UNIVERSIDAD CATÓLICA ANDRÉS BELLO

FACULTAD DE INGENIERÍA

ESCUELA DE INGENIERÍA INFORMÁTICA

SISTEMA DE INFORMACIÓN DE APOYO PARA LA TOMA DE DECISIONES RELACIONADAS CON LA FACTURACIÓN, PEDIDOS Y NECESIDADES DE RECURSOS HUMANOS DE LA EMPRESA FLAMBERGE

Este Jurado; una vez realizado el examen del presente trabajo ha evaluado su contenido con el resultado: 20 (veinte) pts. mención honorífica.

J U R A D O E X A M I N A D O R

Firma:	Firma:	Firma:
Nombre: Fco. Bello	Nombre: ALFREDO GARCIA	Nombre: SUSANA GARCIA M.

REALIZADO POR	Dymara Berroterán Evaristo Gamboa
PROFESOR GUIA	Francisco Bello
FECHA	Octubre 2007

ÍNDICE

AGRADECIMIENTOS i
DEDICATORIA ii
ÍNDICE iii
ÍNDICE DE FIGURAS vi
ÍNDICE DE TABLAS vii
SINOPSIS viii
INTRODUCCIÓN 1
CAPITULO I. PROBLEMA Y OBJETIVO 3
I.1 PLANTEAMIENTO DEL PROBLEMA 3
I.2 OBJETIVO GENERAL 4
I.3 OBJETIVOS ESPECÍFICOS 4
I.4 ALCANCE 4
I.5 LIMITACIONES 5
CAPITULO II. MARCO TEÓRICO 7
II.1 ALMACEN DE DATOS 7
II.1.2 VENTAJAS E INCONVENIENTES DE UN ALMACEN DE DATOS 8
II.1.3 DATAMARTS (DM) 8
II.1.4 DIFERENCIAS ENTRE UN DATA MART Y UN ALMACEN DE DATOS 9
II.1.5 COMPONENTES DE UN ALMACÉN DE DATOS 9
II.1.6 DIFERENCIAS ENTRE UN SISTEMA DE PROCESAMIENTO TRANSACCIONAL EN LÍNEA (OLTP, OnLine Transaction Processing) Y UN SISTEMA ANALÍTICO EN LÍNEA (OLAP, OnLine Analytical Processing) 10
II.1.7 PROCESAMIENTO OLAP 10
II.1.7.1 TIPOS DE PROCESAMIENTO ANALÍTICO EN LÍNEA 10
II.1.7.1.1 Procesamiento relacional OLAP (ROLAP) 10
II.1.7.1.2 Procesamiento multidimensional OLAP (MOLAP) 11
II.1.7.1.3 Procesamiento híbrido OLAP (HOLAP) 11
II.1.7.2 DIFERENCIAS ENTRE ROLAP Y MOLAP 12
II.1.7.2.1 Ventajas y desventajas de ROLAP 12
II.1.7.2.2 Ventajas y desventajas de MOLAP 12
II.2 MODELO MULTIDIMENSIONAL 13
II.3 ESQUEMA ESTRELLA 14
II.4 PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y TRANSPORTE (ETT) 14
II.4.1 Extracción 15
II.4.2 Transformación 15
II.4.3 Transporte 15
II.4.3.1 Carga Inicial 15
II.4.3.2. Carga Histórica 15
II.4.3.3 Carga Incremental 15
II.5 TOMA DE DECISIONES [19] 16
II.5.1 TIPOS DE DECISIONES [10] 16
II.5.2 INDICADORES [25] 16
II.5.2.1 ATRIBUTOS DE UN BUEN INDICADOR 17
II.6 SISTEMA DE SOPORTE A LA TOMA DE DECISIONES (SSD) 17
II.7 PRONÓSTICO [8,28] 18
II.7.1 TENDENCIA, ESTACIONALIDAD Y CICLAJE 19
II.7.2 PROMEDIOS MÓVILES 19
II.7.3 SUAVIZACIÓN EXPONENCIAL 20

II.7.4 REGRESIÓN LINEAL **21**
II.7.5 COMPARACIÓN DE TÉCNICAS ESTADISTICAS DE PRONÓSTICOS **22**
II.7.6 MEDICIÓN DEL ERROR EN EL PRONÓSTICO **22**
CAPITULO III. MARCO METODOLÓGICO **23**
III.1 FASES DEL MODELO **25**
III.1.1 Concepto de software **25**
III.1.2 Análisis de requerimientos **25**
III.1.2.1 Identificación de indicadores 25
III.1.2.2 Evaluación y selección de la herramienta de desarrollo 25
III.1.2.3 Evaluación y selección de los modelos estadísticos 26
III.1.2.3.1 Análisis de los datos 26
III.1.2.3.2 Implementación de técnicas estadísticas 26
III.1.2.3.3 Elaboración de pruebas 26
III.1.2.3.4 Evaluación de los resultados 26
III.1.3 Diseño global **26**
III.1.4 Diseño e implementación del almacén de datos **26**
III.1.4.1 Análisis del origen y destino de datos de datos 26
III.1.4.2 Diseño del almacén de datos 27
III.1.4.3 Implementación del almacén de datos 27
III.1.4.4 Definición de los procesos ETT 27
III.1.4.5 Implementación de los procesos de ETT 27
III.1.4.6 Elaboración de las pruebas ETT y carga del almacén de datos 27
III.1.5 Diseño de la aplicación de predicciones y reportes **27**
III.1.6 Implementación de la aplicación de predicciones y reportes **27**
III.1.7 Elaboración de las pruebas de la aplicación de predicciones y reportes **27**
CAPÍTULO IV. DESARROLLO **28**
IV.1 DESARROLLO DEL SISTEMA DE SOPORTE PARA LA TOMA DE DECISIONES **28**
IV.1.1 CONCEPTO DE SOFTWARE **28**
IV.1.2 ANALISIS DE LOS REQUERIMIENTOS **29**
IV.1.2.1 Identificación de indicadores 30
IV.1.2.2 Evaluación y selección de la herramienta de desarrollo 31
IV.1.2.3 Evaluación y selección de los modelos estadísticos 34
IV.1.2.3.1 Análisis de los datos 34
IV.1.2.3.2 Implementación de las técnicas estadísticas 37
IV.1.2.3.3 Elaboración de pruebas 38
IV.2.2.3.4 Evaluación de los resultados 38
IV.1.3 DISEÑO GLOBAL **39**
IV.1.4 ANALISIS, DISEÑO E IMPLEMENTACIÓN DEL ALMACEN DE DATOS **41**
IV.1.4.1 Análisis del origen y destino de datos 42
IV.1.4.2 Diseño del almacén de datos 42
IV.1.4.3 Implementación del almacén de datos 43
IV.1.4.4 Definición de los procesos de ETT 44
IV.1.4.5 Implementación de los procesos ETT 45
IV.1.4.6 Elaboración de las pruebas ETT 47
IV.1.5 DISEÑO DETALLADO DE LA APLICACIÓN DE PREDICCIONES Y REPORTES **48**
IV.1.6 IMPLEMENTACION DE LA APLICACIÓN DE PREDICCIONES Y REPORTES **49**
IV.1.6.1 Módulo de pronósticos 50
IV.1.6.2 Módulo de reportes 52
IV.1.6.2.1 Reportes 53

IV.1.6.2.2 Reportes Gráficos....53
IV.1.7 PRUEBAS DE LA APLICACIÓN DE PREDICCIONES Y REPORTES....55
IV.1.7.1 Prueba del módulo de pronósticos....55
IV.1.7.2 Pruebas del módulo de reportes....56
IV.1.7.3 Pruebas con los usuarios del sistema....57
IV.2 DESARROLLO DEL SISTEMA WEB DE SEGUIMIENTO Y PROCESAMIENTO DE PEDIDOS....57
IV.2.1 CONCEPTO DE SOFTWARE....57
IV.2.2 ANÁLISIS DE LOS REQUERIMIENTOS....58
IV.2.3 DISEÑO GLOBAL....58
IV.2.4 DISEÑO DETALLADO....58
IV.2.4 IMPLEMENTACIÓN....59
IV.2.5 PRUEBAS DEL SISTEMA TRANSACCIONAL....60
CAPÍTULO V. RESULTADOS....61
CONCLUSIONES....63
RECOMENDACIONES....65
REFERENCIAS BIBLIOGRÁFICAS....66
GLOSARIO....71

ÍNDICE DE FIGURAS

Figura 1. Estructura y procesos de un almacén de datos. 7
Figura 2. DATA MARTS. 8
Figura 3. Infraestructura del almacén de datos. 9
Figura 4. Ejemplo de un cubo multidimensional. 13
Figura 5. Esquema estrella. 14
Figura 6. Elementos de un SSD. 18
Figura 7. Modelo del ciclo de desarrollo elaborado para la realización del proyecto. 24
Figura 8. Gráfica, Cantidad de pedidos Vs. Tiempo 35
Fuente: Elaboración propia. 35
Figura 9. Gráfica, Cantidad de recargos de lluvia Vs. Tiempo 36
Fuente: Elaboración propia. 36
Figura 10. Arquitectura del sistema. 40
Figura 11. Diseño del almacén de datos. 43
Figura 12. Flujo de datos para la Tarea de Extracción de la dimensión Cuenta. 46
Figura 13. Flujo de datos de la tarea Extracción de datos para FactFactura 46
Figura 14. Ventana de pronóstico de recargos de lluvias. 52
Figura 15. Ventana de pronóstico de recargos de lluvias. 53
Figura 16. Ventana del módulo de reportes gráficos. 55

ÍNDICE DE TABLAS

Tabla 1. Diferencias entre OLTP y OLAP. 10
Tabla 2. Diferencias entre ROLAP y MOLAP. 12
Tabla 3. Ventajas y desventajas de ROLAP. 12
Tabla 4. Ventajas y desventajas de MOLAP. 12
Tabla 5. Comparación de técnicas estadísticas de pronósticos. 22
Tabla 6. Lista de casos de uso del sistema. 29
Tabla 7. Descripción de indicadores. 31
Tabla 8. Características de las herramientas de inteligencia de negocios. 32
Tabla 9. Requerimientos mínimos de las herramientas de inteligencia de negocios. 32
Tabla 10. Lista de técnicas seleccionadas para cada pronóstico. 39
Tabla 11. Resultados de los procesos ETT. 47
Tabla 12. Lista de casos de uso. 57

INTRODUCCIÓN

Los sistemas de información están formados por conjuntos de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa. Existen diversos tipos de sistemas de información según la funcionalidad que los caracteriza. Un sistema de soporte a la toma de decisiones es un sistema de información que auxilia a las empresas en la toma de decisiones. Este sistema se encarga de analizar los datos de forma que presente alternativas de solución para las empresas que desean obtener una ventaja competitiva en el mercado.

El origen histórico de los sistemas de soporte a la toma de decisiones se encuentra en dos puntos de referencia de la década de los cincuenta. Por una parte, la aparición de los primeros ordenadores y la transición hacia los sistemas de información acaecida en la década de los setenta. Por otra, los trabajos realizados en el campo de la investigación operativa y la programación lineal que impulsó el departamento de Defensa de los EEUU desde la segunda Guerra Mundial. Ambos caminos de investigación y desarrollo se unieron definitivamente en los primeros años de la década de los setenta proporcionando los primitivos sistemas de soporte a la toma de decisiones.

Estos sistemas de soporte permiten extraer y manipular información de manera flexible así como también permiten al usuario definir interactivamente el tipo de información que necesita visualizar.

Un sistema de soporte a la toma de decisiones se distingue del resto de los sistemas de información porque en él se integran la interfaz de usuario, un modelo de decisiones, base de datos multidimensionales, y análisis de escenarios. Estos sistemas presentan la solución más eficiente para visualizar la actual situación de la empresa a través de informes dinámicos así como permite la posibilidad al usuario de poder visualizar el comportamiento futuro de sus servicios y ventas.

En resumen, el propósito de este trabajo especial de grado es desarrollar un sistema de soporte a la toma de decisiones que permita a la empresa Flamberge a través de una interfaz de usuario generar reportes que proporcionen información para el análisis de la situación actual de la empresa así como la generación de predicciones que les permita realizar una planificación en base a lo que ocurrirá en el futuro.

Para la realización de este proyecto se va a utilizar como metodología un modelo creado específicamente para el desarrollo del mismo. Debido a que la estructura del proyecto necesita que ciertas fases de su desarrollo se hagan de manera iterativa se creó un diseño que permite desarrollar las diversas etapas del ciclo de desarrollo del software combinado con un modelo iterativo, y de esta forma adaptar el proyecto a este modelo de manera que se pueda alcanzar un producto adecuado para el cliente.

En el capítulo I del presente trabajo se muestra el problema que originó la necesidad de elaborar este proyecto, también se podrá observar el objetivo general, los objetivos específicos, las limitaciones y los alcances del mismo.

En el capítulo II se encuentra el marco teórico, donde se exponen los basamentos teóricos recopilados durante la elaboración del proyecto para comprender el problema y los elementos que forman parte del proyecto.

En el capítulo III se explica con detalle la metodología utilizada para la elaboración del proyecto.

En el capítulo IV se presenta el desarrollo del proyecto siguiendo la metodología utilizada, y se especifica el producto elaborado en cada fase del modelo.

En el capitulo V se muestran los resultados obtenidos durante la elaboración del proyecto, las conclusiones, las recomendaciones y por último se presentan los apéndices que fueron necesarios crear para la construcción del proyecto.

CAPÍTULO I. PROBLEMA Y OBJETIVO

I.1 PLANTEAMIENTO DEL PROBLEMA

Flamberge es una empresa dedicada al seguimiento y procesamiento de pedidos a nivel nacional. Su negocio consiste en prestar servicios referentes a diligencias y encomiendas cooperativas. Cada servicio que presta tiene un costo que viene determinado por ciertas características como son el peso del material, el tamaño, el tiempo que tarda el envío, los recargos por lluvia, carga entre otros. La empresa presta servicios de encomiendas y diligencias con distintas clasificaciones que determinan el tipo de servicio requerido, según la necesidad del remitente, la persona que emite la orden de servicio selecciona la clasificación más adecuada de la operación, estas clasificaciones pueden ser: servicio normal, donde la entrega de la encomienda es dentro de un periodo de ocho horas hábiles, servicio de gestión, que se refiere a cuando un envío requiere una operación ante un tercero, servicio rápido, donde la encomienda debe realizarse dentro de un rango de cuatro horas, y el servicio nacional que consiste en una entrega fuera del área de caracas.

Flamberge ofrece sus servicios a todo el país, y según la ubicación de la zona de destino las diligencias y encomiendas tendrán una tarifa. La empresa en los últimos años ha presentado un aumento considerable en el número de clientes y servicios prestados. Debido a este crecimiento y al no llevar un control sobre estos procesos, Flamberge ha registrado pérdidas considerables. Es difícil en la actualidad para los gerentes de la empresa determinar la cantidad de pedidos que se realizarán en fechas especificas del año y esto presenta un gran inconveniente a la hora de planificar y organizar el personal que se necesitará para prestar los servicios de diligencias y encomiendas, produciéndose pérdidas considerables por no contar con el personal suficiente para prestar los servicios. La empresa asigna a cada zona de trabajo una cantidad determinada de empleados para que se encarguen de realizar los pedidos en las zonas correspondientes, esto presenta una dificultad en la actualidad para la empresa ya que hay zonas que requieren mayor cantidad de personal en algunas fechas del año y es difícil determinar este comportamiento. Adicionalmente en épocas donde la demanda de pedidos es elevada no se realiza una buena estimación en el inventario y el material no es suficiente para poder realizar las diligencias y encomiendas, esto ocasiona que ciertos pedidos no puedan ser atendidos.

Flamberge necesita contar con un sistema que le permita hacer predicciones de consumo por zonas, determinar tendencias de subidas y bajadas de pedidos, determinar niveles críticos de consumo según épocas del año, así como contar con el material en inventario necesario para la realización de los pedidos. Actualmente las operaciones no pueden ser monitoreadas y controladas en su totalidad produciendo un desajuste en los procesos de la empresa impidiendo que se realicen estimaciones que disminuyan los costos y las pérdidas. Es necesario un sistema que les permita realizar las predicciones anteriormente mencionadas y así mejorar la calidad de servicio que prestan a sus clientes, y permitirles crecer como empresa al poder abarcar un mercado más amplio.

I.2 OBJETIVO GENERAL

Desarrollar un sistema de apoyo a la toma de decisiones que permita generar indicadores que modelen el comportamiento de consumo y genere predicciones en base a estos comportamientos en el área de facturación, pedidos y necesidades de recursos humanos de la empresa Flamberge.

I.3 OBJETIVOS ESPECÍFICOS

I.3.1 Determinar a través del levantamiento de información los indicadores estadísticos que manejará el sistema de apoyo para la toma de decisiones.

I.3.2 Diseñar e implementar el almacén de datos que utilizará el sistema de apoyo para la toma de decisiones.

I.3.3 Evaluar y seleccionar los modelos estadísticos para la generación de indicadores.

I.3.4 Evaluar y seleccionar herramientas de desarrollo para implementar el modelo estadístico seleccionado.

I.3.5 Desarrollar un sistema basado en la herramienta seleccionada que permita generar los indicadores determinados y las predicciones de comportamiento de consumo.

I.3.6 Desarrollar un sistema transaccional de seguimiento y procesamiento de pedidos.

I.4 ALCANCE

I.4.1 El sistema de apoyo a la toma de decisiones generará reportes apoyado en herramientas estadísticas y técnicas de procesamiento analítico, basado en los indicadores que se determinarán en el levantamiento de información.

I.4.2 El sistema SSD proporcionará información para el apoyo de toma de decisiones a nivel estratégico y administrativo con respecto a los patrones de consumo, tendencias de subidas y bajadas de pedidos, determinación de niveles críticos por zonas y control de inventarios.

I.4.3 El sistema de apoyo para la toma de decisiones permitirá activar alarmas cuando se presenten desvíos o alteraciones sobre las cantidades promedio de consumo por departamento, por zonas o por fechas.

I.4.4 El sistema transaccional permitirá a los clientes de la empresa realizar pedidos y poder realizar el seguimiento del estado de éstos a través de Internet.

I.4.5 El sistema transaccional permitirá a los usuarios de la empresa procesar los pedidos realizados por sus clientes.

I.4.6 Implementación del modelo de datos, las interfaces diseñadas y la aplicación Web del sistema transaccional de seguimiento y procesamiento de pedidos.

I.5 LIMITACIONES

El sistema de apoyo a la toma de decisiones estará fundamentado en los datos históricos que se obtengan del sistema transaccional, y generará indicadores relacionados con las áreas de facturación, pedidos y recursos humanos de la empresa.

I.6 JUSTIFICACIÓN

En la empresa Flamberge cada día tanto sus clientes como los servicios que ofrece van en aumento, por ello surge la necesidad de poder controlar la calidad de servicio que prestan. Actualmente la empresa necesita contar con un sistema que le permita tomar decisiones en cuanto a la calidad de servicios, programación de pedidos y recursos humanos de acuerdo a predicciones de consumo y servicios para poder mejorar sus ingresos y tener bases para realizar mejores planificaciones de sus presupuestos.

El principal fundamento del proyecto se basa en desarrollar un sistema que de apoyo y mejore los procesos de toma de decisiones a nivel gerencial y estratégico de la empresa basándose en reportes por zonas y en indicadores de comportamiento de consumo.

Para poder realizar un sistema de apoyo para la toma de decisiones que logre el propósito antes mencionado es necesaria la ampliación del sistema transaccional. El sistema operacional con el que cuentan actualmente no abarca todos los procesos involucrados en el negocio y no facilita las operaciones diarias ya que los clientes no pueden realizar pedidos

de manera fácil y rápida a través de un medio automatizado. Por consiguiente se desarrollará un sistema transaccional que permita a los clientes el servicio de realizar pedidos por internet. El desarrollo de este sistema transaccional proporcionará información vital para el sistema de soporte para la toma de decisiones; este último, combinará modelos de análisis y de datos que podrá guiar al establecimiento de planes estratégicos para mejorar los servicios, proporcionando a los usuarios de la empresa una herramienta que permita tomar decisiones importantes en base a los resultados obtenidos referentes a predicciones de comportamiento de consumo por zonas y servicios, control de inventario de guías, pronósticos de pedidos, entre otros.

CAPÍTULO II. MARCO TEORICO

II.1 ALMACEN DE DATOS [9,21]

Es un repositorio completo de datos de la empresa, donde se almacenan datos estratégicos, tácticos y operativos, con el objetivo de obtener información estratégica y táctica, sus características principales son:

- orientados a temas: se centra en entidades de alto nivel, no en procesos.
- integrado: los datos se almacenan en un formato consistente.
- variante en el tiempo: los datos están asociados a un instante del tiempo.
- no volátil: este tipo de bases de datos es normalmente de sólo lectura.

El ambiente de un almacén de datos se organiza alrededor de sujetos tales como cliente, vendedor, producto y actividad. Por ejemplo, para un fabricante, éstos pueden ser clientes, productos, proveedores y vendedores.

La alineación alrededor de las áreas de los temas afecta el diseño y la implementación de los datos encontrados en el almacén de datos. Las principales áreas de los temas influyen en la parte más importante de la estructura clave. Las aplicaciones están relacionadas con el diseño de la base de datos y del proceso. La construcción del almacén de datos se enfoca al modelo de datos y el diseño de la base de datos. El diseño del proceso (en su forma clásica) no es separado de este ambiente.

Las diferencias entre la orientación de procesos y funciones de las aplicaciones y la orientación a temas, radican en el contenido de la data a nivel detallado. En el almacén de datos se excluye la información que no será usada por el proceso de soporte de decisiones, mientras que la información de las orientadas a las aplicaciones, contiene datos para satisfacer de inmediato los requerimientos funcionales y de proceso, que pueden ser usados o no por el analista de soporte de decisiones.

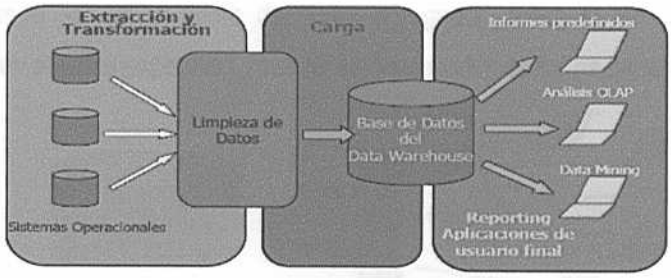

Figura 1. Estructura y procesos de un almacén de datos.

Fuente: Adaptación de [12]

II.1.2 VENTAJAS E INCONVENIENTES DE UN ALMACEN DE DATOS [16]

El almacén de datos aporta facilidad e inmediatez en el manejo de la información, así como logra la cohesión de los departamentos de una empresa. Las ventajas que esta tecnología aporta a la gestión empresarial son:

- Alto retorno de inversión. Una organización debe invertir una gran cantidad de recursos para asegurar el éxito de la implementación de un almacén de datos. Según un estudio realizado por la IDC (*International Data Corporation*) en 1996, en un período de 3 años la mayoría de las empresas obtienen beneficio sobre lo invertido, ROI (*return of investment*).
- Ventaja competitiva. El enorme retorno de la inversión es evidencia de la gran ventaja competitiva que acompaña a esta tecnología. La ventaja competitiva se consigue por permitir a los responsables de la toma de decisiones tácticas y estratégicas el acceso a datos que pueden mostrar información que antes era no disponible, desconocida o sin explotar, en clientes, tendencias y demandas.
- Incremento de la productividad de los responsables de la toma de de decisiones. Los almacenes de datos integran datos de múltiples sistemas incompatibles y proporcionan una vista consistente de la organización. La transformación de los datos orientados a las aplicaciones en información orientada a la toma de decisiones permite a dichos responsables realizar un análisis más preciso y consistente.

II.1.3 DATAMARTS (DM) [16]

Un datamarts es un pequeño almacén de datos diseñado para una unidad de negocio. Un datamart se centra únicamente en los requerimientos de los usuarios asociados con dicha unidad de negocio, por tanto son más fáciles de entender y navegar. Aunque el datamart puede ser un conjunto de datos agregados sobre el almacén de datos, suele ubicarse en un soporte específico y residir en un servidor de forma separada.

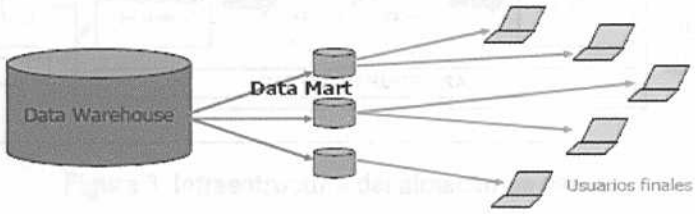

Figura 2. DATA MARTS.

Fuente: Adaptación de [12]

II.1.4 DIFERENCIAS ENTRE UN DATA MART Y UN ALMACEN DE DATOS [4]

- El DM se enfoca a un área temática.
- Un almacén de datos es único para la empresa, pero está compuesto por varios DM.
- El DM tiene un volumen de datos más limitado.
- El DM soporta menos cantidad de usuarios que un almacén de datos.

II.1.5 COMPONENTES DE UN ALMACÉN DE DATOS [11,16]

- Adquisición: Esta parte se encarga de la importación de los datos procedentes de los sistemas OLTP al almacén de datos y de llevar a cabo la preparación de datos. Esta parte puede consumir hasta el 70% del esfuerzo de construcción de un almacén de datos.
- Almacenamiento: El almacén de datos puede ser gestionado por un SGBD relacional o un SGBD multidimensional.
- Acceso: Este nivel trata directamente con el usuario final y consiste en una gran variedad de herramientas: interfaces de consulta "ad hoc" existentes en el entorno de bases de datos y lenguaje SQL, hojas de cálculo, generadores de informes, herramientas de visualización de datos, entornos de soporte a la toma de decisiones, paquetes de análisis estadístico, herramientas de procesamiento analítico en línea, aplicaciones de minería de datos, etc.
- Infraestructura: Los elementos de las tres partes anteriores deben estar soportados por una estructura física. El almacén de datos debe disponer de un hardware potente y de un buen sistema operativo.

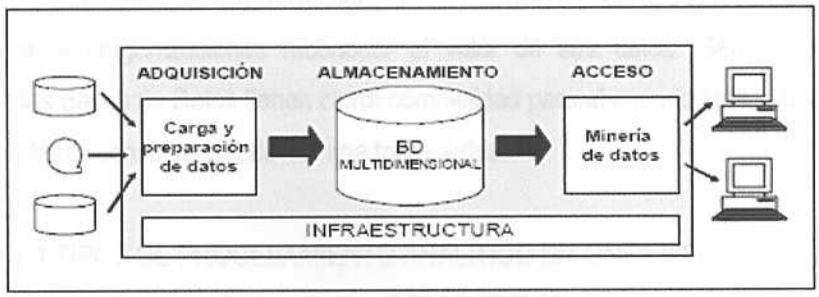

Figura 3. Infraestructura del almacén de datos.

Fuente: [16]

II.1.6 DIFERENCIAS ENTRE UN SISTEMA DE PROCESAMIENTO TRANSACCIONAL EN LÍNEA (OLTP, OnLine Transaction Processing) Y UN SISTEMA ANALÍTICO EN LÍNEA (OLAP, OnLine Analytical Processing) [11]

OLTP	OLAP
Predomina la actualización	Predomina la consulta
La actividad más importante es de tipo operativa (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.
Importancia en el tiempo de respuesta de la transacción instantánea.	Importancia de la respuesta masiva.
Estructura relacional.	Estructura multidimensional.
Usuarios de perfiles medios o bajos.	Usuarios de perfiles altos.

Tabla 1. Diferencias entre OLTP y OLAP.
Fuente: Adaptación de [11]

II.1.7 PROCESAMIENTO OLAP [26]

La tecnología OLAP facilita el análisis de datos en línea en un almacén de datos, proporcionando respuestas rápidas a consultas analíticas complejas. OLAP es utilizado generalmente para ayuda en la toma de decisiones y presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de presentación los usuarios finales pueden ver y entender con mayor facilidad la información de sus Bases de Datos, lo que permite a las organizaciones reconocer el valor de sus datos. Generalmente los esquemas de las Bases de Datos tienen cierta complejidad para el usuario final, debido a ello la concepción de las consultas puede ser una tarea ardua.

II.1.7.1 TIPOS DE PROCESAMIENTO ANALÍTICO EN LÍNEA [26]

II.1.7.1.1 Procesamiento relacional OLAP (ROLAP)

En ROLAP se utiliza una arquitectura de tres niveles. La BD relacional maneja el almacenamiento de datos, el motor OLAP proporciona la funcionalidad analítica, y alguna herramienta especializada es empleada para el nivel de presentación.

El nivel de aplicación es el motor OLAP, que ejecuta las consultas de los usuarios. El motor OLAP se integra con el nivel de presentación a través del cual los usuarios realizan los análisis OLAP. Después de que el modelo de datos para el DW se ha definido, los datos se cargan desde los sistemas transaccionales.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor OLAP, el cual transforma sus datos a consultas en SQL ejecutadas en las BD relacionales y sus resultados son devueltos a los usuarios.

La arquitectura ROLAP es capaz de usar datos precalculados (si estos están disponibles), o de generar dinámicamente los resultados desde la información elemental (menos resumida). Esta arquitectura accede directamente a los datos del DW y soporta técnicas de optimización para acelerar las consultas como tablas particionadas, soporte a la desnormalización, soporte de múltiples reuniones, precalculado de datos, índices etc.

II.1.7.1.2 Procesamiento multidimensional OLAP (MOLAP)

Un sistema MOLAP usa una Base de Datos multidimensional (BDMD), en la que la información se almacena multidimensionalmente. El sistema MOLAP utiliza una arquitectura de dos niveles: la BDMD y el motor analítico. La BDMD es la encargada del manejo, acceso y obtención de los datos. El nivel de aplicación es el responsable de la ejecución de las consultas OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP. La información procedente de los sistemas transaccionales se carga en el sistema MOLAP. Una vez cargados los datos en la BDMD, se realiza una serie de cálculos para obtener datos agregados a través de las dimensiones del negocio, poblando la estructura de la BDMD.

Luego de llenar esta estructura, se generan índices y se emplean algoritmos de tablas hash para mejorar los tiempos de accesos de las consultas. Una vez que el proceso de poblado ha finalizado, la BDMD está lista para su uso. Los usuarios solicitan informes a través de la interfaz y la lógica de aplicación de la BDMD obtiene los datos.

II.1.7.1.3 Procesamiento hibrido OLAP (HOLAP)

Se han desarrollado soluciones de OLAP híbridas que combinan el uso de las arquitecturas ROLAP y MOLAP. En una solución con HOLAP, los registros detallados (los volúmenes más grandes) se mantienen en la BD relacional, mientras que los agregados lo hacen en un almacén MOLAP independiente.

II.1.7.2 DIFERENCIAS ENTRE ROLAP Y MOLAP [26]

	Multidimensional	Relacional
Datos	Detalle y precalculados (agregados)	Detalle y agregados
Estructura	Matrices comprimidas	Tablas relacionales
Administración	Especialista en BDMD	Administrador BD
Acceso	Lenguaje especializado	SQL

Tabla 2. Diferencias entre ROLAP y MOLAP.

Fuente: Adaptación de [26]

II.1.7.2.1 Ventajas y desventajas de ROLAP

Ventajas	Desventajas
• Se puede aprovechar la tecnología relacional, ya que se facilita aprovechar las inversiones realizadas en hardware y en SGBD relacionales. • Uso de la seguridad e integridad de los SGBD relacionales. • Capaz de manejar conjunto de datos muy grandes, por encima de un terabyte. • Pueden utilizarse SGBD relacionales gratuitos. • Puede soportar un gran número de dimensiones. • Es escalable (adición de dimensiones a un esquema existente)	• Es menor en rendimiento frente a MOLAP en BD pequeñas. • Limitación para consultas complejas, debido a que se pueden requerir muchas reuniones para obtener la consulta deseada. • Utilización de mucho almacenamiento en disco.

Tabla 3. Ventajas y desventajas de ROLAP.

Fuente: Adaptación de [26]

II.1.7.2.2 Ventajas y desventajas de MOLAP

Ventajas	Desventajas
• Ofrece buen rendimiento cuando se trabaja sobre datos agregados, totales, subtotales, series temporales y diversos grados de detalle de los datos. • Facilita el estudio a alto nivel de los datos, al ofrecer una mayor flexibilidad y rapidez de acceso para el análisis de los mismos. • Almacenamiento de datos y consultas bastante eficientes. • Mayor rendimiento frente a ROLAP en el procesamiento de consultas en BD pequeñas. • Almacena agregados para facilitar un acceso rápido.	• La asimilación de los conceptos multidimensionales, en especial, cuando se tienen hipercubos de muchas dimensiones. • La construcción y poblado de las estructuras multidimensionales puede demandar mucho tiempo. • Están limitados a tener diez o menos dimensiones debido a la complejidad para el manejo de las mismas. • Debe trabajar con volúmenes de datos limitados menos de 5 GB. • Existen pocas herramientas gratuitas que lo soporten.

Tabla 4. Ventajas y desventajas de MOLAP.

Fuente: Adaptación de [26]

II.2 MODELO MULTIDIMENSIONAL [26]

En un modelo de datos multidimensional los datos se organizan alrededor de los temas de la organización. La estructura de datos manejada en este modelo son matrices multidimensionales o cubos. En un cubo la información se representa por medio de matrices multidimensionales o cuadros de múltiples entradas, que permite realizar distintas combinaciones de sus elementos para visualizar los resultados desde distintas perspectivas y variando los niveles de detalle. Esta estructura es independiente del sistema transaccional de la organización, facilita y agiliza la consulta de información histórica ofreciendo la posibilidad de navegar y analizar los datos. Los ejes del cubo son las Dimensiones, y los valores que se presentan en la matriz, son las Medidas.

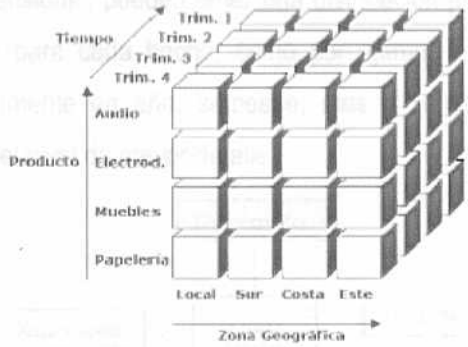

Figura 4. Ejemplo de un cubo multidimensional.

Fuente: [18]

Un modelo multidimensional está compuesto por:

- Medidas: Son características cualitativas o cuantitativas de los objetos que se desean analizar en las empresas. Las medidas cuantitativas están dadas por valores o cifras porcentuales.

- Dimensiones: representan cada uno de los ejes en un espacio multidimensional. Suministran el contexto en el que se obtienen las medidas de un hecho. Algunos ejemplos son: tiempo, producto, cliente, departamento, entre otras. Las dimensiones se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado. Los componentes de una dimensión se denominan niveles y se organizan en jerarquías, verbigracia, la dimensión tiempo puede tener niveles día, mes y año.

- Hecho: es el objeto a analizar, posee atributos llamados de hechos o de síntesis, y son de tipo cuantitativo. Sus valores (*medidas*) se obtienen generalmente por la aplicación de una función estadística que resume un conjunto de valores en un único valor.

II.3 ESQUEMA ESTRELLA [17]

El modelo estrella es un esquema relacional orientado a representación de dimensiones que convergen sobre un hecho común o valor cuantificable, de esta forma se logra implementar en un sistema de bases de datos tradicional el concepto de cubos multidimensionales. Existe una tabla central denominada tabla de hecho que representa el atributo cuantificable (Numérico), alrededor de esta tabla central se definen otras tablas ligadas a esta mediante claves foráneas, estas tablas se denominan tablas de dimensiones y son aquellas que definen cualidades del atributo representado por la tabla de hechos. Finalmente la clave primaria de la tabla de hecho será una clave compuesta por todas las claves foráneas que representan las claves primarias de cada una de las dimensiones asociadas. Las dimensiones pueden tener una distribución jerárquica dependiendo del nivel de detalle deseado para cada hecho, como por ejemplo una dimensión tiempo que se distribuya jerárquicamente en año, semestre, mes y semana, siendo año el nivel más resumido y semana el nivel de mayor detalle.

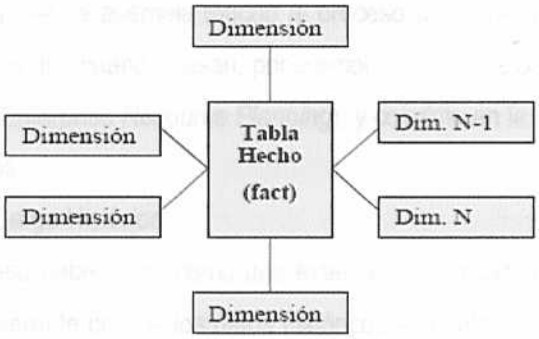

Figura 5. Esquema estrella.

Fuente: [17]

II.4 PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y TRANSPORTE (ETT) [18,20]

ETT es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos, datamart ó bodega de datos. ETT forma parte de la Inteligencia Empresarial (*Business Intelligence*), también llamado "Gestión de los Datos" (Data Management). La idea es que una aplicación ETT lea los datos primarios de unas bases de datos de sistemas principales, realice transformación, validación, el proceso cualitativo, filtración y al final escriba datos en el almacén y en este momento los datos son disponibles para analizar por los usuarios.

II.4.1 Extracción

Se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original.

II.4.2 Transformación

Es la etapa por la que puede atravesar una base de datos para estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos.

II.4.3 Transporte

Consiste básicamente en llevar los datos leídos y estandarizados a la bodega dimensional (puede ser remota o localmente). Si los requerimientos de datos incluyen algunos años de historia es necesario desarrollar tres conjuntos de programas ETL: una Carga Inicial, una Carga Histórica, y una Carga Incremental.

II.4.3.1 Carga Inicial

La carga inicial se asemeja mucho al proceso de conversión entre sistemas que se da en las organizaciones cuando pasan, por ejemplo, de sus viejos sistemas operacionales a un producto ERP (*Enterprise Resource Planning*), y consiste en la extracción, transformación, y carga de los datos.

II.4.3.2. Carga Histórica

Este proceso debe verse como una extensión de la carga inicial, pero la conversión aquí es un poco diferente porque los datos históricos son datos estáticos. A diferencia de los datos operacionales, los datos estáticos ya se archivaron en dispositivos de almacenamiento offline. Es común que con el transcurso del tiempo se eliminen elementos de datos que ya no sirven, se agreguen nuevos, se modifiquen los tipos de ciertos datos o los formatos de los registros, lo que implica que los datos históricos no necesariamente se puedan sincronizar con los datos operacionales. Por lo tanto los programas de conversión escritos para la carga inicial quizás no sean aplicables a la carga de datos históricos sin algunos cambios previos.

II.4.3.3 Carga Incremental

Una vez que el almacén de datos está cargado con datos iniciales e históricos, hay que desarrollar otro proceso para la carga incremental, que se ejecutará mensual, semanal o diariamente. Existen dos formas de diseñar la carga incremental:

- Extraer todos los registros: se extraen todos los registros operacionales, independientemente de los valores que hayan cambiado desde la última carga realizada. En

general esta opción no es viable debido al volumen de los datos, por eso la mayoría opta por extraer deltas.

• Extraer Deltas solamente: sólo se extraen registros nuevos o registros que contengan valores que cambiaron desde la última carga realizada. Diseñar programas ETT para extracciones delta es más fácil cuando las fuentes consisten en bases de datos relacionales y contamos con una columna timestamp para determinar los deltas.

II.5 TOMA DE DECISIONES [19]

Una decisión es un juicio o selección entre dos o más alternativas, que ocurre en numerosas y diversas situaciones de la vida. En líneas generales, tomar una decisión implica definir el propósito, es decir, qué es exactamente lo que se debe decidir, listar las opciones disponibles, evaluar las opciones, escoger entre las opciones disponibles la mejor y convertir la opción seleccionada en acción.

II.5.1 TIPOS DE DECISIONES [10]

Las decisiones estructuradas son aquellas decisiones repetitivas que solucionan hechos con los cuales la organización está familiarizada y tiene experiencia, normalmente este tipo de decisiones son frecuentes al nivel transaccional de la empresa. Las decisiones no estructuradas están relacionadas con casos nuevos, poco conocidos e inesperados, sin antecedentes, este tipo de decisiones son frecuentes en el nivel gerencial de las organizaciones y en las decisiones semiestructuradas una parte del problema es estructurado y algunos otros elementos no lo son.

II.5.2 INDICADORES [25]

Un indicador es toda medida cualitativa o cuantitativa que intenta medir algún suceso o proceso que posee un atributo dado. Un indicador está calculado sobre un solo componente o puede que se trate de un indicador compuesto basado en dos o más medidas. Ellos son necesarios para poder objetivar una situación determinada y a la vez poder evaluar su comportamiento en el tiempo mediante su comparación con otras situaciones que utilizan la misma forma de apreciar la realidad. En consecuencia, sin ellos tendríamos dificultades para efectuar comparaciones. Un indicador requiere siempre del uso de fuentes confiables de información y rigurosidad técnica en su construcción e interpretación.

II.5.2.1 ATRIBUTOS DE UN BUEN INDICADOR

Un buen indicador debe abarcar, como sea posible, el mayor número de las siguientes características:

- Disponibilidad: los datos básicos para la construcción del indicador deben ser de fácil obtención sin restricciones de ningún tipo.
- Simplicidad: el indicador debe ser de fácil elaboración.
- Validez: la validez de los indicadores significa que éstos deben tener la capacidad de medir realmente el fenómeno que se quiere medir y no otros.
- Especificidad: si un indicador no mide realmente lo que se desea medir, su valor es limitado, pues no permite la verdadera evaluación de la situación al reflejar características que pertenecen a otro fenómeno paralelo.
- Confiabilidad: los datos utilizados para la construcción del indicador deben ser fidedignos (fuentes de información satisfactorias).
- Sensibilidad: el indicador debe ser capaz de poder identificar las distintas situaciones de salud aún en áreas con distintas particularidades, independientemente de la magnitud que ellas tengan en la comunidad.
- Alcance: el indicador debe sintetizar el mayor número posible de condiciones o de distintos factores que afectan la situación descrita por dicho indicador. En lo posible el indicador debe ser globalizador.

II.6 SISTEMA DE SOPORTE A LA TOMA DE DECISIONES (SSD) [3,13, 23]

Se define como un conjunto de programas y herramientas que permiten obtener de manera oportuna la información que se requiere durante el proceso de la toma de decisiones que se desarrolla en un ambiente de incertidumbre. Los sistemas de apoyo para la toma de decisiones son posteriores a los sistemas de información (S.I) Transaccionales, dan apoyo a los mandos intermedios y a los gerentes administrativos, son interactivos y amigables.

Los SSD nacen como una necesidad de las empresas para tener información exacta cuando es requerida, reducir costos, realizar cálculos rápidos, eliminar límites cognitivos y de almacenamiento de información y para un soporte rápido y eficiente. Un SSD se distingue del resto de los sistemas porque lo integran los siguientes elementos: Interfaz de usuario, modelo de decisiones, base de datos y análisis de escenarios, con estos elementos integrados como SSD, el apoyo a las decisiones gerenciales es más poderoso que cualquier base de datos o sistema transaccional (contable, de recursos humanos, ingeniería, etc.) por sí solo, ya que su

función es auxiliar en la solución de problemas y en la toma de decisiones tanto semiestructuradas y no estructuradas. Los SSD abordan problemas donde los procedimientos para llegar a una solución podría no estar predefinido con anterioridad. Los SSD tienen más poder analítico que los demás sistemas, contienen explícitamente una variedad de modelos para el análisis de datos, de tal forma que el análisis de los datos sea sencillo para los encargados de tomar decisiones.

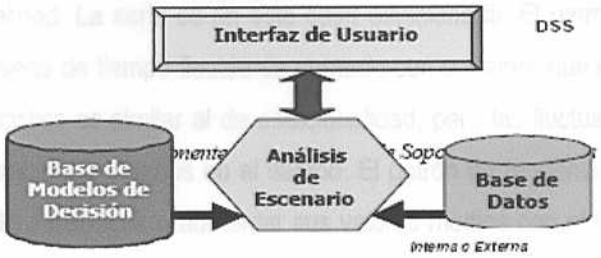

Figura 6. Elementos de un SSD.

Fuente: Adaptación de [13]

II.7 PRONÓSTICO [8,28]

El pronóstico es una técnica que ayuda a predecir lo que ocurrirá en el futuro. El futuro, por lo general, no es determinístico; ninguna técnica de pronóstico será aplicable a todos los procesos de decisión de una organización. Los siguientes elementos determinan, en un principio, que tipo de técnica deberá recomendarse en una situación dada: primero hay que determinar el horizonte de planeación que se predecirá (a un día, a un mes, a un año etc.), segundo, la incertidumbre que rodea todo el marco de toma de decisiones y por último la existencia de series históricas que sean relevantes y confiables.

El pronóstico es un elemento necesario del proceso de planeación, pero no es la planeación en sí. El pronóstico predice lo que pasará si las tendencias históricas no cambian; si esto no sucede (las tendencias históricas cambian), cada pronóstico debe ajustarse porque de otra manera resultará engañoso.

Las técnicas de pronósticos se dividen en dos grandes grupos: cuantitativas y cualitativas. Sólo las primeras requieren de una serie histórica de datos. En la familia de técnicas cuantitativas se encuentran, primordialmente, los promedios móviles, el alisamiento exponencial y el análisis de regresión. Las técnicas cuantitativas de pronósticos se subdividen, a su vez, en técnicas de series de tiempo y causales. Las primeras suponen que la serie histórica tiene un patrón o combinación de patrones que se repiten con el tiempo. Las segundas suponen que el valor de una variable depende es función de otras variables.

II.7.1 TENDENCIA, ESTACIONALIDAD Y CICLAJE

En una serie histórica de datos, conocida como serie de tiempo, existen tres patrones básicos que pueden o no presentarse en la serie y que son fundamentales para la selección de la técnica de pronósticos: la tendencia, la estacionalidad y el ciclaje.

Cuando una serie de tiempo no tiene una tendencia determinada, se dice que existe el patrón de horizontalidad. La serie es en este caso estacionaria. El patrón de estacionalidad existe cuando una serie de tiempo fluctúa de acuerdo con un factor que depende del periodo del año. El patrón cíclico es similar al de estacionalidad, pero las fluctuaciones ocurren más lentamente, son cambios graduados en el tiempo. El patrón de tendencia existe cuando una serie histórica tiende a disminuir o aumentar sus valores medios con el tiempo. Una serie de tiempo puede combinar los patrones de tendencias, ciclaje y estacionalidad. Sin embargo, alguno de estos patrones puede dominar la serie; por lo tanto existen técnicas de descomposición que permiten identificar el elemento dominante.

II.7.2 PROMEDIOS MOVILES

Una de las técnicas cuantitativas de pronósticos es promedios móviles. Esta técnica se utiliza en pronósticos a corto plazo. Es un método no estadístico que requiere de una serie histórica para suavizar o alisar el valor que se pronosticará. Este pronóstico se utiliza, a su vez, para predecir otros valores futuros. En esta técnica se promedia un periodo que contiene varios puntos de datos, dividiendo la suma de los valores de los puntos entre el número de puntos. Así, cada punto tiene la misma influencia. La fórmula viene dada por:

$$PS = \frac{Suma\ de\ demandas\ de\ todos\ los\ periodos\ anteriores}{Numero\ de\ periodos\ de\ demanda}$$

$$PS = \frac{D_1 + D_2 + D_3 + \ldots + D_k}{K}$$

Donde: D_1= demanda del período más reciente; D_2= demanda que ocurrió hace dos períodos; D_k= demanda que ocurrió hace k períodos.

Los resultados de este pronóstico será un promedio que es representativo del verdadero modelo subyacente, especialmente cuando se incrementa el número de pedidos empleados en el promedio, el promedio reduce así las posibilidades de error. Pero si el modelo subyacente cambia en el tiempo el promedio simple no permite detectar este cambio.

II.7.3 SUAVIZACIÓN EXPONENCIAL

La técnica de suavización exponencial supone que el proceso es constante, la misma suposición utilizada en el desarrollo del método de promedios móviles. Sin embargo, está diseñada para atenuar una desventaja del método de promedio móvil, donde se usa el mismo peso sobre todos los datos para calcular el promedio. Específicamente la suavización exponencial coloca un peso más grande en las observaciones más recientes.

Defina α (0< α <1) como la constante de suavización y suponga que los puntos de la serie de tiempo para los periodos pasados *t* son $y_1, y_2, \ldots, y_t$. Entonces y^*_{t+1} el estimado para el tiempo *t* + 1 se calcula como:

$$y^*_{t+1} = \alpha y_t + (1-\alpha) y^*_t$$

La selección de la constante de suavización α es crucial en la estimación de pronósticos futuros. Un valor grande de α implica que las observaciones recientes llevan pesos mayores. Esta técnica es la más utilizada debido a que: los modelos exponenciales tienen una precisión sorprendente, el usuario puede comprender como funciona el modelo, se requieren muy pocos cálculos para usar el modelo, como se usan datos históricos limitados, son pocos los requisitos de almacenamiento en computadores y es fácil calcular pruebas para determinar la precisión del modelo en la práctica.

Para el caso de la técnica de suavización exponencial existe un concepto importante: el rastreo. Debido a que la suavización exponencial supone la continuidad en el futuro de algún patrón histórico, resulta útil desarrollar una medida que se pueda utilizar para determinar cuándo se modifica el patrón básico. La medida más común es la señal de rastreo. Esta señal comprende el cálculo de alguna medición del error a través del tiempo y establece límites, de modo que cuando el error rebase dichos límites, se alerte al pronosticador.

Por ejemplo se puede emplear una señal de rastreo para determinar cuándo debe cambiarse el tamaño de alfa utilizado en una técnica de suavización exponencial. Ya que por lo regular se pronostican una gran numero de elementos, una práctica común consiste en continuar con el mismo tamaño de alfa durante varios periodos, antes de determinar si es necesaria una revisión. Por desgracia, la simplicidad de utilizar un modelo establecido de suavización exponencial es un fuerte motivador para no efectuar un cambio. Aunque en algún punto es necesario actualizar alfa.

II.7.4 REGRESIÓN LINEAL

Otra técnica de pronósticos es la de regresión lineal, se define como una relación funcional entre dos o más variables correlacionadas y se usa para pronosticar una variable con base en la otra. En la regresión lineal la relación entre las variables forma una línea recta. La línea de regresión lineal es de forma:

$$Y = b_0 + b_1 X$$

Donde Y es la variable dependiente que queremos resolver; b_0 es la intersección de Y; b_1 es la pendiente y X es la variable independiente (en el análisis de series de tiempo, X representa unidades de tiempo). Las constantes b_0 y b_1 se determinan a partir de los datos primarios con base en el criterio de los mínimos cuadrados como sigue a continuación:

$$b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \qquad b_0 = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n}$$

Sean (y_i , x_i) los datos primarios donde y_i es la demanda real del tiempo x_i ; i= 1,2,...,n.

Es necesario verificar lo bien que se ajusta el estimador lineal a los datos originales calculando el coeficiente de correlación r, ya que si los datos no presentan alguna tendencia lineal, el pronóstico obtenido no será correcto. El coeficiente de correlación está definido por:

$$r = \frac{\sum_1^n \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_1^n \left(x_i - \bar{x}\right)^2}\sqrt{\sum_1^n \left(y_i - \bar{y}\right)^2}}$$

El coeficiente de correlación puede valer cualquier número comprendido entre -1 y +1. Cuando r=1, la correlación lineal es perfecta, directa. Cuando r=-1, la correlación lineal es perfecta, inversa. Cuando r=0, no existe correlación alguna, independencia total de los valores X e Y.

La técnica de regresión lineal es útil para pronósticos a largo plazo. La restricción principal para usar los pronósticos de regresión lineal es que, supuestamente, los datos pasados y las proyecciones caen sobre una línea recta. En una parte del procedimiento se estima lo adecuado del ajuste de la línea con los datos. Se usa tanto para pronóstico de series de tiempo como para pronóstico de relaciones causales.

II.7.5 COMPARACIÓN DE TÉCNICAS ESTADISTICAS DE PRONÓSTICOS

Técnica	Horizonte de tiempo	Complejidad	Precisión del modelo	Requisitos de datos
Promedios Móviles	Corto	Baja	Media	Bajo
Suavizamiento Exponencial	Corto	Muy Alta	Adecuada	Muy Bajo
Regresión Lineal			Media Alta	Alto

Tabla 5. Comparación de técnicas estadísticas de pronósticos.

Fuente: Adaptación de [28]

II.7.6 MEDICIÓN DEL ERROR EN EL PRONÓSTICO [1,6]

Se han ideado diversos métodos para resumir los errores generados por una técnica particular de pronóstico. La mayoría de estas funciones implican promediar alguna función de la diferencia entre el valor real y su valor de pronóstico. A menudo se denominan residuales a estas diferencias entre valores reales y los valores de pronósticos.

Un método para evaluar una técnica estadística de pronóstico consiste en obtener la suma de los errores absolutos. La Desviación Absoluta de la Media (DAM) mide la precisión de un pronóstico mediante el promedio de la magnitud de los errores de pronósticos (valores absolutos de cada error). La DAM resulta de gran utilidad cuando el analista desea medir el error de pronóstico en las mismas unidades de la serie original. La ecuación siguiente muestra como se calcula la DAM:

$$DAM = \frac{\sum_{t+1}^{n} \left| y_t - y_t^* \right|}{n}$$

Donde Y_t = valor real en el periodo t, Y^*_t = valor del pronóstico en el periodo t y n= cantidad de medidas.

CAPÍTULO III. MARCO METODOLOGICO

El desarrollo de un proyecto debe regirse por una metodología con el fin de reducir riesgos y asegurar el buen cumplimiento de los objetivos generales y específicos que se hayan planteado, para así generar un producto de calidad.

En la realización de este proyecto se creó un modelo que permitiera establecer el orden según el cual realizar las actividades que componen la creación del sistema de apoyo para la toma de decisiones, se establecieron criterios para determinar el paso de una actividad a otra. Fueron tomadas como guía diversas etapas del modelo de cascada para proporcionar una línea base y poder establecer las fases del proyecto, fué necesaria la creación de un nuevo modelo que se ajustara a las necesidades del proyecto y que permitiera la iteración de ciertas fases hasta que se lograra el objetivo de alcanzar un producto adecuado para el cliente. Para ello se incorporo al modelo de cascada las fases iterativas que componen la creación del almacén de datos, para crear estas fases de iteración se utilizó como guía el modelo de entrega evolutiva siguiendo el ejemplo del libro de McConnell Steve (1996), titulado Desarrollo y gestión de proyectos informáticos.

Este modelo se utilizará como metodología a seguir para el desarrollo del proyecto, ya que los requerimientos del sistema están desde un principio bien definidos y se podrá seguir el modelo de cascada secuencialmente. Por tratarse de un sistema de apoyo a la toma de decisiones, las fases relacionadas con el almacén de datos requieren de muchos cambios, y el modelo seleccionado permitirá realizar dichos cambios si son necesarios a través de las iteraciones que se podrán realizar según el modelo de entrega evolutiva adaptado a este proyecto (Ver figura 7).

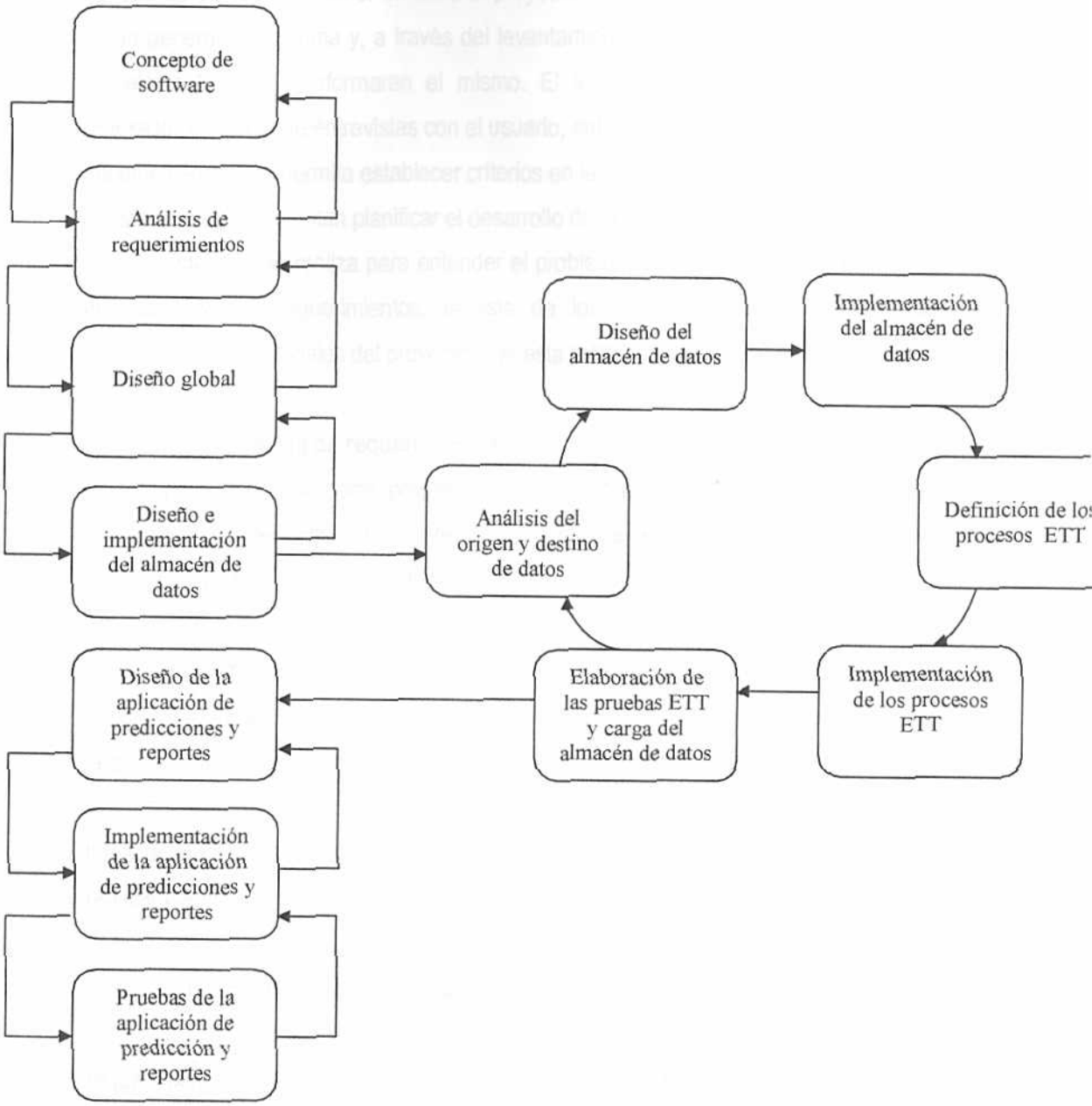

Figura 7. Modelo del ciclo de desarrollo elaborado para la realización del proyecto.

Fuente: Adaptación de [4].

III.1 FASES DEL MODELO

III.1.1 Concepto de software

En esta fase se recogen los requerimientos funcionales y no funcionales del proyecto. La idea de esta fase es tener en claro el proyecto que se va a construir. Se debe tener una visión general del sistema y, a través del levantamiento de información se deben establecer los elementos que conformaran el mismo. El levantamiento de información puede ser realizado por medio de entrevistas con el usuario, entrevistas con los gerentes de la empresa, documentación que permita establecer criterios en la realización del proyecto y la creación de cuestionarios que permitan planificar el desarrollo de las entrevistas.

Esta fase se realiza para entender el problema y buscar soluciones, en ella se hace una lista de los requerimientos, la lista de los casos de uso que representan los requerimientos funcionales del proyecto y de esta manera saber lo que se quiere construir.

III.1.2 Análisis de requerimientos

Esta fase tiene como propósito determinar qué se va a construir y cuáles son los elementos que intervienen en el sistema, así como los detalles de sus funcionalidades, y de esta manera poder obtener una descripción clara de qué sistema se va a desarrollar. Ésta se basa en la lista de requerimientos y lista de casos de uso de la fase anterior, para realizar los distintos documentos que servirán para realizar el diseño del sistema. Adicionalmente se pueden realizar las especificaciones de casos de uso, los modelos de casos de uso, es decir la documentación que sea necesaria para lograr una mejor comprensión sobre lo que desea el usuario y la manera de cómo lograrlo. En la fase de análisis del sistema de soporte para la toma de decisiones se incorporaron adicionalmente otras actividades que son necesarias realizar para cubrir las necesidades del proyecto. Estas actividades son las siguientes:

III.1.2.1 Identificación de indicadores

En esta actividad de seleccionaran los indicadores del sistema según los requerimientos obtenidos en la fase de concepto de software.

III.1.2.2 Evaluación y selección de la herramienta de desarrollo

Se analizarán las distintas herramientas disponibles en el mercado que permitan la realización del proyecto, estas herramientas serán comparadas y evaluadas para seleccionar una o un conjunto de ellas que soporten la estructura del proyecto y que mejor se ajusten a

las características de hardware con las que cuenta la empresa para la cual se desarrollará el sistema.

III.1.2.3 Evaluación y selección de los modelos estadísticos

En esta actividad se investigan las distintas técnicas estadísticas que existen para la elaboración de pronósticos, además se evalúa cada técnica para seleccionar aquella que mejor se ajuste al conjunto de datos. Esta actividad está compuesta a su vez por las siguientes actividades:

III.1.2.3.1 Análisis de los datos

Se realizará un análisis de la naturaleza de los datos para determinar si presentan alguna tendencia con relación a otros periodos, o si los datos tienen algún comportamiento particular que se mantiene en el tiempo como ciclaje o estacionalidad.

III.1.2.3.2 Implementación de técnicas estadísticas

El propósito de esta actividad es aplicar cada una de las técnicas estadísticas de predicciones a los conjuntos de datos obtenidos del sistema transaccional de la empresa.

III.1.2.3.3 Elaboración de pruebas

En esta actividad se realizan las pruebas a los resultados obtenidos y se determinan los porcentajes de error de cada técnica.

III.1.2.3.4 Evaluación de los resultados

Una vez realizadas las pruebas se analizan los resultados obtenidos, y se seleccionará la técnica de pronóstico que mejor se ajuste a cada conjunto de datos.

III.1.3 Diseño global

El propósito de esta fase es elaborar el diseño global de la arquitectura del sistema completo que se va a desarrollar.

III.1.4 Diseño e implementación del almacén de datos

Esta fase se dividió a su vez en varias fases que pueden realizarse iterativamente, hasta obtener el diseño e implementación del almacén de datos que mejor se ajuste a los requerimientos del proyecto.

III.1.4.1 Análisis del origen y destino de datos de datos

El propósito de esta fase es determinar cuáles son las posibles fuentes de datos, así como el análisis del destino de los datos que se utilizarán en el sistema de pronósticos y reportes.

III.1.4.2 Diseño del almacén de datos

En esta fase se identificarán las tablas del sistema transaccional que se utilizarán para extraer los datos que serán cargados en el almacén, además, se diseñará el modelo de datos necesario para almacenar la información sumarizada que, servirá como repositorio de datos para la generación de reportes y predicciones de acuerdo a los requerimientos del sistema.

III.1.4.3 Implementación del almacén de datos

Se desarrollará el almacén de datos. Se procederá a implementar el almacén que se definió en la etapa de diseño.

III.1.4.4 Definición de los procesos ETT

El propósito de esta fase es determinar los procesos que serán necesarios para extraer, transformar y cargar los datos en el almacén.

III.1.4.5 Implementación de los procesos de ETT

Se codifican y programan los procesos ETT del sistema.

III.1.4.6 Elaboración de las pruebas ETT y carga del almacén de datos

Esta fase surge por la necesidad de confirmar que los procesos de extracción implementados en la fase anterior son correctos y así, estar seguros de que los datos que se cargarán en el almacén son consistentes.

III.1.5 Diseño de la aplicación de predicciones y reportes

En esta fase se realizarán los diagramas que sean necesarios para comprender la aplicación de predicciones y reportes a desarrollar y que sirvan de guía en el proceso de implementación del proyecto. Además se diseñará la estructura de los reportes para la presentación de la información, así como las interfaces gráficas de acceso a los reportes y a las predicciones.

III.1.6 Implementación de la aplicación de predicciones y reportes

El propósito de esta fase es desarrollar el código que permita crear la estructura adecuada de la aplicación junto con los reportes, las predicciones y las interfaces diseñadas.

III.1.7 Elaboración de las pruebas de la aplicación de predicciones y reportes

Esta fase consiste en asegurar el buen funcionamiento y calidad del producto a través de la aplicación de varios casos de pruebas al producto final, de esta manera se asegura que el sistema integrado por las interfaces, el código de las predicciones y los reportes funcionen correctamente. Además se realizan las pruebas con los usuarios del sistema.

CAPÍTULO IV. DESARROLLO

En este capítulo se describe la elaboración del proyecto según la metodología explicada en el capítulo III. Además se presentan las etapas de desarrollo del sistema transaccional de seguimiento y procesamientos de pedidos.

IV.1 DESARROLLO DEL SISTEMA DE SOPORTE PARA LA TOMA DE DECISIONES

IV.1.1 CONCEPTO DE SOFTWARE

El concepto de software representa la primera fase del modelo utilizado para la realización del proyecto. En esta fase se realizaron diversas reuniones con los clientes del producto para determinar los requerimientos funcionales y no funcionales del sistema, para definir lo que el cliente quería. En estas reuniones se realizaron entrevistas para determinar los casos de uso que serán necesarios para el desarrollo del proyecto.

Las entrevistas fueron divididas en dos grupos principales. El primer grupo de entrevistas fue para determinar en líneas generales que era lo que el cliente necesitaba. Una vez que se realizó el primer grupo de entrevistas se pudo determinar que los requerimientos del cliente estaban relacionados con el desarrollo de una aplicación que les permitiera saber la situación actual de la empresa a través de informes de datos relacionados con pedidos y facturación, así como la generación de predicciones que les facilite la toma de decisiones en cuanto a las medidas futuras que deben tomar en lo relacionado con necesidades de recursos humanos, guías, inventarios, pedidos y facturación. Seguidamente se identificaron las áreas de negocio que están involucradas en la elaboración del sistema, se determinó que estas áreas son: recursos humanos, facturación y pedidos.

A continuación se procedió a realizar el segundo grupo de entrevistas, éstas se hicieron con el fin de determinar específicamente los requerimientos del sistema en lo relacionado con los reportes y las predicciones. Se realizó una lista de las predicciones que debía generar el sistema y otra lista con los requerimientos relacionados con la generación de reportes.

Luego de establecer los requerimientos funcionales del proyecto se procedió a determinar los requerimientos no funcionales. Como requerimiento no funcional se encontró que el sistema debe ser fácil de utilizar; ésta es una de las principales características que

debe presentar un sistema de apoyo para la toma de decisiones ya que es un sistema que va dirigido a la alta gerencia, normalmente son personas muy ocupadas que no tienen tiempo suficiente para dedicarlo a entender un sistema complicado, además la toma de decisiones es un aspecto delicado de la gerencia por consiguiente el sistema debe ser fácil de usar, para dejar el aspecto difícil que es lo relacionado con la decisión a tomar. Para ello se pensó en construir una aplicación que permitiera al usuario realizar las consultas de los reportes y generar las predicciones sin necesidad de que el usuario se saliera del sistema para ir a otra aplicación. La idea es crear un sistema integrado que permita la generación de predicciones, la generación de reportes y la generación de gráficos. Algunos de los casos de uso determinados en esta fase se listan en la siguiente tabla:

Numero	Nombre del caso de uso	Descripción
1	Generar Gráfico	Permite al gerente de la empresa generar el grafico asociado a un reporte específico.
2	Guardar Reporte	Permite al usuario guardar un reporte.
3	Exportar Reporte	Permite al gerente importar un reporte a Excel.
4	Parametrizar Reporte	Permite al gerente seleccionar nuevos parámetros para un reporte predefinido.
5	Consultar Reporte	Permite al gerente consultar un tipo de reporte predeterminado.
6	Parametrizar Reporte Grafico	Permite al gerente seleccionar nuevos parámetros para un reporte gráfico predefinido.
8	Generar Predicción	Permite al gerente generar una nueva predicción.
9	Consultar Alarma	Permite al usuario consultar las alarmas que se hayan generado en el sistema.

Tabla 6. Lista de casos de uso del sistema.

Fuente: Elaboración propia.

En el Apéndice A.1 se listan todos los casos de uso del sistema.

IV.1.2 ANALISIS DE LOS REQUERIMIENTOS

En esta fase se procedió a realizar el análisis de los requerimientos obtenidos en la fase anterior. Se determinó qué es lo que se va a construir en el proyecto; para esto; se tomó como base la lista de casos de usos, los cuales representan los requerimientos funcionales del sistema. A partir de la lista de casos de uso obtenida en la fase anterior, se procedió a realizar el modelo de casos de uso y la especificación de casos de uso.

Las especificaciones de los casos de uso se realizaron con la finalidad de comprender mejor como interactúa el usuario con el sistema, además éstos permiten conocer en la etapa de implementación lo que se va a desarrollar. En el Apéndice A.2 se muestra el modelo de casos de uso y en el Apéndice A.3 se muestran las especificaciones de casos de uso del sistema.

Además de realizar las actividades básicas del proceso de análisis, también se llevaron a cabo otras actividades necesarias para el desarrollo del proyecto, estas actividades se describen a continuación:

IV.1.2.1 Identificación de indicadores

En base a los requerimientos obtenidos en la fase anterior se identificaron los indicadores necesarios para generar las predicciones y los reportes del sistema SSD. Estos indicadores están relacionados con las medidas que permitirán monitorear los procesos de la empresa.

Los indicadores identificados se muestran a continuación:

Indicador	Descripción
Cantidad de envíos	Contiene información relacionada con el número de envíos realizados según el tiempo al que está asociado. Esta información se obtiene de la suma de los registros existentes por cada envío realizado en la tabla "Cargo" de la base de datos transaccional, en base a la fecha de creación de dicho registro.
Cantidad de guías	Contiene información relacionada con el número de guías utilizadas en los distintos servicios según el tiempo al que está asociado. Dicha cantidad es el resultado de la suma de los registros generados en la tabla "Guía" de la base de datos transaccional, en base a la fecha de creación del mismo.
Cantidad de recargos	Contiene información relacionada con el número de recargos aplicados a un servicio según el tiempo al que está asociado. Resulta de la suma de la cantidad de registros generados en la tabla "Cargo" de la base de datos transaccional, agrupados por su fecha de creación.
Monto facturado	Contiene información relacionada con el monto facturado según el tiempo al que está asociado. Se obtiene de la suma de los montos facturados reflejados en la tabla "Factura" de la base de datos transaccional en base a la fecha de creación de los registros en dicha tabla.
Monto recargo	Contiene información relacionada con el monto generado por los recargos realizados según el tiempo al que está asociado. Esta

	información se obtiene de la suma de los montos registrados por recargos en cada envío realizado en la tabla "Cargo" de la base de datos transaccional, en base a la fecha de creación de dicho registro.
Monto envío	Contiene información relacionada con el monto asociado a los envíos realizados. Resulta de la suma de los montos generados por concepto de envíos en la tabla "Cargo" de la base de datos transaccional, agrupados por la fecha de creación de este registro.

Tabla 7. Descripción de indicadores.

Fuente: Elaboración propia.

IV.1.2.2 Evaluación y selección de la herramienta de desarrollo

El desarrollo del sistema de apoyo para la toma de decisiones debe contar con herramientas que permitan almacenar la información de manera adecuada para su análisis, facilitar la realización de consultas, la creación de gráficos estadísticos así como contar con procesos ETT para cargar la información en la estructura multidimensional. Se determinó que la herramienta necesaria para implementar el sistema de apoyo a la toma de decisiones debería tener las siguientes características:

- Creación de Cubos OLAP.
- Herramienta ETT que permita la extracción, manipulación y almacenamiento de datos en cualquier destino.
- Generación de informes basados en datos contenidos en cubos OLAP.
- Generación de gráficos estadísticos.
- Implementación de modelos estadísticos.

Uno de los aspectos primordiales para la implementación de un sistema de apoyo para la toma de decisiones es el manejador de base de datos, ya que en éste se implementará el almacén de datos de donde se obtendrá la información para su análisis, por ser este aspecto fundamental en el éxito del sistema fue el primero en evaluarse, y en base a estos resultados se prosiguió a seleccionar el lenguaje de programación compatible con el manejador seleccionado.

Para la selección del manejador de base de datos se realizaron dos comparaciones, la primera fue determinar cuál de estos manejadores tenía las características requeridas para la implementación del SSD (Ver tabla 8) y luego de los preseleccionados se determinó cuál de estos se adaptaba mejor a los aspectos técnicos de la computadora donde se instalará el sistema (Ver tabla 9).

Evaluación de las características de las herramientas de bases de datos:

Herramientas de bases de datos	Herramienta ETT	Creación de Cubos	Generación de informes OLAP	Generación de Gráficos
SQL Server 2005 Enterprise Edition	SI	SI	SI	SI
Oracle Business Intelligence Standard Edition 10.g	SI	SI	SI	SI
Essbase Hyperion 7.1	SI	SI	SI	SI
Business Objects	NO	NO	SI	SI

Tabla 8. Características de las herramientas de inteligencia de negocios.

Fuente: Adaptación de [22, 23, 24]

Como se puede observar en la evaluación anterior SQL Server 2005 Enterprise Edition, Oracle Business Intelligence Standard Edition 10.g y Essbase Hyperion 7.1 cuentan con las características necesarias para el desarrollo del sistema, a continuación se presenta la evaluación de las tres herramientas preseleccionadas desde el punto de vista de requerimientos técnicos tomando en cuenta los aspectos relevantes para el desarrollo. Por tratarse de una aplicación de escritorio con un solo usuario final no se tomó en cuenta, entre otros aspectos, que no aplican para el desarrollo del sistema, las capacidades de los manejadores en relación al número de usuarios finales y la concurrencia de usuarios.

Evaluación de los aspectos técnicos:

Manejador de bases de datos	Espacio en Disco	Memoria RAM	Procesador
Oracle Business Intelligence Standard Edition 10.g	1,6 GB	1 G	300 Mhz intel Pentium.
SQL Server 2005 Enterprise Edition	185 MB	512 MB	600 Mhz Pentium III
Essbase Hyperion 7.1	400 MB	1 G	600 Mhz Pentium III

Tabla 9. Requerimientos mínimos de las herramientas de inteligencia de negocios.

Fuente: Adaptación de [22, 23, 24]

SQL Server 2005 Enterprise Edition representa la mejor opción para la implementación del almacén de datos y los cubos OLAP del proyecto debido a que ocupa menos espacio en disco duro y necesita menor cantidad de memoria RAM, SQL server es el indicado para una aplicación de escritorio con un solo usuario final.

Luego de determinar que SQL Server es el manejador de bases de datos adecuado, se procedió a seleccionar un lenguaje de programación que fuese compatible con éste y que trabajara bajo el mismo entorno de desarrollo. C# es un lenguaje de programación que forma parte de la herramienta de desarrollo Microsoft Visual Studio 2005 .NET, que permite la creación de una interfaz amigable para el sistema de apoyo a la toma de decisiones y se puede conectar con las distintas herramientas de reportes compatibles con SQL Server, además C# ofrece una gran variedad de librerías que serán útiles para conectar con cubos OLAP, crear gráficos OLAP y programar fórmulas matemáticas para implementar las técnicas estadísticas de predicciones.

Una vez que se determinó que C# es un lenguaje que permite cubrir las necesidades del proyecto relacionadas con las conexiones a cubos y las fórmulas estadísticas, fue necesario determinar la herramienta que se utilizaría para la creación de reportes.

SQL Server 2005 provee una herramienta de reportes denominada Reporting Services, ésta es un instrumento que permite la creación de informes con datos OLAP así como la creación de gráficos, sin embargo dicho instrumento no puede integrarse a una aplicación de escritorio, sólo funciona en el entorno Web, por ello se buscó una herramienta adicional que pudiera integrarse a una aplicación de escritorio.

Se encontraron diversas herramientas que funcionan como aplicaciones clientes para poder generar y visualizar reportes y gráficos, sin embargo sólo se tomaron en cuenta aquellas herramientas que pudieran conectarse con Visual Studio 2005.

En la etapa de concepto de software se determinó que uno de los requerimientos no funcionales del sistema era que fuese fácil de utilizar, por eso el objetivo de la evaluación de la herramienta de desarrollo fue conseguir una herramienta de generación de reportes y gráficos que pudiera ser consultada desde C#. Crystal Report es la herramienta estándar de generación de reportes en el entorno Windows, Crystal Report puede utilizarse en el ambiente de Visual Studio para la generación de informes y gráficos, sin embargo luego de analizar esta herramienta se observó que Crystal Report no permite la creación de reportes dinámicos en el entorno de Visual Studio, para crear reportes dinámicos es necesario que el usuario utilice el software de Crystal Report en la propia herramienta, otra opción era llamar

estos reportes desde Visual Studio pero los reportes perdían el dinamismo necesario en un sistema de apoyo para la toma de decisiones, por ello se investigaron distintas herramientas que facilitaran la creación de reportes dinámicos y gráficos de tal manera que el usuario pudiese consultarlos sin necesidad de salir de la aplicación de predicciones.

Las herramientas clientes seleccionadas luego de la investigación fueron: OLAP Reporting Tool For Excel y ChartFX. OLAP Reporting Tool For Excel es una herramienta que permite la creación de reportes OLAP dinámicos en donde el usuario podrá cambiar dinámicamente lo que desee ver del reporte. Esta herramienta proporciona una interfaz amigable para la consulta de reportes a cubos OLAP. ChartFX es una librería creada para C chart donde se pueden crear gráficos en base a datos OLAP.

El manejador de base de datos SQL server junto con el ambiente de desarrollo Visual Studio y las herramientas de reportes y gráficos seleccionadas forman un conjunto de herramientas compatibles cuyas características se adaptan perfectamente a las necesidades y requerimientos del sistema de apoyo a la toma de decisiones.

IV.1.2.3 Evaluación y selección de los modelos estadísticos

El pronóstico es una técnica que ayuda a predecir lo que ocurrirá en el futuro. El futuro por lo general, no es determinístico [14]. Las técnicas de pronósticos se dividen en dos grandes grupos: cuantitativos y cualitativos, para el caso de este proyecto se seleccionó el grupo de pronósticos cuantitativos, ya que éstos serán realizados en base a una serie de datos históricos relacionados con facturación, servicios, zonas, guías y recursos humanos (Ver II.7). En el Apéndice B se muestra la lista de los pronósticos que fueron determinados en la fase inicial del proyecto.

Para llevar a cabo la evaluación y selección de las técnicas estadísticas fue necesario crear un conjunto de consultas que permitieran agrupar los datos de manera que se les pudiera aplicar las técnicas de pronósticos. Estas consultas fueron realizadas en SQL server donde se encuentran los datos del sistema transaccional de la empresa Flamberge. El resultado de las consultas fueron copiadas en Excel.

En la familia de técnicas cuantitativas de pronósticos se encuentran, primordialmente, suavización exponencial, análisis de regresión y promedios móviles [14]. Para seleccionar el modelo estadístico adecuado para cada conjunto de datos, se realizaron las siguientes actividades:

IV.1.2.3.1 Análisis de los datos

Los distintos conjuntos de datos no necesariamente presentan un comportamiento igual, por ello hay que analizar cada conjunto de datos por separado para determinar si existe

alguna tendencia en el tiempo, algún comportamiento lineal o alguna estacionalidad, esto es porque las técnicas de pronósticos no pueden ser aplicadas de igual manera si los datos presentan alguna tendencia periódica o algún comportamiento particular que se mantenga en el tiempo (Ver II.7.1).

Para cada conjunto de datos fue realizado un análisis para determinar su comportamiento y saber que técnicas de pronósticos eran las más adecuadas según su naturaleza, ya que si los datos presentan alguna tendencia significativa, las técnicas como suavización exponencial, regresión lineal o promedios móviles no serán buenos estimadores del comportamiento futuro y no es aconsejable su uso porque suprimirían la tendencia de los datos, para este caso habría que separar el rango de tiempo en varios periodos según la estacionalidad y aplicar el método más adecuado para cada periodo de tiempo.

Una de las formas más efectivas de analizar el comportamiento de los datos es a través de la creación de gráficos, los gráficos son una herramienta visual que permite buscar posibles patrones y relaciones entre las variables.

Para encontrar tendencias o estacionalidades en los datos se realizó una gráfica de serie de tiempo para cada uno de los conjuntos de datos. Cada conjunto estaba compuesto por datos de los últimos cuatro años de operaciones de la empresa Flamberge, agrupados por mes. En estas gráficas se pudo observar si los datos presentan alguna tendencia ascendente, descendente o alguna estacionalidad en periodos de tiempo específicos. Por ejemplo el siguiente gráfico fue creado con datos del servicio tipo 1 de la empresa denominado servicio Normal, para observar si existen estacionalidades en el comportamiento de los datos:

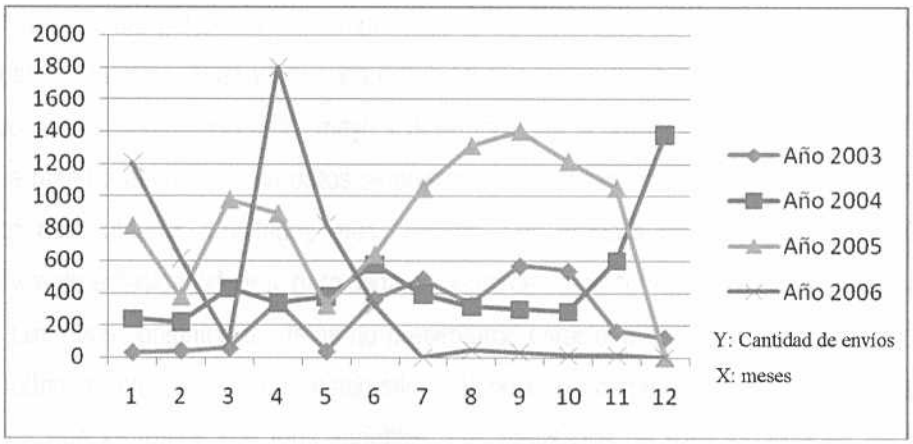

Figura 8. Gráfica, Cantidad de pedidos Vs. Tiempo

Fuente: Elaboración propia.

Como se puede observar en la Figura 8, no existe alguna tendencia significativa de estacionalidad, los datos se comportan de manera distinta en cada periodo del año en todos los años, lo que indica que no existe un periodo estacional o ciclaje que deba analizarse por separado.

Luego de analizar todos los gráficos de la forma mostrada anteriormente se observó que sólo un conjunto de datos presenta una tendencia en periodos de tiempo determinados, éste fue el conjunto de datos relacionados con los recargos de lluvia. A continuación se muestra el análisis de estacionalidad para este caso:

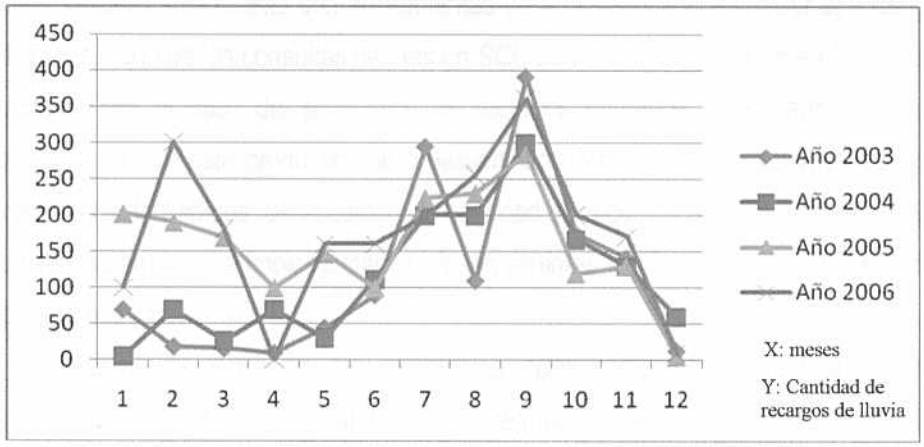

Figura 9. Gráfica, Cantidad de recargos de lluvia Vs. Tiempo
Fuente: Elaboración propia.

Como se puede observar en el gráfico los datos contenidos entre los meses de mayo y diciembre presentan estacionalidad, ya que los datos se comportan de la misma forma en los distintos años, entonces para este caso es necesario separar el conjunto de datos en dos periodos, el primer periodo es el comprendido entre los meses de enero y abril, este periodo no presenta estacionalidad ya que los datos se comportan distinto en cada año, el segundo periodo es el comprendido entre mayo y diciembre que si presenta estacionalidad. Al dividir de esta manera el conjunto de datos se podrán aplicar métodos estadísticos distintos a cada periodo para obtener resultados más certeros. Los gráficos creados para determinar la existencia de estacionalidades para cada conjunto de datos se muestran en el Apéndice C.1

Los otros conjuntos de datos no presentaron estacionalidades, lo que permite que se les puedan aplicar las técnicas estadísticas al periodo completo. Adicionalmente se pudo observar que los datos son muy variables. Los resultados de estacionalidades para cada pronóstico se muestran en el Apéndice C.2

Como los promedios móviles ponderan las variaciones cuando las demandas son bastante estables, surge un problema cuando aumenta el número de periodos a ponderar o la demanda no es estable, ya que los promedios móviles no pueden detectar muy bien las tendencias (Ver II.7.2). Debido a que son promedios, siempre se mantendrán dentro de los niveles pasados y no predicen un cambio hacia un nivel mayor o menor. Por ello el método de promedios móviles fue descartado luego de observar que los datos son muy variables.

IV.1.2.3.2 Implementación de las técnicas estadísticas

Para implementar las técnicas estadísticas se utilizó la herramienta Excel 2007 de Microsoft Office, Excel cuenta con herramientas de análisis que pueden ser aplicados a los datos. El resultado de las consultas hechas en SQL server fueron copiadas en Excel para ser evaluadas. Para el caso del pronóstico de recursos humanos no se aplicaron técnicas estadísticas, ya que este pronóstico será determinado a través de los valores obtenidos en los pronósticos de pedidos, de acuerdo a la cantidad de recursos de personal necesario para la cantidad de envíos. La implementación de las distintas técnicas estadísticas se realizó de la siguiente manera:

El análisis de regresión fue aplicado a cada conjunto de datos utilizando una herramienta de análisis de Excel denominada Regresión Lineal. Esta devuelve estadísticas que describen una tendencia lineal que coincide con los puntos de datos conocidos, mediante una línea recta usando el método de mínimos cuadrados.

Para implementar este método se utilizaron como parámetros las variables independientes que representan el tiempo y las variables dependientes que representan los valores para los cuales se desean determinar las predicciones (Ver II.7.4), como por ejemplo los montos de ventas. El conjunto de datos que representa las medidas como las cantidades de envíos o facturas se asociaron al mes respectivo.

La técnica de suavización exponencial fue implementada sin utilizar la herramienta que Excel proporciona para este tipo de predicción, ya que Excel trabaja con un solo ajuste de valores alfa, y este método proporciona mejores resultados en el pronóstico si se utilizan dos alfas que puedan ser combinadas según el comportamiento de los datos. La idea era buscar alfas que permitieran generar los mejores pronósticos (Ver II.7.3). Se creó una tabla en Excel que contiene los distintos valores de alfa comprendidos entre cero y uno. Luego se creó la fórmula de suavización exponencial de manera que los valores de alfa pudieran ser cambiados y sustituidos en la fórmula repetidas veces hasta encontrar la combinación adecuada. Se probaron todas las posibles combinaciones de alfa para cada conjunto de datos. La tabla de los distintos valores de alfa se muestra en el Apéndice D.1.

IV.1.2.3.3 Elaboración de pruebas

En esta actividad se procedió a realizar distintas pruebas a los resultados obtenidos para conocer la exactitud de los pronósticos con respecto a la data real y verificar si las técnicas estadísticas aplicadas eran las adecuadas para el conjunto de datos.

La regresión lineal no siempre se ajusta a los datos independientemente de que estos no presenten estacionalidades, aparte de determinar en la fase de análisis si los datos eran o no estacionales, hay que realizar pruebas adicionales. Al aplicar regresión lineal mediante la herramienta de análisis de Excel se obtiene, a parte de los valores de los pronósticos, un valor denominado coeficiente de correlación (R) que indica si ocurre un ajuste lineal entre la variable independiente (X) y la variable dependiente (Y) (Ver II.7.4).

Si el valor de R es mayor o igual a 0.90 quiere decir que los valores de X e Y presentan un ajuste lineal casi perfecto. Cuando esto ocurre el método de regresión lineal puede ser utilizado para encontrar las predicciones de las variables dependientes. Los coeficientes de correlación obtenidos en cada conjunto de datos se muestran en el Apéndice E.

Para encontrar la exactitud de las técnicas de pronósticos, se procedió a calcular para cada conjunto de datos el error DAM (Ver II.7.6). En el caso de la técnica de suavización exponencial se calculó un error para cada combinación de alfa realizado.

Los errores fueron calculados mediante una fórmula realizada en Excel. Esta fórmula recibe el valor pronosticado y el valor real de cada elemento del conjunto de datos y calcula el error absoluto total para cada conjunto. Los resultados de los errores DAM para cada conjunto de datos se muestran en el Apéndice F.

IV.2.2.3.4 Evaluación de los resultados

Una vez que se realizaron las pruebas se procedió a evaluar los resultados obtenidos en cada técnica estadística.

Al analizar los resultados obtenidos en los coeficiente de correlación se observó que todos los conjuntos de datos presentan una tendencia lineal, ya que los valores de R eran iguales o mayores a 0.9 lo que quiere decir que Y y X no son independientes.

Seguidamente fue evaluada con detalle la técnica de suavización exponencial para cada conjunto de datos, como en esta técnica se utilizaron diversas combinaciones de alfa, había que seleccionar el alfa que proporcionara la mejor exactitud en el pronóstico, para ello se comparó cada uno de los errores DAM obtenidos en las distintas combinaciones de alfa, y se seleccionó el valor que produjera el menor error en cada conjunto de datos. Los valores de alfa seleccionados para cada tipo de pronóstico se muestran en el Apéndice D.2.

El siguiente paso fue determinar cuál técnica se ajusta mejor a los datos si la regresión lineal o la suavización exponencial. Para esto se evaluaron los resultados obtenidos en las pruebas relacionadas con los errores. Se procedió a comparar los errores obtenidos tanto en la técnica de regresión lineal como en la técnica de suavización exponencial con los valores de alfa seleccionados previamente y de esta manera se determinó cuál de los dos métodos es el más exacto con respecto a los datos reales. Adicionalmente se realizaron gráficos para comparar visualmente las dos técnicas estadísticas con los datos reales. Los gráficos de comparación de técnicas estadísticas se muestran en el Apéndice C.3.

Una vez que se compararon los errores obtenidos en cada técnica se procedió a seleccionar aquella que ocasionara el menor error. Por ejemplo en la siguiente tabla se muestran algunos pronósticos del sistema con la técnica seleccionada.

Nombre del Pronóstico	Técnica Seleccionada
Facturación general	Suavización Exponencial
Recargo de lluvia	Suavización Exponencial
Cantidad de pedidos servicio normal	Suavización Exponencial
Cantidad de pedidos zona 1 servicio normal	Suavización Exponencial
Cantidad de pedidos zona 1 servicio rápido	Regresión Lineal
Cantidad de pedidos zona 1 servicio urgente	Suavización Exponencial
Cantidad de pedidos zona 2 servicio normal	Suavización Exponencial
Cantidad de pedidos zona 2 servicio gestión	Suavización Exponencial
Cantidad de pedidos zona 2 servicio rápido	Regresión Lineal
Cantidad de pedidos zona 2 servicio urgente	Suavización Exponencial

Tabla 10. Lista de técnicas seleccionadas para cada pronóstico.
Fuente: Elaboración Propia.

La lista completa con los pronósticos y las técnicas seleccionadas se muestran en el Apéndice C.4.

IV.1.3 DISEÑO GLOBAL

A continuación se describen los elementos que formarán parte del sistema y que fueron identificados en la fase de Análisis.

- Base de datos transaccional

La base de datos transaccional es la fuente de datos que alimenta al almacén. Esta base de datos posee los datos de las transacciones diarias que se realizan en la empresa.

Esta base de datos se creó en SQL Server 2005, y sus datos fueron migrados desde la base de datos original desarrollada en Access.

- Paquetes de Extracción, Transformación y Transporte (ETT)

Los paquetes son objetos que almacenan información correspondiente a la manera en que los datos serán extraídos, transformados y limpiados desde la fuente hasta el destino en el almacén de datos. Se crearán estos paquetes para poder realizar los procesos ETT y cargar los datos en el almacén. Estos paquetes se desarrollarán utilizando el módulo de Integration Services de SQL Server (SSIS por sus siglas en ingles).

- El Almacén de datos

Permite almacenar los datos provenientes de la base de datos transaccional y basado en un modelo multidimensional permitirá la creación de los cubos de datos. Es la fuente de datos de los procesos de análisis de pronósticos y de reportes OLAP.

Se diseñó e implementó este almacén usando el módulo de Análisis Services de SQL Server 2005 (SSAS por sus siglas en ingles).

- Aplicación de análisis de pronósticos y reportes OLAP

Sistema que incluye las funciones de reportes sobre los cubos así como también las funciones que permiten generar y consultar los pronósticos para un área en específico. Aporta al usuario final herramientas de análisis que le ayudarán a la toma de decisiones en base a esos resultados.

Esta aplicación se implementará en el lenguaje de programación C# usando MVC (Model View Controler) como patrón de diseño.

- Arquitectura del Sistema

Todos los elementos el sistema se relacionan entre si tal como se muestra en la arquitectura del sistema:

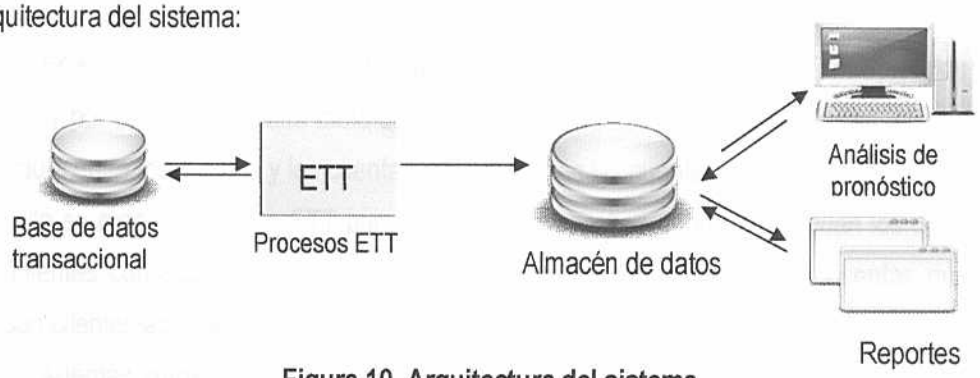

Figura 10. Arquitectura del sistema

Fuente: Elaboración Propia

En esta arquitectura se muestran las relaciones entre los elementos del sistema, en la cual una base de datos transaccional, con los datos de las operaciones diarias de la empresa es la fuente del Almacén de datos. Mediante el uso de procesos ETT se extraen los datos de la base de datos transaccional para ser transformados y limpiados para cargar el Almacén de datos. Una vez que este proceso culmina y los datos ya están presentes en dicho Almacén de datos como cubos multidimensionales, son consultados mediante la librería ADOMD.Net para realizar el análisis de los pronósticos y la generación de consultas por parte del sistema.

IV.1.4 ANALISIS, DISEÑO E IMPLEMENTACIÓN DEL ALMACEN DE DATOS

La fase de análisis, diseño e implementación del almacén de datos está dividida a su vez en otras fases que permiten distribuir de una manera adecuada las actividades necesarias en la creación del almacén de datos.

Es necesario diseñar e implementar un almacén de datos orientado a los temas requeridos que fueron determinados en el proceso de análisis y que sea capaz de soportar las decisiones gerenciales. Las fases que se describen a continuación fueron iteradas tres veces hasta que se completó el modelo e implementación final del almacén de datos (Ver III.1.4)

Durante el proceso de desarrollo del almacén de datos surgió la necesidad de crear una tabla que permitiese almacenar todas las alarmas generadas por el sistema de pronóstico, para poder procesarlas y llevar un control de dichas alarmas para un tiempo determinado. Además fue necesario crear una consulta para los procesos ETT en la que se filtró todas las fechas inválidas ya que el sistema contenia fechas del año 1900. Todo esto ocasionó que se reiniciaran las fases del diseño e implementación del almacén de datos generando una nueva iteración.

En esta segunda iteración surge la necesidad de crear procesos de ETT separados por cliente y por cuenta para cada tabla de hechos. Esto es debido a que existe una relación de jerarquía entre los clientes y las cuentas. No para todas las cuentas existe un cliente asociado por ello se creó un proceso ETT para cada tabla de hechos donde se carguen los datos de los clientes con sus respectivas cuentas y otro proceso para cargar las cuentas que no tengan clientes asociados.

Además surgió una tercera iteración en la que se agregó a la tabla cargas ETT del almacén (en la que se guardan las fechas de cada carga realizada) una columna que indica el número de lote asociado al conjunto de datos procesados para esa fecha.

IV.1.4.1 Análisis del origen y destino de datos

Se determinó que la base de datos transaccional de la empresa Flamberge posee toda la información que se necesita para implementar el almacén de datos, no todas las tablas de la base de datos se utilizarán, sólo aquellas que contienen los datos relevantes para la implementación del sistema de reportes y predicciones. El modelo entidad relación de la base de datos transaccional se muestra en el Apéndice G.

Para el destino de los datos se utilizará el modo de procesamiento ROLAP donde los datos de origen se almacenan en una estructura multidimensional en una base de datos relacional (Ver II.1.7). Esta estructura maximiza el rendimiento de las consultas. Los datos de la estructura ROLAP de la partición están tan actualizados como el procesamiento más reciente de la misma.

IV.1.4.2 Diseño del almacén de datos

En esta fase se identificaron las tablas de la base de datos transaccional que serán utilizadas para obtener los datos de origen, en base a ellas se procedió a diseñar las tablas de hechos y dimensiones que formarán parte del almacén de datos así, como sus respectivos atributos. Una vez que se identificaron los objetos que formarán parte del modelo, se procedió a realizar el diseño del almacén en base a un esquema estrella modelando las áreas del negocio. Este esquema permite la representación de las dimensiones que convergen sobre un hecho común o un valor cuantificable (Ver II.8). Se seleccionó el modelo estrella para proporcionar un proceso de carga más rápido relacionado con las dimensiones, ya que si las dimensiones están normalizadas, habría que realizar tantos Joins como tablas contenga la dimensión para ir poblando sus distintos niveles, en cambio en un modelo estrella las dimensiones no están normalizadas y el proceso de carga es más rápido.

Adicionalmente se determinó que el tiempo de permanencia de los datos en el almacén es de 10 años. Esto es posible gracias a que los datos almacenados en la base de datos transaccional durante los años de operación de la empresa no representan una cantidad considerable de espacio de almacenamiento y tomando en cuenta que los datos almacenados en dicho repositorio están sumarizados, se reduce aún más el tamaño en el mismo lo que puede ser soportado para la infraestructura de la empresa mantener esa cantidad de datos durante esos 10 años.

El diseño del almacén de datos se muestra en la siguiente figura:

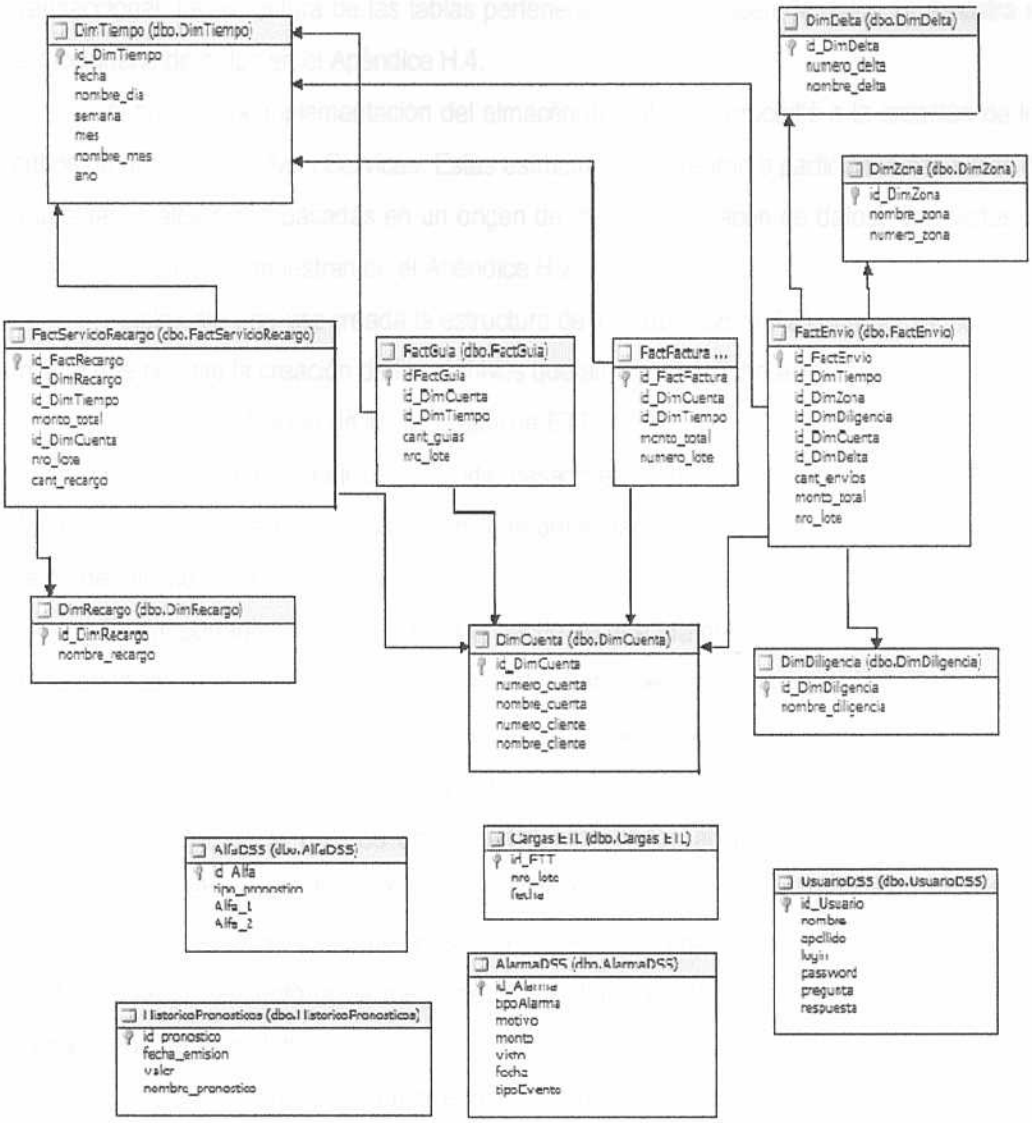

Figura 11. Diseño del almacén de datos.

Fuente: Elaboración Propia

IV.1.4.3 Implementación del almacén de datos

En base al diseño creado del almacén de datos se procedió a la implementación del mismo, usando como herramienta SQL Server 2005. Una vez identificadas las tablas de hechos y dimensiones y sus respectivos atributos, se implementaron mediante la herramienta gráfica de SQL Server, configurando a su vez las relaciones existentes entre las tablas de

hechos y sus tablas de dimensión. En la implementación sólo se crearon las tablas y sus relaciones, sin cargar el almacén con los datos provenientes de la base de datos transaccional. La estructura de las tablas pertenecientes al almacén de datos se muestra en el diccionario de datos en el Apéndice H.4.

Al finalizar la implementación del almacén de datos se procedió a la creación de los cubos de datos en Analysis Services. Estas estructuras se crearon a partir de varias vistas de datos, las cuales están basadas en un origen de datos (el almacén de datos). Las vistas de orígenes de datos se muestran en el Apéndice H.2.

Finalmente una vez creada la estructura de los cubos se realizó el procesamiento del mismo que permite la creación de los archivos que almacenan dicha estructura.

IV.1.4.4 Definición de los procesos de ETT

En esta etapa se realizó un estudio basado en el análisis anterior con la finalidad de definir los procesos de ETT en base al diseño del almacén de datos, los cuales permitirán la carga del mismo con los datos que requiere.

Fue importante tomar en cuenta el hecho de que era necesario definir una estructura que permitiese la actualización automática de los datos de los cubos cada cierto tiempo. Para ello se necesita la creación de un proceso automatizado que ejecute estos procesos ETT y que deberá estar programado para ejecutarse cada cierto tiempo. Por ello se realizó una agenda de carga, en la que los datos serían procesados al almacén y que a su vez no interfiriera con el trabajo diario de la organización ya que consume recursos del sistema. Adicionalmente es importante que no se estén ejecutando operaciones en la base de datos transaccional al momento de la carga de dicho almacén debido a que podría generar una carga errada de esos datos al mismo.

También es importante tomar en cuenta que una vez realizada la primera carga de datos se debe crear un procedimiento que permita cargar los datos nuevos de la base de datos transaccional al cubo. Para ello se concluyó que sería necesario realizar dicha carga por lotes. Cada lote de datos posee un número de lote el cual garantiza que los datos que se almacenen en dichos cubos sean datos nuevos.

Es necesario crear una tabla en el almacén de datos que permita registrar la fecha de la última carga y el número de lote correspondiente al conjunto de datos cargados en esa fecha para poder realizar el filtro de los mismos para realizar la carga de datos nuevos.

Los datos extraídos serán procesados y sumariados con un nivel de granularidad diario para permitirle al usuario tener un nivel de detalle más amplio a la hora de analizar los datos obtenidos.

Finalmente es necesaria la creación de dos paquetes, uno para realizar las operaciones ETT de las tablas de dimensiones y otro para las tablas de hechos. Estos paquetes contendrán componentes de flujo de control que permiten el proceso de flujo de datos desde el origen de datos (la base de datos transaccional) hacia un destino dado (el almacén de datos). Cada uno de estos componentes extraería los datos mediante una consulta realizada a la base de datos transaccional la cual depende de la tabla del almacén.

IV.1.4.5 Implementación de los procesos ETT

Una vez realizada la definición de los procesos ETT se procede a la implementación de los mismos usando como herramienta de desarrollo el módulo de Integration Services de SQL Server 2005, el cual posee herramientas que permiten la creación de paquetes los cuales encapsulan todo el procesamiento ETT de los datos al almacén.

Como parte inicial de la implementación de los procesos ETT se crearon dos paquetes, que encapsulan todas las tareas que permiten realizar estos procesos. Uno de estos paquetes contiene las tareas de flujos de datos para la carga de los datos hacia las tablas de dimensión, mientras que el otro paquete encapsula las tareas que permiten la carga de datos hacia las tablas de hechos. Las tareas de flujos de datos encapsulan el motor de flujo de datos que mueve datos de orígenes a destinos, y proporciona recursos para transformar, limpiar y modificar los datos a medida que se mueven.

En el paquete dimensiones, se implementaron seis tareas de flujos de datos, cada una de estas se encarga de realizar los procesos ETT desde la base de datos transaccional hacia cada una de las tablas de dimensión respectivamente. Por ejemplo, la tarea de extracción DimCuenta encapsula un conjunto de procesos que son: los "Orígenes de Datos", que extraen la data desde la fuente de datos que se le indica, las "Transformaciones", que son diversos procesos con distinta funcionalidades que permiten la limpieza, transformación y modificación de los datos extraídos por el módulo de origen de datos y por último el "Destino de Datos" que proporciona herramientas de conexión para la inserción de los datos ya procesados en un almacén de datos. Todos estos procesos se unen en forma de flujo de procesos para permitir un flujo continuo de datos desde el origen hasta el destino, tal como se muestra en la figura 12.

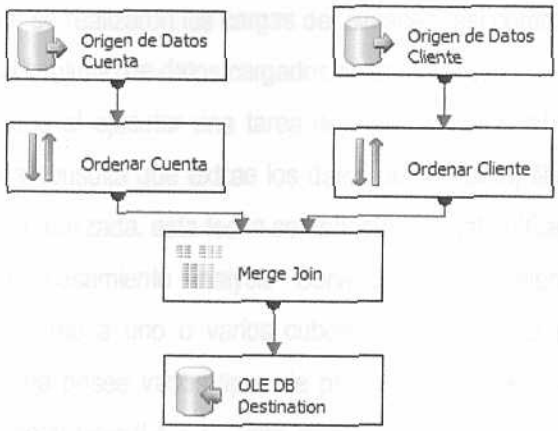

Figura 12. Flujo de datos para la Tarea de Extracción de la dimensión Cuenta.

Fuente: Elaboración Propia.

Las tareas de extracción del paquete de dimensiones se muestran en el Apéndice I.1.

En la implementación del paquete de tabla de hechos se crearon siete tareas de flujos de datos. La siguiente figura muestra un flujo de datos para la tarea de extracción de la tabla de hechos FactFactura.

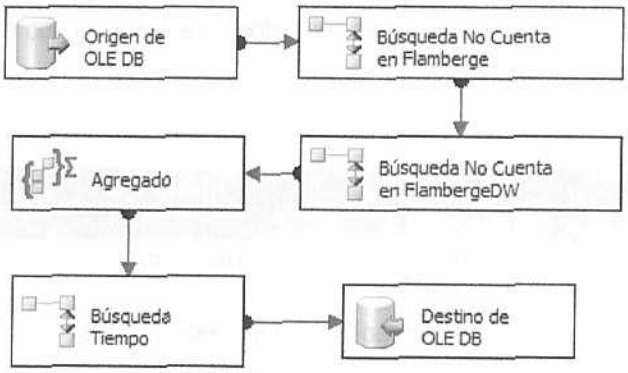

Figura 13. Flujo de datos de la tarea Extracción de datos para FactFactura

Fuente: Elaboración Propia.

Las tareas de extracción del paquete de tablas de hechos se muestran en el Apéndice I.2.

Adicionalmente a estas siete tareas de flujos de datos del paquete de tablas de hechos, se implementaron dos tareas: Tarea Ejecutar SQL y Tarea de Procesamiento de Analysis Services.

La Tarea de Ejecución de comandos SQL, "Actualizar CargaETL", ejecuta un comando "Insert" en la tabla Carga ETL del almacén de datos. Esta tabla mantiene un registro

de las fechas en la que se realizaron las cargas del almacén, así como también el número de lote correspondiente al conjunto de datos cargados en ese proceso.

Esto permite que al ejecutar una tarea de flujo de datos, el módulo de Origen de datos (que ejecuta una consulta que extrae los datos de la fuente) filtre dichos datos por la fecha de la última carga realizada, esta fecha se obtiene de la tabla "CargaETT".

La Tarea de Procesamiento Analysis Services: "Procesamiento Incremental de los Cubos", permite conectarse a uno o varios cubos para realizar el procesamiento de los mismos. Aunque la tarea posee varios tipos de procesamiento de cubos se implementó el procesamiento incremental (Ver II.4.3.3). Para esto, una vez realizada la carga de datos al almacén, la tarea ejecuta una consulta en la que extrae los datos filtrados por número de lote.

IV.1.4.6 Elaboración de las pruebas ETT

Para comprobar que la ejecución de las tareas de ETT funcionan de manera correcta se realizaron pruebas que consistían en ejecutar una consulta en la base de datos que permitiese ver la sumarización de los datos tal como debería ingresarse una vez ejecutadas las tareas. Estas consultas además permitieron verificar el número de tuplas que deberán ser extraídas, y el número de tuplas totales luego de hacer la sumarización. A continuación se muestra un ejemplo con los resultados obtenidos luego de ejecutar los procesos ETT para el almacén de datos referente a la carga de la tabla de hechos Guías:

Procesos ETT	Estado	Tiempo De Ejecución	Cantidad Agregada
Carga de guías	**Completado**	**0,2 segundos**	**16541 filas**

Tabla 11. Resultados de los procesos ETT.

Fuente: Elaboración Propia.

Una vez obtenidos los resultados se comprobó que el número de filas agregadas correspondía con la consulta realizada en la base de datos transaccional y que al hacer una consulta al almacén de datos sobre una cantidad específica por ejemplo "Monto Total" correspondiese con la consulta realizada a la bases de datos transaccional usando los mismos criterios. De esta manera se corroboró que los datos se procesaron correctamente al almacén de datos y que los procesos ETT se ejecutaban correctamente. Los resultados de los procesos ETT se muestran en el Apéndice I.3 y los mensajes de pruebas en el Apéndice I.4.

Luego de completar la implementación de los procesos ETT se procedió a programar la ejecución automática de los paquetes. Esto se realizó a través del servicio Agente de SQL Server en el cual se selecciona el paquete que se desea programar especificando el tipo de programación. Para este caso se determinó el tipo de programación "Periódica" ya que ésta permite que una programación se ejecute varias veces indicando la frecuencia.

Esta agenda comprende la carga de los datos los días sábados de cada semana a las 8:00 de la mañana, ya que permite la carga sin interferir con las operaciones de la empresa. Finalmente luego de realizar las pruebas de los procesos ETT se procedió a cargar los datos en el almacén a través de la ejecución de los paquetes definidos anteriormente.

IV.1.5 DISEÑO DETALLADO DE LA APLICACIÓN DE PREDICCIONES Y REPORTES

En esta fase de diseño se determinó como se iba a construir el sistema, para ello se utilizó como herramienta de modelado UML (*Unified Modeling Language*) o lenguaje unificado de modelado, que permite modelar, construir y documentar los elementos que forman un sistema de software orientado a objetos. Entre los diagramas de UML se decidió realizar los diagramas de actividades, secuencia y clases del sistema.

Primero se crearon los diagramas de actividades de cada especificación de casos de usos definidos en la fase de análisis, estos diagramas facilitan la comprensión del orden en que se realizan las diversas actividades para llevar a cabo un proceso del sistema. Los diagramas de actividades se muestran en el Apéndice J.1.

Para la realización del diagrama de clases se aplicó la técnica de las tarjetas CRC (*Class Responsibility Collaboration*). Estas tarjetas facilitan la creación del diagrama de clases, ya que sirven para identificar las clases del sistema, sus responsabilidades y sus colaboraciones. Las tarjetas CRC del sistema de reportes y predicciones se muestran en el Apéndice J.2.

A partir de las tarjetas se construyó el diagrama de clases, estas tarjetas facilitan la creación de las clases debido a que cada tarjeta representa una clase del sistema, las responsabilidades representan los métodos que deben ser codificados en cada clase, y las colaboraciones representan las relaciones que existen entre ellas. El diagrama de clases del sistema se muestra en el Apéndice J.3.

El diagrama de secuencia es uno de los diagramas más efectivos para modelar la interacción entre los objetos de un sistema, estos diagramas deben ser modelados para cada

caso de uso, por ello cada diagrama de secuencia fue creado en base a las especificaciones de casos de uso definidas en la fase de análisis. Los diagramas de secuencias contienen detalles de implementación que incluyen los objetos y clases que se usan para implementar cada escenario y los mensajes pasados entre los objetos. Los diagramas de secuencias se muestran en el Apéndice J.4.

Seguidamente se realizó el diagrama de navegación del sistema de predicciones y reportes. Este diagrama se muestra en el Apéndice J.5.

Luego de realizar los diagramas UML se procedió a diseñar la estructura de los reportes OLAP y al diseño de las interfaces del sistema de predicciones y reportes. Para la realización de los diseños de las ventanas del sistema se siguieron los siguientes estándares:

- Las pantallas del sistema tienen un menú principal, el cual permanece visible en todas las ventanas del mismo, esto permite la fácil navegación por el sistema.
- El nombre del sistema junto con el nombre de la ventana aparece en la parte superior izquierda de la pantalla.
- El nombre de la funcionalidad de la ventana se muestra como título en el área de datos.
- Los botones del sistema tienen un símbolo según la función que representan.
- Las letras del sistema son de tamaño 14, Century Gothic color azul rey.
- El fondo de las ventanas es de color Lavanda.
- El logo del sistema se muestra como fondo en la ventana principal.

IV.1.6 IMPLEMENTACION DE LA APLICACIÓN DE PREDICCIONES Y REPORTES

El sistema se elaboró con la herramienta de desarrollo Microsoft Visual Studio 2005 .NET y con el lenguaje de programación C# seleccionados en la fase de análisis del sistema. Para realizar los gráficos de reportes se utilizó la librería ChartFX escrita en C#, y la conexión a los cubos OLAP por medio de la librería ADOMD.NET que permite la comunicación de componentes del Framework de Microsoft.NET con el módulo de Analysis Services de SQL Server 2005.

Para comenzar la implementación del sistema se utilizaron los diagramas creados en la fase de diseño. Para la implementación de la interfaz de usuario se utilizaron los estándares definidos previamente.

En el diagrama de clases se pudo observar que las clases de reportes no se relacionan directamente con las clases de predicciones, aunque utilizan clases en común como la clase

de conexión, por ello se separó la codificación del sistema de predicciones y reportes en dos módulos, el módulo de predicciones y el módulo de reportes.

En la implementación de ambos módulos se utilizó el patrón de diseño MVC (*Model View Controller*), este patrón indica la manera adecuada en la que se deben separar las clases del sistema, Por lo tanto se implementaron para cada clase del sistema, su respectiva clase control y su respectiva clase modelo. Además de estas clases es necesaria la creación de las clases que forman parte de la vista del diseño MVC. Estas clases están representadas en el sistema por las ventanas o formularios.

En las clases que forman parte del controlador se codificaron las funciones que permiten realizar las operaciones del sistema, según el patrón MVC las clases que forman parte del controlador no deben contener conexiones a base de datos por ello es necesario crear una clase modelo para cada controlador. En estas clases se codificaron los querys que los controladores necesitan. La clase modelo de cada controlador se comunica con una clase llamada ConexionData que proporciona la comunicación con el almacén de datos para que éstas puedan ejecutar los querys, la clase de conexión es quien se encarga de administrar dichas conexiones.

Cuando un controlador necesita datos para trabajar, se comunica con su modelo respectivo, el modelo se comunica con la clase de conexión para ejecutar el query, seguidamente la clase modelo guarda los datos en su clase filtro correspondiente y devuelve esta clase como objeto a su respectiva clase controlador.

IV.1.6.1 Módulo de pronósticos

El módulo de pronósticos está dividido en tres funcionalidades principales, la primera es la generación de pronósticos a través del método de suavización exponencial, la segunda funcionalidad es la generación de pronósticos a través de la técnica de regresión lineal y la tercera funcionalidad está relacionada con la creación de alarmas.

Para implementar la técnica de regresión lineal se utilizó la siguiente fórmula:

$$\hat{Y} = b_0 + b_1 x.$$

Las variables bo y b_1 se calcularon mediante las formulas de mínimos cuadrados que se utilizan en el cálculo de regresión.

Se creó una clase control de técnicas estadísticas donde se implementaron las funciones necesarias para el cálculo del pronóstico. Para obtener los valores de "*y*" correspondientes a un tiempo especifico fue necesario construir un query que trajera los valores ya sea de pedidos o montos facturados durante un rango de tiempo, el tiempo

escogido para el cálculo del pronóstico fue de un año, entonces cada vez que se calcula la regresión lineal es necesario ejecutar este query para conseguir los valores correspondientes y luego asociarlo a la variable "*x*". Luego de codificar estos métodos se procedió a crear las funciones de las sumatorias correspondientes a las formulas de *bo* y b_1. Una vez codificadas estas funciones, se creó la función encargada de sustituir esos valores en la fórmula de regresión.

Durante las pruebas realizadas a los resultados obtenidos en la técnica de regresión lineal en la fase de análisis se observó que si los valores correspondientes a la variable se ordenaban en el tiempo acumulando los valores del mes anterior con el siguiente y así sucesivamente por cada uno de los valores de "*y*", se obtenía un mejor ajuste lineal, por ello fue necesario crear una función que se encargara de calcular los valores acumulados para cada valor de "*y*", esta función recibe los valores de "*y*" del query, los ordena y calcula los valores acumulados, estos valores acumulados son los que se utilizan para realizar las sumatorias correspondientes a las formulas de *bo* y b_1.

Luego se implementó la técnica de suavización exponencial que viene definida por la siguiente fórmula:

$$y^*_{t+1} = \alpha\, y_t + (1-\alpha)\, y^*_t$$

Para implementar esta técnica es necesario obtener los valores de alfa correspondientes a cada tipo de pronóstico (Ver II.7.3), como cada pronóstico tiene valores de alfas distintos, es necesario almacenar estos valores en la base de datos para poderlos consultar antes de implementar este método. Se crearon funciones en la clase modelo correspondiente a las predicciones que se encargan de ejecutar los querys para obtener los valores reales del mes anterior, el último pronóstico calculado y los valores de alfas correspondientes a cada tipo de pronóstico para sustituirlos en la formula y obtener el valor para y^*_{t+1}.

Como la técnica de suavización exponencial se basa en el pronóstico generado en el mes anterior para poder hallar el siguiente, fue necesario crear una función que permita la creación de pronósticos automáticos de manera que si el usuario no genera el pronóstico manualmente en un mes determinado el sistema lo ejecute automáticamente

La técnica de suavización exponencial supone la continuidad en el futuro de algún patrón histórico, por ello hay que buscar la manera de que si el patrón histórico cambia los valores de alfas se ajusten al nuevo patrón, para disminuir el valor del error (Ver II.7.3).

Por la razón antes mencionada fue necesario crear una función que realice el ajuste de alfa para cada tipo de pronóstico. Esta función calcula los pronósticos utilizando todas las combinaciones de alfa, luego calcula el error DAM para cada conjunto de valores y selecciona el menor error. De esta manera se consigue un ajuste de los valores alfa en caso de que la tendencia histórica cambie a través del tiempo.

Seguidamente se implementó la parte del sistema relacionada con las alarmas, estas alarmas indican al usuario cuando un valor obtenido está fuera del rango, se codificó una función que se encarga de verificar los montos que se obtienen en cada servicio, si la función encuentra algún valor fuera del rango muestra un mensaje indicando el motivo de la generación de la alarma. La explicación de las clases y métodos creados en la aplicación de predicciones y reportes se muestran en el manual del sistema que se encuentra en el Apéndice L.

En la siguiente figura se muestra una ventana del modulo de pronósticos, las demás ventanas se muestran en el manual de usuario en el Apéndice K.

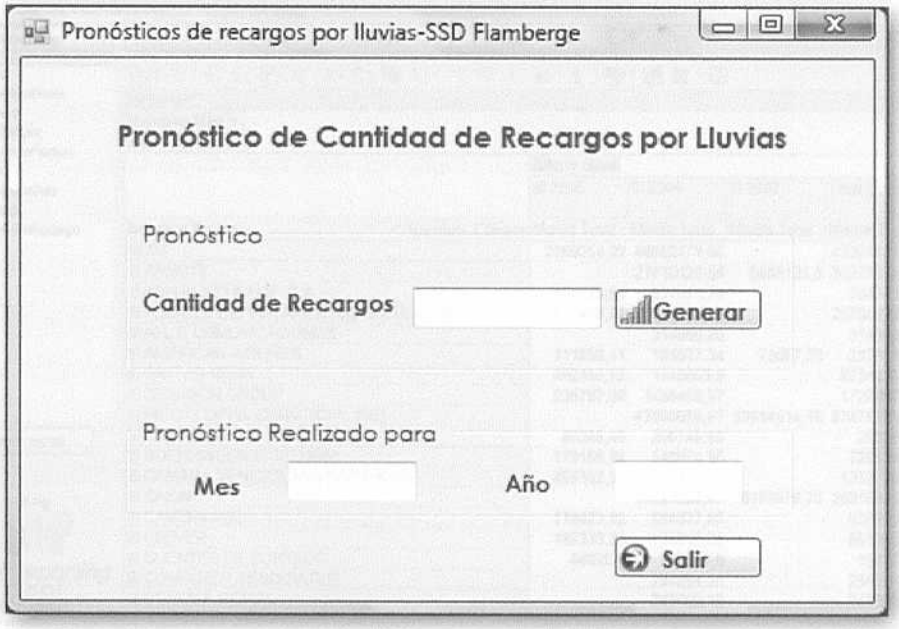

Figura 14. Ventana de pronóstico de recargos de lluvias.

Fuente: Elaboración Propia.

IV.1.6.2 Módulo de reportes

El módulo de reportes del sistema se dividió en dos partes, las cuales comprenden la de Reportes y Reportes Gráficos. Reportes permite al usuario final crear reportes en forma

de tablas dinámicas en la que se podrá modificar la estructura de dichas tablas en tiempo de ejecución. Reportes Gráficos permite al usuario construir un reporte y visualizar el resultado mediante un gráfico.

IV.1.6.2.1 Reportes

Para implementar la funcionalidad de Reportes se usó la herramienta OLAP Reporting Tool for Excel que permite la creación de tablas dinámicas de datos conectados a un cubo. Esta herramienta se parametriza desde la aplicación con los datos necesarios para la conexión a los cubos de SQL Server 2005.

Esta herramienta le permite al usuario consultar reportes predefinidos en el sistema así como también poder crear reportes nuevos en base a sus necesidades. Adicionalmente proporciona herramientas de análisis y filtros predeterminados con los que el usuario podrá visualizar reportes con rangos determinados como por ejemplo visualizar los 10 datos con valores más altos o más bajos, aumentando así la capacidad de análisis de los mismos.

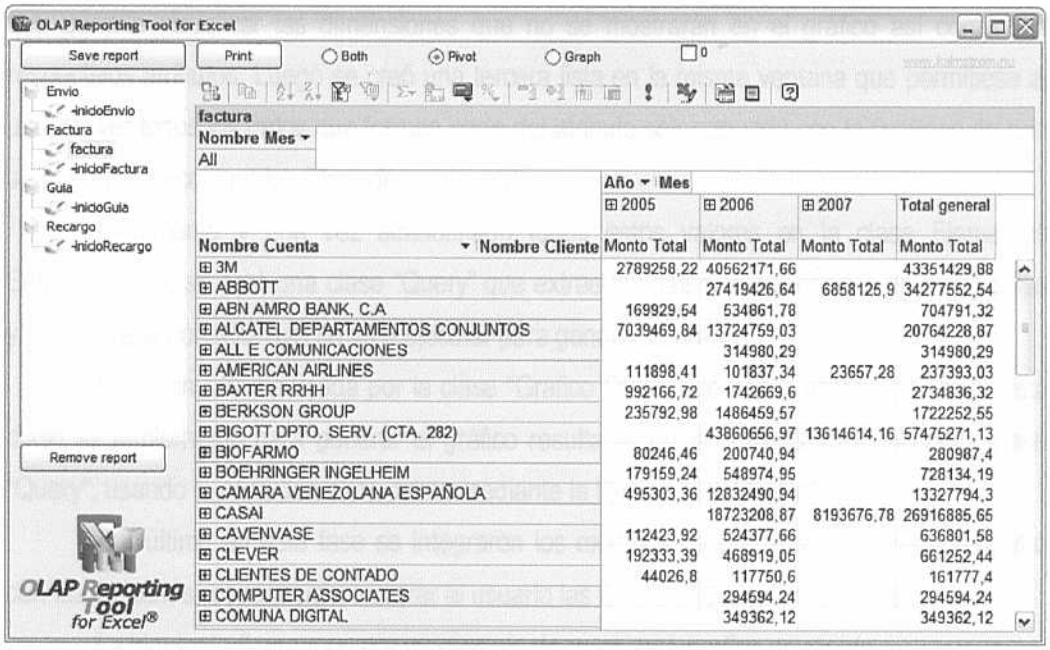

Figura 15. Ventana de pronóstico de recargos de lluvias.

Fuente: Elaboración Propia.

IV.1.6.2.2 Reportes Gráficos

Esta implementación proporciona al usuario la creación de reportes y la visualización de los datos mediante una gráfica.

En primer lugar se implementaron una serie de formularios que permiten al usuario seleccionar los atributos necesarios para la construcción de la consulta gráfica.

Primeramente se desarrolló un formulario que obtiene las dimensiones que se encuentran en el cubo seleccionado por el usuario en la ventana principal, para que éste pueda seleccionar cuáles de estas dimensiones se mostrarán en el gráfico. Esto se hace mediante una petición a la clase de conexión, la cual usa la librería ADOMD.Net que permite conectarse al módulo de Análisis Services de SQL Server, luego, usando las funciones de ADOMD.net se obtienen las dimensiones que forman parte de dicho cubo.

Luego se implementó un segundo formulario que muestra al usuario los atributos de las dimensiones seleccionadas en el formulario anterior, esto con la finalidad de obtener por cual atributo se mostrarán los datos de la gráfica.

Una vez que se obtienen los atributos seleccionados, el usuario tiene la opción de seleccionar datos de atributos que no formarán parte de la consulta gráfica para filtrar dicha consulta por los valores de esos atributos. Esto se logró implementando un formulario final, que permite visualizar las dimensiones que no se mostrarán en el gráfico así como sus respectivos atributos. Luego se creó una tercera lista en la misma ventana que permitiese al usuario ver todos los datos que forman parte del atributo seleccionado con la finalidad de que se puedan seleccionar los filtros de la consulta.

Finalmente y una vez almacenado todos estos valores en la clase Elementos Seleccionados, se creó una clase "Query" que extrae los datos seleccionados de dicha clase y construye la consulta que se va a ejecutar para generar el gráfico.

Esta consulta es usada por la clase "Grafico Olap", que usa la librería ChartFX. Esta clase se implementó para generar el gráfico resultante de la consulta creada por la clase "Query", usando la conexión a los cubos mediante la librería ADOMD.Net.

Por último en esta fase se integraron los módulos de predicciones y reportes junto con las ventanas creadas para mostrar al usuario las funcionalidades del sistema.

La siguiente figura muestra un ejemplo de un reporte grafico generado.

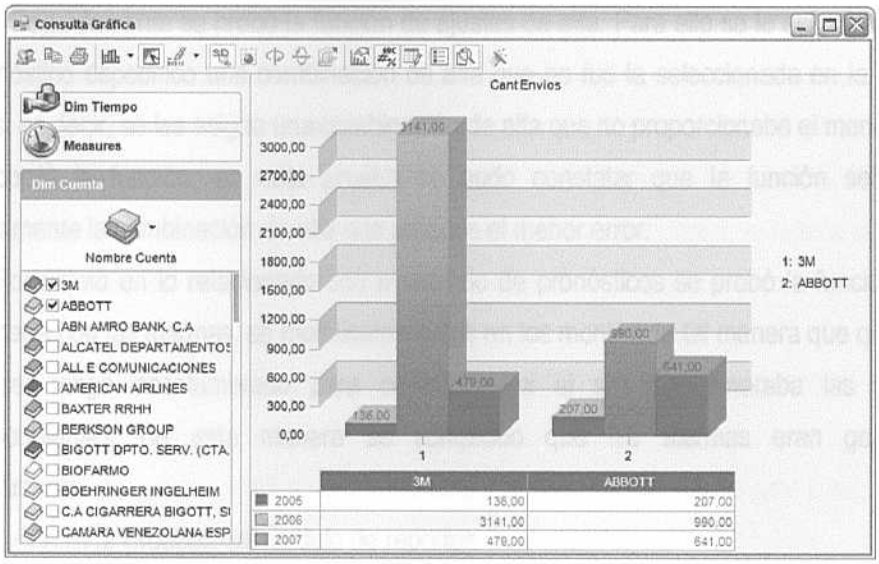

Figura 16. Ventana del modulo de reportes gráficos.

Fuente: Elaboración Propia.

IV.1.7 PRUEBAS DE LA APLICACIÓN DE PREDICCIONES Y REPORTES

En esta fase se realizaron las pruebas del producto final. Se corrieron varios casos de pruebas por cada caso de uso, de manera de verificar el buen funcionamiento de la aplicación y se realizaron las pruebas con el usuario. Estas pruebas se dividieron en tres partes: pruebas al modulo de predicciones, pruebas al modulo de reportes, pruebas del sistema con los usuarios.

IV.1.7.1 Prueba del modulo de pronósticos

En esta actividad se realizaron diversas pruebas para verificar si los resultados que se obtenían por el sistema referentes a los pronósticos eran correctos. Para ello se tomaron los resultados que se obtuvieron al realizar los pronósticos en la actividad de evaluación de técnicas estadísticas perteneciente a la fase de análisis, se tomaron estos pronósticos y se compararon con los pronósticos que se generaron en el sistema. Se tomaron los mismos conjuntos de datos a los que se les realizó los pronósticos en Excel y se procedió a generar los pronósticos en el sistema de predicciones y reportes sobre estos mismos conjuntos; de esta manera se pudo constatar que los pronósticos generados por el sistema eran correctos. Luego se probó la función de generación automática de pronósticos, se eliminaron del sistema los pronósticos generados en la prueba anterior para que la función de generación automática se ejecutara al iniciar sesión y generara las predicciones, seguidamente se verificó que estos pronósticos eran correctos y que cada mes tenía un pronóstico asignado.

Seguidamente se probó la función de ajustes de alfa. Para ello se le asignó a un tipo de pronóstico específico una combinación de alfa que no fue la seleccionada en la fase de análisis, es decir, se les asigno una combinación de alfa que no proporcionaba el menor error, y se corrió la función, en esta prueba se pudo constatar que la función selecciona correctamente la combinación de alfa que produce el menor error.

Por último en lo relacionado con el modulo de pronósticos se probó la funcionalidad relacionada con las alarmas, se modificaron datos en los montos de tal manera que quedaran fuera del rango acostumbrado para comprobar si el sistema generaba las alarmas correspondientes, de esta manera se comprobó que las alarmas eran generadas correctamente.

IV.1.7.2 Pruebas del modulo de reportes

Para verificar si el módulo de reportes funciona correctamente se hicieron las siguientes pruebas.

- Verificación de cantidades

Tanto para las tablas como para los gráficos se verificaron los montos resultantes de una consulta en especifico, usando como base los datos extraídos directamente de los cubos mediante el uso de Expresiones Multidimensionales (MDX) que es el lenguaje de consulta que se utiliza para trabajar con datos multidimensionales y para recuperarlos en SQL Server 2005 Analysis Services.

Estas consultas permiten evaluar dos aspectos, los datos individuales y la suma de los datos. Para evaluar datos individuales las consultas se basan en extraer los datos correspondientes a un elemento del cubo como por ejemplo cantidades de envíos.

De igual forma se hizo para comprobar las sumas de los datos usando consultas MDX donde se muestran los datos correspondientes a un período mayor (Semana, Mes, Año).

Estos datos se verificaron generando consultas usando las herramientas de reporte del sistema y se comprobó que las cantidades coincidieron con las que fueron extraídas directamente de los cubos, lo que permitió evaluar el correcto funcionamiento del Módulo de Reportes del Sistema.

- Verificación de los atributos de las dimensiones

Adicionalmente se verificaron si los atributos de las dimensiones eran correctos. Esto se validó verificando los atributos directamente en el diseño de los cubos y los atributos que se mostraban en las consultas. Con lo cual se pudo evaluar que los atributos se mostraban correctamente en las consultas realizadas en el Módulo de Reportes del Sistema.

IV.1.7.3 Pruebas con los usuarios del sistema

Estas pruebas consistieron en presentar el sistema al usuario de manera que éste pudiera interactuar con él y dar su opinión con respecto a la funcionalidad y facilidad de uso. Por ser la facilidad y la amigabilidad requerimientos no funcionales del sistema, fueron los aspectos esenciales tomados en cuenta durante esta evaluación y como resultado se obtuvo que se cumplía con los objetivos por lo cual no fue necesario realizar cambios al sistema.

IV.2 DESARROLLO DEL SISTEMA WEB DE SEGUIMIENTO Y PROCESAMIENTO DE PEDIDOS

El modelo que se utilizó como metodología para la realización de este sistema fue el modelo de cascada. Este modelo se muestra en el Apéndice M.

IV.2.1 CONCEPTO DE SOFTWARE

El concepto de software representa la primera fase del modelo utilizado para la realización del sistema transaccional. En esta fase se realizaron entrevistas con los usuarios para determinar los requerimientos funcionales y no funcionales del sistema. Como requerimiento no funcional se encontró que el sistema transaccional debería ser fácil de usar para simplificar a los clientes de la empresa la realización de los pedidos a través de internet. La lista de casos de uso asociados a los requerimientos funcionales se muestra a continuación:

Número	Nombre del caso de uso	Descripción
1	Ingresar cliente	Permite al usuario ingresar un nuevo cliente.
2	Buscar guía	Permite al usuario buscar una guía en el sistema.
3	Ingresar centro	Permite al usuario ingresar un nuevo centro de costo asociado a un cliente.
4	Ingresar datos de sesión	Permite al usuario ingresar los datos relacionados con la sesión.
5	Ingresar guía	Permite al usuario ingresar los datos de una guía para realizar un pedido.
6	Consulta especifica de guías	Permite al usuario observar los datos de una guía.
8	Iniciar sesión	Permite al usuario ingresar al sistema.
9	Cambiar password por olvido	Permite al usuario cambiar su password cuando no lo recuerda.
10	Modificar cliente	Permite al usuario modificar los datos de un cliente.
11	Modificar centro	Permite al usuario modificar los datos de un centro de costo.

Tabla 12. Lista de casos de uso.

Fuente: Elaboración Propia.

Durante este proceso de levantamiento de información se determinó que por requisito de la empresa el sistema Web se use dentro de la red LAN de la misma y que los clientes se conectarán a dicha red mediante una conexión VPN ya existente en la organización.

IV.2.2 ANALISIS DE LOS REQUERIMIENTOS

En esta fase se procedió a realizar el análisis de los requerimientos obtenidos en la fase de concepto de software. Con la lista de casos de uso se procedió a realizar el modelo y especificaciones de los casos de uso. En el Apéndice N.2 se muestra el modelo de casos de uso y en el Apéndice N.3 se muestran las especificaciones de casos de uso del sistema.

IV.2.3 DISEÑO GLOBAL

El sistema transaccional está formado por dos grandes estructuras, estas son el sistema web que permite el procesamiento y seguimiento de pedidos y la base de datos transaccional en la cual se almacenan todas las operaciones diarias de la empresa.

Las páginas web se crearon bajo el Framework Java Server Faces (JSF) el cual permite la implementación de aplicaciones web mediante el uso del lenguaje de programación JSP. Estas incluirán todas las funciones que fueron determinadas como necesarias para la construcción de este sistema web en la etapa de análisis de requerimientos. Adicionalmente se implementó la base de datos transaccional que permitirá almacenar las transacciones de la empresa. Esta base de datos se creó en SQL Server 2005.

IV.2.4 DISEÑO DETALLADO

En esta fase se procedió a crear el modelo de datos relacional y los diagramas del sistema necesarios para determinar la manera de cómo se va a construir el mismo. Se utilizó UML como herramienta de modelado. El modelo de la base de datos transaccional se muestra en el Apéndice G. En esta fase se creó el diagrama de clases que se muestran en el Apéndice N.4.

En esta etapa se establecieron los estándares en cuanto al diseño de la interfaz del sistema web. Estos estándares quedaron definidos de la siguiente manera:

- El tamaño de las páginas web está establecido en una resolución de 800-600, esto es para que independientemente el tamaño de resolución de las pantallas los usuarios puedan visualizar completamente la funcionalidad del sistema.

- Todas las páginas tienen una franja superior e inferior de color azul.
- El nombre de la empresa se encuentra en la parte superior derecha y el nombre del usuario que inicio sesión en la parte superior izquierda.
- Todas las páginas tienen un nombre de acuerdo a su funcionalidad.
- El tipo de letra seleccionado para las páginas es Century Gothic, tamaño 12, color azul rey.
- Los enlaces son de tamaño 14 de color azul celeste.

IV.2.4 IMPLEMENTACIÓN

En esta fase se realizó la implementación del sistema de seguimiento y procesamiento de pedidos. Este sistema se elaboró con la herramienta de desarrollo BEA Worshop que facilita la creación de sistemas Web a través de Java Server Faces, este permite el desarrollo de un sistema Web basado en el patrón de diseño MVC (*Model View Controller*). Este patrón se explica con las detalle en la fase de implementación del sistema de soporte para la toma de decisiones. Además se utilizó SQL Server 2005 para la implementación del modelo de datos y Hibernate que es una herramienta de mapeo objeto-relacional creado para la plataforma Java que facilita el mapeo de atributos entre una base de datos relacional y el modelo de objetos de una aplicación mediante archivos XML que permiten establecer estas relaciones. Como servidor Web se utilizó Tomcat, que es un servidor con soporte de servlets y JSF.

Para la codificación de este sistema se procedió en primer lugar a implementar el modelo de datos, luego se crearon las clases del mismo basadas en los diagramas realizados en la fase de diseño. Las clases creadas en el sistema fueron:

Clases DAO: se realizaron por cada tabla de la base de datos transaccional, estas clases acceden a los datos a través de otras clases Java definidas como interfaz que permiten la comunicación entre las clases DAO y Hibernate.

Clases Action: Las clases action del sistema contienen las reglas del negocio de la aplicación, se creó una clase action por cada uno de los procesos del sistema. En estas clases se definieron los métodos y procedimientos necesarios para llevar a cabo las funcionalidades del mismo. Estas clases se comunican con las clases DAO para obtener los datos y poder enviarlos a los action que se encargan de mostrarlos por pantalla a través de la interfaz grafica.

Luego de implementar las clases del sistema se procedió a crear las páginas Web que permitirán la interacción con el usuario. Las páginas del sistema se muestran en el Apéndice N.5.

IV.2.5 PRUEBAS DEL SISTEMA TRANSACCIONAL

En esta fase se procedió a realizar las pruebas de la aplicación y las pruebas con los usuarios finales. En primer lugar se crearon los casos de prueba para ser aplicados al sistema, estos casos de prueba consistían en verificar si las funciones retornaban los valores adecuados. Además se realizó la interacción entre el usuario y el sistema transaccional en donde se observaron las actividades hechas por los usuarios para acceder al sistema y manipularlo, para luego analizar la experiencia que estos tuvieron al utilizar el sistema y si las respuestas del mismo eran las adecuadas.

Como resultado de estas pruebas su obtuvo que el sistema cumple con los requerimientos de la empresa además que su funcionamiento fue el esperado por lo que no fue necesario realizar cambios en el mismo.

CAPÍTULO V. RESULTADOS

Los resultados obtenidos durante el desarrollo de este Trabajo Especial de Grado se muestran a continuación:

1. Con los resultados obtenidos del análisis de requerimientos se pudo determinar cuáles fueron las características mínimas que debía tener el manejador de base de datos a seleccionar.

2. A partir del resultado anterior se buscó y seleccionó la herramienta que cumpliese con estos requerimientos mínimos, que permitiera la implementación de modelos estadísticos y que mejor se adaptara a las necesidades del proyecto.

3. Adicionalmente se encontró que sólo existía una fuente de datos para el sistema de soporte de toma de decisiones (la base de datos operacional de la empresa Flamberge) que almacena toda la información proveniente del sistema transaccional.

4. Como resultado de la búsqueda de un sistema que se adaptara a las especificaciones del proyecto se seleccionó SQL Server 2005 como herramienta de desarrollo del almacén de datos pues posee funciones que permiten el manejo del mismo, así como también de servicios para realizar los procesos de extracción, transformación y transporte de los datos.

5. Se diseño e implementó el almacén de datos de acuerdo a las necesidades del proyecto usando la herramienta SQL Server 2005.

6. Se determinó que no fue necesario modificar la base de datos transaccional.

7. Se seleccionaron los indicadores de gestión del sistema que se requieren para satisfacer las necesidades de análisis de los datos.

8. A partir del estudio realizado en cuanto a los modelos de pronósticos existentes se pudieron seleccionar los modelos estadísticos que más se ajustaban a los datos históricos de las transacciones de la empresa.

9. Al analizar los datos históricos de las transacciones en la base de datos operacional se pudo observar que los mismos son muy variables y no presentan una tendencia significativa.

10. Se desarrolló el sistema de reportes y pronósticos, el cual permite realizar análisis de los datos en cualquier departamento de la empresa, aumentando la calidad y precisión de los procesos de toma de decisiones y que permite la planificación y/o prevención de situaciones futuras.

11. Se desarrollo un sistema transaccional Web que incluye la implementación del modelo de datos, las interfaces diseñadas y las funcionalidades relacionadas con el seguimiento y procesamiento de pedidos.

CONCLUSIONES

A partir de los resultados obtenidos durante la elaboración del presente Trabajo Especial de Grado se puede afirmar lo siguiente:

1. Las herramientas analíticas en línea OLAP son parte importante en las organizaciones pues éstas permiten evaluar y monitorear la situación real de una empresa desde una perspectiva que facilita la exploración de los diversos escenarios de una manera sencilla. Adicionalmente proporciona una visión más profunda de todos los niveles de una organización.
2. Los modelos multidimensionales permiten ver las medidas del negocio desde diferentes perspectivas y además permite identificar fácilmente patrones y tendencias.
3. Los sistemas de soporte a la toma de decisiones son herramientas indispensables para los negocios pues la toma de decisiones es un factor crítico para la supervivencia de las organizaciones, debido a que las decisiones correctas son la clave del éxito empresarial. Debido a que la información es un factor importante que forma parte de la estrategia competitiva de las organizaciones, es necesario contar con herramientas que faciliten el manejo e interpretación de los datos que arrojan los sistemas operacionales y de producción.
4. Para que un sistema de soporte a la toma de decisiones proporcione al usuario lo que realmente requiere es necesario seleccionar correctamente los indicadores que servirán de medida para evaluar los procesos de una empresa. Los indicadores deben reflejar la situación a través del tiempo y permitir al usuario observar el verdadero escenario de la organización sin posibilidad de error.
5. El pronóstico es una herramienta que permite la reducción de riesgos en la toma de decisiones referidas a los distintos aspectos de una empresa. Al incorporar pronósticos en un sistema de soporte para la toma de decisiones se ofrece al usuario la posibilidad de realizar planificaciones basadas en condiciones futuras que rodean a una situación particular.
6. Para seleccionar la técnica estadística adecuada para un conjunto de datos es necesaria la identificación y comprensión de los patrones históricos. Si se pueden

reconocer patrones de tendencia, ciclos o estacionalidades entonces se puede seleccionar la técnica de pronóstico más eficaz para estos patrones.

7. Es importante para el usuario final que el SSD sea una herramienta que permita visualizar de manera clara, sencilla los datos de la empresa ya que éstos forman parte importante en el proceso de toma de decisiones. En esto radica la importancia de las herramientas de desarrollo que permitan la creación de gráficos y tablas para consulta de los datos que permitan un análisis más sencillo de los mismos.

RECOMENDACIONES

1. Crear conciencia en el personal que opera el sistema transaccional sobre la importancia de los datos que son ingresados.

2. Es recomendable que el usuario del sistema SSD ejecute la actualización de los factores alfa para los pronósticos cada tres meses para proporcionar siempre resultados precisos.

3. Agregar al sistema automatizado de procesamiento de los cubos una función en la cual se envíen mensajes vía correo electrónico que permitan al usuario verificar que se ha realizado la carga de los datos a los cubos así como también información adicional como la fecha de la carga y el número de lote que fue procesado.

4. Adicionalmente se recomienda la creación de una función que permita enviar vía correo electrónico a uno o varios usuarios las alarmas que el sistema genere y adicionalmente crear un módulo que permita agregar, consultar modificar y/o eliminar direcciones de correo a los cuales el sistema enviaría estos mensajes. Esta función permitiría llevar un control de los movimientos de la empresa incluso en momentos donde no se pueda revisar el sistema por un período de tiempo.

5. Se recomienda la expansión del sistema Web para poder ser usado en otras plataformas fuera de la red LAN y que los usuarios puedan conectarse al sistema sin necesidad de usar una conexión VPN.

REFERENCIAS BIBLIOGRÁFICAS

LIBROS

[1] Hanke John. (1996). *Pronósticos en los negocios.* Editorial: Prentice Hall.

[2] Kendall Kenneth. (1997). *Análisis y diseño de sistemas.* Editorial: Prentice Hall.

[3] Laudon Kenneth. (2004). *Sistemas de Información Gerencial.* Editorial: Prentice Hall.

[4] McConnell Steve. (1996). *Desarrollo y gestión de proyectos informáticos.* Editorial: McGraw-Hill.

[5] Pressman, R. (2002). *Ingeniería del Software: Un Enfoque Práctico.* Editorial: McGraw-Hill.

[6] Render Barry. (1996). *Métodos cuantitativos para los negocios.* Editorial: Prentice Hall.

[7] Schildt Herbert. (2003). *Manual de referencia C#.* Editorial: McGraw-Hill.

[8] Taha Handy. (2004). *Investigación de operaciones.* Editorial: Prentice Hall.

DOCUMENTOS ELECTRÓNICOS

[9] Buades Gabriel (1999). *Datawarehouse* [Documento en línea].
Disponible:
http://dmi.uib.es/~bbuades/datawarehouse/datawarehouse.ppt.
[Consulta: 2007, Julio 12].

[10] D'Agostino Sandra (2005). *Sistemas y Organizaciones* [Documento en línea].
Disponible:
http://extension.info.unlp.edu.ar/sistemas&organizacion/archivos/clase/CLASE%203%20-%202005.pdf
Consulta: 2007, Mayo 12].

[11] García Griñan (Fecha no disponible). *Almacenes de Datos (Data Warehouse).* [Documento en línea].
Disponible:
http://euler.fmat.ull.es/~bbdd/Ficheros/ALMACEN%20DE%20DATOS.pdf
[Consulta: 2007, Julio 23].

[12] Gonzales José (2005). *Estrategias de inteligencia de negocios* [Documento en línea].
Disponible:
http://www.cainco.org.bo/es/ecainco/actividades/seminarios/sem18042005/TallerBI.pdf
[Consulta: 2007, Febrero 12].

[13] Guzman Alejandro (2000). *¿Son los DSS solo para empresas grandes?*
[Documento en línea].
Disponible:
http://www.asidegroup.com/SonlosDSSsoloparalasgrandesempresas.pdf
[Consulta: 2007, Febrero 12].

[14] Hyperion Solutions Corporation. (2004). *Essbase Administration Services* [Documento en línea].
Disponible:
http://dev.hyperion.com/techdocs/eas/eas_712/eas_install.pdf
[Consulta: 2007, Junio 28].

[15] Magurno (s.f.). *Diseño OO* [Documento en línea].
Disponible:
http://homepage.mac.com/maguca/.Public/ucab/ds/05-disenoOO/diagramaClases.pdf
[Consulta: 2007, Mayo 12].

[16] Ruiz Tomás (2007). *Almacenes de Datos y Bases de Datos XML.* [Documento en línea].
Disponible:
http://alarcos.inf-cr.uclm.es/doc/bbddavanzadas/06-07/Funcionalidad%204.pdf
[Consulta: 2007, Julio 12].

[17] Trujillo Juan (2002). *Diseño de almacenes de datos.* [Documento en línea].
Disponible:
http://www.dlsi.ua.es/~jtrujillo/doctotema3disedw.pdf
[Consulta: 2007, Junio 21].

[18] Zvenger Andrea (2005). *Introducción al soporte de decisiones.* [Documento en línea].
Disponible:
http://www.devoo.net/Introduccion_al_Soporte_de_Decisione.pdf
[Consulta: 2007, Agosto 1].

PÁGINAS WEB

[19] DeGerencia.com (Fecha no disponible). *Toma de decisiones* [Pagina Web en línea].
Disponible:
http://www.degerencia.com/tema/toma_de_decisiones
[Consulta: 2007, Marzo 12].

[20] ETL-Tools.Info (2007). *Almacenes de Datos - Business Intelligence - ETL* [Página Web en línea].
Disponible:
http://etl-tools.info/es/bi/proceso_etl.htm
[Consulta: 2007, Julio 27].

[21] Instituto Nacional de Estadística e Informática (1997). *Manual para la Construcción de un Data Warehouse* [Página Web en línea].
Disponible:
http://www.inei.gob.pe/biblioineipub/bancopub/Inf/Lib5084/INDEX.HTM
[Consulta: 2007, Abril 27].

[22] Microsoft (2006). *SQL Server 2005 Readme and installation requirements* [Página Web en línea].
Disponible:
http://support.microsoft.com/kb/910228/en-us#8
[Consulta: 2007, Junio 28].

[23] MicrosoftTechnet (Fecha no disponible). *Tutorial para crear un paquete ETL sencillo* [Página Web en línea].
Disponible:
http://technet.microsoft.com/es-es/library/ms169917.aspx
[Consulta: 2007, Agosto 10].

[24] Oracle Business Intelligence (Fecha no disponible). *Installation Guide 10g Release 2 (10.1.2.1) for Microsoft Windows and Solaris Operating System (SPARC)* [Página Web en línea].
Disponible:
http://download.oracle.com/docs/html/B16030_01/requirements.htm#CEGEFAGH
[Consulta: 2007, Junio 28].

[25] Rada Gabriel (2007). *Clasificación de indicadores* [Pagina Web en línea].
Disponible:
http://escuela.med.puc.cl/Recursos/recepidem/insIntrod9c.htm
[Consulta: 2007, Abril 14].

[26] Scielo (2006). *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP* [Página Web en línea].
Disponible:
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S012056092006000300016&lng=en&nrm=iso
[Consulta: 2007, Marzo 15].

[27] Universidad Católica Boliviana (2000). *Los sistemas de apoyo a la toma de decisiones* [Página Web en línea].
Disponible:
http://web.syr.edu/~iguzmand/SPdss2.htm
[Consulta: 2007, Marzo 15].

[28] Universidad Nacional del centro (Fecha no disponible). Pronósticos [Página Web en línea].
Disponible:

http://64.233.169.104/search?q=cache:rFodax4D4xcJ:www.eisi.ues.edu.sv/jportillo/ApuntePronosticos.pdf+Facultad+de+Ciencias+Exactas+El+pron%C3%B3stico+es+un+proceso+de+estimaci%C3%B3n+de+un+acontecimiento+futuro+proyectando&hl=es&ct=clnk&cd=1&gl=ve [Consulta: 2007, Marzo 15].

GLOSARIO

-A-

Atenuación exponencial: Es un proceso utilizado para revisar constantemente un pronóstico a la luz de la experiencia más reciente.

-C-

Coeficiente de correlación: Medición que indica hasta qué grado se relacionan linealmente dos variables numéricas entre sí.

Cubo: Estructura multidimensional que forma la base par a las aplicaciones OLAP. A pesar del nombre, la mayoría de los cubos OLAP tienen mucho más de tres dimensiones. En una base de datos multidimensional de OLAP o (MOLAP), los cubos son creados y almacenados físicamente, mientras que en las bases de datos relacionales de OLAP (ROLAP), los cubos son un concepto virtual basado en un esquema de copo de nieve o de estrella Una variedad de cálculos inter-dimensionales y agregaciones son posibles dentro de un cubo.

-D-

Data Warehouse (Almacén de datos): es una colección de datos orientados, integrados, no volátiles, y variantes en el tiempo. El data warehouse contiene datos a nivel atómico y data sumarizada específicamente estructurado para consultas y reportes.

Desnormalización: La técnica de quitar la normalización de los datos para que optimice el rendimiento del sistema.

Desviación Absoluta de la media (DAM): mide la precisión de un pronóstico mediante el promedio de la magnitud de los valores absolutos del error.

Diccionario de datos: Un modulo de software y base de datos que contiene la descripción y definición de la estructura, elementos, interrelaciones y otras características de una base de datos.

Dimensión: Una dimensión es típicamente un valor de texto, como una región, línea de producto, e incluye valores de fechas. Son en otras palabras lo que se quiere medir. Esto define los títulos o etiquetas secundarias que conforman el cuerpo de los reportes.

-E-

Esquema Estrella: es la arquitectura de almacén de datos más simple. En este diseño la tabla de Hechos está rodeada por Dimensiones y juntas forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP.

Esquema Copo de Nieve: Este esquema es obtenido por la normalización de las dimensiones que componen un modelo estrella y tiene como objetivo reducir la información redundante de los modelos dimensionales.

ETT (Extracción, Transformación y Transporte): Se refiere al proceso de obtener datos de un almacén de datos, modificarlos e insertarlos dentro de un almacén de datos diferente.

-G-

Granularidad: El nivel de detalle de los datos almacenados en un almacén de datos.

-H-

HOLAP (Hybrid OLAP): almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

-M-

MOLAP: OLAP en modelo multidimensional. En los servidores MOLAP, el almacén de datos almacena datos multidimensionales en estructuras especiales de datos e implementa las operaciones OLAP sobre estas estructuras.

-N-

Normalización: es un proceso paso a paso de remover redundancias y dependencias de atributos en una estructura de datos. La condición de los datos al completar en cada paso es descrita como "Forma Normal".

-O-

OLAP: es el acrónimo en inglés de procesamiento analítico en línea (*On-Line Analytical Processing*). Es una solución utilizada en el campo de la Inteligencia de Negocios la cual consiste en consultas a estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de Datos o Sistemas Transaccionales (OLTP).

-P-

Promedio móvil: Se obtiene encontrando la media de un conjunto específico de valores y usándola después para pronosticar el siguiente período.

Promedio simple: se obtiene encontrando la media de todos los valores aplicables y empleando después esa media para pronosticar el siguiente período.

-R-

ROLAP: es una implementación OLAP que almacena los datos en un motor relacional.

-S-

Serie estacionaria: Es aquella cuyo valor estacionario no cambia a través del tiempo.

Serie de tiempo: Consiste en datos reunidos, registrados u observados en incrementos sucesivos de tiempo.

Señal de Rastreo: Comprende el cálculo de alguna medición de error a través del tiempo y el establecimiento de límites, de modo que cuando el error acumulativo rebase dicho límite, se alerte al pronosticador.

Sistema de Soporte a la toma de decisiones (SSD): Conjunto de programas y herramientas que permiten obtener de manera oportuna la información que se requiere mediante el proceso de la toma de decisiones que se desarrolla en un ambiente de incertidumbre. Ayudan a la toma de decisiones de los administradores al combinar datos, modelos analíticos sofisticados y software amigable en un solo sistema poderoso que puede dar soporte a la toma de decisiones semiestructuradas o no estructuradas.

-T-

Tabla de Hechos (Fact Table): Las tablas que se extraen de fuentes heterogéneas y se utilizan en los almacenes de datos (Data Warehouse). Contiene el grupo de medidas que dan valor a los atributos de las dimensiones.