventas de tickets.
cant_raciones_exentas_pago	Muestra la cantidad de raciones exactas de pagos.
cant_clientes_aplicacion	Muestra la cantidad total de clientes que acceden a los servicios del comedor por la aplicación control del acceso.
cant_clientes_torpedo	Muestra la cantidad total de clientes que acceden a los servicios del comedor pero que no se registran en la aplicación "Control de acceso" por pérdida de solapín u otros problemas y se registran manualmente en un fichero Excel.

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

cant_total_productos	Muestra la cantidad total de los productos consumidos.
cant_total_platos_planificados	Muestra la cantidad total de platos planificados que forman parte del menú.
cant_total_platos_reales	Muestra la cantidad total de platos reales que forman parte del menú.
cant_frecuencia	Muestra la cantidad de frecuencia con la que se oferta un tipo de producto en el menú semanalmente.
total_clientes	Muestra el total de los clientes que recibieron servicios en el comedor.
raciones_entregadas_al_porcino	Muestra el total de las raciones que se entregan al porcino al terminar los servicios del comedor.
porciento_raciones_entregadas_al_porcino	Muestra el porciento de las raciones que se entregan al porcino al terminar los servicios del comedor.
porciento_raciones_consumidas_dobles	Muestra el porciento de las raciones elaboradas que se consumen en dobles.
porciento_raciones_consumidas_ventas	Muestra el porciento de las raciones elaboradas que se consumen en ventas a los trabajadores.
tickets_recoger_consumidos_en_trabajadores	Muestra cantidad de tickets a recoger que se consumieron por trabajadores.
tickets_recoger_consumo_dobles_y_ventas	Muestra cantidad de tickets a recoger que se consumieron en dobles y en venta a los trabajadores.
diferencia_tickets_e_ingresos_sin_dobles_ni_raciones_vendidas	Muestra la diferencia entre los ingresos y los tickets sin dobles ni raciones vendidas a los trabajadores.
diferencia_tickets_e_ingresos_con_dobles_y_raciones_vendidas	Muestra la diferencia entre los ingresos y los tickets con dobles y raciones vendidas a los trabajadores.

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

porciento_platos_ofertados	Muestra el porciento de platos ofertados entres los reales y los planificados.
porciento_cumplimiento_nutriciona l_desayuno	Muestra el porciento de cumplimiento de la norma nutricional en el evento del desayuno.
porciento_cumplimiento_nutriciona l_almuerzo_comida	Muestra el porciento de cumplimiento de la norma nutricional en los eventos: almuerzo y comida.

Los hechos del mercado de datos se muestran en la tabla 5.

Tabla 5. Hechos

Hecho	Descripción
hech_energia	Representa el control de la energía de los complejos comedores.
hech_planificacion	Representa la planificación de los platos que forman parte del menú.
hech_productos	Representa los productos en bruto que consumen los complejos comedores.
hech_control	Representa el control de las raciones vendidas a los trabajadores y las exentas de pago, así como el control de los clientes a los servicios del comedor.
hech_frecuencia_producto	Representa la frecuencia con que se ofertan los productos en los complejos comedores.
hech_raciones_elaboradas_cons umidas	Representa las raciones elaboradas y las consumidas en los complejos comedores.
hech_aporte_nutricional	Representa el aporte nutricional de cada producto elaborado.

Matriz BUS

Se confecciona la matriz BUS para evitar el solapamiento entre los hechos. La misma ayuda a visualizar las relaciones entre los hechos y las dimensiones para una posterior confección del modelo de datos. Las filas contienen todos los hechos que fueron identificados y en las columnas se

encuentran las dimensiones. En la misma quedan evidenciadas las relaciones existentes entre las tablas del mercado de datos (Ver Figura 4).

Dimens iones/H echos	H1	H2	H3	H4	H5	H6	H7
D1		X	X	X	X		X
D2	X						
D3							X
D4	X	X	X	X			
D5		X	X	X	X		
D6				X			X
D7			X				
D8							X
D9							X
D10				X			

Leyenda de hechos

H1: hech_energia
H2: hech_raciones_consumidas_elaboradas
H3: hech_control
H4: hech_productos
H5: hech_planificacion
H6: hech_frecuencia_producto
H7: hech_aporte_nutricional

Leyenda de dimensiones

D1: dim_tiempo_dia
D2: dim_tiempo_mes
D3: dim_tiempo_semana
D4: dim_complejo_comedor
D5: dim_tipo_evento
D6: dim_tipo_producto_bruto
D7: dim_categoria
D8: dim_tipo_productos_elaborados
D9: dim_norma_nutricional
D10: dim_unidad_medida

Figura 4. Matriz BUS.

Modelo de datos

A partir de la identificación de las dimensiones, los hechos y las medidas del almacén se obtuvo el diseño de la solución que a continuación se muestra, para tener una visión de la estructura lógica del mercado de datos (Ver Figura 5).

Capítulo 2: Análisis y Diseño del Mercado de datos Comedores

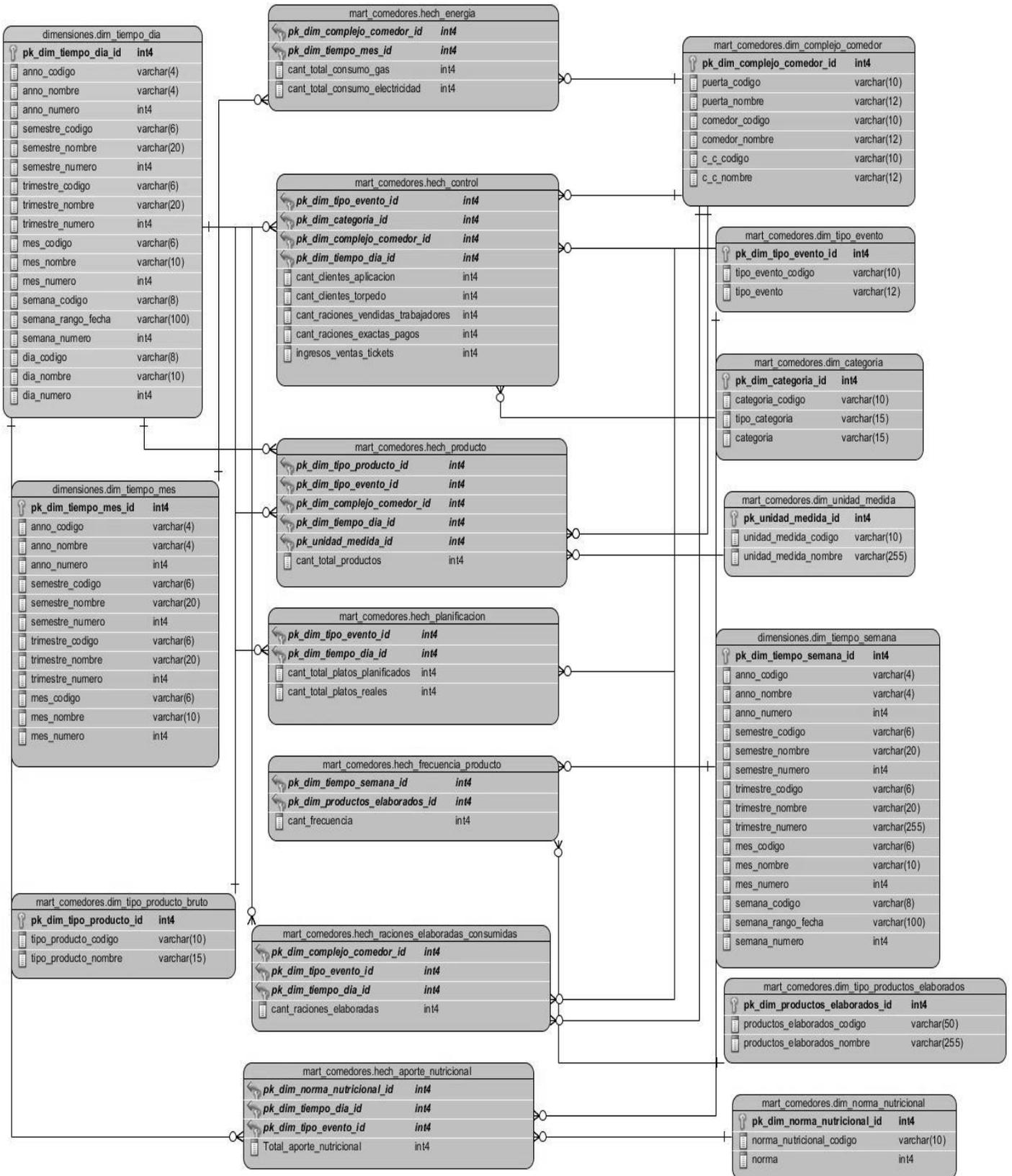


Figura 5. Modelo de datos.

La idea fundamental del modelo de datos dimensional es que los datos del negocio queden representados en forma de cubo de datos. En los cubos cada celda contiene un valor y las aristas del cubo definen dimensiones de análisis. En la solución propuesta se seleccionó el esquema arquitectónico Constelación de hechos, Figura 7, el cual está compuesto por varios esquemas de estrella, donde existen varias tablas de hechos en el centro que se relacionan con una o más dimensiones. Cada dimensión posee una llave primaria que se encarga de mantener la relación que existe entre las tablas de dimensiones y las tablas de hecho.

2.3.2 Diseño del subsistema de integración

Como parte del subsistema de integración se realiza el diseño de las transformaciones el cual puede variar a la hora de hacer la extracción, transformación y la carga de los datos.

Diseño de las transformaciones

Diseño para la carga de las dimensiones estáticas: los datos son cargados del área temporal. Luego se valida que los campos de entrada cumplan con algunas restricciones. En caso de no cumplirse alguna restricción, los datos se almacenan en un fichero Excel para darles tratamiento, si no se insertan en la base de datos mart_comedores. La figura muestra el diseño de la transformación para la carga de las dimensiones (Ver figura 6).

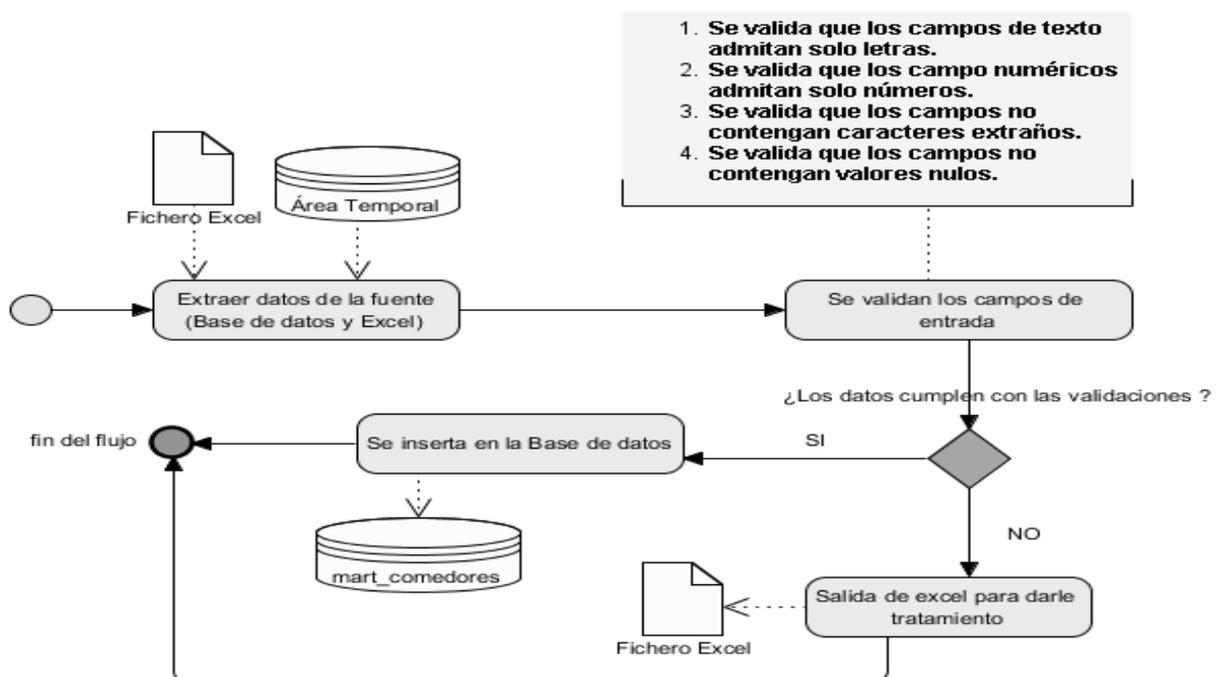


Figura 6. Diseño de las transformaciones para la carga de las dimensiones estáticas.

Diseño para la carga de las dimensiones lentamente cambiantes: inicialmente se cargan los datos de los sistemas fuentes y se validan los campos de entrada verificando que cumplan con algunas restricciones. En caso de no cumplirse alguna restricción, los datos se almacenan en un fichero Excel para parles tratamiento, si no son insertados en la base de datos. En caso que sea necesario insertar una nueva fila, este tipo de dimensión guarda en una tabla de metadatos la fecha de inicio y fin de la inserción y el nombre del esquema donde se encuentra la dimensión a la que se le agrega la nueva fila. La figura muestra el diseño de la transformación para la carga de las dimensiones (Ver figura 7).

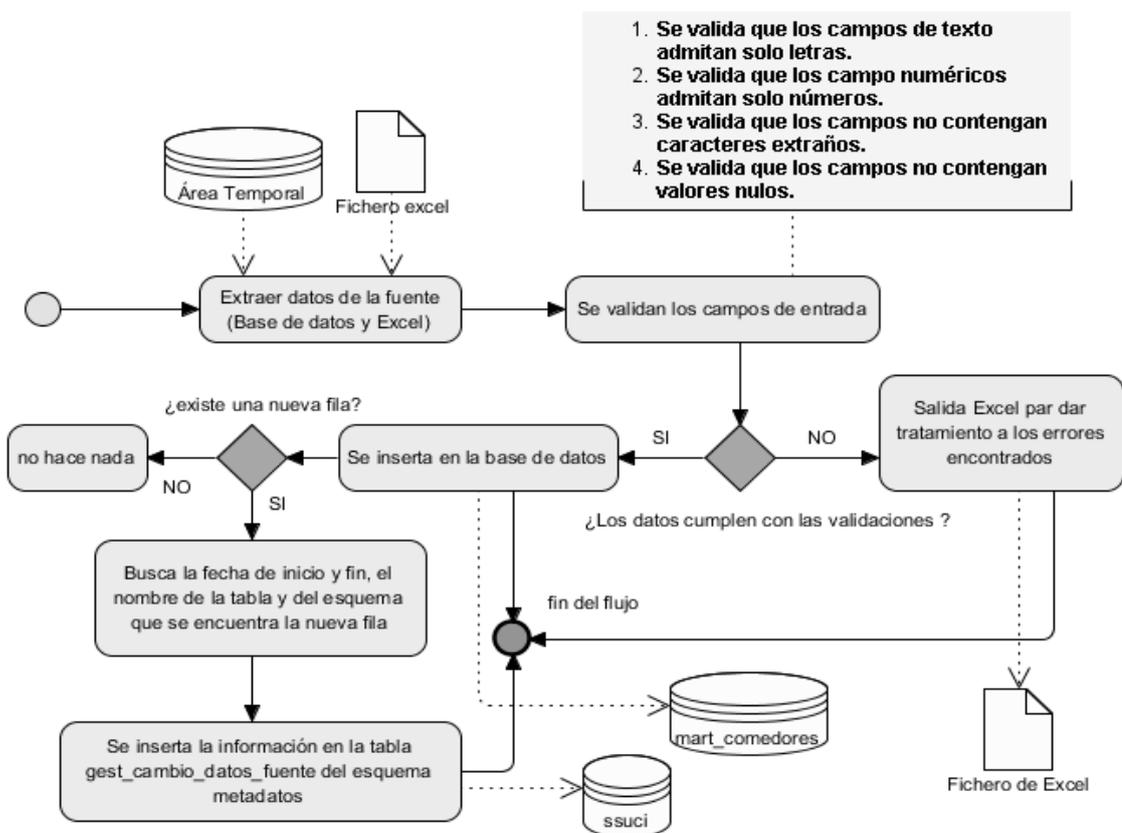


Figura 7. Diseño de las transformaciones para la carga de las dimensiones lentamente cambiantes.

Diseño para la carga de los hechos: una vez ejecutadas las transformaciones de las dimensiones, se procede a realizar la carga de los hechos, para ello se validan un grupo de restricciones definidas por el especialista. En caso de no cumplirse una de estas restricciones, los datos se almacenan en un fichero Excel para darles tratamiento; en caso contrario, todos los datos se insertan en la tabla de la base de datos que tiene como salida dicha transformación. La figura muestra el diseño de la transformación para la carga de los hechos (Ver figura 8).

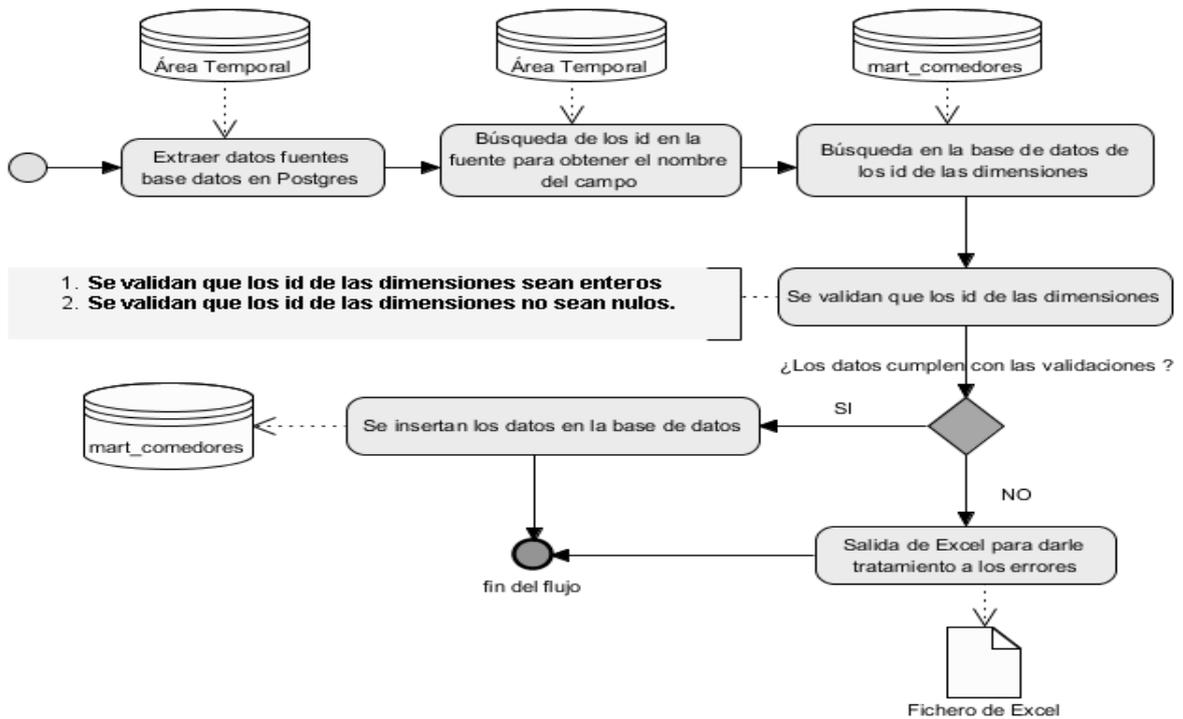


Figura 8. Diseño de las transformaciones para la carga de los hechos.

2.3.3 Diseño del subsistema de visualización

Navegación de la capa de visualización: constituye la estructura gráfica en que está organizada la información. En el área de análisis general (A.A.G) Sala Situacional UCI se encuentra el área de análisis (A.A) Comedores la cual cuenta con seis Libros de trabajo (LT) que guardan la información referente a los temas de análisis que fueron identificados a través de las necesidades del usuario y tienen asociado un número de reportes candidatos (Ver Figura 9).



Figura 9. Diseño del mapa de navegación.

Diseño de los reportes candidatos

Los reportes candidatos responden a las necesidades de los usuarios y corresponden con las salidas que debe producir el sistema. Estos son implementados en la Inteligencia de Negocio (BI).

En el presente trabajo se diseñaron 31 reportes candidatos agrupados en seis libros de trabajos. La tabla 6 muestra un ejemplo del reporte candidato: “Consumo de electricidad por complejo comedor”.

Tabla 6. Reporte candidato “ Consumo de electricidad por complejo comedor”.

Área de análisis (A.A)	Comedores
Libro de Trabajo (L.T)	L.T 6 – Portadores energéticos
Reporte (Tabla de Salida – TS)	T.S 30 – Consumo de electricidad por complejo comedor.
Descripción	El reporte muestra el consumo de electricidad en Kw/h para cada complejo comedor.
Elementos del reporte	<ul style="list-style-type: none"> ➤ Tiempo ➤ Complejo comedor
Frecuencia de emisión	Mensual

2.4 Políticas de seguridad

2.4.1 Roles y permisos

Para el acceso al mercado de datos se define un usuario por cada uno de los roles existentes en el sistema, con el objetivo de garantizar un control de acceso basado en roles y así cada usuario opera en el sistema según los permisos que se le definan al rol (Ver tabla 7).

Tabla 7. Roles y permisos.

Roles	Permisos	
	Lectura	Escritura
Sobre la Base de Datos		
Administrador de base de datos	x	x
Administrador ETL	x	x
Sobre la aplicación		

Administrador	x	x
Especialista	x	-

2.4.2 Salva de la base de datos

Para garantizar la seguridad de la información se realizará un respaldo total de los datos del mercado. La política de respaldo y recuperación que se empleará está condicionada por dos puntos fundamentales:

- **Periodicidad de las salvas:** las salvas de la base de datos se realizarán diariamente. Para ello se utilizará la copia de seguridad diferencial, donde se hace una copia de seguridad total el primer día de cada mes y una copia de seguridad diferencial el resto de los días, cada copia diferencial guardará los archivos que se hayan modificado después de la copia total. Cada copia diferencial anula la copia diferencial anterior.
- **Tablas involucradas:** las tablas involucradas en las salvas son las tablas de hechos y dimensiones del mercado de datos.

Conclusiones del capítulo

La fase de Análisis abordada en el capítulo permitió comprender las necesidades del usuario a través de los 22 requisitos de información que fueron definidos. Se identificaron además 15 requisitos funcionales y 15 no funcionales, los cuales garantizaron la confección del diagrama de casos de uso del sistema. Se diseñó el modelo de datos dimensional estableciendo las relaciones entre las diez dimensiones y los siete hechos del Mercado de datos Comedores, respondiendo así a los requisitos de información del cliente.

El diseño de cada uno de los subsistemas que integran el Mercado de datos Comedores permitió guiar la posterior fase de Implementación.

Capítulo 3: Implementación del Mercado de datos Comedores

Introducción

La implementación es la fase donde se crean las estructuras del mercado de datos, por lo que está muy relacionado con los procesos de diseño. En esta etapa del sistema, se implementa el modelo físico, donde se genera y puebla la base de datos, a través del *script* del mercado y se realizan los procesos de extracción, transformación y carga de los datos. Como parte del subsistema de visualización se implementan y visualizan los reportes candidatos.

3.1 Implementación del modelo de datos

Para una correcta organización del mercado de datos, se implementaron las tablas del modelo de datos en tres esquemas: “dimensiones”, el cual contiene las dimensiones compartidas *dim_tiempo_dia*, *dim_tiempo_semana* y *dim_tiempo_mes*. El esquema “*mart_comedores*” contiene el resto de las tablas, es decir, los hechos y las dimensiones específicas del mercado de datos; y el esquema metadatos donde se almacenan las tablas que contienen información de los datos del mercado (Ver tabla 8).

Tabla 8. Esquemas y tablas del mercado.

mart_comedores	dimensiones	metadatos
<p>Tables (14)</p> <ul style="list-style-type: none"> dim_categoria dim_complejo_comedor dim_norma_nutricional dim_tipo_evento dim_tipo_producto_bruto dim_tipo_productos_elaborados dim_unidad_medida hech_aporte_nutricional hech_control hech_energia hech_frecuencia_producto hech_planificacion hech_producto hech_raciones_elaboradas_consumidas 	<p>Tables (3)</p> <ul style="list-style-type: none"> dim_tiempo_dia dim_tiempo_mes dim_tiempo_semana 	<p>Tables (4)</p> <ul style="list-style-type: none"> gest_cambio_datos_fuentes gestión_histórico_datos fuente trabajo transformacion

3.2 Implementación de los procesos de Extracción, Transformación y Carga

Perfilado de los datos

Capítulo 3: Implementación del Mercado de datos Comedores

El perfilado de datos se realiza con el objetivo de conocer el estado en que se encuentran los datos de la fuente origen y así poder supervisar la calidad de los mismos con el fin de garantizar que la información sea útil y aplicable a una situación de negocio.

Los resultados mostrados a continuación pertenecen a cuatro de las tablas de la fuente: complejo_comedor, categoria_cliente, control_puerta, menú. Para ver todos los resultados arrojados por el perfilado de los datos se debe consultar en el Expediente de proyecto el artefacto “Perfilado de datos”.

- El valor mínimo que puede tomar el id_complejo_comedor es uno y el máximo es tres.
- El valor mínimo y el máximo de caracteres del nombre del complejo comedor es tres.
- El valor mínimo de caracteres de la categoría es de ocho y el máximo es de 10.
- Existen 41602 valores nulos en la tabla control puerta.
- Existen 11073 valores nulos en la tabla menú.
- No existen valores negativos.

Los resultados arrojados por el perfilado de los datos fueron traducidos a reglas de transformación (RT) que forman parte de las reglas de negocio y se les da solución en la implementación de las transformaciones con la limpieza y validación de los datos.

RT1: una vez cargados los datos en el almacén, no pueden existir campos nulos.

RT2: los identificadores deben ser numéricos y auto incrementales.

RT3: los campos tipo_evento, categoria, estado y nombre del producto elaborado no pueden contener números.

RT4: los códigos deben ser numéricos y únicos.

RT5: el nombre del complejo comedor está definido de la siguiente manera CC1 o CC2 o CC3.

RT6: el código de la puerta está definido de la siguiente manera, tiene que tener tres cifras donde el primer número es el número correspondiente al complejo comedor, el segundo corresponde al comedor y el tercero al número de la puerta.

RT7: el nombre del comedor está definido de la siguiente manera C1, C2, C3 o C4.

Implementación de una transformación para la carga de una dimensión lentamente cambiante

Para la implementación de las dimensiones lentamente cambiantes se seleccionó la de tipo 2 “Añadir fila”, pues la misma permite añadir una fila a la tabla donde se produce el cambio con el nuevo valor, manteniendo el historial del viejo valor para evitar la pérdida de información.

Capítulo 3: Implementación del Mercado de datos Comedores

A continuación se muestra la transformación de la carga de la dimensión complejo comedor (Ver Figura 10).

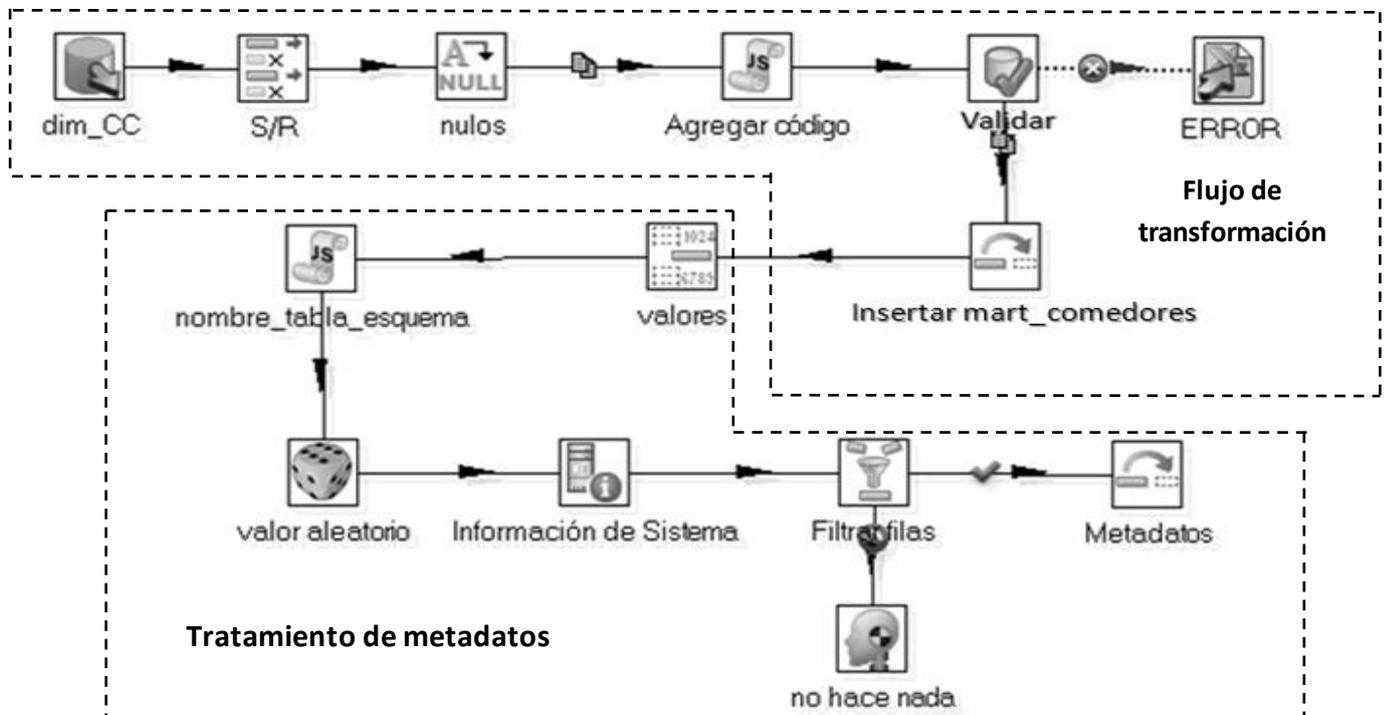


Figura 10. Transformación para la carga de la dimensión dim_complejo_comedor.

Flujo de transformaciones: para poblar la tabla de dimensión dim_complejo_comedor se validó que no existieran campos nulos y que los campos del flujo fueran solo letras, solo números o letras y números dependiendo del tipo de dato del campo. En este caso el campo **nombre de la puerta** solo puede admitir números y en el caso del **nombre del comedor** y **nombre del complejo comedor** puede admitir letras y números pero no caracteres extraños. Posteriormente se validaron los campos mencionados y se insertaron en la base de datos mart_comedores si no es encontrado ningún error. En caso contrario los errores fueron enviados a un Excel con una descripción de los mismo para que el cliente pueda corregir los errores y los datos puedan ser cargados nuevamente.

Tratamiento de metadatos: se le da tratamiento a los metadatos por ser una dimensión lentamente cambiante donde sus datos pueden variar con el tiempo. Este tratamiento se basa en registrar fecha de inicio y fecha de fin en que se agrega una nueva fila, así como el nombre de la tabla de dimensión y el nombre del esquema al que pertenece dicha tabla. Estos cambios son guardados en la tabla gest_cambio_datos_fuente del esquema metadatos.

Implementación de una transformación para la carga de una dimensión estática

Para poblar la tabla de dimensión dim_evento, se añadió el código a la dimensión. Posteriormente se validó que no existieran campos nulos y que el campo **tipo de evento** fuera solo letras. Si no es encontrado ningún error se insertan los valores en la tabla dim_evento de la base de datos mart_comedores. En caso contrario los errores fueron enviados a un Excel con una descripción de los mismo para que el cliente pueda corregir los errores y los datos puedan ser cargados nuevamente (Ver figura 11). Para ver más transformaciones de la carga de las dimensiones estáticas ver anexo 2.

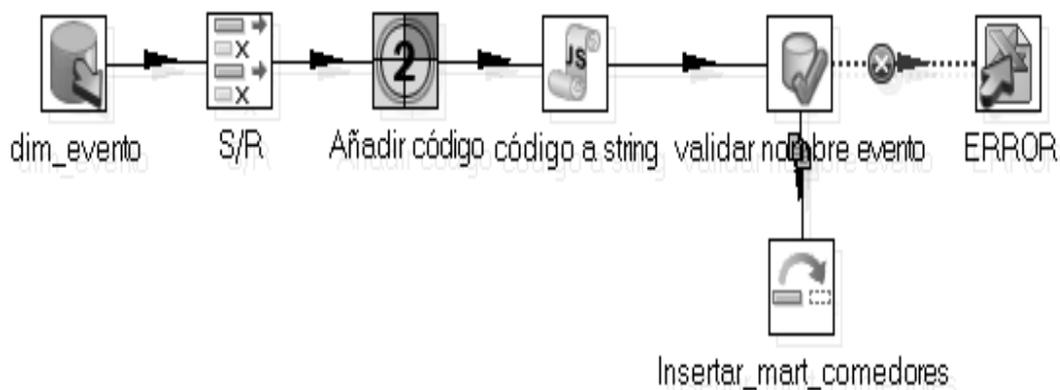


Figura 11. Transformación para la carga de la dimensión dim_evento.

Implementación de una transformación para la carga de un hecho

En la implementación de la transformación para la carga del hech_energia se validó la existencia de campos nulos y se realizó la búsqueda en la base de datos de los indicadores de las dimensiones relacionadas con este hecho. Se validó que los id de las dimensiones fueran enteros y no nulos, si estos están en perfecto estado se procede a ordenar los datos por los id de las dimensiones para posteriormente poder agrupar el consumo eléctrico y de gas licuado por mes e insertarlos en la base de datos mart_comedores. En caso de encontrar alguna inconsistencia, los datos son guardados en un fichero Excel para el tratamiento de los errores, con una descripción de los mismos para que el cliente pueda corregirlos.

La figura muestra un ejemplo de una de las transformaciones para la carga de los hechos (Ver figura 12). Para ver más figuras de la implementación de los hechos debe remitirse al anexo 2.

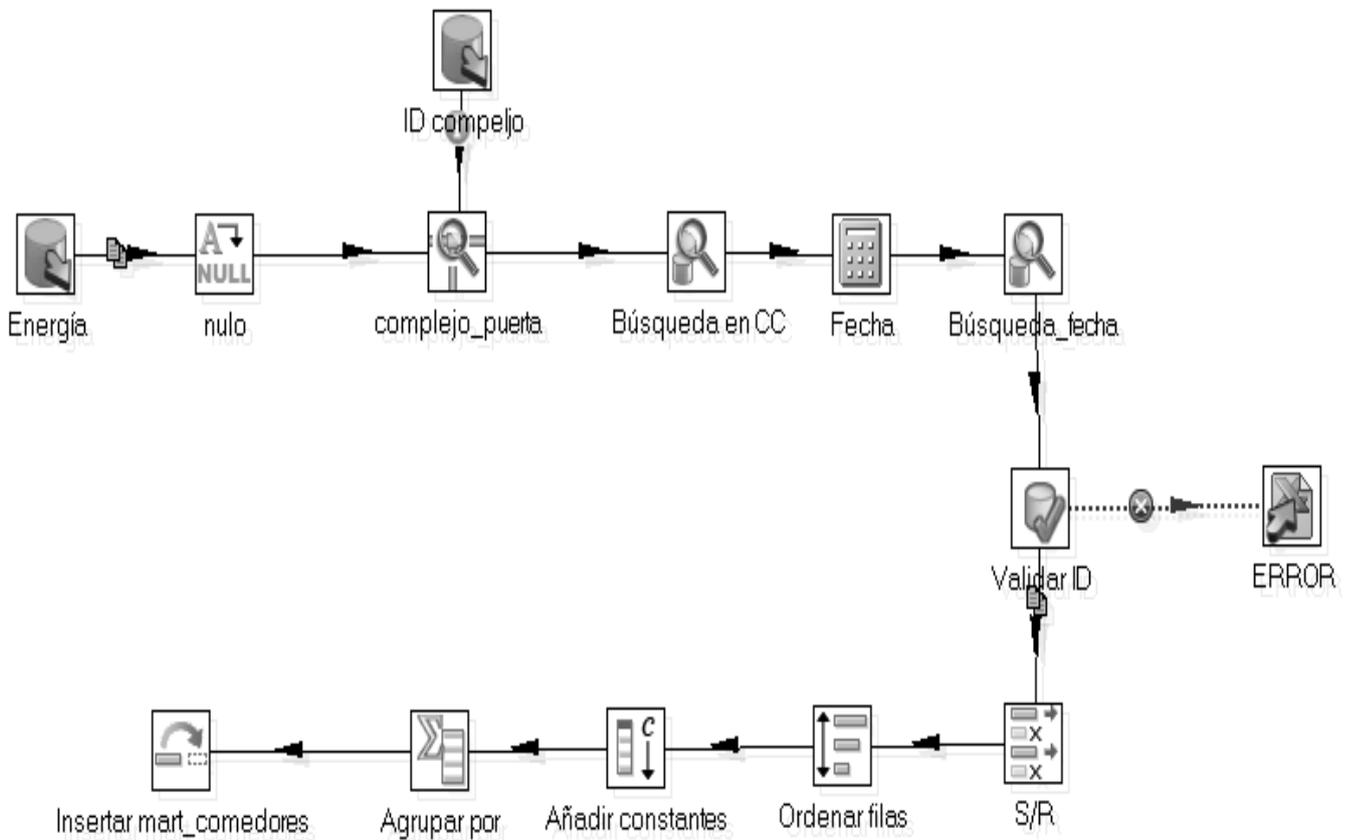


Figura 12. Transformación para la carga del hecho hech_energia.

Implementación de los trabajos

Una vez realizadas todas las transformaciones de las dimensiones y los hechos se implementan las transformaciones de los trabajos para poblar la base de datos. Se implementó un trabajo principal (Ver figura 13) que verifica las conexiones con el servidor de base de datos y el área temporal de donde provienen los datos fuente y manda a ejecutar el trabajo general. Este último ejecuta en un orden lógico las transformaciones de las dimensiones y después la de los hechos (Ver figura 14). Para ver más figuras de la implementación de los trabajos debe remitirse al anexo 2.



Figura 13. Transformación del trabajo principal.



Figura 14. Transformación del trabajo general.

3.3 Implementación del subsistema de visualización

Implementación de los cubos OLAP

Para la implementación de los cubos, se definieron las dimensiones, las medidas y los niveles de jerarquía de cada dimensión. Se creó el esquema Comedores y se implementó un cubo por cada tabla de hecho y un cubo virtual que agrupa el hech_control, el hech_raciones_elaboradas_consumidas y las dimensiones comunes para ambos hechos (Ver figura 15).

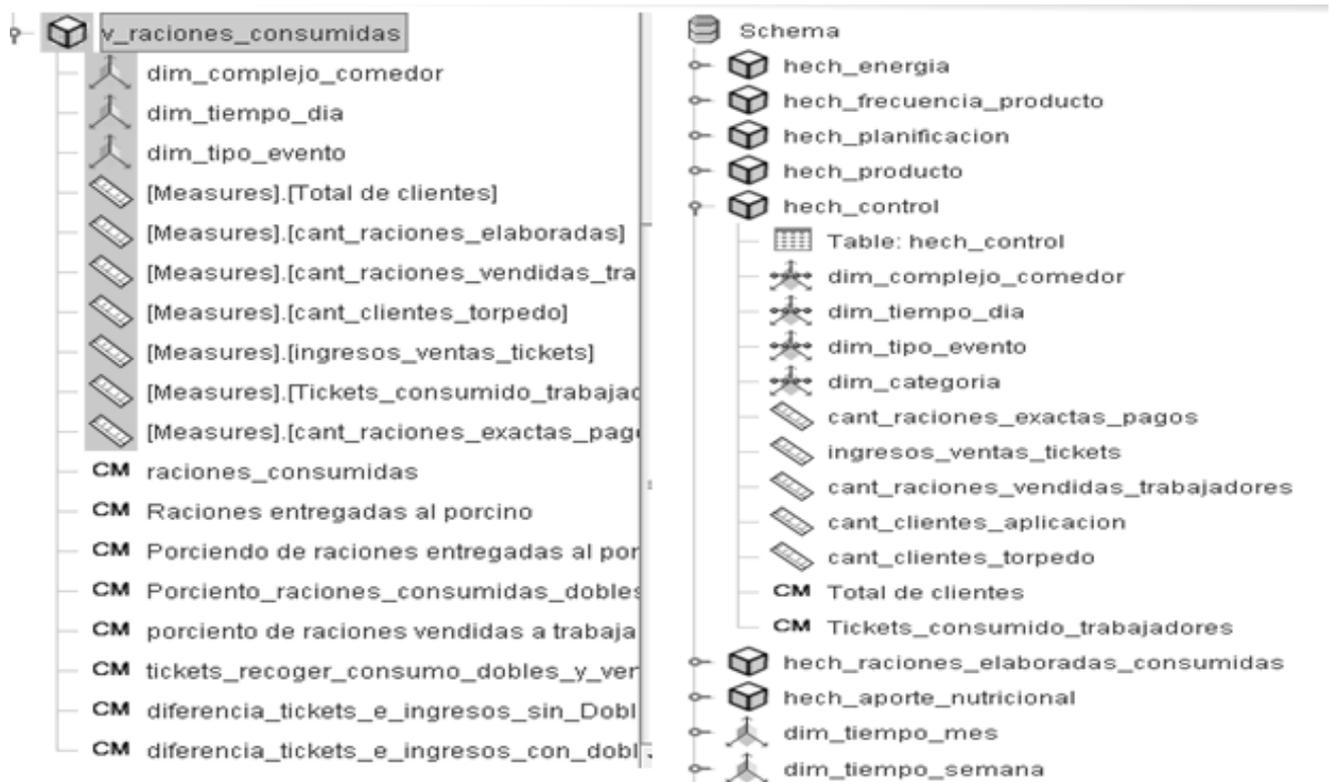


Figura 15. Implementación de los cubos OLAP.

Arquitectura de información

En el capítulo 2 y como parte del diseño del subsistema de visualización, se describió la forma en que quedaría concebida la arquitectura de información a través del mapa de navegación del mercado de datos. A continuación se especifican cada uno de los elementos que componen las estructuras de navegación de la información. De esta manera será presentada la capa de visualización del Mercado de datos Comedores; la misma contiene un Área de Análisis Comedores (A.A Comedores), seis Libros de Trabajos (L.T) (Ver figura 16) y 31 reportes multidimensionales diseñados anteriormente.

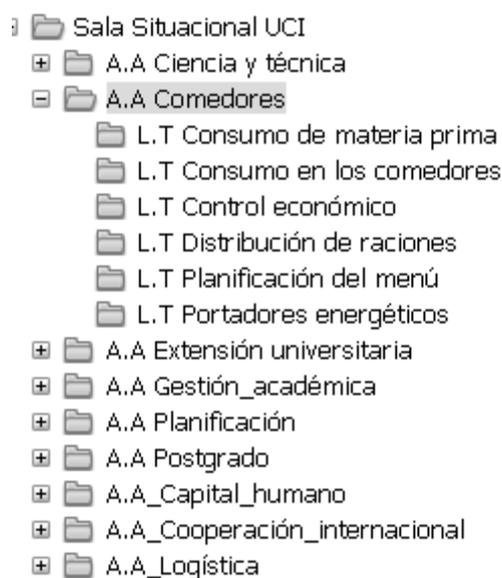


Figura 16. Arquitectura de información del Mercado de datos Comedores.

Descripción del Área General (A.A.G)

A.A.G Sala_Situacional_UCI: agrupa toda la información de todos los mercados de datos de la UCI, formando el almacén ssuci_dhw.

Descripción del Área de Análisis (A.A)

A.A Comedores: agrupa toda la información referente a los indicadores del área de Comedores y contiene los libros de trabajo donde se encuentran los reportes asociados al área.

Descripción de los Libros de Trabajo (L.T)

L.T Consumo en los comedores: agrupa la información referente al acceso de clientes a los eventos del comedor, contiene once reportes que permiten realizar un análisis general de los datos.

L.T Portadores energéticos: agrupa la información referente a los indicadores de energía, como son el gas y la corriente eléctrica, contiene dos reportes que permiten realizar un análisis general de los datos.

L.T Consumo de materia prima: agrupa la información referente a los productos que son utilizados para la elaboración de los platos del menú, contiene un reporte que permite realizar un análisis general de los datos.

L.T Control económico: agrupa la información referente a los ingresos por ventas de tickets, contiene cinco reportes que permiten realizar un análisis general de los datos.

L.T Distribución de raciones: agrupa la información referente al consumo y elaboración de raciones en los complejos comedores, contiene seis reportes que permiten realizar un análisis general de los datos.

L.T Planificación del menú: agrupa la información referente a los platos reales y planificados en cada evento, así como su aporte nutricional, contiene seis reportes que permiten realizar un análisis general de los datos.

Creación y administración de los reportes

En la implementación de los reportes se utilizaron consultas MDX las cuales son administradas por el Pentaho BI Server. MDX un lenguaje de consulta a estructuras multidimensionales (cubos OLAP), su sintaxis es parecida al SQL (Lenguaje de Consulta Estructurada) para base de datos relacionales, con la diferencia de que no maneja los términos tablas y campos, si no cubos, dimensiones, medidas y jerarquías. A continuación se muestra la consulta del reporte “Consumo de electricidad en Kw/h por complejo comedor”, la cual devuelve el consumo de electricidad en kW/h de cada complejo comedor mensualmente; se muestra además el reporte en forma de tabla y un gráfico dinámico que visualiza el consumo mensual por cada uno de los complejos comedores (Ver Figura 17, 18 y 19). Para ver más figuras de las vistas de análisis y gráficos dinámicos debe remitirse al anexo 3.

Editor de consultas MDX

MDX Query Editor

```
select NON EMPTY {[Measures].[cant_total_consumo_electricidad]} ON COLUMNS,
  NON EMPTY Crossjoin(Hierarchize([dim_tiempo_mes].[Todos].Children), [dim_complejo].
[Todos].Children) ON ROWS
from [hech_energia]
```

Figura 17. Consulta MDX del consumo de electricidad.

Capítulo 3: Implementación del Mercado de datos Comedores

		Medidas
Tiempo	Complejo	• Cantidad de consumo de electricidad en Kw/h
2010	CC1	6350310
	CC2	8434264
	CC3	6325206
Septiembre	CC1	2931498
	CC2	3898024
	CC3	2923314
Octubre	CC1	3433278
	CC2	4531744
	CC3	3398454

Figura 18. Reporte Consumo de electricidad por complejo comedor.

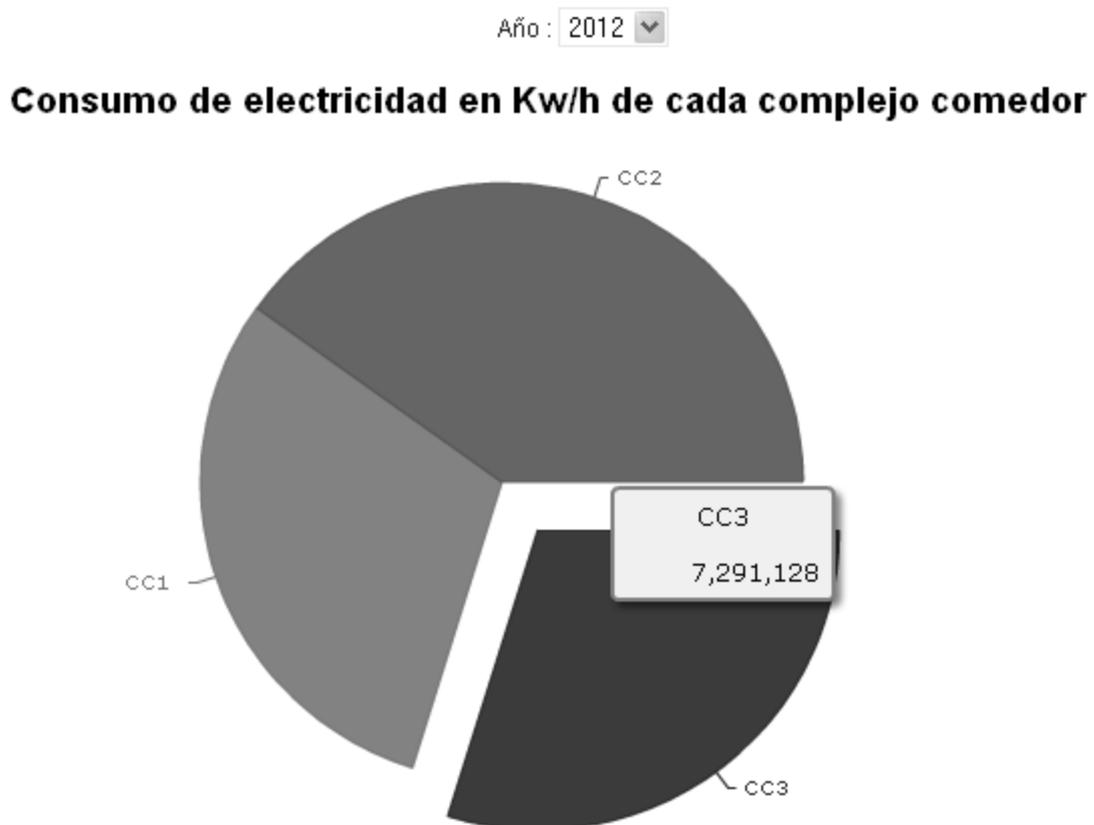


Figura 19. Gráfico dinámico del consumo de electricidad en Kw/h por año.

Capítulo 3: Implementación del Mercado de datos Comedores

Conclusiones del capítulo

La implementación de los tres esquemas: dimensiones, mart_comedores y metadatos, las 17 transformaciones, ocho cubos y 31 reportes del mercado de datos Comedores permitió definir la estructura de los datos a partir del modelo físico, la carga de los datos sin inconsistencias hacia el mercado de datos y la disponibilidad de la información de manera organizada, facilitando el proceso de toma de decisiones.

Capítulo 4: Pruebas al Mercado de datos Comedores

Introducción

Después de realizado el análisis, diseño e implementación del Mercado de datos Comedores, se procede a probar la solución mediante las listas de chequeo, los casos de prueba y las pruebas de aceptación, que garantizan al usuario final del sistema la confiabilidad de los datos cargados en el almacén y que el producto cumpla con los requisitos establecidos por él mismo.

4.1 Pruebas de software

Con el fin de lograr que el mercado cumpla con las especificaciones del negocio se emplea el Modelo V, utilizado por el Centro de Tecnología de Gestión de Datos (DATEC), para crear un estándar y comprobar que el producto cumpla con las especificaciones del negocio. La fase de análisis y diseño se encuentra a la izquierda del modelo y se relaciona con las pruebas que se encuentran a la derecha a través de los planes de pruebas. Este modelo permite que cada vez que se termine una etapa de desarrollo se le apliquen pruebas, para garantizar que el producto no pase con errores a una nueva fase. En la siguiente figura queda estructurado el Modelo V.

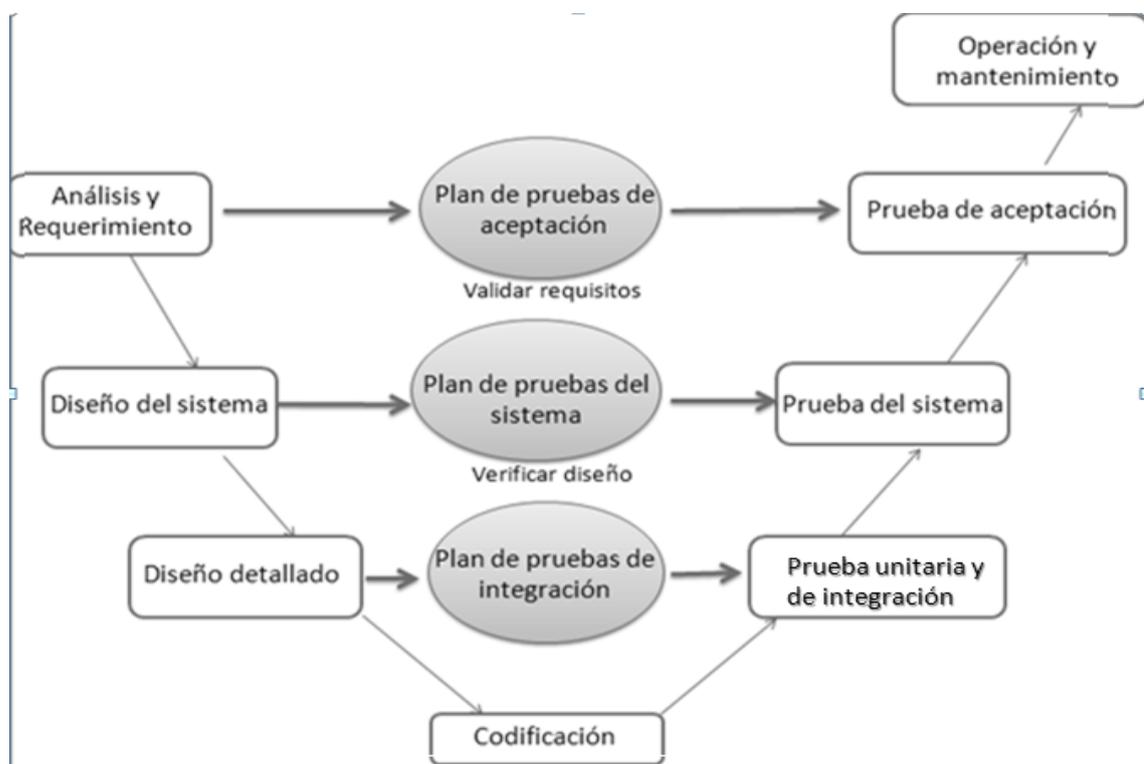


Figura 20. Modelo V.

- **Pruebas unitarias:** estas pruebas son llamadas también pruebas modulares, debido a que permiten determinar si un módulo del programa está listo y correctamente terminado. [28]
- **Pruebas de integración:** se realizan una vez que las pruebas unitarias fueron concluidas exitosamente, su objetivo consiste en probar el sistema como un todo, para detectar errores asociados con la interacción entre los diferentes módulos que lo componen. [28]
- **Pruebas de sistema:** son las pruebas que se realizan para determinar el correcto funcionamiento de un sistema y su cumplimiento con las especificaciones del producto. [28]

4.2 Herramientas para la aplicación de las pruebas

Para realizar las pruebas al Mercado de datos Comedores se utilizaron las siguientes herramientas:

Casos de prueba: el propósito de un caso de prueba es especificar una forma de probar el sistema, incluyendo las entradas con las que se ha de validar, los resultados esperados y las condiciones bajo las que ha de probarse. En la presente investigación se diseñaron seis casos de prueba, uno por cada caso de uso de información, que fue identificado en la etapa de análisis. En la siguiente tabla se muestra un fragmento del caso de prueba correspondiente al CU “Presentar información de los portadores energéticos”, donde se aprecia el primer reporte correspondiente al LT Portadores energéticos con una breve descripción de lo que muestra, además de los indicadores que se miden en el reporte, la respuesta que debe dar el sistema y el flujo central de ejecución del caso de prueba (Ver tabla 9).

Tabla 9. Caso de prueba basado en el caso de uso Presentar información de los portadores energéticos.

Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central
EC 1.1 Presentar información de los portadores energético.	Muestra el consumo de gas y electricidad por cada complejo comedor.	Tiempo. Complejo comedor.	Cantidad total de consumo de gas. Cantidad total de consumo de electricidad.	El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que	Se autentifica. Se entra al sistema. Se despliega hacia la derecha el componente ubicado en el lateral

Capítulo 4: Pruebas al Mercado de datos Comedores

				<p>pueden ser visualizadas para cada reporte.</p>	<p>izquierdo que contiene el navegador.</p> <p>Se selecciona el área de análisis de AA. Comedores</p> <p>Se selecciona el libro de trabajo</p> <p>LT. Portadores energéticos</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte.</p>
--	--	--	--	---	---

Con el propósito de validar los artefactos desarrollados a partir de un grupo de indicadores, se utilizaron varias listas de chequeo, que fueron previamente definidas por el centro.

Listas de chequeo: son un listado de preguntas, en forma de cuestionario que verifican y determinan la calidad de los artefactos generados en el proceso de ETL. La evaluación se desarrolla a través del análisis de un grupo de indicadores, distribuidos en 3 secciones fundamentales:

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.

Capítulo 4: Pruebas al Mercado de datos Comedores

- **Indicadores definidos:** abarca todos los indicadores a evaluar durante la etapa de desarrollo del mercado.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

La estructura de la lista de chequeo está formada por los siguientes elementos:

- **Peso:** define si el indicador a evaluar es crítico o no. El mismo se define con una C si es crítico.
- **Indicadores a evaluar:** son los indicadores a evaluar en las secciones Estructura del documento, Semántica del documento e Indicadores definidos para la etapa a evaluar.
- **Evaluación (Eval):** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de uno en caso de que exista alguna dificultad sobre el indicador y de cero, en caso de que el indicador revisado no presente problemas.
- **No Procede (NP):** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- **Cantidad de elementos afectados (CEA):** especifica la cantidad de errores encontrados sobre el mismo indicador.
- **Comentario (Comt):** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo. Pueden o no existir señalamientos o sugerencias.

A continuación se muestra en la tabla 10 una de las listas de chequeo utilizadas:

Tabla 10. Lista de chequeo para el artefacto de ETL Diccionario de datos.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	NP	CEA	Comt
crítico	1. ¿El entregable contiene las secciones obligatorias de la plantilla estándar definida para el expediente de proyecto?	0			
crítico	2. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0			

Capítulo 4: Pruebas al Mercado de datos Comedores

crítico	3. ¿El objetivo expresa correctamente el propósito del documento?	0			
	4. ¿Se hace un uso adecuado del control del documento?	0			
	5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	1			
	6. ¿En el entregable, la definición de las variables se hace correctamente?	0			
	7. ¿Existe una adecuada correspondencia entre las variables definidas y las descripciones que tienen estas variables?	0			
	8. ¿En el entregable se crea una hoja por cada variable definida?	0			
	9. ¿Queda registrado en el entregable todos los posibles valores que van a tener las variables definidas?	0			
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	NP	CEA	Comt
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0			
Semántica del documento					
Peso	Indicadores a evaluar	Eval	NP	CEA	Comt

crítico	1. ¿Se han identificado errores ortográficos en el entregable?	1			
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0			
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?		NP		

4.3 Resultados de las pruebas

Pruebas unitarias e integración

Terminada la implementación se realizaron pruebas de unidad a los flujos de integración de datos y a los diferentes componentes relacionados con el subsistema de visualización. Para el proceso de estas pruebas se realizaron dos iteraciones donde se encontraron un total de nueve no conformidades que fueron resueltas satisfactoriamente. A nivel de desarrollador se hizo una primera iteración donde se detectaron dos no conformidades que son mencionadas a continuación:

- No se especifican los campos a cargar para los ficheros Excel de errores.
- Poner nombres sugerentes a los pasos de las transformaciones.

Para una segunda iteración realizada por los especialistas del proyecto fueron detectadas las siguientes no conformidades:

- Los nomencladores de la dimensión Categoría no se muestran correctamente (Interno y Externo).
- Falta implementar el reporte de consumo energético.
- No poseen la implementación de los trabajos para las dimensiones y hechos.
- Falta la implementación de algunas transformaciones para la carga de los hechos.
- Utilización incorrecta del paso Insertar/Actualizar para la carga de los hechos.

Todas en su conjunto fueron resueltas satisfactoriamente en un corto período de tiempo.

A continuación se muestra una gráfica donde se reflejan las iteraciones realizadas en el proceso de prueba.

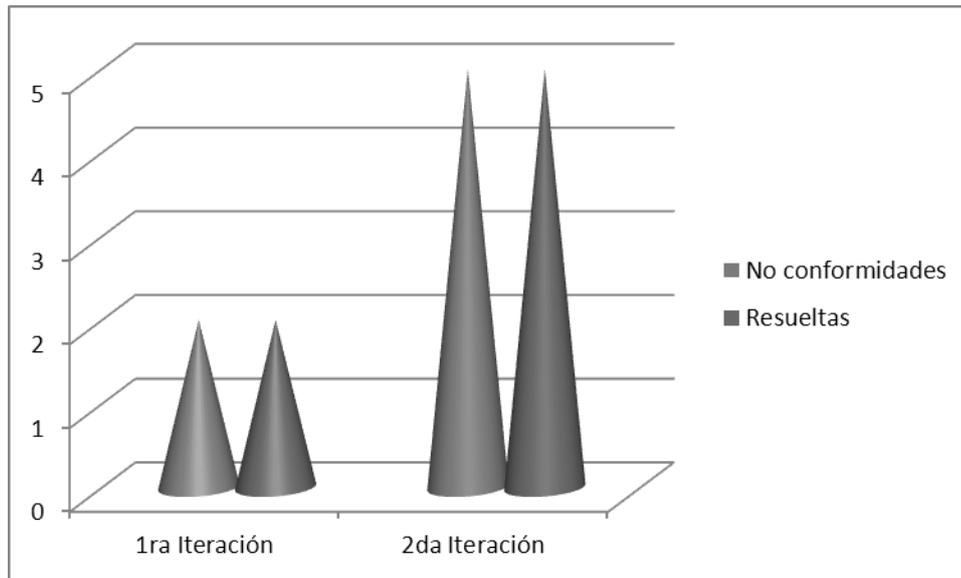


Figura 21. Comportamiento de las no conformidades identificadas en las pruebas unitarias y de integración.

Pruebas de sistema

Las pruebas de sistema fueron realizadas por un grupo de especialistas de calidad del centro DATEC, los mismos se basaron en los casos de prueba diseñados para cada uno de los casos de uso de información correspondientes al mercado de datos. Una vez realizadas se detectaron cinco no conformidades relacionadas con los nombres de las variables de entrada y de salida que no correspondían con el nombre que se mostraba en la aplicación, estas no conformidades fueron resueltas en el tiempo establecido. Posteriormente se realizó una segunda iteración donde se comprobó el cumplimiento con las no conformidades anteriores y que la aplicación funcionara correctamente. En esta segunda iteración se obtuvieron resultados satisfactorios dando paso a que se firmara la carta de liberación interna del producto por parte de los especialistas del proyecto.

Listas de chequeo

Luego de aplicada la lista de chequeo a cada uno de los artefactos de ETL se detectaron dos no conformidades que fueron resueltas satisfactoriamente en un corto período de tiempo. La gráfica que a continuación se muestra representa el comportamiento de los nueve indicadores medidos, donde cinco de ellos son críticos, así como las no conformidades encontradas después de su aplicación.

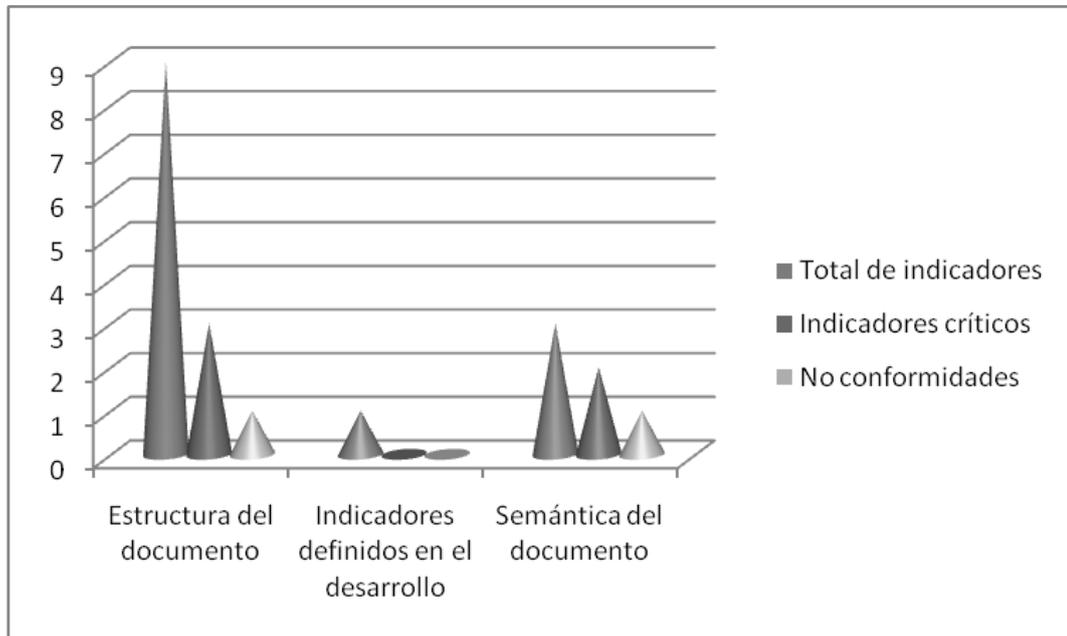


Figura 22. Comportamiento de los indicadores de la lista de chequeo.

Conclusiones del capítulo

Las pruebas de unidad, de integración y de sistema realizadas en el capítulo permitieron encontrar no conformidades que finalmente fueron resueltas garantizando el funcionamiento del mercado de datos.

Las herramientas listas de chequeo y los casos de prueba utilizadas para las pruebas permitieron identificar errores en los artefactos de ETL y verificar el cumplimiento de los objetivos definidos por el cliente.

Conclusiones generales

Una vez concluida la investigación se puede afirmar que se desarrolló el Mercado de datos Comedores para la Sala Situacional de la Universidad de las Ciencias Informáticas, dándole cumplimiento a los objetivos trazados, concluyendo que:

- Mediante la investigación realizada sobre las tecnologías de almacenamiento de datos se determinó que los mercados de datos son una de las soluciones más aceptadas para ayudar a la toma de decisiones en el área de los comedores de la UCI. Así como el estudio de las metodologías y las herramientas permitieron seleccionar la metodología y las herramientas para guiar y dar cumplimiento al desarrollo de la solución.
- El análisis del negocio del área de los comedores permitió definir 22 requisitos de información y 15 requisitos funcionales que fueron agrupados en seis y siete casos de uso respectivamente que contribuyeron a la obtención del diagrama de casos de uso del sistema. El diseño del subsistema de almacenamiento permitió identificar siete hechos, diez dimensiones y 25 medidas dando lugar al modelo lógico y físico de la base de datos.
- La implementación del subsistema de integración y visualización, permitió integrar los datos del área temporal al mercado de datos Comedores, con la implementación de las 17 transformaciones y se obtuvo la capa de visualización de los datos capaz de visualizar los indicadores referentes al área de los comedores para apoyar el proceso de toma de decisiones, con la implementación de los ocho cubos OLAP y los 31 reportes candidatos.
- Las pruebas unitarias, de integración y de sistema realizadas al mercado de datos Comedores permitieron encontrar no conformidades que con la corrección de las mismas se garantizó que el producto cumpliera con las especificaciones del negocio.

Recomendaciones

Se recomienda implementar técnicas de Minería de datos que posibiliten mejorar el proceso de toma de decisiones.

Se sugiere que se ponga en explotación el mercado de datos para tener disponibles el conjunto de indicadores asociados al área de los Comedores, para ayudar a la máxima dirección de la universidad en la toma de decisiones.

Referencias Bibliográficas

1. Willam H. Inmon: Building the Data Warehouse, Technical Publishing Group, 1992.
2. R Kimball: The Data Warehouse Toolkit. Ed. John Wiley, 1996.
3. ESPINOSA, ITZIAR ANGOITIA.DATA WAREHOUSE PARA LA GESTIÓN DE LISTA DE ESPERA SANITARIA. Madrid : s.n.
4. Evelia, Casales Cabrera María. 1 de 2009.
5. Rodríguez Sanz, Miguel y Mingo Postiglioni, Jack Mario. ANÁLISIS Y DISEÑO DE UN DATA MART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO. Universidad Carlos III de Madrid Escuela Politécnica Superior Ingeniería en Informática. Julio, 2010.
6. Dario, Ing. Bernabeu Ricardo.DATA WAREHOUSING:Investigación y Sistematización de Conceptos HEFESTO: Metodología propia para la Construcción de un Data Warehouse. Córdoba, Argentina : s.n., Abril, 2009.
7. INBOT-DIET. [En línea] 1998 © 2007. [Consultado el: 7 de 11 de 2011.] Disponible en: <http://www.club5estrellas.com/inbody/nutricion.htm>.
8. soft nutricional interactivo. [En línea] [Consultado el: 8 de 11 de 2011.] Disponible en: http://www.equilibra.enlinea.ec/profesional/vista_previa.html.
9. Flores Coronado, Adisley y Rovira Prieto, Imirys. Junio, 2007. Disponible en: http://bibliodoc.uci.cu/TD/TD_0356_07.pdf .
10. González Hernández, Yanisbel. PROPUESTA DE METODOLOGIA PARA EL DASARROLLO DE ALMACENES DE DATOS EN DATEC. . Ciudad de La Habana Septiembre, 2011. : Universidad de las Ciencias Informática.
11. Sistemas De Bases de Datos. Segunda Edición, Rosa María Mato García, 2005. Tema 1 y 2.
12. Catarí, Ayala. Romero, Iván. Estudio de herramientas CASE de soporte UML y UML2. Universidad Autónoma “Juan Misael Saracho”. Facultad de Ciencias y Tecnología Tarija - Bolivia. Disponible en: <http://es.scribd.com/doc/25374125/Estudio-de-Herramientas-CASE-de-Soporte-a-UML-y-UML2>.
13. Menéndez, Evelyn Alonso. Herramientas CASE para el proceso de desarrollo de Software. Disponible en: <http://www.monografias.com/trabajos73/herramientas-case-proceso-desarrollo-software/herramientas-case-proceso-desarrollo-software2.shtml>.
14. López, Patricia. Univ. Cantabria – Fac. de Ciencias. Trabajando con Visual Paradigm for UML. Disponible en: <http://personales.unican.es/ruizfr/is1/doc/lab/01/is1-p01-trans.pdf>.

15. Zeis, C.; Ruel, C.; Wessler, M. Oracle® 11g for Dummies®. Indianapolis, Indiana. Wiley Publishing, Inc. 2009. 412.
16. Riggs, S.; Krosing, H. PostgreSQL 9. Administration Cookbook. Birmingham, E.E. U.U. Packt
17. Erich Mario Gómez Perez, Arel Torres Galvez. Administración y Optimización de un sistema de Base de Datos Descentralizado, en Postgres SQL. 2008.
18. Tomo Revista Técnica de la Empresa de Telecomunicaciones de Cuba, S.A. No. 1, s.l. : CICT Impresión: Exclusivas Latinoamericanas ELA, SL , 2010 , Vol. Vol. 7.
19. Sinnexus. [En línea] 2007 – 2011 . [Consultado el: 28 de 11 de 2011.]. Disponible en: http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
20. Medina Mustelier, Doris. Técnicas de Extracción, Transformación y Carga de Datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela. Universidad de las Ciencias Informáticas. Marzo, 2009.
21. W. H. Inmon. The Evolution of Integration
22. Gravatar. Pentaho Data Integration. [En línea]. [Consultado el: 20 de 12 de 2011] Disponible en: <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
23. DataCleaner. [En línea]. [Consultado el: 28 de 11 de 2011.] Disponible en: <http://datacleaner.eobjects.org>.
24. Pentaho BI Plataform Server. [En línea] [Consuldado el: 12 de 01, 2011.] Disponible en: <http://www.pentaho.com/>
25. PENTAHO: Creación de un cubo con Mondrian. . [En línea] [Consultado en: 11 24, 2011.] .Disponible en: <http://www.gravatar.biz/index.php/bi/pentaho-ejemplo-cubo-mondrian/>.
26. García Molina, Jesús. M. José Ortín. Moros, Begoña, Joaquín Nicolás. Toval, Ambrosio. De los Procesos Del Negocio a los Casos de Uso. Grupo de Investigación de Ingeniería del Software2. Departamento de Informática y Sistemas. Facultad de Informática. Universidad de Murcia http://www.willydev.net/descargas/willydev_modeladodenegocio.pdf
27. JACOBSON, Ivar; BOOCH, Grady; RUMBAUGH, James. “El Proceso Unificado de Desarrollo de Software”. 2000. Addison Wesley. Capítulos 7, 8 páginas 125-163, 187-202.
28. Ricardo Mansilla. Control de calidad de software. Pruebas de software. Nov, 2009. <http://www.slideshare.net/cliceduca/pruebas-de-software-2420588>.

Bibliografía

1. CABRERA, M. R. y GONZÁLEZ, Y. D. L. C. V. Análisis, diseño e implementación del mercado de datos para los indicadores específicos de la ocupación para la Oficina Nacional de Estadística. Universidad de las Ciencias Informáticas, 2010.
2. CEJAS, A. H. Sistema de gestión integral de la planificación y control del servicio de comedores UCI: análisis y diseño del módulo “Gestión de Comensales”. Universidad de las Ciencias Informáticas, 2008.
3. CLEANER, D. DATA CLEANER [Consultado el: 15 /11/2011 Disponible en: <http://datacleaner.eobjects.org/>].
4. DATACLEANER. DataCleaner. nº [Consultado el: 28 /11/ 2011]. Disponible en: <http://datacleaner.eobjects.org>.
5. DATAMART. DATAMART. [Consultado el: 13 /11 / 2011 Disponible en: <http://www.dataprix.com/datawarehouse-vs-datamart>].
6. DESVENTAJAS, D. V. Y. DATAMART. Ventajas y Desventajas [Consultado el: 15 / 11 /2011 Disponible en: www.datawarehouse.com].
7. DOMÍNGUEZ, R. R. E. Desarrollo de un Almacén de Datos para el Control del Consumo de Portadores Energéticos en la Oficina Nacional de Estadística. Universidad de las Ciencias Informáticas, 2010.
8. EOBJECTS.ORG. eobjects.org [Consultado el: 28/ 5 /2012 Disponible en: <http://eobjects.org/trac/blog/change-in-preferred-license>].
9. FERNÁNDEZ, L. P. y HERNÁNDEZ, A. R. Diseño de un Almacén de Datos para apoyar la toma de decisiones en el Centro de Informática Médica. Universidad de las Ciencias Informáticas, 2010.
10. GALLARDO, Y. D. L. S. y BRITO, B. W. Extracción, transformación y carga del mercado de datos CIMAvax EGF para el almacén de datos del Centro de Inmunología Molecular. Universidad de las Ciencias Informáticas, 2011.
11. GARCÍA, R. V. y TRABAJO, O. H. Diseño e Implementación de la Base de Datos del Sistema de Gestión Integral de la Planificación y Control del Servicio de Comedores UCI. 2008, nº [Consultado el: 23 /11 / 2011]. Disponible en: http://repositorio_institucional.uci.cu/jspui/handle/ident/TD_1798_08.
12. GEIGER, C.; GALEMMO, N., et al. Mastering Data Warehouse Design. Relational and Dimensional Techniques. publicado el: 12 /11 / 2011 de 2008, última actualización: 12 /11 / 2011.
13. INTEGRATION, D. Using ETL. EAI and EII Tools to Create an Integrate Enterprise [Consultado el: 12/ 11 / 2011 Disponible en: <http://www.bi-bestpractices.com/view-articles/4737>].

14. INTELLIGENCE, B. Business Intelligence fácil: DSS: Tipos de decisiones empresariales [Consultado el: 10 / 11 /2011 Disponible en: <http://www.businessintelligence.info/dss/toma-decisiones-business-intelligence.html>.
15. KIMBALL, R. y ROSS, M. The Data Warehouse 2da ed. Canadá: 2002.
16. MARTINEZ, R. Lanzamiento de PostgreSQL 9.1 [Consultado el: 25 / 4 /2012 Disponible en: <http://www.postgresql.org.es>.
17. MESA, V. B. Extracción, transformación y carga del mercado de datos Racotumumab para el almacén de datos del Centro de Inmunología Molecular. Universidad de las Ciencias Informáticas, 2011.
18. MILANÉS, Y. R. y JUAN, Y. I. S. Mercado de Datos Agricultura, ganadería y silvicultura para el Sistema de Información de Gobierno. Universidad de las Ciencias Informáticas, 2011.
19. MORA, S. L. Data Warehouse Design with UML. España: publicado el: 13 /11 / 2011 de 2005, última actualización: 13 /11 / 2011.
20. MUSTELIE, D. M. Técnicas de Extracción, Transformación y Carga de Datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela. Universidad de las Ciencias Informáticas, 2009.
21. OCHOA, D. G. Diseño e Implementación de un Almacén de Datos Operacionales para la corporación CIMEX. Universidad de las Ciencias Informáticas, 2010.
22. ORACLE. Oracle [Consultado el: 5/2/2010 Disponible en: <http://www.oracle.com>.
23. OSORIO, Y. O. y RODRÍGUE, A. G. Sistema de gestión integral de la planificación y control del servicio de comedores UCI: análisis y diseño del módulo “Administración, análisis económico y gestión de aseguramientos”. Universidad de las Ciencias Informáticas, 2008.
24. PENTAHO. Creación de un cubo con Mondrian [Consultado el: 24 /11/2011 Disponible en: <http://www.gravitar.biz/index.php/bi/pentaho-ejemplo-cubo-mondrian/>.
25. POWER, D. J. Decision support systems: concepts and resources for managers. 2002.
26. PRESSMAN, R. Ingeniería de Software un enfoque práctico. 5ta ed.
27. SÁNCHEZ, L. Z. Metodología para el Diseño Conceptual de almacenes de datos. Valencia: 2008.
28. SIERRA, D. Exposición Herramientas Automatizadas IPN UPIICSA [Consultado el: 26 / 5/ 2012 Disponible en: <http://www.slideshare.net/vanquishdarkenigma/visual-paradigm-for-uml>.
29. VISUAL-PARADING.COM. Visual Parading para UML [Consultado el: 28 /11 / 2011 Disponible en: <http://www.google.com/http://www.visual-paradigm.com/product>.

Glosario de términos

Comensales: usuarios que asisten a los eventos del comedor.

Matriz bus: representación de las relaciones existentes entre los hechos y las dimensiones en forma de matricial.

No conformidades: defecto, error o sugerencia que se le hace al equipo de desarrollo cuando se encuentra alguna dificultad en lo que se está evaluando.

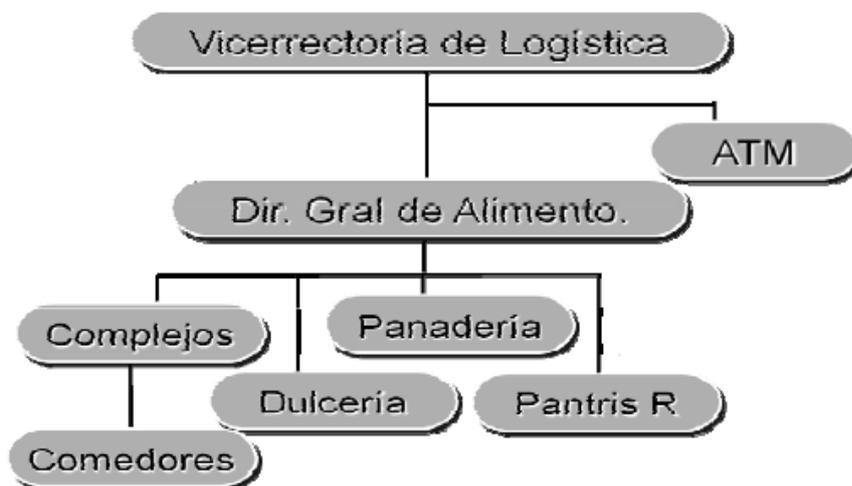
Torpedo: Excel donde se registran los usuarios que acceden a los eventos del comedor sin presentar solapín por causas de pérdida o deterioro de este último.

Raciones exentas de pago: son las raciones que se pagan en los pantry.

Pantry: son comedores que brindan servicios a los directivos de la universidad y visitantes.

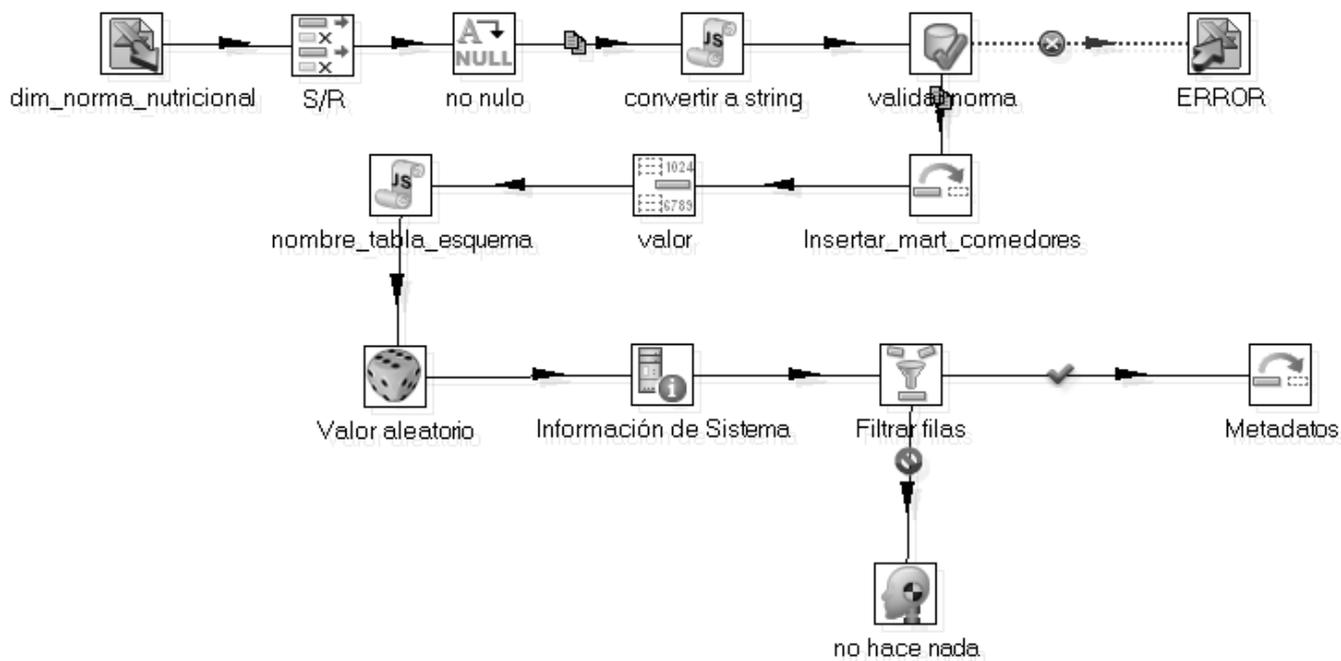
Anexos

Anexo 1: Organigrama de la Vicerrectoría de Logística

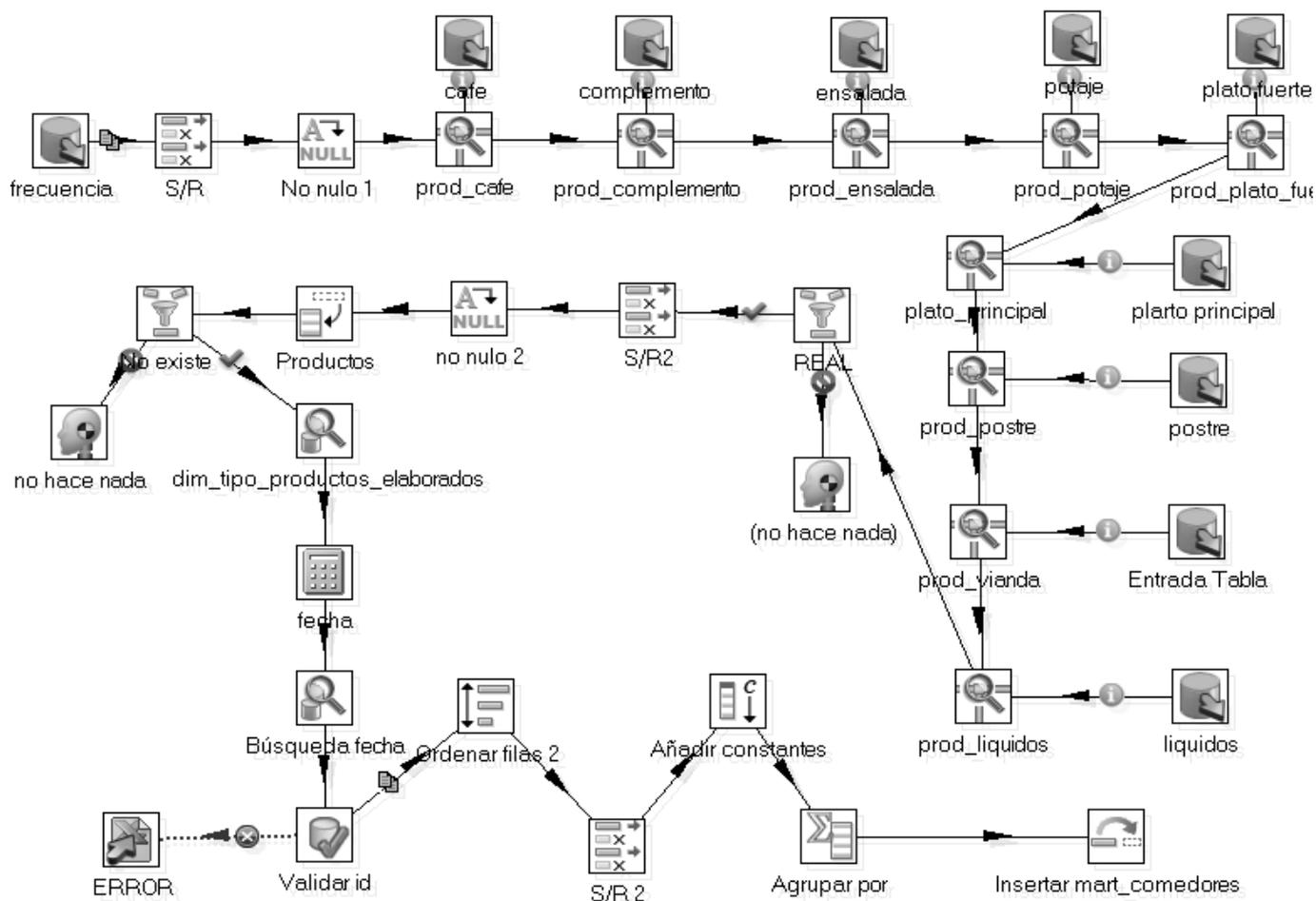


Anexo 2

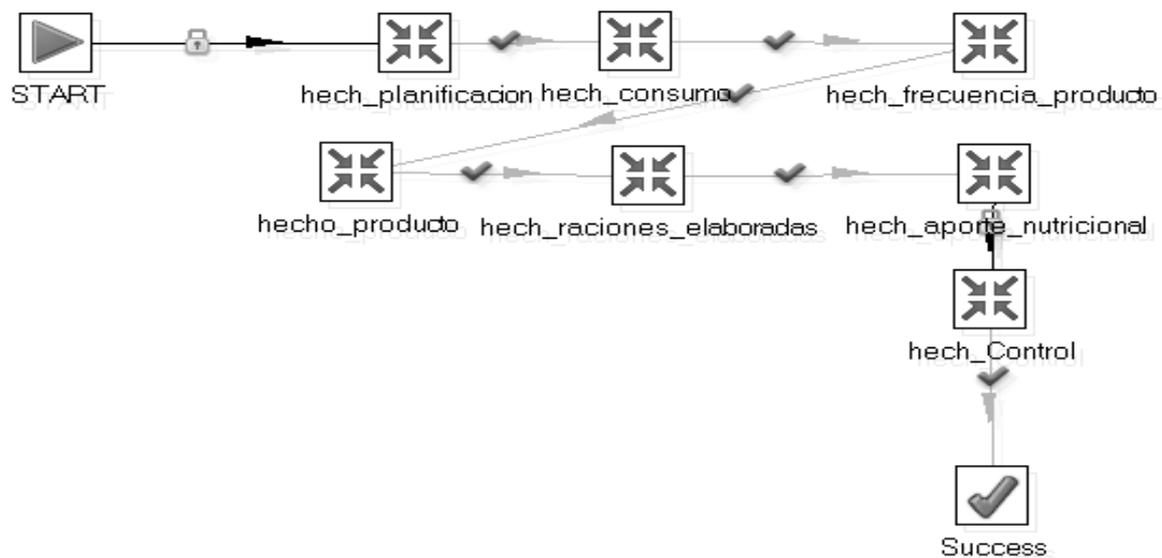
Transformación para la carga de la dimensión estática dim_tiempo_dia



Transformación para la carga del hecho frecuencia_producto



Transformación del trabajo que ejecuta los hechos



Anexo 3: Vistas de análisis y Gráficos dinámicos

Tiempo	Evento	Producto en bruto	Unidad de medida	Medidas
				Cantidad total de productos en bruto
2010	DESAYUNO	LECHE EN POLVO	Kilogramo(s)	155558
		CAFE	Kilogramo(s)	237954
		LECHE LIQUIDA EN CAJA	Unidades	192144
		REFRESCO INSTANTANEO	Kilogramo(s)	188870
		REFRESCO SIROPE	Litro(s)	162512
		POMO DE AGUA	Unidades	209274
		QUESO FUNDIDO	Kilogramo(s)	182602
		MOSTAZA	Kilogramo(s)	150174
		MANTEQUILLA	Kilogramo(s)	173374
		CHOCOLATE	Kilogramo(s)	188670
		QUESO CREMA	Kilogramo(s)	204454
		MAYONESA	Kilogramo(s)	238258
		REFRESCO DE POMO	Unidades	199494
		YOGURT	Unidades	159568
		PAN	Unidades	129430

