



Instituto Tecnológico de Costa Rica
Departamento de Ingeniería en Computación
Programa de Maestría

Prototipo de Mercado de Datos para la División de Control
y Fiscalización de la Dirección General de Aduanas

Proyecto de Graduación para optar al grado de Master Profesional en
Computación con mención en Sistemas de Información

Ricardo Chinchilla Arley
José Luis Vargas Murillo

Cartago, Costa Rica
Marzo, 2002

Aprobación de Proyecto

Diseño de un prototipo de Mercado de Datos para la División de Control y Fiscalización de la Dirección General de Aduanas

Tribunal examinador

M.Sc. Jorge Vázquez Rodríguez
Profesional Externo

Dr. José Enrique Araya Monge
Profesor

M.Sc. Jose Helo Guzmán
Profesor

M.Sc. Daniel Antich Montero
Profesor Asesor

Ing. Kirstein Gatjens Soto, Director
Departamento de Ingeniería en Computación
Programa de Maestría

Dedicatoria

*Al Señor mi Dios, fuente de inspiración y fortaleza
A mis padres, ejemplo de lucha y constancia
A mi esposa, la mayor sacrificada en el proceso
A mi hijo, la esperanza del mañana*

Ricardo

*A Dios por su amor infinito y a mi esposa por su invaluable paciencia y apoyo
A mi hijito de cinco años que me proporciona aliento para ver el futuro
A mi mamá por su ejemplo de perseverancia y esfuerzo*

José Luis

Agradecimientos

Nuestro profundo agradecimiento al M.Sc. Daniel Antich Montero por su desinteresada colaboración en la dirección de este proyecto. Su aporte preciso a nuestro trabajo ha sido de incalculable valor.

Al M.Sc. Fernando Morales Abarca, por sus consejos en la preparación de la propuesta del tema y de este informe final.

Tabla de contenidos

CAPÍTULO 1. INTRODUCCIÓN.....	1
1.1 EL PROBLEMA Y SU IMPORTANCIA.....	1
1.2 OBJETIVOS.....	7
1.3 ALCANCES Y LIMITACIONES.....	7
CAPÍTULO 2. MARCO TEÓRICO.....	10
2.1 ANTECEDENTES.....	10
2.2 TÉRMINOS BÁSICOS.....	12
2.3 METODOLOGÍAS DE DESARROLLO.....	16
2.3 MERCADO DE DATOS.....	25
2.4 PROCESAMIENTO ANALÍTICO EN LÍNEA (OLAP).....	28
CAPÍTULO 3. DESCRIPCIÓN DEL SISTEMA ACTUAL.....	33
3.1 RUTINAS DE EXTRACCIÓN Y GENERACIÓN DE DATOS.....	34
3.2 INFRAESTRUCTURA DE APOYO (EQUIPOS, PLATAFORMAS Y RED).....	43
CAPITULO 4. REQUERIMIENTOS DE INFORMACIÓN Y USUARIOS DEL MERCADO DE DATOS.....	44
4.1 ENTORNO GENERAL.....	44
4.2 DETERMINACIÓN DE USUARIOS Y SUS REQUERIMIENTOS DE INFORMACIÓN.....	45
CAPÍTULO 5: PROPUESTA DE DISEÑO.....	57
5.1 DEFINICIÓN DE DIMENSIONES.....	57
5.2 METODOLOGÍA DE DESARROLLO.....	58
5.3 DISEÑO LÓGICO: MODELO GRÁFICO.....	60
5.4 DISEÑO FÍSICO: DESCRIPCIÓN DEL PROTOTIPO.....	68
5.5 ACCESO AL MERCADO DE DATOS POR INTERNET.....	99
5.6 REQUERIMIENTOS DE HARDWARE.....	107
CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES.....	112
6.1 CONCLUSIONES.....	112
6.2 RECOMENDACIONES.....	114
BIBLIOGRAFÍA.....	116
APÉNDICE A. DICCIONARIO DE DATOS.....	120
APÉNDICE B. CÓDIGO FUENTE, PROGRAMA DE CARGA.....	126
APÉNDICE C. ARCHIVOS IDC PARA CONSULTAS POR INTERNET.....	134
ANEXO A. CONSULTAS SOBRE IMPORTACIONES.....	140
ANEXO B. ENTREVISTA A USUARIOS DE LA INFORMACIÓN.....	145
ANEXO C. LISTA DE CAMPOS DE BASE DE DATOS DADEC.....	147
ANEXO D. CARTA DE APROBACIÓN DE LA EJECUCIÓN DEL PROYECTO POR PARTE DE LA DCF.....	149

Índice de figuras

<u>FIGURA 1: EVOLUCIÓN DE LA TECNOLOGÍA DE DD.....</u>	<u>12</u>
<u>FIGURA 2: REPRESENTACIÓN DE UN HIPERCUBO.....</u>	<u>13</u>
<u>FIGURA 3: CICLO DE DESARROLLO DE SOFTWARE UTILIZADO PARA EL DD SEGÚN GILL.....</u>	<u>20</u>
<u>FIGURA 4: FASES DEL PROYECTO DE INICIACIÓN DEL DD SEGÚN DEVLIN.....</u>	<u>22</u>
<u>FIGURA 5: PROCESO ITERATIVO DEL DD SEGÚN EBEL.....</u>	<u>23</u>
<u>FIGURA 6: ELEMENTOS CLAVE DE UN DD.....</u>	<u>24</u>
<u>FIGURA 7: EQUIPO DE PROYECTO DA DD SEGÚN COREY.....</u>	<u>25</u>
<u>FIGURA 8: MERCADO DE DATOS.....</u>	<u>27</u>
<u>FIGURA 9: ESQUEMA DE DESARROLLO INCREMENTAL DE UN DD SEGÚN WOLF.....</u>	<u>27</u>
<u>FIGURA 10: MULTIDIMENSIONAL ON-LINE ANALYTICAL PROCESSING.....</u>	<u>30</u>
<u>FIGURA 11: ESQUEMA DEL PROCESAMIENTO ANALÍTICO EN LÍNEA RELACIONAL.....</u>	<u>31</u>
<u>FIGURA 12: ESQUEMA GENERAL DEL SISTEMA NACIONAL DE ADUANAS.....</u>	<u>33</u>
<u>FIGURA 13: PROCESOS DE EXTRACCIÓN Y GENERACIÓN DE DATOS.....</u>	<u>34</u>
<u>FIGURA 14: ESQUEMA DE ALMACENAMIENTO EN EL DEPARTAMENTO DE ESTADÍSTICA.....</u>	<u>41</u>
<u>FIGURA 16: ESQUEMA DEL MODELO DIMENSIONAL DEL PROTOTIPO.....</u>	<u>62</u>
<u>FIGURA 17: ESQUEMA DEL MODELO DE JERARQUÍAS DEL PROTOTIPO.....</u>	<u>63</u>
<u>FIGURA 18: DRILL DOWN Y ROLL UP EN LA DIMENSIÓN MERCANCÍA.....</u>	<u>66</u>
<u>FIGURA 19: DRILL DOWN Y ROLL UP EN LA DIMENSIÓN TIEMPO.....</u>	<u>67</u>
<u>FIGURA 21: ESQUEMA DE LA ESTRUCTURA DE LA BASE DE DATOS.....</u>	<u>69</u>
<u>FIGURA 22: ILUSTRACIÓN DE LA JERARQUIZACIÓN DE LA DIMENSIÓN MERCANCÍA.....</u>	<u>70</u>
<u>FIGURA 23: ILUSTRACIÓN DE LA JERARQUIZACIÓN DE LA DIMENSIÓN TIEMPO.....</u>	<u>71</u>
<u>FIGURA 24: PROCESO DE TRANSFORMACIÓN Y CARGA.....</u>	<u>73</u>
<u>FIGURA 25: FORMULARIO DE ARCHIVOS CARGADOS.....</u>	<u>78</u>
<u>FIGURA 26: FORMULARIO DE CONSULTA DE MERCANCÍAS.....</u>	<u>79</u>
<u>FIGURA 27: FORMULARIO DE CARGA DE DATOS AL MD.....</u>	<u>80</u>

<u>FIGURA 28: DISTRIBUCIÓN DE DIRECTORIOS DE PROGRAMAS, DATOS Y RESULTADOS DE LA APLICACIÓN.....</u>	<u>82</u>
<u>FIGURA 29: INTERFASE DE CONSULTA MULTIDIMENSIONAL DEL PROTOTIPO.....</u>	<u>83</u>
<u>FIGURA 30: COMPONENTE PARA LA DEFINICIÓN DEL PERIODO DE ANÁLISIS.....</u>	<u>84</u>
<u>FIGURA 31: FICHA DE SELECCIÓN DE DIMENSIONES DESCRIPTIVAS.....</u>	<u>86</u>
<u>FIGURA 32: FILTRADO DE CONSULTAS POR ATRIBUTOS DE CLASIFICACIÓN.....</u>	<u>89</u>
<u>FIGURA 33: FILTRADO DE CONSULTAS POR ATRIBUTOS DE CLASIFICACIÓN.....</u>	<u>91</u>
<u>FIGURA 34: VISUALIZACIÓN DE INSTANCIAS SELECCIONADAS DE LA DIMENSIÓN ADUANA.....</u>	<u>92</u>
<u>FIGURA 35: FICHA DE SELECCIÓN DE DIMENSIONES DESCRIPTIVAS.....</u>	<u>93</u>
<u>FIGURA 36: VISUALIZACIÓN DE LOS ESTATUTOS SQL DE LA CONSULTA.....</u>	<u>95</u>
<u>FIGURA 37: VISUALIZACIÓN DEL RESULTADO DE UNA CONSULTA.....</u>	<u>96</u>
<u>FIGURA 38: SALIDA DE RESULTADOS A DIFERENTES FORMATOS.....</u>	<u>97</u>
<u>FIGURA 39: ACCESO A BASES DE DATOS POR IIS</u>	<u>101</u>
<u>FIGURA 40: COMPONENTES DE CONEXIÓN DE IIS.....</u>	<u>101</u>
<u>FIGURA 41: PANTALLA PRINCIPAL DE CONSULTA POR WWW.....</u>	<u>104</u>
<u>FIGURA 42: CONSULTA GENERAL POR ADUANA E IMPORTADOR.....</u>	<u>106</u>
<u>FIGURA 43: CONSULTA DE LOS 10 MAYORES IMPORTADORES.....</u>	<u>107</u>

Lista de Abreviaturas

CIF: Siglas en inglés de costo, seguro y flete

DAI: Derechos Arancelarios a la Importación

DBF: Data Base Format

DCF: División de Control y Fiscalización

DD: Depósito de Datos

DE: Departamento de Estadística

DGA: Dirección General de Aduanas

HTML: Hyper Text Mark-up Languaje

IDC: Internet Data Connector

IIS: Internet Information Server

ISAPI: Internet Server Application Program Interface

IV: Impuesto de Ventas

MD: Mercado de Datos

ODBC: Open Database Connectivity

OLAP: On-Line Analytical Processing

RAID: Redundant Arrays of Inexpensive Disks

SAC: Sistema Arancelario Centroamericano

SC: Impuesto Selectivo de Consumo

SIA: Sistema de Información Aduanera

SNA: Servicio Nacional de Aduanas

Capítulo 1. Introducción

1.1 El problema y su importancia

La nueva Ley General de Aduanas (vigente desde julio de 1996), inicia una etapa en materia de control y fiscalización que fundamenta muchos de los procedimientos de análisis y establecimiento de estrategias, en las facultades de acceder a las distintas fuentes de información, internas y externas. Considera de paso, la característica de confianza que deposita en los usuarios autorizados (empresas con beneficios fiscales, agentes aduaneros, depósitos fiscales y otros), procura mantener al mismo tiempo en operación, un sistema aduanero flexible y simplificado.

Desde esa perspectiva, es necesario un cambio en el concepto de control y fiscalización por uno más integral y práctico que use las distintas fuentes de información disponibles y que además sea inteligente y esté apoyado en la tecnología de información apropiada.

La información administrada por el Servicio Nacional de Aduanas (SNA) en todo el país, es actualmente almacenada en el Sistema de Información Aduanera (SIA), el cual contiene los datos de los ingresos fiscales por concepto de impuestos de importación, que constituyen aproximadamente el 50% del presupuesto nacional. La información almacenada es utilizada por muchos usuarios, dentro de los que se destacan el Ministerio de Hacienda, Banco Central y el Instituto Nacional de Estadística y Censos.

El SIA consiste de un conjunto de aplicaciones de base de datos sobre una plataforma Oracle©, que procesa y registra la mayoría de las transacciones aduaneras que recibe. Ejemplo de dichas operaciones son las declaraciones aduaneras de importación o exportación, declaraciones aduaneras de tránsito, entre otras, establecidas en el artículo

110 de la Ley General de Aduanas .

No obstante que el SIA mantiene los registros de las operaciones de los usuarios desde 1993, la información que alberga es controlada por aplicaciones desarrolladas para el registro, seguimiento y control de transacciones, y no están diseñadas para permitir el acceso rápido y preciso requerido en los procesos de análisis y toma de decisiones de la administración aduanera. Por otro lado, es importante apuntar que el SIA es un sistema que opera de forma centralizada en cada aduana, por lo que los datos son almacenados y controlados de forma independiente en cada una de ellas. Por tal motivo, no es posible una visualización global de la información, al no existir una integración interaduanal.

Este sistema tiene 10 años de estar en producción. La versión sobre la que está desarrollado no recibe soporte técnico del fabricante debido a su antigüedad y realizar modificaciones estructurales profundas conllevarían un costo muy elevado.

Tal circunstancia, dificulta también la adopción de estrategias y el establecimiento de directrices de control que permitan definir con menor incertidumbre las empresas, procesos o mercancías que se habrán de seleccionar como sujetos de estudio para propósitos fiscales y aduaneros.

Es necesario contar con una tecnología que permita hacer comparaciones, encontrar patrones de comportamiento, hallar tendencias, establecer relaciones que ayuden a la detección o prevención de delitos aduaneros o el señalamiento de incumplimientos de las leyes y normas del comercio nacional e internacional. Es aquí donde el desarrollo de un Depósito de Datos (DD), definido por [Gonz98a] como un *“[...] repositorio único, completo y consistente de datos obtenidos de diferentes fuentes [...] y que están disponibles al usuario de forma tal que pueda ser comprendido y usado en un contexto de empresa”*, proporcionaría al SNA una herramienta capaz de suministrar

información transformada, filtrada, resumida, agrupada y lista para satisfacer los requerimientos de los usuarios.

El proceso de análisis o toma de decisiones basada en la información extraída del SIA, es una tarea cotidiana dentro de las dependencias encargadas del control aduanero nacional. Una de estas dependencias es la División de Control y Fiscalización (DCF) en la Dirección General de Aduanas (DGA), y está encargada de definir la pauta en la orientación y planeamiento de las funciones fiscalizadoras en general del SNA.

Dentro de las funciones importantes de la DCF se encuentra el velar por la definición y aplicación oportuna y precisa de criterios de control que son los que, a fin de cuentas, determinarán hacia quién, cómo y cuándo deben dirigirse los esfuerzos fiscalizadores.

Los criterios de control emanados de la DCF, se basan en varias fuentes de información, entre las cuales están: 1- el SIA, 2- informes estadísticos, 3- informes periódicos de las distintas aduanas del país con respecto a la actuación de sus usuarios, 4- denuncias, 5- reportes de otras divisiones o departamentos de la DGA, 6- informes de las empresas o auxiliares de la función pública, y 7- expedientes.

Los informes estadísticos son producidos por la División de Estadística Registro y Divulgación de la DGA utilizando la información almacenada en el SIA. Debido a que este sistema no es apropiado para el trabajo de análisis, los datos que se utilizan para satisfacer las solicitudes de información provienen de archivos de texto (planos) extraídos del SIA, generados por funcionarios en las aduanas o desde la misma División de Estadística.

Éstos archivos planos son producidos por rutinas con instrucciones SQL (lenguaje estructurado de peticiones) que extraen los valores de las variables

seleccionadas y registradas en un periodo determinado. Estos valores son descargados en tablas tipo DBF (por sus siglas en inglés de *Data Base File*), a partir de las cuales se extraen otras tablas con información específica, según los requerimientos de la DCF y otras oficinas o dependencias.

Las tablas que componen los informes estadísticos son archivos independientes, están desligados unos de otros y no conforman una estructura de base de datos relacional formal. Por ende, a partir de ellas no es posible hacer fácilmente consultas que asocien sus variables comunes, dificultando a la DCF el análisis global e interrelacionado de los datos, lo cual es vital para la orientación de las investigaciones y el ejercicio del control fiscal.

Se hace necesario entonces, un sistema que permita la integración de la información obtenida del SIA y provea las herramientas necesarias para ser utilizadas en el análisis y la evaluación de la información extraída. Información que contribuirá a disminuir la incertidumbre en el control de los riesgos aduaneros, como son la defraudación fiscal, contrabando y el incumplimiento de la normativa aduanera.

Un Mercado de Datos (MD) lo define [Core97] como un “... *subconjunto de un Depósito de Datos que se extrae para satisfacer las necesidades de una clase particular de usuario*”. Es una tecnología que resolvería la problemática descrita, debido a las facilidades de almacenamiento y procesamiento analítico de que dispone.

Un MD proporciona un esquema de almacenamiento que permite realizar consultas complejas, por medio de la interacción e interrelación de múltiples entidades, manteniendo solo un repositorio con la información agregada y ordenada, según las necesidades de los usuarios. Además, provee de una interfase de consulta que permite al usuario la posibilidad de análisis de la información para la toma de decisiones. Esto con

el propósito de atacar los delitos e infracciones aduaneras, así como procurar el ejercicio del control fiscal.

Los usuarios finales del Mercado de Datos propuesto, serán aquellos funcionarios de nivel medio y alto de la DCF (profesionales, jefes de departamento y jefe de división) que requieren información resumida o detallada del sistema operacional, pero que, a través de una herramienta apropiada, les permita la combinación de múltiples variables y el análisis fundamentado de la información.

El desarrollo de un MD implica un gran esfuerzo que consume tiempo y recursos económicos, técnicos y humanos. [Core97] establece que, debido a la complejidad en la construcción de un MD, un equipo de trabajo se debería componer por los siguientes especialistas:

- Gestor del proyecto del MD.
- Arquitecto del MD.
- Administrador de la base de datos.
- Administrador del sistema
- Especialista en migración de datos.
- Especialista en el sistema heredado.
- Especialista en la transformación y filtrado de los datos.
- Líder de desarrollo de MD.
- Especialista en calidad y pruebas.
- Especialista en infraestructura.
- Usuarios clave.
- Escritor técnico.

- Especialista en herramientas.

El MD propuesto en este documento, no pretende ser una herramienta completa o totalmente desarrollada, ya que para ello se requeriría todo el equipo descrito por [Core97]. Se pretende más bien, el desarrollo de un prototipo orientado inicialmente al análisis de las importaciones definitivas, el cual, de paso, contribuya a darle al personal de la DCF una perspectiva clara sobre la necesidad de contar con una herramienta de este tipo. Esto a fin de cumplir con los objetivos institucionales y a estimar la factibilidad de desarrollarla completamente.

Conforme a lo anterior, las razones para desarrollar el sistema pueden resumirse de la siguiente forma:

1. Pretende ser un proyecto innovador dentro del área de control y fiscalización aduanera.
2. Produce un fuerte impacto al convertirse en un mecanismo fundamental de apoyo a la toma de decisiones.
3. Proporciona información para contrarrestar o prevenir la comisión de delitos aduaneros, las infracciones administrativas o tributarias aduaneras.
4. Representa un nuevo enfoque al tratamiento de la información en la DCF.
5. Pretende mantener la información dentro de una estructura de base de datos multidimensional y ajustada a las necesidades de análisis específicas de la DCF.
6. Procura el aprovechamiento eficiente del recurso humano y tecnológico, actualmente obligado a emplear su tiempo en tareas distintas a las que les corresponden.

1.2 Objetivos

1.2.1 Objetivo general

Desarrollar un prototipo de Mercado de Datos que constituya una herramienta para facilitar a la División de Control y Fiscalización de la Dirección General de Aduanas, el análisis de la información relativa a las importaciones definitivas de mercancías.

1.2.2 Objetivos específicos

- Determinar los potenciales usuarios de la información del Mercado de Datos propuesto.
- Desarrollar una base de datos que constituya el repositorio para el Mercado de Datos.
- Construir los programas de carga y mantenimiento de los datos.
- Diseñar una interfase de consulta para el prototipo de Mercado de Datos.
- Implementar algunas consultas generales al Mercado de Datos a través de Internet.

1.3 Alcances y Limitaciones

El prototipo se limita a los datos suministrados en los informes estadísticos solicitados por la DCF, a la División de Estadística Registro y Divulgación. Los datos en cuestión consisten en aquellos valores contenidos en los campos del SIA correspondientes a las declaraciones aduaneras de importación definitiva, que la DCF ha considerado como necesarios para la atención de la demanda de información. Ejemplos de las variables disponibles son: el código de la aduana, el detalle de la mercancía, el valor CIF, el monto de los impuestos, la partida arancelaria y el importador.

Como fundamento metodológico para el diseño lógico del MD que se propone, se

utilizará el modelo planteado por Sonia Cortez Araniva en su informe de tesis, con la que optó en 1999 al grado de Magíster Scientiae en Computación en el Instituto Tecnológico de Costa Rica. Las razones por la cuales se adopta este modelo son, en primer término, su simplicidad y fácil comprensión y, por otro lado como ella misma lo afirma “*[los otros modelos] no se han orientado [a diferencia de su planteamiento] a sugerir un modelo multidimensional gráfico y sin orientación a una implementación específica*” [Cort99].

La herramienta que se utilizará para la construcción del prototipo será Visual FoxPro 6.0, por considerar que reúne las características necesarias para su desarrollo, además de ser una herramienta que contiene un lenguaje orientado a objetos que facilita la programación, y un motor de base de datos integrado. Por otro lado, es de bajo costo, flexible, rápido y una tiene una interfase gráfica de usuario de relativa facilidad y uso.

La base de datos será implementada bajo el esquema estrella y no tendrá tablas de resumen, por lo que el cálculo de las consultas será efectuado sobre la tabla de hechos general. Por otro lado, el prototipo a desarrollar proporciona una interfase de consulta que pretende ser una aproximación a un OLAP (On-Line Analytical Processing). Esta interfase permitirá consultas multidimensionales, la reducción dimensional y visualizar la información agrupada desde distintas perspectivas, además de ofrecer salidas a distintos formatos de archivo.

El prototipo tendrá un programa de carga de datos que filtrará las información proveniente de las tablas planas, de forma tal que puedan ser descargadas sin dificultad en el repositorio. Contará también, con módulos de mantenimiento de las dimensiones.

Además de lo anterior, se desarrollarán algunas ventanas de consulta así como sus respectivas salidas con los lenguajes de programación interpretados html y Java script. Dichas consultas están basadas en una tabla de agregación precalculada, orientadas a las

necesidades generales de información de los gerentes y jefes de departamento de las aduanas. Esto se hace con el fin de lograr una implementación del sistema en el esquema de web y lograr acceder a él a través de Internet. Con esto se abre la posibilidad de, a futuro, brindar información dinámica a otros usuarios tanto del Ministerio de Hacienda y sus dependencias, como al público en general.

Capítulo 2. Marco teórico

2.1 Antecedentes

Un Depósito de Datos (DD) lo define [Poe98] como

una base de datos de solo lectura donde la información extraída de los sistemas operacionales corrientes de la empresa, es transformada, integrada y resumida para luego ser usada con efectividad en el soporte de decisiones.

[Simo96], por su parte, no considera un DD como simplemente una base de datos, sino que va más allá afirmando que

el enfoque más recomendable para definir un DD es aquel que lo concibe como una solución arquitectónica en la cual se conjugan la mayoría de los elementos individuales que surgen cuando las personas describen su visión de un DD, y se orientan con claridad hacia las características de un ambiente dado.

No importa la definición que se desee aplicar, se debe tener siempre presente que la decisión de implementar este tipo de tecnología podría implicar una transformación profunda en la forma como actualmente se manipulan los datos. De hecho, el DD por sí mismo traerá consigo el consumo de gran cantidad de recursos y requerirá de un largo tiempo de implementación, sin embargo, su utilización es necesaria debido a la necesidad de mejorar la administración general de los datos y de tener una visión más amplia de la información, al disponer de un acceso a la misma sin importar su ubicación. Con el DD se pretende solucionar problemas derivados de la dispersión de los datos en las diferentes aplicaciones transaccionales, muchas veces heterogéneas y proporcionar información para la toma de decisiones.

Históricamente, la idea del DD surge inmediatamente posterior al nacimiento del usuario como tal (1970-75). Éste debe ahora solucionar problemas que antes estaban del lado de los programadores y analistas, al usar soluciones tecnológicas particulares. En la

década de los 80 casi todo lo relacionado con datos ha sido automatizado y se hace prioritario el apoyo a la toma de decisiones con base en dichos datos almacenados. La primera técnica de toma de decisiones se dio con la automatización de reportes. Ya para los noventa se ha dado una gran evolución en cuanto a aplicaciones de usuario final y surgen gran cantidad de sistemas de información, que a diferencia de los sistemas operacionales enfocados en la producción, se especializan en el apoyo a la toma de decisiones, manipulando datos históricos y puntuales y optimizados para la consulta. Sin embargo, se basan en bases de datos transaccionales, las cuales trabajan con datos en tiempo real. Además, muchas empresas tienen diferentes plataformas de bases de datos, por lo que una búsqueda inteligente de información puede tomar gran cantidad de tiempo, degradando la efectividad de la decisión tomada con base en ella. Es aquí donde se hace presente el DD, el cual ha tenido un desarrollo muy discreto, a la sombra de los sistemas de información, pero actualmente ha tenido un repunte debido principalmente a los avances tecnológicos.

Utilizando parte de lo expuesto por [Gill96], la evolución de la tecnología se puede resumir en la figura 1.

Con la incorporación de los DD, las empresas han tratado de minimizar la herencia negativa de los sistemas transaccionales con que cuentan, aumentar el rendimiento, lograr administrar los datos en el tiempo y brindar un efectivo apoyo a la toma de decisiones, con las ventajas competitivas que esto trae consigo.

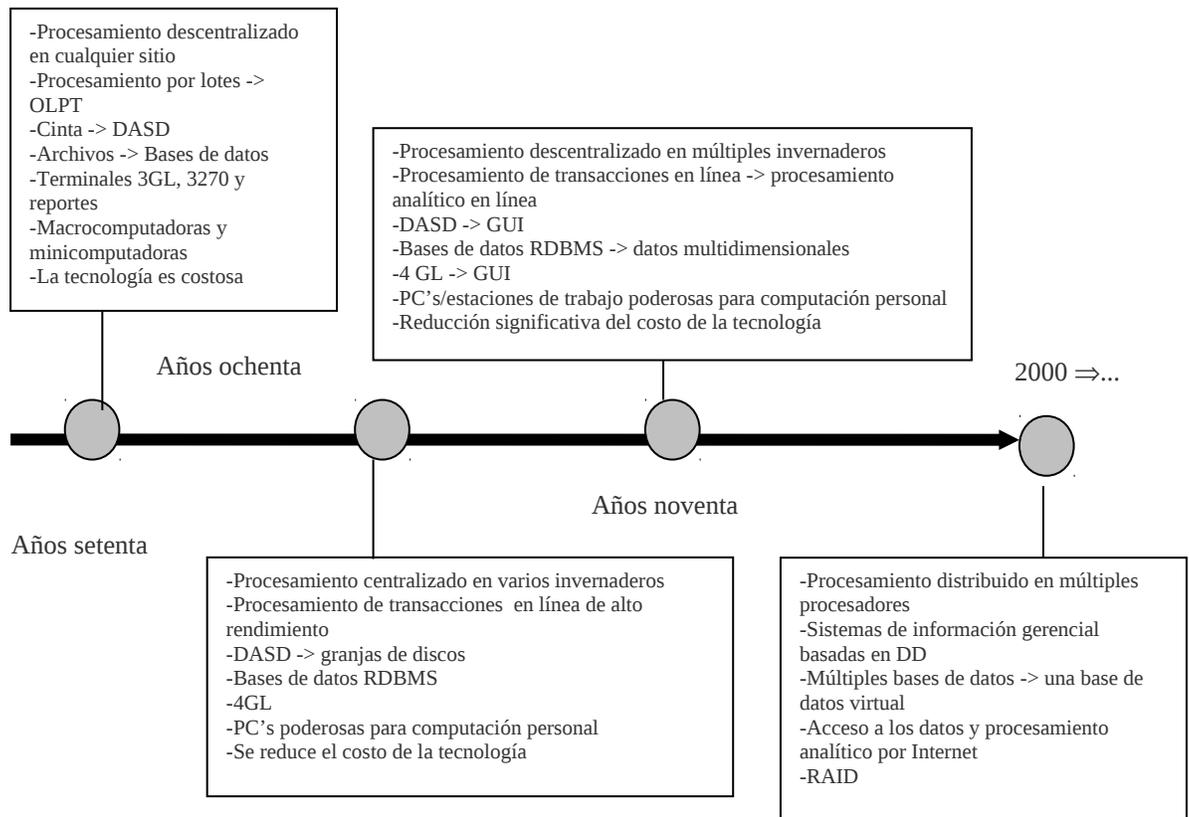


FIGURA 1: Evolución de la tecnología de DD

2.2 Términos básicos

Los DD son también conocidos como hipercubos multidimensionales o cubos nDimensionales, debido a que presentan la información en más de tres dimensiones (>3Dim). Se puede esquematizar de la siguiente manera:

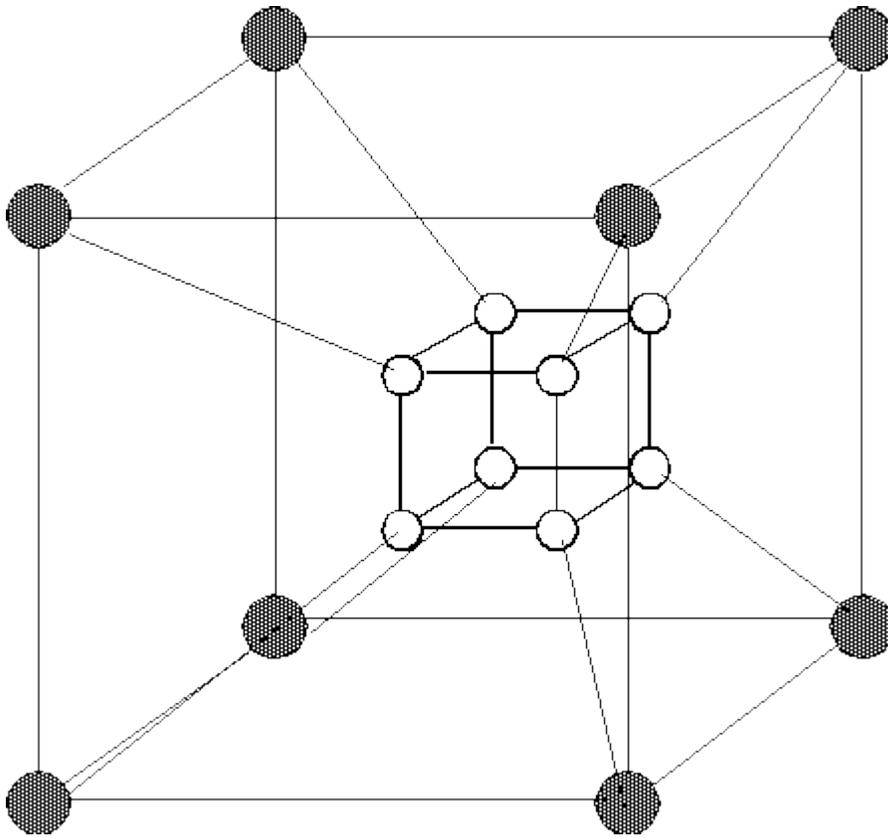


FIGURA 2: Representación de un hiper-cubo

[Gonz98] indica que

el hiper-cubo funciona como una representación intuitiva de un evento [generador de datos de más de 3 dimensiones] debido a que todas las dimensiones coexisten para cada punto del mismo y son independientes entre sí.

La combinación de n Dimensiones y varios niveles o jerarquías por dimensión constituyen la esencia del hiper-cubo.

Datos multidimensionales

Según [Wel95], cualquier dato en teoría puede ser multidimensional. El término se refiere a datos (objetos o eventos) que se pueden describir o clasificar por dos o más atributos.

Bases de datos multidimensionales

Son repositorios grandes de datos, donde éstos son descritos de acuerdo a distintas perspectivas o dimensiones. Según [Cox00], deben cumplir con cuatro características: rapidez, capacidad analítica, soporte multiusuario y multidimensionalidad.

Análisis multidimensional

Podría definirse como la clasificación de la información de acuerdo a categorías. [Gill96] lo define como el análisis simultáneo de múltiples dimensiones de los datos.

Dimensiones

Son las variables sobre las que se califican los datos y pueden ser descriptivas o de hechos. Las descriptivas proporcionan una explicación de los datos que serán analizados. En una librería las dimensiones descriptivas detallarían un producto, como por ejemplo cuadernos, clasificándolos en tamaño, cantidad de hojas, colores, pasta dura o suave, etc. La dimensión de hechos, por su parte, contendría los valores cuantitativos o de medida, como las cantidades, las ventas totales y los precios.

Jerarquías

[Thom97] las describe como una estructura de árbol, con una raíz en la parte superior y las ramas en la parte inferior. Si todas las ramas se encuentran en un mismo nivel, se denomina jerarquía simétrica, pero si esto no ocurre, se denomina jerarquía asimétrica. Por su parte, en [Orac96] se define como una colección de miembros de dimensión con series en cascada de relaciones del tipo uno a muchos, donde miembro es un valor de una dimensión. Por ejemplo, un miembro de la dimensión Tiempo sería el mes “Abril”.

Tabla Dimensional

Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad al modelo.

Rotación

Consiste en cambiar o reacomodar la orientación dimensional de una presentación en pantalla.

Roll-up y Drill-down

Es posible transformar fácilmente la vista dentro de las jerarquías, subiendo o bajando niveles. Aprovechándolas es posible empezar en un nivel de resumen y selectivamente ir obteniendo detalle adicional para explicar observaciones hechas en el nivel de resumen. Cuando se pasa a un nivel más fino, como por ejemplo de mes a semana, se conoce con drill-down. Por el contrario, el pasar de un nivel más fino a otro más global, se conoce como roll-up.

Agregaciones

Para Wolf[00], es la actividad de combinar datos desde múltiples tablas para formar una unidad de información más compleja. Para [Mads96], las agregaciones consisten en la totalización y almacenamiento de datos disponibles en tablas de hechos para mejorar las búsquedas desarrolladas por el usuario final. Estrictamente hablando, las define como *“una totalización precalculada de filas en una tabla de hechos a través de una o más jerarquías de dimensión”*. En una base de datos multidimensional se puede dar tres casos: 1- sin agregaciones, cuando el volumen de datos es pequeño, 2- selectivo, cuando la mayoría de las bases de datos de apoyo a las decisiones son suficientemente grandes

para requerir alguna agregación selectiva, y 3- completa, cuando se requiere una ejecución óptima, es decir que una consulta lea un número mínimo de filas para devolver una respuesta.

La manera más eficiente para la implementación de las agregaciones es la construcción de una tabla separada para cada agregación creada, lo que permite hacer y remover agregaciones independientes y agilizar su mantenimiento.

2.3 Metodologías de desarrollo

Existen varios modelos de implementación y metodologías para el desarrollo de un DD. La utilización de alguno de ellos dependerá de los alcances y limitaciones del proyecto que se desee llevar a cabo.

2.3.1 Modelos de implementación

Esquema de Estrella

Éste consiste en una tabla central construida con llaves foráneas provenientes de un conjunto de tablas llamadas dimensiones. Las llaves foráneas en conjunto, corresponden a la llave primaria de dicha tabla central, la cual, conforme a la terminología de base de datos multidimensional, es llamada también tabla de hechos. Según [Cort99], cada una de las tablas de dimensión tiene una sola llave primaria que corresponde exactamente a uno de los componentes de la llave primaria compuesta en la tabla de hechos.

La tabla de hechos contiene, además de las llaves foráneas, una o más medidas (hechos numéricos) que ocurren como resultado de combinar las dimensiones que definen cada registro. Las tablas de dimensiones contienen información descriptiva.

Esquema Constelación de Hechos

Se constituye de los elementos del esquema estrella, más tablas de agregación para cada nivel de jerarquía de las dimensiones descriptivas.

Esquema Copo de Nieve

Este esquema, al igual que el anterior, es una variación del esquema estrella. Su diferencia radica en que trata de normalizar las dimensiones, además de las tablas de agregación.

2.3.2 Metodologías de desarrollo

2.3.2.1 Metodología de Poe

[Poe98] presenta una visión general para la construcción de un DD por medio de una guía de desarrollo, la cual en primer término plantea una pregunta a primera vista muy obvia: ¿porqué se quiere un Depósito de Datos? La respuesta influirá en los alcances del desarrollo, afectará la calidad de los datos y la arquitectura de los sistemas que se elijan para llevarlo a cabo. Lo que busca esta pregunta es tener claro el objetivo que se persigue.

El segundo punto es entender cómo el DD encaja dentro de la empresa. Para cumplir este punto se debe revisar el entorno automatizado de la empresa, mediante la determinación de las redes instaladas de que dispone, las herramientas de consulta que utiliza, el sistema operacional en producción, las plataformas de uso, entre otros. Se debe tener presente que el fin del DD es proporcionar información para la toma de decisiones, por lo que requerirá una nueva forma de pensar con respecto a la información. En resumen, es necesaria la identificación exacta de los requerimientos, por medio de una

revisión integral de la empresa, tanto en su infraestructura tecnológica como en las necesidades de los usuarios.

[Poe98] en su tercer punto, hace una llamada de atención bastante fuerte: según su criterio, se debe eliminar la idea de que un DD puede implementarse en uno o dos meses, subestimando la cantidad de análisis que se requiere y la potencial complejidad de trabajar con tecnologías nuevas a la vez. El tiempo de desarrollo dependerá de los recursos disponibles, las plataformas sobre las que se trabajará y de la experiencia que se tenga.

El punto cuarto se refiere a la responsabilidad del equipo desarrollador con respecto a las actividades que se deben completar para llevar a cabo un proyecto exitoso. Dichas actividades son: 1- el análisis de los datos y su modelaje, 2- la definición de los requerimientos de los datos, 3- hallar y correlacionar las fuentes de información que se requerirán, 4- reconciliar los conflictos existentes entre los datos de diferentes fuentes, 5- transformar, integrar y calcular, según sean las especificaciones del proyecto y 6- diseñar la base de datos para acceso de sólo lectura.

Un quinto punto se refiere a la educación del usuario con respecto a la diferencia entre los datos transaccionales y los de soporte a la toma de decisiones. En la DCF se manejan datos para la toma de decisiones, sin embargo no se tiene claro aún la diferencia entre los dos tipos de datos.

El sexto punto de la metodología tiene que ver con el presupuesto para el proyecto. Dicho presupuesto debe contemplar la adquisición de herramientas costosas para la automatización de los procesos de transformación, extracción y generación de código, la contratación de personal especializado y experimentado, así como la compra de bases de datos diseñadas para el procesamiento de los datos.

El desarrollo o compra de las herramientas Front-End y de acceso a los datos con base en las habilidades y necesidades de los usuarios, es el sétimo punto. La elección de la herramienta de consulta puede afectar significativamente la aceptación del DD por parte de los usuarios hacia los cuales va dirigido. Ser realista sobre las necesidades y condiciones de los usuarios con respecto a la parte del sistema que está en contacto con ellos, es de vital importancia.

En el octavo y último punto, [Poe98] recomienda buscar toda la experiencia necesaria, ya que se debe lidiar con tecnología nueva y de rápida evolución. Es importante, a fin de contar con experiencia de primera mano, tener contactos con empresas que ya hallan implantado un DD con éxito. Indica que el equipo de expertos debe estar compuesto por: 1- analistas de datos, diseño y administración de base de datos, 2- analistas de negocios con experiencia en trabajo con usuarios, 3- personas con experiencia en los actuales sistemas de información de la empresa, 4- administradores de proyectos con experiencia en el desarrollo de sistemas de apoyo a las decisiones y 5- programadores y diseñadores de interfaces gráficas de usuario (GUI.)

2.3.2.2 Metodología de Gill

Según [Gill96] el desarrollo de un DD requiere de 7 fases: planeación, requerimientos, análisis, diseño, construcción, prueba y desarrollo.

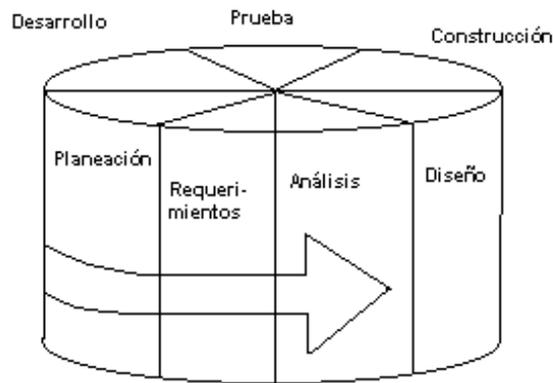


FIGURA 3: Ciclo de desarrollo de Software utilizado para el DD según Gill

La planeación la subdivide, a su vez, en varias etapas: selección de la estrategia de implementación, la selección de la metodología de desarrollo, selección del ámbito de implementación, selección del enfoque arquitectónico, desarrollo de un programa y presupuesto, desarrollo del escenario empresarial y recopilación de metadatos.

En la estrategia de implementación se explican tres enfoques. De arriba hacia abajo (de lo general a lo particular), donde primero se identifican los requerimientos empresariales; recomendada cuando la tecnología está madura y están bien entendidos los problemas a resolver. De abajo hacia arriba (de lo particular a lo general), donde se comienza con experimentos y prototipos; recomendada para desarrollos en las primeras etapas de madurez tecnológica o en mercados de datos. Una combinación de ambos, donde se explota la naturaleza planeada y estratégica del primer enfoque y la rápida implementación del segundo.

En la selección de la metodología de desarrollo se puede utilizar el clásico método de análisis y diseño estructurado, o el método de desarrollo espiral, donde el énfasis está

en la velocidad y la culminación. Este último método se utiliza para la generación de sistemas con intervalos cortos entre versiones sucesivas.

Con respecto a la etapa de requerimientos, ésta consiste en una especificación de las funciones que se obtendrán del DD, los cuales son conducidos tanto por el entorno empresarial como por la tecnología.

En el análisis, se toman los requerimientos y se transforman en especificaciones de apoyo al diseño, por medio de modelos lógicos. Los modelos son transformados en modelos físicos en la etapa de diseño y éstos, su vez, en programas de implementación.

La fase de construcción es responsable de implementar físicamente los diseños desarrollados durante la fase anterior. Una vez finalizada la construcción de la aplicación, ésta debe ser sometida a una serie de pruebas para comprobar su funcionalidad.

2.3.2.3 Metodología de Devlin

[Dev196] determina 4 fases: evaluación, definición, planificación e iniciación del piloto. Indica que un DD requiere de un tiempo de desarrollo máximo de 5 meses, en contraposición a la mayoría de los autores que vaticinan un largo y tortuoso camino para un proyecto de DD exitoso. La ventaja de esta metodología radica en que es bastante práctica y define claramente las reglas del juego, sin profundizar en esquemas teóricos y dar grandes rodeos conceptuales.

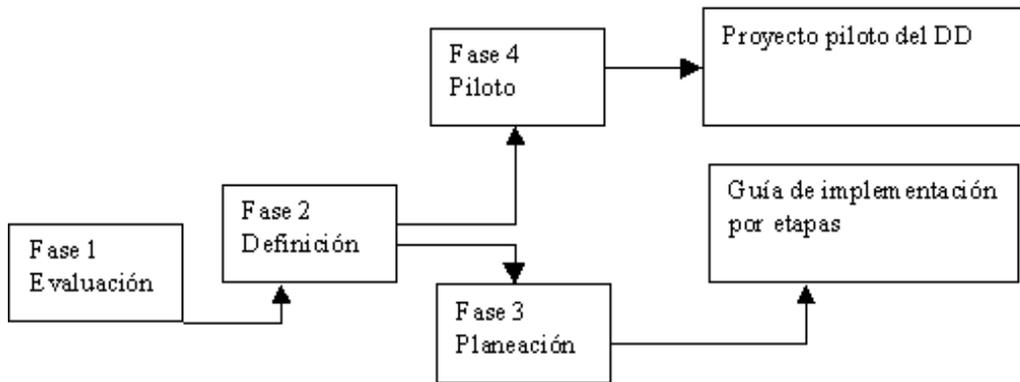


FIGURA 4: Fases del proyecto de iniciación del DD según Devlin

En la primera fase se realiza el estudio de prefactibilidad, la recolección de requerimientos y su análisis, el diseño conceptual inicial y las recomendaciones del caso. En la segunda fase se contemplan los aspectos de preparación y capacitación del equipo, revisión de datos y la construcción de la arquitectura de los datos (definición de alto nivel del modelo empresarial, catálogo inicial del DD y diseño de alto nivel del DD.) En la tercera fase se define la planeación del proceso integral de entrega del DD, el desarrollo de la guía de implementación por etapas y se realizan una serie de recomendaciones finales. La cuarta fase, que se desarrolla de manera paralela a la tercera, busca definir y evaluar el proyecto piloto por medio de su diseño, evaluación de sus alcances y las recomendaciones del caso.

2.3.2.4 Metodología de Ebel

[Ebel98] expone su proceso iterativo de desarrollo basado en los objetivos, necesidades y beneficios que busca el negocio, con tres actividades primarias: planeamiento (determinación de requerimientos, selección de tecnología y modelaje lógico), diseño e implementación (diseño físico e implementación, proceso de

adquisición de datos y procesos de entrega) y apoyo y resultados (operaciones, capacidad, planeamiento, ajuste y auditoria de resultados.)

En primer instancia presenta una visión simplificada, pero que implica una nueva perspectiva del negocio y generalmente requiere de una redefinición de los procesos productivos. Plantea de manera reiterada que el factor tecnológico es el menos importante y es lo último en lo que se debe pensar, ya que el enfoque de un DD no es solo almacenar información, sino la utilización de ésta de manera eficaz para obtener resultados en el negocio.

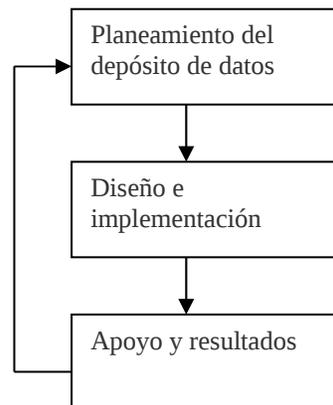


FIGURA 5: Proceso iterativo del DD según Ebel

Existen varios elementos clave para que un DD pueda ser exitoso: 1- el modelo de datos, el cual debe contener la información para responder a cada una de las preguntas del negocio (contiene la definición de datos, las estructuras dimensionales y los datos transaccionales); 2- los metadatos, los cuales son un repositorio que coadyuva en las labores de administración y entendimiento de la información del sistema; 3- la arquitectura de procesamiento, que identifica las entradas y salidas del sistema (contempla los sistemas originarios; la depuración, carga, actualización y borrado de datos; las estructuras y entradas manuales; las herramientas ad-hoc; consultas sobre

demanda; constructores de reportes; minería de datos; réplicas de datos) y 4- los procesos de calidad.

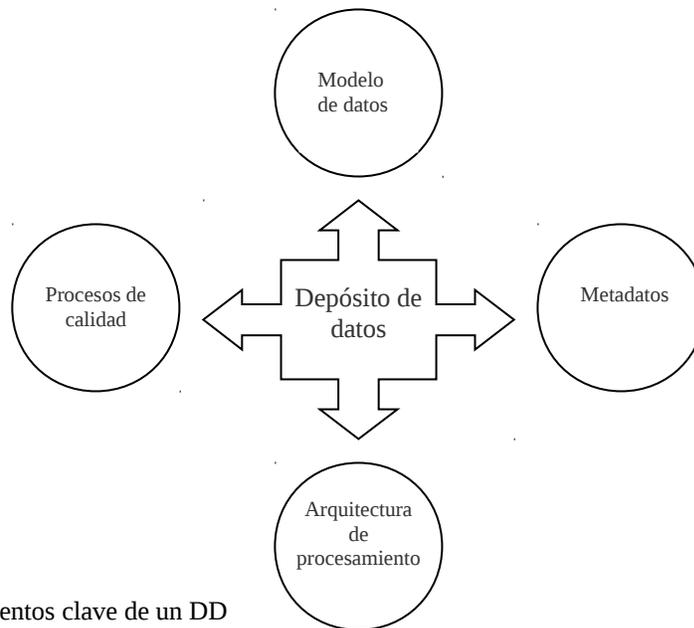


FIGURA 6: Elementos clave de un DD

2.3.2.5 Metodología de Corey

Una metodología novedosa la plantea [Core97] con su ciclo de vida del desarrollo de sistemas hacia atrás, o “la teoría de la evolución”. Para este autor, no se debe realizar una especificación de requerimientos, en contraposición del resto de los autores, ya que, según su postura, los usuarios finales no conocen todavía como analizarán los datos. Es necesario que usuarios clave vean los datos sin limpiar. Se trata de un proceso de crecimiento, donde desarrolladores y usuarios trabajan conjuntamente determinando las herramientas apropiadas tanto para el desarrollo como para la puesta en marcha del DD.

En la siguiente tabla, [Core97] hace una comparación entre las metodologías del ciclo de vida del sistema operacional frente al DD.

Sistema operacional	Teoría de la evolución
1. Definición de los recursos de usuario	1. Escuchar al usuario.
2. Análisis y planes basados en definiciones.	2. Implementar una primera visión comercial del Datawarehouse.
3. Modelizar.	3. Escuchar al usuario.
4. Diseño físico.	4. Desarrollar mecanismos de soporte analítico.
5. Programación.	5. Escuchar al usuario.
6. Asegurar la calidad y la aceptación del usuario.	6. Hacer más visiones comerciales.
7. Implementación.	7. Escuchar al usuario.
	8. Ir al paso 3.

Para este autor, lo primero que se debe hacer es crear el equipo del proyecto el cual debe estar en un continuo diálogo con el usuario, a fin de identificar sus preferencias acerca de las formas en que quiere representar los datos, tal y como se muestra en la figura 7.

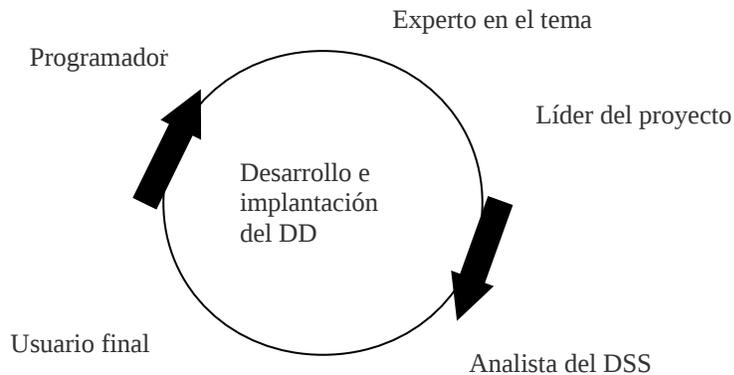


FIGURA 7: Equipo de proyecto da DD según Corey

2.3 Mercado de datos

Un DD contiene información detallada y datos integrados e históricos comunes a toda la organización. Cada departamento de una empresa utilizará el DD para sus análisis internos y es probable que una gran cantidad de datos no sean de su interés, ya que solamente querrá trabajar con un rango definido. Por otro lado, en la medida en que se incrementa el volumen de datos, se aumentará a su vez el costo de procesamiento.

Un Mercado de datos (MD) o “datamart” lo define [Corey97] como “*bases de datos multidimensionales orientadas a una materia específica*”. [Gonza98] lo define como una extensión natural del DD, siendo una base de datos con información de interés para un determinado sector de la organización, contribuyendo con los Sistemas de Apoyo a la Toma de Decisiones de los departamentos con una arquitectura fundamentada en la del DD (figura 8).

[Wolf00], por su parte lo define como

un conjunto de hechos y datos organizados para soporte decisional basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo sólo sentido para el personal de ese departamento y sus datos no tienen porque tener las mismas fuentes que los de otro Datamart.

El modelo del MD está definido por la forma como el usuario necesita ver la información y como quiere que se le presente. Posee las mismas características de un DD pero a un nivel más específico, ya que contiene diferentes combinaciones y selecciones de los datos que se encuentran en el DD y ofrece una mayor personalización de los datos del departamento, permitiendo un manejo más eficiente de la información histórica, ejecución de procesamiento independiente del resto de los departamentos y un costo de

almacenamiento y procesamiento inferior. La figura 8 esquematiza al MD dentro del contexto del DD.

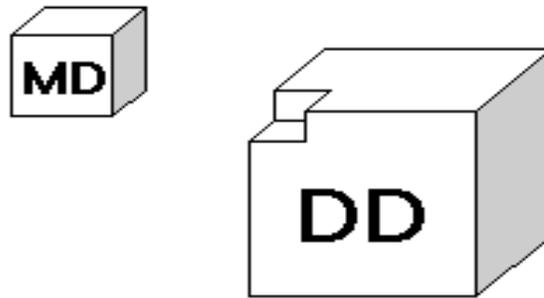


FIGURA 8: Mercado de datos

El MD propuesto vendría a ser parte integral de un futuro DD para la Dirección General de Aduanas y podría ser su punto de partida, si se considera la posición de [Wolf00], quién expone que la construcción de mercados de datos constituye la construcción incremental de un DD total. Dicha construcción incremental es una aproximación práctica para construir un DD a escala empresarial de forma evolutiva. Con base en esta posición, se podría desarrollar un DD como lo muestra la figura 9:

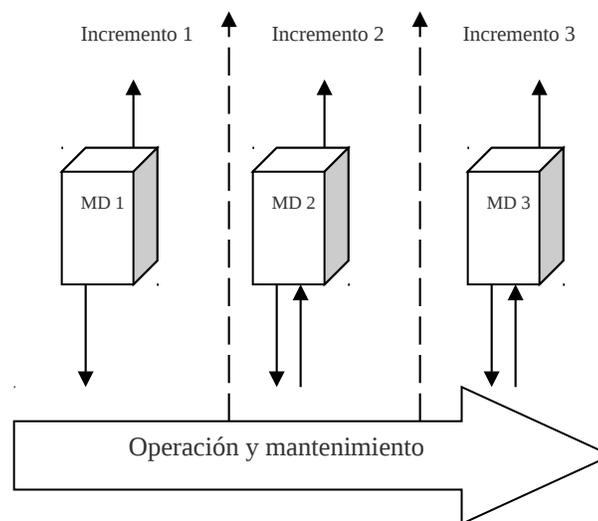


FIGURA 9: Esquema de desarrollo incremental de un DD según Wolf

Partiendo de esta base, un MD es una excelente plataforma sobre la cual se podrá desarrollar un futuro DD.

2.4 Procesamiento analítico en línea (OLAP)

Los DD y los MD almacenan información para su posterior búsqueda y análisis, a diferencia de los sistemas transaccionales, que la almacenan para operación y control. Es por ello que se requiere de la tecnología de Procesamiento Analítico en Línea (OLAP) encargada de realizar dicho análisis.

Las herramientas OLAP permiten el análisis de datos en forma interactiva. [Cox00] lo define como la

tecnología que permite que la información sea analizada de forma rápida, amigable e interactiva y que facilite la construcción de modelos cuantitativos complejos.

Es una característica del software que permite al usuario organizar, visualizar y resumir información desde distintas perspectivas de manera rápida e interactiva.

El OLAP incluye principalmente la agregación de grandes cantidades de datos dispersos y puede abarcar millones de datos con relaciones complejas. Su objetivo es contribuir al análisis de estas relaciones mediante la facilitación de la búsqueda de patrones, tendencias y excepciones. Debe soportar procesamiento lógico y estadístico de los resultados sin que el usuario tenga que programar, además de incluir los requerimientos de seguridad para la confidencialidad y las actualizaciones concurrentes.

El OLAP logra proveer una visión multidimensional de los datos por medio de la consulta interactiva y el análisis de datos. Permite la profundización sobre niveles cada vez más detallados o el ascenso a distintos niveles de resumen. Ofrece opciones de

medición de datos numéricos a través de múltiples dimensiones, creando resúmenes y consolidaciones. Debe responder con rapidez a las consultas, de forma tal que el proceso de análisis no sea interrumpido y la información para la toma de decisiones no pierda valor.

Al estar vinculadas las distintas dimensiones, una instancia de su intersección produce una celda, la cual contiene los valores de medida intersecados de las dimensiones involucradas. Una celda es un punto que corresponde a la intersección definida por el valor de cada una de las dimensiones del vector multidimensional.

En general, estos sistemas OLAP deben:

- Soportar requerimientos complejos de análisis
- Analizar datos desde diferentes perspectivas
- Soportar análisis complejos contra un volumen ingente de datos

MOLAP, ROLAP, HOLAP

Estos tres términos se refieren a la forma de almacenamiento de los datos. Ahora bien, la decisión se debe tomar de acuerdo a si se quiere que el almacenamiento refleje los análisis que se harán (multidimensional), o los datos se almacenarán sin tomar en cuenta para qué serán utilizados (de forma relacional).

El primero de los casos se denomina MOLAP (Multidimensional On-Line Analytical Processing) y aquí los datos se almacenan en forma lógica en arreglos matriciales con estructura multidimensional.

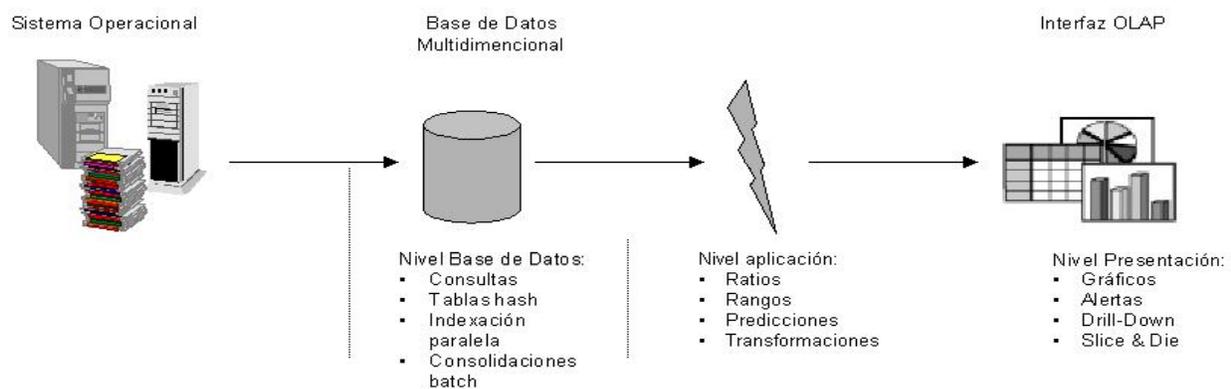
El sistema MOLAP utiliza una arquitectura de dos niveles: Las bases de datos multidimensionales y el motor analítico (figura 10).

- La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.
- El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona un interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

Aquí los datos son redundantes, por lo que se requiere mucha capacidad de almacenamiento y puede contener totales precalculados. La estructura permite la reservación de espacio para todas las combinaciones de todos los posibles valores de todas las dimensiones de cada una de las variables, incluyendo los valores de dimensión que representan acumulados. Es decir, un sistema MOLAP contiene precalculados (almacenados) los resultados de todas las posibles consultas a la base de datos.

MOLAP consigue consultas muy rápidas a costa de mayores necesidades de almacenamiento, y retardos en las modificaciones (que no deberían producirse salvo excepcionalmente), y largos procesos “batch” de carga y cálculo de acumulados.

FIGURA 10: Multidimensional On-Line Analytical Processing



El segundo caso denominado ROLAP (Relational On-Line Analytical Processing), es la arquitectura de base de datos multidimensional en la que los datos se encuentran almacenados en una base de datos relacional, la cual tiene forma de estrella (o copo de nieve). En ROLAP, en principio la base de datos sólo almacena información relativa a los datos en detalle, evitando acumulados (redundancia). Aquí se crea una capa semántica de metadatos en una base de datos relacional, que permite presentar los datos en término de dimensiones. Su arquitectura es de tres niveles: la base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica (figura 11).

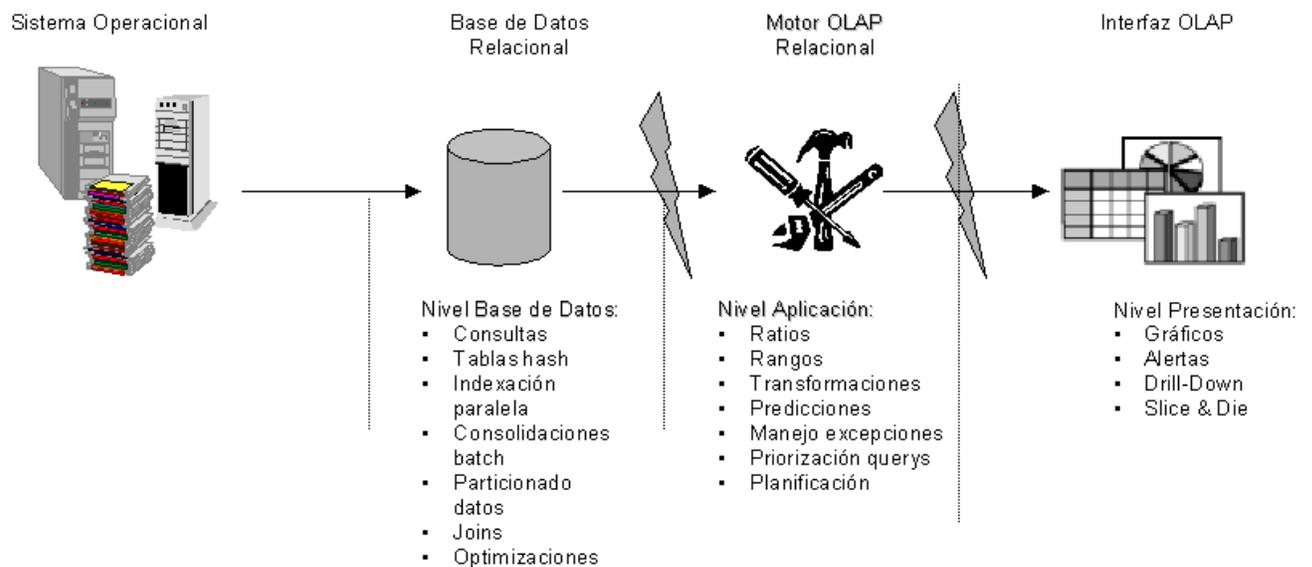


FIGURA 11: Esquema del procesamiento analítico en línea relacional

- El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención de los datos.
- El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

- El motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP.

En ROLAP, al contener sólo las combinaciones de valores de dimensión que representan detalle, es decir, al no haber redundancia, el archivo de base de datos es pequeño. Los procesos “batch” de carga son rápidos (ya que no se requiere agregación), sin embargo, las consultas pueden ser muy lentas, por lo que se aplica la solución de tener al menos algunas consultas precalculadas.

Por último, el HOLAP se refiere al sistema mixto o híbrido, donde la configuración típica es un motor multidimensional basado en una tecnología de almacenamiento relacional. [Gill96] dice que en esta configuración las consultas e información frecuentes o de cómputo intensivo se calculan *a priori* y se almacenan en el depósito multidimensional.

Capítulo 3. Descripción del sistema actual

El Servicio Nacional de Aduanas cuenta con siete aduanas para todas las operaciones de importación y exportación de mercancías en el país (véase figura 12).

El ente rector es la Dirección General de Aduanas, que es una dependencia del Ministerio de Hacienda.

Las siete aduanas se distribuyen de la siguiente forma:

- o Dos aduanas de frontera terrestre: la Aduana Paso Canoas y la Aduana Peñas Blancas.
- o Dos aduanas de puerto marítimas: la Aduana Limón y la Aduana Caldera.
- o Dos aduanas aeroportuarias: la Aduana Santamaría y la Aduana Anexión
- o Una aduana interna: la Aduana Central.

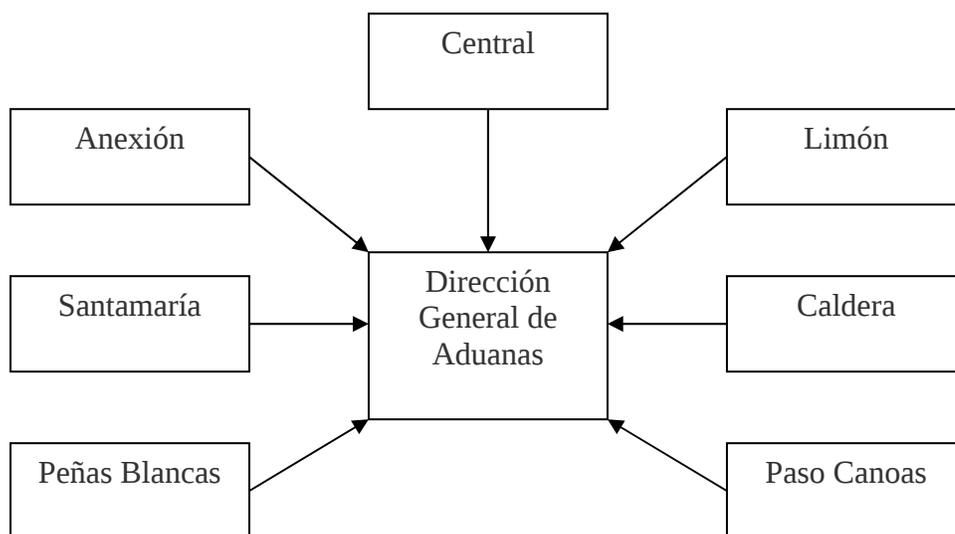


FIGURA 12: Esquema general del Sistema Nacional de Aduanas

Toda información transaccional que se produce en las aduanas, es procesada de forma descentralizada por el SIA. Cada aduana tiene sus propias bases de datos, la información que diariamente recibe por medio de las transmisiones electrónicas o el trámite manual de declaraciones aduaneras, se mantiene bajo su custodia y administración.

Ninguna aduana puede establecer controles, realizar operaciones, ni compartir datos e información de sus sistemas transaccionales con las otras aduanas debido a que sus bases de datos no se encuentran interconectadas.

Las transacciones que a través de las declaraciones aduaneras cada una tramita, están identificadas de manera única conforme al código de la aduana y el número asignado por el sistema.

3.1 Rutinas de extracción y generación de datos

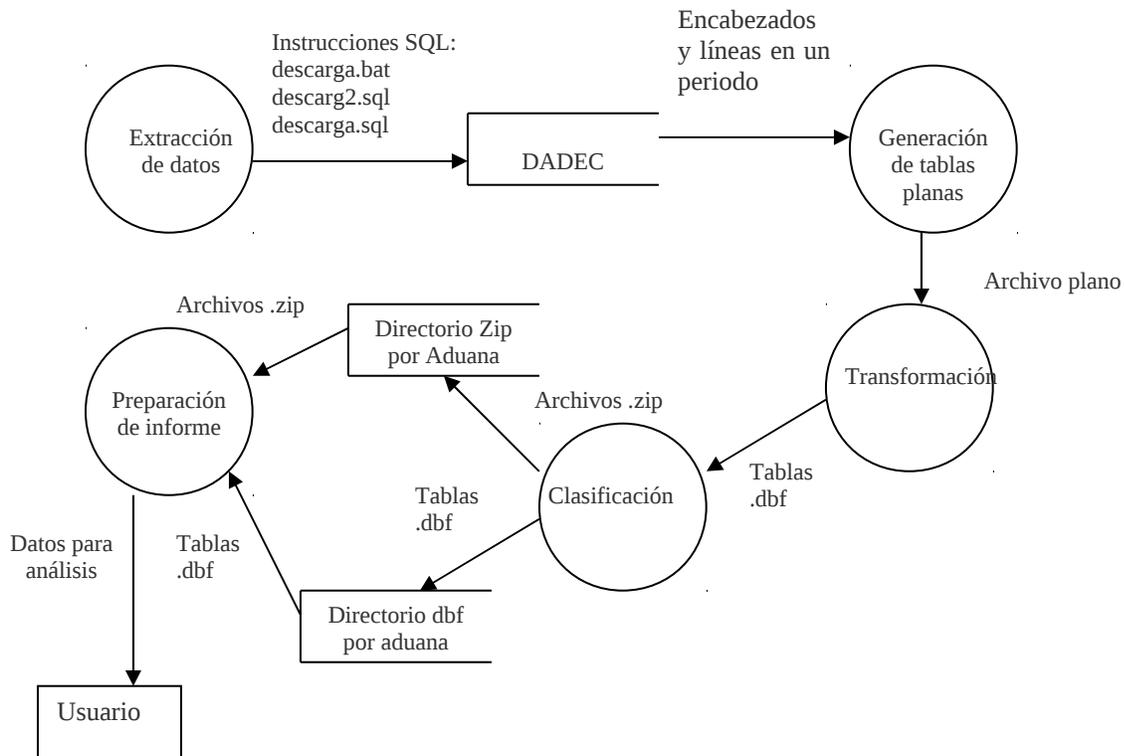


FIGURA 13: Procesos de extracción y generación de datos

En este conjunto de procesos, se manipulan los datos que se extraen del SIA a fin de obtener información para el análisis de casos. La figura 13 ilustra el proceso general.

3.1.1 Rutinas de extracción

Los datos que utiliza el Departamento de Estadística de la División de Estadística Registro y Divulgación para satisfacer las solicitudes de información de los usuarios, provienen de las tablas planas generadas mediante archivos de procesamiento por lotes o “batch”, ejecutados por funcionarios en las aduanas o desde la Dirección General de Aduanas (en San José). Estos archivos invocan otros programas o rutinas que contienen instrucciones SQL, que son las responsables de extraer de las bases de datos ubicadas en las aduanas, los valores de las variables previamente seleccionadas y registrados en un periodo determinado.

Las bases de datos incorporadas en los computadores de cada aduana, utilizan su red interna, la cual permite, a través de los programas del SIA, automatizar la gran mayoría de las operaciones y procesos aduaneros requeridos por los usuarios y la administración misma. Los datos contenidos en dichas bases de datos son controlados exclusivamente por cada aduana, dado que el sistema de información funciona descentralizadamente.

La ejecución de los archivos batch, ya sea por funcionarios autorizados en las aduanas o por funcionarios del Departamento de Estadística, dependerá básicamente de la distancia entre dicho Departamento y la aduana que tiene los datos que se quieren extraer. También entran en consideración las dificultades del enlace telefónico y la calidad de éste para permitir la transmisión de los datos extraídos. Si la aduana está fuera de la cobertura

de la red WAN (tecnología Frame Relay) y el Departamento de Estadística requiere los datos, será necesaria la generación de éstos en la aduana misma y luego su envío mediante disquete.

Los archivos batch referidos y su función son los siguientes:

- Descarga.bat

Este archivo invoca una rutina llamada descarg2.sql que contiene instrucciones SQL y se encarga de extraer de la base de datos llamada DADEC, los datos de las importaciones tramitadas durante un periodo que explícitamente se le indica (normalmente un mes). Ésta información está compuesta por un encabezado y un conjunto de líneas asociadas a éste. En dicho encabezado se encuentran todos los datos generales, no repetitivos de la declaración aduanera y aquellos elementos de información que identifican en forma única cada transacción, por ejemplo el número de la declaración, la aduana a la que pertenece o la agencia aduanera, son datos del encabezado.

Dependiente de éste, es el grupo de líneas que identifica el detalle de las mercancías que ampara la declaración aduanera.

Los elementos de información que se extraen del detalle de la declaración aduanera, esta determinado por las variables o campos que en el mismo programa de extracción se especifican, siendo éstos un subconjunto de los elementos de detalle contenido en cada declaración, no la totalidad de los campos de detalle, dado que no todos son requeridos en las consultas de los usuarios o corresponden a atributos de otras bases de datos de procesos distintos a los de importación.

Las tablas que contienen la información generada, incluyen solo los datos relacionados a la importación definitiva y, para no duplicar la información

extraída de meses y años anteriores, debe corresponder exactamente al periodo que se especifique.

La información generada se transfiere a un archivo plano cuyo nombre es definido por el usuario y debe corresponder con el contenido de la información que se extrae. Usualmente este nombre se construye a partir de los datos que identifican la aduana, el mes y el año en que se extraen. No hay una normalización oficial para la especificación del nombre de estos archivos. Esta situación implica el riesgo de dificultar y hasta impedir la selección del archivo correcto cuando se van a transformar los datos a otro formato de trabajo para atender consultas de los usuarios.

- Descarg2.sql

Este archivo se utiliza para solicitar al usuario la especificación del periodo que desea consultar. Posteriormente invoca la rutina descarga.sql, que finalmente se encarga de extraer los datos de la base de datos DADEC.

3.1.2 Proceso de transformación y validación

Concluida la etapa de generación de archivos planos se continúa con la fase de transformación de formato. Dichos archivos planos conforme se crean se van grabando en un subdirectorio específico de la red de la aduana o de la Dirección General de Aduanas, según sea donde se ejecuten los programas de generación.

La fase de transformación, consiste en convertir cada archivo plano resultante de la extracción aplicada a la base datos DADEC al formato dbf, por medio de un procedimiento escrito en lenguaje Fox. La estructura de la tabla resultante concuerda con

la que previamente se ha establecido según el tipo, tamaño y ubicación de cada campo de datos de la estructura columnar de la que se origina.

Los archivos con estructura dbf generados a partir de los archivos planos, son almacenados en directorios determinados por los códigos de las aduanas a las que se asocian y por el tipo de trámite al que corresponden. Por ejemplo, si el nombre del archivo dbf generado en la Aduana Central es CENT0199.dbf y contiene información de importaciones del mes de enero del año 1999, se almacenará en el subdirectorio importaciones del directorio denominado “Aduana Central”.

Semejante a lo que se hace con los nombres de los archivos de extensión dat, los nombres de los archivos dbf se crean según la aduana, mes y año a los que corresponden los datos.

Los nombres que se asignan a los archivos, no siguen ninguna norma oficial o automatizada de generación, sino que están sujetos al criterio del que los crea, con el riesgo de que se puede confundir la información o perderla por la inadecuada identificación.

3.1.3 Proceso de clasificación

El método o procedimiento que se utiliza para clasificar o distribuir los archivos según su tipo y función, no está definido formalmente.

El funcionario encargado de recopilar y manipular los datos que se extraen de las aduanas utiliza su propio criterio de clasificación y transformación, ya sea que cambie los datos planos a archivos dbf y los someta posteriormente a un proceso de compresión, convirtiéndolos al formato zip, o los deje como dbf. La diferencia está determinada por la ubicación en la que éstos se almacenarán.

No existe una estructura de base datos formalmente establecida para el soporte básico de la información transformada, que garantice la integridad y la consistencia de los datos durante el almacenamiento y respaldo, posterior a dicha transformación. Es común encontrar entonces, información agrupada que ha sido almacenada separadamente en otros directorios de la red o en la misma estación de trabajo del funcionario encargado de la transformación, pero desligada funcional y estructuralmente una con otra, aumentando con ello el riesgo, por parte de los funcionarios encargados de dar respuesta a las consultas, de utilizar datos no actualizados o erróneos, además de dificultar las tareas de mantenimiento y validación.

3.1.4 Proceso de almacenamiento y restauración actuales

El almacenamiento de los datos de las aduanas en el Departamento de Estadística, ya sea que estén disponibles para uso inmediato o comprimidos, se hace utilizando dos criterios: 1- la antigüedad y uso que se demande de ellos en un momento dado y 2- si los archivos corresponden a los datos del año vigente.

En el primer caso, si los datos transformados son de años pasados y no se solicitan con frecuencia o no se están requiriendo en algún proceso de investigación, se almacenan en la carpeta de archivos comprimidos según el tipo de transacción al que correspondan y la aduana a los que se asocian. Si en un momento dado se requieren, se descomprimen y se ponen a disposición del analista que los procesará desde la carpeta de la aduana respectiva.

En el segundo caso, permanecen descomprimidos y disponibles en formato dbf, en espera de ser usados. La herramienta de administración y manipulación de datos que

usa el Departamento de Estadística es el administrador de base de datos Fox, versión 2.5 para DOS.

La idea de separar los datos en directorios según la aduana a la que pertenecen y ubicar en archivos comprimidos aquellos que no sean de uso frecuente, pretende facilitar las tareas de ubicación y manipulación de archivos. No obstante, como se muestra en la figura 14, que esquematiza el almacenamiento de estos datos, si no existe un procedimiento formalmente establecido para dicho agrupamiento y algún sistema automatizado que se encargue de la administración del almacenamiento, identificación y asignación de nombres, no habrá confianza ni seguridad de que la información almacenada sea completa y consistente. Para ilustrar esto veamos algunos casos:

- o Según se muestra en la figura 14, en el directorio de la aduana Caldera falta el archivo *ecal0899.dbf* de exportación correspondiente al mes de agosto. También se muestra como en el directorio ZIP falta el archivo *eelim97.zip* correspondiente a la aduana Limón.
- o Para almacenar la información de importación de la Aduana Santamaría del año 1995, se usa el formato de compresión zip, mientras que para la Aduana de Limón se emplea el formato dbf.
- o En el directorio para los datos de importación de la Aduana Santamaría se usa “SANT” para indicar que se trata de esa aduana (*95_sant.zip*), mientras que para sus mismos datos de exportación se usa el código 05 para identificarla (*95_05.zip*).

En conclusión, el esquema usado para nombrar y almacenar la información duplica innecesariamente los datos, crea inconsistencias y genera confusiones sobre cuales son los archivos que se deben considerar para resolver las consultas.

ESQUEMA DE ALMACENAMIENTO DE DATOS Y
NOMBRES UTILIZADOS EN ESTADÍSTICA DE LA DGA

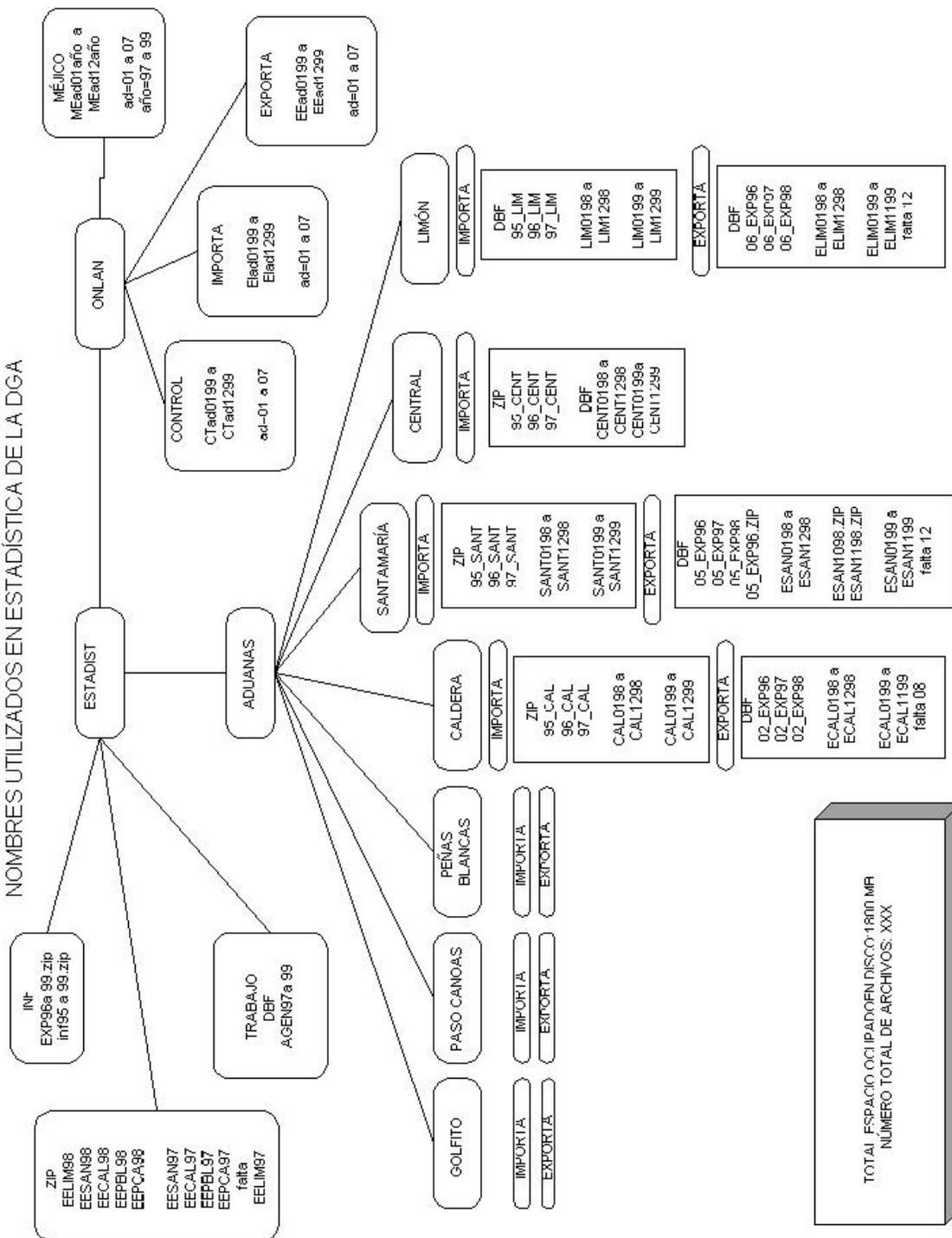


FIGURA 14: Esquema de almacenamiento en el Departamento de Estadística

3.1.5 Procedimiento del Departamento de Estadística para la preparación de la información solicitada por los usuarios.

Los usuarios internos o externos de la institución, que requieren información de las transacciones realizadas en las aduanas, con regularidad hacen la solicitud formal al Departamento de Estadística.

Este departamento, utiliza únicamente las herramientas de manejo de tablas de Fox para construir la estructura de datos según la solicitud planteada, y mediante algún procedimiento manual, coloca la información en las tablas de salida, o si es del caso, utiliza el sistema generador de reportes para algunos de los resúmenes que requieren. También, según la conveniencia del usuario, convierte los resultados obtenidos a una hoja electrónica y los pone a disposición del solicitante, normalmente en disquetes y en la cantidad o formato de compresión que requieran.

En el proceso de preparación de la información, el analista consulta y retrae de los directorios necesarios, los archivos relacionados con la información solicitada, procediendo luego a hacer el análisis procedimental que le proporcionará los resultados buscados. Cuando se trata de responder consultas que involucran datos de varios años, el análisis mencionado implica la ejecución de laboriosas y delicadas maniobras de unión, selección, proyección y filtrado de tablas y registros, que consumen gran cantidad de tiempo y requieren la manipulación directa de los datos. Esto se debe a que los datos de años distintos se almacenan en archivos y directorios separados.

Si los archivos que debe utilizar son de periodos que por su antigüedad están comprimidos, debe proceder primero a su restauración y ajuste, antes de emplearlos en la construcción de consultas.

3.2 Infraestructura de apoyo (equipos, plataformas y red)

La Dirección General de Aduanas cuenta con una red que da servicio a todas las Divisiones y Departamentos de la Institución.

El Departamento de Estadística Divulgación y Registro, como ya se ha dicho, es el encargado de recibir y administrar la información proveniente de las aduanas. Utiliza para ello tres microcomputadoras con capacidad de almacenamiento 30 Gb. cada una, pero sólo una de ellas cuenta con MODEM para la comunicación entre las aduanas más alejadas, que son la Aduana Paso Canoas, la Aduana Peñas Blancas y la Aduana Caldera, situadas en las fronteras terrestres y marítimas.

Este departamento dispone también de espacio de disco en la red de la institución, aunque con mucho menos capacidad.

Capítulo 4. Requerimientos de información y usuarios del

Mercado de Datos

4.1 Entorno general

Es necesario conocer de forma general la estructura organizacional de la División de Control y Fiscalización (DCF), así como el perfil profesional de los funcionarios que conforman cada uno de sus departamentos. Esto con el fin de tener una idea de la naturaleza de los esfuerzos que han tenido que hacer para manipular la información, solicitada al Departamento de Estadística.

La División esta constituida por los departamentos Interno, Externo, Planificación, Legal y Denuncias.

Cada departamento reúne un grupo de profesionales de distintas especialidades, entre los cuales se encuentran contadores públicos, abogados, administradores de empresas, administradores públicos, así como profesionales de otras ciencias económicas y de tecnología de la información.

Obviamente, las funciones de dichos profesionales están orientadas hacia la consecución de los objetivos de cada departamento, de forma tal que, según sean esos objetivos, las necesidades de información estarán delimitadas por los requerimientos de cada caso de estudio o investigación asignado.

Generalmente los funcionarios de los departamentos manipulan los datos usando hojas electrónicas. Deben resumir, agrupar, filtrar, clasificar los datos, y trabajar con innumerables funciones de contabilización que a menudo son complejas y con las cuales deben hacer sus comparaciones y validaciones.

Conforme a la subdivisión departamental señalada y los perfiles mencionados se deberán definir los requerimientos básicos de información de cada departamento, así como las operaciones de manipulación que habrán de realizar para completar su trabajo.

4.2 Determinación de usuarios y sus requerimientos de información

Se adoptaron dos estrategias para la definición de los usuarios potenciales de la aplicación y su requerimientos de información:

- A. La aplicación de una entrevista a todo el universo de casos compuesto por todos los funcionarios de la DCF (30), orientada a establecer las necesidades de información de todos los funcionarios de cada departamento, desde el punto de vista de sus funciones y conforme a los recursos disponibles para el procesamiento de la información (véase entrevista en el Anexo B.)
- B. La revisión exhaustiva de las consultas que durante el año 2000, los departamentos de la DCF hicieron al Departamento de Estadística de la Dirección General de Aduanas. (Véase listado de las consultas hechas a estadística en el Anexo A)

4.2.1. Resultados de la entrevista

Los resultados derivados de la aplicación de la entrevista de la estrategia A se exponen seguidamente, señalando específicamente los requerimientos de información, así como las situaciones y restricciones que cada departamento ha tenido que afrontar para poder utilizarla.

4.2.1.1 Departamento Externo

Esencialmente, este departamento solicita al Departamento de Estadística (DE), información relacionada con las transacciones de importación en un periodo determinado, ya sea de uno o varios importadores (sujetos de estudio) relacionados en determinada investigación o caso.

La información que este departamento pide al DE, por lo general se refiere al máximo detalle aportado por las declaraciones aduaneras de importación asociadas al sujeto de estudio, según los datos transferidos desde los sistemas operacionales de las aduanas. Esto debido a que no hay forma de obtener información ya agrupada y resumida para el análisis. Algunos de estos datos son: números de documento, fechas, consignatarios, referencias de otros documentos de ingreso o inventario, mercancías importadas, valores, pesos, impuestos pagados, impuestos exonerados y otros datos relacionados.

La información requerida al DE también se solicita al importador o importadores en evaluación, a las agencias de aduanas relacionadas y a las aduanas que custodian las declaraciones aduaneras. Se busca en las empresas, las correspondencias, inconsistencias, diferencias y otras medidas de comparación que le permitan evaluar preliminarmente el cumplimiento de las obligaciones fiscales y aduaneras, así como el ajuste de las operaciones de importación a las disposiciones procedimentales establecidas.

La información proporcionada por el DE, antes de ser utilizada como parámetro de comparación y verificación, respecto a la suministrada por las empresas o agencias de aduana, debe ser manipulada por los auditores y analistas según los objetivos que éstos tengan definidos de acuerdo a sus casos. Dicha manipulación, como ya se señaló, se refiere a las operaciones de acomodo, selección, agrupamiento o totalización, etc., que se

hacen utilizando las hojas electrónicas. La información por lo general corresponde a tablas con los registros de las importaciones al mínimo nivel de granularidad (máximo detalle) de las declaraciones aduaneras. Este nivel está determinado, en cada declaración aduanera, por cada una de las líneas asociadas a las mercancías importadas.

4.2.1.2 Departamento Interno

La naturaleza de las funciones definidas para este departamento, determina el perfil de los funcionarios que lo conforman, orientándose éste hacia el conocimiento especializado y práctico de la administración aduanera.

Concretamente, los funcionarios tienen, dentro de sus responsabilidades, verificar el cumplimiento de la normativa aduanera. Esto desde el punto de vista del control que cada aduana debe ejercer sobre los procesos internos que se realizan, y también sobre los resultados derivados de la aplicación de las distintas guías de control y otros procedimientos establecidos en sus planes operativos.

Los procesos internos mencionados, tienen que ver directamente con las operaciones de aceptación de las declaraciones aduaneras, la liquidación y contabilización de los impuestos y el aforo de las mercancías asociadas a éstas. Asimismo, su función está relacionada conjuntamente con otros mecanismos de control y supervisión, según los lineamientos y la normativa reglamentaria aplicable, con la administración que hagan las empresas privadas (denominadas depositarios aduaneros y estacionamientos transitorios), sobre las mercancías importadas no nacionalizadas, almacenadas y custodiadas bajo su responsabilidad.

Conforme a lo indicado, los funcionarios de este departamento requieren información específica de los procesos y de la gestión administrativa que las empresas privadas mencionadas han hecho sobre las mercancías confiadas.

Para ello solicitan al DE datos concernientes a las operaciones efectuadas en un periodo determinado y su relación con las empresas encargadas de la custodia de las mercancías.

Con base en dicha información y con la suministrada por el depositario, evalúan los niveles de inventario y realizan una comprobación por importador, sobre la liquidación y cancelación de impuestos asociado a las mercancías bajo su control.

Generalmente se solicitan datos de baja granularidad, dado que, esencialmente, no tienen suficiente información sobre a cuáles mercancías, importadores o depositarios deben poner atención, según las cantidades y valores de las mercancías que se importan y los montos de impuestos a los que están sujetas.

Lo mismo que ocurre en el Departamento Externo con respecto a las herramientas de trabajo y análisis, sucede para los profesionales del Departamento Interno. Éstos deben recurrir a las hojas electrónicas para ordenar, clasificar, agrupar, resumir y seleccionar los sujetos de estudio que les interesan. No obstante, éstas tareas son arduas, complejas y llenas de riesgos para ellos, por la manipulación tan directa que deben hacer sobre los datos. Dicha manipulación aumenta los riesgos de cometer errores y por tanto, conducir hacia una inadecuada gestión de control, un mal aprovechamiento de los recursos humanos y materiales, y por ende, a un pobre nivel de rendimiento y logro departamental.

4.2.1.3 Departamento de Denuncias

La información que utiliza este departamento normalmente se obtiene de los sistemas transaccionales que operan en cada aduana. Sus funciones básicamente consisten en desarrollar investigaciones y atender las denuncias que le presentan los usuarios o que se generan internamente en la DCF, concretamente sobre las operaciones que implican riesgos de defraudación fiscal o contrabando.

Los datos necesarios para su labor deben ser extraídos de los registros almacenados en el SIA. Dicha extracción puede ejecutarse mediante conexiones vía MODEM o por enlaces Frame Relay.

Las solicitudes de información generalmente se limitan a listados de declaraciones aduaneras tramitadas por algún importador o algún auxiliar de la función pública (agencias de aduana, transportistas, etc.) que posteriormente son pedidas en original para hacer sus comprobaciones e informes basados en ellos.

En ocasiones, los funcionarios del Departamento de Denuncias han requerido al DE, información sobre montos, cantidades e impuestos de mercancías importadas, esto con el fin de hacer una valoración de mercancías por categorías y según el valor de éstas, considerando el costo, el seguro y el flete. Estos tres últimos elementos constituyen el denominado valor CIF.

Al igual que con los departamentos interno y externo analizados anteriormente, la organización, clasificación y agrupamiento de los datos se hace utilizando hojas electrónicas. También consumen gran cantidad de tiempo y esfuerzo al tratar de acondicionar la información a la perspectiva deseada.

4.2.1.4 Departamento Legal

Este departamento prácticamente no hace uso de la información del DE. Los funcionarios de esta dependencia, abogados todos, se relacionan indirectamente con las operaciones de importación de algún sujeto de estudio o auxiliar de la función pública implicado en los casos que ellos atienden. También lo hacen por medio de los informes que los demás departamentos de la división o de la DGA les remiten.

4.2.1.5 Departamento Planificación

Los funcionarios de este departamento constituyen el grupo de usuarios que han requerido del DE la mayor cantidad de información, esto en cuanto al volumen de datos y el máximo detalle posible de información, proporcionado por las declaraciones aduaneras.

La información solicitada por este departamento, generalmente considera el total de registros de importación definitiva de todas las empresas tramitadas en las aduanas durante periodos extensos, como trimestres, semestres o años completos.

Tal magnitud de información transaccional, es requerida debido a que con ella se deben construir las tablas agregadas que se utilizarán para seleccionar los sujetos de estudio, según los criterios específicos definidos por la administración. También permitirán definir los parámetros de selectividad que se aplicarán a las mercancías que se importan.

La determinación de los parámetros de selectividad con los que se establece si las mercancías de importación van a ser objeto de revisión física, documental o no tendrán revisión, debe estar muy bien fundamentada y, por lo general, considera la información de las declaraciones aduaneras tramitadas durante los últimos doce meses.

Los datos en cuestión, deben ser procesados de forma tal que permita la generación de listas de importadores, según diversas formas de agrupación. Estas listas pueden considerar combinaciones múltiples o sencillas de distintas de variables. Por ejemplo, podría requerirse agrupamientos de los datos por aduana, por agencia, por depósito aduanero, por tipo de revisión de las mercancías o por importador. También por la combinación de aduana y agencia, aduana e importador o por los tres, aduana, agencia e importador, y así sucesivamente.

En la preparación de dicha información, se utiliza como herramienta de manipulación, filtrado y agrupación, el Visual Fox 5.0.

Con esta herramienta se busca obtener información relacionada con los importadores de interés, en forma ordenada y clasificada conforme a las necesidades inmediatas del departamento.

Este proceso de manipulación, el cual es manual, debe hacerse en períodos de tres meses. Es un proceso repetitivo no sistematizado que trae inherente riesgos, como la posibilidad de perder información, la duplicación de datos, fallas de integridad y dificultades operativas en la administración de múltiples tablas y archivos intermedios, que no están integrados a una base de datos formal.

4.2.1.6 Jefatura de la División

La información que requiere este nivel u otro superior, debe ser tal que permita tomar decisiones sobre los sujetos de estudio y definir parámetros de selectividad de forma acertada con la menor dispersión posible de recursos. Además, debe ser información que ayude a controlar y medir el grado de cumplimiento de las leyes

aduaneras y fiscales. Debe contribuir también, a la determinación de áreas de alto riesgo en las cuales enfocar la atención.

La información que necesita la jefatura de la División para tomar decisiones de este tipo, debe ser suministrada principalmente por el Departamento de Planificación.

4.2.2 Consulta de datos al Departamento de Estadística

La revisión de las consultas hechas por la DCF al Departamento de Estadística durante el año 2000, permitió establecer seis categorías de peticiones:

Categoría	Descripción ¹
A	Consultas sobre empresas específicas, en periodos distintos o iguales, con posibilidad de incluir muchas variables de detalle.
B	Consultas sobre importadores en las que se pueden incluir mercancías específicas para periodos o años determinados.
C	Consultas de mercancías específicas según distintos niveles de jerarquía y para distintos periodos.
D	Consultas sobre agencias de aduanas que pueden incluir partidas arancelarias específicas y periodos determinados.
E	Consultas sobre aduanas específicas para periodos determinados.
F	Agrupamiento (resumen) por distintas variables de medida en un periodo dado y para empresas específicas.

¹Todas las categorías, excepto la F, se tratan de consultas con niveles de detalle máximos. (mínima granularidad)

La distribución de las consultas durante el año 2000, según los departamentos solicitantes fue la siguiente:

Distribución por depto ²	Control Externo (Ce)	Planificación (Pl)	Control Interno (Ci)	Legal (Dl)
Total consultas	23	16	1	1
Porcentaje (%)	57	39	2	2

²⁾No incluye el Departamento de Denuncias debido a que en el año 2000 aún no se había creado dicho departamento.

Según las categorías definidas, las consultas se agruparon de la siguiente forma:

- Un 57% de las solicitudes corresponden a peticiones de tipo A, o sea, consultas con niveles de granularidad mínimos que involucraban generalmente empresas específicas en un periodo o para periodos distintos.
- Le sigue con un 17 %, las consultas tipo B que consideran peticiones detalladas, especificadas por importador y que pueden incluir mercancías específicas.
- Las consultas tipo D, con un 12 %, consideran las peticiones de información en las que se involucran las agencias de aduana y que pueden también incluir mercancías específicas.
- Con un 7 % están las consultas tipo C, que corresponden a peticiones relacionadas con mercancías específicas según sus jerarquías y para periodos distintos.
- Con un 5% de los casos, están las consultas tipo F, en las que se considera la información agrupada por alguna variable cualitativa (importador, agencia, aduana, etc.) y otra temporal (año, trimestre, mes), en la que se desea obtener clasificaciones de sujetos de estudio resumidos por algún tipo de variable de medida (valor cif, impuestos, etc.)
- Las consultas de tipo E, con un 3%, consideran las consultas detalladas clasificadas por aduanas específicas, en algún periodo de tiempo.

4.2.3 Análisis de resultados de las entrevistas y las consultas al Departamento de Estadística.

Conforme a los resultados, tanto de la aplicación de la entrevista a los usuarios de la DCF, como los obtenidos de las consultas al Departamento de Estadística, se concluye que solamente un 5% logra plantear consultas orientadas hacia la obtención de información resumida, clasificada o aplicada a variables cualitativas y cuantitativas específicas. La información obtenida proporciona una valiosa base para el estudio efectivo de casos, al estar estructurada de forma tal que permite el análisis e interpretación de resultados. Esto facilita una efectiva toma de decisiones basada en información precisa. La información conseguida en estos casos, constituye el tipo de resultados que típicamente son obtenidos de un Mercado de Datos.

El restante 95% de las consultas obtienen resultados muy detallados y voluminosos, que requieren de una manipulación exhaustiva para lograr obtener datos que puedan ser utilizados para la toma de decisiones.

En este alto porcentaje de casos, se solicita el detalle de las declaraciones aduaneras, que incluyendo el número que las identifica, los impuestos, los valores CIF y otras medidas relacionadas. Aunado a ésta situación de escasa especificación de resumen en las consultas de los usuarios, que produce tablas de resultados con gran cantidad de registros, los funcionarios deben usar hojas electrónicas como único medio para la manipulación de datos.

Las hojas electrónicas no tienen capacidad suficiente para manejar la gran cantidad de registros implicados en las consultas tipo A, B, C, D y E, lo que obliga a la separación de los datos que pertenecen a una misma consulta en hojas separadas, lo cual produce serias dificultades de manipulación, asociación y análisis de la información.

Debe señalarse también que la información que se solicita al DE, es utilizada como base para la obtención de muestras de verificación y validación de la documentación y datos que presentan las empresas relacionadas con los estudios que se realizan. De ahí la importancia de que la información proporcionada y analizada sea confiable y completa.

Aparte de las maniobras para manejar la información durante los procesos de validación y comparación, los funcionarios de los departamentos, según sea el estudio que tengan en proceso, deben transformar los datos o complementarlos para adecuarlos a sus necesidades de análisis. Tal es el caso de los campos de fecha que, por lo general, vienen en formato texto y deben ser convertidos al tipo fecha, para después hacer agrupaciones por meses o trimestres según convenga.

Igual sucede con los datos correspondientes a mercancías donde, para poder ver conjuntos específicos de mercancías, según los niveles de clasificación arancelario, es necesario que el código de 10 dígitos que lo compone, según la categoría que se desea consultar. Por ejemplo, los dos primeros para el capítulo, los cuatro primeros para la partida, o todos los diez caracteres para la apertura nacional.

La visualización de los datos de mercancías según el agrupamiento arancelario que se haga, requiere de procedimientos de separación que utilizan funciones para el manejo de cadenas de texto complejas de controlar, consumen mucho tiempo e implican riesgos de mantenimiento e integridad. Los casos y ejemplos de transformación y manipulación, se extienden a otros campos igualmente necesarios para la preparación de los datos según los propósitos de los analistas.

Como se muestra en el cuadro de distribución de las consultas por departamento, el 39% de ellas fueron hechas por el Departamento de Planificación. Sin embargo, debido

a una reorganización de funciones en este departamento, el volumen de datos que ha necesitado posterior al año 2000, ya no implica la consulta puntual sobre segmentos específicos de datos, sino que requiere información completa de periodos extensos, como años o meses, los cuales son necesarios para un agrupamiento y una selección representativa.

4.2.4 Definición de usuarios potenciales del Mercado de Datos

Con base en el análisis expuesto sobre las entrevistas y consultas al Departamento de Estadística, se concluye que solamente el Departamento Legal no requiere datos estadísticos para análisis. Por tal motivo, se establece que los usuarios potenciales del Mercado de Datos propuesto, son los pertenecientes a los departamentos Interno, Externo, Planificación, Denuncias y la jefatura de la división.

Con el modelo a proponer, por ejemplo, el Departamento Externo, que hace el 57% de las consultas, podría cumplir mejor con sus objetivos al tener la posibilidad de construir sus propias consultas, conforme al agrupamiento y selección que más le convenga y sin depender de ningún otro departamento. Esta posibilidad se extiende a los demás departamentos de la división, en virtud de que los datos que utilizan están considerados en el modelo.

Otros usuarios que verían apoyadas sus gestiones de control, el conocimiento general de sus operaciones y el de los auxiliares e importadores con los que se relacionan, serían los gerentes de las aduanas y sus mandos medios. Éstos, a través de una interfase de Internet, tendrían acceso a un conjunto de consultas específicas que les ayudarían a solventar muchas de sus necesidades de información.

Capítulo 5: Propuesta de diseño

5.1 Definición de dimensiones

Todas las consultas que requieren los departamentos de la División de Control y Fiscalización sobre el régimen de importación definitiva de mercancías, están determinadas por las opciones de acceso al SIA permitidas al Departamento de Estadística de la Dirección General de Aduanas.

La información de importación definitiva que es puesta a disposición de dicho departamento, corresponde a todos los registros transaccionales de los importadores que, en sociedad con los auxiliares de la función pública (agencias de aduana), diariamente se transmiten o presentan ante las aduanas del país.

Un subconjunto de las variables o campos que conforman parte de la información transaccional de los importadores en las aduanas, constituirá el grupo que conformará las dimensiones que se utilizarán en la construcción del prototipo de mercado de datos propuesto en este proyecto.

Las dimensiones y las unidades de medida a utilizar en el Mercado de Datos, se determinan a partir de los requerimientos de información definidos en el capítulo 4. Dichas dimensiones tienen correspondencia directa con los campos de la base de datos DADEC (véase Anexo C), desde donde se hace la extracción de los datos que actualmente utiliza el Departamento de Estadística.

Las dimensiones descriptivas que se proponen para el modelo de Mercado de Datos son las siguientes:

- o Aduana
- o Agencia
- o Importador

- o Mercancía
- o Lugar de descarga
- o País de origen
- o País de procedencia
- o Modalidad
- o Trámite
- o Semáforo
- o Tiempo

Las medidas asociadas a las dimensiones indicadas son:

- o Número de bultos
- o Valor Cif (costo, seguro y flete)
- o Derechos arancelarios a la importación.
- o Impuesto selectivo de consumo.
- o Impuesto de la ley 6946.
- o Impuesto de ventas.
- o Impuesto a las mercancías con destino al depósito libre Golfito.
- o Impuesto a las mercancías del tratado libre de comercio con México.
- o Total de impuestos pagados.
- o Total de impuestos exonerados.

Los detalles del nombre, tipo y tamaño de los campos, así como su descripción, se muestran en el diccionario de datos en el Apéndice A.

5.2 Metodología de desarrollo

[Poe98] establece que es importante tener claro el objetivo que se persigue al desarrollar un proyecto de Data Warehouse. Igualmente lo considera [Corey97] al señalar que se debe tener una idea precisa del conocimiento que se puede obtener de un sistema operacional. El prototipo de mercado de datos cuyo desarrollo se propone en este

documento, precisamente está orientado a apoyar las necesidades de información para la toma de decisiones a la jefatura y los departamentos de la División de Control y Fiscalización que lo requieran, así como a las de otros usuarios en las aduanas.

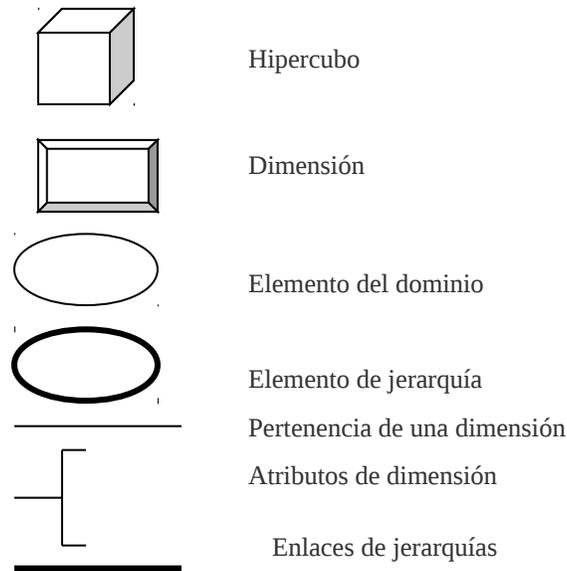
Por tanto, la primera tarea que se llevó a cabo fue determinar las necesidades de información de los usuarios, aplicando para ello entrevistas e investigando en los archivos históricos del Departamento de Estadística, todo lo relacionado a las solicitudes de información. Esto dio como resultado, el conjunto de dimensiones descriptivas y cuantitativas señaladas anteriormente.

Posteriormente, se hizo el análisis de los datos dentro de cada dimensión considerada, modelando con ellos un primera visión del prototipo. La meta era buscar el acomodo de los requerimientos de los usuarios al concepto que se iba desarrollando. En esta etapa se utilizaron datos de prueba reales con los cuales los usuarios exploraban si la interfase que se ofrecía satisfacía sus necesidades. Se continuó con ese tipo de evaluación hasta considerar que la información proporcionada por las dimensiones que formaban el modelo, así como las medidas que las delimitaban, cumplían aceptablemente sus pretensiones.

El concepto metodológico general que se ha empleado en esta propuesta, el cual concuerda con [Corey97] y [Wolf00], ha sido desarrollar una solución de facilitación de información para análisis y toma de decisiones, a partir de sectores específicos, como es el caso de la División de Control y Fiscalización. Posteriormente se pueden considerar nuevas etapas con la inclusión incremental de otras áreas de interés (nuevos mercados de datos), como por ejemplo el de valoración de mercancías, la verificación arancelaria, o los de control de procesos aduaneros.

5.3 Diseño lógico: modelo gráfico

Una vez definidas las dimensiones a modelar, así como el contenido de la tabla de hechos, se elabora el modelo gráfico, utilizando la metodología de [Cort99]. Dicha autora presenta los siguientes símbolos básicos:



5.3.1 Modelos gráficos

El siguiente cuadro muestra los elementos a modelar desde una perspectiva general:

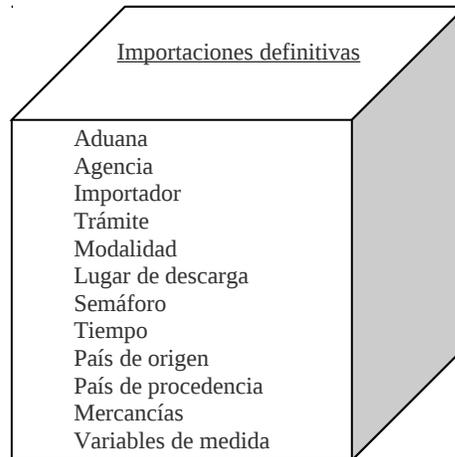
Importaciones	Mercancías	Capítulos	Partidas	Apertura Nacional
definitivas	Período	Años calendario	Trimestres	meses
		Años fiscales		

Tenemos entonces que se implementará un MD para las importaciones definitivas, tomando en cuenta las mercancías importadas, según los capítulos, partidas arancelarias, y el nivel de apertura nacional. Se considera además el período de tiempo en que se realizan éstas. Dicho período puede ser en años fiscales o en años calendario, estos

últimos incluyen trimestres y meses. La descripción anterior define las jerarquías de las dimensiones *mercancía* y *tiempo* del prototipo.

Utilizando la simbología descrita, tendríamos que el modelo general de hipercubos para la propuesta es como sigue:

FIGURA 15: Modelo general del hipercubo



Por su parte, el modelo dimensional para el sistema propuesto sería como se muestra en la figura 16.

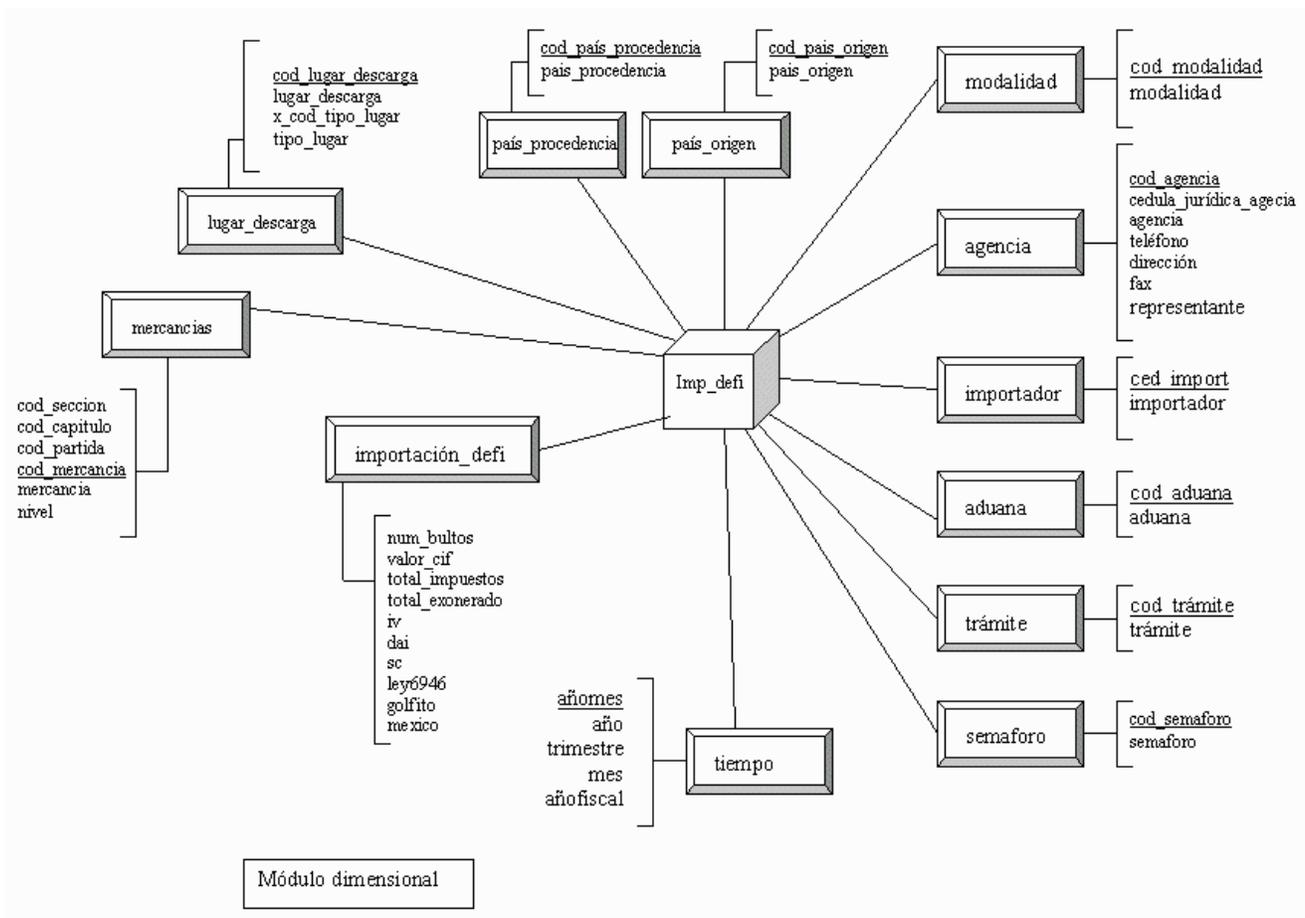


FIGURA 16: Esquema del Modelo Dimensional del prototipo

Aquí se presenta cada una de las dimensiones con sus respectivos atributos, todas ellas unidas al hiper-cubo Importación_defi. Por su parte, el modelo que representa las diferentes jerarquías se muestra en la siguiente figura:

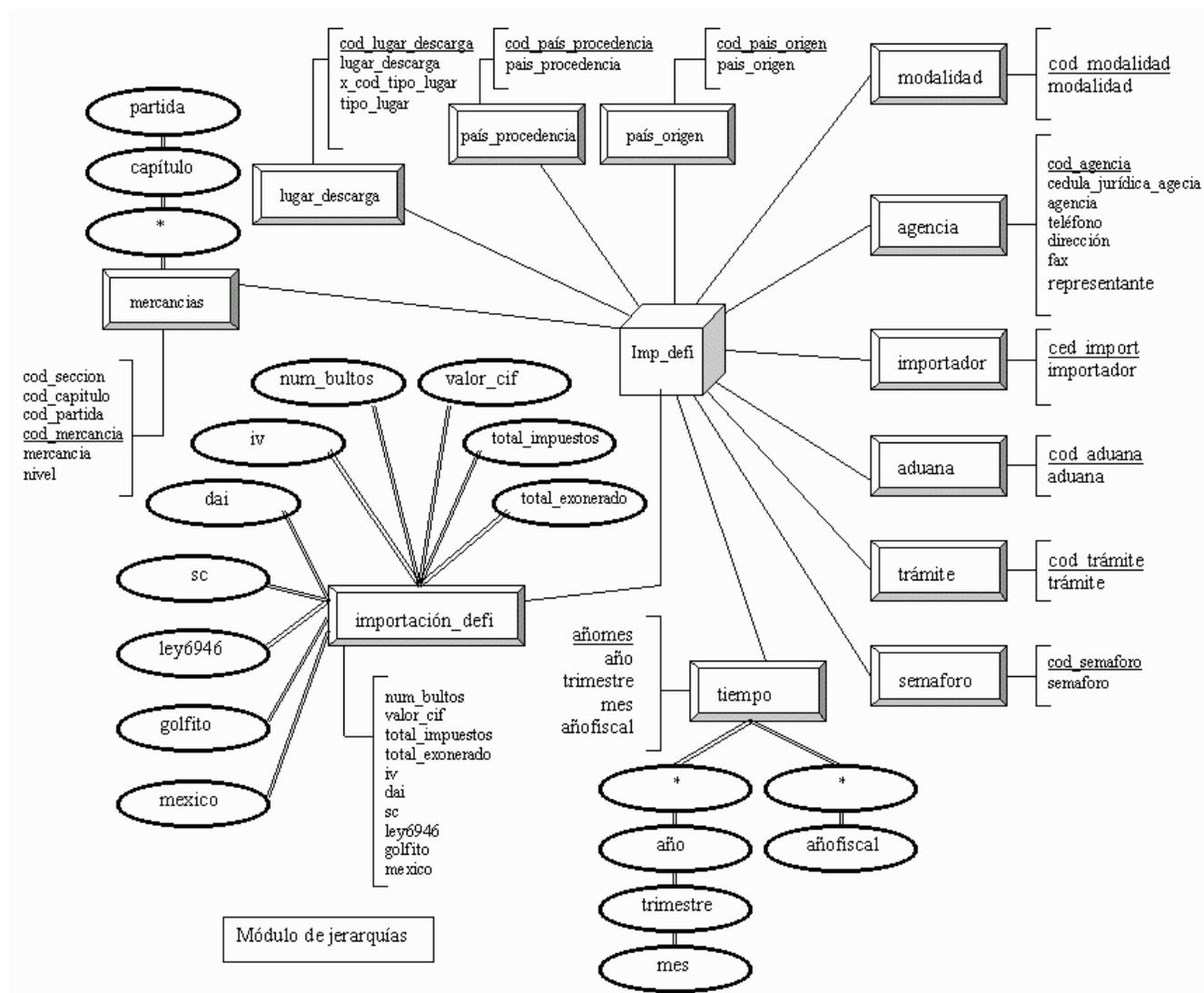


FIGURA 17: Esquema del Modelo de Jerarquías del prototipo

En este modelo se despliegan las jerarquías existentes en el hipercubo. Se puede apreciar el despliegue correspondiente de las dimensiones Tiempo y Mercancías, además de los atributos de clasificación de la tabla de hechos.

5.3.2 Descripción del modelo

La descripción del modelo se detalla a continuación:

Descripción del modelo

Hipercubo: *Importaciones definitivas*

Definición de dimensiones

Importación_defi: Variables de medida en la importación que consideran el valor de las mercancías, los impuestos pagados, las exoneraciones y la cantidad de bultos

Modalidad: Categoría sobre el tipo de mercancía y destinación

Agencia: Nombre y descripción de la agencia aduanal encargada de la importación

Importador: Nombre y descripción del importador de mercancías

Aduana: Nombre y descripción de la aduana por donde se realiza el trámite de desalmacenaje

Tramite: Tipos de tramite para el desalmacenaje

Lugar_descarga: Lugar físico donde se encuentra la mercancía a desalmacenar

Pais_origen: País de donde es originaria la mercancía importada

Pais_procedencia: País del cual procede la mercancía importada

Semáforo: Tipo de revisión al cual se somete la mercancía

Mercancías: Descripción de las mercancías importadas

Tiempo: Período de tiempo en el que se consideran las variables

Definición de Dominio

[Cod_modalidad]: número entero de dos dígitos.

[Modalidad]: nombre de todos los tipos de operación.

[Cod_agencia]: número entero de tres dígitos.

[Ced_jurídica_agencia]: número entero de 11 dígitos.

[Agencia]: nombre de todas las agencias del país.

[Ced_importador]: número entero de 12 dígitos.

[Importador]: nombre de todos los importadores del país.

[Cod_aduana]: número entero de dos dígitos.

[Aduana]: nombre de todas las aduanas del país.

[Cod_tramite]: código alfabético de un carácter.

[Tramite]: nombre de los tipos de trámite aplicados a la mercancía

[Cod_lugar_descarga]: código alfanumérico de tres caracteres.

[Lugar_descarga]: nombre de los tipos de lugar de descarga de mercancías.

[Cod_pais_origen]: número entero de 4 dígitos correspondiente al código ISO de países.

[Pais_origen]: nombre de todos los países de donde son originarias las mercancías.

[Cod_pais_procedencia]: número entero de 4 dígitos correspondiente al código ISO de países.

[País_procedencia]: nombre de todos los países de donde proceden las mercancías.

[Cod_semaforo]: número entero de un dígito, el cual puede ser 0, 1 o 2.
 [Semaforo]: nombre del tipo de revisión aplicado a la mercancía.
 [Cod_capitulo]: número entero correspondiente a los primeros dos dígitos del arancel de mercancías.
 [Cod_partida]: número entero correspondiente a los primeros cuatro dígitos del arancel de mercancías.
 [Cod_mercancia]: número entero correspondiente a los 10 dígitos del Sistema Arancelario Centroamericano.
 [Mercancia]: nombre de las mercancías.
 [Añomes]: número entero de 6 dígitos.
 [Cod_año]: número entero de dos dígitos.
 [Año]: número entero de cuatro dígitos.
 [Cod_trimestre]: nombre de número ordinal correspondiente a los trimestres del año.
 [Trimestre]: número y nombre del trimestre.
 [Cod_mes]: nombre de número ordinal correspondiente a los meses del año.
 [Mes]: número y nombre de los meses del año.
 [Cod_añofiscal]: número entero de dos dígitos.
 [Añofiscal]: número entero de cuatro dígitos.

Definición de jerarquías

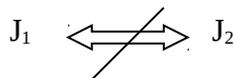
Dimensión Mercancías:

J ₁ :	P ₁ : * → capítulo	Plega todos los capítulos
	P ₂ : capítulo → partida	Plega todas las partidas por capítulo
	P ₃ : partida → mercancía	Plega todas las mercancía por partida

Dimensión Tiempo

J ₁ :	P ₁ : * → años	Plega todos los años calendario
	P ₂ : año → trimestre	Plega todos los trimestres por año
	P ₃ : trimestre → mes	Plega todos los meses por trimestre

J ₂	P ₁ : * → Años fiscal	Plega todos los años fiscales
----------------	----------------------------------	-------------------------------



Las jerarquías J₁ y J₂ son mutuamente excluyentes

Cada uno de los atributos requiere de un dominio, el cual se coloca junto a ellos encerrado entre corchetes.

Se han definido jerarquías en las dimensiones de Mercancías y Tiempo. En el caso de la dimensión Tiempo las jerarquías año calendario y año fiscal son mutuamente

excluyentes, por tanto se pueden hacer consultas sobre una jerarquía determinada, pero no sobre ambas al mismo tiempo.

Cada una de las dimensiones propuestas está estructurada por los atributos descriptivos asociados, uno de los cuales necesariamente corresponde al que identifica unívocamente a cada registro, sea su llave primaria.

Por su parte la tabla de hechos (*importacion_defi*) que contiene las variables de medida, se enlaza con las dimensiones a través de las llaves foráneas provenientes de cada una de ellas. La unión de las llaves foráneas declaradas en la tabla de hechos, conforma su llave primaria, completando así el esquema de estrella.

5.3.3 Representación de operaciones

Drill Down - Roll Up

Es posible, de acuerdo a la estructura de jerarquías diseñada en este modelo, ejecutar las operaciones de Drill Down y Roll Up sobre las dimensiones Tiempo y Mercancías conforme a la siguiente figura:

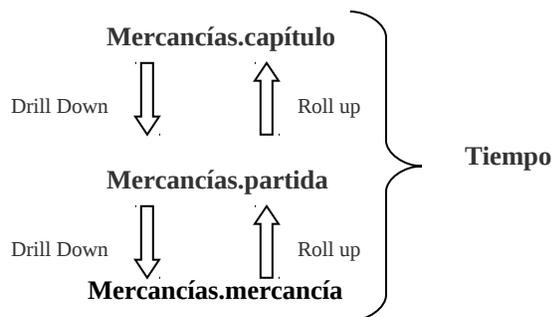


FIGURA 18: Drill Down y Roll Up en la dimensión Mercancía

Mientras se mantiene cualquier jerarquía de la dimensión Tiempo, se puede profundizar o plegar desde la jerarquía de Capítulo hasta la jerarquía de Mercancía y viceversa.

Así mismo, se puede realizar la misma operación para la dimensión Tiempo a través de la combinación de cualquiera de las demás dimensiones consideradas en el modelo:

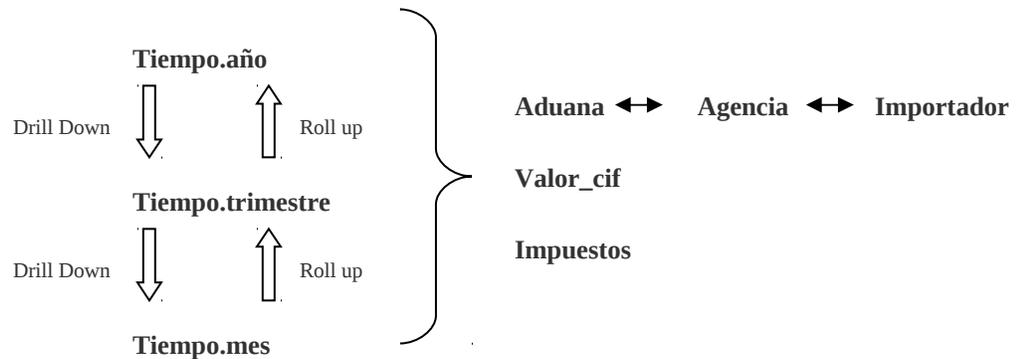


FIGURA 19: Drill Down y Roll Up en la dimensión Tiempo

En este caso se mantiene cualquier combinación de, por ejemplo, Aduana – Agencia – Importador, y los valores de medida Valor_cif e Impuestos, mientras se profundiza o pliega desde la jerarquía Año hasta la jerarquía Mes de la dimensión Tiempo, y viceversa.

En este caso concreto, el mínimo nivel de granularidad de la jerarquía es la de mes. Esta decisión se basa en tres razones: 1- por cuestiones académicas de demostración de la funcionalidad del prototipo, 2- por cuestiones prácticas de ahorro de espacio físico y velocidad de procesamiento, y 3- se determinó que este último nivel satisface las necesidades de información de los usuarios potenciales del prototipo. Sin embargo, la incorporación de el nivel de jerarquía día, o inclusive, hora, no representaría ningún problema técnico ni funcional, y podría ser implementado sin ninguna dificultad.

5.4 Diseño físico: descripción del prototipo

5.4.1 Base de datos

Con base en el modelo lógico representado gráficamente con los símbolos de [Cort99], se desarrolla el modelo físico.

Para la implementación de la base de datos se han creado tablas para cada una de las dimensiones descriptivas del modelo conceptual, las cuales contendrán los atributos de clasificación respectivos y la llave primaria. La figura 20 muestra el mapeo del modelo utilizando el esquema de implementación tipo estrella, donde todas las dimensiones se relacionan con la tabla de hechos con una cardinalidad de uno a muchos.

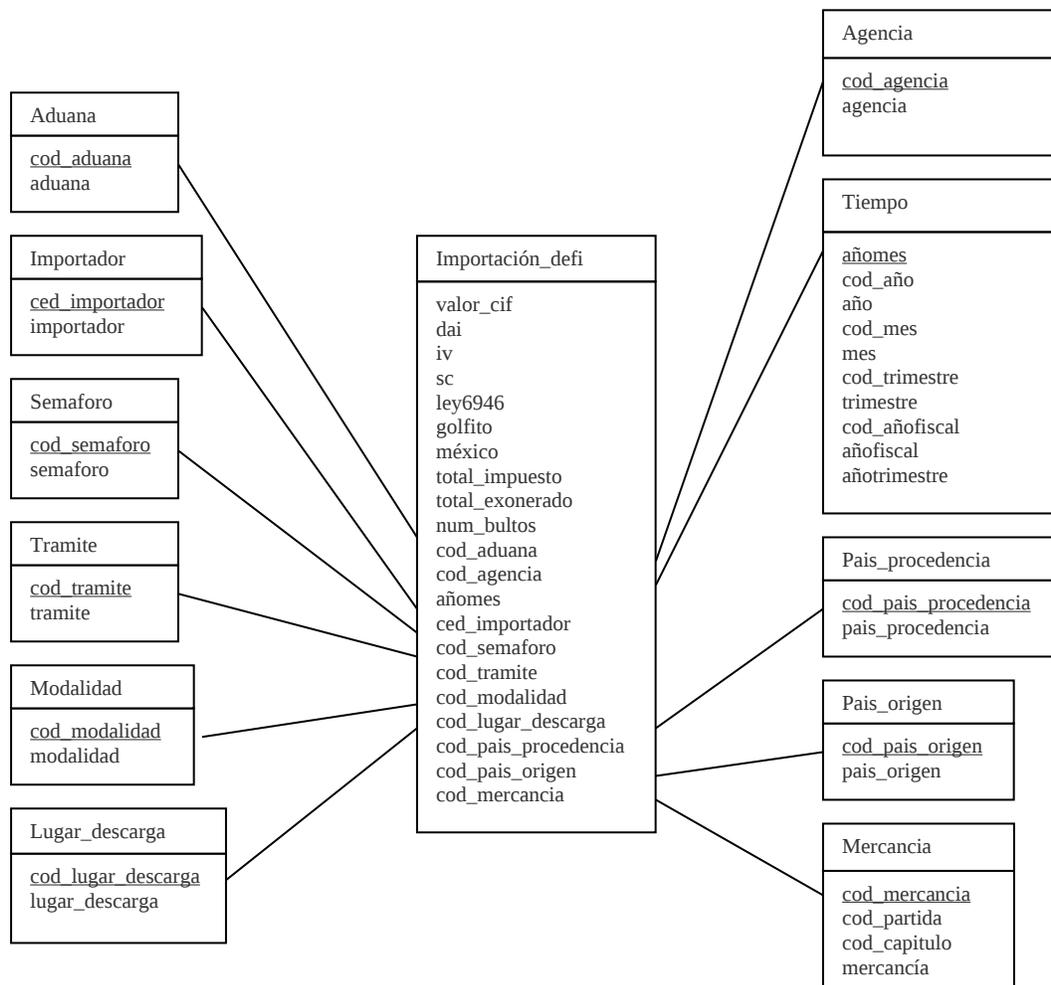


FIGURA 20: Modelo físico de la base de datos

Para este prototipo no se utilizan tablas resumen.

La tabla de hechos contiene las variables de medida. La llave primaria de la tabla de hechos está compuesta por el conjunto total de llaves foráneas, las cuales, a su vez, son llaves primarias de sus respectivas dimensiones $PK=(FK_1+ FK_2+...+FK_n)$.

Tomando en cuenta los elementos anteriores, tendríamos entonces la siguiente figura que representa la estructura condensada del cubo, obtenida mediante Visual Fox:

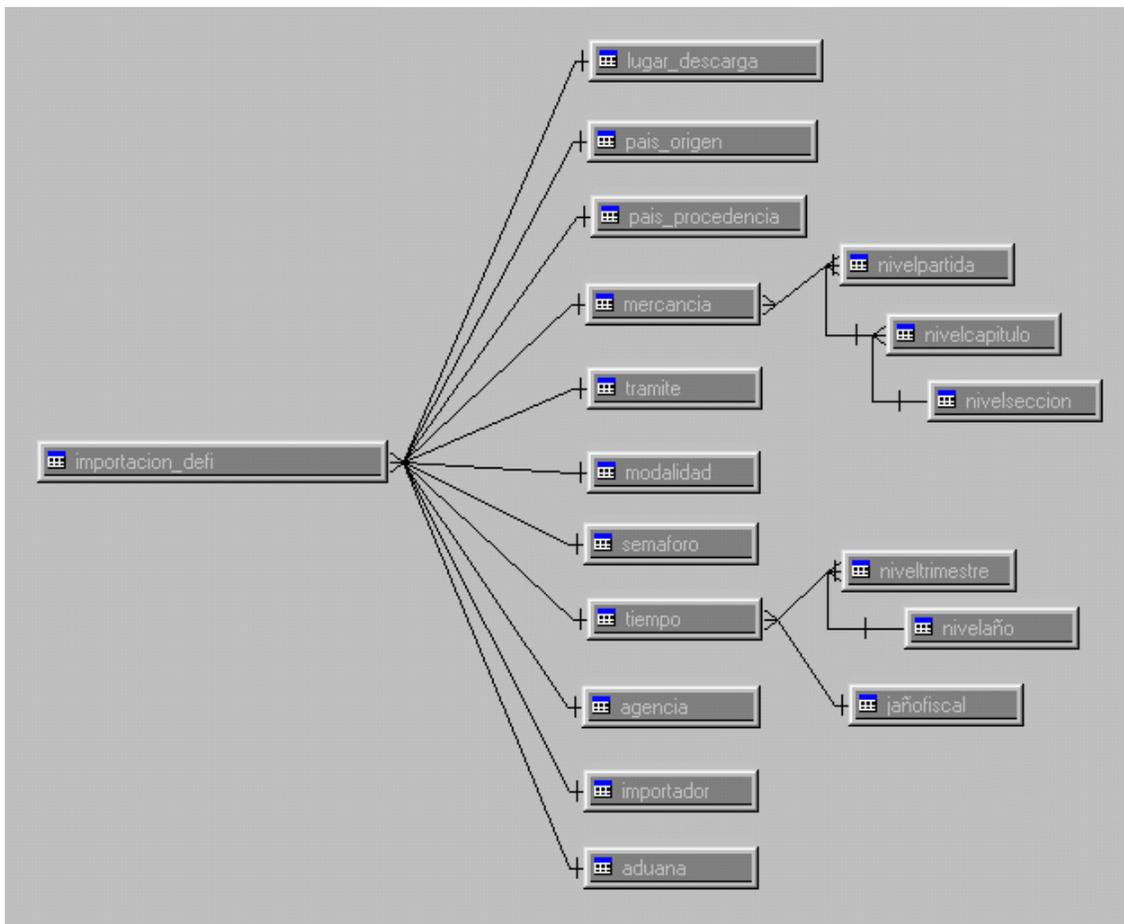


FIGURA 21: Esquema de la estructura de la base de datos

Las tablas Tiempo y Mercancías tienen otras tablas de nivel de jerarquía para la generación de agregaciones y simplificar la selección de datos. Con ello se pretende, a

futuro, agilizar la construcción de las consultas dentro de niveles específicos de jerarquía y obtención de datos concretos.

El mapeo de la dimensión Mercancía se esquematiza en la figura 22, donde se observan los enlaces en cascada con una cardinalidad de uno a muchos de las tablas nivelpartida, nivelcapítulo y nivelseccion.

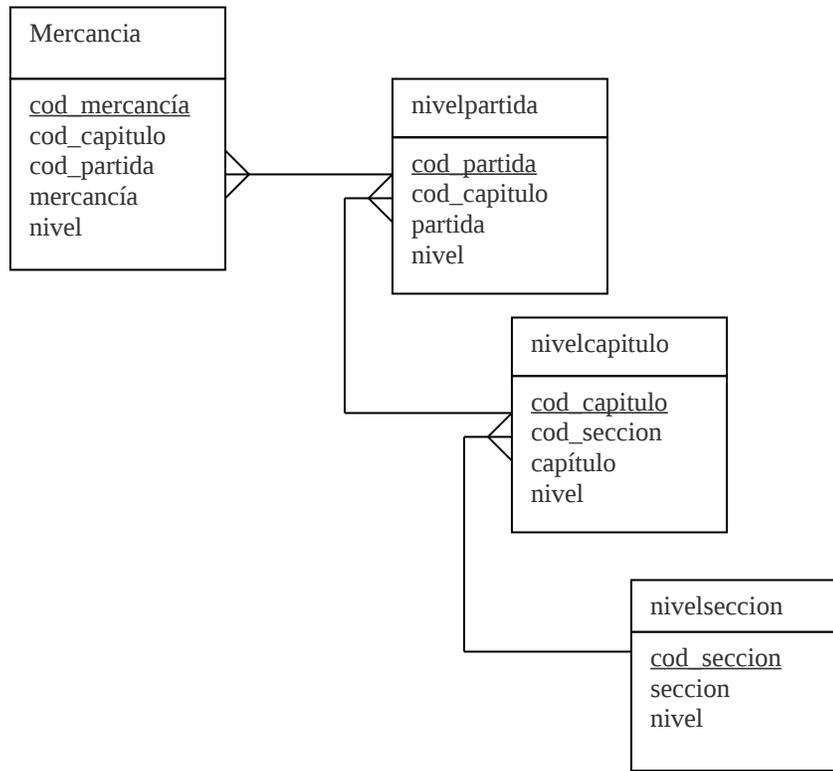


FIGURA 22: Ilustración de la jerarquización de la dimensión Mercancía

Por su parte, la dimensión tiempo, esquematizada en figura 23, presenta las tablas de jerarquización de años calendario y años fiscales y se puede apreciar que son mutuamente excluyentes.

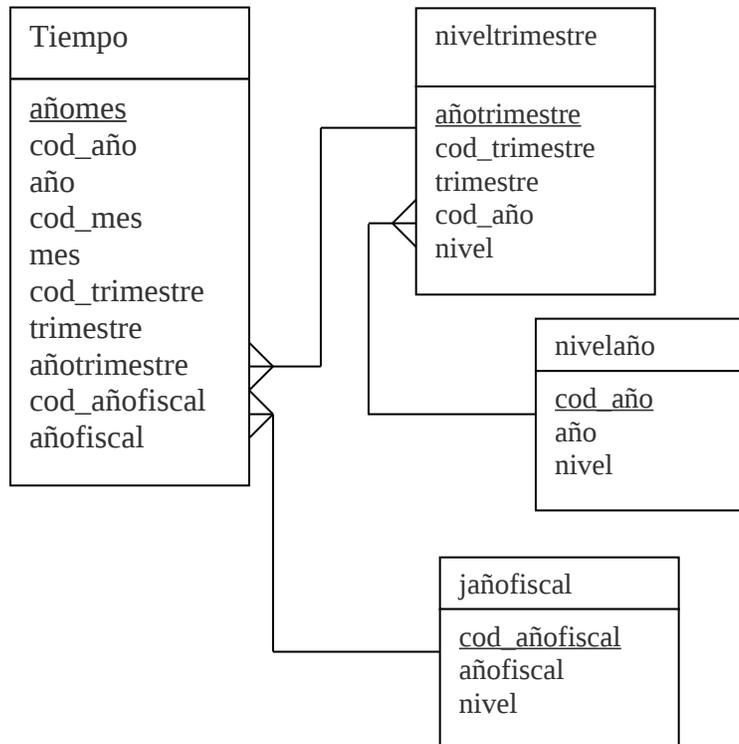


FIGURA 23: Ilustración de la jerarquización de la dimensión Tiempo

Cos respecto a la integridad referencial, se debe recordar que se trabaja con una base de datos histórica, por lo que las dimensiones son raramente modificadas o actualizadas, ejerciendo un estricto control sobre los datos desde los procesos de carga. La integridad referencial es administrada por los dominios de los atributos previamente definidos. De esta forma, no es posible insertar o actualizar un dato si el valor propuesto para un atributo no satisface las reglas de su dominio.

Para todas las dimensiones, la tabla secundaria común es Importación_defi. En la regla de actualización se posibilita actualizar solamente la tabla secundaria, ya que los registros de las tablas primarias no son modificados. En la regla de inserción, se permite insertar registros a la tabla secundaria aunque no existan registros coincidentes en las tablas primarias, esto debido a que los datos son previamente filtrados antes de su carga y

se eliminan los registros sin coincidencias. En la regla de eliminación, es posible quitar registros de las tablas primarias aunque estén vinculados a la secundaria, esto porque, al ser una base de datos histórica, dichos registros nunca serán eliminados.

Por último, los datos a cargar han sido extraídos de un sistema transaccional que aplica su propio esquema de control de integridad referencial en su operación normal. Esto debería garantizar que los datos que ingresan al sistema de carga del prototipo ya han sido depurados al nivel de integridad del sistema transaccional.

5.4.2 Extracción, transformación y carga de datos

La construcción de un sistema de extracción y carga de datos es un paso de gran importancia en la implementación de un MD. [Cata97] indica que uno de los retos más importantes al desarrollar un MD es el diseñar los sistemas de transformación y alimentación que obtienen y convierten los datos en un formato apropiado para el usuario final.

El sistema del modelo propuesto consta de tres subsistemas: el de transformación, el de agregación y el de carga. La figura 24 muestra el proceso con sus respectivos subsistemas. El proceso de extracción, el cual recolecta los datos requeridos por el sistema alimentador, es realizado en el Departamento de Estadística, tal y como se describe en el apartado 3.1 del capítulo 3.

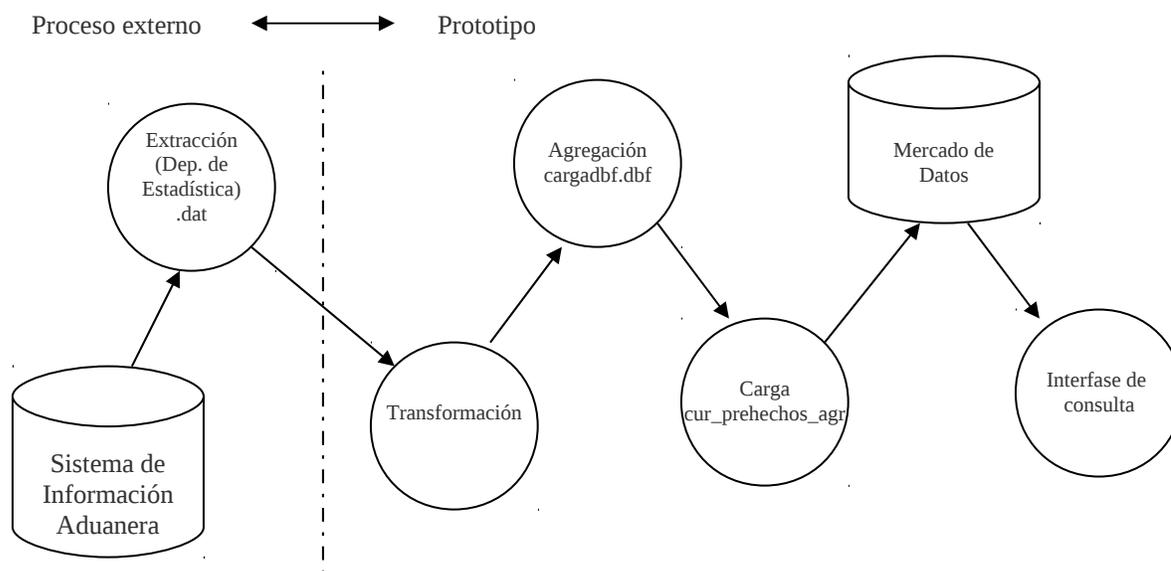


FIGURA 24: Proceso de transformación y carga

El subsistema de transformación convierte los datos en el formato de destino y elimina los campos que no son requeridos. El subsistema de agregación totaliza los datos hasta la jerarquía de Mes de la dimensión Tiempo. El subsistema de carga inserta los datos dentro del MD.

Las metas finales de este sistema en particular, las expone perfectamente [Cata97] al decir que *“son minimizar la complejidad del sistema, proveer de una valiosa capacidad de análisis a los usuarios y crear un sistema estable y de fácil mantenimiento”*.

Un aspecto importante de anotar es el hecho de que, al provenir los datos de una sola fuente (SIA), no es necesario implementar opciones para solucionar los problemas de diferencias lógicas, de tipo o de tamaño en los formatos de los archivos. Por tal motivo, los datos se cargan sin conciliación.

5.4.2.1 Proceso de transformación

El punto de inicio del proceso de transformación es la ubicación de los archivos con extensión dat (.dat) provenientes del subsistema de extracción. Estos archivos, que son de tipo texto, son revisados antes de la carga contra los datos ya incorporados a la tabla de hechos a fin de evitar la duplicidad de datos en ésta. Una vez localizados y revisados, los datos se transforman al formato dbf y se construye con ellos la tabla denominada cargadb, sobre la cual se realizan los siguiente procesos:

- Eliminación de las dos últimas líneas de la tabla que contienen datos relacionados con el proceso extracción, que no son importantes para el proceso de carga de la tabla de hechos.
- Verificación del código del tipo de régimen de importaciones. Se eliminan los registros cuyo código de régimen sea distinto de 01 (importaciones definitivas).

5.4.2.2 Proceso de agregación

En este proceso se totalizan las variables de medida extraídas de las tablas planas con base en la jerarquía *Mes* de la dimensión *Tiempo*, o sea, se totalizan por mes como nivel de mínima granularidad. Posteriormente se agrupan por cada una de las llaves primarias de las dimensiones consideradas. El resultado final es almacenado en un cursor (current set of record), llamado cur_prehechos_agr, el cual constituye el archivo temporal de SQL generado por la instrucción ejecutada.

5.4.2.3 Proceso de carga

Con base en los datos transformados y agregados, se procede a actualizar la dimensión *Tiempo*, agregando los años, trimestres y meses nuevos contenidos en los datos que se van cargar y que, por tal razón, no han sido contemplados en dicha dimensión. Por último, se procede a cargar los nuevos datos en la tabla de hechos.

Para mayor detalle sobre el código fuente del programa de carga, véase el Apéndice B.

5.4.3 La consulta de informacion en el prototipo

El centro neurálgico de la aplicación prototipo lo constituye la herramienta que permite extraer los datos agregados para un posterior análisis. Para el diseño de una interfase de consulta capaz de realizar esta función, fue necesario definir la secuencia lógica de las consultas, a fin de facilitar su operación.

Se decidió establecer, como punto de partida, la definición de el periodo en que se desearía efectuar el análisis. Posteriormente, la selección de las dimensiones y sus instancias, así como el orden de aparición. Por ultimo, las unidades de medida y su ordenamiento.

Para efectuar estas operaciones, existe un procedimiento central, el cual reúne todos las distintas variables que son seleccionadas a lo largo de la construcción de la consulta. Este procedimiento, llamado *bldSQL*, contiene el algoritmo principal, encargado de evaluar y validar el contenido de las variables provenientes de los distintos objetos que componen la interfase.

Todas las variables que son seleccionadas o modificadas en la interfase, son manejadas por el código de un método de la aplicación que permite el cambio interactivo,

reflejado directamente en el resultado de la consulta. Cualquier modificación en la interfase, envía las nuevas variables o valores al procedimiento `bldSQL`, para la construcción interactiva y automática de la sentencia SQL que determina los resultados de la consulta final.

La instrucción SQL final que genera el procedimiento `bldSQL` es la siguiente:

```
lcSelect = "SELECT "+ lcCamposGrl + lcDimMedida+;  
          " FROM "+ lcFromGrl+;  
          lcWhereGrl+;  
          lcGroupBy+;  
          lcOrderBy
```

Cada uno de los elementos, con el prefijo *lc*, contiene cadenas de texto que la configuran, según sean las selecciones, filtros y restricciones definidas interactivamente por el usuario.

5.4.4 Descripción general de la aplicación prototipo

El prototipo de aplicación planteado en este proyecto pretende, principalmente, procesar los datos de importaciones definitiva de mercancías del SIA, los cuales periódicamente son almacenados y administrados por el Departamento de Estadística de la DGA. La finalidad es ponerlos a disposición de los usuarios analistas a través de una aplicación informática que estructure la información desde una perspectiva multidimensional y que, además, posea una metodología de construcción de consultas dinámica, versátil y eficaz.

Para ello, básicamente se provee al usuario una interfaz sencilla, pero que le permita formular consultas multidimensionales y cuyo resultado final pueda exportar y manipular, por medio de otras herramientas de análisis y graficación especializadas.

También, le proporciona métodos para la carga de información, así como, otras herramientas de consulta, gestión administrativa y de mantenimiento necesarias para su funcionamiento.

5.4.4.1 Descripción de las opciones de menú de la aplicación

El prototipo de la aplicación desarrollada, contiene algunas opciones básicas de menú, las cuales son:

- o Archivo
- o Consultas
- o Carga de datos
- o Administración

Archivo

La opción *archivo* del menú, contiene elementos para la impresión de los reportes diseñados automáticamente por la herramienta de desarrollo, según la configuración y especificaciones proporcionadas; también ésta opción permite la salida de la aplicación.

Los reportes mencionados, no pretenden ser más que posibilidades del prototipo para imprimir el contenido de las tablas de dimensión y otras salidas impresas que eventualmente el usuario quisiera ejecutar, o para mostrar la posibilidad de su desarrollo posterior.

Carga de datos

Esta opción contiene las herramientas necesarias para que el administrador del sistema pueda introducir periódicamente los datos que actualizan el MD. También busca facilitar el control de la información que se va introduciendo, ofrece la opción de consulta

de archivos cargados (véase figura 25). El detalle de las operaciones que se realizan a través del formulario de carga de datos, se explicará en la siguiente sección.

ARCHIVOS DE DATOS

Archivos cargados

Código de la aduana

Archivos de datos cargados al sistema

Archivodat	Aduana	año	mes
ei070521	7	21	05
ei071020	7	20	10

Navigation buttons: << < > >>

Cancelar

FIGURA 25: Formulario de archivos cargados

Consultas

Contiene las opciones que dan acceso a la interfase de consulta y a los formularios de cada una de las dimensiones que se consideraron en la aplicación. Éstos posibilitan al usuario la navegación sobre las tablas, la búsqueda de datos, el ordenamiento de campos y, en algunos casos, la impresión de listados.

Los formularios, dependiendo de la cantidad de elementos de clasificación asociados a las dimensiones, posibilitan la navegación dentro de las tablas o muestran su contenido completo. La siguiente figura muestra el formulario de consulta de la dimensión Mercancía.

Sección	Capítulo	Partida	Apertura Nacional
01	01	0103	0103910000
01	01	0103	0103920000
01	01	0104	0104
01	01	0104	010410
01	01	0104	0104101000
01	01	0104	0104109000
01	01	0104	010420
01	01	0104	0104201000
01	01	0104	0104209000
01	01	0105	0105
01	01	0105	01051
01	01	0105	0105110000

GALLOS, GALLINAS, PATOS, GANSOS, PAVOS (GALLIPAVOS) Y PINTADAS, DE LAS ESPECIES DOMESTICAS, VIVOS.

FIGURA 26: Formulario de consulta de Mercancías

La interfase de consulta es el componente principal del prototipo. A través de ella los usuarios pueden, interactivamente, definir las consultas multidimensionales que necesiten, según las perspectivas y variables que requieran en sus estudios o investigaciones o, para el apoyo en la toma de decisiones.

La utilización y aprovechamiento de este módulo se explicará más detalladamente en la sección 5.4.4.3 de este capítulo.

Administración

Esta opción de menú, permite al administrador del sistema hacer correcciones, inclusiones o borrado de datos, según las necesidades de mantenimiento del MD.

Con cierta regularidad los datos descriptivos de las dimensiones contienen errores u omisiones que no son filtrados en la carga de datos, razón por la cual se hace necesaria la posibilidad de hacerles ajustes y reparaciones.

5.4.4.2 Carga de datos

La figura 27 corresponde al formulario mediante el cual se ejecuta la actualización del MD. Está compuesto por tres cuadros de lista, un conjunto de botones de movimiento y un botón de comando que da inicio al proceso de carga, una vez seleccionados los archivos que se desean.

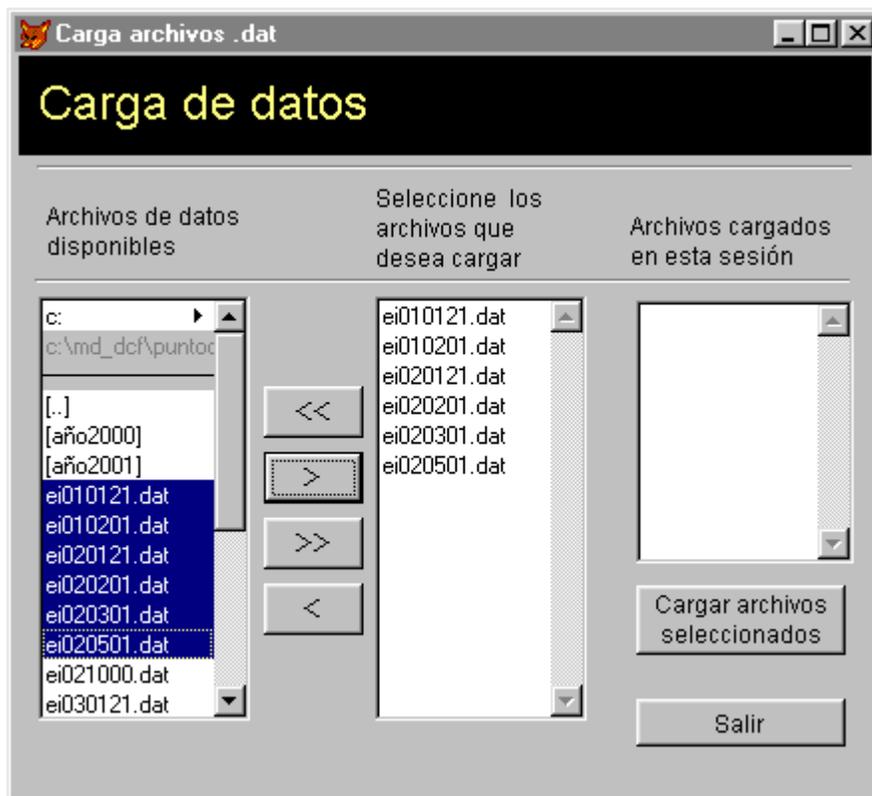


FIGURA 27: Formulario de carga de datos al MD

Archivos disponibles

El cuadro de lista bajo la etiqueta “Archivos de datos disponibles”, tiene la función de permitir al administrador del sistema acceder a los directorios donde se encuentran los archivos de datos necesarios para la actualización del DM. Estos archivos, de tipo texto, se identifican con la extensión dat. Cada archivo esta nombrado siguiendo la siguiente convención:

- Las dos primeras letras del archivo indican que los datos van dirigidos al Departamento de Estadística y que corresponden a la información operativa de importación.
- Los siguientes tres grupos de dos dígitos, señalan la aduana que origina los datos, el mes de que se trata y el año al que se asocian, en ese orden.

Los directorios donde se almacenan los archivos de datos, están estructurados conforme se muestra en la figura 28. Los archivo están agrupados por años y, dentro de éstos, por cada aduana que los origina.

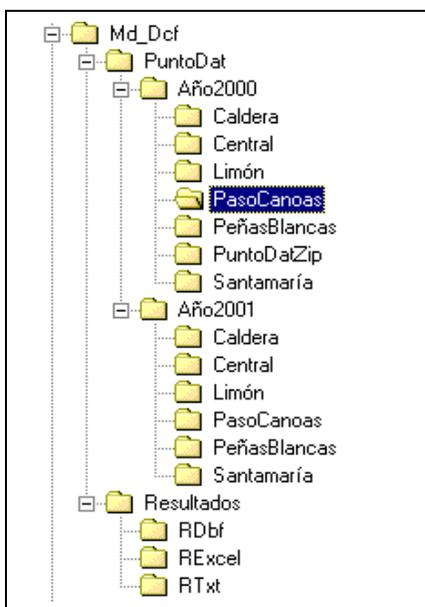


FIGURA 28: Distribución de directorios de programas, datos y resultados de la aplicación

Selección de archivos a cargar

Los elementos que se seleccionen en el cuadro “archivos disponibles”, pueden ser trasladados al cuadro de lista dispuesto a la derecha de éste, mediante los botones de movimiento. El propósito de dicho cuadro es facilitar al administrador del MD, la selección o desección de los archivos que se van de cargar.

A la derecha del cuadro de lista de archivos seleccionados para la carga, se ubica el cuadro de lista de archivos cargados. Éste le indica al operador cuáles archivos han completado exitosamente el proceso de carga en la tabla de hechos.

En esta etapa de la operación de carga, un algoritmo de verificación evalúa si los registros de los archivos seleccionados para carga, están ya incorporados al sistema o no. De estar incluidos, rechaza la operación e indica al operador tal condición.

5.4.4.3 Interfase de consulta

La figura 29 muestra la ventana principal de consulta del prototipo. A través de ella los usuarios pueden construir sus consultas de manera interactiva y sencilla, seleccionando múltiples dimensiones y variables de medida según sus necesidades de investigación y exploración. Es el componente más importante de la aplicación.

Este constructor de consultas, compuesto por un conjunto de cuadros de texto, casillas de verificación, botones de control y ventanas de visualización, actúa coordinadamente conforme el usuario interactúa con ellos al diseñar consultas.

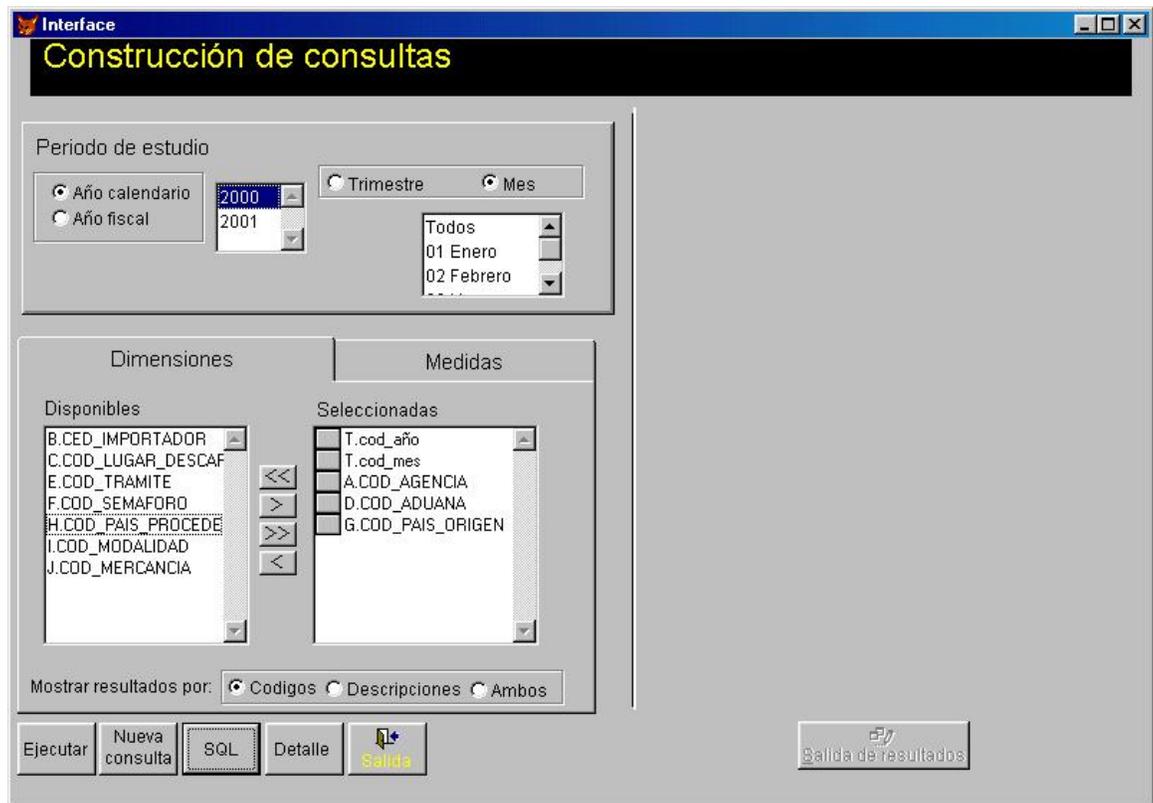


FIGURA 29: Interfase de consulta multidimensional del prototipo.

Contiene cuatro componentes que se pueden describir según su función como: a) El seleccionador de periodos de análisis, b) El marco de páginas para la selección de dimensiones y medidas, c) El área de despliegue de resultados, y d) Los botones de control para la ejecución y visualización de opciones.

El primer componente que se muestra en figura 30, está ubicado en la parte superior de la ventana, y es el que permite al usuario definir el período de análisis de la consulta.



FIGURA 30: Componente para la definición del periodo de análisis.

El usuario debe seleccionar, marcando de las dos opciones disponibles, el tipo de año que desea utilizar para su consulta. Sólo se le permite seleccionar uno de los dos dado que, por el diseño de la base de datos, éstos corresponden a jerarquías mutuamente excluyentes de la dimensión *tiempo*, es decir, no se pueden desplegar ambos a la vez.

En caso de que se seleccione año calendario, se desplegará además de la lista de años naturales disponibles, otro grupo de opciones que permite bajar la consulta al nivel de trimestres o al de meses.

La selección del período de análisis es obligatoria para la ejecución de cualquier consulta. Esto quiere decir que al menos se debe seleccionar uno de los dos tipos de año, de lo contrario, no se podrá continuar con la construcción de la consulta.

Los siguientes son los elementos que componen este bloque:

- **Marcadores de verificación para el tipo de año:** Este componente, ubicado dentro del bloque de controles definidos para establecer el periodo de análisis, opera sobre los atributos de las jerarquías año calendario y año fiscal y permite definir el alcance del periodo en el cual se desea hacer la consulta. Tiene dos opciones de selección: a- consultas por año calendario y b- consultas por año fiscal. La lista que contiene el año calendario, agrupa los meses naturales, en el orden usual, de enero a diciembre, mientras que para el año fiscal, que comprende los meses de setiembre a octubre, no se despliega ninguna lista de meses.
- **Cuadro de lista años:** A la derecha del marcador de verificación se despliega la lista de años disponibles en la base de datos. Este cuadro sólo muestra los años que se han ingresado al MD, esto a fin facilitar al usuario la selección del rango de

años disponibles. También sirve de ayuda visual para tener presente cuáles años de los seleccionados aparecen en el resultado de la consulta.

- **Marcadores de verificación para el agrupamiento de meses o trimestres:** Esta opción se muestra a la derecha del cuadro de lista de años y solo se dispone cuando el tipo seleccionado ha sido “año calendario”. Al activarse despliega la lista de meses o trimestres para selección. Esta opción de agrupar por trimestres o meses no esta disponible para años fiscales.
- **Cuadro de lista de trimestres o meses:** Se abre uno de los cuadros de lista meses o trimestres, posterior a la selección de una de las opciones del marcador de verificación descrito anteriormente. Éstas listas despliegan los meses o los trimestres y permiten una selección total, múltiple o sencilla de los elementos.

El siguiente componente de la interfaz de consulta, mostrado en la figura 31, corresponde al bloque que contiene los dos conceptos principales de una consulta multidimensional en un esquema tipo estrella, estos son, las dimensiones descriptivas y las dimensiones de medidas. La primera de ellas especifica las variables del negocio que determinarán la perspectiva o ángulo desde el cual se visualizarán los datos determinados por la segunda, las medidas.



FIGURA 31: Ficha de selección de dimensiones descriptivas.

Selección de dimensiones y medidas

- **Cuadro de lista de dimensiones disponibles:** Esta ubicado dentro del marco de páginas compuesto por dos fichas, una para las dimensiones descriptivas o simplemente dimensiones y otra para las medidas. En él se listan todas las dimensiones contenidas en el DM, excepto la dimensión *tiempo*, que ya ha sido considerada al definirse el periodo de análisis.

El cuadro de lista de dimensiones es fundamental para la construcción de la consulta. A través de éste, se seleccionan una o múltiples dimensiones, según sean las necesidades del usuario y determinan el resultado de la consulta.

La selección de dimensiones se hace simplemente marcando y trasladando elementos al cuadro de dimensiones seleccionadas ubicado a la derecha, mediante la pulsación de los botones de movimiento dispuestos entre ambos cuadros, o haciendo doble clic sobre la dimensión que se desea seleccionar. También, puede seleccionar dimensiones marcándolas y arrastrándolas con el puntero del ratón al cuadro de dimensiones seleccionadas. Igualmente, puede deseleccionar

dimensiones que ya no desea en el resultado de la consulta, regresándolas al cuadro de dimensiones disponibles, mediante las mismas operaciones, pero, invertidos los sentidos de las flechas si utiliza los botones de movimiento para hacerlo.

Opciones para la visualización de resultados

Debajo de los cuadros de lista de dimensiones disponibles o seleccionadas, se muestra el grupo de opciones que permiten al usuario elegir si desea ver en los resultados, los códigos de las dimensiones que participan, sus descripciones o ambas. El propósito de facilitar tales opciones, es contribuir a darle al analista en los resultados, mayor claridad, en el caso de elegir “Descripciones” o “Ambos”, o mayor discreción al omitir datos explícitos en el caso de elegir sólo “Códigos”.

Otra posibilidad muy útil que tiene el usuario para configurar dinámicamente los resultados de las consultas es, que el orden de los agrupamientos de las dimensiones seleccionadas, lo puede establecer justamente después de elegir las. Para ello, tiene a su disposición, los botones desplazables colocados junto a cada dimensión seleccionada, en el cuadro de lista ubicado a la derecha de la ficha Dimensiones. Esto es, conforme los mueva, lo harán las dimensiones a las que se asocian, determinando así el orden de despliegue de los resultados de la consulta según la siguiente concordancia: de arriba hacia abajo en el cuadro de lista de dimensiones seleccionadas, equivale a acomodar de izquierda a derecha en la tabla de resultados, las dimensiones en cuestión.

Selección de instancias específicas de dimensión

Una característica importante de los cuadros de lista de la ficha Dimensiones, es que a partir de los atributos de clasificación que determinan las llaves primarias de cada dimensión, los cuales están contenidos en éstos cuadros, se pueden filtrar instancias específicas de las dimensiones que se seleccionen, de forma tal que, la consulta se puede restringir a ciertos valores de las dimensiones elegidas. El resultado de la selección de ocurrencias específicas, determina el número de filas de la salida y por lo tanto, la claridad y la facilidad de análisis que tendrá el usuario con ella. La figura 32 muestra como se despliegan y seleccionan valores específicos de la dimensión Agencia [de Aduanas] mediante el cuadro de dialogo que surge al hacer clic derecho sobre cualquier dimensión disponible o seleccionada.



FIGURA 32: Filtrado de consultas por atributos de clasificación

Este cuadro de diálogo está estructurado con dos partes principales. Cualquiera de las dos, permite al usuario seleccionar instancias específicas de dimensión, pero de forma distinta. En la parte superior de la ventana, a través de un cuadro de texto, el usuario puede digitar los códigos específicos de las instancias de dimensión que desea, las cuales se van disponiendo en la lista previa ubicada debajo de dicho cuadro de texto, hasta que se ordene, mediante la acción sobre el botón “Iniciar búsqueda y selección”, hacerlo en la lista que muestra todas las instancias. Si

los valores digitados se encuentran en la lista inferior, se resaltarán en azul y los que encuentre serán considerados y filtrados en el resultado final.

Por su parte, el usuario también puede seleccionar valores específicos directamente de la lista inferior, sin tener que digitarlos y sin perder la selección que mediante digitación hizo, siempre y cuando haga la selección directa antes de hacerlo mediante la digitación.

Mención aparte requiere la selección de instancias específicas para la dimensión Mercancías.

Dado que esta dimensión contiene las jerarquías capítulo, partida y mercancías, construidas según las disposiciones legales, nacionales e internacionales que definen con exactitud que grupos de mercancías las conforman, fue necesario construir la ventana de consulta que se muestra en la figura 33.

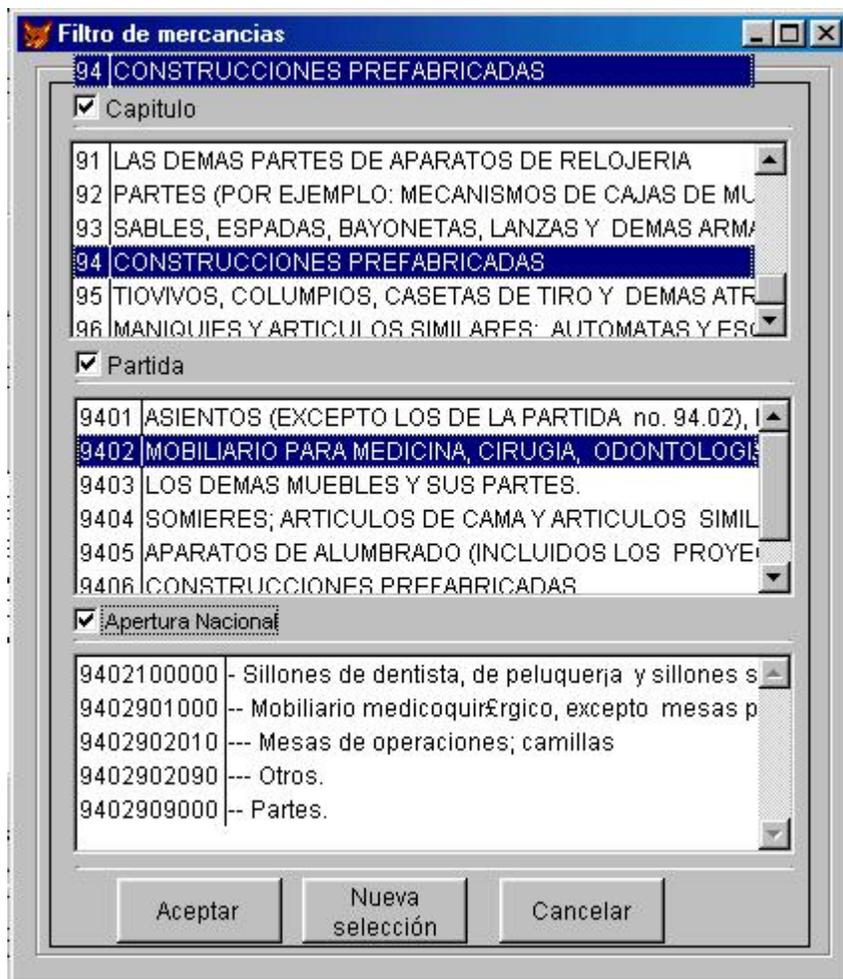


FIGURA 33: Filtrado de consultas por atributos de clasificación

Mediante esta ventana, que se activa haciendo clic derecho sobre la dimensión Mercancía, se podrán seleccionar los códigos arancelarios de cada jerarquía, al mismo tiempo que se observan las descripciones correctas y exactas de las mercancías a las que se asocian.

También, la forma en que están programados los objetos que la componen, permite seleccionar cualquiera de las jerarquías, independientemente de si un nivel superior en la jerarquía está o no seleccionada.

Salvo si hay instancias específicas seleccionadas a escala superior, el siguiente nivel mostrará, nada más, las instancias del o los niveles inferiores que estén contenidos en los valores de instancia de los niveles más altos que haya escogido.

Visualización de instancias específicas de dimensión

Para ayudar al usuario a recordar las instancias de dimensión que ha seleccionado se le facilita el botón de comando “Detalle”. Éste botón, conjuntamente con la selección en el cuadro de lista de dimensiones seleccionadas, de aquella que desee examinar, permite ver las instancias que ha seleccionado de una dimensión seleccionada. El resultado de dicha acción es un cuadro de consulta como el que se muestra en la figura 34.

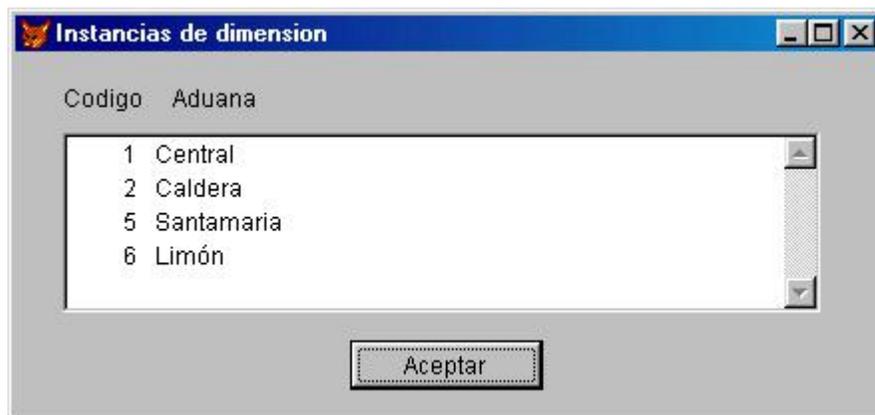


FIGURA 34: Visualización de instancias seleccionadas de la dimensión Aduana.

- **Cuadros de lista de medidas:**

Al igual que la ficha Dimensiones, la de Medidas esta compuesta por dos cuadros de lista, uno a la izquierda que contiene todas las medidas disponibles en el MD y otro a la derecha que muestra las medidas que se han seleccionando, ver figura 35. La selección de medidas se hace de la misma forma que se describió para las

dimensiones, o sea, marcándolas en la lista y usando los botones de movimiento, doble clic o arrastrándolas, se pasan al cuadro de medidas seleccionadas.



FIGURA 35: Ficha de selección de dimensiones descriptivas.

También el cuadro de lista de medidas seleccionadas permite el acomodo de las variables en el resultado final de la misma forma que se hace con las dimensiones.

Por medio de los botones desplazables el usuario puede acomodar dichas variables de arriba hacia abajo y en el resultado se ordenarán concordantemente, de izquierda a derecha.

Es importante señalar que las variables de medida, siempre se acomodarán a la derecha de la última dimensión seleccionada y en ese orden se desplegarán en los resultados. Esto significa que, la variable de medida que más arriba se encuentre en el cuadro de lista de medidas seleccionadas, es la que se colocará primero a la derecha de la última dimensión, la siguiente medida se ubicará a la derecha de esa medida y así sucesivamente hasta la última variable de medida seleccionada.

Debido a que los valores de las medidas en el DM son variables continuas, solo se permite escoger una de ellas para definir el orden de salida específico que tendrá el

resultado final, el cual puede ser ascendente o descendente. Si se selecciona alguna de esas opciones, todo el orden de la tabla resultante estará determinado por esa selección. Pero hay una salvedad, si se selecciona la opción “ninguno”, el orden de la salida estará determinado por la ubicación de las dimensiones en el cuadro de lista de dimensiones seleccionadas, y por omisión, cada una de ellas se ordenará ascendentemente y siguiendo un orden de izquierda a derecha.

Las dos carpetas Dimensiones o Medidas, incluyen botones de selección que permiten a los usuarios pasar de izquierda a derecha y viceversa, los atributos de dimensión disponibles en el DM, permitiendo con ello que las configuraciones de la consulta hecha, puedan ser modificadas rápidamente y sin complicaciones.

La versatilidad y capacidad interactiva de la interfase de consulta, son definitivamente características relevantes de la aplicación.

Ejecución de la consulta

Una vez establecido el periodo de análisis, las condiciones de ordenamiento y visualización de las dimensiones que se desean consultar, la distribución de éstos en la tabla de salida, los filtros para instancias de dimensión específicas (si se seleccionan), y las variables de medida buscadas, estaría la consulta lista para su ejecución.

El formulario que constituye la consulta principal, dispone del botón de comando “Nueva consulta” que le permitiría al usuario cancelar totalmente la consulta que esté en proceso de construcción o haya sido ya ejecutada y comenzar de nuevo otra.

Como se muestra en la figura 36, en la mitad derecha del formulario de consulta se puede desplegar mediante la acción del botón SQL, una ventana que muestra la instrucción de ese lenguaje, que el programa automáticamente construye a partir de las

selecciones que usuario a través de las herramientas del programa. El propósito de mostrar dicha instrucción es permitir el examen de su contenido para efectos únicamente académicos. El botón SQL apaga y enciende a gusto del usuarios la ventana que muestra dicha instrucción.

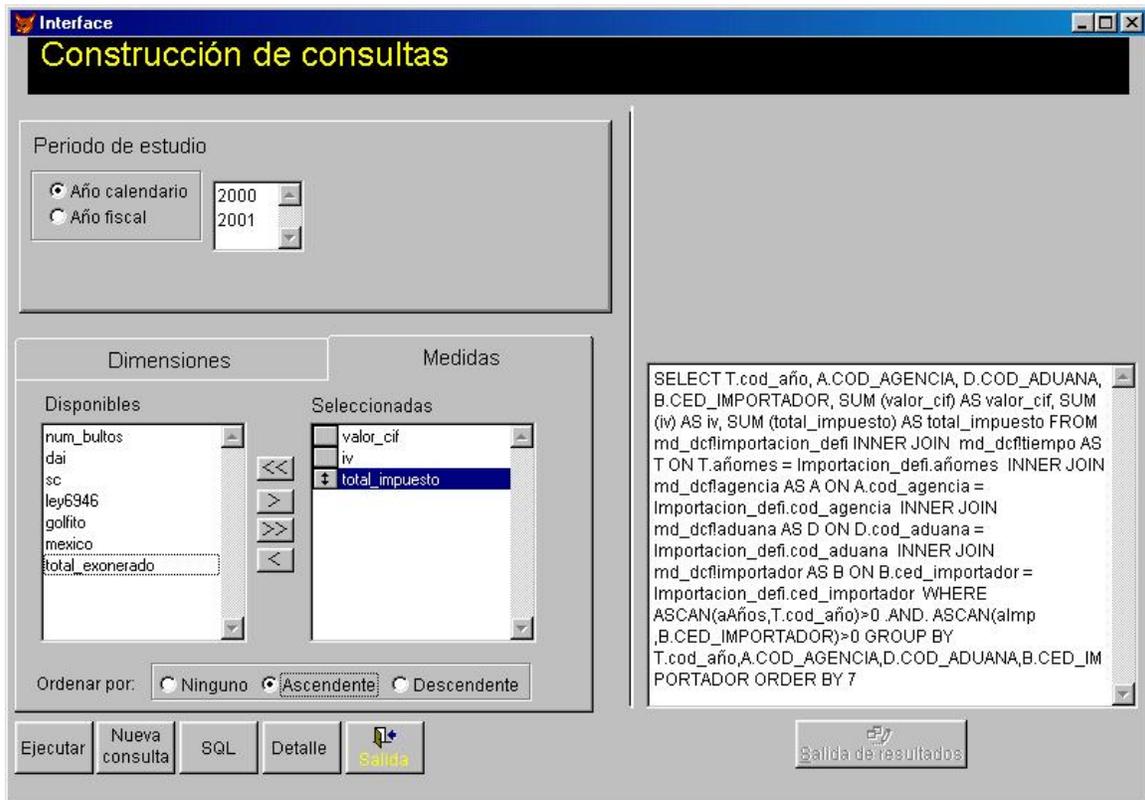


FIGURA 36: Visualización de los estatutos SQL de la consulta

La figura 37 muestra en la parte derecha de la interfaz una cuadrícula con el resultado de la consulta desarrollada. Se observa en el ejemplo, que la descripción de los títulos de las columnas corresponden a la selección de la opción “Descripción” del cuadro de dialogo Dimensiones seleccionadas. También se observa que el usuario ha decidido desplegar los resultados de la consulta en orden descendente según el valor cif de las importaciones hechas en cada aduana, durante los meses y años indicados.

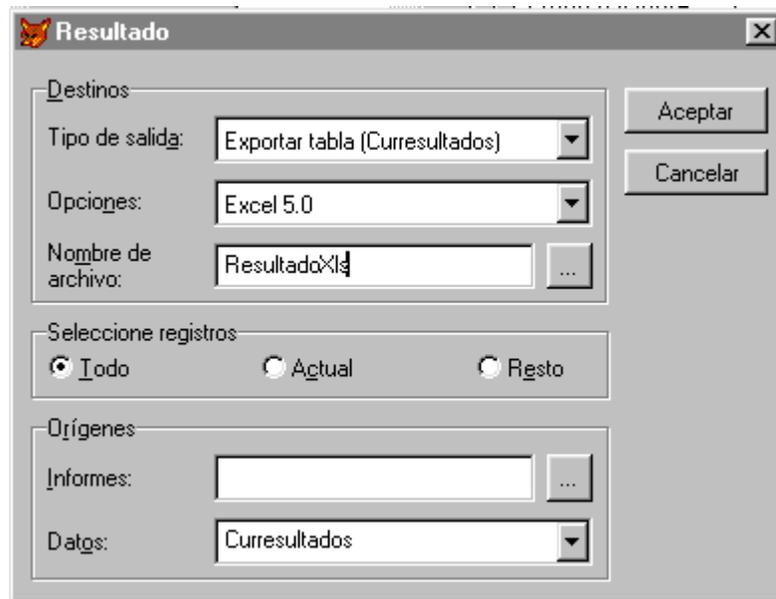


FIGURA 38: Salida de resultados a diferentes formatos

5.4.4.4 Pruebas preliminares de carga y ejecución

Con el fin de estudiar el comportamiento del prototipo, se realizaron pruebas preliminares de carga y ejecución de consultas. Para ello se utilizó un equipo de cómputo con las siguientes características:

- Procesador Pentium III de 700 Mhz.
- Disco duro de 20 Gb.
- Memoria RAM de 64 Mb.
- Caché de 256 Kb.

Pruebas de carga

Con el propósito de conocer el desempeño de la aplicación, se prepararon los 72 archivos planos que corresponden a las importaciones definitivas tramitadas durante el año 2000 en todas las aduanas del país.

El total de líneas de transacción con el máximo nivel de detalle que dichos archivos generaron fue de un millón noventa mil líneas.

Utilizando el módulo de carga de la aplicación desarrollada y conforme a la capacidad de éste para resumir los datos al nivel de la jerarquía Mes, de la dimensión Tiempo, dio como resultado un total de 976 458 líneas.

El tiempo transcurrido en el proceso de carga de los 72 archivos planos fue de una hora veintidós minutos, lo cual es considerado aceptable dadas las prestaciones del equipo utilizado y el volumen de datos de carga empleado.

Consultas de prueba

Basados en los datos cargados al MD correspondientes al año 2000, se hicieron las siguientes consultas:

CONSULTA 1: ¿Cuál es la aduana que tiene el mayor monto acumulado de valor CIF según las mercancías importadas durante el año 2000?.

o DURACIÓN DE LA CONSULTA: 2:57 minutos.

Dimensiones involucradas: aduana, tiempo filtrado por año. Variables de medida: valor_cif

CONSULTA 2: Listado de las mercancías importadas durante el año 2000, a través de la Aduana Santamaría y la Aduana Limón, al nivel de la jerarquía arancelaria de apertura nacional, ordenadas de mayor a menor según el monto de valor CIF

o DURACIÓN DE LA CONSULTA: 6:53 minutos.

Dimensiones involucradas: tiempo filtrado por año, aduana filtrada por Santamaría y Limón, mercancía filtrada por apertura nacional. Variables de medida: valor_cif.

CONSULTA 3: Nombre y cédula jurídica del importador, en la Aduana Santamaría, cuya mercancía importada, clasificada al nivel arancelario de apertura nacional, tiene el mayor valor CIF en el año 2000. Debe obtenerse el código arancelario de dicha mercancía a 10 dígitos.

o DURACIÓN DE LA CONSULTA: 4:19 minutos.

Dimensiones involucradas: importador, tiempo filtrado por año, aduana filtrada por Santamaría, mercancía filtrada por apertura nacional. Variables de medida: valor_cif.

5.5 Acceso al Mercado de datos por Internet

Con el desarrollo de un enlace por Internet al MD, se pretende facilitar el acceso de la información a los usuarios alejados en las aduanas y al nivel ejecutivo del Ministerio de Hacienda y el Servicio Nacional de Aduanas. Además, pretende servir de base para una futura expansión del prototipo, en un eventual proyecto de Data Warehouse a escala general.

5.5.1 Internet Database Connector (IDC)

Para el acceso a los datos contenidos en el MD, se utiliza un servidor con Microsoft NT Server como sistema operativo, Internet Information Server (IIS) como web server y la herramienta ODBC (Open DataBase Connectivity) Internet Database Connector (IDC) para el enlace a los datos. Este es el método más simple y directo para acceder datos de Visual FoxPro según lo afirma [Bazi00]. Por su parte [Cort97] añade

que IDC ofrece un mecanismo directo de alto rendimiento para la integración del contenido de una base de datos dentro de una página Web.

IDC es un componente integral de IIS y constituye una herramienta simple, basada en una secuencia de comandos que permite el acceso a tablas de Visual FoxPro o cualquier otro origen de datos que se pueda acceder por medio de ODBC. Una aplicación IDC consiste de tres documentos: 1- el formulario HTML que envía los parámetros a consultar, 2- el segundo documento, recibe los parámetros y contiene la sentencia SQL de la consulta, y 3- el tercer documento es un archivo de extensión HTML (.htx) con una sintaxis especial para hacer el despliegue de los resultados. Cada vez que un usuario hace una petición a un archivo IDC, la consulta asociada a él se ejecuta como un programa de enlace de librería dinámica para un Internet Server Application Program Interface (DLL/ISAPI), y se comunica a la base de datos Visual FoxPro por medio de ODBC.

Conceptualmente, el acceso a la bases de dato es realizado por el IIS, como se muestra en la figura 39. El visualizador del cliente (browser) envía las consultas al servidor Web a través del protocolo de transferencia de hipertexto (http). El servidor Web responde con un documento en formato HTML montado sobre una plantilla. El acceso a la base de datos es proporcionado por el componente IDC, que es un programa DLL/ISAPI (httpodbc.dll) que utiliza ODBC para establecer dicho acceso.

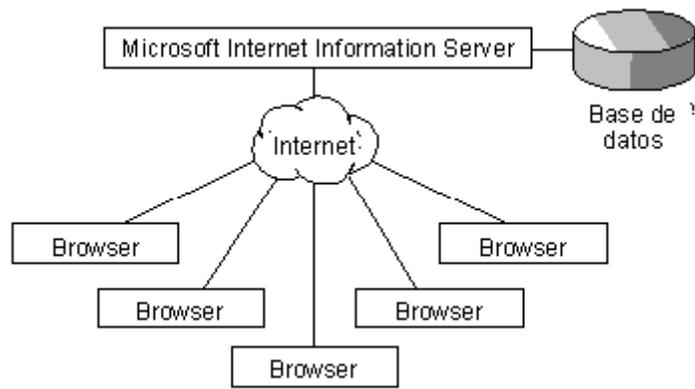


FIGURA 39: Acceso a bases de datos por IIS

En la figura 40 se muestran los componentes para conectarse a bases de datos desde el IIS.

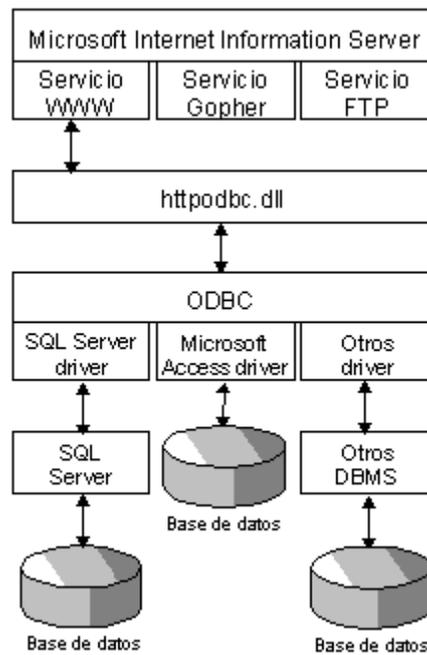


FIGURA 40: Componentes de conexión de IIS

Los archivos IDC contienen la información necesaria para conectarse a la fuente de datos ODBC apropiada y para ejecutar la sentencia SQL. También contiene el nombre y ubicación del archivo HTX. Un archivo IDC tiene la siguiente estructura:

Datasource: QUEVFP (nombre del origen de datos) Template: <nombre del archivo de plantilla htx donde se enviarán los datos> SQLStatement: +<estatutos SQL para el acceso a los datos>
--

El archivo HTX es la plantilla para el documento HTML que será devuelto al visualizador, después de que la información de la base de datos haya sido fusionada dentro de dicho documento por el programa IDC, lo cual hace de forma dinámica. El contenido de un archivo HTX da el formato a los datos recuperados. Contiene las etiquetas HTML <%BeginDetail%> y <%EndDetail%>, las cuales proporcionan una estructura de bucle que se ejecutará a través del cursor resultante. Cualquier HTML que esté entre estas dos etiquetas, se genera para cada uno de los registros y los nombres de los campos se pueden incrustar empleando la sintaxis <%nombrecampo%>.

5.5.2 Construcción del enlace

Para la construcción del enlace del MD a Internet, se definieron las consultas que podrían requerir los altos ejecutivos del Ministerio de Hacienda, los gerentes o las jefaturas de departamento de las aduanas. Se toman estos usuarios clave como base para la definición debido a que, al encontrarse geográficamente distanciados de la DCF, no es posible el acceso directo y personal a la interfase de consulta principal.

Las consultas disponibles que se plantean son (vease figura 41):

- Consultas generales
 - Aduana e importador
 - Aduana y agencia
 - Aduana y lugar de descarga
- Consultas específicas por aduana y lugar de descarga
 - Aduana e importador
 - Aduana y agencia
 - Aduana y lugar de descarga por agencia
- Consulta por mercancías
 - General por aduana
 - Aduana e importador
 - Aduana y agencia
 - Aduana y lugar de descarga
 - Aduana y país de origen
- Consulta por tipo de revisión
- Mayores importadores mensuales

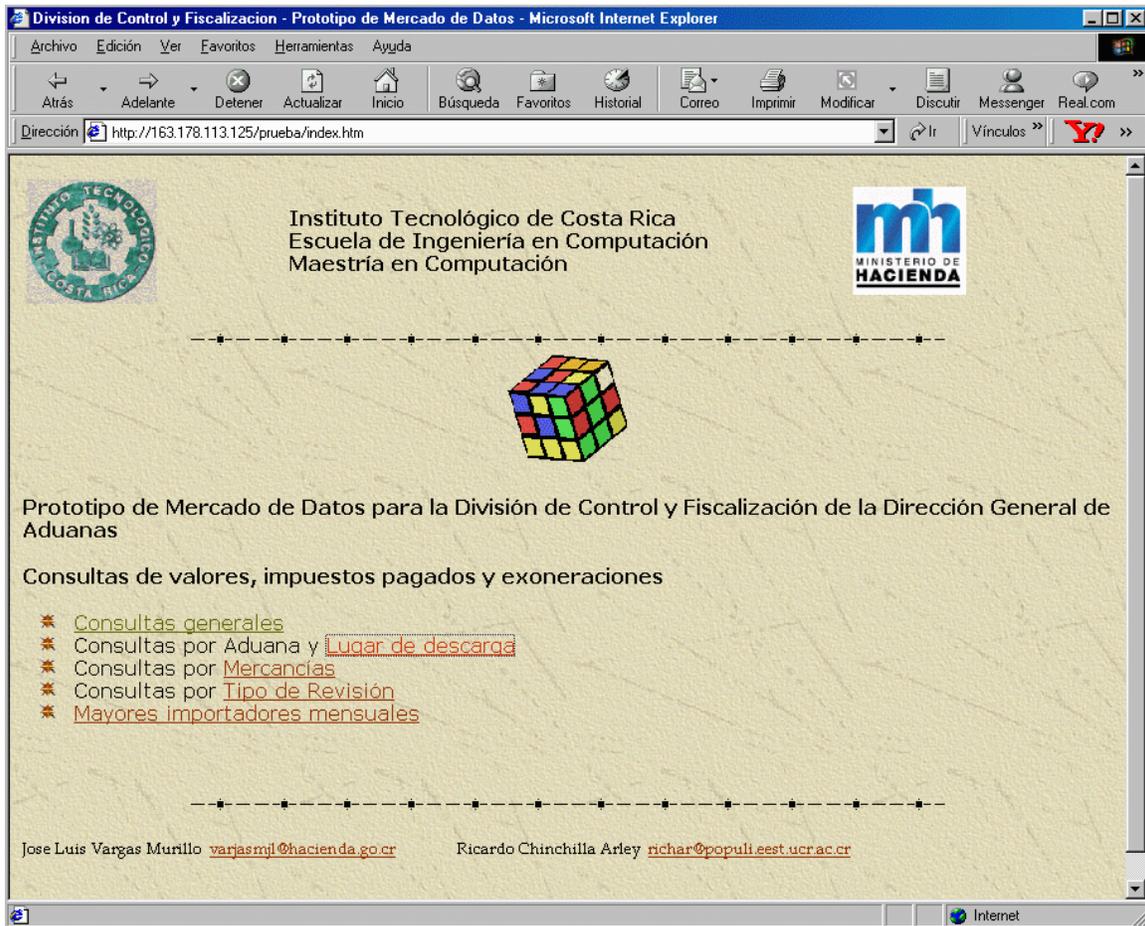


FIGURA 41: Pantalla principal de consulta por WWW

Una vez especificadas, se llevan a cabo varios pasos clave.

El primer paso es la construcción de una tabla libre agregada, llamada Consulta-Web, que utiliza la siguiente instrucción de SQL como mecanismo de agregación:

```

SELECT T.Cod_Año,T.Año,T.Cod_mes,T.Mes,D.Cod_Aduana,D.Aduana,A.Cod_Agencia,A.Agencia,;
B.Ced_importador,B.Importador,C.Cod_Lugar_Descarga,C.Lugar_Descarga,F.Cod_Semaforo,F.Semaforo,;
G.Cod_Pais_Origen,G.Pais_Origen,H.Cod_Pais_Procedencia,H.Pais_Procedencia,J.cod_Mercancia,;
SUM (valor_cif) AS valor_cif,;
SUM (dai) AS dai,;
SUM (sc) AS sc,;
SUM (ley6946) AS ley6946,;
SUM (iv) AS iv,;
SUM (golfito) AS golfito,;
SUM (mexico) AS mexico,;

```

```

SUM (total_impuesto) AS total_impuesto,;
SUM (total_exonerado) AS total_exonerado,;
SUM (num_bultos) AS num_bultos;

FROM md_dcf!tiempo AS T;
INNER JOIN md_dcf!importacion_defi ON T.añomes = Importacion_defi.añomes;
INNER JOIN md_dcf!aduana AS D ON D.cod_aduana = Importacion_defi.cod_aduana;
INNER JOIN md_dcf!agencia AS A ON A.cod_agencia = Importacion_defi.cod_agencia;
INNER JOIN md_dcf!importador AS B ON B.ced_importador = Importacion_defi.ced_importador;
INNER JOIN md_dcf!lugar_descarga AS C ON C.cod_lugar_descarga = Importacion_defi.cod_lugar_descarga;
INNER JOIN md_dcf!semaforo AS F ON F.cod_semaforo = Importacion_defi.cod_semaforo;
INNER JOIN md_dcf!pais_origen AS G ON G.cod_pais_origen = Importacion_defi.cod_pais_origen;
INNER JOIN md_dcf!pais_procedencia AS H ON H.cod_pais_procedencia = importacion_defi.cod_pais_procedencia;
INNER JOIN md_dcf!Mercancia AS J ON J.cod_Mercancia = Importacion_defi.cod_mercancia;

GROUP BY T.Cod_Año,T.Cod_Mes,D.Cod_Aduana,A.Cod_Agencia,B.Ced_Importador,C.Cod_Lugar_Descarga,;
F.Cod_Semaforo,G.Cod_Pais_Origen,H.Cod_Pais_Procedencia,J.cod_Mercancia;

INTO TABLE ConsultaWeb

```

Los archivos html, idc y htx deben residir en un directorio determinado con derechos de ejecución en IIS y dentro de c:/InetPub/wwwroot.

El segundo paso es la configuración del origen de datos ODBC mediante el administrador de orígenes de datos ODBC de NT. Para ello se define un origen tipo sistema, utilizando Microsoft VisualFoxPro Drivers con el nombre QueVFP direccionado al directorio de la tabla libre.

Como tercer paso se encuentra la creación de los archivos IDC, los cuales especifican los parámetros que contienen información empleada para acceder una base de datos ODBC. Para la interfase se construyeron los archivos IDC necesarios para cada consulta, éstos reciben los parámetros enviados por los formularios de consulta (mayores detalles de los archivos IDC generados, se muestran en el Apéndice C). Para el despliegue de los datos, se utilizan plantillas HTX para cada una de las consultas.

Por último, se crean los formularios de captura de la consulta. Dichos formularios envían parámetros ingresados por el usuario en elementos *select* tipo *multiple*, y controlados por Java Script. Estos últimos realizan dicho control por medio de funciones llamadas

por el manejador de eventos *onChange*. Por ejemplo, para el control de los Tipos de Revisión, se utiliza la función *revi*:

```
function revi()  
{  
  if (document.forms[0].revision.options[0].selected==true){  
    document.forms[0].revision.options[1].selected = true  
    document.forms[0].revision.options[2].selected = true  
    document.forms[0].revision.options[3].selected = true }  
  else document.forms[0].revision.options.value.multiple = true;  
}
```

De esta forma se permite ya sea la selección de uno, varios o todos los elementos en forma múltiple (véase ejemplo de figura 42).

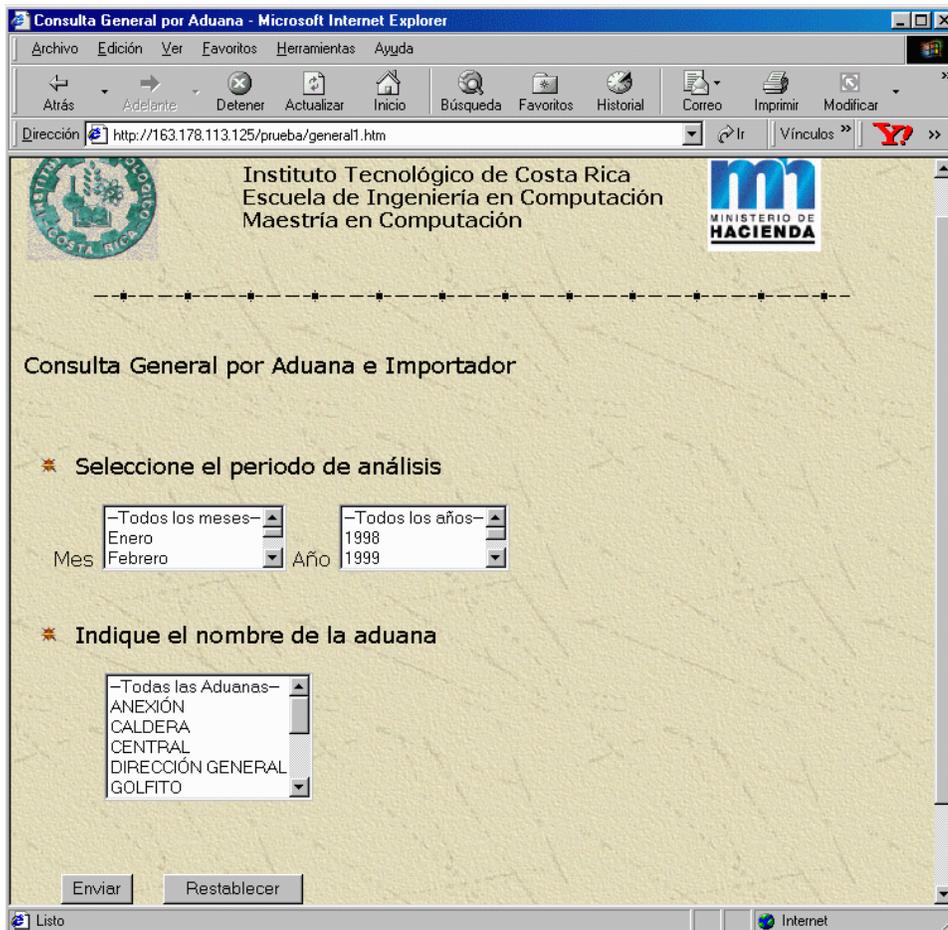


FIGURA 42: Consulta general por aduana e importador

El resultado de una consulta sería desplegado dentro de una plantilla HTX, tal y como fue descrito. Un ejemplo de despliegue de datos puede verse en la figura 43, donde se realiza una búsqueda de los 10 mayores importadores del país.

Resultados de búsqueda

Año	Mes	Aduana	Importador	Valor_cif
2000	Enero	LIMÓN	ROHM Y HAAS C.A. S.A.	\$240250
2000	Enero	LIMÓN	INVERSIONES CINCO MIL LTDA	\$216166.3
2000	Enero	LIMÓN	COMPANIA MUNDIMAR, S.A.-	\$206812.24
2000	Enero	LIMÓN	STANDART FRUIT COMPANY DE C.R. S.A.	\$190790.4
2000	Enero	LIMÓN	BIS COSTA RICA S.A.	\$173321.69
2000	Febrero	CENTRAL	CAJA COSTARRRICENSE DE SEGURO SOCIAL	\$171410.22
2000	Noviembre	LIMÓN	LABORATORIOS GRIFFITH	\$164324.59
2000	Enero	LIMÓN	VIDRIERA CENTROAMERICANA,S.A.	\$141442.84
2000	Noviembre	LIMÓN	CORRUGADOS BELEN.S.A.	\$136332.83
2000	Noviembre	LIMÓN	TRANSMERQUIN DE COSTA RICA, S.A.	\$99609.69

FIGURA 43: Consulta de los 10 mayores importadores

5.6 Requerimientos de hardware

Definir requerimientos de hardware actualmente resulta una tarea bastante efímera, debido a que cada año se desarrollan computadoras nuevas y más rápidas. Es una tendencia de la industria de alta tecnología que hace que la velocidad de los microprocesadores se doble cada año, sino antes. Sin embargo, hay cinco aspectos generales que deben ser considerados a la hora de tomar una decisión en la adquisición y utilización de equipo de cómputo.

El primer aspecto es la utilización de la capacidad de procesamiento de datos en los DD que, a diferencia de los sistemas transaccionales donde es fácil de predecir, en aquellos es esporádica y nunca se puede saber con certeza cómo y cuánto afectará los recursos del sistema. Esto último debido a que no es posible determinar por adelantado cuales serán las consultas que se realizarán y su complejidad. Se debe recordar que, en un procesamiento de tipo analítico, se utilizan grandes conjuntos de datos. La conclusión a la que se llega rápidamente y la cual anota [Core97], es que no deben mezclarse nunca los sistemas de DD con los sistemas transaccionales.

Generalmente, por razones presupuestarias, se desea que un mismo sistema de computadoras realice tanto las actividades de DD como las transaccionales, pero ésta práctica no es recomendable y, a la postre, puede traerse abajo el sistema completo. Es necesario entonces, contar con un sistema computacional exclusivo para la administración de un MD. Se recomienda para el prototipo propuesto, un servidor con al menos 2 procesadores de alto rendimiento, de 1 a 2 Mb de memoria caché para cada uno de ellos, al menos 1 Gb de memoria RAM y un servidor web que administre las consultas que se realizan a la tabla agregada general que presta servicios por Internet.

El segundo aspecto se refiere al almacenamiento de los datos en disco y su acceso. [Core97] cita un refrán popular: “la fuerza de la cadena depende del eslabón más débil” y en un DD dicho eslabón lo constituye el flujo de entrada y salida de datos. El mismo autor indica que a pesar de que han aumentado las capacidades de E/S de los discos, la evolución no ha sido tan rápida como la de los procesadores. Por este motivo, el acceso a los discos es la operación más lenta que realiza la computadora. Dicha lentitud, debido a los grandes volúmenes de datos que maneja, se hace más palpable en el ambiente de DD. Además, un fallo en un disco compromete no solo la disponibilidad de los datos, sino también la fiabilidad de los datos que tiene almacenados.

Una tecnología que ha surgido con la finalidad de resolver los problemas de almacenamiento y el rápido acceso a los datos, es el llamado vector redundante de discos independientes (de l inglés Redundant Array of Independent Disk), o RAID.

La tecnología RAID, que tiene varios niveles de aplicación, utiliza dos o más discos donde se reparte y almacena la información. Por ejemplo, en el nivel 5 la información se almacena por bloques y un bloque de cada disco se dedica a paridad, es decir, los datos de paridad se añaden como otro sector que rota por los discos igual que los datos ordinarios. Esto permite recuperar la información almacenada en un disco que falla. Para ello se conectan al servidor mediante un canal E/S como si fueran un solo disco.

La importancia de la redundancia radica en que se manejan volúmenes de datos que fácilmente pueden llegar a los cientos de gigabytes y que requieren varios discos, lo que aumenta la posibilidad de falla. El RAID soporta el relevo de discos sin afectar la disponibilidad de la información por parte de los usuaios.

El modelo propuesto requiere entonces, contar con un arreglo RAID a nivel 5 (requiere un mínimo de 3 discos) dotado de controladores SCSI con al menos 50 Gb. internos y poseer una amplia capacidad de crecimiento.

El tercer aspecto a tomar en cuenta es la escalabilidad y el rendimiento. En un DD éstas son características necesarias que deben considerarse a la hora de tomar una decisión sobre la adquisición de hardware, dados los grandes volúmenes de datos que se procesan y lo valiosos que estos pueden llegar a ser.

La escalabilidad tiene dos aspectos: la capacidad de crecer en tamaño y la capacidad de mantener el rendimiento durante el crecimiento. Este crecimiento debe ser balanceado y para ello hay que buscar productos que sean modulares en su diseño, que permitan además incrementar la cantidad de memoria caché, buses, controladores y canales, proporcionalmente a la cantidad de discos que se agreguen. Todo esto de acuerdo a sus capacidades óptimas de operación y en procura de evitar que algún componente se convierta en un cuello de botella.

El cuarto aspecto se refiere al respaldo de los datos. La tecnología de cinta magnética continúa siendo uno de los mecanismos de almacenamiento masivo para respaldo más eficientes, dado su relativo bajo costo y reducido tamaño. Debido a la gran cantidad de datos a respaldar, es recomendable la utilización de la tecnología de cintas de respaldo, adquiriendo unidades de cinta DAT o DTL de 35/70 Gb.

El quinto aspecto a tomar en cuenta se refiere a los costos. [Mads98] indica que el costo por almacenamiento puede ser más del 50% del precio del sistema total e indica que dicho costo se deriva de dos aspectos: la inversión inicial de almacenamiento y el costo por crecimiento incremental. En cuanto al primer aspecto, aconseja adquirir sólo el hardware necesario para el volumen de datos inicial y presupuestar las necesidades futuras,

pues con el avance tecnológico los precios tienden a bajar, logrando minimizar tanto el costo inicial como el costo incremental. Este último aspecto, puede corresponder a incrementos pequeños, como la compra de discos individuales, o incrementos grandes como la compra de arreglos completos.

Capítulo 6. Conclusiones y recomendaciones

6.1 Conclusiones

- Las entrevistas y el análisis de consultas determinó que los usuarios potenciales del Mercado de Datos propuesto son los funcionarios de los departamentos Interno, Externo, Planificación, Denuncias y la jefatura de la división, así como los gerentes de aduanas y sus mandos medios.
- La aplicación de la entrevista dejó de manifiesto los problemas de información existentes en cada uno de los departamentos de la División de Control y Fiscalización, así como los problemas derivados de la inadecuada manipulación de los datos a través de las herramientas de análisis de que disponen. Dicha manipulación se debe, por un lado a que las herramientas no han sido diseñadas para el manejo de grandes volúmenes de datos, y por otro lado, a la falta de capacitación en su uso.
- La base de datos constuida para el prototipo propuesto, la cual se basa en un esquema tipo estrella, ha demostrado tener la capacidad de multidimensionalidad al lograr integrar las distintas dimensiones y medidas, establecidas como las requeridas por los usuarios en un Mercado de Datos de mercancías importadas.
- A pesar de que el diseño de una base de datos ya ha sido tratado a profundidad por muchos autores y se han definido las metodologías necesarias para su desarrollo, en el caso de las bases de datos multidimensionales no se contaba con un modelo para el diseño conceptual. La metodología definida por [Cort99] constituye un aporte en esta materia y define un rumbo que facilita el desarrollo de esta tecnología.

- Los programas de carga desarrollados para la aplicación, logran transformar y cargar los datos provenientes de las tablas planas extraídas del SIA de manera satisfactoria y confiable.
- La interface de consulta construida facilita el análisis multidimensional de los datos, mediante la construcción de tablas de resultados que almacenan la información de las distintas dimensiones de forma agrupada y utilizando reducción dimensional. Dichos resultados, obtenidos de manera interactiva, le permiten al usuario tener distintas perspectivas de la información conforme a sus necesidades y de acuerdo a las dimensiones y unidades de medida que se consideran en la consulta.
- La implementación de la solución propuesta contribuirá significativamente en la labor de investigación y en el ejercicio del control fiscal. Tal contribución se verá reflejada, en primer término, en el rendimiento y tiempo de análisis y en segundo término en la calidad de los hallazgos derivados de su aplicación. Actualmente se tarda días e inclusive semanas en el análisis de un caso. Con el prototipo, que facilita la búsqueda y el análisis interactivo de la información, este tiempo se reduce a pocos minutos.
- El modelo propuesto logra la incorporación de algunas consultas generales a través de Internet, por medio del acceso a una tabla agregada especialmente construida para este fin. Su implementación se ha enfocado en resolver problemas puntuales de información de aquellos usuarios que requieren acceder el Mercado de Datos de forma remota.

6.2 Recomendaciones

- Es necesario, antes de implementar la solución propuesta, capacitar a los usuarios en el uso eficiente de ésta y en el aprendizaje de los conceptos y temas relacionados con el análisis multidimensional para el apoyo a la toma de decisiones. Al estar los usuarios bien identificados, es posible focalizar esta capacitación de acuerdo a sus necesidades particulares y la naturaleza de sus labores.
- La DGA debe realizar esfuerzos en la capacitación de usuarios en materia de herramientas de análisis y desarrollo multidimensional.
- Es necesario la evaluación periódica de la aplicación, analizando las recomendaciones de los usuarios e incorporandolas al Mercado de Datos.
- El prototipo utiliza el esquema estrella en la base de datos. Se recomienda, para una implementación final, incorporarle tablas agregadas y registros sumariados para cada una de las combinaciones posibles. Esto con el fin de permitir una mayor velocidad en el procesamiento de consultas.
- El prototipo evita la carga de archivos duplicados, sin embargo para mayor seguridad es necesario implementar un sistema de administración y control de los archivos de transacciones mensuales remitidos por las aduanas. Es recomendable eliminar las inconsistencias, ambigüedades y duplicación de la información de la gestión aduanera, y que la administración de dicha información sea ordenada, normalizada y automatizada.

- Es necesario la incorporación de funciones graficadoras y de sumarización a la interfase de consulta, a fin de convertirla en un OLAP relacional y aumentar el potencial de análisis de los usuarios.
- El acceso por Internet a consultas preestablecidas, tal y como se ofrece actualmente en el prototipo, puede ser ampliado incorporando mayor flexibilidad en los formularios y en los programas asociados que efectúan la extracción de información de la tabla agregada.
- A futuro, es recomendable una conexión directa al Mercado de Datos completo, y no a una sola tabla agregada.
- Una vez instalado el prototipo y evaluadas sus capacidades y posibilidades de apoyo a la administración, es recomendable su desarrollo futuro sobre tecnologías de base de datos multidimensionales más robustas.

Bibliografía

[Bazi00] Bazian, Menachen. Visual FoxPro 6: edición especial. México: Prentice-Hall, 2000

[Bisc93] Bischoff, Joyce. “Designing efficient DB2 decision support systems”. Von Halle, Barbara; Kull, David, eds. Handbook of data management. Boston: Auerbach Publications, 1993

[Cata97] Cataldo, Joseph. “Care feeding data warehouse”. Database programming and design. December, 1997
<http://www.dbpd.com/vault/9712toc.htm>

[Cohe96] Cohen, Daniel. Sistemas de información para la toma de decisiones. México: McGraw-Hill, 1996

[Core97] Corey, Michael; Abbey, Michael. Oracle Data Warehousing: guía práctica para analizar, construir e implantar con éxito un sistema data warehouse. Madrid : McGraw-Hill, 1997

[Cort97] Cortés, Carlos; Meléndez Jaime; Tobar, Mercedes. “Internet Database Connector”. Interfaz CGI para Servidores Web y Sistemas de Administración de Bases de Datos. San Salvador: Universidad Centroamericana José Simeón Cañas , 1997
<http://168.243.1.4/investigacion/bdweb/reportes/idc.html>

[Cort99] Cortez Araniva, Sonia Elizabeth. Desarrollo de un Modelo para el Diseño de Bases de Datos Multidimensionales. Cartago: ITCR, 1999

[Cox00] Cox Alvarado, Alexander. Propuesta de un álgebra para modelar bases de datos multidimensionales. Cartago: ITCR, 2000

[Dane96] Danesh, Arman. Aprendiendo JavaScript en una semana. México: Prentice-Hall Hispanoamericana, 1996

[Data96] Data cube: a relational aggregation operator generalizing Group By, Cross-Tab, and Sub-Totals. Microsoft Technical report MSR-TR-95-22, June, 1996

[Devl96] Devlin, Barry. Data warehouse: from architecture to implementation. New York: Addison-Wesley, 1996

[Ebel98] Ebel, Doug. Data Warehousing: start to start. NCR Business Solution Architect. S.l.: NCR Corporation, 1998
<http://www.sdw.bull.com/ncr/dwstart.pdf>

[Gido99] Gido, Jack; Clements, James. Administración exitosa de proyectos. México: Thomson, 1999

[Gill97] Gill, H.S; Rao, P.C. Data Warehousing: la integración de información para la mejor toma de decisiones. Prentice-Hall, 1997.

[Gonz96] González Alvarado, Carlos. Sistemas de bases de datos. Cartago: Editorial Tecnológica de Costa Rica, 1996

[Gonz98a] González Alvarado, Carlos. Depósito de datos. Cartago: ITCR, (1998)

[Gonz98b] González Alvarado, Carlos. Procesamiento analítico en línea. Cartago: ITCR, (1998)

[Gude98] Guderian, Dave; Leer, Doug; Molini, Steve. "Keeppeing the data warehouse in track" Database programming and design. January, 1998

[Hari96] Harinarayan, V.; Rajaraman, A.; Ullman, J. "Implementing data cubes efficiently" SIGMOD'96 6/96, Montreal, Canada, 1996
<http://www-db.stanford.edu/pub/harinarayan/1995/cube.ps>

[Herr00] Herrán Gascón, Manuel de la; Castellar-Busó, Vicent. "Cómo diseñar grandes variables en bases de datos multidimensionales" Revista digital universitaria, jul, 2000
<http://www.revista.unam.mx/vol.1/art5/index.html>

[Imie96] Imielinski, Tomasz; Manilla, Hikki. "Database Approach to Knowledge Discovery." Communications of ACM. November, 1996

[Kamf98] Kamfonas, Michael. "Where version meet dimensions" Database programming and design. September., 1998

[Kimb97] Kimball, Ralph. "A dimensional modeling manifesto". DBMS. August, 1997

[Koch96] Kochhar, Neena; Kramer, Debby. Introduccction to Oracle: SQL and PL/SQL using procedure builder. Redwood Shores, CA: ORACLE Corporation, 1996
<http://education.oracle.com>

[Laud96] Laudon, Kenneth; Laudon, Jane. Administración de los sistemas de información: organización y tecnología. México: Prentice-Hall, 1996

[Lewi94] Lewinson, Lisa. "Data Mining: Tapping Into the Mother Lode." Database programming and design. feb., 1994

[Mads96] Madsen, Mark. "Warehouse Design in the Agregate". Database programming and design. July, 1996
<http://www.dbpd.com/vault/julytoc.htm>

[Mads98] Madsen, Mark. "Big warehouse, big decisions". Database programming and design. April, 1998

[Mayn83] Mayne, Alan; Wood, Michael. Introducción a las bases de datos relacionales. Madrid: Ediciones Díaz de Santos, 1983.

[Micr98] Microsoft Corporation. Microsoft SQL Server: getting started with SQL Server 7.0. Microsoft Corporation, 1998 (Books on Line)

[Orac96] Oracle Corporation. Design a database for OLAP. March/April, 1996
<http://www.oramag.com/oracle/96-mar/26meth.html>

[Pars96a] Parsaye, Kamran. "OLAP and Data Mining: Bridging the Gap". Database programming design. February, 1996
<http://www.dbpd.com/vault/parsfeb.htm>

[Pars96b] Parsaye, Kamran "Surveying Decision Support: New Realms of Analysis". Database programming desing. April, 1996

[Pere99] Pérez Martínez, Cayetano. Replicación de bases de datos por Internet. México: Instituto Politécnico Nacional, 1999
<http://148.204.20.3/PROYECTOS/TESIS/cayetano/DTeMae.htm>

[Poe98] Poe, Vidette. Building a data warehouse for decision support. New Jersey: Prentice-Hall, 1998

[Rees97] Reese, Joseph. "Making the datawarehouse work overtime". Database programming design. September, 1997

[Rodr99] Rodríguez, Nuria; Martínez, William. Planificación y evaluación de proyectos informáticos. San José: EUNED, 1999

[Senn93] Senn, James. Análisis y diseño de sistemas de información. México: McGraw-Hill, 1993

[Simo96] Simon, Alan. "Beyond the warehouse". Database programming desing. december, 1996

[Sing94] Singleton, J.; Schwartz, M. "Accesando datos dentro de la estructura de un depósito de información" IBM Systems Journal. Vol.33, no. 2, 1994

[Stra99] Strahl, Rick. Internet applications with Visual FoxPro 6.0. Whiterfish Bay, WI: Hentzenwerke Publishing, 1999

[Thom97] Thomsen, Erik. OLAP solution: building multidimensional information system. New York: John Wiley & Sons, 1997

[Wel95] Weldon, Jay; Weldon, Louise. "Managing multidimensional data". Database programming and design. august., 1995

[Wel96] Weldon, Jay; Weldon, Louise. *"Choosing Tools for Multidimensional Data"*. Database programming and design. February., 1996

[Wolf00] Wolff, Carmen. *"Implementando un DataWarehouse"*. Revista Ingeniería Informática: revista electrónica del DIICC. Edición 5, año 3 [2000]
<http://www.inf.udec.cl/revista/edicion5/cwolff.htm>

[Zimm96] Zimmerman, Scott; Brown, Christopher. Kit de construcción de sitios Web para Windows 95. México: Prentice-Hall Hispanoamericana, 1996

<http://www.oracle.com/olap>

<http://www.oramag.com/oracle>

<http://www.olapreport.com/glosary.htm>

Apéndice A. Diccionario de Datos

Diccionario de Datos

Importador

Campo	Tipo	Long.	Descripción
Identificación del importador <u>ced_importador</u>	C	12	Número de cédula o pasaporte del Importador.
Importador Importador	C	40	Razón Social Completa con Clase de Sociedad (ej. S.C. , R.L), o dos Apellidos y Nombre en caso de persona física.

Agencia

Campo	Tipo	Long.	Descripción
Código de la agencia <u>cod_agencia</u>	C	3	Código de la agencia que tramita
Cédula jurídica <u>ced_jurídica_agencia</u>	C	11	Cédula jurídica de la agencia aduanera
Agencia <u>Agencia</u>	C	11	Nombre de la agencia aduanera
Teléfono <u>x_teléfono</u>	C	11	Número telefónico de la agencia aduanera
Dirección <u>x_dirección</u>	C	11	Dirección física de la agencia aduanera
Fax <u>x_fax</u>	C	11	Número de fax de la agencia aduanera
Representante <u>x_representante</u>	C	11	Nombre completo del representante de la agencia aduanera

Aduana

Campo	Tipo	Long.	Descripción
Código de aduana <u>cod_aduana</u>	N	2	Código de la aduana donde se realizó el trámite de desalmacenaje.
Aduana	C	20	Nombre de la aduana o puesto

aduana			
--------	--	--	--

País_origen

Campo	Tipo	Long.	Descripción
País de Origen <u>cod_pais_origen</u>	C	4	Código ISO del país de origen de la mercancía.
Nombre del país <u>pais_origen</u>	C	50	Nombre del país de donde es originaria la mercancía

País_procedencia

Campo	Tipo	Long.	Descripción
País de procedencia <u>Cod_pais_procedencia</u>	C	50	Código ISO del país de procedencia de la mercancía.
Nombre del país <u>pais_procedencia</u>	C	4	Nombre del país de embarque de las mercancías

Importación_defi

Campo	Tipo	Long.	Descripción
Total de bultos <u>numero_bultos</u>	N	12	Total de bultos solicitados a despacho.
Impuesto de ventas <u>Iv</u>	N	11,2	Monto correspondiente al Impuesto de Ventas
Derechos Arancelarios a la Importación <u>Dai</u>	N	11,2	Monto correspondiente a los Derechos Arancelarios a la Importación.
Impuesto Selectivo de Consumo <u>Sc</u>	N	11,2	Monto correspondiente al Impuesto Selectivo de Consumo
Valor de costo, seguro y flete <u>valor_cif</u>	N	10,2	Es el valor manifestado en el formulario por parte del declarante, constituido por la suma en dólares de los montos: monto de la factura, flete, seguro y otros gastos.

Ley 6946 Ley6946	N	11,2	Monto correspondiente al impuesto de la Ley 6946
Golfito	N	11,2	Impuesto pagado por las mercancías compradas en el Depósito Libre de Golfito
mexico	N	11,2	Impuesto pagado por las mercancías importadas bajo el Tratado de Libre Comercio con México
Total de impuestos total_impuestos	N	11,2	Monto total de colones a pagar por concepto de impuestos de importación de la mercancía
Total de exoneración total_exoneraciones	N	11,2	Monto total de colones exonerados del pago de impuestos

Mercancías

Campo	Tipo	Long.	Descripción
Sistema Arancelario Centroamericano cod_mercancía	C	10	Código del Sistema Arancelario Centroamericano (SAC), correspondiente a la mercancía declarada (diez dígitos).
Capítulo cod_capítulo	C	2	Código correspondiente a los primeros dos dígitos del Arancel de mercancías
Partida cod_partida	C	4	Código correspondiente a los primeros cuatro dígitos del Arancel de mercancías
Descripción Mercacias	M	4	Descripción de las mercancías importadas

Definición de jerarquía

J ₁ :	P ₁ : * → sección	Plega todas las secciones
	P ₂ : sección → capítulo	Plega todos los capítulos por sección
	P ₃ : capítulo → partida	Plega todas las partidas por capítulo

Lugar_descarga

Campo	Tipo	Long.	Descripción
Tipo de ubicación	C	1	Código del tipo de ubicación en que se

X_cod_tipo_lugar			encuentra la mercancía.
Código de ubicación cod_lugar_descarga	C	3	Código de la ubicación específica de la mercancía.
Descripción del tipo de ubicación X_cod_lugar_descarga	C	20	Descripción de los tipos de ubicación existentes
Descripción de la ubicación específica Tipo_lugar	C	20	Descripción de cada uno de los tipos de ubicación específica

Semáforo

Campo	Tipo	Long.	Descripción
Revisión cod_semaforo	C	1	Código del tipo de revisión efectuada a la mercancía
Descripción Semáforo	C	25	Descripción del tipo de revisión: sin revisión, revisión documental o revisión física.

Trámite

Campo	Tipo	Long.	Descripción
Código de trámite cod_tramite	C	1	Código del trámite que se lleva a cabo con la mercancía.
Trámite Tramite	C	20	Tipo de trámite aplicable a la mercancía

Modalidad

Campo	Tipo	Long.	Descripción
Modalidad cod_modalidad	C	2	Código de tipo de modalidad.
Descripción Modalidad	C	60	Tipo de naturaleza de la operación

Tiempo

Campo	Tipo	Long.	Descripción
Año y mes <u>Añomes</u>	C	6	Año y mes
Código de año Cod_año	C	2	Dos últimos dígitos del año
Año Año	N	4	Año al que pertenecen los datos
Código del trimestre Cod_trimestre	C	1	Número ordinal entre 1 y 4
Trimestre trimestre	N	1	Trimestre al que pertenecen los datos
Código del mes Cod_mes	C	2	Número ordinal entre 1 y 12
Mes Mes	N	2	Mes al que pertenecen los datos
Código del año fiscal Cod_añofiscal	C	2	Dos últimos dígitos del año
Año fiscal Añofiscal	N	4	Año fiscal, comprendido entre el último trimestre de un año al tercer trimestre del siguiente año (del el 1 de octubre al 31 de septiembre)

Definición de jerarquía

J ₁ :	P ₁ : * → años	Plega todos los años
	P ₂ : año → trimestre	Plega todos los trimestres por año
	P ₃ : trimestre → mes	Plega todos los meses por trimestre
J ₂	P ₁ : * → Años fiscal	Pliega todos los años fiscales

Apéndice B. Código fuente, programa de carga

```

*****
*-- Form:          cargapd (c:\md_dcf\cargapd.scx)
*-- ParentClass:  form
*-- BaseClass:    form
*
DEFINE CLASS cargapd AS form

    Top = 22
    Left = 46
    Height = 372
    Width = 433
    DoCreate = .T.
    Caption = "Carga archivos .dat"
    Name = "CARGAPD"

    ADD OBJECT cmdcargapd AS commandbutton WITH ;
        Top = 267, ;
        Left = 312, ;
        Height = 35, ;
        Width = 105, ;
        WordWrap = .T., ;
        Caption = "Cargar archivos seleccionados", ;
        Name = "cmdCargaPD"

    ADD OBJECT cmdcerrar AS commandbutton WITH ;
        Top = 324, ;
        Left = 312, ;
        Height = 24, ;
        Width = 105, ;
        Caption = "Salir", ;
        Name = "cmdCerrar"

    ADD OBJECT moverlist1 AS moverlist WITH ;
        Top = 122, ;
        Left = 12, ;
        Width = 279, ;
        Height = 216, ;
        BorderColor = RGB(192,192,192), ;
        Name = "Moverlist1", ;
        lstSource.Height = 213, ;
        lstSource.Left = 0, ;
        lstSource.Top = 0, ;
        lstSource.Width = 106, ;
        lstSource.Name = "lstSource", ;
        lstSelected.Height = 213, ;
        lstSelected.Left = 169, ;
        lstSelected.MoverBars = .F., ;
        lstSelected.Top = 0, ;
        lstSelected.Width = 106, ;
        lstSelected.Name = "lstSelected", ;
        cmdAdd.Top = 78, ;
        cmdAdd.Left = 113, ;
        cmdAdd.Height = 26, ;
        cmdAdd.Width = 50, ;

```

```

cmdAdd.Name = "cmdAdd", ;
cmdAddAll.Top = 111, ;
cmdAddAll.Left = 113, ;
cmdAddAll.Height = 26, ;
cmdAddAll.Width = 50, ;
cmdAddAll.Name = "cmdAddAll", ;
cmdRemove.Top = 144, ;
cmdRemove.Left = 113, ;
cmdRemove.Height = 26, ;
cmdRemove.Width = 50, ;
cmdRemove.Name = "cmdRemove", ;
cmdRemoveAll.Top = 45, ;
cmdRemoveAll.Left = 113, ;
cmdRemoveAll.Height = 26, ;
cmdRemoveAll.Width = 50, ;
cmdRemoveAll.Name = "cmdRemoveAll"

```

```

ADD OBJECT label2 AS label WITH ;
    WordWrap = .T., ;
    Caption = "Archivos de datos disponibles", ;
    Height = 33, ;
    Left = 16, ;
    Top = 74, ;
    Width = 102, ;
    Name = "Label2"

```

```

ADD OBJECT label3 AS label WITH ;
    WordWrap = .T., ;
    Caption = "Seleccione los archivos que desea cargar", ;
    Height = 44, ;
    Left = 181, ;
    Top = 65, ;
    Width = 103, ;
    Name = "Label3"

```

```

ADD OBJECT lstarchivoscargados AS listbox WITH ;
    Height = 132, ;
    Left = 312, ;
    Top = 124, ;
    Width = 109, ;
    Name = "lstArchivosCargados"

```

```

ADD OBJECT lblarchivoscargados AS label WITH ;
    WordWrap = .T., ;
    Caption = "Archivos cargados en esta sesión", ;
    Height = 30, ;
    Left = 309, ;
    Top = 79, ;
    Width = 108, ;
    Name = "lblArchivosCargados"

```

```

ADD OBJECT app_mediator AS _formmediator WITH ;
    Name = "APP_MEDIATOR"

```

```

ADD OBJECT shape1 AS shape WITH ;
    Top = 1, ;
    Left = 2, ;
    Height = 50, ;
    Width = 428, ;
    BackColor = RGB(0,0,0), ;
    Name = "Shape1"

ADD OBJECT label1 AS label WITH ;
    FontSize = 18, ;
    BackStyle = 0, ;
    Caption = "Carga de datos", ;
    Height = 35, ;
    Left = 12, ;
    Top = 12, ;
    Width = 197, ;
    ForeColor = RGB(255,255,128), ;
    Name = "Label1"

ADD OBJECT shape3 AS shape WITH ;
    Top = 56, ;
    Left = 11, ;
    Height = 1, ;
    Width = 409, ;
    FillColor = RGB(192,192,192), ;
    BorderColor = RGB(128,128,128), ;
    Name = "Shape3"

ADD OBJECT line1 AS line WITH ;
    Height = 0, ;
    Left = 12, ;
    Top = 57, ;
    Width = 408, ;
    BorderColor = RGB(255,255,255), ;
    Name = "Line1"

ADD OBJECT shape2 AS shape WITH ;
    Top = 113, ;
    Left = 10, ;
    Height = 1, ;
    Width = 409, ;
    FillColor = RGB(192,192,192), ;
    BorderColor = RGB(128,128,128), ;
    Name = "Shape2"

ADD OBJECT line2 AS line WITH ;
    Height = 0, ;
    Left = 11, ;
    Top = 114, ;
    Width = 408, ;
    BorderColor = RGB(255,255,255), ;

```

```
Name = "Line2"
```

```
PROCEDURE barradeprogreso
    PARAMETERS Tarea, Porcentaje
    LOCAL loTherm, lcTask, lnPercent, lnSeconds, lnDuracion
    loTherm = NewObject("_thermometer", "_therm", "", "Cargando
archivo:") &&THIS.Parent.txtTitle.Value)
    lcTask = Tarea &&THIS.Parent.txtTask.Value
    lnDuracion=20
    WITH loTherm
        .Show()
        FOR i = 1 TO lnDuracion
&&THIS.Parent.spnDuration.Value
            lnPercent = m.i/lnDuracion*100
&&THIS.Parent.spnDuration.Value*100
            .Update(lnPercent, lcTask+" "+TRANS(lnPercent))
            lnSeconds = SECONDS()
            DO WHILE lnSeconds+1>SECONDS()
            ENDDO
        ENDFOR
        .Complete()
    ENDWITH
ENDPROC
```

```
PROCEDURE Init
    SET DEFAULT TO C:\MD_DCF
    SET PATH TO C:\MD_DCF\PUNTODAT\AÑO2000\Central;
C:\MD_DCF\PUNTODAT\AÑO2001\Central;;
                C:\MD_DCF\PUNTODAT\AÑO2000\PeñasBlancas;
C:\MD_DCF\PUNTODAT\AÑO2001\PeñasBlancas;;
                C:\MD_DCF\PUNTODAT\AÑO2000\Caldera;
C:\MD_DCF\PUNTODAT\AÑO2001\Caldera;;
                C:\MD_DCF\PUNTODAT\AÑO2000\Santamaría;
C:\MD_DCF\PUNTODAT\AÑO2001\Santamaría;;
                C:\MD_DCF\PUNTODAT\AÑO2000\Limón;
C:\MD_DCF\PUNTODAT\AÑO2001\Limón;;
                C:\MD_DCF\PUNTODAT\AÑO2000\PasoCanoas;
C:\MD_DCF\PUNTODAT\AÑO2001\PasoCanoas
```

```
THISFORM.moverlist1.lstSource.RowSourceType = 7
THISFORM.moverlist1.lstSource.RowSource = "*.dat"
ENDPROC
```

```
PROCEDURE cmdcargapd.Click
    LOCAL nFile, lcArchivoPuntoDat, DirActual, DirActual2
    SET SAFETY OFF
    nFile=1
    DO WHILE nFile <= THISFORM.moverlist1.lstSelected.ListCount
        IF THISFORM.moverlist1.lstSelected.Selected(nFile)
            lcArchivoPuntoDat=
THISFORM.moverlist1.lstSelected.List(nFile)

        THISFORM.moverlist1.lstSelected.RemoveItem(nFile)
            IF used("ControlDat")
                SELECT ControlDat
```

```

                set order to archivodat
ELSE
    SELE 0
    USE ControlDat order archivodat
ENDIF
IF !SEEK(SUBSTR(lcArchivoPuntoDat,1,8))
    INSERT INTO ControlDat
(ArchivoDat,aduan_a_cod,mes,año) VALUES (lcArchivoPuntoDat,;

int(val(substr(lcArchivoPuntoDat,3,2))),substr(lcArchivoPuntoDat,5,2),su
bstr(lcArchivoPuntoDat,7,2))

    THISFORM.lstArchivosCargados.AddItem(lcArchivoPuntoDat)
ELSE
    #DEFINE MSG_LOC "El archivo
"+lcArchivoPuntoDat+ ".dat ya ha sido cargado al sistema"
    #DEFINE TITLE_LOC "Inserción
denegada"

    =MESSAGEBOX(MSG_LOC,64+0+0,TITLE_LOC)
    LOOP
    ENDIF

    IF used("cargadbf")
        SELECT cargaDbf
        ZAP
    ELSE
        SELE 0
        USE c:\md_dcf\md_dcf!cargadbf
EXCLUSIV
        ZAP
    ENDIF

    append from &lcArchivoPuntoDat SDF
    go bott
    skip -2
    delete rest
    delete all for cod_regimen <>"01" && solo
se trabaja con las importaciones definitivas.
    pack
    SELECT
SUBSTR(fecha,5,4)+SUBSTR(fecha,3,2)AS añosmes FROM c:\md_dcf\md_dcf!
CargaDBF GROUP BY añosmes;
        INTO TABLE AñoMesPd

    Thisform.barradeprogreso(lcArchivoPuntoDat,25)
    SELECT
SUBSTR(fecha,5,4)+SUBSTR(fecha,3,2)añosmes, cod_aduana, cod_agencia,
ced_importador,;
        cod_mercancia, cod_pais_origen,
cod_pais_procedencia, ubicacion+cod_ubicacion cod_lugar_descarga,;
        cod_regimen, cod_modalidad, cod_tramite,
cod_semaforo, SUM(valor_cif) valor_cif, SUM(dai) dai, SUM(iv) iv,
SUM(sc) sc,;
        SUM(ley6946) ley6946, SUM(golfito)
golfito, SUM(mexico) mexico, SUM(total_impuesto) total_impuesto,;

```

```

SUM(num_bultos) num_bultos,
SUM(total_exonerado) total_exonerado;
FROM c:\md_dcf\md_dcf!CargaDBF;
GROUP BY añomes, cod_aduana,
cod_agencia, cod_importador, cod_mercancia, cod_pais_origen,
cod_pais_procedencia,;
cod_lugar_descarga, cod_regimen,
cod_modalidad, cod_tramite, cod_semaforo;
INTO CURSOR cur_prehechos_agr
*
Llamada al programa de actualización
de la dimensión TIEMPO
DO actualizaTiempo
IF used("importacion_defi")
SELECT importacion_defi
APPEND FROM DBF("cur_prehechos_agr")
Thisform.barradeprogreso("Proceso de
carga en progreso, por favor espere unos minutos",25)
ELSE
SELE 0
use c:\md_dcf\md_dcf!
importacion_defi EXCLUSIV
APPEND FROM DBF("cur_prehechos_agr")
Thisform.barradeprogreso("Terminando
proceso de carga, por favor espere unos minutos",25)
ENDIF
ELSE
nFile=nFile+1
ENDIF
ENDDO
#DEFINE MSG_LOC1 "Proceso de carga concluido"
#DEFINE TITLE_LOC1 "Carga de datos"
=MESSAGEBOX(MSG_LOC1,64+0+0,TITLE_LOC1)
THISFORM.cmdCerrar.SetFocus
*SET PATH TO
ENDPROC

PROCEDURE cmdcerrar.Click
IF TYPE("THISFORM.Parent") = "0"
THISFORMSET.Release
ELSE
THISFORM.Release
CLOSE TABLES ALL
ENDIF
ENDPROC

PROCEDURE moverlist1.lstSource.DblClick
THIS.Parent.lstSelected.AddItem(This.List(This.ListIndex))
*This.RemoveItem(This.ListIndex)
ENDPROC

PROCEDURE moverlist1.lstSelected.DblClick
THIS.Parent.lstSource.AddItem(This.List(This.ListIndex))

```

```

        This.RemoveItem(This.ListIndex)
    ENDPROC

PROCEDURE moverlist1.cmdAdd.Click
    THISFORM.LockScreen = .T.
    FOR nCnt =5 TO THIS.Parent.lstSource.ListCount
        IF THIS.Parent.lstSource.Selected(nCnt)

THIS.Parent.lstSelected.AddItem(THIS.Parent.lstSource.List(nCnt))
        ENDIF
    ENDFOR
    THISFORM.LockScreen = .F.
ENDPROC

```

```

PROCEDURE moverlist1.cmdAddAll.Click
    THISFORM.LockScreen = .T.
    FOR i = 5 to THIS.Parent.lstSource.ListCount
        IF !RIGHT(THIS.Parent.lstSource.List(i),3)# "dat"

THIS.Parent.lstSelected.AddItem(THIS.Parent.lstSource.List(i))
        ENDIF
    ENDFOR
    THISFORM.LockScreen = .F.
ENDPROC

```

```

PROCEDURE moverlist1.cmdRemove.Click
    THISFORM.LockScreen = .T.
    FOR nCnt =1 TO THIS.Parent.lstSelected.ListCount
        IF THIS.Parent.lstSelected.Selected(nCnt)
            THIS.Parent.lstSelected.RemoveItem(nCnt)
        ENDIF
    ENDFOR
    THISFORM.LockScreen = .F.
ENDPROC

```

```

PROCEDURE moverlist1.cmdRemoveAll.Click
    THISFORM.LockScreen = .T.
    *FOR i = 1 to THIS.Parent.lstSelected.ListCount
    *
THIS.Parent.lstSource.AddItem(THIS.Parent.lstSelected.List(i))
    *ENDFOR
    THIS.Parent.lstSelected.Clear
    THISFORM.LockScreen = .F.
ENDPROC

```

```

ENDDDEFINE
*
*-- EndDefine: cargapd
*****

```

Apéndice C. Archivos IDC para consultas por Internet

Archivos IDC para consultas por Internet

1. Consulta general por Aduana e Importador (gene01.idc)

Datasource: QUEVFP

Template: general1.htx

SQLStatement:

```
+select año, mes, aduana, importador, sum(valor_cif) valor_cif, sum(total_imp  
total_imp, sum(total_exon) total_exon, sum(dai) dai, sum(sc) sc, sum(ley6946  
ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico, sum(num_bultos)  
+from consultaweb  
+where aduana in ('%aduana%')  
+and mes in ('%mes%')  
+and año in ('%año%')  
+order by in ('%variable%')  
+group by cod_año, cod_mes, cod_aduana, ced_import
```

2. Consulta General por Aduana y Agencia (gene02.idc)

Datasource: QUEVFP

Template: general2.htx

SQLStatement:

```
+select año, mes, aduana, agencia, sum(valor_cif) valor_cif, sum(total_imp  
total_imp, sum(total_exon) total_exon, sum(dai) dai, sum(sc) sc, sum(ley6946  
ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico, sum(num_bultos)  
+from consultaweb  
+where aduana in ('%aduana%')  
+and mes in ('%mes%')  
+and año in ('%año%')  
+order by valor_cif DESC  
+group by cod_año, cod_mes, cod_aduana, cod_agenci
```

3. Consulta General por aduana y Lugar de Descarga (gene03.idc)

Datasource: QUEVFP

Template: general3.htx

SQLStatement:

```
+select año, mes, aduana, lugar_desc, sum(valor_cif) valor_cif, sum(total_imp  
total_imp, sum(total_exon) total_exon, sum(dai) dai, sum(sc) sc, sum(ley6946  
ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico, sum(num_bultos)  
+from consultaweb  
+where aduana in ('%aduana%')  
+and mes in ('%mes%')  
+and año in ('%año%')  
+order by valor_cif DESC  
+group by cod_año, cod_mes, cod_aduana, cod_lugar_
```

4. Consulta de Lugar de Descarga por Aduana e Importador (lugar01.idc)

Datasource: QUEVFP

Template: lugar1.htx

SQLStatement:

```
+select año, mes, aduana, importador, lugar_desc, sum(valor_cif) valor_cif,
sum(total_imp_u) total_imp_u, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%año%')
+and importador in ('%impo%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, ced_import, cod_lugar_
```

5. Consulta de Lugar de Descarga por Aduana y Agencia (lugar02.idc)

Datasource: QUEVFP

Template: lugar2.htx

SQLStatement:

```
+select año, mes, aduana, agencia, lugar_desc, sum(valor_cif) valor_cif,
sum(total_imp_u) total_imp_u, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%año%')
+and agencia in ('%agencia%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, cod_agenci, cod_lugar_
```

6. Consulta de Aduana y Lugar de Descarga por Agencia (lugar03.idc)

Datasource: QUEVFP

Template: lugar2.htx

SQLStatement:

```
+select año, mes, aduana, agencia, lugar_desc, sum(valor_cif) valor_cif,
sum(total_imp_u) total_imp_u, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
```

```
+and año in ('%ano%')
+and lugar_desc in ('%desca%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, cod_agenci, cod_lugar_
```

7. Consulta de Mercancías por Aduana (merca00.idc)

Datasource: QUEVFP

Template: merca.htx

SQLStatement:

```
+select año, mes, aduana, cod_mercan, sum(valor_cif) valor_cif, sum(total_imp)
total_imp, sum(total_exon) total_exon, sum(dai) dai, sum(sc) sc, sum(ley6946)
ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico, sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%ano%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, cod_mercan
```

8. Consulta de Mercancías por aduana e Importador (merca01.idc)

Datasource: QUEVFP

Template: merca1.htx

SQLStatement:

```
+select año, mes, aduana, importador, cod_mercan, sum(valor_cif) valor_cif,
sum(total_imp) total_imp, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%ano%')
+and importador in ('%impo%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, ced_import, cod_mercan
```

9. Consulta de Mercancías por Aduana y Agencia (merca02.idc)

Datasource: QUEVFP

Template: merca2.htx

SQLStatement:

```
+select año, mes, aduana, agencia, cod_mercan, sum(valor_cif) valor_cif,
sum(total_imp) total_imp, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
```

```
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%ano%')
+and agencia in ('%agencia%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, cod_agenci, cod_mercan
```

10. Consulta de Mercancías por Aduana y Lugar de Descarga (merca03.idc)

Datasource: QUEVFP

Template: merca3.htx

SQLStatement:

```
+select año, mes, aduana, lugar_desc, cod_mercan, sum(valor_cif) valor_cif,
sum(total_imp_u) total_imp_u, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%ano%')
+and lugar_desc in ('%desca%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, cod_lugar_, cod_mercan
```

11. Consulta de Mercancías por Aduana y País de Origen (merca04.idc)

Datasource: QUEVFP

Template: merca4.htx

SQLStatement:

```
+select año, mes, aduana, pais_orige, cod_mercan, sum(valor_cif) valor_cif,
sum(total_imp_u) total_imp_u, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%ano%')
+and cod_mercan in ('%merca%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, cod_pais_o, cod_mercan
```

12. Consulta por Tipo de Revisión (revi.idc)

Datasource: QUEVFP

Template: rev_i.htx

SQLStatement:

```
+select año, mes, aduana, importador, semaforo, sum(valor_cif) valor_cif,
```

```
sum(total_imp_u) total_imp_u, sum(total_exon) total_exon, sum(dai) dai, sum(sc)
sc, sum(ley6946) ley6946, sum(iv) iv, sum(golfito) golfito, sum(mexico) mexico,
sum(num_bultos)
+from consultaweb
+where aduana in ('%aduana%')
+and mes in ('%mes%')
+and año in ('%año%')
+and semaforo in ('%revision%')
+and importador in ('%impo%')
+order by valor_cif DESC
+group by cod_año, cod_mes, cod_aduana, ced_import, cod_semafo
```

13. Mayores importadores mensuales (mayor.idc)

Datasource: QUEVFP

Template: mayor.htx

SQLStatement:

```
+select top 10 año, mes, aduana, importador, sum(valor_cif)
+ from consultaweb
+ order by valor_cif DESC
+ group by cod_año, cod_mes, cod_aduana, ced_import
```

Anexo A. Consultas sobre importaciones

Consultas sobre importaciones

No.	Grupo	Solicitan te	Información requerida	Objetivo
1	A	CE	Detalle de las importaciones de 4 empresas específicas en periodos distintos para cada una	Comportamiento de las operaciones
2	B	CE	Importadores del capítulo 33 durante un año específico	Estudio
3	C	CE	Importaciones de la partida arancelaria (Sac) 1515210000 en un periodo específico, en todas las aduanas. Requiere en el informe las siguientes variables: número de declaración de importación (en adelante denominada DA), el código de la agencia aduanera, el nombre del importador, cédula jurídica del importador, el origen de las mercancías, la procedencia de las mercancías, el total de impuestos pagados y el tipo de pago	Estudio
4	B	PL	Importadores (cédula jurídica y nombre) de una partida específica en un periodo dado	Complemento a estudio en proceso
5	A	Pl	Importaciones de varias empresas en un periodo determinado. (se aportan las cédulas jurídicas de ellas) No indican que variables requieren.	Complemento a estudio
6	A	Ce	Importaciones de una empresa específica en todas las aduanas, para tres meses específicos de un año específico. Indica las variables que requiere, entre ellas la aduana de ingreso y las declaraciones aduaneras asociadas, también requiere el desglose por cada tipo de impuestos (ad valorem, consumo, ventas, ley 6946)	Proporcionar información a una oficina extranjera.
7	A	Ce	Listado <i>cronológico</i> de importaciones de una empresa específica detallada por aduana de ingreso en un periodo específico. Indica las variables que requiere, entre ellas los impuestos desglosados. Pide que se <i>ordene por número de declaración!</i>	Estudio (Investigación)
8	A	Ce	Importaciones de las empresas específicas (más de una) en un periodo dado. Indica las variables, entre ellas la aduana de ingreso y los números de las declaraciones involucradas.	Respuesta a consulta de Tributación Directa.
9	A	Ce	Listado total de importaciones de un empresa específica en un periodo dado. Detalla las variables que se requieren (peso neto, valor CIF, e impuestos desglosados)	Estudio
10	A	Ce	Operaciones de una empresa en un periodo dado. Requiere las declaraciones , las mercancías, los impuestos el peso, el valor Cif y los impuestos	Atención a una denuncia.
11	B	Pl	Listado de declaraciones, los importadores, la aduana, la modalidad <i>anticipada</i> de ingreso y el resultado de aforo de cada declaración (sin revisión, revisión documental y revisión física)	Complemento a estudio.
12	A	Ce	Importaciones de una empresa en un periodo. Requiere la aduana y las declaraciones, además de los valores CIF, FOB, peso, impuestos desglosados y totales, exoneraciones y tipo de pago.	Seguimiento a estudio.
13	A	Ce	Importaciones de un grupo de empresas en todo el país par un periodo dado. Da las variables que se requieren.	Estudio
14	A	Ce	Importaciones de un grupo de empresas en todas las aduanas en un periodo dado. Indica variables.	Estudio
15	C	Ce	Importaciones de una empresa determinada en un año	Respuesta al OCI

			específico, para una partida arancelaria determinada a 10 dígitos (nivel :“apertura nacional”)	
16	D	Ce	Importaciones clasificada por empresa (cj) en un periodo.	Estudio de seguimiento.
17	A	Ce	Importaciones de una empresa en un periodo. Da variables.	Seguimiento.
18	A	Ce	Importaciones de una empresa determinada en un periodo. Da las variables requeridas	Seguimiento
19	A	Ce	Importaciones de una empresa para un periodo determinado. Da las variables requeridas.	Seguimiento
20	B	Ce	Importadores de la partida 0203 (nivel de partida: 4 dígitos) y las subpartidas (todo lo que este debajo del nivel partida, o sea los 6 dígitos restantes). Requiere las declaraciones aduaneras, el importador, la agencia de aduanas, la aduana de ingreso y la descripción de las mercancías	Atención denuncia.
21	A	Ce	Importaciones de un empresa en un periodo dado, detallada por aduana, desglosando los impuestos y otros datos.	Estudio.
22	D	Ce	Importaciones tramitadas por un grupo de agencias de aduana en todo el país en periodos distintos, para cada agencia. Indica variables de interés.	Estudio
23	A	Ce*	Ingresos de un importador en un periodo comprendido entre varios meses pero de dos años distintos.	Estudio
24	C	Ce	Detalle de las importaciones de una empresa determinada para las mercancías comprendidas en el nivel partida (primeros 4 dígitos) durante un año específico.	Estudio
25	C	Pl	Importaciones de una empresa durante <i>el último año</i> y las mercancías del mismo periodo pero del nivel capítulo 24 (primeros 2 dígitos)	Denuncia
26	A	Ce	Importaciones de un grupo de empresas por aduana. Dan las variables.	Estudio
27	A	Pl	Operaciones de un grupo de empresas durante un trimestre de un año específico. El resumen se utilizó para generar un indicador de valor contra peso, y impuestos contra peso.	Estudio de Ce. Fue primero resumido por partida.
28	E	Pl	Importaciones en una aduana específica y para un trimestre específico. Da variables.	Datos para gira
29	A	Pl	Operaciones de un importador en un periodo dado.	
30	A	Pl	Movimiento de mercancías de una empresa determinada para un periodo dado.	Investigación
31	D	Ce	Operaciones aduaneras de agencias específicas durante meses de años contiguos (dic, ene, feb). Entre otros datos pide <i>número de entero</i> , factura, fecha y el nombre del importador.	Auditorias
32	D	Pl	Volumen de operaciones de agencias de aduanas en términos de la cantidad de declaraciones aduaneras tramitadas durante un año específico.	Parámetros.
33	A	Al	Detalle de importaciones realizadas por una empresa determinada en un periodo de dos años específicos	Estudio legal
34	C	Ci	Mercancías de una partida arancelaria específica importadas durante un periodo específico por cualquier aduana. Agrupado por empresa, agencia, aduana y número de declaración.	Investigación.
35	F	Pl	Información sobre valor Cif de importación para un grupo de periodos y de un conjunto de empresas	Determinación de sujetos de estudio.
36	F	Pl	Igual al anterior para otras empresas	Idem
37	A	Pl	Movimiento de una empresa específica durante un año a partir de una fecha dada	Investigación
38	A	Pl	Movimientos de internación de mercancías de un grupo de empresas durante los últimos tres meses.	Investigación.

39	D	PI	Cantidad de declaraciones aduaneras tramitadas por cada agencia durante un periodo determinado en todas las aduanas.	Determinación de sujetos de estudio.
40	C	PI	Detalle de las mercancías (partida arancelaria a 10 dígitos) importadas durante un trimestre dado con el valor cif total de cada una de ellas y los países de origen y procedencia asociados a ellas. Así como las agencias que las tramitaron.	Determinación de sujetos de estudio.
41	D	PI	Cantidad de declaraciones aduaneras tramitadas por las agencias de aduana (dar código y nombre) de una aduana específica durante un periodo específico. Indicar el porcentaje que representa cada una de ellas del total.	Determinación de sujetos de estudio.

DISTRIBUCIÓN DE CONSULTAS POR GRUPO AF

GRUPO	CANTIDAD	%	DESCRIPCIÓN
A	22	53,7	Consultas completas (mínimo nivel de granularidad), para una o más empresas específicas, en distintos periodos.
B	4	9,8	Consultas completas por importadores de partidas arancelarias específicas, en un determinado periodo o año.
C	6	14,6	Consultas completas, para partidas arancelarias específicas dentro de distintos niveles de la jerarquía y en diferentes periodos.
D	6	14,6	Consultas completas por agencias de partidas arancelarias específicas, en un determinado periodo.
E	1	2,4	Consultas completas para aduanas específicas, en un periodo determinado.
F	2	4,9	Agrupamiento (resumen) por variables de medida en un periodo dado y para empresas específicas.
TOTALES	41	100	

DISTRIBUCIÓN DE CONSULTAS POR SOLICITANTE DENTRO DE LA DCF

DTB	Control Externo	Planificación	Control Interno	Legal
Cantidad	23	16	1	1
Porcentaje	57	39	2	2

Anexo B. Entrevista a usuarios de la información

Entrevista a usuarios de la información

1. ¿Como la información que usted ha solicitado ha contribuido a tomar decisiones?
2. Cual fue el proceso al que sometieron la información proporcionada para que fuese útil?
3. Que resultados obtuvieron del análisis de dicha información?
4. ¿Cómo analizaron los datos que se les proporcionó?
5. ¿Los datos suministrados estaban presentados en una estructura que facilitaba su utilización o análisis?
6. ¿Fue necesario transformarlos o complementarlos para poder utilizarlos?
7. ¿Fue necesario el desarrollo de una estructura especial para acomodar los datos o utilizaron herramientas convencionales de manipulación y ordenamiento como hojas electrónicas.?
8. ¿La información suministrada fue suficiente para los propósitos para los que la solicitaron?
9. ¿Se requirió de otras fuentes para completar la información?
10. ¿ Que otra información fue necesario buscar y donde?
11. ¿ En que formatos se le presentó la información?
12. ¿ La información solicitada ha sido oportuna para la empresa o institución?
13. ¿Qué sugeriría a la administración para mejorar la calidad y variedad de información que ponga a disposición?

Anexo C. Lista de campos de base de datos DADEC

Cant_Aforo
Ced_Importador
Cif_Aforo
CMonto1
Cod_Aduana
Cod_Agencia
Cod_Estadistico
Cod_Impuesto
Cod_Partida_Aforo
cod_Ubicacion
Cond_Mercancia
Fech_Hora_Cancel
Fob_Aforo
Mon_Autoriz_Nota
Mon_Cascada
Mon_Cif_Aforo
Mon_Exen_Nota
Mon_Fob_Aforo
Mon_Liquid
mon_liquid_vieja
Monto1
Monto_Ex1
Monto2
Monto_Ex2
Monto3
Monto_Ex3
Monto4
Monto_Ex4
Monto5
Monto_Ex5
Monto6
Monto_Ex6
Nom_Importador
Num_Linea
Num_Poliza
Pais_Origen_Aforo
Pais_Procedencia
Peso_Aforo
TCant_Polizas
TCant_Lineas
Tiene_Lin
Tipo_Poliza
Tipo_revision
Tipo_Ubicacion
Tot_Liquid_Lin
Total
Total_Ex
TTotal_Cif
TTotal_Liq
TTotal_Peso
XMonto
XMonto1

**Anexo D. Carta de aprobación de la ejecución del proyecto por
parte de la DCF**