

“Desarrollo de Data Warehouse Comercial con Software de Control de Stock y Reportes”



Autores:

Flores, Diego Alejandro

Martínez Chirichian, Rodrigo

Tutor: Ciolli, María Elena

Proyecto de Grado

Ingeniería en Sistemas

Instituto Universitario Aeronáutico

Índice

Introducción.....	6
Empresa.....	7
1.1 Presentación de la empresa.....	7
Nombre de la empresa: Strega.....	7
Rubro de la empresa:.....	7
Misión – Visión - Valores.....	7
1.2 Organigrama:.....	9
1.3 Descripción de áreas funcionales:.....	9
1.4 Planteamiento del problema.....	10
1.5 Objetivos de investigación.....	13
1.5.1 Objetivos generales.....	13
1.5.2 Objetivos específicos.....	14
1.6 Justificación del estudio.....	14
1.7 Limitaciones de la investigación.....	15
Desarrollo de Software “Stockia”:	16
2.1. Descripción de apostadores y usuarios.....	16
2.1.1. Información demográfica del mercado.....	16
2.1.2. Resumen de apostadores.....	16
2.1.3 Resumen de usuarios.....	16
2.1.4. Entorno de usuarios.....	17
2.1.5. Lay-Out de las Sucursales.....	17
2.1.6. Tecnologías Disponibles.....	17
2.1.7. Perfil de Apostadores.....	18
2.1.8 Perfil de Actores.....	19
2.1.9 Necesidades de usuarios.....	19
2.1.10 Alternativas y competencia.....	20

2.2. Descripción del Producto	20
2.2.1 Perspectiva del producto	20
2.2.2 Supuestos y dependencias	21
2.2.3 Costo y presupuesto	22
2.2.4 Rangos de calidad.....	22
2.3. Precedencia y Prioridad	22
2.4 Requerimientos.....	23
2.4.1 Descripción del sistema de control de Stock.....	23
2.4.2 Requerimientos Funcionales:.....	23
2.4.3 Requerimientos No Funcionales:	24
2.4.4 Requerimientos del sistema	24
2.4.5 Requerimientos de entorno:	25
2.5. Presentación del proyecto	25
2.5.1 Introducción.....	25
2.5.2 Metodología	25
2.5.3 Objetivos y Alcance.....	26
2.5.4 Hipótesis:	26
2.5.5 Alcances y Limitaciones:.....	26
2.5.6 Justificación del Proyecto.....	27
2.5.7 Supuestos y restricciones	27
2.5.8 Resultados del proyecto.....	28
2.5.9 Planificación	28
2.6 Organización del proyecto	30
2.6.1 Estructura organizacional.....	30
2.6.2 Contactos externos	30
2.6.3 Roles y responsabilidades	30
2.7. Proceso de dirección	31

2.7.1 Plan de la fase	31
2.7.2 Plan de monitoreo y supervisión del proyecto	31
2.7.3 Plan de control de agenda	32
2.7.4 Plan de control de calidad	32
2.7.5 Plan de comunicación	33
2.8. Riesgos del proyecto	33
2.8.1 Introducción.....	33
2.8.2 Riesgos	33
2.9. Flujo de requisitos.....	37
2.9.1. Caso de Uso Trazo Grueso Vs. Trazo Fino	38
2.9.2 Caso de Uso: Trazo Grueso.....	38
2.9.3 Caso de Uso: Trazo Fino	40
2.10. Flujo de Análisis.....	44
2.10.1 Diagrama de entidad relación (DER).....	45
2.10.2 Diagramas de colaboración.....	45
2.11. Flujo de diseño	46
2.11.1 Introducción.....	46
2.11.2 Diagramas de secuencia	47
2.12. Descripción de la arquitectura	48
2.12.1. Introducción.....	48
2.12.2. Arquitectura.....	48
2.12.3. Arquitectura de la Aplicación	50
2.13. Flujo de pruebas	52
2.13.1. Introducción.....	52
2.13.2. Tipo de Prueba.....	52
2.13.3. Prueba: Proceso ABM de Productos.....	53
Análisis, Diseño y Construcción de Data Warehouse	55

3. Marco Teórico.....	55
3.1 Conceptos relacionados:.....	56
4. Herramientas de Data Warehouse	67
4.1 Motor de bases de datos MySQL.....	67
4.1.1 MySQL Workbench – Administrador de base de datos.....	67
4.2 Solución BI.....	68
4.3 Netbeans.....	72
5. Metodología para el desarrollo de un Data Warehouse.....	75
5.1 Planeación y administración del proyecto	75
5.2 Análisis de requerimientos	80
5.3 Modelamiento dimensional.....	82
5.4 Diseño técnico de la arquitectura	88
5.5 Procesos de Extracción, Transformación y Carga	89
5.6 Selección e instalación de productos	95
5.7 Características de aplicaciones para usuarios finales.....	97
5.8 Mantenimiento y crecimiento de un Data Warehouse.....	100
6. Desarrollo del proceso de construcción del Data mart	103
6.1 Planeación y administración del proyecto	103
7. Análisis de requerimientos.....	106
7.1 Análisis de requerimientos	106
7.2. Documentación de requerimientos.....	106
8. Modelamiento dimensional	108
8.1 El Data mart	108
8.2 Definición de la granularidad.....	109
8.3 Dimensiones	109
8.4 Tablas de hechos	112
8.5 Diseño del modelo dimensional	115

9. Diseño técnico de la arquitectura.....	116
9.1. Datos.....	116
9.2. Backroom: Extracción, Transformación y Carga	119
9.3 Frontroom.....	126
9.4 Automatización de actualización de repositorio	126
9.5 Infraestructura de Data Warehouse	127
10. Reportes implementados	128
10.1 Ver ingresos por producto por local en el tiempo	128
10.2 Ver ingresos por rubro por local en el tiempo.....	129
10.3 Ver ingresos por local.....	130
10.4 Ver cantidad de desechos por local en el tiempo.....	131
10.5 Ver productos menos vendidos por local en el tiempo	133
11. Mantenimiento y crecimiento del Data mart.....	134
12. Conclusiones y Recomendaciones	135
Conclusiones.....	135
Recomendaciones	136
Anexos:	137
ANEXO I – Problema de acceso a los datos de Strega.....	137
ANEXO II – Manual de uso Software “Stockia”	142
BIBLIOGRAFIA	157

Introducción

Desde hace varios años, el rubro alimenticio viene teniendo un crecimiento sostenido en toda la Argentina, permitiendo que se desarrollen a lo largo del país distintos polos alimenticios, siendo Córdoba uno de los más importantes. Es por esto que las decisiones que se tomen en este rubro, son claves para que el crecimiento no se detenga y el negocio se pueda expandir hacia nuevos mercados.

Hoy en día contamos con una gran masa de datos, pero para poder sacar ventajas de ello es necesario analizarlos y extraer información que nos permitan detectar nuevas oportunidades de mejora para nuestro emprendimiento, servicio o negocio.

Strega es una cadena de delivery de comida rápida que está instalada en Córdoba desde hace ya más de diez años, otorgándole cierto prestigio frente a la competencia. En los últimos años ha tenido un crecimiento sostenido, lo que llevó a que su oferta de locales se amplíe hasta 10 a lo largo de toda la ciudad. Gracias a este crecimiento, el volumen de datos almacenados se vio incrementado notablemente, y dado que la empresa no tiene un área de Inteligencia de Negocios, no se los aprovecha de manera eficiente para poder hacer un análisis de la situación actual de la empresa y tomar decisiones en la marcha que permitan aprovechar nuevas oportunidades de negocio.

En el presente proyecto, se realizará el análisis de las fuentes de datos con las que cuenta la empresa, detectando aquellas necesarias y que no están siendo registradas, con el fin de implementar un Data Warehouse para dar soporte a la toma de decisiones de los directivos. Las áreas puntuales con las que trabajaremos son Marketing y Publicidad, brindándole información de fácil lectura e interpretación.

Empresa

En este bloque daremos a conocer aspectos importantes de la empresa Strega, como serán su historia, su desarrollo en el tiempo, para de esta forma comprender el modo en que la misma trabaja y como las necesidades del comienzo fueron cambiando. No obstante, su objetivo, misión y valor se mantienen y establecen una sólida base en la que se desarrolló el negocio.

También encontraremos información sobre el problema que impulsó el desarrollo del software y como éste podrá mejorar la gestión de los locales y, alineado con los objetivos de la empresa, aumentar la productividad y eficiencia en los distintos procesos.

1.1 Presentación de la empresa

Nombre de la empresa: Strega



Rubro de la empresa:

La empresa se dedica al rubro gastronómico, principalmente al delivery de comida, aunque también cuenta con locales para poder comer allí.

Misión – Visión - Valores

La incursión en el mercado gastronómico de cocina rápida nace de la adquisición de un negocio familiar. Después de más de 20 años y motivados por la necesidad de seguir apostando a un producto familiar, pero ordenado y organizado, nace Strega como marca.

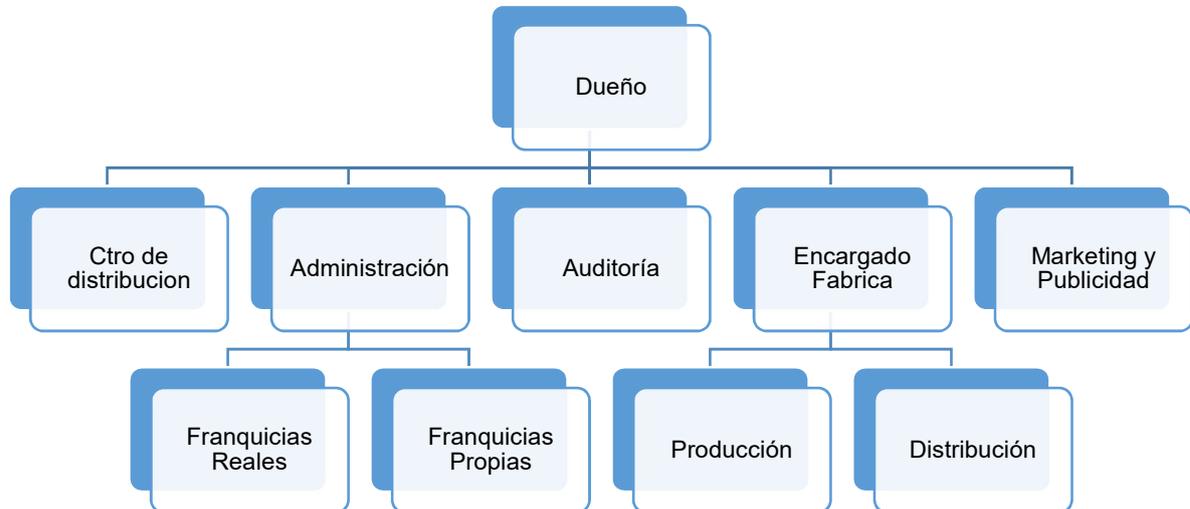
Misión: Strega tiene como propósito ofrecer menús con ingredientes frescos y de calidad, brindando un alto nivel de servicio y atención para sus clientes. Como empresa pretende comercializar los productos en distintos puntos de la provincia de Córdoba y del país, a través de la red de Franquicias.

Visión: Strega pretende lograr el reconocimiento entre sus clientes por sus productos de buen sabor y calidad, acompañado del mejor servicio.

Valores: Máxima calidad y seguridad en nuestros productos: verificamos cada una de las fases de producción, desde la selección de los ingredientes frescos hasta la preparación del producto en el local.

- Excelente servicio de atención al cliente: respondemos con la inmediatez en salones y reparto, presentando un producto final de excelencia.
- Mejora continua: siempre estamos aprendiendo con el fin de anticiparnos respondiendo a las necesidades cambiantes de nuestros clientes y empleados.
- Crecimiento constante: trabajamos para proveer un crecimiento rentable sostenido para nuestros franquiciados; exigiéndonos continuamente para hacer crecer el negocio y generar ganancias.

1.2 Organigrama:



1.3 Descripción de áreas funcionales:

Centro de distribución: Se encarga de la distribución de la materia prima que utilizan todos los locales para elaborar las distintas comidas.

Administración: Es el área encargada de llevar los números de la empresa, administrar y coordinar la comunicación con los franquiciados, tomar decisiones y comunicar al gerente el funcionamiento de la empresa, etc.

Auditoria: Ésta área se encuentra tercerizada, y se encarga de auditar todo el funcionamiento de la empresa para permitir, tanto a franquiciados como a locales propios, establecer una misma calidad en todos sus productos.

Fábrica: Toda la materia prima que utilizan los locales para la elaboración de las distintas comidas, se elabora en un solo lugar, permitiendo así establecer la mejor calidad en las comidas a lo largo de todos los locales, y que el sabor no cambie entre un local y otro.

Marketing y publicidad: Ésta área se encuentra tercerizada, y se encarga de elaborar todas las campañas publicitarias, imágenes promocionales, logotipos, diseño de locales, etc.

1.4 Planteamiento del problema

Hace muchos años las transacciones en Strega se gestionan mediante el software PIXO. Este software almacena los datos en una base de datos cuya gestión pertenece a la empresa proveedora del software, y en la misma se almacenan todas las ventas, recetas y gestiones que se producen en los locales a excepción del módulo de stock, el cual no es utilizado por locales por ser complicado y difícil de manejar, y en su reemplazo, los locales utilizan un Excel donde llevan el control propio de cada sucursal, lo que genera un problema debido a que es manejado por muchas personas y no se tiene un control preciso de la información ni de los insumos y materias primas del local.

Recientemente, y debido a la gran expansión que tuvo el negocio, comenzaron a surgir problemas que afectan directamente a las ventas del negocio:

Problema	No hay registro de stock de locales, provocando que en ocasiones los locales no cuenten con materia prima y tengan que buscar.
Afecta	Cliente, venta, stock
Impacto	Afecta al cliente ya que el menú no se puede ofrecer. También afecta a la empresa ya que por política los menús deben estar disponibles siempre
Beneficio Posible	Solución Software que permita ingresar los pedidos y mantener actualizado el stock

Problema	En ocasiones los pedidos de las sucursales no son correctos, se solicitan menos insumos y otros no se tienen en cuenta.
Afecta	Cliente, venta, stock, empresa
Impacto	Al igual que el ítem anterior afecta al cliente y a la empresa. El local se ve perjudicado por esta mala gestión ya que tendrá que asumir los gastos de conseguir los faltantes y absorber las pérdidas generadas por el cliente que no compra.
Beneficio Posible	Solución Un sistema de alertas, que informe cuando un producto está próximo a escasear.

Problema	Se tienen por mes muchos desechos ya sea por productos vencidos, en mal estado, etc.
Afecta	Cliente, venta, stock, empresa
Impacto	Se tienen muchas pérdidas por falta de control en los productos que se piden para stock, no se controlan las fechas de vencimiento y los productos se ponen en mal estado, provocando pérdidas de dinero ya se por el producto a desechar o por ventas que no se pueden realizar por falta de insumos.
Beneficio Posible	Solución Permitir cargar por software la cantidad de desechos semanales y mensuales, informando la causa del desecho, para luego generar informes y poder anticiparse al hecho.

Problema	No tienen forma sencilla de medir los beneficios de acciones publicitarias.
Afecta	Ventas, Empresa
Impacto	No se sabe con exactitud la eficiencia en el contrato de publicidad.
Beneficio Posible	Solución Un Data Warehouse nos permitiría medir cada local en un tiempo determinado y comparar con meses anteriores si hubo incremento en ventas. De esta forma se puede reforzar publicidad.

Problema	Para medir la situación de las sucursales, realizan Excel que le permitan ver resumen de ventas por local, entre otras cosas. Estas tareas son manuales y llevan su tiempo, además, al no ser automáticas hay posibilidad de errores humanos.
Afecta	Ventas, empresa
Impacto	Hay una fecha en donde los empleados deben presentar la cantidad de ventas, pero de necesitarse esa información urgente, no lo pueden cumplir.
Beneficio Posible	Solución Un Data Warehouse permite obtener información actualizada de manera rápida, sencilla y organizada.

- 1- La publicidad que se estaba realizando no llegaba al segmento de clientes a los que el negocio necesitaba apuntar.
- 2- La administración del stock de materias primas en los locales no era eficiente y se tenían muchos desechos, generando pérdidas por el desecho en sí y además porque muchas ventas no se podían realizar

debido a que no se contaba con la materia prima en el momento para realizar el producto.

Idea Solución:

Luego de muchas reuniones con los dueños de la franquicia, se analizó la posibilidad de explotar los datos que tenían de ventas, con lo cual surge la idea de implementar una herramienta de Inteligencia de Negocios que ayude a generar nuevas estrategias de Marketing gracias al análisis de datos y la generación de reportes. Por otro lado, nos dimos cuenta que necesitaríamos los datos de control de stock y desechos eliminados, motivo por el cual se ofreció la posibilidad de desarrollar una herramienta de control de stock amigable, fácil de usar y que cuente con las funciones básicas, que permita llevar el stock de una manera más eficiente, controlando las salidas y los desechos que se generan, y añadiendo también un Módulo de Reportes que, luego de realizada la carga de las bases en el Data Warehouse, permita la generación de reportes con gráficas y datos para mejorar el análisis de la información y la posterior toma de decisiones.

1.5 Objetivos de investigación

1.5.1 Objetivos generales

- Análisis, Diseño y Construcción de un Data Warehouse que permita centralizar todos los datos de las distintas sucursales para su posterior análisis.
- Desarrollo de un software de Control de Stock, de ahora en más llamado “Stockia”, para utilizar como complemento al sistema de ventas que actualmente se utiliza en los locales y así generar otra base de datos que servirá de entrada para el Data Warehouse. Integrado al desarrollo, se añadirá un módulo que permita al usuario la visualización de los reportes generados con el Data Warehouse.

1.5.2 Objetivos específicos

- Centralizar las bases de datos de las sucursales en una sola para tener una mejor variedad y un mejor volumen de datos. Este proceso puede incluir un rediseño de las bases con las que cuenta la empresa.
- Generar reportes semanales y mensuales que permitan visualizar el funcionamiento de la empresa y tomar decisiones fundadas.
- Analizar datos históricos para aprovechar oportunidades estacionarias.
- Analizar los datos de las distintas sucursales para expandir el negocio a nuevos mercados.
- Mejorar la gestión del stock de los locales mediante la implementación de un software de Control de Stock, optimizando la organización de los insumos y materias primas, para poder realizar anticipadamente los pedidos a proveedores, evitando faltantes en el momento en que el local se encuentra operando.
- Evitar costos innecesarios en la adquisición de licencias de software, por lo que se deberá tratar de usar software sin licencias y open source para disminuir al mínimo los costos del proyecto.

1.6 Justificación del estudio

Luego de la última reunión con los dueños, se les enumeraron los múltiples beneficios de explotar los datos de ventas que actualmente tienen cargados:

- Mayor control de las ventas del local.
- Posibilidad de generar reportes para que el área de Marketing los explote y poder generar estrategias de ventas.
- Atraer nuevos clientes en base a publicidad segmentada.
- Posibilidad de realizar minería de datos para identificar patrones de ventas.

En base a estos beneficios, se acordó con la empresa realizar el análisis y el diseño del Data Warehouse, generando los reportes necesarios, pero sin realizar la implementación en los locales, debido a que primero tienen que analizar la posibilidad con los encargados de administrar las bases de datos. La principal justificación es la de trabajar en conjunto con el área de Marketing, para

facilitarles datos, reportes, gráficos y estadísticas de venta para poder generar nuevas campañas publicitarias, atraer nuevos clientes y expandir el negocio fuera del área local. Además, con este tipo de reportes, la gerencia se puede encargar de tomar decisiones fundadas con respecto a la administración de los locales, de los productos, apertura de nuevos locales, etc. Además, se pueden dar ciertos tipos de herramientas, números certeros y reportes a los franquiciados para que puedan administrar mejor sus locales, y así generar una mejor renta tanto para los franquiciados como para los dueños de la franquicia.

1.7 Limitaciones de la investigación

La principal limitación que se tuvo en el presente trabajo de Investigación fue de tipo legal, debido a que la empresa no nos facilitó las bases de datos con las ventas de los locales, argumentando que, si estos datos se filtraban, podían llegar a tener problemas impositivos. Debido a esto, se generó una base de datos mediante procedimientos SQL con transacciones de ventas en base a estadísticas, teniendo especial cuidado en reflejar lo más certera posible los números del negocio dependiendo de qué local se trate. Luego por el lado de los datos que sí nos brindaron, nos topamos con bases de datos en formato DBF, lo que provocó que, al migrarlas, se perdieran muchos datos de clientes. Lo que se hizo para solucionar este problema, fue limpiar la base de datos antes de migrarla, y dejar la nueva base de datos en formato MySQL con información certera y nutrida, sin basura y con datos esenciales. A partir de ese momento, se recomendó a los dueños de la franquicia migrar la base de datos que actualmente tienen a una nueva con motor MySQL, con la principal ventaja que es libre de licencias y que actualmente se utiliza en muchos sistemas, obteniendo una respuesta negativa de parte de ellos, argumentando que tendrían que cambiar el sistema que actualmente tienen y no está en sus planes en el corto plazo realizarlo. Por este motivo, se decidió migrar los datos que tienen actualmente de clientes, pero al haber incompatibilidad entre el formato DBF y MySQL, se hará todo el trabajo de investigación, pero sin realizar la implementación en los locales.

Desarrollo de Software “Stockia”:

2.1. Descripción de apostadores y usuarios

2.1.1. Información demográfica del mercado

Abarca únicamente las sucursales de Paraná, Pueyrredón y Obispo Salguero. No se tiene en cuenta el almacén central de distribución.

2.1.2. Resumen de apostadores

Nombre	Descripción	Responsabilidades
Gerente	Es el dueño de la empresa, a quien se le presenta la propuesta de desarrollo del sistema de control de stock.	-Realizará seguimiento del progreso del proyecto -Aprobará las erogaciones de dinero para el proyecto -Proveerá de información necesaria para el desarrollo del proyecto.

2.1.3 Resumen de usuarios

Nombre	Descripción	Responsabilidades
Encargados de sucursal	Este usuario es el principal responsable de las tareas de control de stock. Tendrá acceso total a las funciones del sistema.	Responsable de gestionar el almacén, realizando pedidos, descontando productos utilizados por día, manteniendo actualizados proveedores, almacenes.
Dueño de Sucursal	Al igual que el encargado, tendrá acceso total del sistema, este rol es creado ya que	Es el reemplazo del encargado de almacén en caso de que no haya uno o que el mismo no

	en la actualidad es el dueño quien reemplaza al encargado de sucursal en caso de que el mismo no asista o no haya uno.	asista. Es el responsable de las mismas tareas.
--	--	---

2.1.4. Entorno de usuarios

Los responsables de interactuar con el sistema, cuentan con una PC de uso compartido en el caso del encargado de sucursal y de uso individual en el caso del dueño, motivo por el que si bien las máquinas poseen el hardware necesario para correr el programa las características de las mismas no son un estándar en todas las sucursales.

Detallamos brevemente los componentes de las PC. Las mismas poseen paquete office, acceso a internet y Windows XP.

2.1.5. Lay-Out de las Sucursales

Las sucursales mantienen estructuras diferentes dependiendo el lugar donde se encuentren. En determinadas situaciones se asigna más espacio de salón, mesas y sillas para la atención en sitio. Sin embargo, todas las sucursales cuentan con:

- Mostrador: donde se atiende de forma personal o telefónica.
- Salón: donde se encuentran mesas y sillas para que el cliente pueda ordenar y comer en sitio.
- Cocina: donde se preparan los pedidos, se lavan y encuentran los utensilios.
- Almacén: donde se encuentran almacenados los productos.

2.1.6. Tecnologías Disponibles

La distribución de máquinas en las sucursales es simple, pero al mismo tiempo podemos encontrar dos situaciones:

- Que cuenten con una máquina situada en mostrador. En la misma corre la aplicación PIXO para la gestión de ventas y en la que correrá también el sistema de control de stock. Cuenta con acceso a internet y posibilidad de acceso remoto.
- Que cuenten con dos máquinas, en la primera corre el programa PIXO y la segunda funciona como server, en la que se almacenan temporalmente (15 días) los datos de ventas, productos, menús, entre otras cosas. El objetivo de este servidor es evitar consultar la Base de Datos principal generando demoras y sobre procesamiento en el servidor principal.

Características de los terminales:

Terminal 1: Server

Microprocesador Intel Core i5

4GB de memoria RAM

Placa de video de 1gb GeForce 810

Terminal 2: funcionamiento de PIXO

Microprocesador Intel Core i3

2GB de memoria RAM

Placa de video integrada

2.1.7. Perfil de Apostadores

Representante	Dueño
Descripción	Gerente general.
Tipo	Posee conocimiento de toda la empresa, lleva a cabo el desarrollo del plan estratégico.

Responsabilidades	- Realizará seguimiento del progreso del proyecto. - Evaluará presupuestos para gastos en tecnología.
Criterio de éxito	Que el sistema sea implementado en el tiempo pactado y que cubra los requerimientos definidos con anterioridad.
Involucramiento	Proveer información y tomar decisiones en presupuestos.
Entregables	Establece el contacto con dueños de sucursales, para el desarrollo de requerimientos.

2.1.8 Perfil de Actores

Representante	Usuarios de sucursales
Descripción	Encargado de sucursal
Tipo	Posee conocimiento de sucursal, gestión de empleados, stock, control de cumplimiento de normas de calidad definidas por Strega.
Responsabilidades	Actualización de productos en almacén, gestión de proveedores.
Criterio de Éxito	- Mantener orden de almacén. - Proveer de productos en todo momento. - Mantener Stock actualizado.
Involucramiento	-Principal involucrado en la definición del sistema, posee conocimientos suficientes del manejo de la sucursal y los problemas relacionados al control de stock.
Entregables	-Posee información con respecto a la metodología utilizada actualmente. -Especificación del requerimiento.

2.1.9 Necesidades de usuarios

- Llevar actualizado el registro de pedidos.
- Actualizar stock.
- Recibir alertas por faltantes.

- Generar informes semanales y mensuales de ventas y desechos.

2.1.10 Alternativas y competencia

La empresa se encuentra con dos alternativas, la primera es el uso del módulo de control de stock con el que cuenta PIXO, la segunda es continuar utilizando herramientas Office para la gestión de pedidos y el control de stock. Lo que nos dice la experiencia es que la gestión con la herramienta office no ha sido del todo satisfactoria, ya que el manejo de planillas es más complejo y poco amigable.

2.2. Descripción del Producto

El producto contará con las funcionalidades básicas de un controlador de stock y además permitirá realizar:

- ABM de insumos
- ABM de recetas
- Gestión de cuenta de usuario
- Consultas
- Generación de informes de ventas y desechos con datos obtenidos de un Data Mart

El software es una aplicación de escritorio, desarrollada en lenguaje orientado a objetos y que interactúa, por un lado, con una base de datos mysql para la gestión de base de datos y por otro lado con una base de datos de un Data Mart.

2.2.1 Perspectiva del producto

Herramienta Actual	<p>La empresa se encuentra gestionando los pedidos y el control de stock mediante el uso de planillas Excel.</p> <ul style="list-style-type: none"> • Control de los productos disponibles en almacén. • Control de entrada de productos. • Control de salida de productos. • Control de inventario. • Configuración de umbral de cantidad a alertar.
---------------------------	--

Herramienta Propuesta	<p>El sistema que vamos a desarrollar les permitirá continuar con los registros que realizaban de la forma antigua. Además, mediante un módulo de identificación limitará la modificación de datos solamente a las personas con usuarios y claves, de esta forma se mantiene la integridad de la información.</p> <p>Permitirá mediante definición de umbrales detectar faltantes en productos, dando lugar a un mantenimiento preventivo.</p> <p>Se podrán registrar las salidas de insumos y el motivo de la misma.</p> <p>Se podrá consultar en cualquier momento el estado del stock.</p> <p>Se podrán generar informes de ventas y desechos en cualquier momento para poder tomar decisiones en la marcha.</p>
------------------------------	---

2.2.2 Supuestos y dependencias

El sistema a desarrollar se presenta como una necesidad y un beneficio para todos los locales, no obstante, el análisis no se realiza en todos los locales sino en 3 sucursales, dos a cargo del gerente de Strega y otra sugerida por él. Bajo este escenario suponemos que los requerimientos serán bien representados por los dueños de los locales y representen la realidad del resto.

El apoyo político para la implementación de la herramienta va a ser lo más importante, el uso de la misma debe presentarse como obligatorio junto al resto de los procesos y reglamentaciones que las franquicias deberán cumplir. La mayor resistencia al cambio la podemos tener de los locales que se encuentran trabajando hace tiempo y que no han sido involucrados, para esto se dará una capacitación del nuevo sistema y se recalcarán los beneficios que se obtienen del uso del mismo.

2.2.3 Costo y presupuesto

Seleccionamos las tareas que creemos más importantes para definir los costos, entre ellas se incluyen:

- Costo por desarrollo e implementación del sistema.
- Adquisición de nueva tecnología, en caso que sea necesario.
- Capacitación de usuarios finales.
- Mantenimiento técnico de software.

2.2.4 Rangos de calidad

El proyecto que presentaremos no es de gran medida, por lo que será sencillo realizar un correcto control y seguimiento. Si bien no nos basaremos en un modelo específico para definir la calidad de software, prestaremos principal atención a la satisfacción de la necesidad del usuario en el menor tiempo posible y con los menores errores, persiguiendo con esto que el sistema sea óptimo en usabilidad, eficiencia y confiabilidad.

En cuanto al mantenimiento post implementación, se utilizarán en las primeras semanas indicadores de continuidad de servicio para determinar la cantidad de impactos negativos que pueden presentarse, los mismos se registrarán y serán analizados para determinar el origen del error.

2.3. Precedencia y Prioridad

- 1- Actualización de estado de almacén en locales: este sin duda es el principal problema que se intenta solucionar. Ya que es un factor que impacta directamente en el cliente final, y la percepción de éste con el negocio es primordial.
- 2- Organización en procesos: Se busca estandarizar el control de stock en todos los locales. De esta forma se ofrece mayor apoyo al franquiciado, al darle todas las herramientas necesarias para una correcta gestión del local.

- 3- Disminuir eventos innecesarios: La herramienta permitirá trabajar de forma preventiva, evitando llegar a situaciones no deseadas como falta de stock en algunos productos.
- 4- Mejora en calidad del servicio: si los apartados anteriores se cumplen, el impacto positivo se verá reflejado en una mayor satisfacción del cliente, ya que disminuirán las demoras en preparación y no habrá faltantes en menús.

2.4 Requerimientos

2.4.1 Descripción del sistema de control de Stock

Strega necesita un sistema de control de stock que le ayude a organizar cada almacén de los locales, a fin de solucionar la problemática planteada anteriormente, que integre también un módulo para generar informes de ventas y desechos. Nosotros nos centraremos en el desarrollo del software de control de stock y en la generación de los informes.

El sistema, de ahora en más Stockia, estará manipulado por el encargado de sucursal o dueño, se implementará de forma local en cada sucursal, trabajando sobre bases de datos locales. El encargado podrá ingresar a Stockia cada pedido de mercadería, modificar cantidades, eliminar productos, realizar consultas de productos y ser alertado cuando un producto sea inferior a un umbral pre establecido. Las salidas de insumos no generarán comprobantes, sino que quedarán registradas en la base de datos con fecha, cantidad y motivo de la salida del insumo.

2.4.2 Requerimientos Funcionales:

- Identificación: Stockia permitirá el ingreso a personas autorizadas para la manipulación del programa. Esto se hará asignando un usuario y clave que permitirá diferenciar un encargado de un administrador. Solo se definirán estos dos roles que tendrán acceso a la totalidad de las funciones.

- **Modificaciones:** Stockia permitirá modificar los datos que están dados de alta, es decir, permitirá hacer un CRUD de insumos, recetas y usuarios.
- **Consultas:** Stockia permitirá realizar consultas de stock, para verificar las cantidades que se encuentran en cada almacén. También se podrá hacer consultas de transacciones realizadas como por ejemplo las salidas de productos.
- **Gestión de Stock:** Todas las salidas diarias serán cargadas al cerrar la caja, de esta forma se podrá obtener por reportes la cantidad de productos utilizados por día, información que servirá luego para realizar los pedidos. Estas salidas quedarán registradas bajo el usuario que la realizo y el día.
- **Generación de informes:** Stockia debe permitir generar informes con gráficos y tablas de las ventas de los locales en un tiempo determinado, como así también de los desechos generados en cada local.

2.4.3 Requerimientos No Funcionales:

- El diseño de la base de datos debe ser pensada para que a futuro se pueda integrar con el sistema PIXO que utiliza la empresa.
- El sistema debe tener una interfaz amigable, ser fácil de usar y rápido. Para evitar que el usuario lo vea como una pérdida de tiempo.
- Para desarrollar el sistema se utilizará el lenguaje de programación Java (JEE), bajo un entorno de desarrollo Netbeans, con un motor de base de datos MySql server.

2.4.4 Requerimientos del sistema

Los requisitos necesarios con los que se debería contar para un buen funcionamiento del sistema son computadoras personales con las siguientes características:

- Sistema operativo Windows XP o superior.
- Procesador Pentium IV o superior.
- Memoria RAM 1gb Mb.
- Disco duro de 30 Gb.
- Navegadores de Internet (Mozilla)
- Configuración de equipos de red, router, switch, hub.

El equipamiento con que la organización cuenta en este momento sobrepasa las necesidades y requerimientos para el pleno funcionamiento del sistema propuesto por lo que no es necesaria la adquisición de Hardware para la implementación del mismo.

2.4.5 Requerimientos de entorno:

No hay requerimientos específicos para el entorno. Se verificó que las condiciones de ambiente sean adecuadas para el correcto funcionamiento de los terminales, en este análisis se tuvo en cuenta que la cocina se encuentra separada del mostrador por una pared, de esta forma se mantiene una temperatura adecuada y una limpieza necesaria para no afectar el hardware.

Los modem-router se encuentran también en el sector del mostrador, permitiendo un fácil acceso al servicio de WI-FI y brindándole también, las condiciones analizadas en el párrafo anterior.

2.5. Presentación del proyecto

2.5.1 Introducción

El documento de proyecto intentará establecer las bases sobre las que se trabajará. El objetivo principal de este documento es permitir una mejor organización en las tareas a realizar para poder cumplir con las entregas y tiempos pactados. Además, se presenta una guía que nos permitirá hacer un seguimiento y control del proyecto en base a las definiciones realizadas en un principio.

Para esto se contemplan entre otras cosas los supuestos, restricciones y planificaciones correspondientes, para de este modo tener presente donde estamos, donde deberíamos estar y a donde debemos ir.

2.5.2 Metodología

Teniendo en cuenta la dimensión del proyecto hemos seleccionado la metodología tradicional en cascada.

Nuestra elección se justifica en que es un proyecto pequeño y fácil de modelar, en donde los requerimientos son fácilmente definidos con ayuda del cliente y se

puede realizar este análisis de requerimientos en una primera fase. Esto nos permite organizar y planificar el tiempo de desarrollo y entrega sin depender de los tiempos de nuestro cliente.

2.5.3 Objetivos y Alcance

Objetivo General:

Satisfacer las necesidades de nuestro cliente, brindándoles un software con interfaz intuitiva, amigable y además funcional.

Objetivos específicos:

- Permitir un mejor control de los insumos de cada sucursal.
- Generar indicadores, para poder medir la eficiencia en la utilización de estos recursos.
- Mantener el orden de los productos, para facilitar los pedidos a casa central.
- Trabajar con alertas preventivas, que eviten que los empleados tengan que salir del local a comprar productos por faltantes.
- Poder tomar decisiones en base a los informes de ventas y desechos generados por el sistema.

2.5.4 Hipótesis:

El crecimiento de la empresa en el último tiempo no permite tener una visión correcta de la realidad y se continúa trabajando como una pequeña empresa familiar. La realidad es que, si bien Stockia le va a permitir mejorar en aspectos de control y orden, será necesario rever los procedimientos de control de stock y pedidos de insumos para mantener a todos los locales alineados y obtener mayores beneficios de la implementación.

2.5.5 Alcances y Limitaciones:

Alcances:

- Alta, Baja y Modificación de Productos.
- Alta, Baja y Modificación de Recetas.
- Consulta de productos en stock.
- Restricción de acceso a datos.

- Obtener mayor control y orden del almacén.
- Generación de reportes.

Limitaciones:

- Definir metodologías para la carga de productos.
- La falta de costumbre de esta tarea puede al comienzo traer resultados erróneos.
- Stockia se utilizará en forma simultánea a PIXO (el software actual de la empresa).
- La base de datos con la que se encuentran trabajando los locales deberá ser la misma, para mantener una integración de la información.

2.5.6 Justificación del Proyecto

La empresa actualmente se encuentra trabajando con el Software Pixo, el cual le sirve principalmente para las ventas. Si bien, cuenta con un módulo de control de stock el mismo no se está utilizando. Debido a esto las sucursales no tienen posibilidad de prevenir faltantes, en ocasiones cuando faltan productos se ven obligados a comprar en un supermercado o almacén lo que les falta o pedirse entre sucursales mercadería, este problema se hace presente repetidas veces.

Uno de los principales motivos por el cual no se utiliza el módulo de control de stock es debido a que, al comienzo, la cantidad de pedidos no era significativa y no justificaban el proceso de registro de recursos, esa costumbre se mantuvo y actualmente la empresa ha crecido al igual que sus ventas y franquicias, y cada local comienza a sentir la falta de organización a la hora de generar solicitudes. Por otro lado, la herramienta no se presenta amigable para la función de control de stock, y el hincapié de las capacitaciones es principalmente sobre la gestión de ventas.

2.5.7 Supuestos y restricciones

En cuanto a las encuestas y análisis, se trabajará con tres locales, dos locales propios del dueño, y el segundo una franquicia a tercero. Se considerará la información recopilada de estos, un estándar en la situación de todos los locales.

Se ha conversado del alcance del proyecto con el dueño y no se encontraron inconvenientes en cuanto al acceso de información ni de disponibilidad de tiempo.

2.5.8 Resultados del proyecto

En la lista de entregables pactados con el dueño se encuentran:

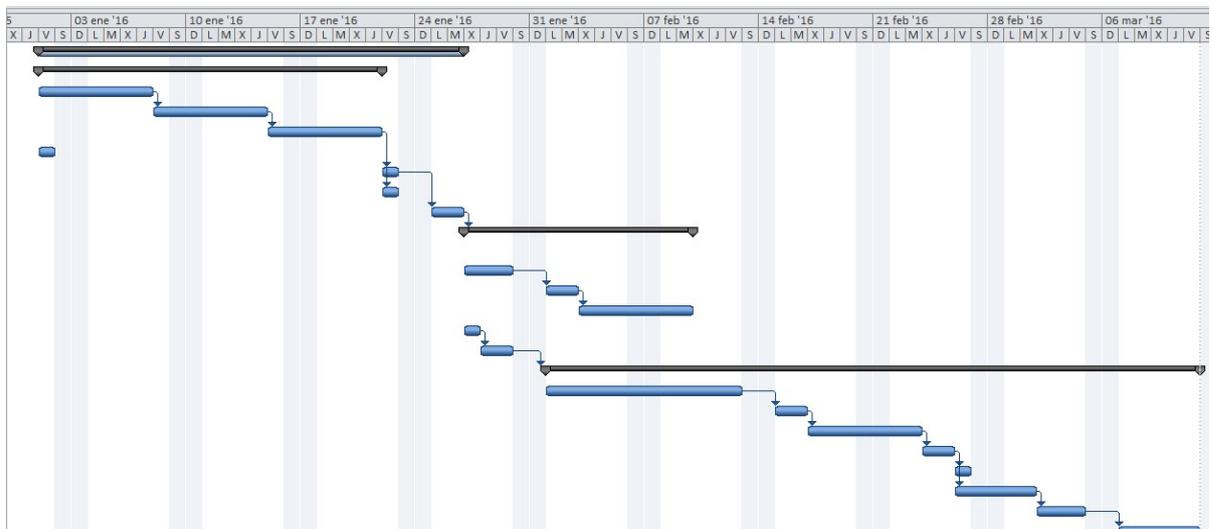
- Plan de desarrollo de software: Es el bloque de “Presentación de proyecto”
- Visión de negocio: en el que se intenta describir las características de Stockia, con el fin de alinearlos a la necesidad planteada por el cliente.
- Lista de riesgos: se pondrán en discusión situaciones detectadas que pueden generar un problema o complicación.
- Un modelo de prueba: con el fin de testear el sistema, se seleccionarán tareas diarias y una situación particular.
- Producto final.

Acostumbrados de trabajar con clientes, hemos definido estos entregables ya que consideramos que los mismos no contienen lenguaje técnico y permitirán al cliente entender más fácilmente la realidad que se intenta modelar, por esto también será de utilidad para nosotros ya que la retroalimentación será más enriquecedora.

2.5.9 Planificación

Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
Sistema de Control de Stock		vie 01/01/16		
Plan de desarrollo	18 días	vie 01/01/16	mar 26/01/16	
Recopilación de información	15 días	vie 01/01/16	jue 21/01/16	
Encuestas, entrevistas	5 días	vie 01/01/16	jue 07/01/16	
Análisis	5 días	vie 08/01/16	jue 14/01/16	4
Relevar infraestructura	5 días	vie 15/01/16	jue 21/01/16	5
Presentación del producto	1 día	vie 01/01/16	vie 01/01/16	
Entregable: plan de desarrollo	1 día	vie 22/01/16	vie 22/01/16	6
Entregable: prototipo	1 día	vie 22/01/16	vie 22/01/16	6

Revisión	2 días	lun 25/01/16	mar 26/01/16	8
Documento de diseño de Software	10 días	mié 27/01/16	mar 09/02/16	10
Análisis de requerimientos	3 días	mié 27/01/16	vie 29/01/16	
Diseño de sistema	2 días	lun 01/02/16	mar 02/02/16	12
Diseño de programa	5 días	mié 03/02/16	mar 09/02/16	13
Entregable: Visión de negocio	1 día	mié 27/01/16	mié 27/01/16	
Revisión	2 días	jue 28/01/16	vie 29/01/16	15
Implementación	30 días	lun 01/02/16	vie 11/03/16	16
Codificación	10 días	lun 01/02/16	vie 12/02/16	
Pruebas	2 días	lun 15/02/16	mar 16/02/16	18
Correcciones	5 días	mié 17/02/16	mar 23/02/16	19
Pruebas	2 días	mié 24/02/16	jue 25/02/16	20
Entregable: Producto	1 día	vie 26/02/16	vie 26/02/16	21
Implementación	3 días	vie 26/02/16	mar 01/03/16	21
capacitación	3 días	mié 02/03/16	vie 04/03/16	23
Seguimiento implementación post	5 días	lun 07/03/16	vie 11/03/16	24



2.6 Organización del proyecto

2.6.1 Estructura organizacional

Recursos:

- Rodrigo Martínez
- Diego Flores

En el proyecto estaremos compartiendo roles, esto lo decidimos de tal forma ya que es nuestro primer proyecto y creemos más conveniente compartir las tareas para que las mismas sean más completas al contar con distintos puntos de vista.

Dentro de las tareas destacamos:

- Líder de proyecto
- Analistas funcionales
- Analistas de riesgo
- Diseñador de sistema
- Desarrollador
- Implementadores
- Testing

2.6.2 Contactos externos

Recursos:

- Dueño de empresa - Adrián Coraglio
- Jefe de área de marketing – Alejandro Coraglio
- Dueño de Franquicia – Sebastián Benassi

2.6.3 Roles y responsabilidades

- Líder de proyecto: Encargados de gestionar el proyecto en su totalidad, velando por el cumplimiento de tiempos pactados.
- Analistas funcionales: Encargado de mantener vínculo con el cliente, elaborará el análisis funcional de la aplicación a desarrollar.

- Analistas de riesgo: Encargado de detectar los riesgos y la mitigación de los mismos.
- Diseñador de sistema: Descompone y organiza el sistema en elementos que puedan elaborarse por separado.
- Desarrollador: Encargados de la programación del software, codificación.
- Implementadores: Detecta las condiciones necesarias para la implementación del sistema. Realiza la implementación y hace seguimiento y correcciones.
- Testing: Encargados de realización de pruebas.

2.7. Proceso de dirección

2.7.1 Plan de la fase

Los tiempos requeridos para las presentaciones y entregas de los trabajos, no pueden estimarse en este punto de desarrollo del sistema, ya que están vinculados a los tiempos impuestos por el docente. Es decir, que no pueden estimarse los tiempos que insumirán las correcciones de cada entregable ni el número de versiones que serán necesarias para optimizar los documentos.

Recursos del proyecto

Los recursos principales del proyecto serán los integrantes del grupo, como explicamos con anterioridad las tareas se compartirán con el fin de tener distintos puntos de vista.

Además, contamos con un tutor de tesis que realizará las correcciones necesarias permitiéndonos trabajar con el menor error posible.

2.7.2 Plan de monitoreo y supervisión del proyecto

Alineado con el proyecto el plan de monitoreo es de principal importancia dado que permitirá detectar a tiempo aquellos desvíos que pueden generar pérdidas tanto de tiempo como de dinero. Ayudando a cumplir con los entregables.

Al igual que el proyecto principal, este plan es de gran importancia ya que toda la planificación o desvíos de tiempo pueden generar graves problemas a la consecución del proyecto.

Aquellos ítems que consideramos deben estar bajo control, teniendo en cuenta la dimensión del proyecto son:

- Agenda
- Calidad
- Reportes – comunicación
- Conformidad

2.7.3 Plan de control de agenda

Este plan consta del seguimiento de la agenda ya presentada, el líder de proyecto será el encargado de velar por el cumplimiento de los tiempos pactados. Las actividades de mayor importancia son aquellas que involucran al usuario, entregables, es por esto que el mayor foco estará puesto en disminuir desvíos de tiempo con respecto a las fechas de entrega pactadas.

Se dará lugar a cambios radicales en la agenda solo con el acuerdo del cliente, esta medida deberá ser impulsada por una necesidad puntual del cliente, de modo contrario se finalizará la agenda y se procederá a generar otro proyecto como solución al nuevo requerimiento.

El líder de proyecto se encargará de las decisiones frente a demoras, se analizará la causa y dependiendo de las mismas se actuará para evitar reiteraciones o en caso de imprevistos difícil de repetirse se reasignarán recursos para volver la agenda a tiempo.

2.7.4 Plan de control de calidad

Se definió evaluar la calidad del producto a partir de la percepción del cliente con respecto a los entregables.

Cada entregable contará con un seguimiento preventivo para evitar errores o incoherencias en el proceso de desarrollo. Previo a la entrega se realizará una inspección para evitar que errores lleguen al cliente, esta inspección consta de revisiones del entregable y en caso de producto software se realizarán testing de los casos más representativos.

2.7.5 Plan de comunicación

Creemos que mantener un flujo de información con el cliente es uno de los puntos clave para el éxito del proyecto. Una comunicación breve pero continua con respecto al progreso del proyecto genera seguridad en el cliente, por otra parte, cada entregable contara con un reporte en el cual se incluirán las funcionalidades, objetivo y fechas del próximo entregable. En el mismo habrá un apartado de pendientes el que se completará con la devolución del usuario y deberá encontrarse solucionado en la siguiente entrega.

2.8. Riesgos del proyecto

2.8.1 Introducción

El presente documento tiene como objetivo identificar los posibles riesgos que pueden aparecer durante el desarrollo del mismo.

El riesgo es la posibilidad de que se produzca un contratiempo, por lo tanto, aparte de identificarlos, se trata de especificar el posible impacto que puede llegar a tener si éste se manifiesta, crear una estrategia para mitigarlo y un posible plan de contingencia para evitar que el impacto en el proyecto sea grande y ponga en peligro el éxito del mismo.

Este documento presenta los riesgos que se pueden presentar y permite plantear posibles alternativas de solución.

2.8.2 Riesgos

Nombre del riesgo: Nivel de conocimiento en informática de los empleados.

MAGNITUD	Media
DESCRIPCIÓN	El proceso de capacitación a los usuarios puede volverse en un problema, si no tienen el suficiente nivel de conocimiento para manejar el sistema de manera correcta. Esto puede traer como consecuencia, un aumento en el

	tiempo estimado para las capacitaciones, retrasando todo el trabajo posterior de pruebas e implementaciones finales.
IMPACTO	Incremento de tiempos de capacitación. Retraso en tareas de pruebas. Demora en la implementación del producto final.
INDICADORES	Entrevistas a empleados y dueños de las franquicias.
ESTRATEGIA DE MITIGACIÓN	Brindar periódicamente cursos interactivos o capacitaciones breves a usuarios con escaso conocimiento en informática. Comenzar de manera anticipada con las capacitaciones a usuarios e identificar a los usuarios que presenten falencias en el uso de los sistemas, para reforzar las capacitaciones.
PLAN DE CONTINGENCIA	Estimar el posible retraso en los tiempos de pruebas e implementación, seguir con los plazos y planes estipulados, y brindar una capacitación intensiva a los usuarios que presenten más problemas a la hora de utilizar el sistema.

Nombre del riesgo: Presupuesto Excedido.

MAGNITUD	Medio
DESCRIPCIÓN	El riesgo se produciría en caso de que se estime un presupuesto incorrecto en la etapa de relevamiento e investigación, derivando esto en el surgimiento de un exceso de lo estimado y aprobado.

IMPACTO	Este riesgo impacta sobre los tiempos de desarrollo del sistema, ya que se haya sujeto a la adquisición del presupuesto necesario.
INDICADORES	Se advierte una notable reducción del monto destinado al desarrollo del sistema faltando aún etapas para la culminación o, el surgimiento de un costo extra que no había sido tenido en cuenta.
ESTRATEGIA DE MITIGACIÓN	Entregar presupuestos en la primera etapa de planificación. Reuniones semanales para la revisión de presupuesto planificando y detallando los gastos.
PLAN DE CONTINGENCIA	Si el exceso en los costos estimados es producto de una situación inesperada, se deberá informar la misma a los encargados de la organización, para que evalúen los pasos a seguir, ya sea la aprobación del gasto extra o la demora en la implementación del sistema. Si el error se originó por un error en la etapa de relevamiento, se deberá analizar alguna alternativa de desarrollo que permita reducir los costos (ej: analizar la adquisición de otras tecnologías, de otros equipos), sin que ello perjudique el objetivo final que es: la satisfacción de las necesidades y requerimientos de los usuarios e interesados.

Nombre del riesgo: Impacto del cambio sobre los usuarios.

MAGNITUD	Media
DESCRIPCIÓN	El riesgo se debe a que todo cambio en el método de trabajo provoca una sensación de angustia e incertidumbre en los empleados.

IMPACTO	<p>Este riesgo impacta sobre los tiempos de implementación del sistema, debido a necesidad de capacitar a los usuarios.</p> <p>En caso de que esta capacitación no sea tomada en cuenta por los desarrolladores, se correrá el riesgo de que una vez implementado el sistema, los usuarios al no comprender su funcionamiento expresen su disconformidad, adjudicando fallas operativas al funcionamiento del sistema.</p>
INDICADORES	<p>La resistencia al cambio, las fallas en la utilización y la negación ante su funcionamiento.</p>
ESTRATEGIA DE MITIGACIÓN	<p>En la etapa de planificación del sistema se deberá analizar el método a utilizar para capacitar a los empleados estimando costos e insumos necesarios.</p> <p>Además, sería recomendable que, considerando límites, los usuarios pudieran participar del desarrollo, conociendo los avances, logrando de esta manera su involucramiento e identificación con el sistema al sentirse participes en su desarrollo.</p>
PLAN DE CONTINGENCIA	<p>Si se presenta la situación en que el usuario se sienta afectado por la incorporación de esta nueva herramienta de trabajo, la única solución posible es la capacitación del usuario una vez implementado el sistema, remarcando las ventajas que le brindaría su utilización (ejemplo: la seguridad y la efectividad que brinda esta herramienta, la disminución de los tiempos en la realización de las tareas, entre otras).</p>

Nombre del riesgo: Falencias en la etapa de análisis.

MAGNITUD	Alta
-----------------	------

DESCRIPCIÓN	El riesgo se debe a que la etapa fundamental para un correcto desarrollo de un sistema, corresponde a un análisis minucioso y detallado, necesario para identificar las funcionalidades que deberá incluir el sistema y las características distintivas del entorno en el cual se desarrollará.
IMPACTO	Este riesgo impacta sobre los tiempos de desarrollo e implementación del sistema.
INDICADORES	El surgimiento de requerimientos o funcionalidades que no fueron identificados durante el análisis y que entorpecen la continuidad en el desarrollo al tener que darles resolución, y así evitar falencias mayores.
ESTRATEGIA DE MITIGACIÓN	La etapa de análisis deberá ser considerada la más importante en nuestro trabajo, porque si bien es la que más tiempo requiere, es la fundamental a la hora de comenzar con el desarrollo, ya que nos servirá de guía y apoyo a lo largo de nuestro trabajo.
PLAN DE CONTINGENCIA	Si surge una nueva necesidad o funcionalidad que no fue tomada en cuenta durante la etapa de análisis, por un error o negligencia de los desarrolladores, se deberá evaluar la posibilidad de trabajar horas extras para salvar dicha situación, sin que por ello se ocasionen demoras en la entrega final del sistema.

2.9. Flujo de requisitos

Este capítulo permite identificar y describir los casos de uso más importantes del sistema, que permiten identificar las distintas funcionalidades del software.

Un caso de uso es una secuencia de interacciones que se desarrollarán entre un sistema y sus actores en respuesta a un evento que inicia un actor principal sobre el propio sistema. A su vez, los diagramas de casos de uso sirven para

especificar la comunicación y el comportamiento de un sistema mediante su interacción con los usuarios y/u otros sistemas.

El alcance de este documento, abarca desde una identificación de los casos de uso, su descripción (a trazo grueso y fino) y la esquematización de los mismos, incluyendo, además, actores y sus relaciones.

2.9.1. Caso de Uso Trazo Grueso Vs. Trazo Fino

Para identificar los casos de uso, primero se definen todos los casos de uso del sistema para luego expresarlos en “trazo grueso”. Para el actual sistema, especificaremos los 3 casos de uso más importantes que son “Alta de Insumos”, “Modificación de Insumos” y “Baja de Insumos”, para lo cual:

- Se ignoran detalles sobre la forma de interacción entre el actor y el sistema;
- Solo se incluyen las alternativas más relevantes;
- No se incluyen detalles.

Luego se expresan en “trazo fino” aquellos que se definen y detallan una vez que se ha tomado la decisión de implementarlos. Aquí, se tienen en cuenta todos los detalles que no fueron contemplados en el trazo grueso:

- Se detallan las interacciones de los usuarios con el sistema;
- Se incluyen alternativas (excepciones o errores que provienen de requerimientos de los usuarios);
- Especificaciones con más detalle del comportamiento interno del sistema.

2.9.2 Caso de Uso: Trazo Grueso

2.9.2.1 Alta de Insumos

Especificación de casos de uso	
Nombre: Alta de insumos	Autor: Diego Flores Versión: 1.0

Nivel del CU:	X Negocio Sistema de Información
Objetivo: Ingresar un nuevo insumo a la base de datos del sistema.	
Prioridad: Alta	Actor Principal: Encargado
Tipo: Concreto	Actor Secundario:
Descripción: En caso de que el producto no haya sido dado de alta previamente en el sistema, se debe ingresar el nuevo producto al sistema.	
Fecha creación: 11/11/2015	Autor última modificación: 11/11/2015

2.9.2.2 Modificación de Insumos

Especificación de casos de uso	
Nombre: Modificación de insumos	Autor: Diego Flores Versión: 1.0
Nivel del CU:	X Negocio Sistema de Información
Objetivo: Modificar algún dato de un producto ya existente en el sistema.	
Prioridad: Alta	Actor Principal: Encargado
Tipo: Concreto	Actor Secundario:

Descripción:	
Se necesita cambiar un dato como puede ser la cantidad, el nombre o el precio de un producto que ya se encuentra dado de alta en el sistema	
Fecha creación: 11/11/2015	Autor última modificación: 11/11/2015

2.9.2.3 Baja de Insumos

Especificación de casos de uso	
Nombre: Baja de insumos	Autor: Diego Flores Versión: 1.0
Nivel del CU: X Negocio	Sistema de Información
Objetivo: Dar de baja un producto ya existente en la base de datos.	
Prioridad: Alta	Actor Principal: Encargado
Tipo: Concreto	Actor Secundario:
Descripción:	
Se necesita dar de baja o eliminar un producto que ya existe en la base de datos.	
Fecha creación: 12/11/2015	Autor última modificación: 12/11/2015

2.9.3 Caso de Uso: Trazo Fino

2.9.3.1 Alta de insumos

Especificación de casos de uso

Nombre: Alta de insumos	Autor: Diego Flores Versión: 1.0
Nivel del CU: <input type="checkbox"/> Negocio <input checked="" type="checkbox"/> Sistema de Información	
Objetivo: Dar de alta un nuevo producto en el sistema.	
Prioridad: Alta	Actor Principal: Encargado
Complejidad: Media	Actor Secundario: No aplica
Tipo: Concreto	Extiende: Usa:
Precondiciones: Ninguna.	
Post Condiciones: Si el caso de uso se realizó correctamente, se debe visualizar el nuevo producto con todos sus atributos en el listado superior de productos.	
Actor	Curso Alternativo
1. Una vez logueado, el Encargado de Sucursal selecciona la opción Insumos.	
2. El encargado deberá completar los campos ID Producto, Nombre y Cantidad y deberá seleccionar una categoría y la unidad en que se medirá el stock. Todos los datos son obligatorios.	
3. Una vez completados los datos requeridos, el encargado de sucursal oprime el botón "Agregar".	
4. El sistema valida que los datos, almacena la información e informa que los datos han sido guardados.	4.A Faltó de ingresar un dato en el producto 4.A.1 El sistema informa con un mensaje el error

	4.A.2 El encargado ingresa el campo faltante
5. Fin del caso de uso	

2.9.3.2 Modificación de insumos

Especificación de casos de uso	
Nombre: Modificación de insumos	Autor: Diego Flores Versión: 1.0
Nivel del CU: <input type="checkbox"/> Negocio	<input checked="" type="checkbox"/> Sistema de Información
Objetivo: verificar todas las planificaciones creadas	
Prioridad: Alta	Actor Principal: Planificador
Complejidad: Baja	Actor Secundario: No aplica
Tipo: Concreto	Extiende: Usa:
Precondiciones: El Encargado debe estar registrado en sistema con un usuario y contraseña, y el producto a modificar debe existir en la base de datos.	
Post Condiciones: El sistema permitirá ver los datos modificados en la lista de productos en la parte superior.	
Actor	Curso Alternativo
1. Una vez ingresado con usuario y contraseña, el Encargado de Sucursal selecciona la opción Insumos.	
2. Se selecciona un insumo de la lista, al seleccionarlo, los datos se cargan en los box y quedan disponibles para ser editados.	

3. Luego de modificarse los datos, el encargado de sucursal oprime el botón “Actualizar” para guardar los cambios en el producto modificado.	
4. El sistema valida que los datos ingresados, almacena la información e informa que los datos han sido guardados.	4.A Se dejó alguno de los campos en blanco 4.A.1 El sistema muestra mensaje de error.
5. Fin del caso de uso	

2.9.3.3 Baja de Insumos

Especificación de casos de uso	
Nombre: Baja de insumos	Autor: Diego Flores Versión: 1.0
Nivel del CU: <input type="checkbox"/> Negocio <input checked="" type="checkbox"/> Sistema de Información	
Objetivo: Eliminar un producto ya existente en el sistema.	
Prioridad: Alta	Actor Principal: Planificador
Complejidad: Media	Actor Secundario: No aplica
Tipo: Concreto	Extiende: Usa:
Precondiciones: El Encargado debe tener usuario y contraseña para ingresar al sistema, y el producto a dar de baja debe existir en la base de datos.	
Post Condiciones: Si el caso de uso se ha realizado correctamente el sistema muestra el listado de productos verificando que no se encuentra más en la base de datos el producto eliminado.	

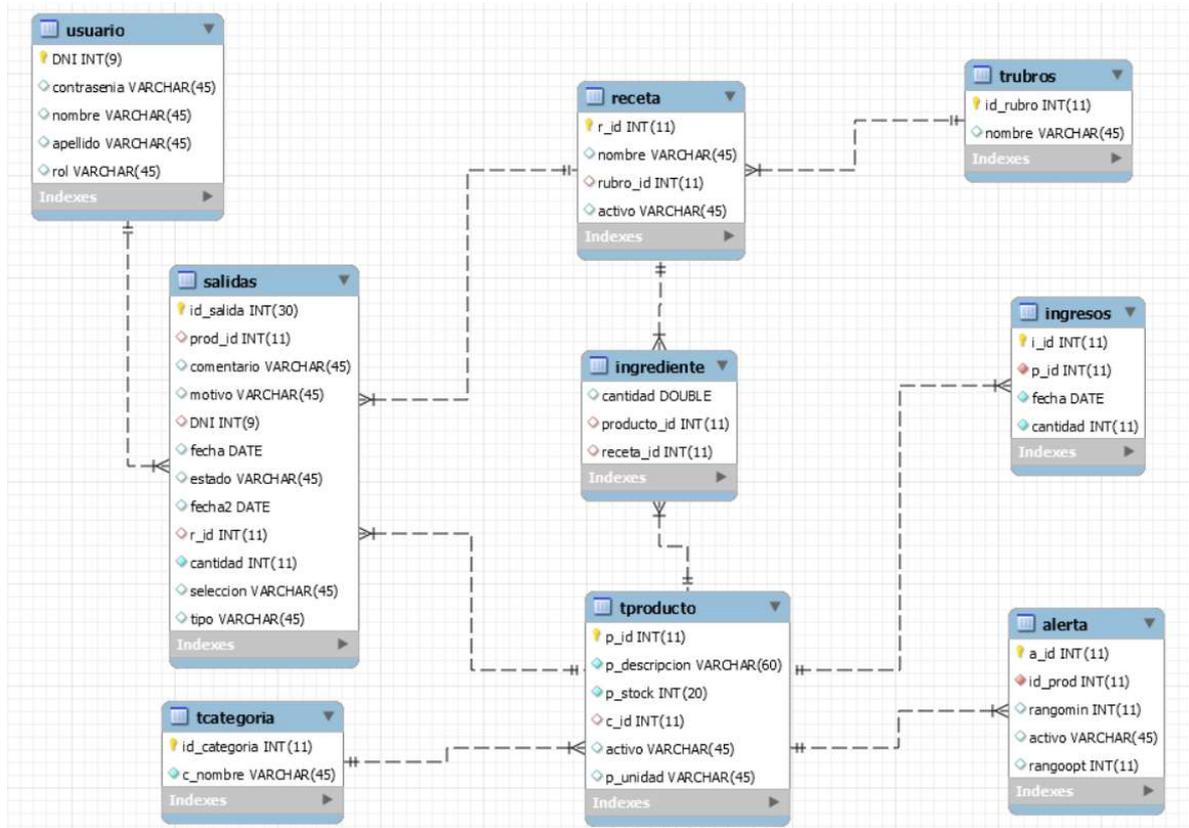
Actor	Curso Alternativo
1. Una vez ingresado con usuario y contraseña, el Encargado de Sucursal selecciona la opción Insumos.	
2. Se selecciona de la lista de insumos el insumo que se desea eliminar y se oprime el botón "Borrar".	
3. El sistema elimina los datos e informa que se realizó la tarea.	
4. Fin del caso de uso	

2.10. Flujo de Análisis

En esta sección se detallan las características fundamentales y los artefactos pertenecientes a la etapa de análisis.

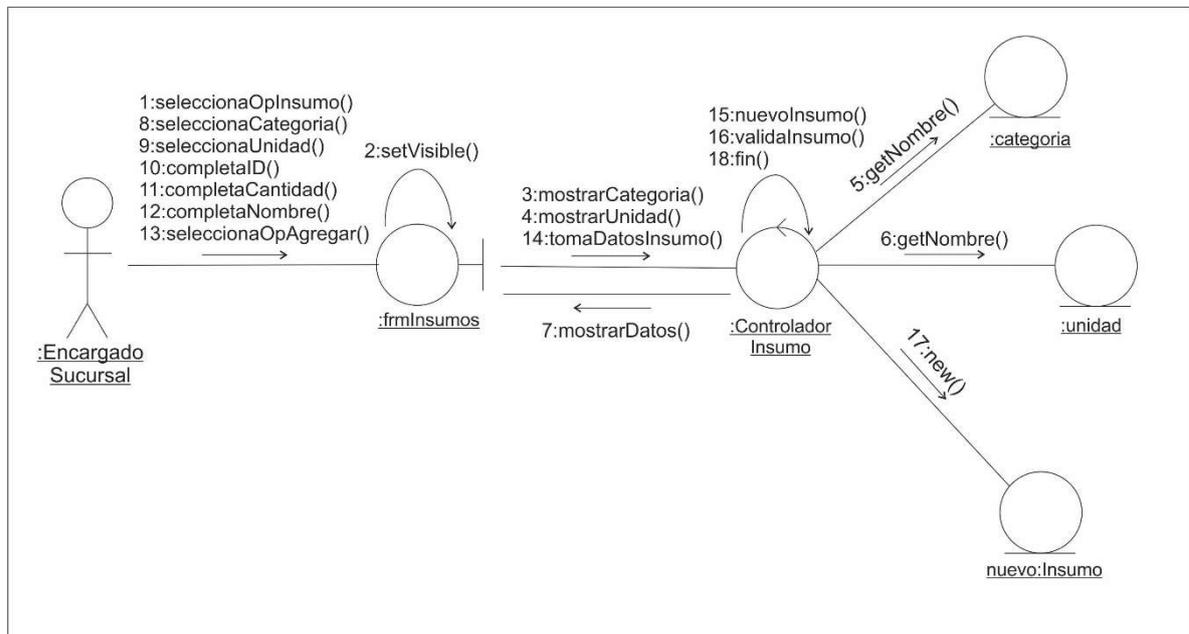
Se incluyen Diagramas de clases de análisis y Diagramas de Colaboración.

2.10.1 Diagrama de entidad relación (DER).

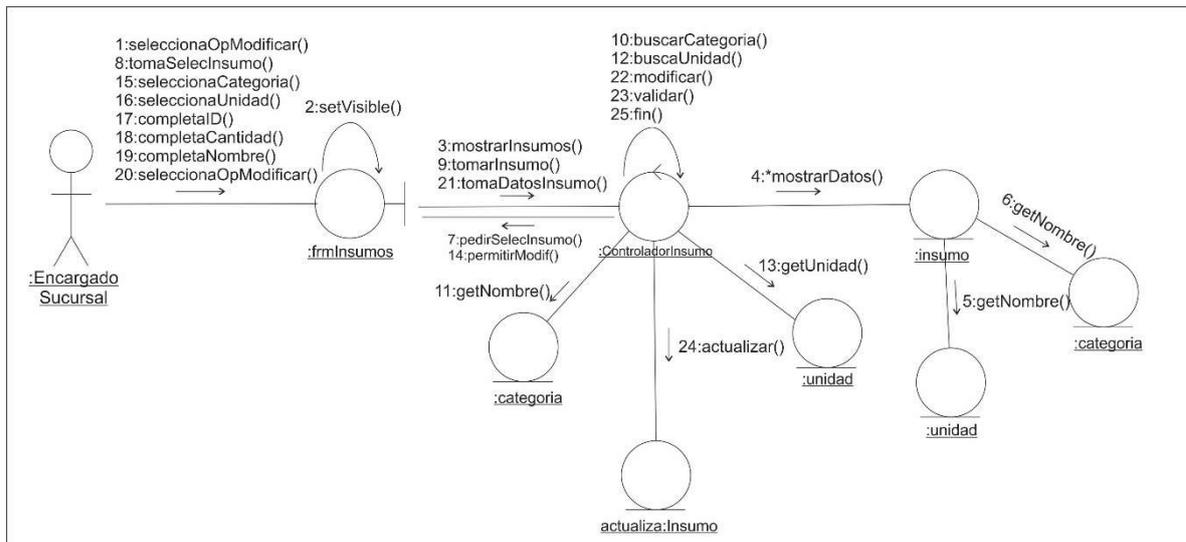


2.10.2 Diagramas de colaboración

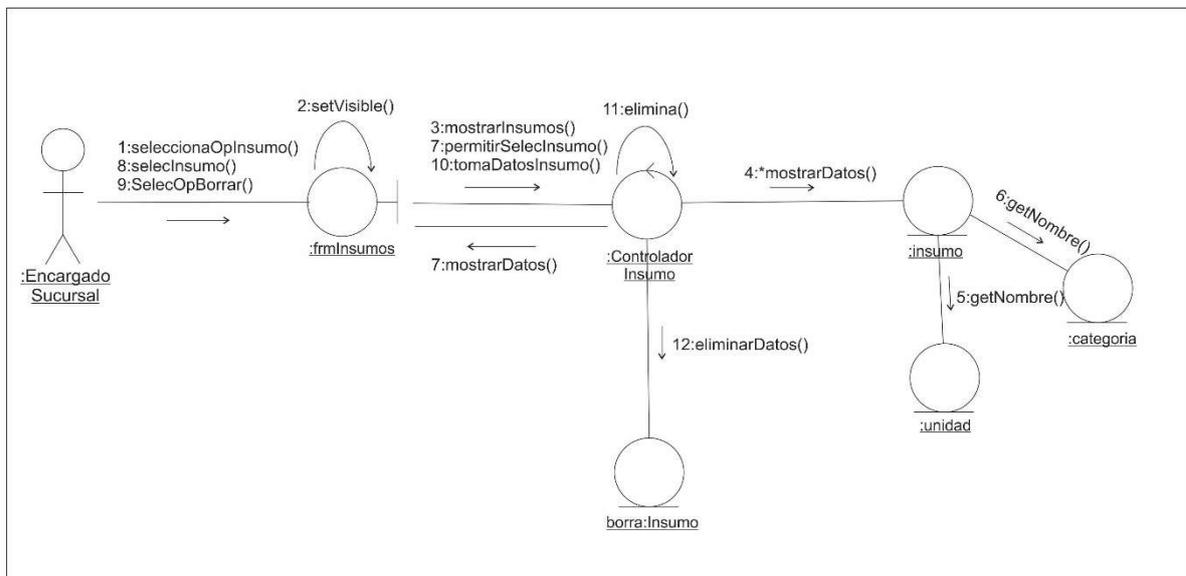
2.10.2.1 Caso de uso: Alta de producto



2.10.2.2 Caso de uso: Modificar producto.



2.10.2.3 Caso de uso: Borrar producto.



2.11. Flujo de diseño

2.11.1 Introducción

Un diagrama de clases sirve para visualizar las relaciones entre las clases que involucran el sistema y está compuesto por:

Clases:

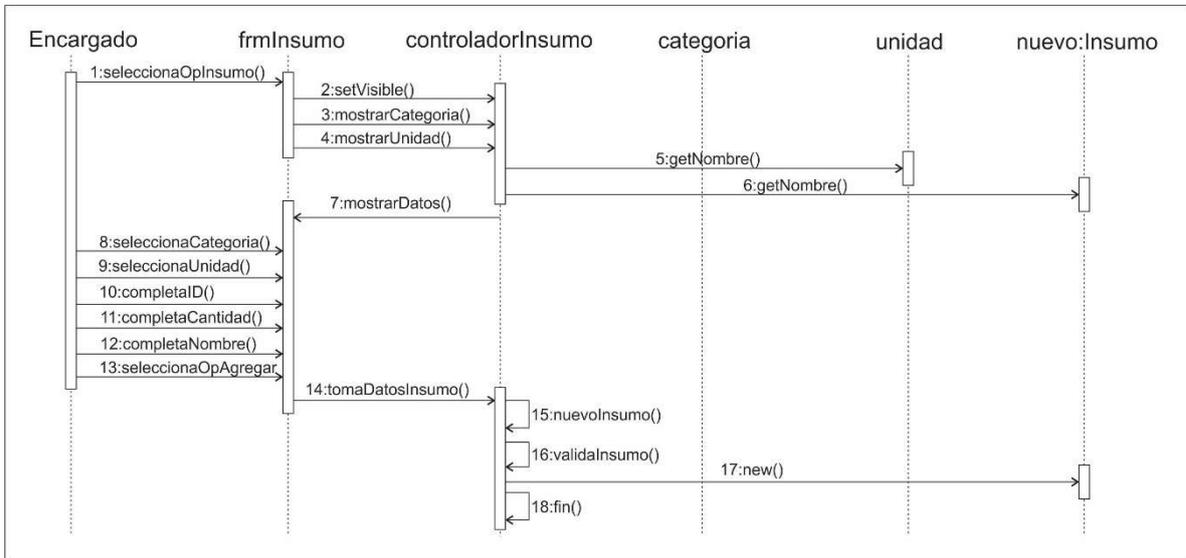
- ◆ Atributos
- ◆ Métodos

◆ Visibilidad

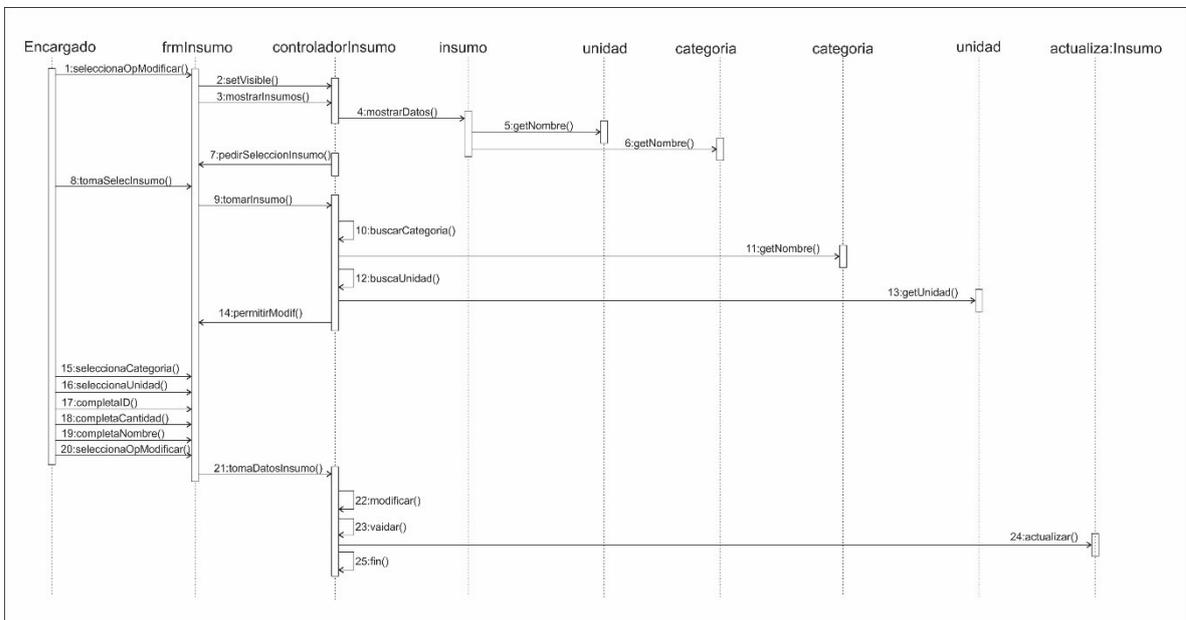
A continuación, se agregan los artefactos correspondientes a esta etapa.

2.11.2 Diagramas de secuencia

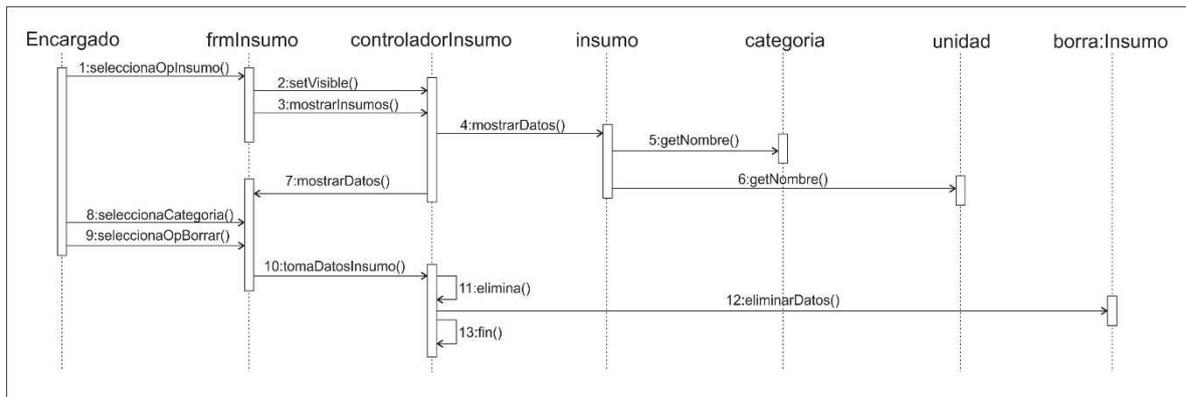
2.11.2.1 Caso de uso: Alta de producto



2.11.2.2 Caso de uso: Modificar producto



2.11.2.3 Caso de uso: Borrar producto



2.12. Descripción de la arquitectura

2.12.1. Introducción

En este apartado daremos a conocer la arquitectura del software a desarrollar, con el fin de entender la lógica que se utilizará para el desarrollo. Quedaran en evidencia los beneficios de la arquitectura utilizada y del patrón MVC para el desarrollo.

2.12.2. Arquitectura

El patrón de arquitectura de software seleccionado es MVC (Modelo – Vista – Controlador). En líneas generales, MVC es una propuesta de diseño de software utilizada para implementar sistemas donde se requiere el uso de interfaces de usuario. Surge de la necesidad de crear software más robusto con un ciclo de vida más adecuado, donde se potencie la facilidad de mantenimiento, reutilización del código y la separación de conceptos.

La arquitectura de Modelo – Vista – Controlador es el patrón de diseño que especifica cómo debe ser estructurada una aplicación, las capas que van a componer la misma y la funcionalidad de cada una. En este patrón, la capa intermedia de una aplicación puede ser dividida en tres grandes bloques funcionales: Modelo, Vista, Controlador.

Vista:

Se puede deducir de su nombre, la vista es la encargada de generar las respuestas que deben ser enviadas al cliente.

Cuando esta respuesta tiene que incluir datos proporcionados por el controlador, el código HTML de la página no será fijo si no que deberá ser generado de forma dinámica, por lo que sus implementaciones correrán a cargo de una página JSP. Las paginas JSP resultan mucho más adecuadas para la generación de las vistas que los servlets pues, al ser documentos de texto, resulta sencilla la inclusión de bloques estáticos HTML y de fácil mantenimiento por diseñadores web con escasos conocimientos de programación.

Controlador

Se puede decir que el controlador es el cerebro de la aplicación. Todas las peticiones a la capa intermedia que se realcen desde el cliente son dirigidas al controlador, cuya misión es determinar las acciones a realizar para cada una de estas peticiones e innovar al resto de los componentes de la aplicación (Modelo y Vista) para que realicen las acciones requeridas en cada caso, encargándose también de la coordinación de todo el proceso.

Modelo:

En la arquitectura MVC la lógica de negocio de la aplicación, incluyendo el acceso a los datos y su manipulación, esta encapsulada dentro del modelo. El Modelo lo forman una serie de componentes de negocio independientes del controlador y la vista, permitiendo así su reutilización y el desacoplamiento entre las capas.

En una aplicación J2EE el modelo puede ser implementado mediante clases estándares java o a través de Enterprise JavaBeans

Capa de presentación:

Es la que se encarga de que el sistema interactúe con el usuario y viceversa, muestra el sistema al usuario, le presenta la información y obtiene la información del usuario en un mínimo proceso. En el mundo de la informática es conocida como interfaz gráfica y debe tener las características de ser amigable u fácil de usar. Esta capa se comunica únicamente con la capa intermedia o de negocio.

Capa de negocio

Es donde residen las funciones que se ejecutan, se reciben las peticiones del usuario, se procesa la información y se envían las respuestas tras el proceso. Se denomina capa de negocio o capa de lógica de negocio, porque es aquí donde se establecen todas las reglas que deben cumplirse, Esta capa se comunica con la de presentación, para recibir las solicitudes y presentar los resultados, y con la capa de acceso a datos, para solicitar al gestor de base de datos almacenar o recuperar datos de él.

Capa de acceso a datos

Esta capa es la encargada de almacenar los datos del sistema y de los usuarios. Su función es almacenar y devolver datos a la capa de negocio, aunque para esto también es necesario en algunos casos, que tengan procedimientos almacenados y funciones dentro de la capa. En una arquitectura de tres capas, esta capa es la única que puede acceder a los mismos. Está formada por uno o varios sistemas gestores de bases de datos, localizados en un mismo servidor o en varios.

Estas capas, pueden estar localizadas todas en un mismo ordenador, si el programa o software informático que se desarrolla es de baja complejidad, pero si el mismo fuera de gran complejidad cada una de las capas puede estar situada en diferentes ordenadores.

2.12.3. Arquitectura de la Aplicación

Respecto a la tecnología implicada en el desarrollo del proyecto SPO, corresponde señalar a aquellas tecnologías que se relacionan con la plataforma Java Enterprise Edition JEE 6.

En particular, entre las tecnologías de JEE 6 que se usarán, podemos mencionar:

- Enterprise JavaBeans (EJB).
- Java Servlet
- Java Server Page (JSP)

- Java Server Pages Standard Tag Library (JSTL).
- Java Server Faces (JSF)
- Java Message Service (JMS).
- Java Transaction API (JTA).
- JavaMail API y JavaBeans Activation Framework (JAF).
- Tecnologías XML (JAXP, JAX-RPC, JAX-WS, JAXB, SAAJ, JAXR)
- JPA, JDBC API
- Java Naming and Directory Interface (JNDI)
- Java Authentication and Authorization Service (JAAS)

A su vez, las tecnologías declaradas implicarán el uso de una arquitectura multicapa característica de JEE:

- La capa del cliente (Client-tier) que será la capa destinada a mostrar la interfaz gráfica de usuario. La aplicación SPO, bajo plataforma Java EE, será una aplicación Web renderizada en un navegador web. Esta capa se ejecutará en la computadora cliente.
- La capa de la lógica de negocio (Business-tier) y la capa de la lógica de presentación (Web-tier). Estas capas se ejecutarán en el servidor de aplicaciones (Servidor Glassfish).
- La capa de los datos (Data-tier) que será la capa destinada a la gestión de los datos y se ejecutará en un servidor de base de datos (Servidor MySQL).

Así mismo se utilizarán dos librerías de componentes:

- Primefaces: conjunto de componentes enriquecidos que facilitan la creación de las aplicaciones web.
- JQuery: permite simplificar la manera de interactuar con los documentos HTML, manipular el árbol DOM, manejar eventos, desarrollar animaciones y agregar interacción con la técnica AJAX a páginas web

2.13. Flujo de pruebas

2.13.1. Introducción

En este apartado se analizarán las distintas pruebas a las que se someterá el sistema, para garantizar su correcto funcionamiento y que cumpla con los requerimientos previamente especificados.

Se utilizará la metodología de caja negra, la cual permite centrarse en la funcionalidad del software estudiándolo desde el punto de vista de las entradas que recibe y las salidas o respuestas que produce, dejando de lado el código fuente o su funcionamiento interno.

2.13.2. Tipo de Prueba

Se evaluarán los resultados de las pruebas de Sistema, que consiste en:

- Comparar los resultados obtenidos con los esperados.
- Identificar los problemas que pueden llegar a surgir, determinar el impacto de cada uno y que acciones tomar para resolverlo de la manera más eficiente.
- Indicar si el plan de prueba debe realizarse total o parcial, y si será necesario contemplar nuevos planes de prueba no contemplados anteriormente.

Entrada, Salidas y Participantes:

- Productos de entrada:
 - Resultado de las pruebas de Sistema
 - Plan de Prueba.
- Productos de Salida:
 - Evaluación del resultado de las pruebas.
- Participantes:
 - Analistas (Flores – Martínez)

2.13.3. Prueba: Proceso ABM de Productos

2.13.3.1 Caso de Prueba: Alta de Producto

Entrada

- ◆ Se debe registrar un nuevo producto en la base de datos.
- ◆ Se carga un nuevo producto indicando Nombre, Cantidad, Precio y Umbral de aviso.

Resultado

- ◆ El Producto debería grabarse en la tabla Productos, generando un nuevo ID para el Producto.
- ◆ El Producto debería poder visualizarse en el listado de stock actual.

Condiciones

- ◆ Los Productos pueden ser vistos por todos los usuarios existentes en la tabla Usuarios.

Procedimiento de Prueba

1. Seleccionar el botón “Productos” dentro de la pantalla principal del sistema. Se debe abrir una nueva ventana con el listado de productos que actualmente hay en stock.
2. Se cargan los datos del nuevo producto en las box correspondientes con todos los datos que se piden.
3. Una vez que se cargan los datos, seleccionar la opción “Agregar” para ingresar el nuevo producto en la base de datos.
4. En el listado de Productos de la parte superior se debe visualizar el nuevo producto agregado con todos sus atributos y la cantidad actual en stock.

2.13.3.2 Caso de prueba: Modificar Producto

Entrada

- ◆ Se necesita modificar el stock, el nombre o algún atributo de un producto en particular de la tabla Productos.
- ◆ El sistema permite visualizar el stock actual de Productos, para seleccionar el que se desea modificar directamente del listado.

Resultados

- ◆ La actualización del producto debería verse reflejada en la tabla Productos.
- ◆ El campo actualizado debería visualizarse en el listado de productos dentro de la ventana “Productos”.

Condiciones

- ◆ Cualquier usuario integrante de la tabla “Usuarios” debería poder cambiar el stock actual de Productos.

Procedimiento de Prueba

1. Seleccionar el producto a modificar del listado de Productos en la parte superior.
2. Las box con los atributos de cada producto se llenan con los atributos del producto seleccionado.
3. Se modifica el/los campo/s directamente desde la box y se presiona el botón “Modificar”.
4. El cambio debería verse reflejado en el listado de Productos de la parte superior.

Análisis, Diseño y Construcción de Data Warehouse

3. Marco Teórico

Inteligencia de negocios hace referencia a la realización de nuestros negocios de forma inteligente. Las técnicas milenarias han permitido tomar decisiones “por instinto”, y si bien hay muchos casos de éxito que han sido resonantes, poco conocemos de los numerosos casos que han fracasado.

En la actualidad las empresas han empezado a darle mayor importancia a las herramientas de soporte para la toma de decisiones como lo son los Data Warehouse.

Data Warehouse, es una de las herramientas que recién mencionamos y como su nombre lo indica es un almacén de datos que utiliza las bases de datos existentes en un negocio o empresa para nutrirse de datos. La principal función de un Data Warehouse es la transformación de datos en información, permitiendo analizar las variables objetivo desde diversos puntos de vistas o dimensiones.

En esta tesis implementaremos Inteligencia de Negocios en una empresa de delivery de comida. Nuestro objetivo es la creación de un área que trabaje sobre los requerimientos o necesidades que surjan de Marketing y Auditoría.

En este capítulo se brinda una descripción de todos los conceptos relacionados con la Inteligencia de Negocios que permiten llevar a cabo el presente Proyecto de Investigación.

3.1 Conceptos relacionados:

Para comprender el contexto en que se desarrollara el presente trabajo de tesis, es importante entender lo que es la Inteligencia de Negocios (Business Intelligence – BI) y todo lo que ésta implica, puesto que dichos conceptos van de la mano con el trabajo a desarrollarse.

3.1.1 Business Intelligence:

Desde un punto de vista pragmático, y asociándolo directamente con las tecnologías de la información, podemos definir Business Intelligence como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas...) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.

Más específicamente, es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

- Los principales productos de BI que existen hoy en día son:
- Cuadros de Mandos Integrales (CMI)
- Sistemas de Soporte a la Decisión (DSS)
- Sistemas de Información Ejecutiva (EIS)

Por otro lado, los principales componentes de orígenes de datos en el BI que existen en la actualidad son:

- Data Warehouse
- Data Mart

Los sistemas y componentes del BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa típicamente que, en un Data Warehouse, los datos están desnormalizados para apoyar consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y borrado de datos. En este sentido, los procesos ETL (extracción, transformación y carga), que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados.

- ✓ En definitiva, una solución de BI completa permite:
- ✓ Observar ¿Qué está ocurriendo?
- ✓ Comprender ¿Por qué ocurre?
- ✓ Predecir ¿Qué ocurriría?
- ✓ Colaborar ¿Qué debería hacer el equipo?
- ✓ Decidir ¿Qué camino elegir?

3.1.2 Data Warehouse:

Tras las dificultades de los sistemas tradicionales en satisfacer las necesidades informacionales, surge el concepto de Data Warehouse, como solución. Este término, se traduce literalmente como Almacén de Datos. No obstante, si el Data Warehouse fuese exclusivamente un almacén de datos, los problemas seguirían siendo los mismos que en los Centros de Información.

La ventaja principal de este tipo de sistemas se basa en la estructura de la información. Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales. Se caracteriza por ser:

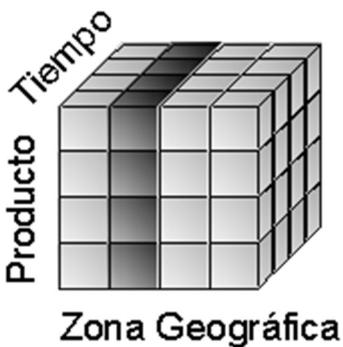
- Integrado: los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes

entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

- Temático: sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del Data Warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.
- Histórico: el tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- No volátil: el almacén de información de un Data Warehouse existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

E.F. Codd, considerado como el padre de las bases de datos relacionales, ha venido insistiendo desde principio de los noventa, que disponer de un sistema de bases de datos relacionales, no significa disponer de un soporte directo para la toma de decisiones. Muchas de estas decisiones se basan en un análisis de naturaleza multidimensional, que se intentan resolver con la tecnología no orientada para esta naturaleza. Este análisis multidimensional, parte de una visión de la información como dimensiones de negocio.

Dimensiones de negocio vistas como cubos

<p>Estas dimensiones de negocio se comprenden mejor fijando un ejemplo, para un sistema de gestión de expedientes, las jerarquías que se podrían manejar para el número de los mismos para las dimensiones: zona geográfica, tipo de expediente y tiempo de resolución. La visión general de la información de ventas para estas dimensiones definidas, la representaremos, gráficamente como el cubo de la derecha.</p>	
<p>Un gerente de una zona estaría interesado en visualizar la información para su zona en el tiempo para todos los productos que distribuye, lo podría tener una representación gráfica como el cubo de la derecha.</p>	
<p>Un director de producto, sin embargo, querría examinar la distribución geográfica de sus productos, para toda la información histórica almacenada en el Data Warehouse. Esto se podría representar como la figura de la</p>	

<p>derecha, que refleja la selección de datos por la jerarquía de producto.</p>	
<p>O se podría también examinar los datos en un determinado momento o una visión particularizada.</p>	

Para comprender el concepto de Data Warehouse, es importante considerar los procesos que lo conforman. A continuación, se describen dichos procesos clave en la gestión de un Data Warehouse:

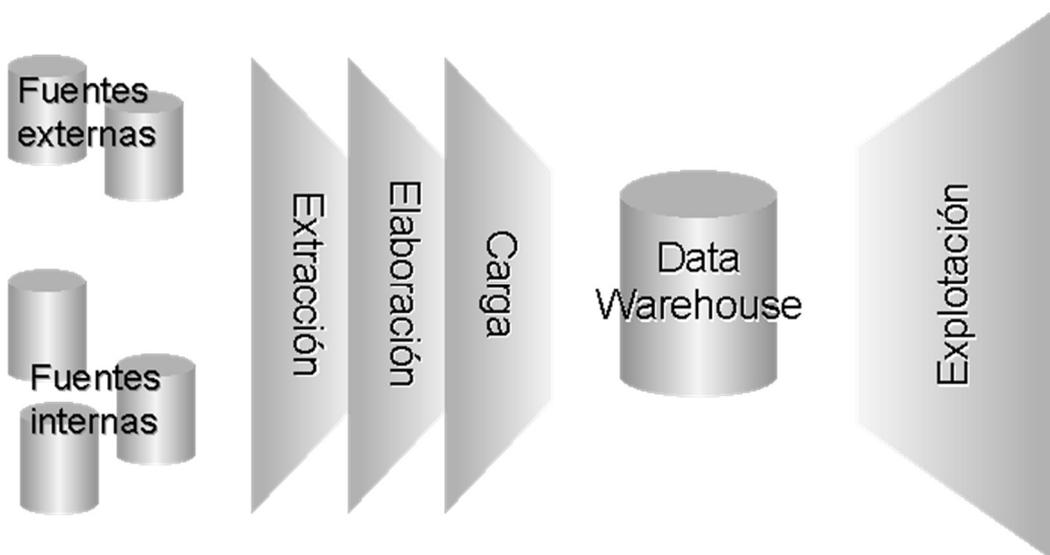


Figura 1 Proceso ETL

- Extracción: obtención de información de las distintas fuentes tanto internas como externas.
- Elaboración: filtrado, limpieza, depuración, homogeneización y agrupación de la información.

- Carga: organización y actualización de los datos y los metadatos en la base de datos.
- Explotación: extracción y análisis de la información en los distintos niveles de agrupación.

Desde el punto de vista del usuario, el único proceso visible es la explotación del almacén de datos, aunque el éxito del Data Warehouse radica en los tres procesos iniciales que alimentan la información del mismo y suponen el mayor porcentaje de esfuerzo (en torno a un 80%) a la hora de desarrollar el almacén.

Las diferencias de un Data Warehouse con un sistema tradicional las podríamos resumir en el siguiente esquema:

SISTEMA TRADICIONAL	DATA WAREHOUSE
Predomina la actualización	Predomina la consulta
La actividad más importante es de tipo operativo (día a día)	La actividad más importante es el análisis y la decisión estratégica
Predomina el proceso puntual	Predomina el proceso masivo
Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Importancia del dato actual	Importancia del dato histórico
Importante del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva
Estructura relacional	Visión multidimensional
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos

Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa relacionada con el negocio
---	---

Tabla 1 Comparación de modelos

Una de las claves del éxito en la construcción de un Data Warehouse es el desarrollo de forma gradual, seleccionando a un área como piloto y expandiendo progresivamente el almacén de datos a las demás áreas. Por ello es importante elegir este usuario inicial o piloto, siendo importante que sea un departamento con pocos usuarios, en el que la necesidad de este tipo de sistemas es muy alta y se pueda obtener y medir resultados a corto plazo

En el libro “The Data Warehouse Toolkit” de Kimball & Ross, se plantean 6 requerimientos u objetivos que debe perseguir o alcanzar un Data Warehouse, estos son:

- Debe hacer que el acceso a la información de la organización sea simple y fácil, así como el contenido debe ser entendible y los datos deben ser intuitivos y obvios para el usuario final, y no para el desarrollador.
- Debe presentar la información de la organización de forma coherente y consistente, debe ser creíble y estar perfectamente ensamblada de una variedad de recursos alrededor de la organización.
- Debe ser adaptable y resistente al cambio y al inevitable paso del tiempo.
- Debe ser de fácil acceso, pero a la vez seguro, para proteger y resguardar información importante y sensible de la empresa.
- Debe servir como base para una mejor toma de decisiones, desde el momento en que se muestra la información.
- Toda la organización debe aceptar el Data Warehouse como una poderosa herramienta y continuar con su uso luego de terminada la implementación del mismo.

Procesos de extracción, transformación y carga de los datos

Para obtener el almacén de datos, se desarrollan en forma secuencial la extracción de los datos, su transformación y finalmente su carga en el Data

Warehouse. El proceso de extracción consiste en la obtención de la información desde las distintas fuentes (bases de datos y archivos operacionales) tanto internas como externas mediante herramientas de gestión de datos.

A continuación, es necesario transformar los datos en los requeridos para el almacén. El proceso consiste en filtrado, limpieza, depuración, homogeneización e integración de la información. Esto debe hacerse, ya que las bases de datos operacionales, diseñadas para el soporte de varias aplicaciones de producción, frecuentemente difieren en el formato, por lo cual pueden tenerse los mismos elementos de datos, pero nombres y formatos y codificaciones incoherentes. Todas estas inconsistencias deben resolverse antes de realizar el último paso de este proceso que corresponde a la carga de los datos en el Data Warehouse.

Explotación

Es importante recordar que el almacén de datos no es un fin en sí mismo, sino que es un medio para solucionar una necesidad: el análisis de información y la toma de decisiones a través de los datos de la empresa, objetivo que se logra con el proceso de explotación del almacén de datos.

En esta etapa es donde se desarrolla la inteligencia del Negocio, y, por lo tanto, es un componente esencial del Data Warehouse, ya que es el punto de contacto directo con el usuario final, quien es el encargado de tomar decisiones o acciones que redundan en beneficio de la compañía y en el ROI (Retorno de la Inversión) del Data Warehouse.

Las herramientas utilizadas para el desarrollo de inteligencia del negocio pueden incluir software de consultas, generadores de reportes, procesamiento analítico en línea, herramientas de minería de datos, etc., dependiendo de los tipos de usuarios y sus requerimientos particulares. Sin embargo, se hace necesaria la integración de varias herramientas puesto que una sola no satisface todos los requerimientos.

Los niveles de aplicaciones típicas en esta etapa son: Consultas e Informes, OLAP, minería de datos, Sistemas de Información Ejecutiva, y Visualización geográfica.

En este trabajo se utilizará solamente una herramienta de Consultas e Informes desarrollada en JAVA para que los dueños puedan consultar los informes y reportes obtenidos.

3.1.3 Data Mart:

Otro concepto de mucha importancia en lo que es BI es el de Data Mart. Un Data Mart, es un subconjunto del Data Warehouse, con un alcance de contenido limitado. Éste se usa para un solo departamento de una organización y/o un problema particular de análisis dentro de la organización. Un Data Mart por sí solo, no es un Data Warehouse, ya que un Data Warehouse tiene más usuarios y más temas que un Data Mart, y provee una vista completa de las áreas funcionales de la organización. Un Data Mart, al igual que un Data Warehouse, consiste en una base de datos.

3.1.4 Sistema de Soporte de Decisiones:

Un Sistema de Soporte de Decisiones (DSS- Decision Support System) contiene todos los servicios y procesos, para seleccionar, manipular, y analizar información y presentar resultados. Debe de permitir acceso transparente a la data en varias partes del Data Warehouse y proveer una interfaz común para los diferentes grupos de usuarios. Un DSS también puede ser definido como un sistema computacional diseñado para apoyar en los procesos de la toma de decisiones en una organización. Un DSS es la ventana del usuario a los datos almacenados en el ambiente del Data Warehouse.

3.1.5 OLAP:

En lo que es consultas para la presentación de los datos, es importante explicar el concepto de OLAP (Online Analytical Processing). OLAP proporciona un modelo de datos intuitivo y conceptual, para que los usuarios que no tengan experiencia como analistas puedan comprender y relacionar los datos mostrados. Este modelo es llamado análisis multidimensional, siendo habilitado para ver los datos a través de múltiples filtros, o dimensiones. Los sistemas OLAP organizan los datos directamente como estructuras multidimensionales, incluyendo herramientas fáciles de usar por usuarios para conseguir la

información en múltiples y simultáneas vistas dimensionales. OLAP es también rápido para el usuario.

Rápidos tiempos de respuesta permiten que los gerentes y analistas puedan preguntar y resolver más situaciones en un corto período de tiempo. Una dimensión es una vista de los datos categóricamente consistente. Una característica de las dimensiones es la habilidad de hacer slice-and-dice. Slice (rebanada) y dice (cubo) hacen particiones de los datos en una base de datos multidimensional de acuerdo a los valores de ciertas dimensiones. Otra capacidad inherente en el diseño de OLAP es la rotación y anidamiento (Pivoting-and-Nesting) de las dimensiones. El pivoted permite rotar los datos desde las columnas hasta las filas. También es importante mencionar el concepto de drill, que en los sistemas OLAP tiene un significado muy específico. Drill down es la acción de seleccionar un miembro para ver el siguiente nivel inferior de detalle en la jerarquía. Drill up es seleccionar un miembro para ver el siguiente nivel superior, esto es, una acción de bottom-up. La mecánica o funcionamiento de las interfaces OLAP, especialmente pointing-and-clicking (apuntar y seleccionar) para hacer drill-down dentro de las capas de interés se hace posible por la velocidad con que las consultas son resueltas. Esta funcionalidad permite por completo a los gerentes y analistas un nuevo proceso para tratar con grandes cantidades de datos, un proceso conocido con el nombre de análisis ad hoc. En resumen, los sistemas OLAP organizan los datos por intersecciones multidimensionales. Esta organización, acompañada por una herramienta de interface para rotar y anidar dimensiones, permite a los usuarios visualizar rápidamente valores en detalle, patrones, variaciones y anomalías en los datos que estarían de otra manera ocultos por un análisis dimensional simple. A mayor número de dimensiones (dentro de los límites razonables), mayor es la profundidad del análisis.

3.1.6 Otras definiciones:

- Dimensión: es un grupo de miembros consistentes categóricamente representados como una arista específica de un cubo OLAP, por ejemplo, el tiempo, clientes, productos.

- Jerarquía: es la organización de niveles dentro de una dimensión que refleje: cómo los datos añadidos están agregados nivel a nivel, y el camino que permita hacer drill-down de arriba abajo dentro de la dimensión. Por ejemplo: año, trimestre y mes.
- Miembro: es el nombre o etiqueta para cualquier miembro de cualquier nivel en una jerarquía. Los miembros inferiores son llamados algunas veces miembros hoja o miembros de nivel cero.
- Generación jerárquica: este término se utiliza para describir las relaciones entre miembros de una jerarquía. Lo más común es usar nombres de familia, como los siguientes:
 - Hijo: es un miembro directamente subordinado o por debajo de otro miembro en una jerarquía.
 - Padre: es un miembro que está directamente encima de otro miembro en una jerarquía.
 - Hermano (Sibling): es un miembro que está al mismo nivel de uno o más miembros compartiendo el mismo padre.
 - Descendiente: cualquier miembro en cualquier nivel en relación a otro miembro específico.
 - Ancestro: cualquier miembro de cualquier nivel superior en relación a otro miembro.

4. Herramientas de Data Warehouse

4.1 Motor de bases de datos MySQL

MySQL es el servidor de bases de datos relacionales libre más popular, desarrollado y proporcionado por la compañía sueca MySQL AB, que mantiene el copyright del código fuente del servidor SQL, así como también de la marca. Una de las razones para el rápido crecimiento de popularidad de MySQL, es que se trata de un producto Open Source, y, por lo tanto, va de la mano con este movimiento.

MySQL es un sistema de gestión de bases de datos relacional, licenciado bajo la GPL de la GNU. Su diseño multihilo le permite soportar una gran carga de forma muy eficiente. MySQL AB distribuye una versión comercial de MySQL, que no se diferencia de la versión libre más que en el soporte técnico que se ofrece, y la posibilidad de integrar este gestor en un software propietario, ya que, de no ser así, se vulneraría la licencia GPL.

MySQL es una herramienta Open Source. Es posible descargar el software de MySQL de Internet y usarlo sin pagar por ello. Inclusive, es posible estudiar el código fuente y cambiarlo de acuerdo a cualquier necesidad.

MySQL usa la licencia GPL (Licencia Pública General GNU), para definir qué es lo que se puede y no se puede hacer con el software para diferentes situaciones. Sin embargo, si se quiere tener otra licencia para incorporar código de MySQL en una aplicación comercial, es posible comprar una versión de MySQL con una licencia comercial.

4.1.1 MySQL Workbench – Administrador de base de datos

MySQL Workbench es una aplicación para el diseño y documentación de bases de datos (sucesora de la aplicación DBDesigner4) pensada para ser usada con el sistema de gestión de bases de datos MySQL (recién adquirido por Sun Microsystems). Existen dos versiones del producto, una es open source y la otra es una versión comercial. La versión comercial proporciona algunas funcionalidades que pueden resultar de interés en algún ámbito, aunque la

versión open source es más que suficiente para la realización de la práctica, sumado a que la misma no genera ningún costo adicional en el trabajo.

4.2 Solución BI

En la actualidad la exigencia de mercado y la demanda de clientes sumado con la competencia global, ha vuelto necesario obtener información de manera rápida y con la mayor veracidad, por esta razón cada vez más organizaciones acuden a soluciones de inteligencia de negocios y así mejorar la toma de decisiones y la agilidad comercial.

Por esta exigencia, la comunidad open source brinda programas libres para generar inteligencia artificial que se orientan en la solución y se centran en los procesos, es así como una plataforma permite crear soluciones complejas a problemas de negocio, las soluciones que muestra este tipo de herramientas se componen de infraestructura de herramientas de análisis e informes integrados, lo cual asegura que se pueda crear las reglas de negocio necesarias. Una muestra clara de solución es la herramienta Pentaho.

4.2.1 Pentaho

Open Source Pentaho Business Intelligence, se define como una plataforma “orientada a la solución” y “centrada en procesos”, su objetivo principalmente se basa en cubrir amplias necesidades de análisis de datos y reportes empresariales, agilitando de esta manera la toma de decisiones. Las soluciones de Pentaho están escritas en Java al igual que su ambiente de implementación, lo que permite que sea flexible y cubra una amplia gama de necesidades ya sean específicas o sofisticadas. Brinda la posibilidad de que la información sea clocada en la intranet de la empresa.

Pentaho abarca las soluciones brindando diferentes opciones como son: reportes, análisis, dashboards, minería de datos, consultas, procesos ETL.

Características de Pentaho

Pentaho es una plataforma de Inteligencia de Negocios orientada a la solución y centrada en los procesos, que permite presentar información rápida y oportuna,

realizar análisis de datos para generar informes empresariales, todo gracias a la explotación de datos.

Las soluciones Pentaho, está escrita en Java y su ambiente de implementación está basado también en Java, construido bajo un servidor de aplicaciones compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic y Oracle AS, lo que asegura la escalabilidad, integración y portabilidad. Además, maneja servicios críticos tales como programación, seguridad, integración automatización y flujo de trabajo.

- ✓ Maneja la Integración con los procesos del negocio
- ✓ Permite administrar y programar reportes
- ✓ Maneja administración para la seguridad de usuarios

Se puede trabajar con base de datos: vía JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird; no posee dependencia de sistema operativo; sus componentes están expuestos vía Web Services lo que apoya en la integración con Arquitecturas Orientadas a Servicios (SOA).

Arquitectura de Pentaho

La solución Business Intelligence OpenSource Pentaho es una alternativa a las soluciones de negocio tradicionales más completas por lo que incluye varios componentes para agilizar la manera de obtener información y analizarla. Los informes se generan mediante la herramienta JfreeReport en formatos (HTML, Excel, pdf). En cuanto a la generación de PDFs lo realiza mediante Apache FOP, gracias a la librería JPivot, con la cual se pueden ver tablas OLAP a través de un browser.

Integra Pentaho Report Design Wizard, que es una herramienta que permite el diseño de informes, para la generación de gráficos es a través de JFreeChart, una librería para generar los gráficos, interfaces para acceder a diferentes fuentes de datos, exportación a PNG, JPEG y PDF y soporte para servlets, JSPs, applets y aplicaciones clientes. Las soluciones que se elaboran se pueden publicar sin ningún inconveniente como servicio web, gracias al motor de servicios web Apache Axis, esta es una característica fundamental de Pentaho.

Los reportes que se han definido son ejecutados por el **Pentaho Reporting Engine** (PRE), que es una librería de reportes en Java que puede ser embebida tanto en aplicaciones Java de escritorio o servidor. El Reporting Engine maneja tanto el Pentaho Report Designer como el Pentaho BI-Server, que es una aplicación web de Business Intelligence, usada para publicar reportes y otros contenidos vía web

Formatos de salida para exportar Reportes:

Los reportes de Pentaho Reporting pueden ser exportados en: PDF, Excel, XHTML, De texto enriquecido (RTF), CSV, XML (vía API).

Características

- ✓ Permite la creación de reportes con la herramienta Report Designer.
- ✓ Permite agregar un dominio de metadatos fácil de usar con una fuente de datos a través del editor de metadatos.
- ✓ Pentaho Web Console, permite tener una fácil gestión de informes y vistas de análisis.
- ✓ Maneja conectividad entre las herramientas de cliente y el servidor de informes de BI, lo que permite que el contenido creado con el Editor de Metadatos, Report Designer y Design Studio, pueda ser publicado directamente en el servidor de BI.

Ésta herramienta no se utilizó debido a que se integró la generación de reportes en la aplicación Java desarrollada para la gestión y el control del stock dentro del negocio, permitiendo la unificación de los datos y la generación de los reportes directamente desde la misma, reduciendo así los tiempos operativos y facilitando el uso de la misma.

4.2.2 Kettle – Pentaho Data Integration

Pentaho Data Integration (PDI) es un motor de transformación, diseñado para realizar procesos de integración de datos.

KETTLE - “K Extraction Transformation Transportation Load E”, es una potente herramienta capaz de manipular y transformar información con gran volumen de datos. Realiza procesos de extracción, manipulación, validación y carga de datos desde múltiples fuentes de origen y en diferentes entornos, lo que permite evitar grandes cargas de trabajo manualmente difícil de mantener y de desplegar.

Hace uso de una interfaz gráfica “SPOON” que permite diseñar “jobs” o tareas de transformación de datos para ser procesados bajo las herramientas de Kettle (PAN y KITCHEN).

Arquitectura de Pentaho Data Integration

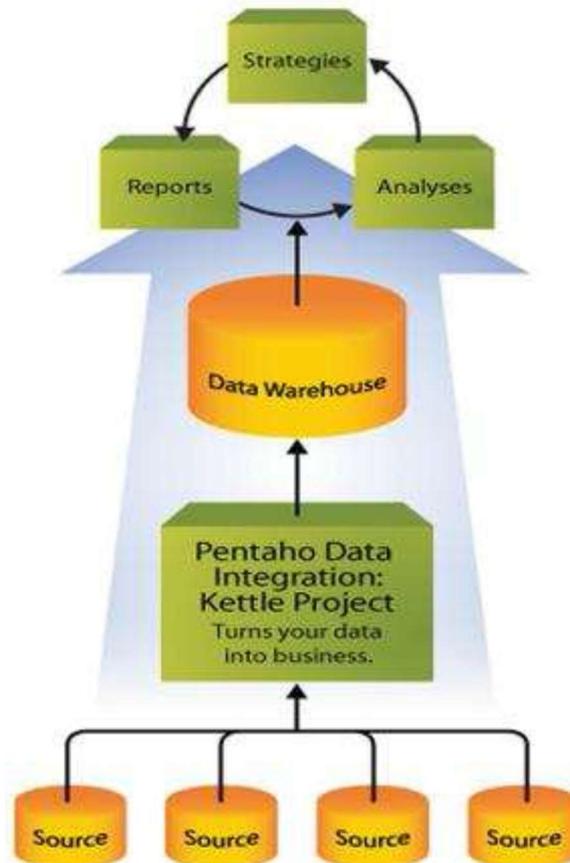


Figura 2 Pentaho Data Integration

Propiedades básicas

- Incluye un entorno gráfico de desarrollo
- Maneja tecnologías: Java, XML, JavaScript
- Es independiente del sistema operativo
- Está Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones)
- Incluye cuatro herramientas:
 1. SPOON: Diseñar transformaciones ETL usando el entorno gráfico
 2. PAN: Ejecuta transformaciones diseñadas con spoon, es un motor de transformación de datos que permite realizar acciones de la lectura y escritura sobre diversas fuentes de datos
 3. CHEF: Permite crear trabajos o “Jobs” programados
 4. KITCHEN: Es una aplicación que permite ejecutar “Jobs” o tareas planificadas en modo “batch” (serie, lote) que corren automáticamente en intervalos de tiempo establecidos.

Muchas organizaciones manejan gran cantidad de datos alojada en distintas aplicaciones y base de datos diferentes, Pentaho Data Integration abre, limpia e integra esta información poniéndola disponible para el usuario, generando consistencia al poder manejar una sola versión de todos los recursos de información

Características

- Proporciona algoritmos que automatizan los procesos de transformación de datos a la forma en que la minería de datos puede explotarlos. Los resultados se pueden visualizar en modo gráfico ya sea agrupado, segmentado, de árbol de decisión, bosque aleatorio, redes neurales y componentes de análisis.
- Utiliza filtros especiales para realizar procesos de discretización, normalización, re-muestreo “re-sampling”, selección y transformación de atributos.
- Kettle es una herramienta de PDI que permite implementar los procesos extracción, transformación y carga de datos (ETL).
- Los datos y sus transformaciones están separados. Las transformaciones y Jobs son almacenadas en archivos de formato XML, donde se detallan las acciones a realizar en los datos. Para construir las transformaciones, se utilizan los componentes, que se enlazan entre sí mediante saltos (steps), para determinar el flujo de datos entre los diferentes componentes.
- Incluye herramientas para realizar consultas, generación de reportes, análisis interactivo, tableros de mando, extracción, transformación y carga de datos (ETL), integración de datos, minería de datos (data mining), y un servidor para la plataforma de BI.

4.3 Netbeans

Netbeans es un entorno de desarrollo gratuito y de código abierto (Open Source) que permite el uso de un amplio rango de tecnologías de desarrollo tanto para escritorio, aplicaciones Web o para dispositivos móviles. Permite el desarrollo en lenguajes de programación como Java, PHP, Groovy, C/C++, HTML5; Además

puede instalarse en diferentes sistemas operativos tales como Windows, Linux y Mac OS.

Características principales

- Permite mantener actualizado el entorno Java, ya que brinda soporte a casi todas las novedades en este lenguaje, que es uno de los más utilizados en la actualidad.
- Incluye asistentes para la creación y configuración de distintos proyectos, incluida la elección de algunos frameworks.
- Posee un editor de código simple de usar, multilenguaje, con sugerencias de código, acceso a clases directamente desde el código, control de versiones, localización de ubicación de la clase actual, comprobaciones sintácticas y semánticas, plantillas de código, herramientas de refactorización, etc.
- Simplifica la gestión de grandes proyectos con el uso de diferentes vistas, asistentes de ayuda, y estructurando la visualización de manera ordenada, lo que ayuda en el trabajo diario.
- Herramientas para depurado de errores: incluye un debugger que permite visualizar el funcionamiento del proyecto paso por paso para verificar y controlar errores.
- Optimización de código: incluye la herramienta Profiler que ayuda a optimizar las aplicaciones para que se ejecuten más rápido y con el mínimo uso de memoria.
- Acceso a base de datos: desde la misma aplicación permite conectarse a distintos sistemas gestores de bases de datos como Oracle y MySQL, para poder visualizar las tablas, realizar consultas y modificaciones, y todo integrado en el propio IDE.
- Plugins: gracias a la comunidad Open Source, existen infinidad de librerías para instalar en la aplicación, simplificando el uso de la misma, y a su vez, potenciando y agregando funcionalidades útiles para la gestión del proyecto.

Se decidió usar esta herramienta principalmente porque es de código abierto y no implica gastos en el proyecto, y además porque estamos familiarizados con la misma, lo que redujo considerablemente las horas de desarrollo que se utilizaron en la creación del programa de “Stockia”.

4.3.1 Librería para reportes JasperReports y JFreeChart

- JasperReports: es una poderosa y flexible solución de código abierto para la generación y gestión de informes. Se instala como un módulo o plugin dentro de Netbeans y permite, por medio de un entorno gráfico, diseñar todo tipo de reportes para luego ser mostrados por pantalla, con la posibilidad de exportarlos a PDF, Excel o Word.
- JFreeChart: es un plugin que se integra a Netbeans, permitiendo la generación de gráficos estadísticos dentro de aplicaciones Java para incluirlos dentro de reportes y luego imprimirlos en pantalla.

Éstas herramientas se utilizaron en la aplicación de “Stockia” desarrollada en Java, para permitir al usuario generar reportes que se nutren de los datos del Data Warehouse, y así satisfacer los requerimientos definidos en el proyecto de investigación.

5. Metodología para el desarrollo de un Data Warehouse

La metodología para el desarrollo del proyecto será la establecida por Ralph Kimball, que es considerado como uno de los padres de este concepto. Kimball se ha dedicado al desarrollo de su metodología para que este concepto sea correctamente aplicado en las organizaciones, y se asegure la calidad de este tipo de proyectos. Durante su carrera ha innovado, escrito libros, educado y ha sido consultor en el campo de las bodegas de datos.

Kimball ha establecido ciertos procesos para llevar al éxito un proyecto de Data Warehouse o Data Mart. Para su desarrollo se incluyen varias tareas que pueden ser realizadas en paralelo o en forma secuencial. El correcto desarrollo de cada una de las fases planteadas en esta metodología garantiza la calidad y el correcto proceso de desarrollo.

5.1 Planeación y administración del proyecto

Definición del proyecto

Existen varios escenarios posibles en los que surge un proyecto de Data Warehouse para una empresa. Es importante identificar el escenario para determinar el alcance y definición del proyecto. Los escenarios, originados por una demanda del proyecto en una empresa son los siguientes:

- Demanda de un sector del negocio: En este escenario, un ejecutivo del negocio tiene el propósito de obtener mejor información con un mejor acceso para tomar mejores decisiones.
- Demasiada demanda de información: En este escenario, existen múltiples ejecutivos del negocio buscando mejor información.
- En busca de demanda. En este escenario usualmente está involucrado el presidente de una empresa, quien no identifica la necesidad de realizar un proyecto de Data Warehouse para su negocio, pero desea incorporar este sistema por requerimientos o necesidades del negocio.

Al identificar el escenario, es posible determinar si existe demanda para el proyecto y de donde proviene la misma. El primer caso es el ideal, ya que se tienen objetivos claros y con un alcance determinado de lo que se quiere del proyecto. El segundo escenario es riesgoso, pues para implementar un almacén de datos que soporte varios requerimientos de diferentes áreas de la empresa, se necesita mucho tiempo, dinero y soporte interno de la organización que perdure en el tiempo. En el tercer escenario se deben buscar los requerimientos de los cuales puede derivar la solución y basar en ellos el proyecto. En todos los escenarios es determinante contar con sponsors o patrocinadores internos del proyecto para lograr el éxito. Si no se cuenta con un patrocinador interno de la empresa involucrado en la demanda es preferible posponer el proyecto. Por último, luego de identificar el escenario, es importante conocer si la empresa está lista para realizar este proyecto.

Determinar la preparación de la empresa para un proyecto de bodega de datos

De acuerdo a Ralph Kimball existen cinco factores que deben existir en una organización para iniciar un proyecto de bodega de datos.

- Patrocinio de la gerencia del negocio: Al contar con este patrocinio, se tiene una visión del impacto que tendrá el almacén de datos en la empresa. Los gerentes son líderes influyentes dentro de la organización y determinarán el apoyo y soporte al proyecto de los demás miembros de la organización. Es preferible tener varios patrocinadores que uno solo, en caso de cambios en la organización o necesidad de un mayor apoyo.
- Motivación del negocio: Al implementar un almacén de datos se busca encontrar un sentido de emergencia por parte de la organización, causado por una motivación del negocio. Un ejemplo de motivadores son la competencia y la visión competitiva. Otras organizaciones han encontrado el motivador en una crisis. Un motivador importante también es un mercado potencial. Lo importante para un proyecto de almacén de datos es alinearse con uno de estos motivadores estratégicos del negocio.

- Acompañamiento del departamento de tecnología y de negocio: El éxito de un proyecto de almacén de datos se produce gracias a un esfuerzo de las áreas de tecnología y de negocio, compartiendo responsabilidades.
- Presencia de cultura analítica: Es importante que las decisiones de la organización se basen en hechos, más que en simples intuiciones, y que estas decisiones sean determinantes y recompensadas.
- Factibilidad: Es preferible que la infraestructura que soporte el almacén de datos esté presente y sea robusta. La primera factibilidad debe ser la de los datos. Si estos se encuentran sucios o no cumplen con estándares, el proyecto tendrá retrasos respecto al cronograma planeado.

Desarrollo del enfoque preliminar

Luego de haber determinado la preparación de la organización para el proyecto, se debe centrar el proyecto en su enfoque, y justificarlo para recibir el apoyo y presupuesto de desarrollo. Para determinar el enfoque, se deben responder preguntas como: ¿Se busca el enfoque y presupuesto para cubrir el levantamiento de requerimientos y diseño? ¿O para una primera versión del almacén? ¿O para el proyecto completo?

Para definir este enfoque, la base deben ser los requerimientos del negocio, no un cronograma. Para la definición del enfoque es importante seguir los siguientes parámetros:

- La definición del enfoque es responsabilidad del departamento de tecnología y del negocio: El enfoque usualmente se establece para desarrollar requerimientos específicos del negocio, en un tiempo determinado.
- El enfoque inicial del proyecto debe ser factible y manejable: Es preferible empezar “pequeño” y luego continuar el proceso de forma iterativa, lanzando pequeños desarrollos del proyecto.
- Enfoque inicial en un solo requerimiento del negocio soportado por una sola fuente de datos.
- Limitar el número de usuarios que tendrán acceso al almacén de datos inicialmente.

- Establecer criterios de éxito del proyecto mientras se define el enfoque:
Se refiere a entender lo que la gerencia espera del proyecto.

Una vez que el área de tecnología y negocios han acordado un enfoque, este se debe documentar.

Desarrollar la justificación del negocio

Luego de haber definido el enfoque, la justificación debe ser establecida. Esto significa que se identifican anticipadamente los costos y beneficios asociados al proyecto. Una forma de hacer esto es con el factor retorno de la inversión (ROI), que consiste en comparar el retorno financiero esperado (beneficios del negocio) contra la inversión esperada (costos).

Se deben considerar las siguientes inversiones y costos:

- Compras de licencias de software y hardware.
- Costos de mantenimiento: muchos productos de hardware y software requieren mantenimiento.
- Recursos internos de desarrollo.
- Recursos externos requeridos.
- Capacitación para desarrolladores y usuarios.
- Soporte a usuarios.
- Costos de crecimiento: Por cambios en requerimientos y actualizaciones.

Se deben considerar los siguientes retornos y beneficios:

Los proyectos de almacén de datos típicamente tienen un impacto en el incremento de ingresos y ganancias, más que en reducción de costos.

- Incremento de ingresos por nuevas ventas a nuevos y antiguos clientes.
- Incremento de ganancias por aumento de respuestas a la publicidad.
- Incremento de niveles de servicio al cliente.
- Descubrimiento de nuevas oportunidades.

Planeación del proyecto

El proyecto de almacén de datos debe tener un nombre que lo identifique como tal, luego se identifican roles que pueden ser cubiertos por uno o varios integrantes del equipo y a su vez, cada miembro del equipo también puede desempeñar varios roles, dependiendo de los requerimientos y del tamaño del proyecto. Los siguientes roles se identifican para el proyecto:

- Patrocinadores de negocio.
- Gerente del proyecto: Responsable de la gerencia de tareas y actividades cotidianas.
- Líder de negocios del proyecto: Con el gerente del proyecto monitorea el proyecto y lo comunica a la organización. Tiene un alto entendimiento de los requerimientos del negocio.
- Analista del sistema de negocios: Lidera las actividades de definición de requerimientos.
- Modelador de datos: responsable del análisis de datos y el modelo dimensional.
- Administrador de bases de datos de la bodega (DBA): Responsable de determinar agregaciones, particiones y soporte a la base de datos.
- Diseñador de proceso ETL: Responsable del diseño de la extracción, transformación y carga de la bodega.
- Desarrolladores de aplicación al usuario.
- Educador del almacén de datos.

Desarrollo del plan de proyecto

El objetivo de planear es el de proveer el detalle suficiente para hacer seguimiento al progreso del proyecto. Se identifican actividades, recursos y tiempos para el desarrollo. También permite monitorear los procesos y tener un plan de riesgos.

Administración del proyecto

Se consideran las reuniones de equipo, monitoreo del estado, el enfoque y estrategias de comunicación. Para las reuniones se debe seguir una agenda y

mantener un ambiente de comunicación entre el equipo. El monitoreo se debe realizar periódicamente, analizando el estado del proyecto en diferentes estados del tiempo.

5.2 Análisis de requerimientos

Acercamiento a la definición de requerimientos

Para entender mejor los requerimientos se debe empezar por hablar con los usuarios del negocio. No se debe preguntar a estos usuarios qué datos quieren que aparezcan en el Data Mart, sino hablar con ellos sobre sus trabajos, objetivos y retos e intentar conocer cómo toman decisiones, actualmente y en el futuro. Se debe considerar lo que requiere el negocio comparando estos requerimientos con los datos disponibles en las bases de datos que servirán como fuente para lograr el soporte de éstos.

Preparación de la entrevista

Se deben determinar roles y responsabilidades en el equipo entrevistador. Es preferible que el mismo equipo conduzca las entrevistas a usuarios del negocio y al equipo de tecnología de la empresa. Los roles que se deben manejar comprenden a un líder, encargado de dirigir el cuestionario, el cual debe tener habilidades en el conocimiento del negocio y comunicaciones. También debe existir un “escribano” encargado de tomar notas durante las entrevistas. Se debe tomar el mayor detalle posible del contenido. Al finalizar las entrevistas, esta persona debe hacer preguntas para aclarar dudas y obtener una retroalimentación de los entrevistados.

Investigación previa a entrevistas

Antes de iniciar el proceso de levantamiento de requerimientos, se deben analizar los reportes anuales de la compañía, para determinar las decisiones y hechos estratégicos. También es útil obtener planes de negocios de la compañía. También se debe analizar la competencia de la compañía y sus principales

fortalezas y debilidades. Si ha existido un intento anterior de desarrollar un almacén de datos de la compañía, este también se debe analizar.

Selección de los entrevistados

Se deben seleccionar personas representativas de cada área de la organización. Es importante observar el organigrama de la compañía para determinar los candidatos a entrevista. Los principales entrevistados deben ser los administradores ejecutivos del negocio, para comprender la estrategia en un alto nivel de la empresa. Luego es importante entrevistarse con los analistas del negocio de cada área quienes conocen el manejo de información que se lleva a cabo.

Desarrollo del cuestionario

El líder de la entrevista debe desarrollar el cuestionario antes de iniciar la entrevista. Se deben desarrollar varios cuestionarios que serán aplicados dependiendo del rol de los entrevistados dentro de la empresa. El cuestionario debe ser de una sola página, para evitar exceso de tiempo de entrevistas.

Es preferible iniciar las entrevistas en un nivel medio de jerarquía de la organización, en vez de iniciar desde la parte superior con las altas gerencias, pues en los mandos medios se maneja un mayor nivel de detalle respecto a los datos que sirven para luego definir la granularidad del almacén. Es importante que durante la entrevista se especifique terminología, la definición exacta de esta tendrá un gran impacto en la granularidad y dimensionalidad del modelo. Es posible que una palabra signifique muchas cosas, por eso lo importante es identificarlas y documentar estas inconsistencias en el vocabulario para luego confrontarlas con los entrevistados.

Inicio y desarrollo de la entrevista

La entrevista debe iniciarse con una introducción, para recordar al usuario sobre el proyecto y el equipo desarrollador. Los objetivos del proyecto y de la entrevista deben ser nombrados y los miembros del equipo presentados. Para documentar

información útil se debe preguntar a los usuarios sobre sus trabajos, por qué los hacen y cómo los hacen. Se deben realizar preguntas en un alto nivel y luego irse al detalle para obtener respuestas cada vez más específicas.

Al entrevistar ejecutivos, el principal objetivo es obtener una visión y entender globalmente el negocio. Al entrevistar administradores y analistas de la empresa, se buscan los objetivos y visión de cada departamento. En el área de auditoría y administración de datos se busca saber si existen los datos para poder dar soporte a los requerimientos encontrados en las entrevistas previas. Se debe entender las definiciones de los campos de las bases de datos, granularidad, volúmenes de datos, y otros detalles de estas fuentes de información. Al cierre de las entrevistas se debe preguntar por los criterios de éxito del proyecto, de esta forma se entienden las actitudes y expectativas frente al proyecto. Estos criterios deben ser medibles y cuantificables.

Análisis de las entrevistas

Si algún miembro del equipo conoce los sistemas operativos fuente de la empresa, debe explicarlos al resto del equipo para determinar la factibilidad de implementar los requerimientos encontrados. Se deben resaltar los descubrimientos y requerimientos clave para el proyecto. Se deben analizar y repasar los reportes y análisis reunidos en las entrevistas, lo cual comúnmente conlleva a una aproximación del descubrimiento de dimensiones para el modelo. Para finalizar, es importante documentar los requerimientos obtenidos y comunicarlos a los usuarios para adquirir su aprobación y compromiso.

5.3 Modelamiento dimensional

Modelo entidad relación

El modelo entidad relación (ER) es una técnica poderosa para diseñar lógicamente sistemas para el procesamiento de transacciones OLTP (procesamiento transaccional en línea). Siempre va encaminado a la eliminación de la redundancia, lo que permite que la manipulación sobre la base de datos tenga que hacerse en un solo lugar y sea mucho más rápido ya que en el

momento en que la transacción requiera insertar, adicionar, borrar o modificar un dato es necesario que lo haga en un solo lugar y no en múltiples. Esto contribuye también al almacenamiento de grandes cantidades de datos dentro de las bases de datos relacionales, a través del proceso de normalización. Por eso es perfecto para la inserción y actualización de la información. Este es un modelo excelente para registrar transacciones y administración de tareas operativas, sin embargo, para el modelamiento de un almacén de datos presenta varios problemas. Los usuarios finales no entienden ni recuerdan un diagrama entidad relación.

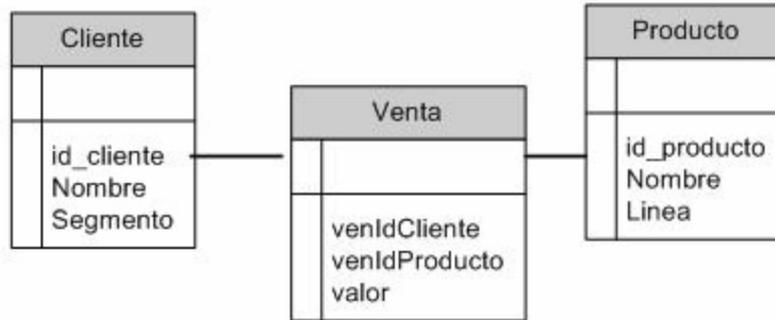
Nos es posible que los usuarios finales naveguen sobre el modelo. El uso del modelo entidad relación va en contra del objetivo principal de un almacén de datos, de proporcionar datos de forma intuitiva y con un buen desempeño y tiempos de respuesta.

Modelo Dimensional

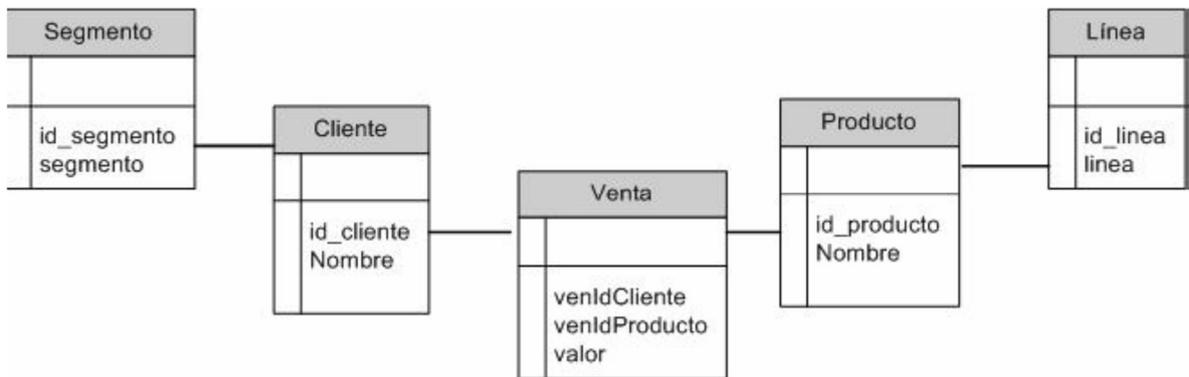
El modelo dimensional es una técnica de diseño lógico que busca presentar los datos de una forma intuitiva y que proporcione acceso de alto desempeño. Cada modelo dimensional se compone de una tabla con múltiples llaves foráneas, llamada tabla de hechos (facttable), y un conjunto de tablas más pequeñas, llamadas tablas de dimensión.

Los atributos de las tablas de dimensión son las fuentes de las restricciones de búsqueda necesarias para consultar un almacén de datos. Son utilizadas como título de atributo de las filas resultantes de queries de SQL.

Existen dos modelos dimensionales que predominan en las soluciones de almacén de datos: el modelo estrella y el modelo copo de nieve. En el modelo estrella, como se ve en la figura, se tiene una tabla de hechos y en ella llaves foráneas a cada una de las tablas de dimensión que tiene el modelo. Es decir, cada tabla dimensional está directamente relacionada a la tabla de hechos.



Una dimensión es modelada de forma copo de nieve cuando los campos de baja cardinalidad de la dimensión han sido removidos a tablas separadas y unidas a la tabla original con llaves foráneas como se muestra en la figura. En este modelo la tabla de hechos no tendrá llaves foráneas a todas las demás tablas como en el caso de la estrella. Las nuevas tablas no estarán conectadas con la tabla de hechos sino con las dimensionales establecidas.



Generalmente el copo de nieve no es recomendado en ambientes de bodegas de datos. Este modelo será más complejo para los usuarios y la navegación por el modelo será más lenta.

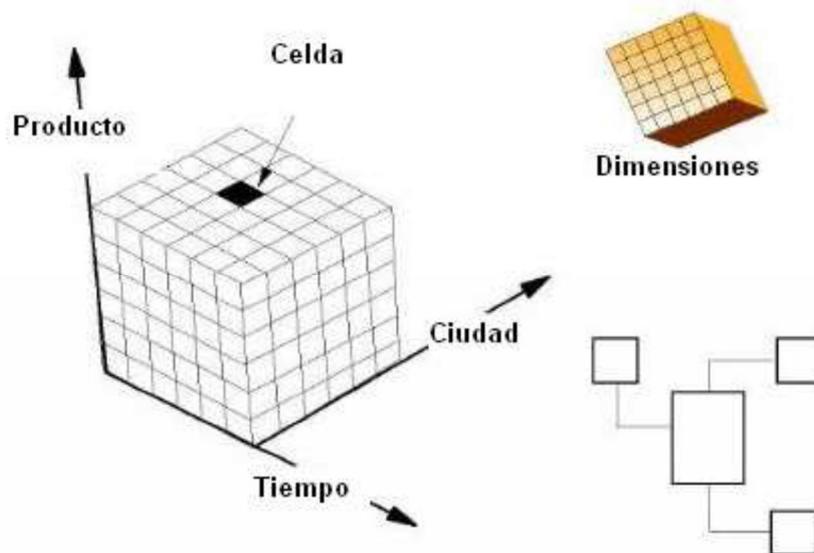
Diseño de dimensiones y hechos

En el desarrollo de un almacén o Data Mart, comúnmente es necesario unir Data Mart. Esto se logra creando una arquitectura de bus de Data Mart. Como se utilizarán las mismas tablas de dimensiones, es importante que las tablas de

dimensiones y hechos cumplan con las mismas especificaciones, como su granularidad. Estas dimensiones son llamadas dimensiones conformes (*conformed dimensions*). Se caracterizan por cumplir estas condiciones:

1. Una tabla de dimensión puede ser usada con cualquier tabla de hechos de la misma base de datos.
2. Las interfaces de usuario y contenido de datos son consistentes para cualquier uso de la dimensión.
3. Hay una interpretación consistente de atributos, por lo tanto, se obtiene la misma interpretación de la tabla en cualquier Data Mart.

La granularidad es un factor que se debe tener en cuenta para el diseño de las tablas. Un almacén de datos debe mantener sus dimensiones basadas en datos con la mayor granularidad posible. De esta forma se facilita la expansibilidad de los Data Mart para contener nuevos atributos, ya sea en las tablas de dimensiones o en la de hechos. Además, se permite de esta manera realizar técnicas de minería sobre la bodega, las cuales comúnmente requieren una alta granularidad.



Al diseñar las tablas de hechos y dimensiones, la idea principal es permitir que cada dato del negocio sea representado como un cubo. Donde las celdas del cubo contienen valores medidos y los bordes del cubo definen las dimensiones

de los datos. Comúnmente al modelar datos de negocios se obtienen entre 4 y 15 dimensiones. En el ejemplo de la figura de arriba, se modela un cubo con las dimensiones Producto, Tiempo y Ciudad. A la derecha de la figura aparece el modelo dimensional que representa al cubo, con las tres dimensiones unidas a la tabla de hechos.

Hechos

Los hechos son medidas de las variables que se consideran. Un hecho puede ser el valor de una factura con sus respectivas relaciones: la factura es generada a un cliente, correspondiente a un producto, creada en una sucursal. Al seleccionar los hechos para el diagrama dimensional, se debe sospechar que cualquier valor numérico, especialmente si es de tipo flotante, es probablemente un hecho. Es de especial utilidad que cada hecho sea aditivo, para los análisis propios de la bodega.

Dimensiones

Los atributos de tipo texto que describen cosas son organizados en dimensiones. Es necesario establecer un criterio puramente de diseño y basado en los requerimientos del negocio para establecer los atributos que se incluyen como dimensiones y los que se pueden descartar al realizar la bodega de datos.

Los atributos dimensionales, servirán como fuente para las restricciones y encabezados de filas en los reportes. Todos los atributos dimensionales son igualmente candidatos a ser encabezados de filas en los reportes. Al agregar restricciones a una búsqueda o reporte, se está haciendo un drilling down, es decir se está estableciendo una nueva restricción en una búsqueda para obtener un mayor nivel de detalle. Un drill down eficiente mezcla atributos de las diferentes dimensiones, para realizar reportes realmente robustos.

Llaves subrogadas

Todas las llaves de las tablas de la bodega de datos deben ser llaves subrogadas, es decir no deben significar nada respecto a las características de

su contenido ni a su fuente en los sistemas fuente. No se deben utilizar las llaves originales de un sistema fuente del cual fueron extraídas. Estas llaves subrogadas se manejan como enteros.

Método de diseño de una tabla de hechos

Para el diseño de la tabla de hechos, de acuerdo a la metodología de Ralph Kimball se deben seguir los siguientes pasos, de cada paso depende el siguiente. Se deben tomar decisiones respecto a:

1. El Data Mart.
2. La granularidad de la tabla de hechos.
3. Las dimensiones
4. Los hechos

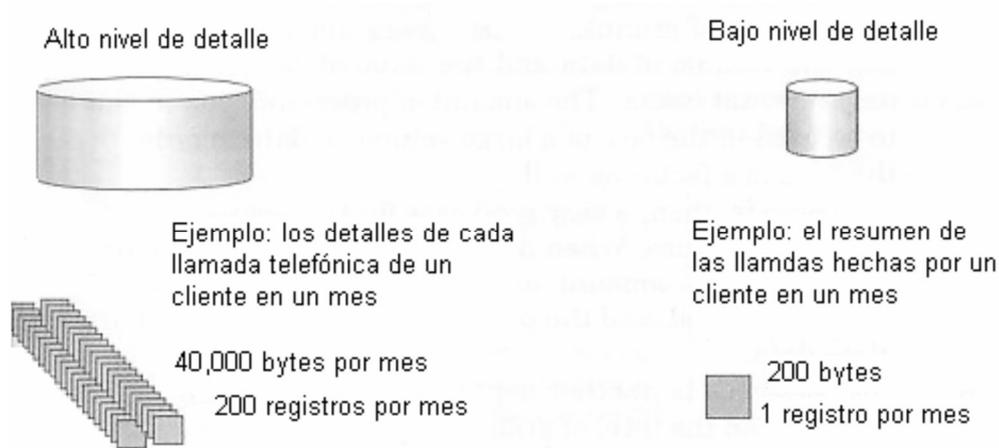
1. Selección del Data Mart.

Para un correcto desarrollo de un almacén, es preferible seleccionar e implementar primero los Data marts que dependan de una sola fuente y luego continuar con los que deben extraer datos de múltiples fuentes. En el caso más simple, seleccionar el Data Mart es lo mismo que seleccionar la fuente legacy de datos.

2. Declaración de granularidad de la tabla de hechos.

Es necesario definir claramente lo que es un registro de la tabla de hechos en el diseño dimensional propuesto. La granularidad es la respuesta a la pregunta ¿Qué es un registro de la tabla de hechos?

La granularidad se refiere al nivel de detalle existente en las unidades de los datos de la bodega. Entre más detalle haya, menor será el nivel de la granularidad. Entre menos detalle haya, mayor será la granularidad. Es un factor determinante en el desarrollo de la bodega, debido a que de ella depende el volumen de datos que será almacenado en la bodega y el tipo de queries que pueden ser realizados.



Generalmente la granularidad de la tabla de hechos es seleccionada como la más baja o granular posible. Existen varias ventajas para seleccionar este tipo de granularidad como una transacción individual o una línea de documento. De esta forma será más fácil responder a nuevas consultas y agregar nuevos datos.

3. Selección de dimensiones.

Generalmente la granularidad determina unas dimensiones mínimas e iniciales. Al agregar nuevas dimensiones los atributos de estas deben cumplir con la misma granularidad que se ha definido. La granularidad de una dimensión no puede ser menor que la granularidad de la tabla de hechos.

4. Selección de los hechos.

La selección de granularidad de la tabla de hechos también permite seleccionar los hechos. En el caso de tabla de hechos de transacciones habrá un solo hecho, el monto de la transacción. Si el caso es una tabla de hechos de snapshot puede contener diversos resúmenes de las actividades realizadas en la toma del snapshot. Como cantidad vendida, valor, costo.

Los hechos siempre deben ser específicos a la granularidad de la tabla de hechos.

5.4 Diseño técnico de la arquitectura

En los sistemas de información la definición de una arquitectura permite hacer un desarrollo más confiable y eficiente. Con la definición de la arquitectura se

mejora la comunicación entre las diferentes áreas del proyecto, el planeamiento del proyecto, la flexibilidad y el mantenimiento del mismo.

Aspectos de arquitectura

Para hacer el diseño de la arquitectura se debe comenzar analizando los sistemas legacy actuales, estos deben ser consistentes y manejar de forma correcta sus transacciones, pues en la metodología del desarrollo del DWH (Data Warehouse) se toma como hecho que estos sistemas son confiables.

5.5 Procesos de Extracción, Transformación y Carga

Este proceso comprende varios aspectos determinantes para el almacén de datos. Por lo tanto, se debe seguir un plan para su correcto desarrollo. Se establecen varios pasos que conducen al desarrollo del proceso y se describen a continuación.

Paso 1. Plan de alto nivel

El proceso de diseño se inicia con un esquema simple de los componentes del plan que son conocidos: las fuentes y los destinos de los datos. Se identifica de donde provienen los datos y las características y problemas con dichas fuentes. Con este esquema es posible comunicar la complejidad del proyecto a la gerencia y miembros del equipo de desarrollo del proyecto.

Las aplicaciones de ETL realizan tres pasos: extracción, transformación y carga al Data Mart. Estos pasos se deben ver en un esquema de alto nivel: Tomar los datos de las fuentes, transformarlos y cargarlos en los destinos.

Paso 2. Herramientas ETL

Las extracciones típicamente se escriben en el lenguaje de la fuente de los datos. Existen herramientas que realizan todo el proceso de extracción, transformación y carga que buscan minimizar el tiempo requerido para estas tareas. Estas herramientas implican un costo por licencias y posibles incompatibilidades o dificultades con transformaciones complejas que fuesen requeridas para el proceso. Ya se haga el proceso con código desarrollado, o herramientas

existentes, es determinante realizar prácticas que mejoren el rendimiento del proceso, como ordenar los datos o cargarlos de forma rápida para cargas masivas en las bases de datos.

Paso 3. Plan detallado

El plan se inicia seleccionando las tablas en las que se va a trabajar, en qué orden y secuenciar las transformaciones para cada conjunto de datos. Se debe graficar un diagrama con estas estructuras.

Todas las tablas de dimensión deben ser cargadas antes que las tablas de hechos. Se debe iniciar el desarrollo de la aplicación ETL con la dimensión más simple y continuar con las demás hasta llegar la tabla de hechos.

Paso 4. Poblar una tabla de dimensión simple

La principal razón para iniciar el proceso con una dimensión estática y simple es la facilidad para poblar esta tabla. A continuación, se presenta información para realizar el proceso completo de ETL relevante a cualquier tabla destino (dimensiones y hechos).

- Extracción de una dimensión: Se deben entender las características de la fuente de datos. Si la fuente es relevante para la dimensión destino, como se actualiza esta fuente, si hay alguna hora en el día en la cual no se deba acceder a esta fuente, son preguntas que se deben resolver en esta etapa. También se deben generar reportes “de extracción”, que permitan extraer únicamente los campos o datos requeridos de cada fuente.

Existen dos fuentes principales de datos: archivos y bases de datos. Si la fuente se encuentra en una base de datos, se crea un solo proceso que haga fluir los datos desde la fuente a la sección de transformación y carga. Si la fuente es un archivo, se requieren varios pasos en el proceso: extracción al archivo, ubicación del archivo en el servidor del proceso, transformar los datos del archivo y cargar el archivo en la base de datos utilizada para transformar y cargar datos.

- Transformación de una dimensión: Existen transformaciones comunes y sencillas, como el cambio de un tipo de dato a otro, o cambiar letras minúsculas por mayúsculas, otras más complejas se indican a continuación.

- Asignación de llaves substitutas (*surrogate* en inglés): Las llaves substitutas son las llaves primarias de las tablas del modelo dimensional, que son diferentes y no deben tener ninguna relación con las llaves primarias de las tablas fuentes ni de los datos mismos. Típicamente se utilizan enteros para estas llaves.

- Combinación de diferentes fuentes: Si un dato depende de diferentes fuentes se debe mapear el dato destino y los datos orígenes para controlar esta transformación.

- Carga

Para mejorar el rendimiento, la primera carga a la bodega se debe hacer con las características de población masiva de datos del motor de base de datos del modelo dimensional. Los principales motores de bases de datos tienen esta característica y su uso depende del producto utilizado. Otra forma de mejorar el rendimiento de carga es suprimir los índices en la base de datos, que son costosos para el rendimiento de la carga, y este costo no se recupera en mejora de rendimiento del análisis multidimensional de la bodega.

Paso 5. Implementación de la lógica del cambio de una dimensión

Al cambiar los datos de una dimensión, ya sean nuevos o actualizaciones a los datos previos, es preferible construir la extracción de tal forma, que se extraigan únicamente los datos que han cambiado. Al determinar los cambios, se debe contar con reglas del negocio que determinen como manejar estos cambios en los atributos. Si se determina que la modificación es legítima y es posible actualizar el dato, se utiliza la técnica de una dimensión cambiante.

El primer paso para preparar la dimensión consiste en determinar si ya existe un registro antiguo del dato. Al encontrar una modificación se aplica alguno de los siguientes tipos de respuesta al cambio.

- Tipo 1: Sobrescribir.

Se toma el nuevo dato de la fuente y se sobrescribe la tabla de la dimensión con el nuevo contenido.

- Tipo 2: Crear un nuevo registro de la dimensión.

Se crea un nuevo registro de la dimensión con una nueva llave subrogada.

- Tipo 3: Bajar el antiguo valor.

Se crea un nuevo campo en la dimensión que contenga el antiguo valor del dato modificado y se actualiza la dimensión con el nuevo contenido.

Cada tabla de dimensión usará una, dos o las tres técnicas para administrar los cambios de datos.

Paso 6. Poblar las dimensiones restantes

Para poblar el resto de dimensiones se sigue el proceso del paso 4, y se cargan ya sea con el uso de una herramienta de carga existente o con el desarrollo de una que soporte conexiones ODBC o JDBC.

Paso 7. Carga histórica de hechos

En el proceso de ETL debe existir un paso para reemplazar las llaves primarias de las fuentes por las llaves subrogadas que se han asignado a cada dimensión, y que deben ir como llaves foráneas en la tabla de hechos.

Para mantener una integridad referencial en la bodega, - que significa que por cada llave foránea en la tabla de hechos exista un registro en la dimensión correspondiente – se realizan dos búsquedas de llaves subrogadas en el proceso de extracción. Una ocurre cuando se utiliza la técnica 2 del paso 5, y se ha creado un nuevo registro de una dimensión que contiene una nueva llave subrogada y un campo modificado con el resto del contenido original. La segunda búsqueda ocurre cuando se están procesando los registros de la tabla de hechos. En este momento se debe buscar en todas las dimensiones los registros que coincidan con el dato que se está buscando relacionar con la tabla de hechos, y una vez encontrado el registro de la dimensión, agregar su llave subrogada a la llave foránea de la tabla de hechos.

Este proceso funciona bien para almacenes de datos que no manejan grandes cantidades de registros, pues al ser grande el tamaño de datos es preferible mantener una tabla de búsqueda que registre la llave primaria en la fuente y la

llave subrogada en la dimensión. De esta manera se mejora el rendimiento al poblar la tabla de hechos.

Una vez que se cuenta con todas las llaves foráneas de la tabla de hechos, esta tabla se empieza a cargar.

Paso 8. ETL de una tabla de hechos incremental

Al realizar cargas semanales o periódicas al Data Mart desde las fuentes, es decir al actualizar el almacén, se deben extraer y procesar únicamente las transacciones que han ocurrido luego de la última carga. Para seleccionar únicamente estas nuevas transacciones se pueden usar los logs de las bases de datos, y de esta forma identificar los datos que han cambiado en el período analizado. El log captura cada cambio que haya sido realizado a un registro. También es posible crear logs en las bases de datos fuentes que registren únicamente los cambios que son importantes para la bodega, utilizando triggers.

Paso 9. Operación y automatización de la bodega de datos

Idealmente el proceso ETL de una bodega de datos se ejecuta de manera automática y no atendida. Las operaciones típicas en la bodega que se deben tener en cuenta en este paso son las siguientes:

- Definición de tarea – Flujos y dependencias
- Horarios de tareas – basadas en tiempos y hechos
- Monitoreo
- Logging
- Manejo de excepciones
- Manejo de errores
- Notificaciones

Estas características deben ser proporcionadas para que el almacén de datos pueda ser un sistema de producción.

Procesos automáticos:

- Extracción. Para que todo el proceso ETL sea automático, el paso de extracción debe notificar al siguiente cuando empezar. Este proceso también debe registrar

su trabajo en una metadata, que contenga el nombre de la tarea realizada, el usuario, la fecha, tiempo tomado para la tarea y cualquier otro hecho que quiera ser monitoreado.

- Calidad y limpieza de datos. En todos los sistemas se encontrarán problemas de calidad y limpieza de datos, por lo tanto, se deben establecer políticas que determinen estándares de calidad aceptables y procedimientos a realizar cuando estos problemas se identifiquen.

Mejoramiento de datos.

Para garantizar el uso de los mejores datos posibles para el almacén, se deben tener en cuenta los siguientes pasos:

- Identificar la fuente de datos con la mejor calidad: Es posible que se encuentren varias fuentes con los mismos datos, pero en algunas se tenga mejor calidad de los mismos.
- Identificar variaciones en palabras: Como errores de ortografía, mayúsculas y minúsculas.
- Discutir problemas de datos con el equipo.
- Arreglar los problemas de datos en las fuentes cuando sea posible, en vez de hacerlo en el proceso ETL o directamente a la bodega.
- No arreglar todos los problemas. Si existen muchos problemas en las fuentes, arreglarlos en el proceso ETL irá en contra del rendimiento, estos problemas deben ser responsabilidad de los sistemas fuentes.
- Realizar tareas de limpieza sobre los datos.

Aseguramiento de la calidad

Una vez se tienen los datos, es importante determinar si este contenido es realmente correcto. Se pueden hacer varios procesos para determinar esto:

- Cruce de datos. Se ejecutan varias queries contra las fuentes de datos y se verifica que el resultado de estos queries sea el mismo que el dado con los datos seleccionados del proceso ETL
- Examen manual. Si existen varias fuentes para un mismo dato, o eventos de ese tipo, es preferible examinar los datos manualmente y compararlos con el resultado esperado.
- Validación del proceso. Al utilizar el almacén de datos es posible encontrar diferentes resultados de los que se harían con simples queries sobre las fuentes. Esto se da debido a la limpieza y transformaciones hechas a los datos en el proceso ETL. Por lo tanto, es importante identificar las causas de las diferencias y determinar cuál resultado es realmente el correcto.

5.6 Selección e instalación de productos

Selección de Productos

Una vez se tiene completa la arquitectura, se deben tener en cuenta dos componentes fundamentales para la selección de productos:

- Requerimientos técnicos.
- Requerimientos de negocio.

Mantener el negocio enfocado

Para la selección de la herramienta se debe hacer un riguroso estudio de cada una de ellas, mirar sus ventajas y desventajas, y saber el alcance que se quiere con cada una para saber las cosas que se pueden y no se pueden hacer.

Otro punto importante en la escogencia de la herramienta son los requerimientos de los usuarios: los requerimientos del negocio, mirar el análisis y el grado de procesamiento que se puede obtener con cada una ya sea un motor de base de datos, una herramienta de ETL o una herramienta OLAP o de reportes. También se deben mirar las limitaciones de los usuarios en lo referente al HW, pues no tiene sentido contratar una herramienta que consume muchos recursos cuando no se tiene la plataforma para soportarla.

Para la escogencia de las herramientas también se debe tener presente el plan del proyecto, la arquitectura para mirar los alcances técnicos y la documentación de los requerimientos que seguramente ayudan a la mejor opción. El momento de la escogencia de las herramientas tiene su momento indicado durante el desarrollo del proceso de Data Warehouse. Si se hace muy temprano no se tendrá un real entendimiento de los requerimientos del negocio y seguramente no se hará la mejor opción. Si espera y se hace un tiempo después, tendremos más información acerca los requerimientos, las plataformas y se puede escoger de esta forma una mejor opción.

Áreas a evaluar para la escogencia de las herramientas

Se tienen cuatro áreas fundamentales para la evaluación y escogencia de las herramientas:

- **Plataforma de hardware:** Se deben evaluar las diferentes plataformas para mirar su capacidad y escoger las herramientas teniendo esto en cuenta las limitaciones que se podrían presentar. Las plataformas de servidores donde se ejecutarán las herramientas de análisis dimensional, la plataforma donde tenemos los Data Mart y los demás servidores de aplicaciones deben ser analizadas cuidadosamente.
- **Plataformas de Base de Datos:** Para proyectos pequeños se puede utilizar un motor de BD sencillo como mysql que no es muy pesado y que podrán soportar las plataformas. Para proyectos grandes se debe considerar un motor de BD con más características, que otorgue mayor seguridad y que permita realizar más y mejores aplicaciones, por ejemplo, PostgreSQL, Oracle, SQL Server, interbase, entre otras.
- **Herramientas de Data Staging:** Esta es el área de trabajo del Data Warehouse donde la información se extrae de las fuentes, se les hace limpieza a los datos, se hacen las transformaciones y se envían a los Data marts. Estas son las herramientas más costosas y complejas del proyecto.

- Herramienta de acceso a los datos: Esta elección es difícil porque en el mercado no existe un líder, y además se terminan usando varias de estas herramientas por la gran variedad de requerimientos que se desean cumplir.

Cuando se usa software propietario, es fácil la utilización por que se solicita soporte técnico para que ayude en la capacitación y manejo de la herramienta y aprovechar algunos beneficios adicionales que pueda ofrecer. Para la escogencia de la herramienta se debe tomar un tiempo prudente, debemos recordar que las herramientas tecnológicas cambian continuamente por eso hay que hacer lo posible para mantenerse siempre actualizado.

Una posibilidad con el software propietario es que se puede tener acceso a versiones trial por 90 días para tener la oportunidad de conocer el producto y decidir si definitivamente se ajusta o no a las necesidades del proyecto. Cuando las herramientas son SW libre, se deben manejar recursos como foros on-line y hacer contacto con otros usuarios del producto para de esta manera tener algo de soporte técnico.

5.7 Características de aplicaciones para usuarios finales

En los pasos explicados anteriormente, se ha analizado el diseño e implementación; ahora se profundizará más en el front room. El objetivo del front room es proporcionar la interfaz que mostrará al usuario reportes y análisis multidimensionales que tomará como base en la toma de decisiones. Una aplicación de usuario final, provee un diseño y estructura a los reportes tomando como base los datos del DWH.

Especificación de aplicaciones para usuario finales

Hay algunos pasos importantes en el proceso de especificación de las aplicaciones de usuario final:

- Determinar el conjunto inicial de plantillas de reportes: El primer paso es tener la lista de reportes que se va a desarrollar en la aplicación. Para esto se deben hacer 3 tareas fundamentales:

- Identificar los posibles reportes: mirando la documentación que se tiene del proyecto, mirando los requerimientos del negocio se encuentra un gran potencial para determinar los posibles reportes. Con esta información se hace una lista de reportes a planificar, especificando el nombre, las filas y columnas con los elementos de datos y las medidas. Por ejemplo, tener la lista de la información de los departamentos que van a estar involucrados en los reportes y análisis.
- Consolidar la lista de candidatos: Una vez se tiene la lista de reportes, de debe categorizar dicha lista. Las categorías se reconocen rápidamente debido a la forma como están estructuradas las compañías. Para analizar como categorizar los reportes, se pueden tener en cuenta las siguientes preguntas: ¿Cómo es el negocio?, ¿Cuál es la tendencia?, entre otras.
- Priorizar la lista de reportes: Una vez la lista de reportes se tiene establecida, se debe definir con el usuario final para establecer prioridad en los reportes que ellos van a utilizar.
- Determinar la navegación en los reportes: Se debe desarrollar un modelo que permita al usuario encontrar la información de una manera más rápida y eficiente. Se debe crear un modelo inicial para la navegación y poder analizar de una manera más fácil los reportes y análisis multidimensionales.
- Determinar el estándar de plantillas de reportes: En esta parte se establecen estándares en cuanto a los elementos de la base de datos. Esta estandarización es muy importante por ayuda para entender la naturaleza de los reportes. La estandarización se enfoca principalmente a los elementos de datos del DWH que son utilizados para alimentar los reportes.
- Determinar la especificación de estas plantillas: Cada organización tiene un diferente grado de detalle en la documentación del proyecto. También se debe determinar si el equipo que trabaja en el proyecto hace formalmente la documentación de las aplicaciones. Si la respuesta es

negativa se debe presionar para que la tenga al día y a disposición para cualquier usuario que la quiera mirar.

Formato de especificación: La especificación de la aplicación para el usuario final tiene 2 partes fundamentales:

- La definición, que provee información básica acerca de las plantillas de reportes.
- El diseño, provee una representación visual de lo que debería mostrar el reporte.

La definición de las plantillas de reportes incluye el nombre, la descripción o propósito, frecuencia con que debe ser actualizado, parámetros, entrada, entre otros.

Tipos de aplicaciones para usuarios finales: Existen varios tipos de aplicaciones:

1. Basadas en Web: Estas aplicaciones son accedidas a través de un browser de Internet. Los usuarios podrían conectarse y ver los reportes vía intranet o Internet entrando a la página del Data Warehouse.
2. Herramienta independiente: Con la herramienta se diseñan algunas plantillas de reportes que el usuario va a poder acceder a través de una interfaz. Esta implementación no es muy flexible, pues en el caso que se quiera añadir reportes sale costoso, pero de igual forma es muy rápida. Estos reportes son muchas veces almacenados en archivos compartidos para que todas las personas la puedan acceder.
3. Herramienta de interfaz ejecutiva: Proporciona una estructura de acceso a las plantillas de reportes a través de una serie de pantallas. Estas implementaciones permiten fácilmente la navegación en la plantilla escogida.
4. Interfaz por código: Estas herramientas proporcionan un API que permite diseñar una interfaz. Esta es una buena posibilidad, pues se utiliza una herramienta de desarrollo gráfico y la navegación se puede ajustar mejor a las necesidades del usuario.

5.8 Mantenimiento y crecimiento de un Data Warehouse

Administración del entorno de Data Warehouse

Cuando una empresa adquiere sus sistemas de información, el cambio que tendrán estos sistemas es muy poco, sin embargo, cuando se desarrolla un proyecto de Data Warehouse se debe pensar en el mantenimiento posterior a la implementación, pues estas aplicaciones tienen gran tendencia a crecer a medida que crece la información de la organización.

La inversión en el mantenimiento del Data Warehouse es bastante importante, sin embargo, estas aplicaciones retornan la inversión que se les hace.

Diagnóstico de resultados incorrectos

Cuando se tienen problemas en el Data Warehouse y no se tienen los resultados esperados, se debe revisar alguno de los pasos anteriores en el ciclo de vida. También se tiene una serie de preguntas que, en el caso de no ser positivas, hay más razones para revisar los pasos anteriores.

- ¿El proyecto tiene el apoyo de los altos ejecutivos de la organización?
- ¿El equipo de desarrollo del proyecto entendió correctamente los requerimientos del usuario?
- ¿Los usuarios finales entienden el modelo dimensional o tienen que recurrir a las complicadas tablas del Data Mart directamente?
- ¿Para la escogencia de las herramientas análisis de datos, los usuarios finales fueron considerados en el proceso?
- ¿Los datos utilizados para alimentar el Data Mart son correctos?
- ¿Son coherentes los reportes que se están obteniendo del DWH?, de no ser así, ¿el equipo de desarrollo entiende las inconsistencias y tiene informado a los usuarios finales?
- ¿Los datos en el DWH se renuevan de manera oportuna?

- ¿La base de datos está bien diseñada o las consultas tardan mucho tiempo para darnos el resultado?
- ¿Tiene el DWH un equipo constante para la construcción de plantillas de reportes que ayuden al usuario a hacer mejores análisis?
- ¿Los usuarios finales han recibido suficiente entrenamiento en las herramientas generadoras de análisis multidimensionales o de reportes?
- ¿Conocen los usuarios del negocio a quien recurrir del equipo del DWH en el caso que tengan problemas con este?
- ¿Alguien responde cuando ellos necesitan consultoría para resolver problemas?

Enfoque en usuarios del negocio

Durante las semanas posteriores a la terminación del proyecto, se debe tener en cuenta que la cercanía del usuario debe ser esencial, pues se le debe prestar la mayor atención y ayuda posible para que el proyecto comience a mostrar sus resultados rápidamente. No se debe esperar que los usuarios llamen a los miembros del equipo, lo correcto es estar con ellos porque si los problemas duran mucho tiempo, el usuario se podría sentir decepcionado con el producto.

Administración de operaciones

Cuando el Data Warehouse hace grandes cargas, se debe tener la implementación bien robusta para que el proceso sea completamente exitoso; se le debe hacer mantenimiento a la infraestructura técnica, al manejo y mantenimiento de la base de datos y al manejo de los datos y metadata.

- Manejo de la infraestructura técnica: esta es una de las pocas áreas donde el Data Warehouse se parece a los sistemas operacionales de la empresa. Se debe monitorear frecuentemente la infraestructura técnica para más adelante no tener problemas. Las áreas críticas donde puede haber problemas son los servidores de aplicaciones, los servidores de BD, entre otros.

- Desempeño de la base de datos: Se pueden sacar indicadores cada vez que se cargan datos o cuando se ejecuten rutinas que involucren esta plataforma. Con estos indicadores se puede medir el desempeño de esta plataforma.
- Mantenimiento de los datos y metadatos: Los datos en el Data Warehouse cambian mucho, es decir por la gran cantidad de procesos y por la transaccionabilidad de las operaciones diarias, continuamente se están anexando datos al mismo y los usuarios finales cada vez están necesitando más y más análisis de datos para mejorar la calidad en la toma de decisiones.

El impacto de estos cambios debe analizarse de forma detenida, y determinar si tienen algún efecto en contra de la funcionalidad de la plataforma del DWH.

Todos estos cambios en la plataforma se deben documentar para tener detallado el seguimiento que se le hace y además para que todos los miembros del equipo de desarrollo estén actualizados acerca del estado del proyecto.

6. Desarrollo del proceso de construcción del Data Mart

Se optó por realizar un Data Mart y no un Data Warehouse, dado que la magnitud de la empresa y de la base de datos que actualmente tienen no llega a justificar la realización de un Data Warehouse completo. Se sigue la metodología estudiada de Ralph Kimball, dado que establece claros procesos para todo el ciclo del desarrollo del proyecto y garantiza la calidad y eficiencia de la solución de inteligencia de negocios.

6.1 Planeación y administración del proyecto

6.1.1 Objetivo del proyecto

Realizar el análisis, diseño y construcción de un Data Mart que soporte el análisis multidimensional de ventas y entradas/salidas de insumos de la empresa Strega, utilizando herramientas de libre distribución.

6.1.2 Definición del proyecto

Para el proyecto se consiguió un alto interés en los dueños de Strega, dado que necesitan elevar las ventas del negocio a través de nuevas estrategias de marketing y reducir al máximo los costos evitando pérdidas innecesarias en el manejo de insumos. Para esto, se planteó la necesidad de sacarle todo el provecho a las bases de datos que todos los días se nutre de información de ventas, clientes, insumos, etc.

Esta necesidad se satisface al implementar la solución de inteligencia de negocios del proyecto, con un Data Mart encargado de transformar los datos diarios de ventas en información útil, utilizando herramientas de data warehouse para extraer, transformar y analizar estos datos para luego generar reportes útiles a los dueños de los locales y la empresa. Toda la solución se basa en software libre, tanto en las etapas de desarrollo como en las etapas de análisis y producción, permitiendo reducir al máximo los costos de implementación y poner a su disposición la inteligencia de negocios.

6.1.3 Alcance del proyecto

El proyecto tiene todos los componentes de un Data Mart, pero al querer extraer la base de datos de los locales, se tuvo un inconveniente con los dueños de la empresa ya que no querían brindarnos la base de datos completa por problemas legales y comerciales. Llegamos a un acuerdo con los dueños de Strega y con el franquiciado del local de Pueyrredón para que nos brindara las estadísticas de ventas de este local de los últimos 6 meses. Luego con esta información, y en base a un análisis de la misma, se replicaron los datos para los locales de Paraná y Obispo Salguero, para centrarnos solamente en los 3 locales que más venden y que están ubicados en Nueva Córdoba. Estas bases de datos se encuentran cargadas en un servidor MySQL que usaremos como base para generar el Data Mart. Por el otro lado, está la base de datos en MySQL que corresponde al stock del local, lo que permite llevar un mejor control de los insumos y de los desechos que se generan diariamente por insumos en mal estado, vencidos, etc. Para el proceso de extracción, transformación y carga se utilizó Spoon, integrado a la suite de Pentaho Data Integration - Kettle, que contiene todas las herramientas para conectarse a un servidor MySQL y permite, luego de una transformación de los datos para normalizarlos, cargarlos en un Data Mart para generar los reportes que satisfagan los requerimientos. Todas las herramientas antes mencionadas son soportadas por los Sistemas Operativos más utilizados, incluyendo Windows y Linux, y tienen la ventaja de ser open source, lo que no genera costo alguno en la realización de este proyecto.

6.1.4 Justificación del proyecto en el negocio

Además de ser un proyecto académico, el proyecto busca mejorar las ventas y la eficiencia en el manejo del stock de Strega. Como justificación para el negocio se prevén los siguientes beneficios:

- Aumentar la cantidad de ventas por día:
Al realizar un análisis de ventas en base a los productos que se venden, la empresa podrá realizar campañas de marketing más eficientes, al

enfocar determinadas promociones basándose en los reportes y las estadísticas que surjan del Data Mart.

- Mejorar el servicio y generar nuevos clientes en meses bajos en ventas:
Se pueden mejorar las ventas en los meses desde diciembre a marzo, gracias al acceso y análisis de la información. A través de campañas de marketing apuntadas a un determinado tipo de cliente, se pueden mejorar las ventas en estos meses equiparándose así a los meses de mejores ventas del año.
- Mejorar el manejo del stock de cada local:
Con el adecuado uso del software “Stockia” que se desarrolló, y utilizando estos datos en el Data Mart, se puede mejorar la cantidad de insumos no utilizados y descartados por desechos en cada local, al mejorar la administración de los mismos. Se realizará un análisis de los productos que más se desperdician para administrarlos de manera más eficiente y reducir al mínimo las pérdidas.
- Generación de reportes de ventas semanales y mensuales:
Con el módulo de reportes implementado en el software “Stockia”, se busca tener una herramienta al alcance de los dueños del negocio que les permita tener información real y actualizada de las ventas, para poder tomar decisiones en la marcha y con información y datos fundados.

7. Análisis de requerimientos

7.1 Análisis de requerimientos

Para realizar el análisis de requerimientos se tuvieron en cuenta las entrevistas con los dueños de la empresa y con uno de los franquiciados, quienes serán los beneficiados con la solución de Inteligencia de Negocios a implementar. Los requerimientos del negocio se describen a continuación:

- Ver cantidad de ventas por producto por local en el tiempo
- Ver ingresos por local en el tiempo
- Ver ingresos por rubros por local en el tiempo
- Ver productos menos vendidos por local
- Ver cantidad de desechos por local

Estos requerimientos fueron los acordados a implementar con la empresa, todos ellos soportados con datos que se encuentran actualmente en la base de datos que utilizan los locales, sumados a los datos que se van a tener con el software “Stockia”.

7.2. Documentación de requerimientos

La documentación de estos requerimientos implica una descripción de cada uno y las fuentes de donde se extraerán los datos para tener la información correspondiente a cada caso. El detalle de cada fuente de datos se encuentra en el capítulo 9.

7.2.1 Ver cantidad de ventas por producto por local en el tiempo

Descripción: Esta consulta permite explorar la cantidad de ventas de cada local, discriminándolas por el producto que se desee analizar, permitiendo filtrar por mes y año.

Fuentes de datos: Base de datos: Data Warehouse – Tablas: “factventas”, “dimarticulo”, “dimlocales” y “dimtiempo”.

7.2.2 Ver ingresos por local en el tiempo

Descripción: Esta consulta permite explorar los ingresos que se obtuvieron de las ventas de cada local, permitiendo filtrarlas por mes y año.

Fuentes de datos: Base de datos: Data Warehouse – Tablas: “factventas”, “dimtiempo” y “dimlocales”.

7.2.3 Ver ingresos por rubros por local en el tiempo

Descripción: Esta consulta permite explorar los ingresos que se obtuvieron de las ventas de cada local, discriminándolas por rubros de productos y pudiendo filtrarlas por mes y año.

Fuentes de datos: Base de datos: Data Warehouse – Tablas: “factventas”, “dimlocales” y “dimtiempo”.

7.2.4 Ver productos menos vendidos por local

Descripción: Esta consulta permite identificar los 3 productos que obtuvieron menos ventas de cada local pudiendo filtrarlos por mes y año.

Fuentes de datos: Base de datos: Data Warehouse – Tablas: “factventas”, “dimlocales” y “dimtiempo”.

7.2.5 Ver cantidad de desechos por local

Descripción: Esta consulta permite ver la cantidad de desechos que se producen en cada local, pudiendo filtrarlo por mes y año.

Fuentes de datos: Base de datos: Data Warehouse – Tablas: “facttransacc”, “dimtiempo” y “diminsumos”.

8. Modelamiento dimensional

Para iniciar el modelamiento dimensional, se debe tener en cuenta el principal objetivo de cualquier almacén de datos o Data Mart: el análisis de la información. Este análisis se realiza por medio de reportes, por lo tanto, al modelar el Data Mart se debe tener como objetivo la información deseada en los reportes.

Dada la definición de los requerimientos, se identifican reportes de ventas que contengan información sobre: productos, rubros, locales y periodos de tiempo en que se realizan. A su vez, el reporte de insumos debe contener la información de: desechos, locales y productos. Estas categorías son dimensiones en el modelo dimensional, y se crearon dos tablas de hechos para diferenciar los datos: por un lado, una donde aparece la unidad de medida que es el valor de las ventas llamada “factventas”, y por otro lado una donde aparece como unidad de medida la cantidad de insumos en stock. La relación de la tabla de hechos a las tablas de dimensiones es de uno a muchos. Esto queda en evidencia lógicamente por el modelo del negocio, que tiene ventas en muchos locales a lo largo de Córdoba, a varios tipos de clientes, de varios productos en diferentes periodos de tiempo.

8.1 El Data Mart

El modelo diseñado e implementado es un Data Mart. Este concepto se diferencia al de Data Warehouse ya que el primero busca centrarse en un único objetivo o en el análisis de un área específica de la empresa, en este caso son las ventas de los locales y los insumos.

Sin embargo, no es posible clasificar a un Data Mart como un “pequeño almacén de datos”, dado que el concepto de Data Mart o almacén es solo de diseño. Un Data Mart de un área de una gran empresa puede ser más complejo y contener un mayor número de datos que todo un almacén de datos de una empresa de menor tamaño, o con diferentes bases de datos.

El Data Mart implementado en este proyecto tiene dos fuentes de datos que lo pueblan, ambas realizadas en MySQL: una base de datos de las ventas de los 3

locales de Nueva Córdoba, y otra de los insumos de cada local con la información de desechos y pedidos de insumos. Los detalles de estas fuentes de datos se encuentran en el capítulo 9.

8.2 Definición de la granularidad

Se definió la granularidad de la tabla de hechos como la más baja o granular posible. Esta granularidad corresponde a una venta de un producto a un cliente en una fecha determinada, es decir una transacción individual, y del ingreso o egreso de un insumo en el stock del local. De esta forma será posible llegar al grado de detalle de consultar una venta específica o un insumo, aunque este no sea el objetivo del Data Mart.

La medida de la tabla de hechos de ventas es el valor de una venta, y la medida de la tabla de hechos de insumos es la de un ingreso o egreso de insumo con la granularidad establecida.

8.3 Dimensiones

Se definen las dimensiones que soportan los requerimientos definidos, cumpliendo con la granularidad de la tabla de hechos. Las siguientes secciones relacionan las tablas diseñadas para la base de datos con su dimensión correspondiente.

8.3.1 Dimensión artículos

Tabla: dimarticulo

Contiene toda la información de los artículos que se venden en el local.

Campo-nombre	Tipo/Tamaño	Descripción	Llave primaria
CodDim	INT(8)	Llave primaria de dimarticulo	Si
nomDim	VARCHAR(20)	Nombre del producto	No

preDim	DOUBLE	Precio del producto	No
desDim	VARCHAR(20)	Corresponde al rubro específico del producto	No

8.3.2 Dimensión clientes

Tabla: dimclientes

Contiene toda la información de los clientes a los que alguna vez se le vendió, aunque sea un producto.

Campo-nombre	Tipo/Tamaño	Descripción	Llave primaria
codDim	INT(8)	Llave primaria de dimcliente	Si
descDim	VARCHAR(20)	Nombre del cliente	No

8.3.3 Dimensión locales

Tabla: dimlocales

Contiene toda la información de los 3 locales de Strega de Nueva Córdoba.

Campo-nombre	Tipo/Tamaño	Descripción	Llave primaria
CodDim	INT(8)	Llave primaria de dimlocales	Si
nomDim	VARCHAR(20)	Nombre del local	No

8.3.4 Dimensión tiempo

Tabla: dimtiempo

Contiene información de las fechas en que se realizan las ventas. Esta tabla contiene información sobre el año, mes, día. Se tienen estos diferentes campos para una misma fecha para permitir al usuario navegar por la dimensión dependiendo del grado de profundidad de fecha que prefiera.

Campo-nombre	Tipo/Tamaño	Descripción	Llave primaria
idtiempo	Date	Llave primaria de dimtiempo	Si
Ano	INT(4)	Descripción del año en que se realizó la venta	No
Mes	INT(2)	Mes en números en el que se realizó la venta	No
Día	INT(2)	Día en números en el que se realizó la venta	No
Dia_sem	INT(2)	Día de la semana en la que se realizó la venta	No
Dia_sem_desc	Varchar	Nombre del día de la semana en la que se realizó la venta	No
Dia_sem_desc_corto	Varchar	Nombre corto del día de la semana	No

		en la que se realizó la venta	
Mes_desc	Varchar	Nombre del mes en el que se realizó la venta	No
Mes_desc_corto	Varchar	Nombre corto del mes en el que se realizó la venta	No

8.3.5 Dimensión insumos

Tabla: diminsumos

Contiene toda la información de entrada y salida de insumos que se realizan en el local.

Campo-nombre	Tipo/Tamaño	Descripción	Llave primaria
CodDim	INT(8)	Llave primaria de factventas	Si
desDim	Varchar	Corresponde a la descripción del insumo	No
uniDim	INT(8)	Cantidad que ingresó o salió del insumo	No

8.4 Tablas de hechos

Como se mencionó, se realizarán dos tablas de hechos, una especificando las ventas y otra con las entradas/salidas de insumos.

8.4.1 Tabla de hechos de ventas

En esta tabla, el hecho es la venta que se realiza en el local, también es llamado una medida. Esta medida cumple con la granularidad definida, es decir es el valor de una transacción individual de una venta.

Los demás campos son llaves foráneas, para relacionar esta tabla con sus dimensiones.

Tabla: factventas

Registra información de las ventas que hace cada local.

Campo-nombre	Tipo/Tamaño	Descripción	Llave primaria
CodFact	INT(8)	Llave primaria de factventas	Si
CodSec	INT(8)	Corresponde a la secuencia de cada venta	No
CodArticulo	INT(8)	Articulo vendido	No
CantFact	INT(8)	Cantidad vendida del articulo	No
fechaFact	Date	Fecha de la venta	No
CodCliente	INT(8)	Código del cliente que realizó la compra	No
CodLocal	INT(8)	Código de local donde se realizó la venta	No

8.4.1 Tabla de hechos de insumos

En esta tabla, el hecho es la entrada o salida de insumo en el local, también es llamado una medida. Esta medida cumple con la granularidad definida, es decir es el valor de una transacción individual de una entrada o salida.

Los demás campos son llaves foráneas, para relacionar esta tabla con sus dimensiones.

Tabla: facttransacc

Registra información de las entradas y salidas de insumos que se hacen en el local.

Campo-nombre	Tipo/Tamaño	Descripción	Llave primaria
Id_fact	INT(8)	Llave primaria de factinsumos	Si
cantidad	INT(8)	Cantidad de ingreso o salida por transacción	No
Tipo	INT(8)	Indica si la transacción fue un ingreso o salida	No
Fecha	Date	Fecha de la transacción	No
Id_trans	INT(8)	Id de la tabla de ingresos o salidas	No
Id_prod	INT(8)	Id del insumo que ingresó o salió	No

8.5 Diseño del modelo dimensional

Luego de haber determinado las dimensiones y las tablas de hechos, se diseña el diagrama multidimensional implementado en la base de datos MySQL. Este diagrama es un modelo de estrella dado que las dimensiones no se relacionan entre sí, sino que se relacionan únicamente con las tablas de hechos, por lo que solo existe una tabla de dimensión para cada dimensión, lo que ofrece menor granularidad y mayor detalle. Éste modelo se adapta a la información que se tiene en las distintas tablas que se utilizan como base, su estructura es simple y permite que la extracción de datos sea más ágil y rápida.

9. Diseño técnico de la arquitectura

9.1. Datos

9.1.1 Problema con acceso a los datos Strega

[Ver Anexo I.](#)

9.1.2 Diagrama base de datos Strega

La figura 1 muestra las tablas con las que trabajaremos y sus respectivas relaciones. Los datos abarcan registros de 3 sucursales, Pueyrredón, Paraná y Obispo Salguero.

La tabla ventas se utiliza como cabeza de una factura donde se encuentran los datos de clientes, fecha y local. Por su parte la tabla secuencia representa el cuerpo de la factura, contiene el detalle de los productos vendidos, las cantidades y el precio.

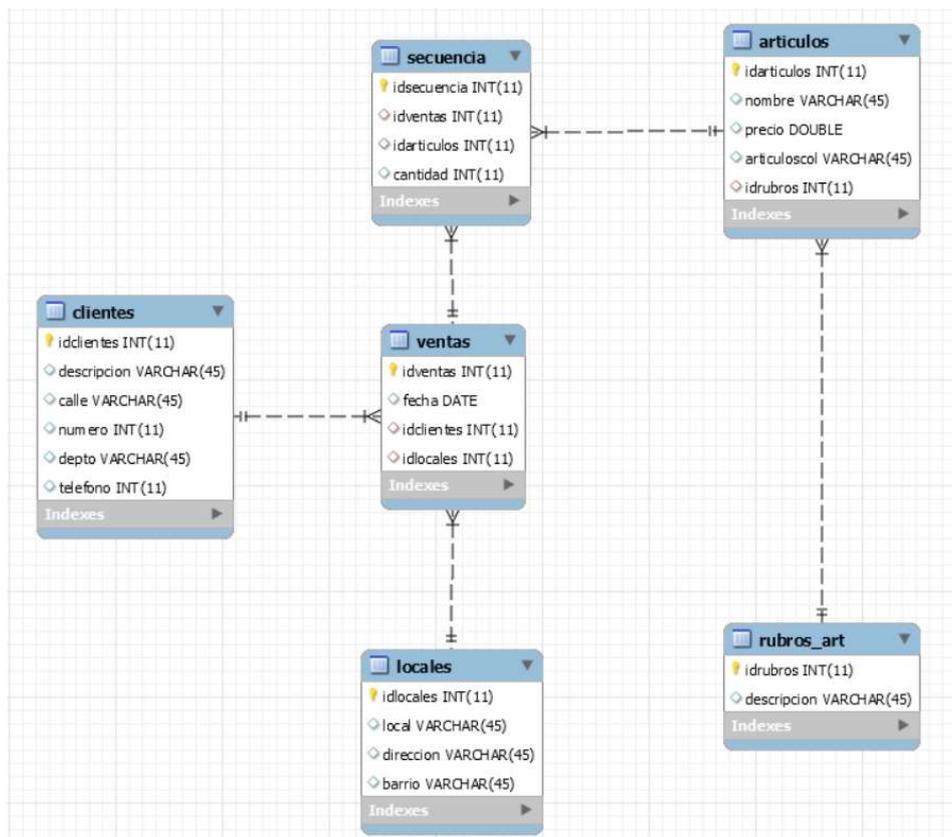


Figura 1 Diagrama Bases de Dato de Strega

9.1.2 Diagrama base de datos software “Stockia”

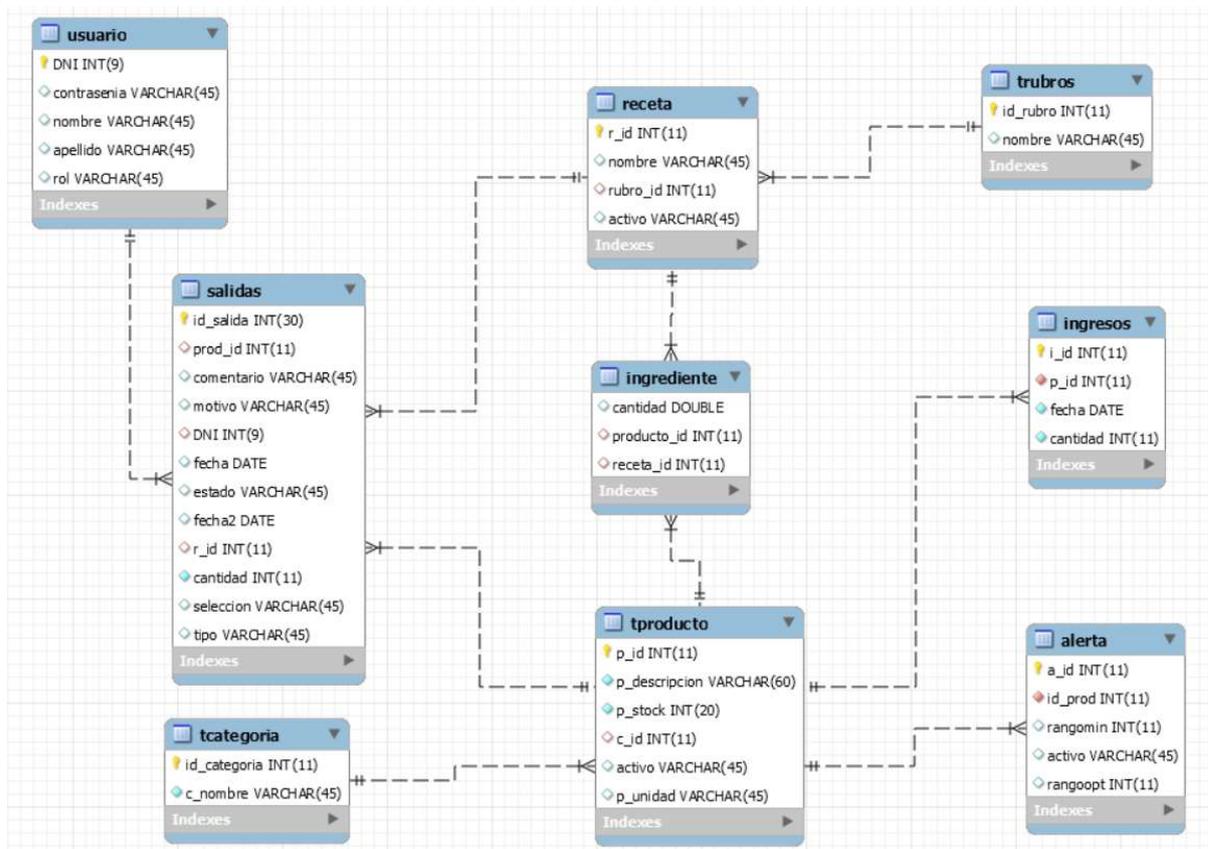


Figura 2 Diagrama Bases de Dato de Stockia

Del Software de Control y Stock se utilizarán las tablas:

- Salidas: contiene las salidas impactadas por el local
- Ingresos: contiene los ingresos de insumos registrados por el local
- Tproducto: contiene detalle de los insumos del local
- Tcategoria: categoriza los insumos con los que trabaja cada local

9.1.3 Mapeo de los datos

Para cargar los datos en el modelo dimensional se requiere la información de las tablas mencionadas anteriormente, además, para tener la información completa se requiere de las tablas que se crearan. En la siguiente tabla se pueden ver los datos del modelo dimensional y su respectiva fuente de datos.

Tablas de la estrella	Fuente de datos
-----------------------	-----------------

Dimclientes	Cientes
Dimarticulos	Artículos, Rubros_art
Dimlocales	Locales
DimInsumos	Tproducto, tcategoria
Factansacc	Ingresos, Salidas
Factventas	Caja, Caja_det

Tabla 1 Mapeo de datos

9.1.4 Tablas de hecho y sus relaciones:

En la siguiente imagen observamos la tabla de hechos factventas y sus correspondientes dimensiones. Factventas contiene los datos de ventas de los locales relacionadas con las dimensiones tiempo, cliente, artículo, locales.

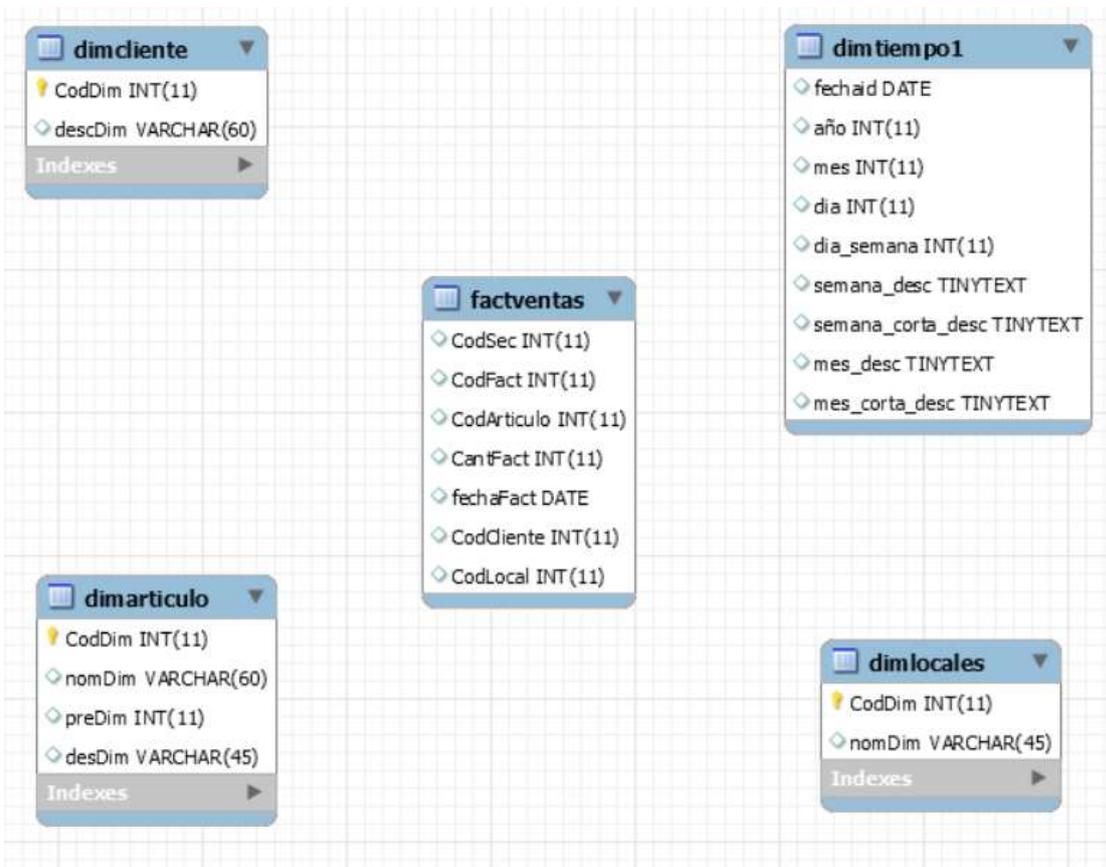


Figura 3 Data Mart Ventas Strega

Por otro lado, en la tabla de hechos facttransacc encontramos información de ingresos y salidas del local con sus correspondientes dimensiones artículo, insumos y tiempo.

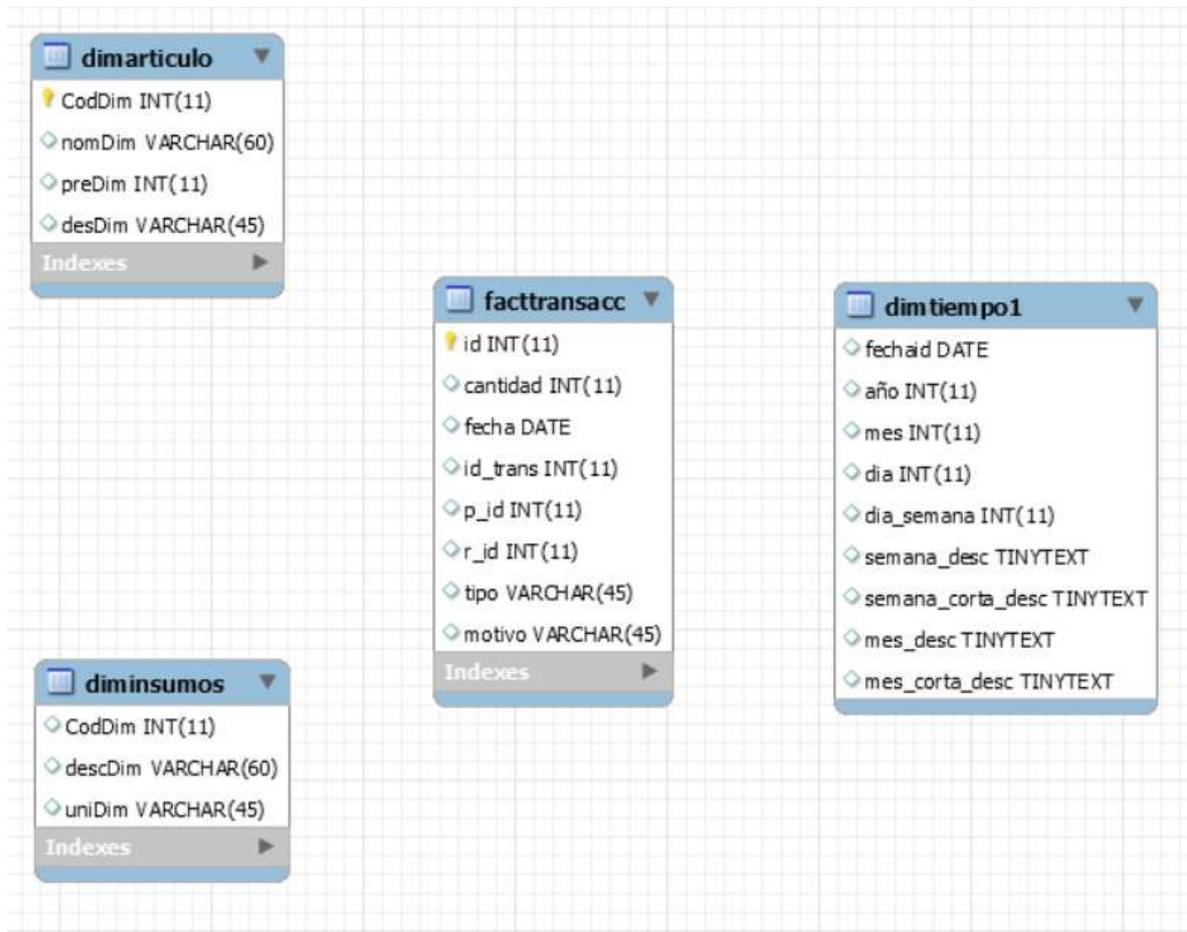


Figura 4 Data Mart transacciones por sucursal

9.2. Backroom: Extracción, Transformación y Carga

Una vez identificadas los recursos de datos con los que contamos, comenzamos planear los procesos ETL (Extracción, Transformación y Carga).

A continuación, una imagen que nos da la idea del trayecto de los datos desde su fuente origen (source) hasta su destino (destination).

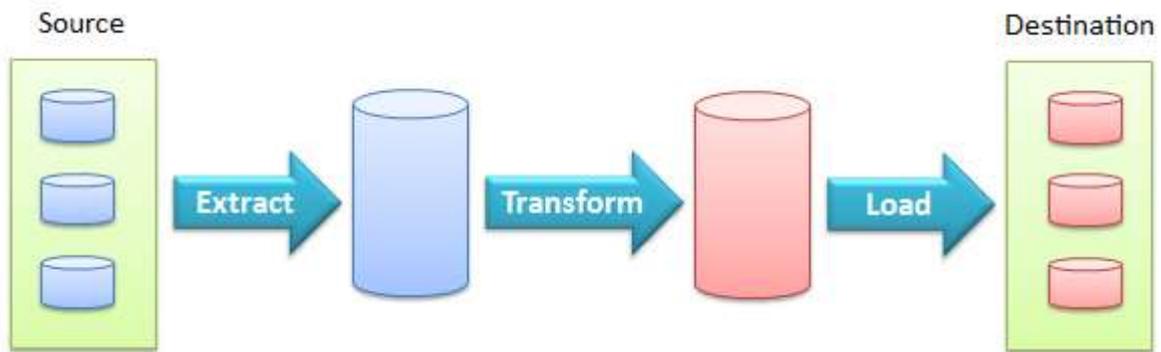


Figura 5 Proceso ETL

- Strega compañía: contiene los datos de todos los locales.
- Software “Stockia”: donde se encuentran todas las transacciones (ingresos y salidas de insumos) de cada sucursal. El software “Stockia” trabaja con bases locales para cada sucursal y.

Extracción: Proceso en el que se extraen los datos que nos permitirán llenar las tablas de dimensiones y hecho de nuestro Data Mart según se definen en el apartado anterior 7.1.1 Mapeo de datos.

Vale destacar qué como recomendación, decidimos volcar estos datos en un staging. En el staging se encuentran los datos crudos, normalizados solo en el caso de productos y sus correspondientes rubros. El beneficio de esto es que nos permite re-ejecutar proceso de ETL sin afectar a las fuentes ni generar sobrecarga a las bases con consultas.

En la transformación seleccionamos los campos, de cada tabla, que nos ayudaran a dar respuesta a los requerimientos definidos al comienzo. Luego el último proceso se encarga de cargar el Data Mart con los datos listos para ser consultados. En este último proceso se realiza primero las cargas de las dimensiones y al último las tablas de hecho.

Para realizar este proceso utilizamos las herramientas de Pentaho.

En el proyecto creamos dos ETL asíncronos, uno para cada fuente de datos. Como se mencionó con anterioridad los datos de todas las sucursales se encuentran centralizados por Strega, no así con los datos del software de control de Stock que trabajan con bases locales.

Para facilitarnos la lectura llamaremos:

- ETL Strega: al proceso que utilizara las fuentes Strega
- Y ETL Stockia: al proceso que utilizara las fuentes del software de control de stock.

9.2.1 Proceso de Extracción, Transformación y carga de las bases de Strega

ETL Strega

A continuación, se explica el proceso. Con el fin de automatización, se creó un JOB/trabajo, el cual agrupa las diferentes tareas (extracciones, transformaciones, etc.). En el JOB se encuentran los siguientes pasos:

- Start: salto que permite iniciar el proceso, en el mismo se puede configurar inicio de ejecución automático, en nuestro caso no se hace uso.
- Dos transformaciones:
 - Dimensión: se encuentran definidos los procesos ETL para cargar las dimensiones.
 - Hecho: se encuentran definidos los procesos ETL para cargar los hechos.
- Éxito: un Dummy que marca el fin.



Figura 6 Proceso ETL Strega

Descripción de la transformación dimensión:

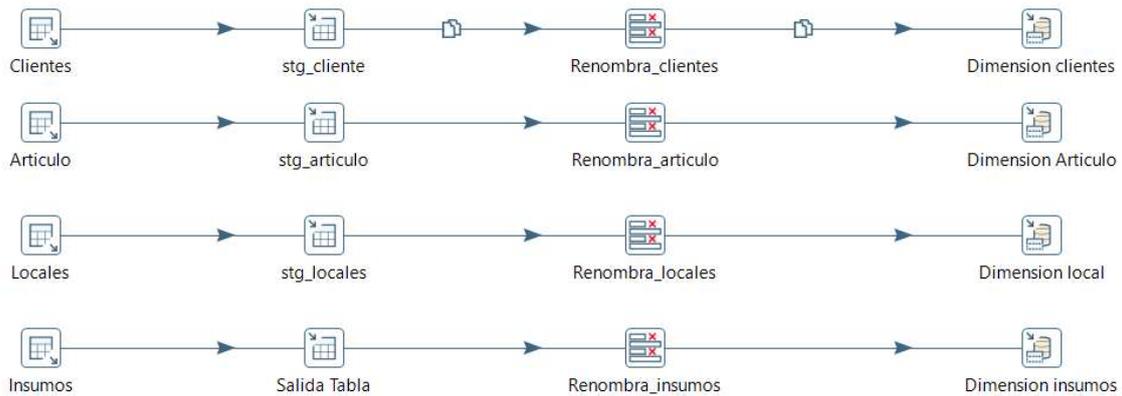


Figura 7 Proceso ETL Strega dimensiones

En la imagen se distinguen las fuentes (Clientes, Artículos, Locales), el staging el cual contiene los datos crudos, en el caso de artículos normalizados con sus rubros. La tercera columna se utiliza para seleccionar los campos de tabla que son necesarios y que terminaran actualizando las dimensiones.

Descripción de la transformación Hecho factventas (ver imagen):



Figura 8 Proceso ETL Strega tabla de hechos

Similar al proceso anterior, las diferencias están en que obtenemos los datos de dos tablas distintas Ventas y Secuencias, esta unión se realiza con un INNER JOIN y luego se cargan los datos en el staging. La otra diferencia es que los datos se cargan al final en una tabla de hechos- factventas-.

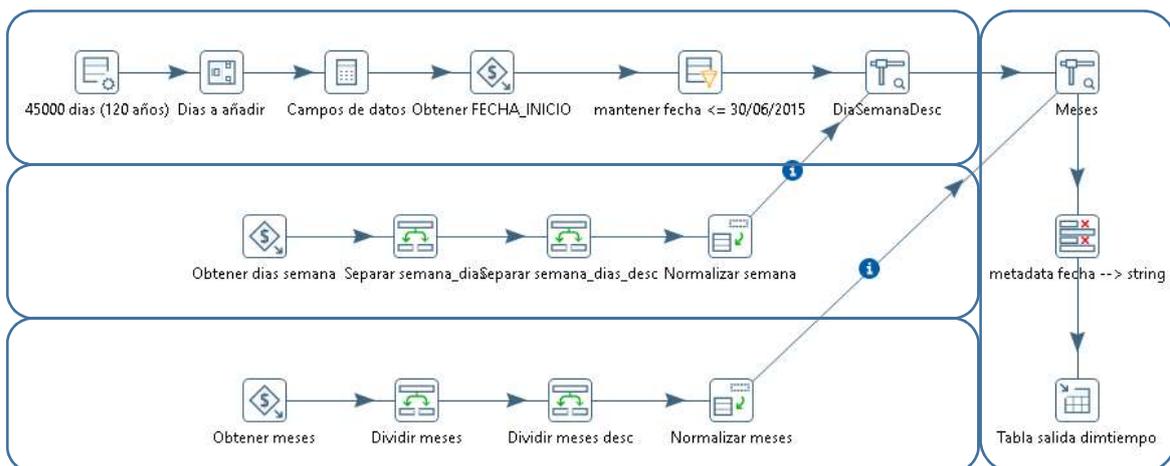
Dimensión tiempo

La dimensión tiempo es un caso especial y en cada proyecto se pueden definir diversas formas de generarla. En nuestro caso se optó por la ejecución de un

proceso que genere los datos de un periodo de tiempo determinado. El proceso elegido se ejecuta una sola vez, mediante Spoon (herramienta de Pentaho), para definir los meses con los que se trabajará. Para que la misma contenga un buen nivel de detalle, se realizó un ETL que llene la dimensión tiempo con los siguientes datos:

- Fechaid: es la clave de la tabla y permite vincularla con la tabla de hechos.
- Ano: Año separado de Fechaid
- Mes: Mes separado de Fechaid
- Día: Día separado de Fechaid
- Semana_dia: día de la semana obtenido de Fechaid
- Semana_desc: nombre del día de la semana obtenido de Fechaid
- Semana_desc_corto: nombre corto del día de la semana obtenido de Fechaid
- Mes_desc: nombre del mes obtenido de Fechaid
- Mes_desc_corto: nombre corto del mes obtenido de Fechaid

En este caso el proceso se configuró para generar los datos correspondientes a los primeros 6 meses del año 2015, ya que es el intervalo de tiempo en el que tenemos las ventas del local de Pueyrredón.



Descripción del ETL dimtiempo

Se divide el trabajo en 4 procesos para poder explicarlo mejor:

- Antes de realizar el primer proceso, se definen 5 parámetros que ayudarán a las tareas dentro del ETL:
 - DIAS_SEMANA: Define el nombre de los días de la semana.
 - DIAS_SEMANA_CORTO: Define el nombre de los días de la semana, pero abreviados.
 - FECHA_INICIO: se define la fecha de inicio que se tomará para generar los registros.
 - MESES: se definen los meses del año.
 - MESES_CORTO: se definen los meses del año, pero abreviados.
- El primer proceso horizontal hace referencia a los días de la semana:
 - Se comienza con una tarea que define la fecha de inicio, la cual será usada como punto de inicio para generar los datos de la tabla.
 - Luego se define el valor mínimo que se usará para generar cada registro de la tabla, en este caso será 1 día.
 - Se definen los campos de datos en un formato que luego se pueda separar para normalizarlos.
 - La siguiente tarea define la fecha hasta la cual se generarán los datos, en este caso son 6 meses (30/06/2015).
 - Luego mediante la tarea “DiaSemanaDesc” se llena la tabla con el nombre del día de la semana de cada registro de la tabla.
- En el segundo proceso horizontal, se realizarán las tareas correspondientes para definir el nombre de cada día de la semana de acuerdo a la fecha de cada registro.
 - Se obtienen los días de la semana mediante la tarea “Obtener días semana” y por medio de un proceso especial, se define a que día de la semana corresponde cada número del mes.
 - Se nombra cada día correspondientemente, se normalizan los datos y se utilizan como entrada para la tarea “DiaSemanaDesc”.
- En el tercer proceso horizontal se hace referencia a las tareas que definen a que mes del año corresponde cada registro de la tabla:
 - Se obtienen los meses de cada registro de la tabla, se los divide y luego se los nombra a cada uno.

- Luego se normalizan los datos y se utilizan como entrada en la tarea “Meses”.
- En el cuarto proceso vertical, se procede a utilizar como entrada las tareas de los procesos anteriores, para convertir los datos en string y cargarlos finalmente en la tabla dimtiempo.

Proceso de ETL de las tablas correspondientes al Software “Stockia”

ETL “Stockia”

Como el objetivo de automatización abarca a todo el proyecto, en este proceso también se creó un JOB:



Figura 9 Proceso ETL Stockia

Este proceso define una nueva tabla de hechos -facttransacc- que contiene los datos de las transacciones de ingreso y salida de productos o insumos registrada por cada sucursal. Esto significa que para cargar datos de transacciones que tienen como fuente dos bases distintas, será necesario primero organizar los datos para luego unirlos en una misma tabla. A continuación, se describe la imagen anterior:

- Start: salto que permite iniciar el proceso, en el mismo se puede configurar inicio de ejecución automático, en nuestro caso no se hace uso.
- Transformación CargaDimInsumo: contiene los procesos ETL para la carga de la Dimensión Insumos.
- PreparalIngresos y PreparaSalidas son JOB que extraen los datos de las tablas ingresos y salidas respectivamente, los transforman y los unen en un staging que integra la información de ambas tablas.
- CargaFact2: carga la tabla de hechos -facttransacc-.
- Éxito: un Dummy que marca el fin.

9.3 Frontroom

El Data Mart de ventas de Strega está estructurado de forma que se puede ver la información multidimensional de las ventas con respecto a los clientes, artículos, rubros, local de venta y en medidas de tiempo (por día, mes y año).

Por otro lado, cada local puede ver transacciones de ingreso y salidas de insumos y productos en el tiempo.

Respecto a los reportes también se puede decir que serán actualizados semanalmente, cuando el usuario ejecute la aplicación de carga para actualizar la información contenida en la bodega.

9.4 Automatización de actualización de repositorio

Kitchen es un programa que puede ejecutar trabajos diseñados con Spoon y almacenados como XML o en el repositorio de base de datos. Para poder automatizar la ejecución del JOB creado en Spoon se recurrió a la documentación de Pentaho.

Para ayudarnos con este proceso de automatización se creó un archivo .bat en el que se definen los pasos y parámetros necesario para que Kitchen ejecute el JOB de forma correcta.

A continuación, mostramos que parámetros fueron necesarios definir:

{ubicación del programa kitchen.bat}

{ubicación y nombre del trabajo.kjb}

{nivel de ejecución}

{> Destino y nombre del log}

La actualización del repositorio la podrá realizar el usuario desde el software "Stockia", sección reportes (apartado explicado en anexo manual control stock).

9.5 Infraestructura de Data Warehouse

Las bases de datos y los servidores OLAP se ejecutan sobre PC de configuración muy similar, por esta razón se puede ver que tanto los servidores de BD como los servidores OLAP tiene la siguiente configuración:

Sistema Operativo: Windows 7.

Memoria RAM: 4gb.

Disco Duro: 500 Gb.

Procesador: Intel Core i3.

Una plataforma con esta capacidad tiene la facultad de soportar las bases de datos MySQL, la herramienta de ETL y el software “Stockia”.

10. Reportes implementados

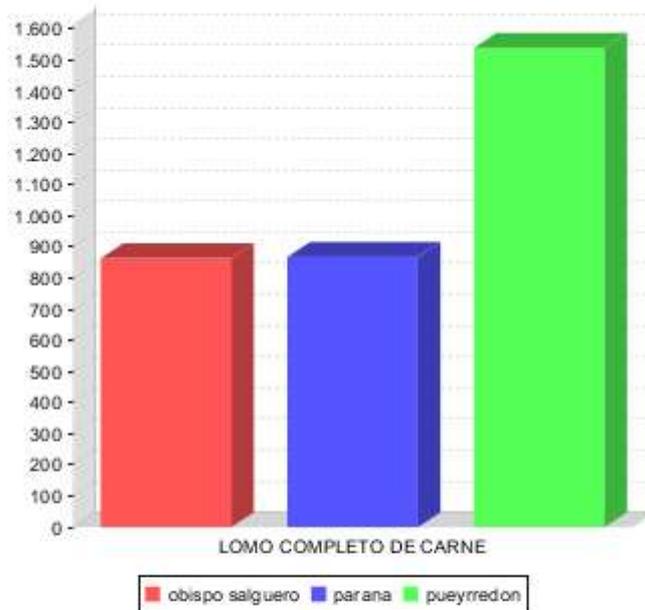
En esta unidad se mostrarán los reportes implementados que satisfacen los requerimientos, como así también la consulta SQL que genera el reporte y un breve análisis que se puede realizar sobre el reporte obtenido.

10.1 Ver cantidad de ventas por producto por local en el tiempo

El siguiente reporte le permite al usuario obtener la cantidad de ventas pudiendo filtrarlas por producto y por local, para así saber cuáles son los productos que más se venden y ofrecer promociones para aumentar las ventas.

```
1 SELECT
2     sum((factventas.`CantFact`)*(dimarticulo.preDim)) AS Ingresos,
3     dimlocales.`nomDim` AS Locales,
4     dimtiempol.`year` AS Ano,
5     dimtiempol.`month_desc` AS Mes,
6     dimarticulo.`nomDim` AS Producto
7 FROM
8     `factventas` factventas
9     INNER JOIN `dimtiempol` dimtiempol ON factventas.`fechaFact` = dimtiempol.`dateid`
10    INNER JOIN `dimlocales` dimlocales ON factventas.`CodLocal` = dimlocales.`CodDim`
11    INNER JOIN `dimarticulo` dimarticulo ON factventas.`CodArticulo` = dimarticulo.`CodDim`
12 WHERE
13     dimtiempol.month = ${P{mes_p}} and dimarticulo.nomDim = ${P{producto_p}}
14 AND dimtiempol.year = ${P{ano_p}}
15 GROUP BY
16     Locales,
17     Mes
18 ORDER BY
19     Locales ASC,
20     dimtiempol.month ASC
```

Cantidad de ventas por producto por local



Local	Mes	Producto	Ventas
obispo salguero	Marzo	LOMO COMPLETO DE CARNE	864
parana	Marzo	LOMO COMPLETO DE CARNE	868
pueyrredon	Marzo	LOMO COMPLETO DE CARNE	1540

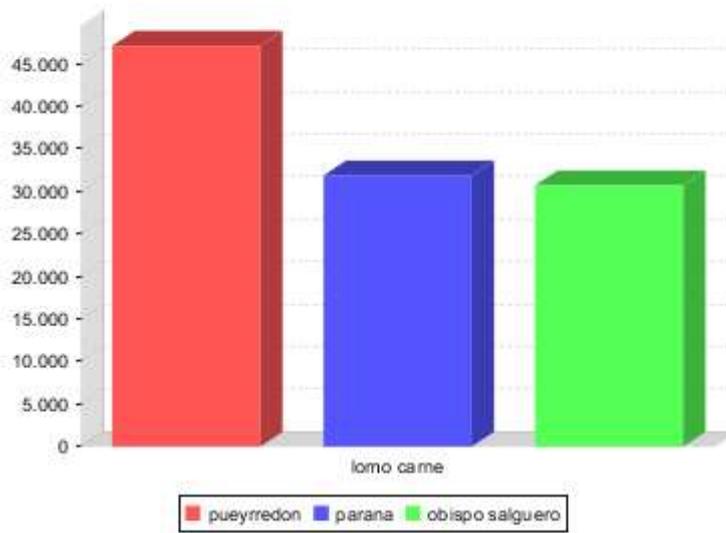
10.2 Ver ingresos por rubro por local en el tiempo

Este reporte permite obtener los ingresos cuantificados en \$ (pesos) de cada rubro que se comercializa en el local, para obtener un reporte que permita saber los rubros más vendidos y poder ejecutar campañas de marketing en base a estos datos.

```

1 SELECT
2     sum((factventas.`CantFact`)*(dimarticulo.preDim)) AS Ingresos,
3     dimlocales.`nomDim` AS Locales,
4     dimtiempol.`month_desc` AS Mes,
5     dimtiempol.`year` AS Ano,
6     dimarticulo.`desDim` AS Rubro
7 FROM
8     `factventas` factventas
9     INNER JOIN `dimtiempol` dimtiempol ON factventas.`fechaFact` = dimtiempol.`dateid`
10    INNER JOIN `dimlocales` dimlocales ON factventas.`CodLocal` = dimlocales.`CodDim`
11    INNER JOIN `dimarticulo` dimarticulo ON factventas.`CodArticulo` = dimarticulo.`CodDim`
12 WHERE
13     dimtiempol.month = ${mes_p} and dimtiempol.year = ${ano_p} and dimarticulo.desDim = ${rubro_p}
14 GROUP BY
15     Locales,
16     Mes, Rubro
17 ORDER BY
18     Ingresos DESC
    
```

Ingresos por rubro por local



Locales	Mes	Año	Rubro	Ingresos (\$)
pueyrredon	Marzo	2015	lomo carne	47088
parana	Marzo	2015	lomo carne	31856
obispo salguero	Marzo	2015	lomo carne	30710

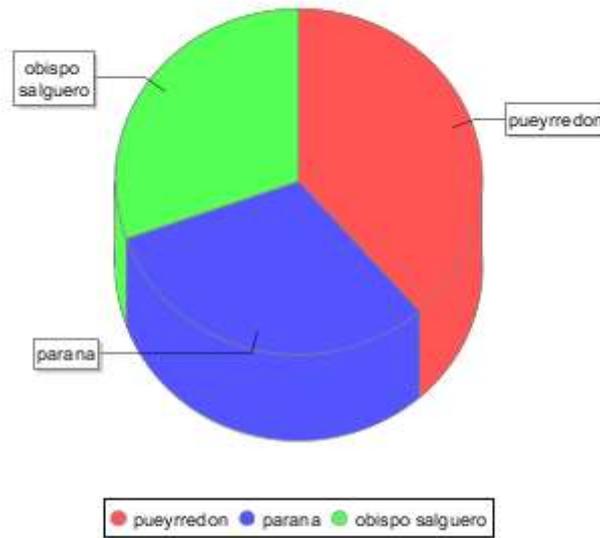
10.3 Ver ingresos por local

Este reporte permite ver los ingresos cuantificados en pesos (\$) de los 3 locales, pudiendo filtrarlo por mes y año. Esta consulta permite realizar una tendencia de ventas y ver si, en algún mes en particular hubo una baja, poder tomar decisiones correctivas en los meses siguientes para poder aumentar los ingresos.

```

1 SELECT
2     sum((factventas.`CantFact`)*(dimarticulo.preDim)) AS Ingresos,
3     dimlocales.`nomDim` AS Locales,
4     dimtiempol.`year` AS Año,
5     dimtiempol.`month_desc` AS Mes
6 FROM
7     `factventas` factventas
8     INNER JOIN `dimtiempol` dimtiempol ON factventas.`fechaFact` = dimtiempol.`dateid`
9     INNER JOIN `dimlocales` dimlocales ON factventas.`CodLocal` = dimlocales.`CodDim`
10    INNER JOIN `dimarticulo` dimarticulo ON factventas.`CodArticulo` = dimarticulo.`CodDim`
11 WHERE
12     dimtiempol.month = ${mes_p}
13 GROUP BY
14     Locales,
15     Mes
16 ORDER BY
17     Ingresos ASC,
18     dimtiempol.month ASC
    
```

Ingresos por local



Local	Mes	Ingresos (\$)
pueyrredon	Marzo	353406
parana	Marzo	284438
obispo salguero	Marzo	278104

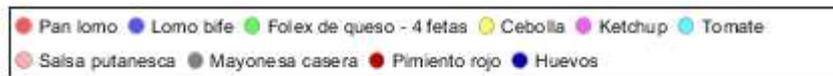
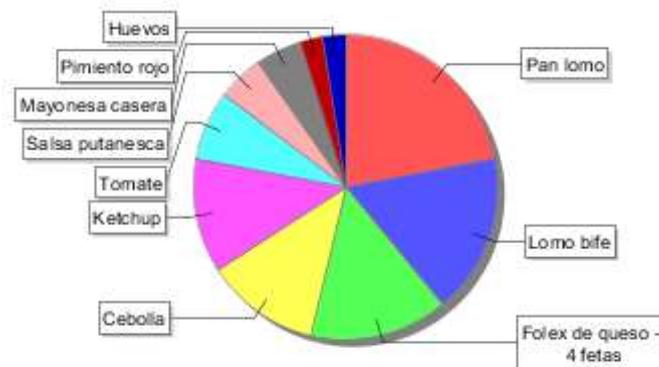
10.4 Ver cantidad de desechos por local en el tiempo

Este reporte permite verificar la cantidad de desechos que se tiran por cada local, pudiendo filtrarlo por mes, año y local. En base a los números que arroje, se pueden tomar acciones desde el área de Auditoria en un local que tenga una cantidad elevada de desechos, pudiendo ser que se estén realizando ventas que no se están contabilizando en el local, y los insumos se pasan por desechos.

```

1 SELECT
2     sum(facttransacc.`cantidad`) AS Cantidad,
3     diminsumos.`descDim` AS Insumo,
4     dimtiempol.`month` AS Mes,
5     dimtiempol.`year` AS Año,
6     facttransacc.`motivo` AS Motivo
7 FROM
8     `facttransacc` facttransacc
9     INNER JOIN `dimtiempol` dimtiempol ON facttransacc.`fecha` = dimtiempol.`dateid`
10    INNER JOIN `diminsumos` diminsumos ON facttransacc.`p_id` = diminsumos.`CodDim`
11 WHERE
12     facttransacc.motivo = "Desecho" and dimtiempol.month = $P{mes_p} and
13     dimtiempol.year = $P{ano_p}
14 GROUP BY
15     Insumo
16 ORDER BY
17     Cantidad DESC
    
```

Cantidad de desechos por local



Insumo	Mes	Año	Motivo	Cantidad
Pan lomo	5	2015	Desecho	18
Lomo bife	5	2015	Desecho	14
Folex de queso - 4	5	2015	Desecho	12
Cebolla	5	2015	Desecho	10
Ketchup	5	2015	Desecho	10
Tomate	5	2015	Desecho	6
Salsa putanesca	5	2015	Desecho	4
Mayonesa casera	5	2015	Desecho	4
Pimiento rojo	5	2015	Desecho	2
Huevos	5	2015	Desecho	2

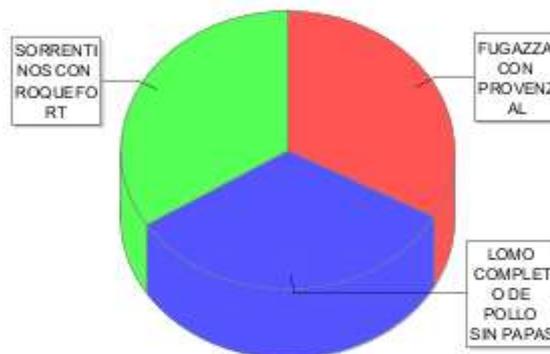
10.5 Ver productos menos vendidos por local en el tiempo

Este reporte permite ver los 3 productos menos vendidos por cada local, pudiendo realizar el filtro por mes y año, y así en base a estos datos, promocionar estos productos para poder aumentar sus ventas y generar nuevos ingresos.

```

1 SELECT
2     sum(factventas.`CantFact`) AS Ventas,
3     dimtiempol.`month` AS Mes,
4     dimtiempol.`year` AS Año,
5     dimarticulo.`nomDim` AS Articulo,
6     dimlocales.`nomDim` AS Locales
7 FROM
8     `factventas` factventas
9     INNER JOIN `dimtiempol` dimtiempol ON factventas.`fechaFact` = dimtiempol.`dateid`
10    INNER JOIN `dimlocales` dimlocales ON factventas.`CodLocal` = dimlocales.`CodDim`
11    INNER JOIN `dimarticulo` dimarticulo ON factventas.`CodArticulo` = dimarticulo.`CodDim`
12 WHERE dimtiempol.`month` = 3
13 GROUP BY
14     Articulo
15 ORDER BY
16     Ventas ASC
17 LIMIT 3
    
```

Productos menos vendidos por local



● FUGAZZA CON PROVENZAL ● LOMO COMPLETO DE POLLO SIN PAPAS
● SORRENTINOS CON ROQUEFORT

Locales	Articulo	Mes	Ventas
pueymedan	FUGAZZA CON PROVENZAL	Enero	140
pueymedan	LOMO COMPLETO DE POLLO	Enero	140
pueymedan	SORRENTINOS CON	Enero	144

11. Mantenimiento y crecimiento del Data Mart

Con respecto al mantenimiento y crecimiento del almacén de datos, se encuentran disponibles los scripts para que se hagan las actualizaciones semanales, para mantener los datos al día y disponibles para cualquier momento que se necesiten, teniendo en cuenta los horarios de atención. Además, se puede correr el script de actualización en cualquier momento desde el software Stockia, se advierte que su ejecución puede generar bajo rendimiento en las bases ya que los procesos de ETL afectan a las bases de forma directa.

En cuanto al crecimiento, la tabla con mayor incremento es la tabla de factventas del almacén de datos y en menor medida la tabla de facttransacc. No existe ningún riesgo que estos datos se pierdan o entren en conflicto en la aplicación, la base de datos MySQL permite garantizar que soporta la cantidad de información que puede ser generada por la empresa.

12. Conclusiones y Recomendaciones

Conclusiones

- Luego de concluir con éxito la etapa de pruebas de todo proyecto de software, debido al problema que tuvimos con las bases de datos, el Data Warehouse está listo para su implementación en cada local, pero la misma no se llevará a cabo, de acuerdo a lo que se pactó con los dueños de Strega. Los dueños de la empresa evaluarán la decisión y quedará en sus manos la implementación del mismo. El proyecto se mostró tanto a ellos como al franquiciado del local de Pueyrredón, y están convencidos de que se ha realizado un producto de muy buena calidad, que, en el caso de concretarse la implementación, ayudará de manera eficaz en la toma de decisiones, ya que se rigió por una metodología especializada para la materia que garantiza el éxito del proyecto.
- Para que un proyecto de Data Warehouse tenga éxito dentro de una empresa, es necesario que las áreas intervinientes tengan un alto grado de compromiso y confianza con el mismo, para que, en adelante las decisiones se tomen en base a datos confiables y reales, y no en base a intuiciones y suposiciones.
- El uso de software libre en Inteligencia de Negocios tiene la ventaja de ser desarrollado a la medida de las necesidades de una empresa. Esto permite que el análisis sea más personalizado y adaptado al negocio,

diferente de las soluciones comerciales, que son los negocios los que deben adaptarse a ellas.

- Con el uso de software libre en este proyecto, se puede competir de igual a igual con las soluciones comerciales en términos de rendimiento y funcionalidad, y a la vez, los costos que se ahorran en licencias se pueden destinar para otros proyectos en los que sea necesario adquirir este o cualquier otro tipo de licencia.
- La estructura de una base de datos de un proyecto de Inteligencia de Negocios es esencial para el éxito del mismo. Aunque la estructura de Entidad-Relación funciona muy bien para almacenar datos operativos, no es la recomendada para este tipo de proyectos. Lo recomendable es utilizar un esquema en estrella o copo de nieve, siempre dependiendo las dimensiones y las necesidades del Data Warehouse que se va a desarrollar
- Una solución de Inteligencia de Negocios no debe ser pensada para reducir costos, sino todo lo contrario, para aumentar las ventas y los ingresos de la empresa.

Recomendaciones

- Como primera recomendación, se deben migrar las bases de datos del programa que actualmente usan a una tecnología más nueva como MySQL que además de ser libre, dado el tamaño de la empresa, es el motor de base de datos que más se ajusta a sus necesidades y permite una mayor compatibilidad con el Data Warehouse diseñado.
- Se deberían identificar justificadores necesarios del negocio al iniciar un proyecto de Inteligencia de Negocios en una empresa, esto es desde el punto de vista financiero, identificando un retorno de la inversión realizada. Esto permitirá un mayor entusiasmo por parte de los dueños de la empresa y de los actores involucrados.

Anexos:

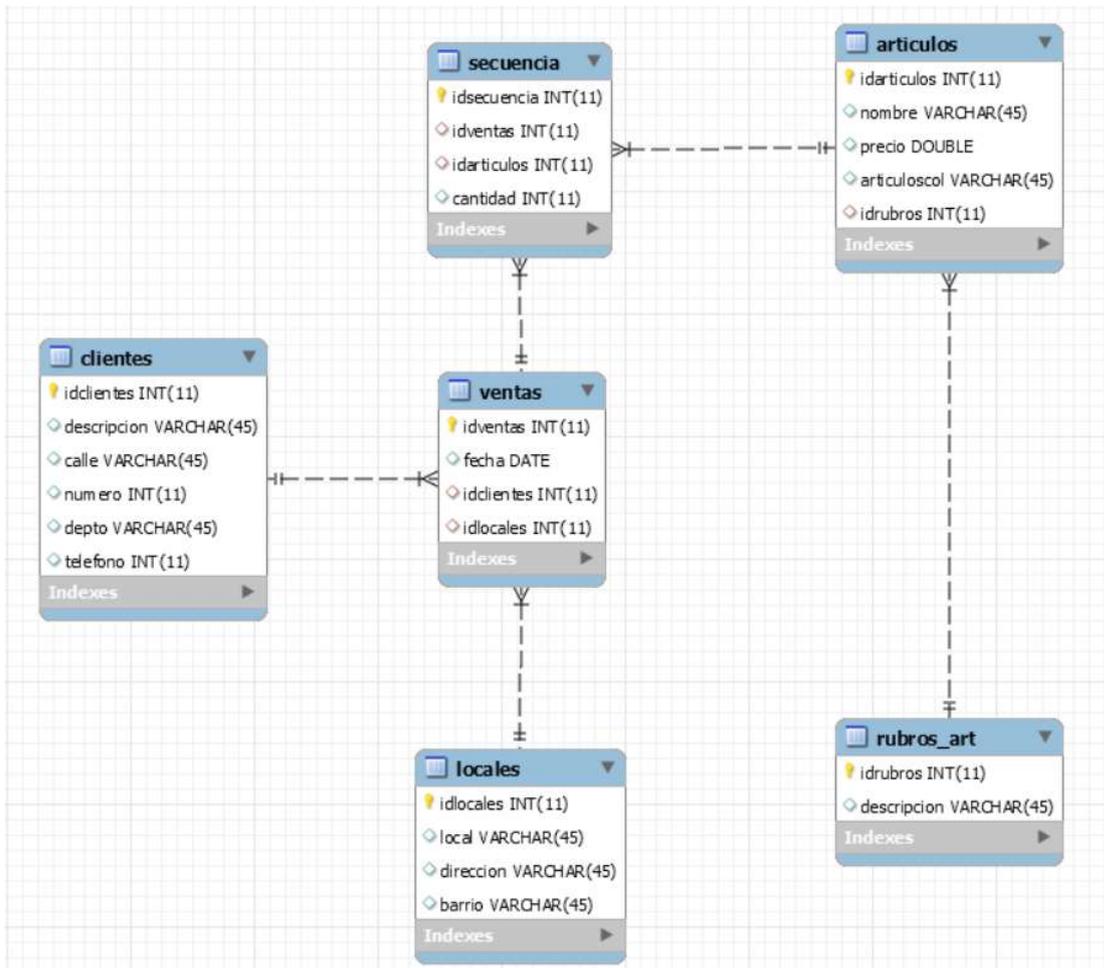
ANEXO I – Problema de acceso a los datos de Strega

Los datos son lo esencial y la parte para el desarrollo del proyecto, por eso nos aseguramos de ser claros al momento de presentar la idea. Al momento de presentar la idea con los dueños de la empresa, no se incluyeron a otras áreas como Contaduría o Legales. En el momento de necesitar las bases de datos de la empresa, los dueños hablaron con su contadora y por un tema legal de declaraciones nos rechazó el pedido, imposibilitándonos el acceso a las bases de datos en su totalidad.

Frente a esta situación y con el análisis previo que realizamos, decidimos buscar una alternativa. Sabíamos que los datos nos pueden dar un escenario del estado real de cada sucursal y de la empresa en general, por lo que decidimos comenzar un proceso de definición de datos a partir de un escenario final. En base a esto, el dueño de la sucursal de Pueyrredón nos dio los datos de ventas de su local de los últimos 6 meses, y con esa información, pudimos obtener números claros que nos sirvieron como punto de partida para generar las bases de datos de ventas de los 3 locales de Nueva Córdoba, que son los que se usaran para este proyecto. Utilizando estos números como punto de partida, se diseñó una estructura sencilla de base de datos para generar las bases que íbamos a necesitar. Adicionalmente, sabemos que las ventas de Paraná y Obispo Salguero son un 16% y un 12% menores respectivamente que el local de Pueyrredón, por lo tanto, se generaron los datos primero del local que más

vende, para luego seguir con los demás locales. Si bien la base de datos de Strega cuenta con más de aproximadamente 15 tablas, por este inconveniente legal, en el diagrama solamente se muestran las tablas de interés para nuestro Data Mart.

El diagrama que se realizó en base a los datos que necesitamos obtener es el siguiente:



Para nuestro proyecto se requiere la información relacionada con las ventas, datos contenidos en las tablas:

- Ventas: contiene las transacciones diarias de ventas.
- Secuencia: contiene los detalles de cada venta.
- Clientes: contiene la base de datos de clientes.

Para dar un contexto a los datos se crearon las siguientes tablas que ayudaran a ordenar la información y darles sentido a los reportes finales:

- Artículo: tabla que contiene la información de los artículos que se venden. Adicionalmente se creó un campo auxiliar llamado “porcentaje” que refleja el porcentaje de aparición de este artículo en el total de las ventas del mes, pero no lo pusimos en el diagrama de la base de datos ya que luego de generar los datos, este campo se eliminó para que quede lo más limpia posible la tabla.
- Locales: tabla que contiene la información de los locales que utilizaremos para el desarrollo del Data Mart.
- Rubros_art: tabla que agrupa los artículos en rubros.

Proceso de generación de los datos:

En base al análisis de las ventas que tuvo el local de Pueyrredón se realizaron las siguientes conclusiones:

Del 100% de las ventas del local:

- El 50% fueron ventas de un solo producto.
- El 30% fueron ventas de dos productos diferentes.
- El 10% fueron ventas de tres productos diferentes.
- El 6% fueron ventas de cuatro productos diferentes.
- El 4% restante fueron ventas de 5 productos diferentes.
- No hubo ventas de más de 5 productos diferentes.
- En promedio se realizaron entre 50 y 80 ventas diarias de domingo a jueves, y de viernes a sábado las ventas ascendían entre 80 y 117, siendo éste último valor el máximo de ventas que se realizó en un mismo día.
- Los 10 productos más vendidos en el local empezando primero por el más vendido y luego bajando en ventas son: Muzzarella, Lomo completo de carne, Burger, Especial, Lomo Strega de carne, Tarta de jamón y queso, Tarta de verduras, Rucula, Sorrentinos con fileto, Napolitana especial.
- Los productos restantes que se comercializan tienen una tendencia de ventas similar entre ellos.

Con este análisis, se creó una Procedure en MySQL para generar los datos aleatoriamente siguiendo los porcentajes especificados, controlando las cantidades de ventas por día y los productos más vendidos. Dentro de la misma, se utilizó el campo “porcentaje” para determinar, en base a la cantidad de ventas mensuales, cuantas veces se debía vender ese producto en el mes. Las tablas que se llenaron con este script fueron Caja y Caja_det, que son las necesarias para la creación del Data Mart

Luego de generarlas, se realizaron pruebas las siguientes pruebas:

```
1 select count(v.idventas) Cantidad, l.`local` from server.ventas v
2 inner join locales l on v.idlocales=l.idlocales
3 group by v.idlocales
4 order by Cantidad DESC;
```



Cantidad	local
14078	pueyrredon
12527	obispo salguero
11824	parana

- De esta consulta se puede verificar que las ventas del local de Obispo Salguero son un 13% aproximadamente menores a las del local de Pueyrredón, y por otro lado, las del local de Paraná son aproximadamente un 16% menores a las de Pueyrredón.

```
1 select count(v.idventas) cantidad, a.nombre from server.ventas v
2 inner join secuencia s on v.idventas=s.idventas
3 inner join articulos a on s.idarticulos=a.idarticulos
4 group by a.nombre
5 order by cantidad DESC
6 limit 10
```

articulos (2x10)	
cantidad	nombre
3758	MUZARELLA
3412	LOMO COMPLETO DE CARNE
2356	BURGER
1525	ESPECIAL
1297	LOMO STREGA CARNE
841	TARTA DE JAMON Y QUESO
809	TARTA DE VERDURAS
807	RUCULA
497	SORRENTINOS CON FILETO
495	NAPOLITANA MED ESPECIAL

- De esta consulta se pueden verificar los 10 productos más vendidos corresponden a los establecidos anteriormente.

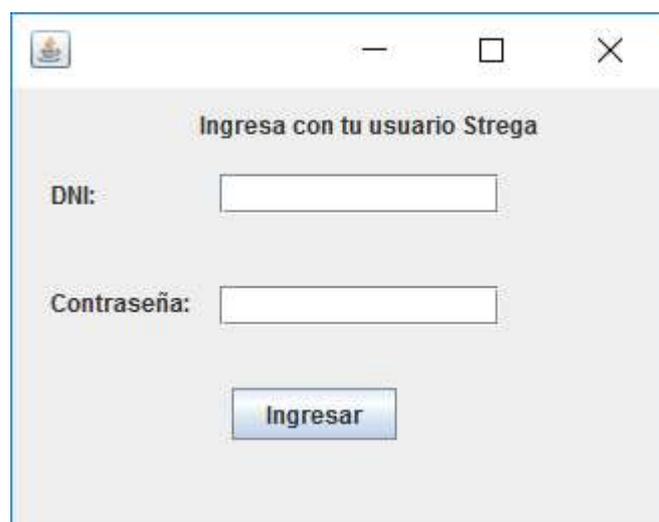
Luego de realizar estos controles, se generaron ciertas tendencias para generar aún más credibilidad en los datos, como aumento de ventas por feriados, partidos de futbol, principios de mes, etc. Vale aclarar que los porcentajes planteados tanto de ventas por día como de cantidad de artículos por venta son aproximados, para que así los datos sean aún más reales. Finalmente se dejan los datos disponibles para continuar con los procesos de ETL correspondientes.

ANEXO II – Manual de uso Software “Stockia”

MANUAL: CONTROL STOCK Y REPORTES DE VENTAS

LOGIN:

En el inicio del programa encontraremos una pantalla de login. La misma presentara el menú adecuado dependiendo del perfil ingresado.



The image shows a screenshot of a login window. The window title bar includes a small icon on the left and standard minimize, maximize, and close buttons on the right. The main content area has a light gray background. At the top, it says "Ingresa con tu usuario Strega". Below this, there are two input fields: "DNI:" followed by a text box, and "Contraseña:" followed by a text box. At the bottom, there is a blue button with the text "Ingresar".

Figura 1 Login

MENU PERFIL ADMINISTRADOR:

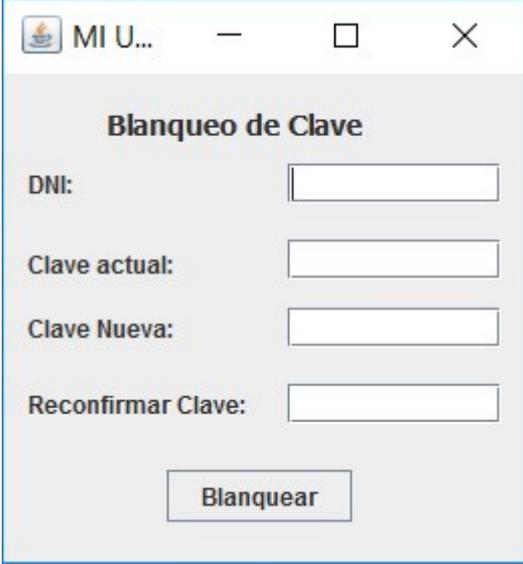
Figura 2 Menú perfil administrador

En la figura anterior encontramos los botones:

- a) Mi usuario
- b) Insumos
- c) Productos
- d) Salidas
- e) Adm. Usuarios
- f) Reportes
- g) Actualizar Datos
- h) Salir

A continuación, se explica la utilidad de cada uno de los botones anteriores:

- a) MI USUARIO: al ingresar a la opción Mi usuario se abre la pantalla (Figura 3), en la misma el usuario podrá realizar el blanqueo/cambio de su clave personal.



The image shows a web browser window with the title "MI U...". The main content is a form titled "Blanqueo de Clave". The form contains four text input fields, each with a label to its left: "DNI:", "Clave actual:", "Clave Nueva:", and "Reconfirmar Clave:". Below these fields is a single button labeled "Blanquear".

Figura 3 Mi Usuario

b) INSUMOS: En esta pantalla se presentan en una lista todos los insumos/ materia prima con los que trabajan los locales.

El panel introducción de datos nos permite interactuar con la lista, realizar altas, bajas o modificaciones de insumos. Seleccionando un insumo en la lista los datos del mismo se transfieren al panel de introducción de datos, en el mismo se pueden modificar los mismos y actualizarlos con el botón "ACTUALIZAR", se pueden eliminar con el botón "BORRAR". Para agregar datos se deben completar los campos del panel introducción de datos y presionar el botón "AGREGAR" (verificar que el ID no exista ya que en ese caso obtendremos un error y no podremos agregar el producto).

Pulsando el botón "CARGAR INGRESOS" del panel Ingresos de Stock nos aparecerá la pantalla (ver Figura 4a) podremos ingresar cantidades por insumos.

The screenshot shows a software window titled 'INSUMOS'. It contains a table of items, a data entry section, and a stock input section.

ID	Descripcion	Categoria	Stock	Unidad	Rango Opt	Rango Min
1	Lomo bife ...	CARNES	126	UNIDADES	-	-
2	Pan lomo	CARNES	91	Unidades	10	2
3	Folex jamo...	SEMI ELAB...	116	Unidades	-	-
4	Mayonesa c...	SEMI ELAB...	100	Kgrs.	-	-
5	Tomate	VERDURA	124	Kgrs.	-	-
6	Lechuga	VERDURA	100	Kgrs.	-	-
7	Huevos	HUEVOS	97	Unidades	100	50
8	Papas fritas	PAPAS FRI...	94	Kgrs.	-	-
9	Anana 490 ...	ENVASADO...	100	Kgrs.	-	-
10	Cebolla	VERDURA	100	Kgrs.	-	-

Introducción de Datos

ID Producto : Rubro: AGREGAR

Nombre : ACTUALIZAR

Cantidad : Unidades: BORRAR

Limpiar Campos ALERTAS

Ingresos de Stock

Cargar Ingresos

Figura 4 Insumos

ALERTAS: el botón “ALERTAS” nos permite configurar por cada insumo, cuales son las cantidades máximas y mínimas que creemos conveniente tener en el almacén, activar la alerta o dejarla sin uso. Los insumos se deberán buscar por su ID. Ver Figura 5.

The screenshot shows a window titled 'INGRES...' with a standard Windows title bar. Inside the window, there is a form with the following elements:

- A label 'Seleccione Insumo:' followed by a dropdown menu where 'Lomo bife' is selected.
- A label 'Ingrese Cantidad:' followed by an empty text input field.
- Two buttons at the bottom: 'Agregar' on the left and 'Volver' on the right.

Figura 4a Cargar Insumos

The screenshot shows a window titled 'ALERTAS' with a standard Windows title bar. Inside the window, there is a form with the following elements:

- An input field labeled 'ID del insumo:' followed by a 'BUSCAR' button.
- A section titled 'Información del insumo' containing:
 - A dropdown menu labeled 'Alerta:' with a '-' symbol selected.
 - An input field labeled 'Rango Óptimo:'.
 - An input field labeled 'Rango Mínimo:'.
- An 'ACTUALIZAR' button at the bottom of the form.

Figura 5 Alertas

- c) PRODUCTOS: En la pantalla se permite seleccionar algún producto ofrecido por el local, presionando el botón “BUSCAR” obtendremos la receta reflejada en la lista de ingredientes. Además de buscar podremos

eliminar un producto o crear nuevo producto. Presionando el botón “NUEVO PRODUCTO” (Figura 7), podremos ingresar un nombre del nuevo producto, su correspondiente rubro y luego, Figura 8, definir su receta, ingresando uno por uno los ingredientes y sus respectivas cantidades.

En los paneles Modificación de ingredientes y alta de ingredientes se podrán actualizar, agregar o eliminar los ingredientes de cada producto

The screenshot shows a window titled "RECETAS" with a standard Windows-style title bar (minimize, maximize, close). The interface is divided into several sections:

- Seleccione producto:** A dropdown menu currently showing "LOMO COMPLETO DE CARNE". Below it are three buttons: "Buscar", "Eliminar Producto", and "Nuevo Producto".
- Lista de Ingredientes:** A table with three columns: "INGREDIENTE", "CANTIDAD", and "CANTIDAD". The table is currently empty.
- Modificacion de Ingrediente:** A section with two input fields: "Ingrediente" (empty) and "Cantidad:" (empty). To the right are two buttons: "Actualizar" and "Eliminar".
- Alta de Ingrediente:** A section with a dropdown menu for "Ingrediente" (showing "Lomo bife"), a "Cantidad:" input field, and a unit label "BOLSAS". To the right is an "Agregar" button.

Figura 6 Productos

RECETAS

Ingrese el nombre del Producto

Seleccione Rubro:

Lomitos carne

Crear Nuevo Producto

Figura 7 Nueva receta

RECETAS

pruebas

Seleccione Ingrediente:

Lomo bife

Seleccione Cantidad:

Agregar Eliminar

Volver

Ingredientes

INGREDIENTES	CANTIDAD

Figura 8 Definición de ingredientes

- d) **SALIDAS:** se podrán impactar dos clases de salidas, la primera representa a las ventas realizadas ingresando la cantidad por producto, la segunda permite impactar salidas de insumos/materia prima indicando el motivo. Presionando el botón “CARGAR” los productos e insumos se listan en la lista salidas a impactar. Una vez finalizada la carga de los

productos e insumos será necesario impactar la misma presionando el botón "IMPACTAR", luego aparecerá la pantalla, ver figura 10, en la cual la persona que intenta impactar en la base de datos deberá validar su usuario con DNI y contraseña.

El éxito del proceso es representado por el mensaje de la figura 11 el cual indica que los productos se descontaron correctamente. En caso de ingresar usuario o contraseña incorrectos aparecerá la figura 12.

The screenshot shows a window titled "SALIDAS" with a standard Windows-style title bar. The main content area is titled "Salida De Productos". It contains two main sections. The first section has a "Producto:" label followed by a dropdown menu showing "LOMO COMPLETO D...", a "Cantidad:" label with an empty text input field, and a "Cargar" button. The second section has an "Insumo:" label with a dropdown menu showing "Lomo bife", a "Cantidad:" label with an empty text input field, a "Cargar" button, a "Motivo:" label with a dropdown menu showing "Desecho", and a "Comentario:" label with an empty text input field. Below these sections is a table titled "Salidas a impactar". The table has two columns: "PRODUCTOS" and "CANTIDAD". The table body is currently empty. At the bottom of the window are two buttons: "IMPACTAR" and "DESHACER TODO".

Figura 9 Salidas

The image shows a window titled "IMPACTA USUA..." with a standard Windows-style title bar (minimize, maximize, close). Inside the window, there are two text input fields. The first is labeled "DNI:" and the second is labeled "Contraseña:". Below these fields is a button labeled "Validar".

Figura 10 Valida Usuario

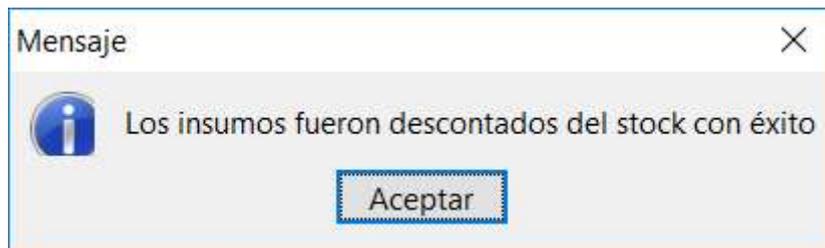


Figura 11 Mensaje exitoso

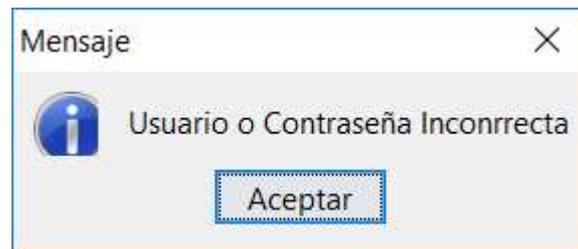


Figura 12 Mensaje Usuario Contraseña incorrecta

- e) ADMINISTRACION DE USUARIOS: la siguiente función le permite a un administrador agregar usuarios, eliminar usuarios o actualizar nombre y apellido. En el caso de agregar usuario deberá presionar el botón "NUEVO USUARIO" donde deberá completar los datos que se solicitan en la figura 14.

Administración de usuarios

DNI	nombre	apellido	rol
33333333	rodrigo	martinez	encargado
33700794	rodrigo	martinez	administrador
33975429	diego	flores	administrador

Modificación de Datos de Usuario

Nombre:

Apellido:

DNI:

Rol:

Figura 13 Administración de usuarios

Creación de Usuario

DNI:

Nombre:

Apellido:

Rol:

Clave:

Confirmar clave:

Figura 14 Creación de usuario

- f) **REPORTES:** Aquí el administrador podrá obtener dos tipos de reportes:
- Reportes online de ventas registradas por tiempo o por vendedor. Se trabaja con las bases online por lo que se podrán obtener los datos en tiempo real.
 - Reportes offline por lo que antes de trabajar se deberán actualizar los datos mediante el botón “ACTUALIZAR INFORMACION”.

**Figura 15** Reportes

VENTAS – CONSULTAS ONLINE: En la siguiente pantalla se podrá obtener información de las ventas registradas en el local. La consulta nos permite listar la información con el filtro de DNI o buscando por fecha.

Title 1	Title 2	Title 3	Title 4

Figura 16 Consultas Online

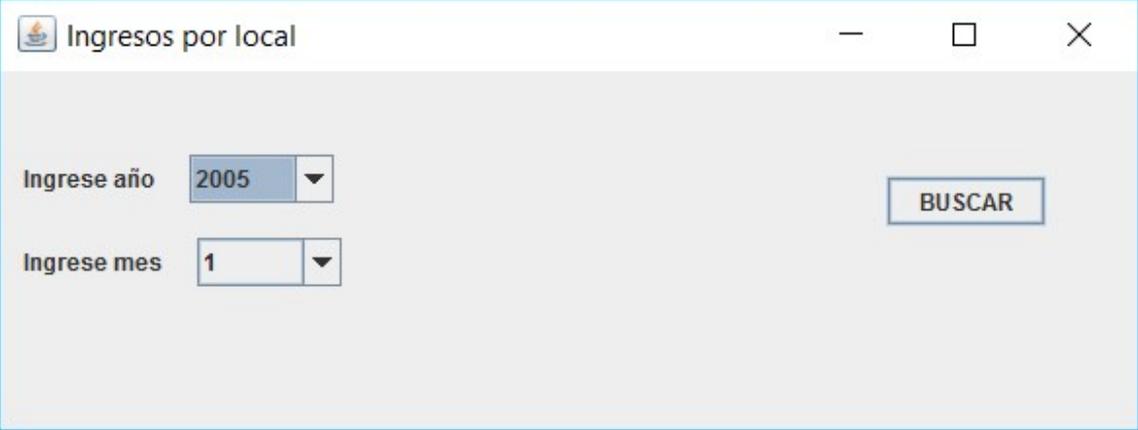
VENTAS – REPORTES DIRECTIVOS:

Para obtener reportes con información reciente, es necesario actualizar los datos mediante el botón “ATUALIZAR INFORMACION” (permite actualizar la tabla de hechos de la sucursal), inmediatamente aparecerá el mensaje figura 21 y una vez concluido el proceso otra ventana indicará el éxito del mismo, figura 22.

Los reportes directivos se encuentran definidos junto a los administradores con antelación, los mismos son:

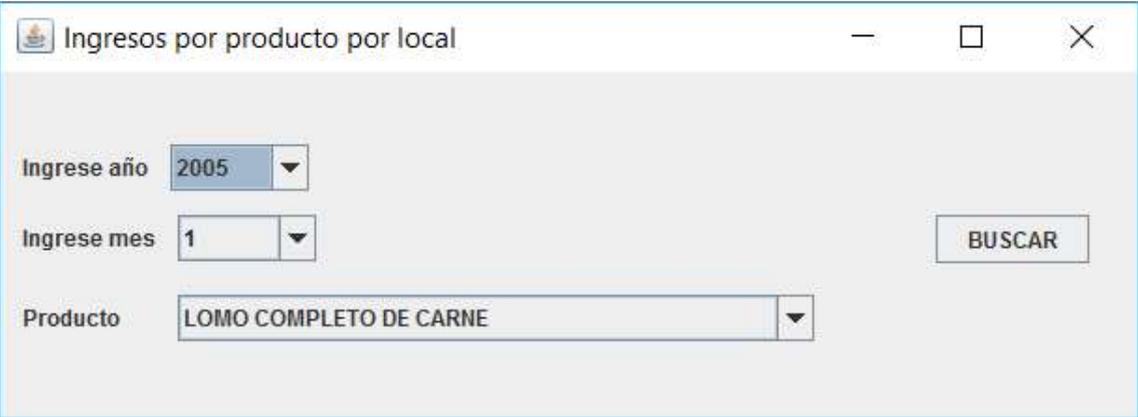
1. Ingresos por local, luego de esta opción podremos elegir el año que deseamos filtrar, figura 17
2. Ingresos por producto por local, en este caso podremos elegir el mes y el producto que deseamos filtrar figura 18.
3. Ingresos por rubro por local, ídem a ingresos por local. Figura 19

4. Cantidad de desechos. Figura 20
5. Productos menos vendidos Figura 21



The screenshot shows a window titled "Ingresos por local". It features two dropdown menus for filtering data. The first dropdown, labeled "Ingrese año", has "2005" selected. The second dropdown, labeled "Ingrese mes", has "1" selected. To the right of these dropdowns is a button labeled "BUSCAR".

Figura 17 Reporte ingresos por local



The screenshot shows a window titled "Ingresos por producto por local". It features three dropdown menus for filtering data. The first dropdown, labeled "Ingrese año", has "2005" selected. The second dropdown, labeled "Ingrese mes", has "1" selected. The third dropdown, labeled "Producto", has "LOMO COMPLETO DE CARNE" selected. To the right of these dropdowns is a button labeled "BUSCAR".

Figura 18 Reporte Ingresos por producto por local

The screenshot shows a web form titled "Ingresos por rubro por local". It contains three input fields: "Ingrese año" with the value "2015", "Ingrese mes" with the value "1", and "Ingrese Rubro" with the value "lomo carne". A "BUSCAR" button is located to the right of the fields.

Figura 19 Reporte Ingresos por rubro por local

The screenshot shows a web form titled "Cantidad de desechos". It contains two input fields: "Ingrese año" with the value "2015" and "Ingrese mes" with the value "2". A "BUSCAR" button is located to the right of the fields.

Figura 20 Cantidad de desechos

The screenshot shows a web form titled "Cantidad de desechos". It contains three input fields: "Ingrese año" with the value "2015", "Ingrese mes" with the value "5", and "Local" with the value "parana". A "BUSCAR" button is located to the right of the fields.

Figura 21 Productos menos vendidos

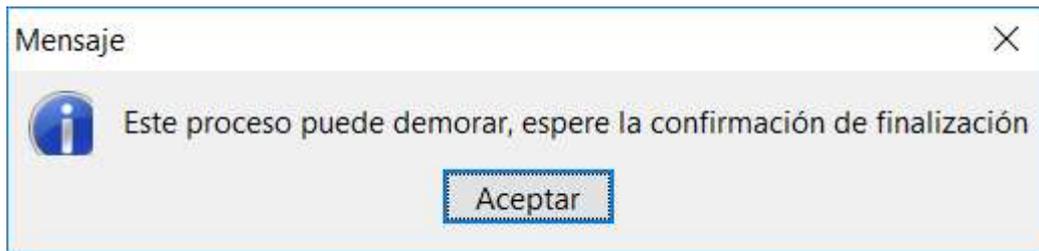


Figura 21 Mensaje actualización de información

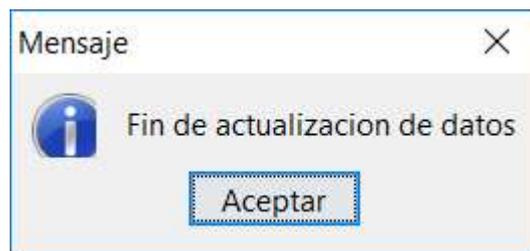


Figura 22 Mensaje de actualización exitosa

- g) ACTUALIZAR DATOS: Esta opción solo disponible para el administrador permite cargar las dimensiones del datawarehouse y generar la tabla de hechos de ventas.

BIBLIOGRAFIA

<http://es.slideshare.net/tegsistemas/modelo-de-calidad-del-software>

<http://es.slideshare.net/juanksi28/modelo-cascada-y-espinal>

<http://www.alegsa.com.ar/Dic/requerimientos.php>

Ingeniería del Software - Roger S. Pressman - Ed. McGrawhill

Análisis y Diseño de Sistemas - Henry F. Korth - Ed. McGrawhill

Análisis y Diseño de Sistemas de Información - Jeffrey Whitten - Ed. Irwin

https://es.wikipedia.org/wiki/Modelo%28%93vista%28%93controlador#Frameworks_MVC

<http://www.desarrolloweb.com/articulos/que-es-mvc.html>

<https://wikiuml.wikispaces.com/Diagrama+de+Colaboraci%C3%B3n?responseTocken=4d209e2a7454416b73378baf77d3194>

<http://www.elclubdelprogramador.com/2012/03/28/uml-diagramas-de-colaboracion/>

<http://www.e-mas.co.cl/categorias/informatica/analisisyd.htm>

<http://www.lawebdelprogramador.com/cursos/MySQL/6678-Manual-de-MySQL-Workbench.html>

https://es.wikipedia.org/wiki/Inteligencia_empresarial

<http://www.dataprix.com/que-es-un-datawarehouse>

<http://www.dataprix.com/tecnicas-explotacion-dwh>

<http://openbi.ning.com/>

http://www.sinnexus.com/business_intelligence/

<http://www.pentaho.com/>

<http://community.pentaho.com/>

<https://sourceforge.net/projects/mondrian/files>

<https://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/>

<https://churriwifi.wordpress.com/2010/05/10/16-3-construccion-procesos-etl-utilizando-kettle-pentaho-data-integration/>

<https://www.adictosaltrabajo.com/tutoriales/kettle/>

<http://www.open-source-guide.com/Solutions/Developpement-et-couches-intermediaires/Etl/Pentaho-data-integration>

<http://wiki.pentaho.com/display/EAI/Kitchen+User+Documentation>

<http://community.jaspersoft.com/project/ireport-designer>

<https://ajdgeniz.wordpress.com/2010/06/20/ejemplo-de-reportes-netbeans-java-mysql/>

<https://www.youtube.com/watch?v=VZ1Y7cxQFFY>

<https://www.youtube.com/watch?v=SvHWBFrLhPs&t=794s>

<https://www.youtube.com/watch?v=gV0nEevE8lk>

<https://hipertextual.com/archivo/2013/10/10-paginas-con-iconos-gratis/>

<http://findicons.com/>

<http://www.iconarchive.com/>