



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
CAMPUS "ACATLAN"

EL ODS COMO BASE DE UN DATA WAREHOUSE.
CASO PRÁCTICO ADUANAS.

SEMINARIO TALLER EXTRACURRICULAR

QUE PARA OBTENER EL TITULO DE:

LICENCIADO EN MATEMATICAS APLICADAS
Y COMPUTACION

PRESENTA:

OREA GUZMÁN ARTURO

ASESOR:
ING. RUBÉN ROMERO RUIZ

FEBRERO 2004





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

CONTENIDO

Introducción.....	5
Capítulo Uno. Marco Teórico.....	8
1.1 La necesidad de almacenes de datos.	8
1.1.1 Origen y naturaleza de los datos.....	9
1.1.2 La necesidad de obtener Información del Negocio.....	10
1.1.3 Convertir datos en Información.	11
1.1.4 Sistemas Transaccionales.....	11
1.2 ¿Qué es un Data Warehouse?	13
1.2.1 Componentes de un Data Warehouse	14
1.2.2 Características de un Data Warehouse	16
1.2.3 Metadatos	20
1.2.4 El Valor de retorno sobre inversión de un Data Warehouse.....	21
1.3 Diferencias entre una base de datos OLTP y un DW.....	23
Capítulo Dos. Fundamento Estructural.....	26
2.1 Planeación:.....	26
2.1.1 Consideraciones iniciales	26
2.1.2 Equipo de trabajo.....	30
2.2 Recolección de Requerimientos	31
2.2.1 Requerimientos Existentes.....	31
2.2.2 Requerimientos Futuros	33
2.3 Etapa de Análisis.....	34
2.3.1 Identificación de Fuentes de Datos	34
2.3.2 Obtener la granularidad apropiada.....	36
2.3.3 Implicaciones del diseño físico.....	39
Capítulo Tres. Diseño y Construcción.....	42
3.1 Diseño del Modelo Entidad Relación del Repositorio (ODS)	42
3.1.1 Modelo ODS.	42

3.1.2 Diseño del modelo Entidad Relación.....	45
3.2 Generación de bases de datos a partir de los modelos entidad relación. 48	
3.3 Proceso de Extracción Transformación y Carga.....	49
3.4 Extracción de datos.	50
3.4.1 Especificaciones de extracción.....	52
3.5 Transformación de datos.....	52
3.5.1 Transformación e integración.	53
3.6 Mapeo.....	56
3.7 Carga de datos.....	56
3.7.1 Mejoras en el proceso de carga.....	57
3.7.2 Carga y recuperación.....	58
Capítulo Cuatro. Validación del modelo.....	61
4.1 Validación de datos.	61
4.2 Pruebas de consistencia.	61
Capítulo Cinco. Automatización.....	67
5.1 Afinación de procesos.....	67
5.2 Automatización.....	75
Capítulo Seis. Explotación del ODS.....	82
6.1 Herramientas para la explotación de los datos del ODS.	82
6.1.1 Niveles de información.....	89
6.1.2 Usuarios dentro de cada nivel de información.	91
6.2 Herramientas de Reporteo:.....	98
6.2.1 Reportes.....	98
6.3 Herramientas OLAP: Molap, rolap, holap.....	100
6.3.1 Las reglas del Dr. E. F. Codd.....	103
6.3.2 Modelos de almacenamiento.....	111
6.4 Minería de datos.....	113

6.4.1 Relación de la minería de datos con otras disciplinas	119
6.4.2 Ventajas de la minería de datos.	120
6.4.3 Áreas de aplicación:	121
6.4.4. Modelos predictivos y modelos descriptivos	123
6.4.5 Tipos de técnicas:.....	127
6.4.6 Visualización.....	129
Capítulo Siete. Aduanas.....	132
7.1 Antecedentes	132
7.1.1 Reseña histórica	133
7.1.2 La Administración General de Aduanas.....	142
7.2 Necesidad de información.....	150
7.3 Descripción del Proyecto:	151
7.3.1 Requerimientos.....	152
7.3.2 Beneficios	152
7.3.4 Revisiones y ajustes.....	157
7.3.5 Infraestructura.....	159
7.4 Áreas Usuarías.....	160
7.5 Modelo de base de datos resultante para el Data Warehouse	162
7.6 Esquema de ETL del Data Warehouse	166
7.6.1 Proceso implementado en las aduanas	168
7.6.2 Factores críticos de éxito.....	168
7.7 Experiencia.....	169
7.8 Proyección.....	171
Conclusiones.....	173
Glosario.....	174
Glosario sobre Aduanas.....	176
Bibliografía	185

Introducción

El hombre piensa en imágenes mentales no en palabras, asociando a cada imagen mental el cúmulo de experiencias adquiridas a lo largo del ejercicio de sus funciones dentro de una organización, es a través de estas asociaciones que se efectúa la toma de decisiones y es con el propósito de llevar a cabo una toma de decisiones mejor sustentada, que surge de manera indefectible la necesidad de crear un Data Warehouse dentro del Servicio de Administración Tributaria, finalmente el propósito de esta obra es mostrar desde la perspectiva del autor los escenarios y las exigencias que implicó el establecimiento de las bases para implementar un Data Warehouse en una organización que se mueve hacia la digitalización, muy singular, en relación a que se trata de la autoridad fiscal en México, un país que navega entre los novedosos y sofisticados conceptos de procesamiento de datos e Inteligencia de Negocio y las limitaciones propias de una Nación en vías de desarrollo que lucha con denuedo por alcanzar los niveles que en materia de legislación fiscal y de comercio exterior mantienen los Estados altamente desarrollados de los cuales es socio comercial.

Lo que usted está a punto de leer es el producto de varios años de búsqueda, investigaciones, aprendizaje y análisis de los principios que rigen el novedoso mundo de la obtención y manejo de información dentro de las organizaciones, así como de su implantación dentro de una entidad gubernamental, donde cabe resaltar que para desarrollar una estrategia de acercamiento a las necesidades que surgen, derivadas de una economía que empuja a tomar decisiones más rápidas y efectivas, a través del análisis de toda su información desde un nivel general hasta un nivel detallado, mejorando sustancialmente la consecución de los objetivos planteados en la visión de la Administración tributaria federal se necesita primordialmente

ampliar los matices con que los responsables de definir las estrategias organizacionales perciben el desempeño de las instituciones a su cargo, sin dejar atrás que la información proviene de fuentes de datos cada día más complejas lo que representa una imagen parcial de la realidad.

El siguiente paso se hace evidente al notar el valor que tiene el conocimiento para las instituciones públicas, el cual proporciona una sensible mejoría en el cumplimiento de los fines inherentes a los organismos del Estado. A este respecto, la concepción de Data Warehouse es vista como un elemento fundamental del concepto de Inteligencia de Negocio.

Hoy en día, la inteligencia de negocio se considera una nueva disciplina que pretende convertir el gran volumen de datos que normalmente tiene almacenada una organización en información útil para la toma de decisiones. Este es un proceso apoyado por herramientas modernas de tecnología de información, que permiten explorar, conocer y descubrir información institucional para transformarla en conocimiento para la organización.

CAPÍTULO I

Capítulo Uno. Marco Teórico

En este Capítulo se definen los conceptos básicos para comprender los temas venideros, se presentan las diferencias entre las bases de datos operacionales y las de soporte a la toma de decisiones, se muestran las características generales de los sistemas de Data Warehouse; y finalmente se describen algunos de los impactos y beneficios esperados al implementar un Data Warehouse.

1.1 La necesidad de almacenes de datos.

El crecimiento explosivo de las bases de datos en las organizaciones de gobierno, empresas y científicas entre otras, ha rebasado por mucho nuestra habilidad para interpretar y digerir los datos contenidas en ellas, por ello surge la necesidad de contar con metodologías, técnicas y herramientas para analizar la información. El desarrollo e implementación de un Data Warehouse es un avance en este sentido, ya que proporciona un apoyo en la toma de decisiones, basándose en información integrada y global de la organización, facilita la aplicación de técnicas estadísticas de análisis y modelos para encontrar relaciones ocultas entre los datos obteniendo un valor agregado para la organización, ayuda a efectuar análisis de tendencias basado en datos históricos, simplifica el acceso a la información, apoya al cumplimiento de los objetivos de las áreas sustantivas, entre otras aplicaciones más.

Actualmente el desconocimiento y la información incompleta es gran amenaza para una organización, ya que provoca que se tomen decisiones con poca fundamentación, creyendo equivocadamente que se conoce el contexto.

El Data Warehouse al reunir los datos generados por los componentes de la organización, pretende ser la única fuente de información integrada consistente y oportuna de consulta, que apoyado en tecnología de punta, colabore en la optimización de procesos de la organización, hacer que los conceptos inherentes a ella tengan el mismo significado para todos, asimismo es de gran ayuda en el monitoreo de las operaciones. Finalmente contribuye a disminuir supuestos erróneos y a tomar riesgos innecesarios derivados de información poco confiable o desactualizada, dejando vulnerable a la organización.

1.1.1 Origen y naturaleza de los datos.

Las organizaciones diariamente capturan grandes volúmenes de datos, que generalmente registran las operaciones realizadas a través de registros que a su vez representan el estado actual del negocio.

Vistos superficialmente, estos datos brindan información de lo que sucede dentro de la organización. Sin embargo, pueden brindarnos más elementos de análisis que eso, por ejemplo, los datos recolectados pueden ayudar a entender como opera la organización y apoyar la planificación de sus operaciones en el futuro. Otra manera de referirse a los datos es como *datos operacionales*, *datos transaccionales*, *datos crudos*.

Ejemplo de datos operacionales:

- Una cadena de tiendas de música captura información de las ventas por cada compra, devolución o cambio en sus sucursales, cada uno de estos movimientos es un dato operacional y será registrado en una base de datos.
- Otro dato operacional es el registro correspondiente al registro de un alumno en una materia para un determinado ciclo escolar.
- Algunos sumarios sencillos son obtenidos a partir de los datos operacionales por ejemplo se hace una operación de suma para los

datos comprendidos en un rango de tiempo. Ejemplo: Ganancia durante el mes de enero de 2002 = \$200,000.

1.1.2 La necesidad de obtener Información del Negocio

La Información debería ayudar a conocer el comportamiento de la organización ya sea en un punto de su historia, en un periodo o en un conjunto de ellos, así como aislar los aspectos que la afectan directa o indirectamente, apoyando a la dirección a administrarla de mejor manera.

Parte del proceso para obtener información del negocio a partir de datos operacionales comprende:

- Examinar los datos en diferentes contextos y desde varios puntos de vista.
- Determinar como los hechos se relacionan con otros datos.
- Entender como la información refleja las metas y objetivos de la empresa.

Ejemplo:

Un negocio de música quiere conocer mejor el estado de las ventas en sus sucursales. Se hará preguntas como:

- ¿Qué música está vendiendo la sucursal?
- ¿Cuáles son los 10 productos más vendidos?
- ¿Qué música no se vende?
- ¿Qué efecto tiene la promoción de productos?

La necesidad de obtener información valiosa del negocio es grande y las herramientas y sistemas tradicionales no están diseñados para hacer frente a estos cuestionamientos.

1.1.3 Convertir datos en Información.

Cuando una empresa reconoce la importancia de obtener Información del negocio, se hacen numerosas y frecuentes las solicitudes de datos e información, generalmente se vuelve una tarea compleja y altamente demandante satisfacerlas, debido en primer lugar por la necesidad de consultar grandes y complejas bases de datos y en segundo lugar por intentar analizar esta información y distribuir los resultados.

Es por eso que se debe implementar algún tipo de Sistema de Soporte de Decisiones (DSS) basado en un Data Warehouse dedicado a brindar datos e información que pueda ser empleada para analizar a la organización sin interrumpir la operación de sus sistemas ni navegar a través de miles o millones de registros que por si mismos no ofrecen elementos analíticos valiosos.

Implementar un sistema de esta naturaleza requiere una inversión considerable tanto en recursos humanos como materiales, sin embargo el valor de retorno sobre inversión (ROI¹) puede ser muy alto dependiendo de las medidas que sean tomadas a partir de las explotaciones y descubrimientos que se hacen de la organización y al disminuir riesgos ocasionados por la falta de ésta.

1.1.4 Sistemas Transaccionales

Los sistemas transaccionales o sistemas operacionales son conocidos por sus siglas en inglés como Sistemas OLTP (OnLine Transaction Processing), y son los que capturan y almacenan las transacciones de una organización y generalmente son la principal fuente de datos de un Data Warehouse.

¹ Ver Glosario.

Una organización puede tener uno o más sistemas OLTP que controlen sus procesos. Pueden estar en equipos separados, en diferentes redes e incluso pueden ser externos a la empresa.

Características de los sistemas OLTP:

Procesamiento en tiempo real.

Estos sistemas trabajan al tiempo que se efectúan las transacciones, continuamente cambian para representar el estado actual de la empresa. Hay altas, bajas y cambios de manera inmediata conforme el sistema procesa nuevas transacciones.

Están estructurados para recibir datos y editarlos. En otras palabras encontraremos bases de datos fuertemente normalizadas, lo cual es ideal para procesamiento de transacciones, pero implica una gran cantidad de juntas (joins) y esto a su vez gran complejidad al diseñar consultas.

La consulta y el análisis no son prioritarios, obtener reportes directamente de un OLTP nos brinda los datos más actuales, sin embargo tiene un efecto negativo en el desempeño del sistema y puede llevarnos a obtener reportes inconsistentes debido a la volatilidad propia de un sistema OLTP.

Ejemplos:

- Sistemas de cajeros automáticos y en general sistemas bancarios. El tiempo de respuesta inmediato al igual que el registro de cada transacción son indispensables para evitar fraudes. Tratar de obtener información directamente del OLTP seguramente entorpecería de alguna manera la operación del mismo.
- Ventas en línea, como el que ofrece Ticketmaster. La cantidad y naturaleza de eventos que administra obliga a que el sistema opere sobre disponibilidad, esencialmente dos personas no pueden comprar el mismo lugar. De manera que mientras se efectúan transacciones

ciertos lugares quedan bloqueados temporalmente para que otros usuarios en línea o vía telefónica no los ocupen, ésta y otras limitaciones impedirían emplear el sistema OLTP como soporte en la toma de decisiones.

1.2 ¿Qué es un Data Warehouse?

Se tiene un gran interés por conocer y analizar los datos de los sistemas OLTP sin embargo consultarlos directamente puede ocasionar muchos problemas, debido en primer lugar a que los sistemas operacionales están optimizados para efectuar inserts y updates asociados con las operaciones del negocio, las consultas efectuadas sobre ellos merman su desempeño. Además la volatilidad derivada de su propósito produce reportes inconsistentes.

Dadas las limitaciones para obtener información directamente de los sistemas OLTP, la mejor forma de cubrir las necesidades de analizar a la organización es recurrir a un Data Warehouse. La transferencia y adecuación de datos de los sistemas operacionales a los usuarios es la función primaria de un sistema Data Warehouse. La figura 1 esquematiza esta transferencia.

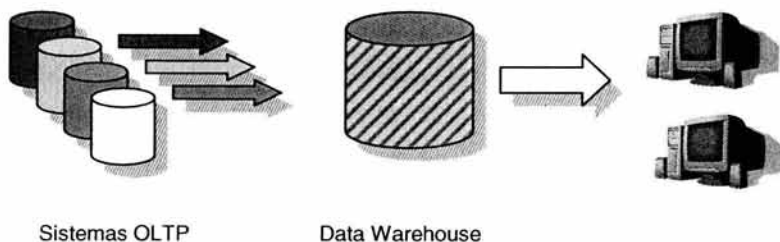


Figura 1. Obtención y distribución de datos.

Un Data Warehouse es un repositorio central que contiene la información más valiosa de la organización. Los datos que aquí se almacenan han pasado por un proceso de calidad que asegura su consistencia. Además el repositorio está construido de tal manera que el acceso sea lo más rápido posible.

Su construcción se va haciendo por etapas que normalmente corresponden a las principales áreas operativas de la organización.

El objetivo del Data Warehouse o Almacén de datos es reunir o consolidar las bases de datos diferentes, que se mantienen en los diferentes departamentos o áreas funcionales de la organización como subsistemas de información independientes, en una gran base de datos, recogiendo datos dispares y, en muchas ocasiones infrutilizados, procedentes de fuentes internas repartidas por toda la organización².

1.2.1 Componentes de un Data Warehouse

Un Data Warehouse está compuesto de la siguiente manera:

El diseño del repositorio central de información subdividido en áreas de negocios y temas específicos.

Fuentes de datos: Este componente es el que normalmente está presente originariamente en las organizaciones, y a partir del cual se realiza la captura de datos que se contemplará en el Data Warehouse. Estas fuentes de datos pueden ser sistemas operacionales, sistemas operacionales departamentales y fuentes externas.

Extracción y Transformación: Este es el componente responsable de que la información pueda moverse con las transformaciones necesarias, desde

² Inducción a la administración de Transformaciones y Consistencia en Bases de Datos, AGTI, Versión 1.0. SAT. 2003.

las fuentes de datos, al Data Warehouse. En este sentido hay que decir que por Data Warehouse puede entenderse tanto el sistema completo como únicamente las bases de datos en las que se almacenan tanto la información extraída de los sistemas anteriores como los metadatos.

Servidor de datos: También podría denominarse componente de gestión. Los servicios que debe ofrecer incluyen un servicio de mantenimiento de datos y un servicio de distribución para exportar datos del Data Warehouse a servidores de bases de datos descentralizadas y a otros sistemas de soporte de decisiones de usuario. El componente de gestión también ofrece servicios de seguridad (archivo, respaldo, recuperación) y monitorización. Generalmente estos servicios utilizan los medios suministrados por el software del sistema operativo y de bases de datos disponible. El componente de sistema manejador de bases de datos consiste en el software que se utilice para mantener y extraer datos.

Metadata: La metadata o metadatos, es la base fundamental para establecer la completa integración de los datos de la organización. Los metadatos son útiles para analizar como afectarán los cambios a la aplicación, sus módulos y otras aplicaciones con las cuales interactúa. Los metadatos proveen un modelo integrado de datos y una descripción de cómo la base de datos se relaciona a este modelo, el cual puede ser enlazado con las distintas aplicaciones que usan estos datos.

Modelo de datos: Su propósito es organizar los datos para mantener los objetivos de la organización y proveer un buen comportamiento de la operación de la base de datos.

Reportes: Un conjunto de reportes y gráficos que explotan la información de los modelos y la hacen disponible en forma inmediata al usuario. Proyecta

servicios de asesoría con el propósito de atender las necesidades particulares de los usuarios.

1.2.2 Características de un Data Warehouse

De acuerdo a Bill Inmon las características de un Data Warehouse son:

Integrado:

Los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse en varios niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

La integración se realiza de varias maneras, en estándares de nombres, en las unidades de medida, en los tipos de datos, etc. La falta de integración de los ambientes OLTP se debe a que al desarrollar las aplicaciones se tomaron decisiones que condujeron a diseños personalizados, sin apegarse a ningún estándar. En consecuencia algunas aplicaciones almacenaron el sexo de la persona como M y F, otras como 1 y 0, otros X e Y o como "masc" y fem, realmente no es tan importante que formato llegue al Data Warehouse, lo importante es que llegue de manera única, las unidades de medida por su parte están en metros, centímetros, pulgadas, yardas, pies, y al igual que en el caso del genero se deben transformar los datos haciendo las conversiones apropiadas para que se cargue una sola medida estándar.

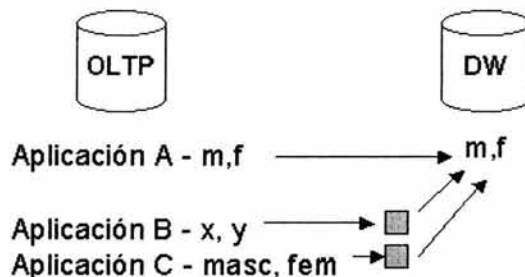


Figura 2. Integración de datos.

Temático:

Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. La información se tiene estructurada con base a los aspectos que son de interés para la organización, en contraste con el ambiente OLTP que es diseñado conforme a los requerimientos de las aplicaciones. El Data Warehouse además está organizado por sujetos tales como contribuyente, mercancías, proveedor, etc.

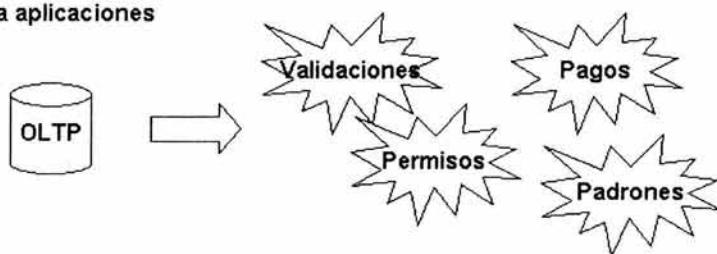
Orientado a aplicaciones

Figura 3. Un sistema OLTP está orientado a aplicaciones.

Orientado a temas

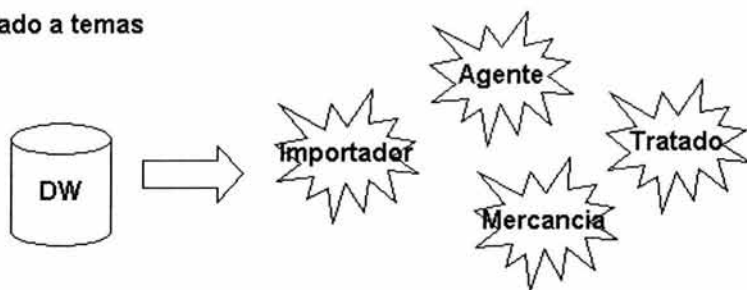


Figura 4. Un Data Warehouse está orientado a temas o sujetos.

Histórico:

El tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

Aspecto	OLTP	Data Warehouse
Contenido de datos	Valor actual de los datos.	Datos consolidados.
Vigencia de datos	Corto horizonte de tiempo.	Horizonte histórico.
Identificador de temporalidad	Un elemento de información puede o no contar con la variable de tiempo.	Se cuenta con la variable de tiempo.

Tabla 1. Comparativa de vigencia de datos entre un sistema OLTP y un Data Warehouse.

No volátil:

La información de un Data Warehouse tiene como propósito principal ser consultada. La información se considera permanente o estática, la actualización del Data Warehouse contempla en la mayoría de casos únicamente la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre los que ya existían. La importancia de esta característica radica en que para hacer análisis se requiere de información estable. Los procesos de actualización en la base de datos OLTP impiden esta estabilidad, el Data Warehouse tiene prácticamente solo dos tipos de operaciones, la carga de datos y la consulta de éstos, la actualización de datos es inusual.



Figura 5. Operaciones típicas sobre un sistema OLTP y un Data Warehouse.

Optimizado para Consulta:

En un Data Warehouse los datos son almacenados en estructuras optimizadas para distribución y consulta.

El propósito de un Data Warehouse es exponer información del negocio a los usuarios. Los usuarios analizan la información y se apoyan en ella para tomar decisiones.

Algunos de los beneficios esperados con la implementación de un Data Warehouse son:

1. La toma de decisiones de los altos mandos de la organización permite disponer de información de manera oportuna.

2. Contar con información integrada que proporcione un panorama global de la organización.
3. Facilitar la generación de información estadística, análisis de tendencias, encontrar relaciones ocultas entre variables, lo que permite obtener un alto valor agregado de la información.
4. Ayudar a comprender errores del pasado.
5. Permitir identificar comportamientos atípicos, como fraudes, errores en la captura de datos en los sistemas operacionales u otros defectos en las aplicaciones pueden quedar expuestos, etc.

Un sistema Data Warehouse es la base para la organización y análisis de los grandes volúmenes de información que genera la organización. Sin embargo la definición del modelo y su implementación requiere de tecnología y personal especializado que atienda y soporte las necesidades particulares de cada organización.

1.2.3 Metadatos

La manera de mantener el control sobre los datos que contiene el Data Warehouse y de dónde se obtuvieron es mediante el uso de metadatos, que conforman un componente muy importante. Al diferencia de la información que contiene el Data Warehouse los metadatos no provienen de sistemas OLTP, son producto de documentar transformaciones, datos origen, datos destino, estructura de ellos, nombres en base de datos, su significado dentro de la organización.

Los metadatos funcionan como un directorio que indica los contenidos del Data Warehouse e indican el mapeo de datos, especificando las transformaciones efectuadas en el camino del ambiente transaccional hacia el Data Warehouse.

Entidad	Tabla	Nombre de negocio		Campos	Dato
R 511 V N	Observaciones del pedido	Synop11			
		Patente aduanal	FK	paadtue1	CHAR(4)
		Número de pedimento	FK	npuemde1	CHAR(7)
		Clave de aduana de despacho	FK	cadldea1	SMALLINT
		Secuencia de la observación		soebcsu1	SMALLINT
		Observaciones		observa1	CHAR(120)
		Fecha de selección automatizada		fsapcee1	DATE

Tabla 2. Ejemplo de metadatos

La tabla muestra una manera simple de documentar en una hoja de calculo el contenido de una tabla y la información que representa en relación a la organización.

1.2.4 El Valor de retorno sobre inversión de un Data Warehouse.

En terminos muy generales el valor de retorno sobre inversión permite saber si un proyecto genera ganancias o pérdidas y en que tiempo.

En muchas empresas es un parámetro que puede determinar la aprobación o rechazo de un proyecto, sin embargo el ROI en un Data Warehouse puede ser difícil de estimar.

Los sistemas de información pueden clasificarse en tres grandes grupos:

- Operacionales
- Informativos
- Estratégicos

Para efectos de calcular el valor de retorno sobre inversión de un Data Warehouse se puede clasificar como informativo o estratégico, dependiendo del uso que se le dé.

Como sistema informativo, el DW contiene toda la información de la organización. Al contemplar el ROI consideramos que los costos del DW deben ser menores que los beneficios financieros. La estimación de los

costos es simplemente sumar los gastos asociados al proyecto; Hardware, Software, Personal dedicado, Capacitación, Consultoría, Etc. Los beneficios sin embargo pueden ser complicados de estimar, y están asociados con los objetivos planteados al iniciar el proyecto de DW y con la explotación lograda de la información.

Metas como la satisfacción absoluta del cliente, resultan muy complicadas de determinar, por la cantidad de elementos subjetivos que involucra, si se desea recurrir a esta dimensión, esta variable se debe convertir en algo cuantificable de lo contrario solo se especulará al respecto. Por otra parte metas como mejorar la eficiencia del departamento de control de inventario, reducir productos rechazados debido a errores al procesar pedidos, disminuir gastos de publicidad al identificar clientes potenciales, permitirá cuantificar fácilmente los beneficios financieros.

Es importante mencionar que no siempre es imprescindible cuantificar totalmente el ROI, puede ser suficiente fijar metas generales, por ejemplo al implementar la primera etapa de su Data Warehouse, la empresa Bimbo simplemente estableció como metas mejorar el canal de distribución y comenzar la búsqueda de áreas de oportunidad.

Como sistema estratégico el proyecto de DW es puesto en práctica para obtener ventajas competitivas, expandir o mantener participación en el mercado y encontrar áreas de oportunidad, en ese sentido los beneficios deberían ser lo suficientemente evidentes como para que una empresa apruebe el proyecto inmediatamente.

1.3 Diferencias entre una base de datos OLTP y un DW.

Ahora que se ha mencionado la misión de un Data Warehouse y la de un Sistema OLTP, conviene contrastar las principales características de uno y otro. En la Tabla 4 se encuentra un sumario de estos puntos.

Característica	SISTEMA OLTP	DATA WAREHOUSE
Función	Soporta las transacciones de la organización.	Soporta el análisis, la exploración y la decisión estratégica
Operación predominante	Altas, bajas y cambios.	Predomina la consulta.
Granularidad	Detalle	Detalle a agregado.
Integración de información.	Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Tiempo	El estado actual del negocio.	Histórico, Actual, Proyectado.
Respuesta	Trabaja en tiempo real. Respuesta de la transacción instantánea. Devuelve datos puntuales.	Procesamiento masivo de datos. Tiempo de respuesta es de unos cuantos segundos hasta varias horas.
Estructura de almacenamiento.	Estructura relacional. Altos niveles de normalización.	Visión multidimensional. Modelos estrella, copo de nieve, tablas con cierto nivel de desnormalización.
Perfil de usuario	Usuarios de niveles operativos a medios mandos.	Usuarios de niveles ejecutivos, gerenciales.
Alcance de la información.	Explotación de la información limitada al segmento del negocio que soporta el sistema OLTP	Explotación de toda la información relacionada con el negocio.

Tabla 3. Comparativa entre un sistema OLTP y un Data Warehouse

Los sistemas de información actuales generan grandes volúmenes de datos, ya que almacenan todas las transacciones efectuadas, dichos sistemas son conocidos por sus siglas en inglés como sistemas OLTP (OnLine Transaction Processing), son optimizados para responder a las necesidades operativas

de la organización, es decir Altas, Bajas y Cambios. En consecuencia su diseño lógico y físico obedece a la exigencia de un alto nivel de performance, lo que implica bases de datos fuertemente normalizadas, sin embargo, esto provoca que la exploración de información se torne compleja. Por tanto cuando se pretende obtener información estratégica de un sistema OLTP surgen problemas como la interrupción o el alza en el tiempo de respuesta de la operación de los sistemas, del mismo modo la normalización provoca que las consultas involucren varios joins que consumen mucho tiempo hombre en su diseño y tiempo maquina en su ejecución. Esto nos sugiere que no es conveniente obtener información estratégica de los sistemas operacionales, en lugar de ello se debería recurrir a un modelo que no interrumpa la operación, que sea fácil de explotar y que nos ofrezca información rica en contenido, Un Data Warehouse nos ofrece estas características ya que desde un principio se diseña para cumplir como objetivo principal el responder preguntas, ayudar en el análisis de tendencias, preparar fácilmente comparativas, facilitar diversos análisis exploratorios y desde luego a no obstruir los procesos operativos.

El contar con un Data Warehouse no garantiza tener la información de la manera que se requiere para ser usada, ya que en principio solo es un depósito de datos, pero ofrece mejores posibilidades de tomar decisiones más atinadas ya que aporta elementos para fundamentar una decisión.

Capítulo II

Capítulo Dos. Fundamento Estructural

Los elementos en este Capítulo muestran puntos importantes para iniciar el desarrollo de un Data Warehouse con el propósito de disminuir los retrasos ocasionados por imprevistos causados por desconocer las fuentes de datos y su operación, por no contar con el personal adecuado para esta tarea, por no determinar el nivel de detalle apropiado a las necesidades de la organización, etc. Asimismo se describe parcialmente el proceso de obtener los requerimientos del usuario y como tratar de anticiparse a los venideros.

2.1 Planeación:

La planeación consiste en tener en cuenta lo que debe ser diseñado y que puede ser parte de la infraestructura tecnológica y del modelo de datos. En ese sentido el presente trabajo no pretende exponer de manera detallada las actividades necesarias en el diseño e implementación del Data Warehouse, en lugar de ello se mencionan los puntos más importantes de acuerdo mi experiencia en la implementación del Data Warehouse de Aduanas complementado con material del SAT³ y material bibliográfico estudiado.

2.1.1 Consideraciones iniciales

El desarrollo de un Data Warehouse puede ser complejo, los motivos para que falle generalmente son:

- Poca planeación.
- Falta de experiencia del personal involucrado.
- Administración deficiente.

³ Servicio de Administración Tributaria, órgano desconcentrado de la Secretaría de Hacienda y Crédito Público.

-
- Presupuesto insuficiente debido a una estimación incorrecta de costos.
 - Análisis limitados que no contemplan las necesidades de la organización.
 - Errores en el manejo de los datos origen y en los procesos para integrarlos al Data Warehouse que provocan información inconsistente.

Las consideraciones de diseño y desarrollo que se deben tomar al iniciar el proyecto de Data Warehouse son:

- **El costo del almacenamiento.** Dependiendo del tamaño de la organización puede ser suficiente contar con un servidor pequeño o requerir equipos muy costosos para poder atender las consultas en tiempo y oportunidad. Para mantener un nivel apropiado de eficiencia cuando hay un número elevado de consultas concurrentes.
- **La tasa de crecimiento.** Es importante evaluar el crecimiento del Data Warehouse, si es muy elevado se necesita recurrir a sistemas escalables.
- **Los programas de extracción y transformación que van a ser requeridos.** Se debe evaluar si se van a desarrollar en la organización o se va a recurrir a desarrollos de terceros, la decisión depende fundamentalmente del tiempo, el presupuesto y la capacidad del equipo de trabajo con que se cuente. Estos desarrollos generalmente se hacen en los lenguajes de desarrollo del manejador de base de datos adquirido, en el caso de Aduanas con 4GL de Informix, Lenguaje "C", SQL-EC, SQL, SH, etc.
- **Aplicaciones que explotarán el Data Warehouse** se puede optar por desarrollos propios o de terceros, sin embargo los desarrollos internos requieren demasiado tiempo, en contraste con un desarrollo de terceros que permitirá obtener resultados más rápidamente.

Se deben identificar todas las tareas involucradas mediante un plan de trabajo, muchas requieren previsión y coordinación entre un gran número de personas y organizaciones. Se requiere que el analista y el desarrollador sistemáticamente revisen la lista de pendientes para poder concluir adecuadamente el plan de trabajo.

Algunos de estos puntos son:

- Identificar el problema central por el cual se iniciará el proyecto de Data Warehouse. Las razones más importantes son: No contar con información integrada, histórica, obtener resultados limitados provenientes de reportes de aplicaciones de propósito operacional, no estar en condiciones de realizar o aplicar técnicas avanzadas para análisis de datos.
- Contar con un equipo de personas dedicadas de tiempo completo a esta actividad y con conocimientos apropiados.
- Identificar las necesidades de los usuarios y prever las futuras.
- Identificar las fuentes de datos que alimentarán el Data Warehouse.
- Estimar los volúmenes de datos que se van a generar y el impacto en la infraestructura tecnológica elegida.

Si se comienza a trabajar sin tomar en cuenta las consideraciones iniciales se ocasionarán retrasos, problemas y productos de poca calidad. Algunos ejemplos son:

- Elaboración de diseños de base de datos que no están apegados a las necesidades de los usuarios, debido a que se inició el diseño físico de la base de datos sin conocer el nivel de detalle requerido de los datos. Y aunque es común que éste cambie a lo largo del ciclo de vida del Data Warehouse es deseable que los cambios no sean tan frecuentes, al menos se espera que una primera entrega satisfaga algunas necesidades básicas de los usuarios.

- Tiempo de diseño extendido indefinidamente. Se recomienda dividir las etapas o iteraciones del Data Warehouse, es preferible iniciar un modelo sencillo pero robusto que se entregue en tiempo y forma a uno muy complejo que se entrega tardíamente.
- Definir prioridades. Es deseable tener claro el orden en que se deben alcanzar estas etapas. Para poder asignar prioridades a las especificaciones que deben alcanzarse. Por ejemplo asignar prioridad de desarrollo por Departamento, Contable, Cobranza, Atención a clientes, etc.
- Un Data Warehouse desconocido por los usuarios analíticos. El modelo inicial debe cubrir por lo menos las necesidades de información básicas de los usuarios.
- Modelo demasiado rígido que resulta inflexible. Se deben estimar los posibles requerimientos futuros. Si se hace una buena previsión sobre los posibles requerimientos futuros será más fácil preparar un modelo que se adapte de mejor manera a los cambios.
- Extracciones de datos complejas, o no factibles por costo de procesamiento, o lentitud. Se deben identificar las fuentes de datos. Sin una correcta apreciación de la estructura y disponibilidad de las fuentes de datos, se puede llegar a un modelo ideal, pero inoperante debido, por ejemplo, a que la extracción de los datos no sea posible por interrumpir la operación o que no sea posible obtener estructuras eficientes debido a aplicaciones legadas poco confiables.

En otras palabras, si no se toman en cuenta los factores anteriores es poco probable obtener un diseño realista que sea implementado en el tiempo esperado.

2.1.2 Equipo de trabajo

El equipo de trabajo es uno de los elementos centrales del Data Warehouse, es importante tener presente lo siguiente:

- Cada integrante debe ser de tiempo completo.
- Cada integrante debe conocer tener los conocimientos generales que adelante se indican, y ser especialista en algunos de ellos.
- Es muy importante la creación y correcto seguimiento de los planes de trabajo, es sumamente fácil desviarse en pequeñas tareas y alejarse de las metas principales.

Perfil de los integrantes del equipo.

Los integrantes del equipo deben tener conocimientos de:

- Diseño de base de datos.
- Modelado de datos.
- De la operación y reglas del negocio.
- Programación.
- Habilidades de análisis del negocio.
- Administración de proyectos.
- Experiencia en el manejo de herramientas de análisis.
- Habilidades en el manejo de presupuesto.

La cantidad de personas implica algunas cosas según Inmon, el número de integrantes del equipo no debe ser menor a cinco, ni mayor a diez. *“El equipo, no debe tener menos de cinco personas. Si hay menos de cinco personas el equipo será insuficiente. El equipo no debe constar de más de*

diez personas, de lo contrario habrá gente administrando gente y no haciendo trabajo productivo".

2.2 Recolección de Requerimientos

La recolección de requerimientos es la etapa en que se desea conocer las necesidades de los usuarios de la información. Nos guía sobre que aspectos de la organización deben estar presentes en el Data Warehouse y cuales pueden ser considerados como secundarios o incluso posponer o evitar su incorporación. Sin embargo no es una tarea tan sencilla como pudiera parecer, no siempre el usuario es capaz de expresar lo que necesita, el enfoque de sistemas transaccionales que soportan la operación (OLTP) y el de un sistema de soporte de decisiones (DSS) como un Data Warehouse son muy distintos, razón por la cual, aun, un usuario con vasta experiencia en sistemas transaccionales y en los informes que éste arroja le puede resultar complejo entender las ventajas de contar con un Data Warehouse y como aprovecharlas.

2.2.1 Requerimientos Existentes

Un Data Warehouse responde de mejor manera a los requerimientos cuando se ha construido basado en un modelo de datos que represente a la organización que necesita la información. Sin embargo no hay que olvidar que antes de la existencia del Data Warehouse las organizaciones ya requerían información y muy probablemente cuentan con procesos que logran en mayor o menor medida cubrir la demanda de información.

Sólo porque el Data Warehouse represente una nueva y mejorada fuente de información no significa que el equipo que diseña el Data Warehouse pueda ignorar los procesos de obtención de información existentes, por el contrario se debe de recopilar tanta información de estos procesos como sea posible,

esto incluye pero no limita, el analizar reportes, pantallas, hojas de calculo, análisis manuales, etc. Además es recomendable llevar a cabo entrevistas y mantener sesiones que ayuden a obtener los requerimientos.

Sin embargo es importante mencionar que por más cuidado que se tenga al realizar estos pasos, sólo será posible obtener una pequeña parte de los requerimientos esenciales del usuario final, esto debido a que generalmente el usuario esta limitado a expresar lo que necesita en función de lo que ya conoce⁴, desconoce las posibilidades de un Data Warehouse hasta que comienza a verlo en operación, otra actitud del usuario final que se presenta con bastante frecuencia es el expresar que todo lo que se les de sirve, esto derivado de que llegan a estar tan alejados de la información que una vez que les es ofrecida consideran que deben solicitar todo porque puede ser que no vuelvan a tener la oportunidad de solicitar más, la consecuencia inmediata es que no se definan prioridades y que el equipo de desarrollo asuma prioridades equivocadas.

Es frecuente que el usuario final tenga un gran dificultad en explicar cuales son sus requerimientos. Por eso es recomendable que la primera etapa del Data Warehouse esté orientada a cubrir las necesidades básicas, incluyendo los informes a los que el usuario final está acostumbrado obtener.

En este caso se recomienda limitar la recolección de requerimientos por tiempo, de tal manera que una vez que ha concluido el periodo establecido no se recaban más requerimientos. Los requerimientos faltantes serán incluidos en etapas futuras de desarrollo. Uno de los retos a vencer por parte del equipo de desarrollo en la especificación de las primeras etapas es cuando se han definido los alcances de una etapa y se comienzan a ampliar más allá de los previamente establecidos, la consecuencia es que el alcance

⁴ Inmon, Bill 1996

de la etapa continua ampliándose hasta el punto extremo de que la entrega nunca ocurre.

Es importante reconocer que no se va a obtener la totalidad de los requerimientos, sin embargo lo importante es involucrar de lleno al usuario final en esta etapa y no menospreciar los sistemas de información existentes ya que proporcionan una idea de lo que el usuario necesita en una primera etapa del Data Warehouse.

2.2.2 Requerimientos Futuros

Una característica especial de diseño de un Data Warehouse, es que se debe anticipar a los requerimientos futuros. Razón por la cual es necesario contar con estrategias de incorporación rápida de campos, tablas, etc. Esto contribuye a mejorar los tiempos de respuesta a las solicitudes de información y detectar los requerimientos de respuesta inmediata de los que no los son.

Bill Inmon sugiere los siguientes puntos para prever situaciones futuras y mejorar la adaptación a los cambios:

- Acudir a los analistas del ramo al que pertenece la organización y consultar los aspectos que requieren para sus pronósticos.
- Acudir a los administradores y analizar sus reportes estratégicos en un intento de ver que modificaciones se tienen contempladas hacer a corto y mediando plazo.
- Observar otras organizaciones y ver lo que están haciendo y lo que están dejando de lado al implementar sus soluciones.
- Dejar el mayor grado de granularidad posible.
- Desnormalizar a mediano nivel, de manera tal que se puedan aprovechar mejor las relaciones entre llaves primarias y foráneas.

- Consultar al usuario final para conocer sus requerimientos e incorporarlos de acuerdo a un esquema de prioridades.
- Construir el Data Warehouse iterativamente, es decir por etapas que representen cambios, rápidos de implantar.

Es claro que no es posible anticipar todos los requerimientos futuros, pero en la medida que se tengan contemplados será más sencillo que el Data Warehouse evolucione, en buena medida gracias a que los usuarios ayudan a brindarle riqueza a su contenido.

2.3 Etapa de Análisis

En este apartado tomaremos en consideración los aspectos de identificación de fuentes de datos, la granularidad⁵ de la información y algunas implicaciones y factores a tomar en cuenta cuando se diseña el modelo.

2.3.1 Identificación de Fuentes de Datos

Este elemento es generado en la propia organización, a partir de él se realiza la captura de datos derivados de la propia actividad de la organización, y es uno de los orígenes de información del Data Warehouse. Estas fuentes son el mayor proveedor de datos y pueden ser el sistema organizacional o sistemas parciales que reflejan la operación de un área específica.

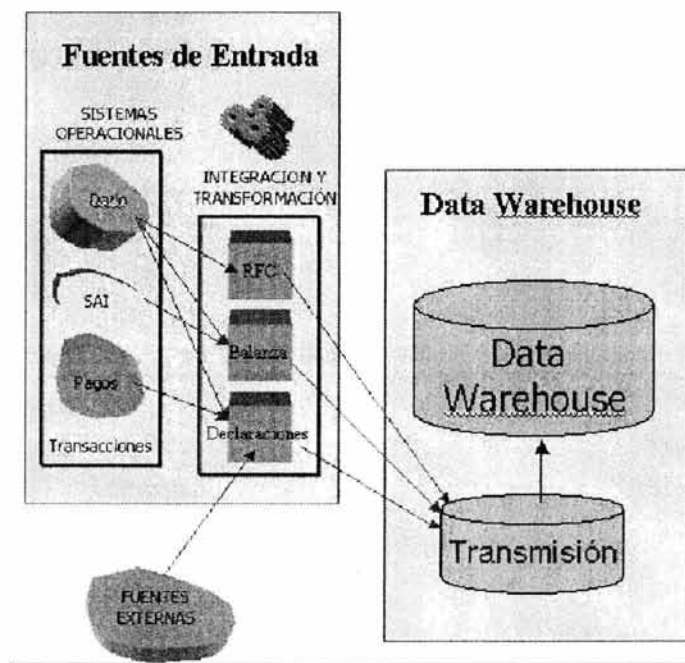
Por su tipo estas fuentes de datos pueden ser principalmente hojas de cálculo, bases de datos, salidas de sistemas manejadores de archivos, archivos planos.

En algunas ocasiones tenemos más de un origen para un atributo o un conjunto de ellos en el Data Warehouse y hay que decidir de dónde se va a

⁵ Consultar Glosario

tomar la información, hay varios factores que inciden en el momento de decidir la fuente de datos, como son:

- El grado de actualidad de la fuente de datos.
- La compatibilidad de la fuente de datos con la estructura del modelo propuesto.
- La presencia de inconsistencias en una fuente u otra.
- La necesidad de procesamiento sobre la fuente para cambiar formatos, separar cadenas de texto, concatenando otras, etc.
- La calidad de datos de la fuente, se necesita saber si hace falta algún proceso de limpieza de datos⁶ derivado de errores de captura, o de los sistemas origen.



⁶ Ver glosario.

Figura 6. Los sistemas departamentales proveen de los datos al Data Warehouse, previo proceso de transformación⁷.

Es importante conocer los momentos del día, semana o periodicidad definida en lo que es posible extraer datos, debido a que estos procesos pueden interferir fuertemente con la operación del sistema transaccional.

Mientras mejor se conozcan las fuentes, menos complicaciones surgirán en el proceso de integración porque será posible detectar a tiempo los cuellos de botella derivados de imprevistos en las fuentes de datos.

2.3.2 Obtener la granularidad apropiada.

Uno de los temas más importantes que enfrenta la administración del Data Warehouse es el de la granularidad de los datos del Data Warehouse.

La granularidad es el nivel de detalle contenido en una unidad de datos, por ejemplo el nivel de detalle de un registro de una tabla.

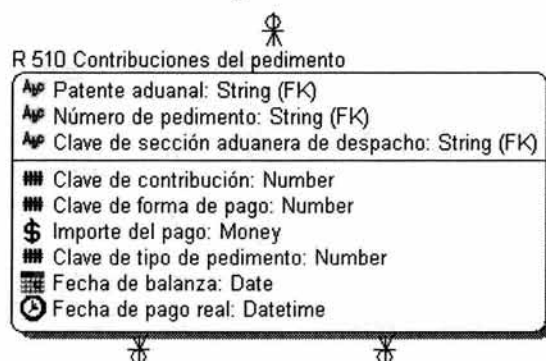


Figura 7 Tabla de contribuciones del pedimento. Granularidad de detalle a nivel documento.

La Tabla de contribuciones del pedimento muestra una tabla con información a nivel documento (pedimento), la de normas oficiales mexicanas, lo hace a nivel detalle de

⁷ Documentación de la Administración de Transformación y Consistencia en Base de Datos del SAT.

la mercancía. Mientras que por cada documento hay solo un registro, en la primer tabla en la segunda puede haber uno o más registros.

^

R 559 Normas oficiales mexicanas

App	Patente aduanal: String (FK)
App	Número de pedimento: String (FK)
App	Clave de sección aduanera de despacho: String (FK)
App	Fracción arancelaria: String
HH	Secuencia de la fracción arancelaria: Number
App	Clave de la norma oficial mexicana: String
App	Identificador de la norma oficial mexicana: String
App	Exención o certificado: String
App	Número de certificado: String
DD	Fecha de balanza: Date
CD	Fecha de pago real: Datetime

Figura 8 Tabla de normas oficiales. Granularidad a nivel mercancía.

A mayor detalle, mayor nivel de granularidad.

El nivel de granularidad impacta fuertemente en:

- El costo del Data Warehouse, a mayor granularidad mayor el costo y viceversa.
- La funcionalidad del Data Warehouse. mientras más alto sea el nivel de granularidad, mayor es la funcionalidad y viceversa.
- El volumen de datos del Data Warehouse, a mayor granularidad mayor volumen de datos y viceversa.
- El desempeño dentro del Data Warehouse, a mayor granularidad, mayor el degradamiento del desempeño y viceversa.

Debido a la que la granularidad afecta varios aspectos del Data Warehouse y de desarrollo, se debe llegar a un consenso entre el área usuario y la de desarrollo, explicando las ventajas y desventajas del nivel de granularidad elegido.

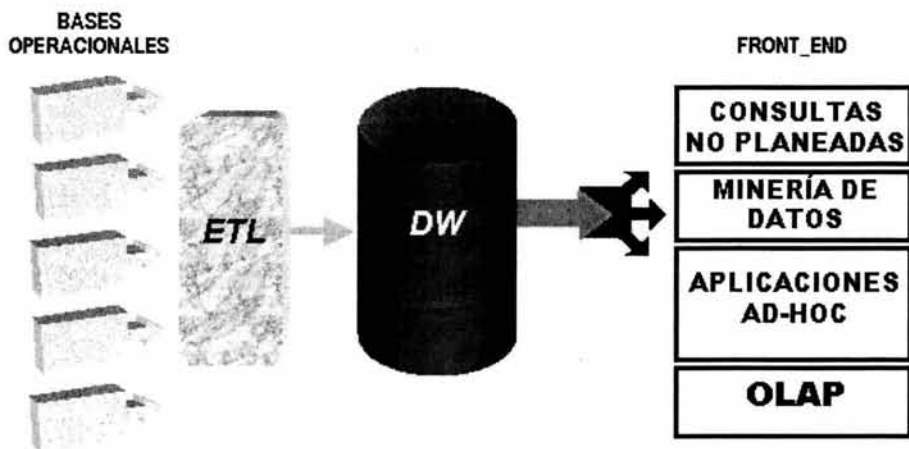


Figura 9. Un nivel de granularidad adecuado ayuda a que el Data Warehouse soporte varias aplicaciones y permite ver la información desde múltiples puntos de vista.

Que un Data Warehouse pueda soportar diferentes maneras de analizar la información y las aplicaciones que la explotan, depende en gran medida del nivel de granularidad. Sin embargo es normal que a lo largo del desarrollo cambie.

Los aspectos más relevantes respecto a la granularidad son⁸:

- Orientar el modelo de datos para soportar múltiples maneras de ver la información.
- Estimar el volumen de datos que se va a generar.
- Tener en cuenta el costo de operar altos volúmenes de datos derivados de un alto nivel de granularidad.

⁸ Inmon, Bill. Metodología

- Cuidar que el desempeño permanezca en límites adecuados de operación.

El resultado esperado de analizar estos factores es determinar el nivel de granularidad inicial.

2.3.3 Implicaciones del diseño físico

El diseño físico de la base de datos es el proceso de llevar los aspectos de diseño analizados a un manejador de base de datos (DBMS⁹). De este diseño se derivan los siguientes puntos:

La estructura y conformación de datos en el DBMS.

El contenido del Data Warehouse.

La estructura de llaves primarias y foráneas.

El volumen de datos que se almacenan.

Cumplir con el nivel de granularidad de la información especificado.

La carga de datos.

Se debe lograr un balance entre la infraestructura existente y los requerimientos, para producir un modelo adecuado.

⁹ Ver glosario.

CAPÍTULO III

Capítulo Tres. Diseño y Construcción.

3.1 Diseño del Modelo Entidad Relación del Repositorio (ODS)

3.1.1 Modelo ODS.

Dentro del desarrollo de un Data Warehouse hay varios modelos que responden en mejor o peor medida según la operación y características de la organización que lo va a implementar, el presente capítulo versa sobre el uso y adaptación de un modelo conocido como ODS.

Comúnmente el componente ODS es empleado como un área temporal para procesar y transformar los datos y posteriormente integrarlos en otra base de datos. Sin embargo en nuestro caso de estudio conforma la base de operación del Data Warehouse de Aduanas. Posee un alto grado de granularidad pero ha sido fuertemente desnormalizado. Su actualización es diaria y a diferencia de las definiciones tradicionales contiene información histórica, además hay una serie de políticas implementadas para los usuarios que pueden consultarlo mediante consulta no-planeada¹⁰, a partir de él se obtienen reportes detallados, y se generan otros productos orientados a temáticas especializadas.

Otras definiciones de ODS son:

ODS: Abreviación del inglés Operational Data Store, o en español almacén de datos operacionales. Se trata de una base de datos que sirve como un área provisional para un Data Warehouse con el propósito de almacenar

¹⁰ Ver Glosario.

datos operacionales de un periodo específico que puedan ser consultados rápida y eficientemente. En contraste con un Data Warehouse, que contiene gran cantidad de datos estáticos, un ODS contiene información que se actualiza conforme se llevan a cabo transacciones en la organización y ejecutará consultas sencillas sobre pequeños conjuntos de datos. Lo que implica que contiene solo datos actuales mientras que un Data Warehouse contiene tanto datos históricos como actuales¹¹.

El ODS se diseña para proveer información integrada y colectiva para el entorno operacional; Se puede actualizar fácilmente y contiene solo información actual. El ODS es en muchas formas la contraparte operacional del Data Warehouse.

Bill Inmon, clasifica cuatro clases de ODS de acuerdo a la manera en que opera y a la velocidad en la que la información derivada de las transacciones se carga en el ODS.

ODS Clase I:

Se caracteriza porque hay un retraso de entre dos y tres segundos entre el momento en que se registra la transacción en el sistema OLTP y el que se carga el ODS. El usuario no nota retraso alguno.

ODS Clase II:

En esta clase las transacciones se almacenan y se envían al ODS después de haberse ejecutado en el OLTP. Típicamente hay un retraso de dos a cuatro horas. El usuario nota el retraso entre la operación y el ODS.

¹¹ <http://itmanagement.webopedia.com/TERM/O/ODS.html>

ODS Clase III:

Básicamente es el mismo mecanismo de la clase II, sin embargo hay un retraso de una noche o 24 horas hasta que la transacción se ve reflejada en el ODS.

ODS Clase IV:

Es aquel en el que los datos provienen del Data Warehouse, a diferencia de las otras clases, los datos ya han pasado por algunos procesos analíticos para cuando se cargan en el ODS, sin embargo el tiempo de respuesta del ODS es muy parecido al de un OLTP.

Aparentemente las únicas diferencias entre las clases de ODS son la frecuencia de actualización del mismo. Pero en realidad hay otras muy importantes. Una es el costo de desarrollo y soporte, el ODS Clase I, es mucho más difícil de construir y operar que un ODS clase II o clase III. Debido a su elevado costo el ODS Clase I, debe haber una fuerte justificación para desarrollarlo, lo cual o sucede en los tipos II y III.

El ODS Clase IV es una herramienta muy potente para el usuario. Permite obtener información previamente procesada rápidamente. En un entorno de interacción directa con un gran número de usuarios, se convierte en una opción invaluable.

Otra diferencia importante es el grado de transformación realizada mientras los datos se integran al ODS. En la Clase I, solo se realizan transformaciones mínimas, de hecho la información se carga tal cual fue generada. En contraste en un ODS Clase II o III se pueden efectuar transformaciones que permitan obtener un mayor grado de integración. Generalmente en los datos se pueden hacer sumarios, cambios en los formatos, hacer conversiones de datos, etc. En un ODS Clase I simplemente no se cuenta con el tiempo suficiente para hacer esto, su alto costo provoca que pierda atractivo.

El modelo empleado para Aduanas se aproxima al ODS Clase II, ya que presenta un desfase de un día respecto al momento en que se realizan las transacciones. Cómo se mencionó previamente se decidió extender el modelo para que contuviera datos históricos, además alimentará otros productos de soporte de decisión, cubos multidimensionales, reportes, minería de datos, etc. El ODS tiene la funcionalidad de base del Data Warehouse, actualmente permite consultarlo para generar reportes predefinidos sobre demanda, consulta no-planeada, cubos molap, rolap.

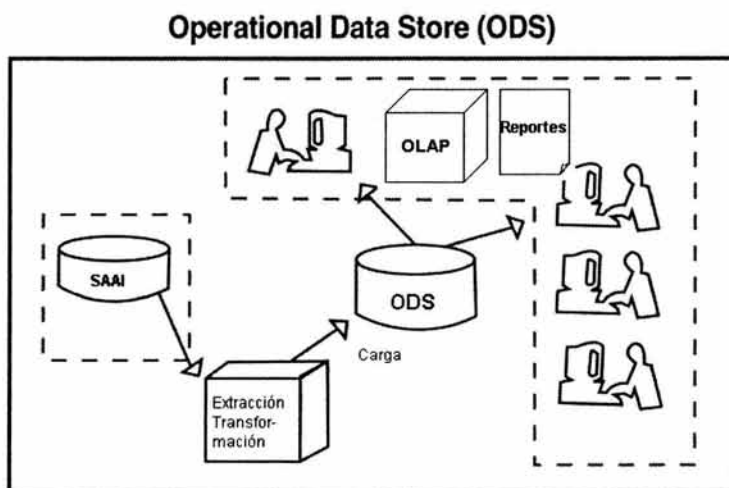


Figura 10 El ODS como base de un Data Warehouse.

3.1.2 Diseño del modelo Entidad Relación.

En el Capítulo dos se mencionaron varios de los aspectos de diseño y los elementos relacionados al mismo, sin embargo una vez tomadas las decisiones relativas al nivel de granularidad, aspectos de la organización que se van a tomar en cuenta para integrar en el ODS, y que se han identificado las fuentes de datos y sus características, se diseña el modelo entidad relación que genere el ODS.

La recomendación de no desnormalizar demasiado el modelo original, contribuye a alcanzar un modelo que siga siendo flexible, con la intención de agregar posteriormente atributos no tomados en cuenta en una etapa inicial. Conviene preparar un diccionario de datos que ayude a identificar cada

Diccionario de Datos de Registros de SAAI M3. (V1.1)

cs_tipo_seccion	Catálogo de Tipos de Aduana.				
Columna	Tipo	Null?	Descripción de Columna	PK?	FK?
c_tipo_seccion	smallint	not null	Llave que define el tipo de sección aduanera	si	n
d_cve_tipo_aduana	char(1)	not null	Clave que define el tipode sección aduanera	no	n
d_descripción	char(30)	not null	Descripción del tipo de sección aduanera	no	n
n_plazo_deposito	smallint	not null	Plazo, en meses, que se tiene para el depósito ante la aduana	no	n
f_ini_vig	date	not null	Fecha de inicio de vigencia	no	n
f_fin_vig	date	null	Fecha de fin de vigencia	no	n
b_estado	char(1)	not null	Bandera que indica el estado de la clave	no	n

Tabla 4 Diccionario de datos de tabla origen.

elemento de las fuentes de datos origen.

En la construcción del modelo del ODS, no debe olvidarse tener en cuenta la estructura de las fuentes de datos origen, ya que puede ocurrir que no sea posible hacer las extracciones, o que se vuelvan muy complejas debido a las transformaciones necesarias para poder cargar datos al ODS.

La elaboración del un modelo entidad relación que cubra los requerimientos y recoja el modo de trabajar y reglas de la organización es un proceso creativo para el que no existe un procedimiento definido¹². Sin embargo sí es posible seguir recomendaciones que ayuden en el diseño. Estas recomendaciones no son reglas que funcionen siempre sino que en algunos casos son adecuadas y en otros no.

El modelo entidad relación del ODS, no debe perder o quitar importancia a las entidades, por el contrario en el nuevo modelo debe facilitarse la obtención de información integral de las misma, ya que siguen siendo los

¹² De Miguel, 2001.

objetos principales sobre los que se recoge información; por lo general son personas, lugares, cosas o eventos. Las entidades aparecen reflejadas en las entrevistas con el usuario como nombres.

Observe la Figura 11, muestra una tabla *Observaciones a nivel partida*, del sistema origen el cual está fuertemente normalizado, después de analizar los requerimientos iniciales se diseñó una tabla en el ODS llamada *Observaciones de la partida* que integra la información de tablas fuente para obtener una sola tabla que facilite la búsqueda de información al usuario, se agregaron fechas, agente aduanal y clave de la sección aduanera de despacho, eliminando la necesidad de hacer cruces con otras tablas para obtener la misma información.

Los atributos que detallan a las entidades asignándoles propiedades deben evaluarse para presentar los atributos que son importantes para el análisis.

Es recomendable agregar los atributos identificadores en cada tabla porque disminuye la necesidad y número de cruces (joins) con varias tablas y en muchos casos se convertirán en los campos índices del modelo físico. Una recomendación adicional para identificar la importancia de determinados atributos descriptivos es saber cuales son obligatorios y cuales opcionales, de igual manera para determinar si la estructura que los almacene será normalizada o desnormalizada se deberán identificar los atributos univaluados y los multivaluados.

Observaciones a nivel partida





 Identificador de pedimento: Number (FK)
 Fracción: String
 Secuencia de la fracción: Number
 Secuencia de las observaciones: Number
Observaciones: String
Posición del registro: Number

Figura 11 Fuente de datos.

3.2 Generación de bases de datos a partir de los modelos entidad relación.

Transformar un modelo entidad relación en un esquema relacional implica cambiar los nombres lógicos de los atributos en nombres físicos, indicar las llaves primarias, foráneas y obtener o crear las sentencias SQL que generen las tablas apropiadas.

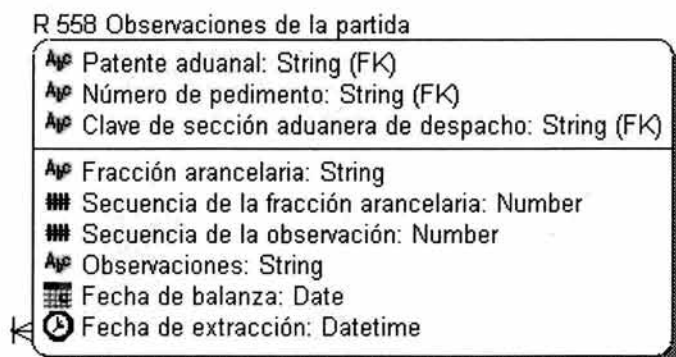


Figura 12. Entidad Observaciones de la partida.

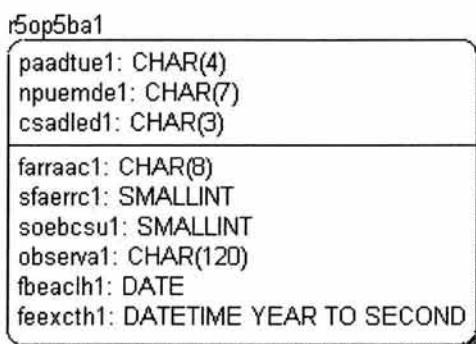


Figura 13. Tabla Observaciones de la partida.

La Figura 12 presenta la entidad *Observaciones de la partida*, dicha entidad da origen a la tabla *r5op5ba1* mostrada en la Figura 13. Mediante la instrucción SQL de la Figura 14 se crea esta tabla.



```
Seleccionar C:\WINNT\System32\cmd.exe - telnet cpnn3
DBSCHEMA Schema Utility          INFORMIX-SQL Version 8.21.UD5X2
Copyright (C) Informix Software, Inc., 1984-1998
Software Serial Number INFN010494
< TABLE "ddwh".r5op5ba1 row size = 158 number of columns = 9 index size = 0 >
create operational table "ddwh".r5op5ba1
<
  paadtue1 char(4),
  npuemde1 char(7),
  csadled1 char(3),
  farraaci char(8),
  sfarrrc1 smallint,
  soebcsu1 smallint,
  observat char(120),
  fbeaclh1 date,
  fpreaer1 datetime year to second
);
revoke all on "ddwh".r5op5ba1 from "public";

cpnn3(1) $
```

Figura 14. Sentencia SQL

3.3 Proceso de Extracción Transformación y Carga

El proceso de Extracción, Transformación y Carga consiste en obtener datos de una o más bases de datos transaccionales, de aplicaciones legadas, de archivos o de la mezcla de estas fuentes. Transformar estos datos para eliminar inconsistencias, agregar información sumariada y finalmente cargar estos datos en el Data Warehouse. El resultado final es contar con una base de datos dedicada que contenga información estable, integrada, no volátil. A su vez esta información es presentada con la dimensión del tiempo, de manera tal que podemos ver información de un día, una semana, o de determinado mes, permitiendo efectuar análisis de tendencias¹³.

¹³ *Ascential DataStage Designer Guide*

3.4 Extracción de datos.

Los datos de sistemas transaccionales o de archivos es la principal fuente de datos para el Data Warehouse. Las bases de datos transaccionales pueden ser archivos indexados, sistemas de bases de datos relacionales, etc. La extracción de datos es el proceso que se obtiene los datos desde las bases de datos transaccionales o fuentes externas¹⁴.

La extracción de datos es el primero de tres pasos que se llevan a cabo para poblar de datos al ODS.

A pesar de la simpleza del enunciado, es uno de los procesos más complejos cuando se implementa un Data Warehouse. El tiempo de análisis y diseño no debe limitarse.

En ocasiones los procesos de extracciones no están bajo el control del equipo de desarrollo del Data Warehouse, puede ser que estén administrados por personal a cargo de los sistemas OLTP, o administradores de base de datos. En tal caso el equipo de Data Warehouse debe proporcionar las especificaciones de extracción necesarias para delegar esta tarea a la gente indicada.

Las especificaciones de extracción deben:

- Determinar dónde se localizan los datos en los sistemas operacionales y en los legados. Puede ser difícil obtener parte de la información. Si esto es así se debe determinar la razón y contar con planes de contingencia para obtener la información en periodos de tiempo razonable y previamente establecido.

¹⁴ *Ascential DataStage Designer Guide*

-
- Definir los estándares que se van a usar. Se debe establecer un estándar para nombrar los archivos, el separador de campos dentro de los archivos, el formato para fechas, etc.

 - Determinar la frecuencia de los procesos de extracción. Este punto debe acordarse entre las áreas que administran los sistemas transaccionales, de usuarios y la de desarrollo del Data Warehouse, ya que generalmente los usuarios desean tener la información tan actual como sea posible, en tanto que los administradores de los sistemas operacionales prefieren extracciones semanales o nocturnas. El área de desarrollo del Data Warehouse puede hacer las observaciones pertinentes buscando un balance entre la operación y la extracción de datos, entre los recursos materiales y el tiempo requerido y a la complejidad y necesidad de los procesos de extracción.

 - Documentar ampliamente las transformaciones de datos a efectuar y solicitar al área de operaciones que informe cualquier cambio en las estructuras de datos de los sistemas transaccionales. Es frecuente que los tipos de datos origen y destino tengan tipos de datos distintos o convenciones para el manejo de archivos diferentes.

Si se cumplen los puntos mencionados se puede ahorrar mucho tiempo en corregir errores derivados de supuestos equivocados, tales como recibir un campo fecha en formato mm-dd-aa, cuando se esperaba como dd/mm/aaaa, recibir archivos separados por comas, cuando se esperaban separados por pipes (|), campos de texto entre comillas cuando se esperaban sin éstas, etc.

3.4.1 Especificaciones de extracción.

Es recomendable crear archivos de extracción delimitados con *pipes*, encabezados y cifras de control. Los datos deberían extraerse en un formato que pueda ser fácilmente accedido por los procesos de transformación. Efectuar conversiones en tiempo de transformación o carga degradarán sustancialmente el desempeño de los procesos. Agregar cifras de control como número de registros, totales, fechas, etc. ayuda a verificar que los procesos trabajan adecuadamente.

En lo posible resulta conveniente extraer registros en archivos dirigidos a tablas del Data Warehouse de manera uno a uno. Esto es, que si se está pensando en alimentar a una tabla del Data Warehouse a partir de dos o más tablas del sistema transaccional, descargar primero esta información integrada en un archivo, que posteriormente se cargará directamente a la tabla del Data Warehouse. Idealmente un archivo de extracción contiene datos para una sola tabla en el Data Warehouse.

Transformar los datos durante la extracción. Se debe hacer toda la transformación posible mientras se efectúa la extracción, por ejemplo convertir tipos de datos, agregar campos calculados, etc. Los sistemas legados tienden a tener tipos de datos numéricos difíciles de descifrar y formatear, es recomendable convertirlos durante la extracción, las fechas también suelen tener inusuales formatos de fechas que causan considerables problemas de conversión. Es recomendable determinar un formato estándar, aplicarlo durante la extracción y conservarlo a lo largo de todo el Data Warehouse. Se debe determinar el manejo de valores nulos y como los tomarán los procesos de extracción y transformación.

3.5 Transformación de datos

Debido a que la información del Data Warehouse generalmente proviene de múltiples fuentes, los datos pueden tener formatos distintos o ser

inconsistentes. La transformación es el proceso que convierte los datos según son requeridos en la estructura y definición del Data Warehouse. Los datos se transforman mediante rutinas basadas en reglas, por ejemplo los códigos de producto se pueden estandarizar mediante una regla que verifique que el dato cumpla con el formato especificado, en caso no hacerlo efectuar las operaciones adecuadas para ajustar el valor¹⁵.

3.5.1 Transformación e integración.

Una de las metas principales del Data Warehouse es tomar datos de varias fuentes dispersas y brindar una visión completa de la organización.

Algunos de los problemas que se presentan en este punto son:

Datos en diferentes tablas con los mismos nombres, pero cuyo contenido representa diferentes aspectos de la organización

Datos en tablas con diferente nombre, pero que representan lo mismo.

Tablas en el Data Warehouse que requieren datos de varias fuentes.

En el otro extremo, puede haber entornos en los que no se necesita efectuar transformaciones. En estos casos simplemente se extraen los datos y se cargan al Data Warehouse sin procesos intermedios. La transformación se limita a un mapeo campo a campo de los campos origen al campo destino en el Data Warehouse.

Las transformaciones más frecuentes son:

Selección de datos. Es la transformación que selecciona todo el registro o parte de él para insertarlo en una tabla del Data Warehouse. Debido al gran volumen de datos encontrados en los sistemas origen, es deseable capturar solo los campos requeridos.

¹⁵ <http://itmanagement.webopedia.com/TERM/E/ETL.html>

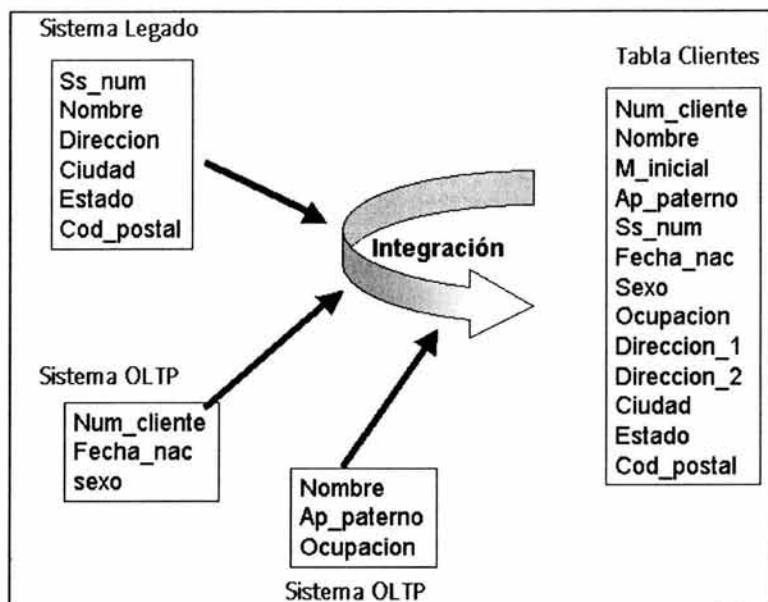


Figura 15 Integración de múltiples fuentes.

Separación de datos. La separación ocurre cuando se divide un campo en dos o más. El propósito de la separación es simplificar a los usuarios ver la información o soportar otros usos de la información. Por ejemplo la siguiente cadena 31462000400650, representa un pedimento, sin embargo para efectos de análisis y relación de la información con otras tablas se divide este valor en sus componentes; Agente aduanal = 3146, pedimento 2000400, clave de sección aduanera de despacho = 650. De esta forma se puede rastrear los pedimentos que ha tramitado este Agente aduanal, o averiguar en que Aduanas tiene la mayor cantidad de operaciones.

Concatenación de datos. Es la operación inversa a la separación. La concatenación une la información de dos o más campos. Permite que un campo se extienda con más detalles sobre el sujeto. Por ejemplo en el Data Warehouse se tiene la información separada del Agente aduanal 3146, el pedimento 2000400 y la sección aduanera de despacho 650. Sin embargo

para poder identificar de manera única a un pedimento es necesaria la unión de los tres campos, ya que cada uno de ellos por separado es repetible. De manera que para obtener el identificador único se deben de concatenar los campos y obtener 31462000400650.

Normalización o desnormalización. Puede requerirse normalizar aspectos de la organización cuando un esquema estrella requiere mucho espacio en disco o cuando tablas demasiado grandes tienen un pobre desempeño en el tiempo de respuesta. La desnormalización puede buscarse para reducir el número de joins implicados en las consultas más frecuentes.

Conversión de datos. Aun cuando la mayoría de las conversiones se deben tratar de efectuar durante la extracción, es posible hacerlo posteriormente. Las conversiones más comunes son: Conversión a campos numéricos, conversión de fechas, valores nulos, convertir cadenas a todas minúsculas o mayúsculas, conversiones de sistema inglés a métrico, conversión de moneda, etc.

Reconciliación. Es el proceso que elimina variaciones en el contenido de la información, mejorando la consistencia de los datos. Por ejemplo en un sistema legado¹⁶ el género se guardaba como 1/0 y el sistema actual lo guarda como H/M, se debe implementar un proceso que estandarice estos datos. Otro ejemplo es la manera en que se guardan nombres, estado civil, etc.

Agregación. Es el proceso de transformación más importante al poblar el Data Warehouse. Se agregan estadísticas como totales, promedios, resumizados, a partir de un conjunto de registros de detalle. Típicamente los datos extraídos de los sistemas operacionales están en su forma más

¹⁶ Ver glosario.

detallada. Sin embargo el soporte para la toma de decisiones usualmente requiere algún nivel de agregación. Sin agregación las consultas deberán efectuar los totales y otras operaciones en tiempo de ejecución. Sumarizar millones de registros tomará más tiempo, independientemente de la configuración del Data Warehouse y de los recursos técnicos con que cuente. Por otro lado, es posible reducir fuertemente la cantidad de registros que almacena el Data Warehouse si los usuarios solo necesitan totales, otro beneficio es que se facilita la navegación y análisis de la información¹⁷.

3.6 Mapeo

El mapeo es un proceso de diseño y documental que se debe efectuar para identificar correctamente los procesos de transformación que se llevaron a cabo y los de carga que se ejecutarán. Ayuda a identificar fallas al ejecutar varios procesos involucrados y mejora el tiempo que lleva dar mantenimiento al Data Warehouse.

Una manera común de hacer el mapeo es utilizar una hoja de cálculo como la que aparece en la Figura, y especificar los campos destino y que proceso atravesaron desde que se extrajeron del sistema origen.

3.7 Carga de datos

La carga de datos es el proceso de poblar el Data Warehouse. Es una tarea fundamental, que generalmente involucra grandes volúmenes de datos y procesos repetitivos, típicamente deben hacerse en breves periodos de tiempo, por ejemplo diariamente durante momentos de baja actividad, semanalmente, en días no laborables, etc.

¹⁷ *Ascential DataStage Designer Guide*

	A	B	C	D	E	F	G	H
1	DESTINO	DATO DESTINO	DATO	TIPO	PROCESO	TABLA ORIGEN	DATO ORIGEN	TIPO
2	R3DQPC01	patente aduanal	pasdue1	char(4)	Seleccionado a través de un catálogo	ci_autoriza_adu	n_patente	smallint
3		número de pedimento	npuede1	char(7)	formateo de campo, rellenado a ceros a 7 posiciones	nm_documento	n_numerocto	integer
4		clave de sección aduanera de despacho	csadied1	char(3)	sin modificaciones	ci_seccion	d_cve_seccion	char(3)
5		clave de documento	cdloacv1	char(3)	sin modificaciones	ci_clave_documento	d_clave_documento	char(3)
6		tipo de cambio	tcasmo1	money(9,5)	sin modificaciones	dd_pedim_gral	d_tipo_cambio	money(9,5)
7		registro federal del contribuyente	rfeog1	char(13)	sin modificaciones	dd_pedim_gral	d_rfc_imp_exp	char(13)
8		curp del contribuyente	curpcn1	char(18)	sin modificaciones	dd_pedim_gral	d_curp_imp_exp	char(18)
9		nombre del contribuyente	ncoombt	char(120)	sin modificaciones	dd_imp_exp	d_nomb_vazsoc	varchar(120)
10		curp del agente o apoderado aduanal	curpaat1	char(18)	sin modificaciones	dd_pedim_gral	d_curp_apad_mand	char(18)
11		clave de prevalidador	cpiraev1	char(6)	sin modificaciones	cs_prevalidador	d_cve_prevalidador	char(6)
12		bandera de justificación	bjausd1	char(1)	sin modificaciones	nm_documento	b_justificado	char(1)
13		clave de banco	cbaanv1	smallint	sin modificaciones	dd_pedim_gral_op	c_banco	smallint
14		número de caja	ncuame1	smallint	sin modificaciones	dd_pago_pedimento	n_caja	smallint
15		bandera de turno de pago	btpaue1	char(1)	sin modificaciones	dd_pago_pedimento	b_turno	char(1)
16		clave de tipo de pedimento	ctplae1	smallint	sin modificaciones	nm_documento	c_tipo_pedimento	smallint
17		fecha de validación	tveach1	datetime	sin modificaciones	nm_documento	f_recabe_pedim	datetime
18		fecha de balanza	fbaech1	date	"01"/month(f_pago)/f_year(f_pago)	dd_pago_pedimento	f_bal	date
19		fecha de pago real	fpreact1	datetime	capturado	dd_pago_pedimento	f_pago	datetime

Tabla 5. Mapeo

3.7.1 Mejoras en el proceso de carga

Hay varias cosas que se pueden hacer para mejorar el desempeño de los procesos de carga.

- El entorno debe configurarse para mejorar acceso a disco.
- Hay un parámetro en la mayoría de los manejadores de base de datos conocido como logging, este se encarga de llevar una bitácora de cada transacción hecha, sin embargo si este parámetro se deja activado mientras se tiene un proceso de carga provocaría que por cada registro insertado se agregara información a la bitácora. Por eso es recomendable desactivarlo o configurarlo para que trabaje a un nivel más general, sin que audite cada registro que es agregado. Es mejor el uso de bitácoras que registren cifras de control, por ejemplo total de registros previos a la carga, registros insertados, total de registros al final de la carga, la hora de inicio y fin de cada carga, etc..
- Borrar índices antes de iniciar un proceso de carga. Se recomienda crear los índices después de realizar la carga de datos. Crear índices mientras se efectúa la carga afectará negativamente el desempeño del proceso.

- Es buena idea deshabilitar las restricciones de integridad referencial antes de un proceso de carga.

3.7.2 Carga y recuperación.

En el caso de que la carga falle, debe tenerse una estrategia de recuperación. Sin importar los procesos de aseguramiento de la calidad envueltos en la transformación y carga, y a pesar de las pruebas exhaustivas que se hayan hecho, las cargas fallarán inevitablemente por una u otra razón.

Es responsabilidad del equipo de Data Warehouse desarrollar estrategias de recuperación para los casos en que el proceso de carga falle.

La estrategia más simple es borrar y crear la tabla nuevamente e iniciar el proceso de carga desde el principio. Es aceptable si se realiza la carga en una tabla vacía. Si la tabla contiene datos, se debe hacer un respaldo previo al proceso de carga.

Una segunda estrategia es cargar incrementalmente los datos en una tabla temporal previo a cargar en la tabla destino. Esta es la estrategia más segura. En caso de que la carga falle, la tabla temporal se puede borrar y volver a crearse sin afectar la tabla destino del Data Warehouse. Cuando la carga se completa exitosamente, es prueba de que la carga se puede hacer directamente en la tabla destino, o bien los registros de la tabla temporal pueden insertarse directamente en la tabla destino. La limitación más fuerte es el tiempo que puede tomar en caso de contar con poco tiempo para hacer la carga.

Una tercer estrategia es agregar un campo de fecha para identificar exactamente que registros pertenecen a que proceso de carga. Si la carga

falla se efectúa un proceso de borrado sobre los registros recién agregados. Una vez que se han borrado los registros y se ha identificado el problema se puede reiniciar el proceso de carga desde cero.

Los procesos de extracción, transformación y carga son las tareas que permiten que el Data Warehouse se mantenga actualizado y que responda mejor a los requerimientos del usuario. Este capítulo presentó sugerencias para llevar los datos de su origen en los sistemas legados hasta su destino en el Data Warehouse. Se mencionaron algunos de los obstáculos y sugerencias para vencerlos y así como otras recomendaciones para mejorar el desempeño de algunos procesos.

CAPÍTULO IV

Capítulo Cuatro. Validación del modelo.

La validación del modelo consiste en efectuar pruebas que permitan detectar errores en los procesos de extracción, transformación, carga y transmisión de datos de las fuentes.

4.1 Validación de datos.

La validación de datos se debe efectuar para comprobar la correcta operación de los procesos de extracción y transmisión de información, es parte del proceso de calidad del Data Warehouse que asegura la consistencia de la información, y ayuda en el objetivo de reunir y consolidar bases de datos diferentes.

4.2 Pruebas de consistencia.

Después de efectuar los procesos de extracción, transformación y carga es necesario asegurarse de que la información ha llegado completa y correctamente. Se deben diseñar consultas que permitan verificar la información, tomando en cuenta que el nivel de granularidad de la información puede ser diferente.

Un primer tipo de verificación simple pero que resulta útil es el tamaño de archivos enviados y recibidos en los casos en que interactúan varios servidores y en que es necesario transmitir por medio de una red los datos entre ellos.

Un tipo de extracción de datos se considera no excluyente si no discrimina a los registros por condiciones especiales como pueden ser tipos de persona física o moral, grandes contribuyentes, personas con más de 100 operaciones anuales, u otros, en este caso la comprobación es tan simple

como comparar el número de registros leídos de la fuente con el número de registros escritos en el destino, tan solo acompañada de los criterios usuales como son fechas de extracción.

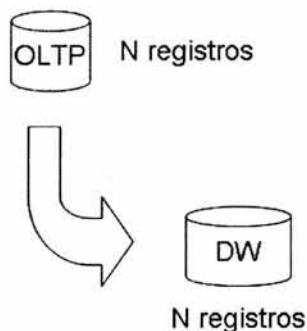


Figura 16. Descargas completas.

El siguiente tipo de extracción es de tipo parcial, ya que a diferencia del caso anterior toma segmentos de la información de acuerdo a criterios específicos, cómo los antes mencionados. En este caso la verificación de información debe incluir sentencias SQL que también discriminen los registros de la fuente y se verifiquen totales contra el destino, aplica la ejecución de funciones agregadas como contadores de acuerdo a los grupos formados.

Un tercer caso involucra la aplicación de reglas de negocio, por ejemplo a partir de la tabla fuente, se desea conocer los totales de impuestos de un contribuyente, sin embargo de acuerdo a las características de éste se aplicarán diferentes criterios para obtener el total. El problema más común en este caso es la generación incorrecta de cifras. Se deben prever los diferentes casos posibles, establecer una muestra que los satisfaga y comprobar los valores esperados mediante consultas contra la fuente datos y el destino.

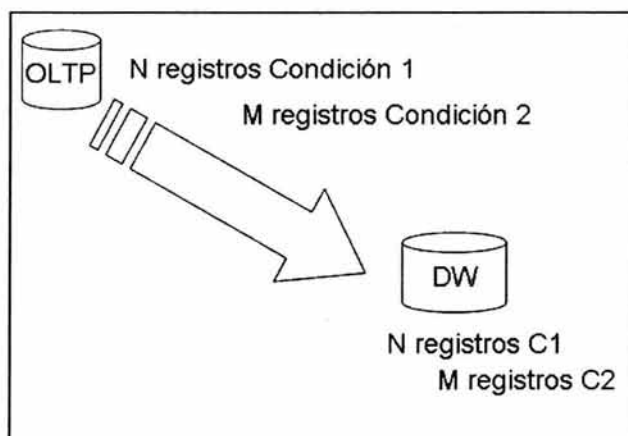


Figura 17. Descargas parciales

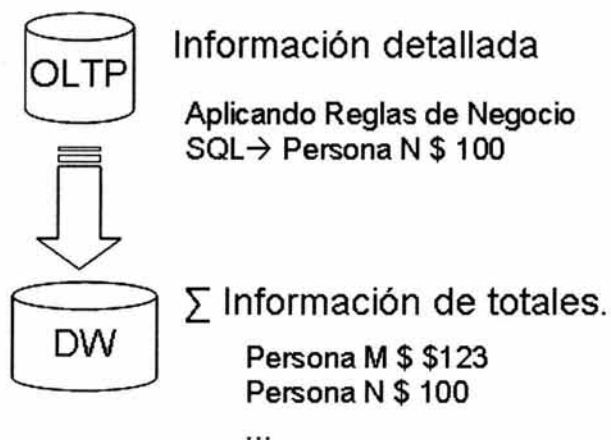


Figura 18. Comparación de información entre la fuente y el destino.

En la realidad se presentan combinaciones de los 3 casos básicos, en ocasiones los procesos involucran varios pasos, por ejemplo a partir de una

fuente se extraen los registros completos hacia una tabla intermedia destino, a partir de esta se discriminan y agrupan datos y al final se obtiene una tabla de grandes sumarios, en cada paso es necesario establecer puntos de revisión de cifras; totales de registros, contadores, totales por grupos de datos, etc.

Una vez que se ha identificado el caso o la combinación de casos se debe probar el esquema completo del Data Warehouse. Comparar el contenido de los registros de las fuentes contra el contenido destino, realizar los ajustes necesarios en el proceso ETL y volver a ejecutar las revisiones.

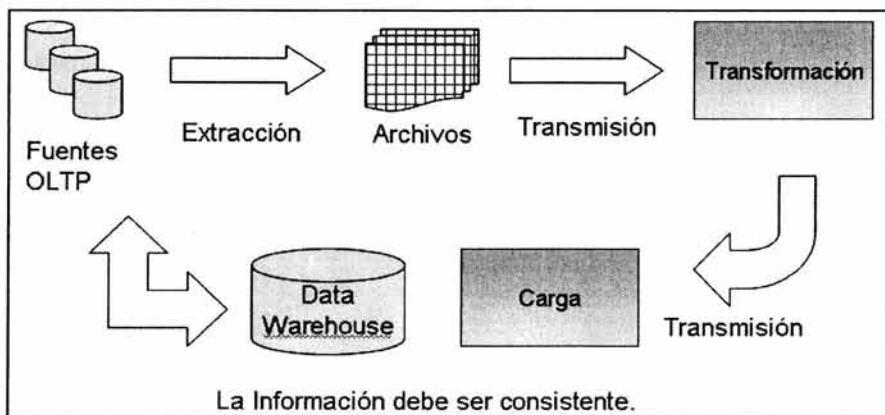


Figura 19. Diagrama de procesamiento y transmisión de datos, de las fuentes origen al Data Warehouse.

Es importante mencionar que los procesos de ETL deben tener reglas claras de rechazo de registros. Haciendo una analogía entre los procesos de ETL y un policía al que se le indica dividir en dos grupos a la gente que pasa por un corredor según tenga vestimenta color rojo o azul de acuerdo a una regla de negocio que indique que sólo se presentará este par de casos, se supone

que éste hará bien su trabajo, hasta el momento que aparezcan personas con colores distintos o sin ropa. Una cláusula que indique que acción tomar en los demás casos le ayudaría al policía a saber que hacer. En nuestro haber, estos casos se presentan frecuentemente en forma de registros duplicados, valores en blanco, nulos, valores inválidos, etc. Provocados por errores en las aplicaciones transaccionales, errores de captura, modificaciones a los sistemas, de base de datos, etc. Contar con grupos de rechazados disminuye errores masivos de carga y ayuda a identificar más fácilmente las causas de problemas, lo que redundará en retroalimentar a los sistemas transaccionales para corregir errores, modificar el conjunto de reglas de negocio de los procesos ETL, etc.

CAPÍTULO V

Capítulo Cinco. Automatización

5.1 Afinación de procesos

La afinación de procesos es una etapa muy extensa y cubrirla con gran profundidad queda fuera de los alcances de este capítulo, por lo tanto nos limitaremos a mencionar algunos factores y aspectos técnicos que tienen que ser tomados en cuenta para disminuir el tiempo y mejorar el uso de los recursos computacionales.

Acceder a la información origen de la organización, extraerla, transformarla y ponerla a disposición para consulta en el Data Warehouse, por si misma ya es una compleja tarea, que sin embargo está sujeta a mejoras, a esto es a lo que llamamos Afinación de procesos, y toca a todas las fases del Data Warehouse y comprende desde analizar y modificar las operaciones más elementales en un equipo de computo como son modos de lectura y escritura en disco, hasta evaluación de lenguajes de programación y diseño, modificación a estructuras de Base de datos. En ocasiones se puede determinar necesario el cambio de manejador de base de datos o de sistema operativo, modos (uso de protocolos) de transmitir información, etc. A pesar de la sencillez de la idea conseguir las mejoras no es tan simple. Debido a la importancia del manejo de recursos humanos y materiales se considera importante dedicar este apartado, dónde se mencionan diferentes puntos a considerar, teniendo presente la posibilidad de recurrir a especialistas y reunir a diferentes niveles de la organización como son directivos, líderes de proyectos, encargados de procesos transaccionales, etc. para lograr una serie de modificaciones que por si solas o en su conjunto mejoren el desempeño de los procesos.

5.1.1 Identificación de factores que inciden en la afinación a corto plazo de los procesos

Algunos son:

- Asignar prioridades.
- Identificar cuellos de botella.
- Identificar procesos que presentan el mayor número de fallas.
- Identificar en que capa se están presentando retrasos (manejador de base de datos, transmisión –red, sistema operativo, necesidad de hardware de mayor capacidad, etc.)

5.1.2 Identificación de factores que inciden en la afinación a mediano plazo de los procesos

La mayoría de los temas tienen que ser manejados con profundidad para el afinamiento de procesos, algunos se desarrollan en el sistema operativo de los equipos, otros en los lenguajes de programación, en el manejador de base de datos, en los procesos de transmisión de archivos y el hardware asociado, etc. Debido a la extensión de cada tema se recomienda acudir a especialistas que ayuden a resolver las problemáticas expuestas en cada rubro.

5.1.3 Lenguajes de programación:

Se debe evaluar las bondades y desventajas de los lenguajes en que se desarrollan los programas asociados al Data Warehouse, por ejemplo posiblemente no valga la pena afinar y programar en un lenguaje complejo – de difícil mantenimiento- una tarea que apenas consume unos cuantos segundos, por otro lado puede resultar muy benéfico dedicar grandes esfuerzos optimizar una tarea que involucra el procesamiento de altos volúmenes de información y de la cual dependen otras tareas para poder llevarse a cabo, se recomienda tener presente los siguientes puntos al considerar los lenguajes de programación:

- Características de lenguajes
- Ventajas y desventajas generales
- Facilidad de uso vs. eficiencia
- Recursos humanos que pueden desarrollar en él.
- Facilidad de mantenimiento.
- Soporte técnico, consultoría disponible.
- Plataformas en las que se puede ejecutar
- Portabilidad.

5.1.4 Recomendaciones generales:

Utilización de notificaciones o señales, si un proceso puede captar eventos también conocidos como interrupciones de software, como ocurren (sincrónica o asincrónicamente), si las genera el sistema operativo o un proceso en ejecución, se puede lograr que interactúe con otros procesos. El uso de los procesos que puedan ser ejecutados en tiempos específicos, la implementación de procesos tipo demonios (UNIX) que permiten la ejecución de procesos independientes de operadores, terminales y que están disponibles en cualquier momento para realizar una tarea específica ayudan a lograr cierta automatización.

Es recomendable profundizar en la arquitectura Cliente Servidor. Esto nos puede ayudar a generar aplicaciones que brinden servicios a varios equipos, por ejemplo: la implementación de servicios automáticos como sockets que son procesos que están a la espera de ser contactados por un proceso cliente que puede operar en la misma plataforma o en otra, para realizar una tarea específica como la transmisión de archivos entre equipos.

Implementar servicios o procesos demonios.

Que corran automática, o permanentemente en espera de señales que indiquen que deben ejecutar determinadas tareas.

Programación en Red.

La programación en red involucra dos o más procesos interdependientes que pueden verse o comunicarse entre ellos utilizando diferentes técnicas IPC (comunicación entre procesos) disponibles en el sistema operativo como: Archivos, bloqueo de registros¹⁸, tuberías¹⁹, flujos (Streams)²⁰, colas de mensajes, espacios nombrados (Name Spaces), semáforos, memoria compartida, sockets o procesos que proveen comunicaciones entre procesos de diferentes sistemas (equipos)

A continuación se describen sugerencias generales relacionadas con los manejadores de Base de datos:

1) Organización de datos.

Se recomienda verificar si las definiciones para espacios en disco en donde residirán las tablas de la base de datos son los más convenientes para los tipos y el uso que se piensa dar a los datos, por ejemplo en algunos manejadores existen conceptos de diferentes tipos de espacios en disco para tipos de datos específicos y que pueden mejorar el acceso a los mismos.

Es importante el almacenamiento en disco y su configuración, no es lo mismo tener un solo disco donde trabaja el manejador de base de datos con sus tablas y espacios temporales, más el uso de programas de transformación (uso de temporales), carga (lectura de archivos y uso de rdbms) que distribuir la carga de trabajo en varios discos, o en arreglos de discos en cluster.

¹⁸ Ver glosario

¹⁹ Ver glosario

²⁰ Ver glosario

2) Tipos de datos y manejo interno de ellos.

La elección de tipos de datos adecuados para el almacenamiento de la información, nos puede brindar ahorro de espacio y mejor el acceso.

Por ejemplo: el uso de enteros o boléanos en banderas en lugar de campos char, el uso de tipos varchar en lugar de char.

3) Índices.

Los tipos de índices cluster o el uso de índices compuestos pueden mejorar el acceso.

4) Fragmentación.

La fragmentación de tablas e índices en algunos manejadores de bases de datos en diferentes espacios pueden mejorar el acceso a los datos, siempre y cuando se cuente con discos que trabajen en paralelo.

5) Paralelismo.

El uso de paralelismo dentro de un proceso que se ejecuta en un manejador de base de datos, puede mejorar en acceso a los datos.

6) Uso del Optimizador.

La evaluación del plan de búsqueda de un query, puede indicarnos si se están realizando por índices y cual es el costo de la misma, lo cual puede traer como consecuencia el cambio en las condiciones de acoplamiento entre las tablas que componen las búsquedas.

7) Actualización de Estadísticas.

El uso adecuado de la actualización de estadísticas de acceso a la información puede ayudar al optimizador de consultas del manejador de base de datos a tomar mejores decisiones en el momento de elegir el camino de acceso a los datos, se deben realizar pruebas para cada una de las opciones disponibles como la actualización a nivel base de datos, tabla y

columnas, así como la profundidad planeada alta, media y baja, El tiempo de la ejecución de puede llevarse la ejecución va de unos minutos hasta horas, básicamente depende de los volúmenes de información existentes en la base de datos.

8) Personalización o parametrización del Optimizador.

Se recomienda el uso de directivas al optimizador de accesos solamente cuando se han efectuado pruebas y se ha detectado que el optimizador no reconoce el mejor camino para resolver consultas.

9) Manejo de Integridad referencial por parte del RDBMS.

El uso adecuado de la verificación de integridad puede ayudar a disminuir el tiempo de actualización a una base de datos, en los casos de reemplazo de tablas o cargas masivas de información es recomendable eliminar la verificación de integridad, al finalizar dichas operaciones se debe reactivar.

10) Aislamientos.

El manejo correcto del nivel de aislamiento de transacciones (read uncommitted, read committed, repeatable read, serializable read) en una base de datos, puede ayudar a mejorar los accesos, dependiendo del grado elegido el manejador tendrá que destinar una cantidad de variable de recursos para proteger la concurrencia de procesos.

11) Procedimientos Almacenados.

El uso de procedimientos almacenados reduce significativamente el acceso a disco, en aquellos procesos enfocados a la explotación de información, ya que únicamente se evalúan cuando se guardan en la base de datos, a diferencia de lo que sucede con las instrucciones de un programa externo, dónde se evalúan cada vez que se ejecutan.

12) Implementar técnicas de Data Warehouse.

Como tablas dimensionales, normalización y desnormalización, granularidad y fragmentación o partición de datos.

Identificación de factores que inciden en la afinación a largo plazo de los procesos

1. Volumen de la información de la organización: Se tiene que determinar en lo posible, el espacio requerido para cada elemento del Data Warehouse, dependiendo de los volúmenes de operación y de información requerida se puede hablar de megabytes, gigabytes o terabytes de espacio requerido en disco.

2. Capacidad de los equipos de cómputo. Se debe evaluar la capacidad de los equipos de computo, es decir dimensionar si son capaces de soportar el procesamiento adicional que será introducido y conocer los límites de los mismos.

3. Sistemas operativos disponibles. Elegir el más adecuado para las necesidades de procesamiento.

4. Lenguajes de programación disponibles. Evaluar cuales son los idóneos para cada una de las partes de la solución propuesta.

5. Manejadores de base de datos. Conocer a profundidad los que están siendo utilizados en repositorios operacionales y hacer uso del indicado para el repositorio final del Data Warehouse.

6. Infraestructura de comunicaciones disponible. Se debe evaluar la capacidad actual de la infraestructura de red, ya que puede resultar insuficiente para los volúmenes de transmisión de datos, y para soportar las consultas hechas al Data Warehouse una vez que éste se libere.

7. Nivel de conocimientos del personal de sistemas. Puede ser necesario capacitar ampliamente a los integrantes del equipo de desarrollo o ser necesario recurrir a una consultoría.

8. Recursos económicos. Se debe analizar exhaustivamente el costo de implementar una solución de este tipo ya que la organización debe estar conciente de la inversión necesaria en la implementación de una solución.

9. Disponibilidad de herramientas comerciales. Se debe hacer la evaluación de las herramientas disponibles que ayuden a cubrir parte de los procesos, disminuyendo tiempo de desarrollo, sin olvidar los aspectos de facilidad de uso, soporte técnico, licenciamiento, requerimientos y costos.

Es recomendable avanzar todo lo posible en la afinación del modelo de Data Warehouse ya que no hacerlo puede afectar la implementación en diversas formas, retrasos para contar con la información, saturación de la red, caída en el desempeño de los sistemas OLTP, falta de recursos para terminar el proyecto, etc. Es cierto que no existe como tal una receta a seguir, se debe proceder con mucho cuidado y se tiene que contar con tiempo para evaluar cada factor, una decisión mal tomada puede repercutir en altos costos (horas hombre, recursos económicos) para la organización. Se recomienda realizar un análisis de costo beneficio de cada elemento del Data Warehouse que deba ser afinado, así como de su relación con otros para donde conseguir una integración de tareas, recursos, tiempos y responsables.

5.2 Automatización

En una organización donde se intente implementar un Data Warehouse es imprescindible automatizar en la medida de lo posible los procesos de ETL, con esto podemos lograr ahorros de horas hombre y evitar el desgaste del personal en tareas repetitivas además disminuye la posibilidad de errores humanos.

La mala distribución del tiempo en algunas de las etapas para integrar una fuente de datos al Data Warehouse, trae consigo que dediquemos más tiempo a actividades secundarias y descuidemos las importantes, un ejemplo de una mala distribución se aprecia en las etapas para la incorporación de fuentes al Data Warehouse.

Etapas.

Identificar la fuente requiere el 40% del tiempo total

Generar los modelos de extracción 25%

Carga y transformación 25%

Publicación 10%

Para la identificación se debe hacer el análisis de los modelos de datos y su documentación

Extracción: programas de extracción, Verificar Integridad de los datos origen con respecto a la información centralizada, especificar los medios de transmisión de los datos.

Carga y Transformación: se debe manejar adecuadamente los registros duplicados, la depuración de registros y las reglas de rechazo. La implementación adecuada de las reglas de negocio es una tarea que requiere mucho tiempo en la implementación adecuada de las mismas.

Publicación: Dificultad en la integración de la información a los repositorios finales. Para determinar el grado de automatización de los procesos de ETL conviene tomar en cuenta sus posibles modos de operación:

Extracción

Para mejorar los procesos de extracción es necesario conocer que tipo de extracción se ajusta más a nuestras necesidades en seguida se enlista:

Extracciones estáticas: son descargas totales de una tabla o base de datos en un periodo de tiempo y que servirán para poblar al Data Warehouse, esta descarga puede contener todos los datos de los sistemas operaciones o un subconjunto de ellos, se consideran estáticas por que la tabla destino no cambian en un periodo de tiempo que puede ser una semana o mes, hasta que la información es reemplazada por otra extracción. Este tipo de extracciones son utilizadas para poblar inicialmente el Data Warehouse.

Extracciones dinámicas: También conocidas como incrementales y en esta se extrae solo la información que ha cambiado, este tipo de extracciones son útiles cuando se requieren reflejar más rápidamente los cambios ocurridos en los sistemas operacionales y es necesario mantener una completa historia de los datos. En otras palabras deben permitir adicionar la información que no existía en el Data Warehouse, esto puede involucrar operaciones de agregación, borrado y actualización de datos.

Algunas técnicas para capturar el cambio de datos puede ser la utilización de triggers²¹, logs de captura²², capturas basadas en campos de tiempo o la comparación de archivos.

Sugerencias

²¹ ver glosario

²² ver glosario

En los procesos de extracción de los datos origen se tienen que considerar:

- Los horarios de operación de los sistemas transaccionales, la competencia por recursos de hardware,
- El volumen de datos a extraer para dividir las extracciones en partes mas pequeñas, de ser posible comprimidas y que se ejecuten con la mayor frecuencia posible sin que degraden a los sistemas operacionales.
- Robustecer los métodos de comunicaciones entre equipos para lograr conexiones más rápidas y seguras.
- Aprovechar las características de los manejadores de bases de datos para efectuar cargas rápidas y las distribuciones más acordes a los modelos propuestos.
- La capacidad para introducir modificaciones a la definición de cada una de las extracciones en catálogos fácilmente actualizables.
- Considerar técnicas para el transporte de la información entre diferentes sistemas, tales como el uso dispositivos de respaldo, la utilización protocolos de comunicaciones como FTP (file transfer protocol), la utilización de servicios de transferencia (sockets), etc.

Integración (carga de datos y transformación).

Estos son los procesos que requieren de un ambiente de trabajo más robusto, debido a que se van a manejar los datos de diversas fuentes de datos, se deben considerar los siguientes factores:

El volumen de datos y su crecimiento para efectuar las cargas con la mayor frecuencia posible y averiguar si los datos pueden manejarse de forma comprimida para disminuir el consumo de ancho de banda y a su vez el tiempo de transmisión y espacio ocupado en disco.

Evaluar el uso de programas en C, Awk scripts, uso de herramientas de transformación desarrolladas por terceros, en base a la complejidad de la tarea y a la facilidad de implementar una solución.

Alternativas para el proceso de extracción.

Realizar por medio de un operador una extracción estática y total o de un subconjunto de las tablas origen, enviar en un dispositivo como DDS (Digital Data Storage) o vía FTP a la información a la localidad donde se concentrará.

Realizar extracción estática y programada de un subconjunto de tablas en un horario especificado que no afecte la operación utilizando funciones de envío remoto RCP (Copia remota de archivos).

Desarrollar procesos que se ejecuten diariamente en un horario que no afecte a la operación y que realice las extracciones de forma dinámica de las bases de datos locales y las concentre en una sola localidad, este proceso puede ser implementado como un demonio²³ que se encargue de controlar la extracción y transmisión de la información.

La transmisión se puede realizar con sockets²⁴ para automatizar el envío, aumentar la seguridad y para minimizar el espacio en disco y el consumo de ancho de banda de la red se manejarán los datos en archivos compactados. Para solucionar los problemas de comunicaciones se debe contar con un proceso que verifique cifras de control y cuente con un retransmisor de datos.

²³ Ver glosario

²⁴ Ver glosario

Se debe contar con un esquema de replica²⁵ de catálogos automatizados a nivel nacional. Implementar las definiciones de extracción en un esquema de catálogos que pueda ser replicado.

Opciones para el proceso de carga.

Desarrollar un proceso de carga sustitutivo que reemplace toda la base de datos cuando se ejecute, utilizando manejadores de bases de datos tradicionales.

Desarrollar un proceso de carga incremental que reemplace solo las tablas seleccionadas, se utilizará un manejo tradicional de base de datos.

Desarrollar procesos de carga diaria que incrementaran la base de datos de forma diaria, considerara la opción de reemplazo de información, la carga de tablas es parametrizada y controlada por catálogos de carga, permitirá la utilizar información compactada, se deberán utilizar manejadores orientados al Data Warehouse, con funciones de carga rápida. La Fragmentación de los datos a todos los espacios disponibles. La validación de las reglas del negocio estará inmersa en la programación.

Opciones para procesos de transformación.

Implementar un esquema de transformación mensual si se cuenta con un manejador de base de datos tradicional.

Implementar un esquema de transformación semanal en un manejador de base de datos tradicional con amplia disponibilidad de procesamiento.

²⁵ Ver glosario

Implementar un esquema de transformación diaria en un manejador de base de datos orientado al Data Warehouse, que permita el uso de descargas rápidas y de forma diaria, se aprovechará la fragmentación de los discos, las reglas de negocio estarán parametrizadas.

CAPÍTULO VI

Capítulo Seis. Explotación del ODS

El proceso de explotación del Data Warehouse es un punto que conviene revisar a fondo en virtud de las posibilidades que ofrece para tomar decisiones mejor informadas y estratégicas para el desarrollo de la organización.

Haciendo una breve remembranza cabe mencionar que a partir de los setentas se comienza a vislumbrar el concepto de sistemas para la toma de decisiones, que fueron en principio un conjunto de herramientas de análisis acerca de la información disponible de la operación de la organización y eran utilizados casi en exclusiva por especialistas en programación y análisis de datos. La llegada de las bases de datos relacionales, las computadoras personales y la graficación por computadora, el acceso a la información y su representación por parte de los usuarios finales comenzó a ser cada día más frecuente.

6.1 Herramientas para la explotación de los datos del ODS.

En la actualidad, la disponibilidad de herramientas y el desarrollo de la tecnología han permitido que se genere y almacene una enorme cantidad de información acerca de las operaciones de las organizaciones. Sin embargo es muy poco probable que toda esta información se utilice oportuna y adecuadamente solo por el hecho de que se encuentra almacenada. Este gran monto de información es poco relevante para la toma de decisiones, por ello es necesario convertirlo en información importante para la organización; adicionalmente la información debe presentarse fácil y accesiblemente a los usuarios finales.

Al conjunto de métodos y herramientas que permiten transformar los datos en información relevante y accesible para los tomadores de decisiones es a

lo que llamamos explotación del Data Warehouse o bien Inteligencia de Negocios (BI, por sus siglas en inglés).

Existen diversas herramientas para la explotación del Data Warehouse, dentro de las que destacan las OLAP y sus variantes, reporte y minería de datos.



OLAP es un acrónimo de OnLine Analytical Process (Procesamiento analítico en línea) permite ofrecer vistas multidimensionales de los principales aspectos de la organización, responde preguntas tales como ¿quién?, ¿cómo?, ¿cuánto?, ¿dónde?, ¿qué?

Las herramientas OLAP funcionan sobre un sistema de información (transaccional o almacén de datos), permiten realizar agregaciones y combinaciones de los datos de maneras mucho más complejas y ambiciosas, con objetivos de análisis más estratégicos. Proporcionan facilidades para manejar y analizar los datos. El resultado se presenta de una manera matricial, híbrida o mediante gráficas.

Las herramientas de reporte y los sistemas de informes o consultas avanzadas por su parte están basados, generalmente, en sistemas relacionales u objeto-relacionales, utilizan los operadores clásicos: concatenación, proyección, selección, agrupamiento, (en SQL y extensiones), atiende generalmente las necesidades más elementales de información. Los resultados se presentan de manera tabular o de gráficas.

La minería de datos es la integración de inteligencia artificial, estadística y bases de datos, capacidad de procesamiento masivo de datos.

Las herramientas de minería de datos son muy variadas: ayudan a identificar patrones, modelos, descubrir relaciones, regularidades, tendencias, etc. Producen reglas o patrones. Permiten conocer los datos y variables que

generan diversos efectos en la organización. Son un importante apoyo para establecer modelos que reproducen situaciones del negocio de tal suerte que es posible llevar a cabo pronósticos, evaluación de riesgos y pueden llegar a responder el ¿por qué?.

Es conveniente mencionar que ninguna herramienta debe ser implantada de manera discrecional, sin antes analizar su posicionamiento tecnológico y si ésta se encuentra alineada con los objetivos estratégicos de la organización, el valor que aporta y el retorno de inversión que ofrece.

La explotación del Data Warehouse debe arrancar teniendo como base aquellos puntos donde se puedan aportar mejoras sustanciales a los procesos de mayor peso, en este sentido se recomienda el desarrollo incremental de soluciones.

Se debe implementar una estrategia que mediante el uso de herramientas permita explotar la información y entregarla adecuadamente a los diferentes niveles de usuarios de la organización. El éxito de un Data Warehouse no radica tanto en su construcción, sino en su explotación para mejorar los procesos de la organización, sus operaciones y la toma de decisiones. Lograr que el Data Warehouse sea usado efectivamente, requiere entender sus impactos en la organización, además disponer de los datos, contar con un grupo de personas que sepan que hacer con ellos, interpretar resultados y definir acciones a partir de éstos:

Efectos sobre los recursos humanos:

El Data Warehouse intenta facilitar el acceso a la información a los usuarios cuando la necesitan. Este enfoque de entrega de información tiene varias implicaciones:

- La gente de la organización puede necesitar aprender nuevas destrezas.

- Los análisis extensivos a través de información muy detalla y poco significativa, así como las demoras para obtener información serán eliminados o bien disminuidos. Dado que la información estará lista para ser consultada, los requerimientos probablemente aumentarán.
- Los departamentos informáticos tienen posibilidad de detectar nuevas áreas de oportunidad.
- Reducción de reportes impresos.
- La evolución del Data Warehouse esta en función de la interacción con sus usuarios.

Impactos en la Organización.

Procesos y Decisiones.

Se deben considerar los beneficios potenciales de los siguientes impactos:

- Los procesos que implican la toma de Decisiones pueden mejorarse por la disponibilidad de información. La gente de niveles ejecutivos puede tomar decisiones más rápidamente al estar más informada.
- Disminuye el tiempo perdido esperando información que al final puede ser incorrecta o no encontrada.
- Las secuencias de procesos pueden ser optimizados para ganar eficiencia y reducir costos al volverse más claros y entendibles.
- La información se estructura y pone a disposición en términos propios de la organización. Los defectos de las aplicaciones actuales y legadas pueden quedar expuestos, siendo posible corregir aplicaciones y evaluar de mejor manera la certeza de los datos.

Comunicación e Impactos Organizacionales.

Algunas consecuencias inmediatas de contar con un Data Warehouse como principal fuente de información son:

- Hay mayor confianza hacia las decisiones tomadas, ya que están basadas en información confiable.
- La información y conceptos conducen a un lenguaje común. Se mejora la confianza y cooperación entre distintos sectores de la organización.
- La transparencia, acceso y conocimiento de los datos generan mayor confianza en los sistemas operacionales.

¿Quiénes y para qué lo usan?

A continuación veremos ejemplos de aplicaciones típicas y algunos casos puntuales en distintas industrias.

Las necesidades de información en la Organización que ha implementado el Almacén de datos en términos generales, podemos decir que corresponde a entornos en los que se identifican grandes volúmenes de datos, asociados a: cantidad y usuarios de distintos niveles, de clientes, variedad de productos o servicios ofrecidos y cantidad de operaciones.

A continuación se describen ejemplos de usos típicos y casos concretos en distintos sectores productivos.

Sector tiendas de autoservicio:

Utilizan su Data Warehouse para consultar información de meses o años de venta tomada en los puntos de venta. Con ésta pueden efectuar en forma más precisa y eficiente compras, determinación de precios, manejo de inventarios, disposición mercancías en estantes, etc.

A las promociones se les da seguimiento, se analizan, corrigen y mejoran. Las modas y tendencias son cuidadosamente estudiadas a efectos de maximizar utilidades y reducir costos de inventario. El producto es reasignado por sucursales o regiones según sus ventas y tendencias, es posible implementar eficientemente prácticas de mercadotecnia “en

consignación”, en esta modalidad la cadena de tiendas paga al proveedor conforme los productos son vendidos en el punto de venta.

Esta información proporciona ventaja competitiva y permite ejercer mayor poder de negociación sobre los proveedores, dado que la cadena puede llegar a saber más que el fabricante sobre sus productos: ¿quién los compra?, ¿dónde?, ¿cuándo?, ¿con que otros productos?, etc.

Uno de los desarrollos de este tipo con mayor cantidad de datos pertenece a la cadena comercial Wal*Mart que cuenta con unas 2500 sucursales en el mundo y que cada una de ellas tiene una variedad de entre 50,000 y 80,000 artículos. Otros desarrollos de gran magnitud en la industria minorista de autoservicio son los de Kmart y Sears.

Sector Manufactura de Bienes de Consumo Masivo.

Las empresas de este sector necesitan un manejo ágil de la información para mantenerse competitivas en la industria. Buscan predecir la cantidad de producto que se venderá a un determinado precio y por consiguiente producir la cantidad adecuada para una entrega justo a tiempo. Al mismo tiempo se organiza la distribución a cadenas minoristas con inmensas cantidades de productos en consignación, que serán pagados hasta llegar al consumidor final.

Las cadenas minoristas y sus proveedores utilizan el Data Warehouse para compartir información, permitiéndole conocer el nivel de producto en los estantes y eventualmente hacerse responsables del reemplazo de inventario de la cadena minorista.

Un ejemplo es el de Whirlpool. Este fabricante de electrodomésticos, utiliza su Data Warehouse para dar seguimiento a sus clientes y a sus aparatos instalados. Las mayores aplicaciones del sistema son para marketing, ventas, mantenimiento, garantía y diseño de productos. Permite mantener un

inventario de partes más ajustado y mejorar las condiciones de negociación con sus proveedores. Si, por ejemplo, un determinado motor se identifica como poseedor de una tasa de falla superior, Whirlpool puede utilizar información para hacer renegociaciones de garantía con sus proveedores.

Una anécdota bastante conocida ocurrió durante el verano de 1993 los ingenieros de Whirlpool detectaron una tasa de falla muy alta en una manguera de conexión en una serie de lavadoras. A partir de allí se detuvo la producción, se identificaron los clientes y se enviaron técnicos a reemplazar la parte defectuosa antes de que entrara en falla. Esto no solo tuvo un impacto muy importante en satisfacción de clientes sino que se redujeron costos de garantía por reemplazo planificado y especialmente se evitaron costosos reclamos por daño a la propiedad debidos a pérdidas de agua.

Algunas empresas del sector que cuentan con este tipo de sistemas son: Coca Cola, Nike, Procter & Gamble, Hallmark, Maybelline, 3M, Walt Disney.

Sector Transporte y Carga de Pasajeros.

El Data Warehouse se utiliza para almacenar y acceder a meses o años de datos de clientes y sistemas de reservas para realizar actividades de marketing, planeación de capacidad, monitoreo de ganancias, proyecciones, análisis de ventas, costos, programas de calidad y servicio a clientes.

Las empresas de transporte de cargas llevan datos históricos de años, de millones de cargamentos, capacidades, tiempos de entrega, costos, ventas, márgenes, equipamiento, etc.

Las aerolíneas han logrado implementar programas de viajeros frecuentes, compartir información con los fabricantes de naves, para la administración del transporte de cargas, para compras y administración de inventarios, etc. Hacen un seguimiento de partes de repuesto, cumplimiento con las regulaciones aeronáuticas, desempeño de los proveedores, seguimiento de

equipaje, historia de reservas, ventas y devoluciones de tickets, reservas telefónicas, desempeño de agencias de viajes, estadísticas de vuelo, contratos de mantenimiento, etc.

Algunas de las empresas que cuentan con Data Warehouse de gran magnitud son: Union Pacifica, Delta, Lufthansa, British Airways, American Airlines, Canadian Airlines.

Sector Telecomunicaciones

Estas empresas utilizan sus Data Warehouses para operar en un mercado crecientemente, global que, a su vez, atraviesa fuertes cambios tecnológicos. Se almacenan datos de millones de clientes: Sus circuitos, facturas mensuales, duración de llamadas, servicios utilizados, equipamiento vendido, configuraciones de redes, etc. así como información de facturación, utilidades y costos con propósito de marketing, contabilidad, reportes gubernamentales, inventarios, compras y administración de redes.

Muchas otras industrias y sectores utilizan actualmente o comienzan a implementar un proyecto de Data Warehouse; entidades gubernamentales, empresas de servicios públicos, de entretenimiento, editoriales, fabricantes de autos, empresas de petróleo y gas, laboratorios, etc.

6.1.1 Niveles de información

Es de gran utilidad clasificar la información por niveles y a los usuarios por tipos, para entender mejor las necesidades de éstos y atenderlas mejor de acuerdo al tipo de actividad que desarrollan²⁶.

²⁶ Servicio de Administración Tributaria. Administración de Transformación y Consistencia en Base de Datos. Documentación interna.



Figura 20. Niveles de información de la organización y proporción de gente en cada una.

La Información Estratégica es la información calificada, sintetizada y consistente que facilita la toma de decisiones a la alta dirección en apoyo del logro de los objetivos de la organización.

Las características de la información que se encuentra en este nivel son las siguientes:

- Información significativa y valiosa para la toma de decisiones.
- Alta sumarización.
- Totales Sumarizados Históricos.
- Permite la comparación por periodos de interés.
- Objetivos a lograr numéricamente hablando por período definido.

La Información Táctica permite realizar análisis en el nivel medio de la organización para mantener las actividades dentro de los parámetros planeados, a través de resultados cuantitativos y comparables, que permiten medir los objetivos de cada área y mantener consistencia y congruencia en todos los niveles de la institución.

La Información Operativa permite mantener la continuidad de la operación mediante aplicaciones intuitivas que proporcionan el nivel de seguridad requerido y que automatizan los procesos correctos en las dimensiones necesarias.

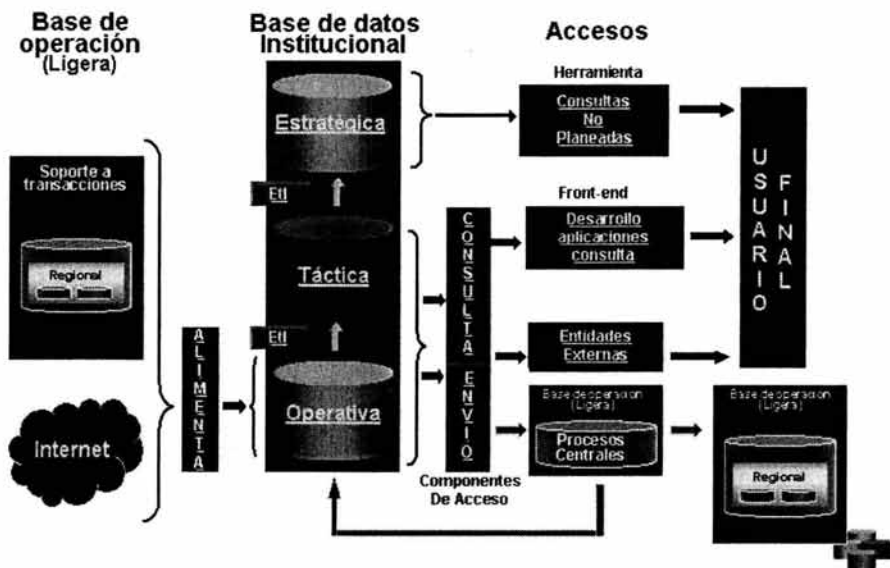


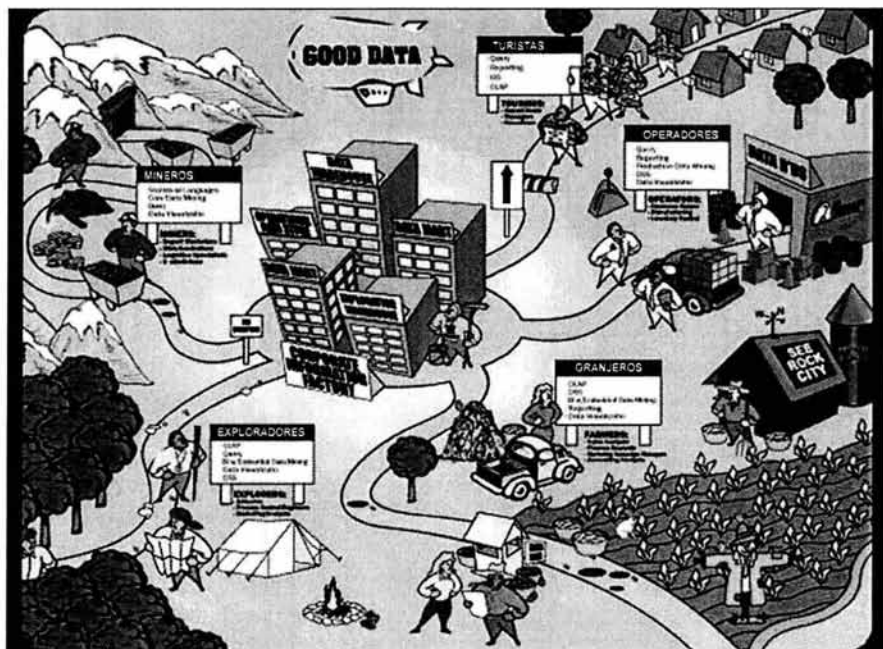
Figura 21 Después de integrarse al Data Warehouse, la información se entrega al usuario final.

6.1.2 Usuarios dentro de cada nivel de información.

Es importante identificar a los usuarios y el nivel de información que necesitan conocer, hay usuarios que solo necesitan información puntual, otros por el contrario necesitan saber la tendencia de las métricas más importantes de la organización.

De acuerdo a Bill Inmon se tiene el Universo de usuarios con lo siguientes elementos:

- Turistas.
- Granjeros.
- Exploradores.
- Mineros.
- Operadores.



²⁷ Figura 22 Universo de usuarios de la información.

Identificar los niveles de información y los usuarios que corresponden a cada uno de ellos es parte del camino para dar la respuesta acertada en el momento preciso empleando la información correcta

²⁷ Fuente: <http://www.billinmon.com>

Los sistemas de información tienen que estar orientados hacia objetivos estratégicos, para poder comunicarlos y distribuirlos consistentemente a todos los niveles de la organización.

Usuarios en la capa estratégica.



Figura 23 Turistas: usuarios de la capa estratégica.

TURISTAS

- Están pendientes de la información y del conocimiento que se produce por el negocio.
- Usualmente cubren de manera global la información pero con poca profundidad.
- Ven al mundo en términos de negocio.
- Tienen la perspectiva global del negocio.
- Necesitan información fundamental para dirigir las decisiones encaminadas a lograr el buen funcionamiento de la institución.



Figura 24 Granjeros: usuarios de la capa estratégica.

- **GRANJEROS**

- Monitorean el efecto de las decisiones en las reglas de negocio implementando indicadores de medición.
- Retroalimentan a los usuarios tácticos en relación a la efectividad de sus predicciones.
- Ven al mundo en términos de los aspectos importantes de la organización (tiempo, recaudación, geografía) y de métricas (utilización, beneficio, costos).
- Los granjeros saben que datos necesitan, como y cuando quieren que se les muestren y en que medio.
- Realizan esfuerzos para obtener información sólida y estratégica para su análisis, ya que entienden el valor de datos integrados y valoran la limpieza y calidad en ellos.

Usuarios en la capa táctica.

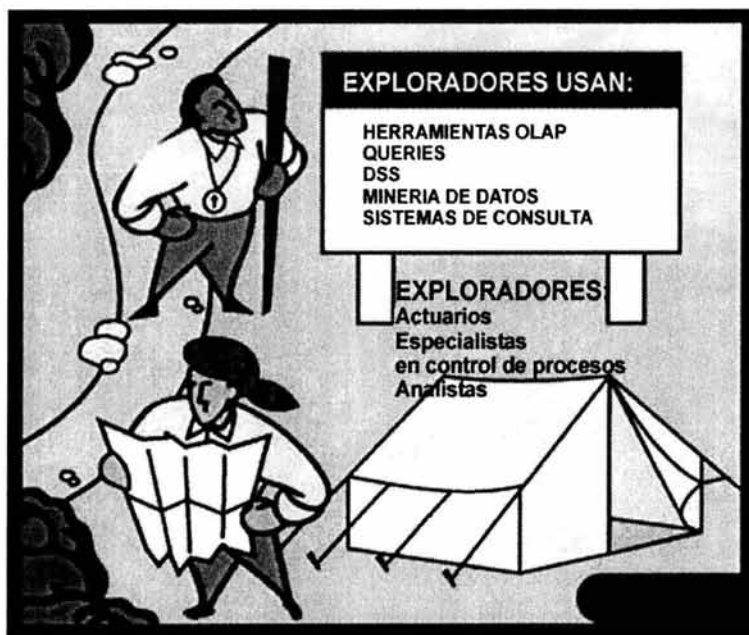


Figura 25 Exploradores: usuarios de la capa táctica.

- **EXPLORADORES.**

Procuran entender que hace que los procesos funcionen buscando comportamientos ocultos dentro de los datos.

- Tienen un poca o nula idea de lo que esperan encontrar antes de ejecutar consultas.
- Tratan de predecir el futuro, basados en eventos pasados ya que conocen el contenido de los datos de la organización a fondo.
- Ven al mundo en función a los datos y las relaciones entre ellos.
- Intentan encontrar relaciones entre porciones de datos y eventos, muchas veces basadas en consultas lanzadas al azar.
- Los exploradores frecuentemente equivocan sus hipótesis, pero cuando aciertan el beneficio puede ser muy grande para la institución.



Figura 26 : Mineros: usuarios de la capa táctica.

- **MINEROS.**
Buscan en grandes volúmenes de datos a detalle la confirmación de hipótesis o la búsqueda de patrones de comportamiento.
- Tienen una idea preconcebida de lo que esperan antes de ejecutar la búsqueda.
- Operan sobre una base de datos planos y desnormalizados que se ha adecuado para su análisis.
- Demuestran un patrón de comportamiento de los datos razonablemente predecible.
- Los mineros buscan a detalle de manera minuciosa.
- Necesitan tantos datos históricos como sea posible.
- Buscan la “pepita de oro” que dará el giro a la organización.

Usuarios en la capa operativa.



Figura 27 Usuarios de la capa operativa

- **OPERADORES**

Utilizan la inteligencia derivada de los Exploradores y Granjeros para implementar las reglas de negocio.

- Funcionan como fuente primaria de alimentación a los procesos de información. Proveen los insumos de información en términos de disponibilidad, actualización de datos y ejecución de consultas.
- Ven al mundo en términos de procesos.
- Los operadores necesitan información fresca, a detalle, día a día para la realización de su trabajo.
- Necesitan tiempos de respuesta óptimos en transacción y desempeño de las herramientas que utilizan.

Front-End de un Data Warehouse:

Generalmente el usuario conoce poco sobre el funcionamiento y procesos del Data Warehouse de la organización, por el contrario conoce las aplicaciones o vistas de la organización generadas a partir de él. El front-end generalmente toma alguna de las siguientes formas:

- Generación de reportes predefinidos
- Reportes Ad-hoc
- Generación de gráficas.
- Análisis multidimensional.
- Minería de datos



Figura: Productos obtenidos del Data Warehouse hacia el usuario final.

Éstos se implementan mediante sistemas de consulta y análisis de la Información y distribución.

6.2 Herramientas de Reporteo:

6.2.1 Reportes

Los reportes en línea generados a partir de la información del Data Warehouse, ayudan a los usuarios a entender el desempeño de la

organización. Les ayuda a tomar acciones inmediatas sobre la información más reciente. Crea un contexto común para los departamentos y niveles existentes, además de asegurar que se vea una sola versión de la verdad.

Los informes son un componente esencial para la explotación del Data Warehouse, ya que suelen ser sencillos de manejar y aprovechar por los usuarios, los apoya a alcanzar y cubrir sus metas más inmediatas.

<i>Customer Ranking by Net Profit Margins</i>						Date: 14/02/03
Historical Data:				Count	Percent	
Average Product Margin: 51%				Platinum:	81	59%
Average Net Margin: 36%				Gold:	45	33%
Average Discount: 15%				Silver:	11	8%
					138	
Silver Customers: Where our Net Profit Margins are higher than 35%						
Platinum Customers: Where our Net Profit Margins are Between 25% and 35%						
Gold Customers: Where our Net Profit Margins are lower than 25%						
Order Date	Customer Name	Discount %	Product % Margin	Net Profit Margins	Ranking	
1996	New Wave Wilderness 1	1	67	65.66	Platinum	
	Over the Top Cycles 3 AB	14	67	52.66	Platinum	
	OutBack Ply 1	11	50	39.73	Platinum	
	Ultra Sports 4	13	52	36.46	Platinum	
	Key Mart 4	10	46	37.62	Platinum	
	Ultra Sports 3	10	47	37.30	Platinum	
	ActMUp Fitness 4	10	47	37.22	Platinum	
	OutBack Ply 3	10	47	36.97	Platinum	

Tabla 6. Ejemplo de reporte tabular

Un ejemplo de reporte es: ¿cuántas operaciones se realizaron la semana pasada por día, cuántas se realizaron ayer?

6.2.2 Consultas No Planeadas

Son modelos que permiten a los usuarios ejecutivos realizar directamente (sin depender de expertos en sistemas) todas sus consultas y reportes sobre información detallada de modelos que se encuentran en las bases de datos y en el Data Warehouse provistas para tal efecto. Su versatilidad permite la generación de reportes para ser distribuidos a través de una red intranet, mediante el uso de herramientas especializadas.

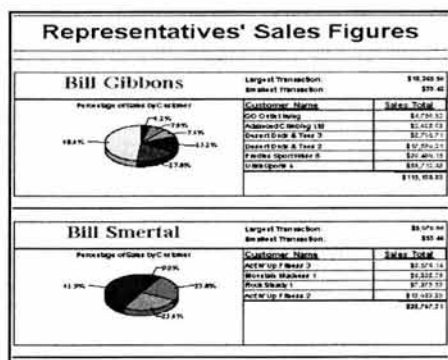


Figura 28 Reporte con gráficas, baja a mediana sumarización.

6.3 Herramientas OLAP: Molap, rolap, holap

Análisis Multidimensional

Hechos, dimensiones, métricas.

Antes de definir lo que es OLAP y en general el modelo multidimensional, conviene aclarar lo que son los hechos, las dimensiones y las métricas. Un hecho es una actividad que es objeto de análisis dentro de una organización. Las dimensiones son los aspectos que caracterizan al hecho, también son los aspectos o sujetos importantes dentro de la organización, las métricas son la parte cuantificable del hecho.

Ejemplo:

En alguna aduana se efectúa la importación²⁸ de mercancía.

Hecho: La importación.

Dimensiones:

Aduana dónde ocurre el hecho: Nuevo Laredo.

Agente Aduanal: 3125.

Contribuyente: General Motors.

²⁸ Ejemplo ficticio.

Tipo de operación: Definitiva.

País Origen: Estados Unidos de América.

Fracción Arancelaria: 40111001.

Métricas:

Cantidad de mercancía.

Valor en aduana.

Valor comercial.

Total de impuestos.

La diferencia entre OLTP y OLAP es que el primero se concentra en las transacciones, mientras el segundo usa los datos de manera analítica.

El número de solicitudes concurrentes no es tan grande como el de un OLTP. El análisis frecuentemente no comienza con una meta predefinida sino que navega en múltiples direcciones en el espacio de datos dependiendo de los resultados que se van obteniendo. El espacio de datos se estructura por criterios como son el tiempo, el cliente, el producto, servicio, región, etc. Estos criterios se conocen como dimensiones, las dimensiones constituyen un espacio de datos multidimensional o cubo multidimensional. Además las dimensiones se estructuran por jerarquías que forman árboles que a su vez pueden recorrerse hacia arriba o abajo (drill up, drill down). Métricas como las ventas pueden agregarse a lo largo de estas jerarquías o a niveles más generales.

El análisis OLAP (On Line Analytical Processing), surge por la necesidad de consolidar información, verla y analizarla desde diferentes puntos de vista, a lo largo de momentos del tiempo específicos. En contraste con el Data Warehouse, generalmente basado en una base de datos relacional, el análisis OLAP brinda vistas de datos multidimensionales de información sumariada, provee acceso rápido a información estratégica y permite

realizar análisis con ella. La información se ve reflejada en términos de la organización y no de la forma en que se almacena en los sistemas de base de datos. Aun así la tecnología OLAP y el Data Warehouse se complementan mutuamente, el Data Warehouse almacena y administra la información y la tecnología OLAP la convierte en información estratégica, fácil de consultar²⁹.

Se pueden considerar los sistemas OLAP como pertenecientes a los sistemas de información, utilizados para proporcionar al nivel estratégico información útil para la toma de decisiones.

En un modelo de datos OLAP, la información es vista como cubos, los cuales consisten de categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo de datos multidimensional simplifica a los usuarios formular consultas complejas, arreglar datos en un reporte, cambiar de datos resumidos a datos detallados y filtrar o rebanar los datos en subconjuntos significativos.

Aspecto	OLTP Base de datos Transaccional	OLAP Data Warehouse
usuarios	Operadores	Analíticos.
Función	Operaciones diarias. Procesamiento de transacciones	Soporte de decisiones. Procesamiento analítico.
Diseño	Orientado a las aplicaciones	Orientado al usuario.
Datos	Actuales, atómicos,	Históricos, resumidos, multidimensionales.
Uso	Rutinario.	Sobre demanda.
Acceso	Lectura/escritura. Transacciones simples.	Lectura Consultas complejas.
Necesidades	Rapidez de respuesta, consistencia en los datos	Optimización de consultas. Datos organizados.

²⁹ <http://www.olapcouncil.org/research/whtpapy.htm>

Tabla 7 Comparativa entre un sistema Transaccional OLTP y OLAP.

6.3.1 Las reglas del Dr. E. F. Codd

En 1999 El Dr. E. F. Codd & Associated publicó un artículo llamado "Providing OLAP (On-line Analytical Processing) to user-analysts: An IT Mandate". Para ese momento el Dr. Codd era bien conocido como un respetado investigador en base de datos además de contar con el crédito de ser el inventor del modelo de base de datos relacional. Sin embargo las reglas OLAP fueron controversiales por considerarse ser patrocinadas por Arbor Software (hoy Hyperion), más que basadas en matemáticas. Aun así han servido de fundamento para el desarrollo de numerosos productos. La investigación de estos criterios llevó a la elaboración de 12³⁰ reglas³¹ de funcionamiento que son las siguientes:

1. Vista conceptual multidimensional de los datos.

La manera de analizar el universo de información del usuario es multidimensional por naturaleza, por lo tanto, los modelos de datos y la vista del usuario también debe serlo. Este esquema dimensional facilita el análisis intuitivo. Los modelos tradicionales requieren más tiempo y esfuerzo para obtener los mismos resultados.

2. La transparencia de cara al usuario.

El usuario no debería saber de dónde proviene la información de su herramienta OLAP, ni cómo opera internamente. Esto es crucial para mantener la productividad del usuario al tiempo que se brinda la funcionalidad y manera de trabajar multidimensional, asegurando que no hay complejidades adicionales que son innecesarias.

³⁰ An IT mandate

F. Codd, S.B. Codd, and C.T. Salley, Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate; 1993

³¹ <http://www.businessdecision.com/esp/Decouvrir/Glossaire.htm#>

3. **Accesibilidad.** El análisis debe poder realizarse sobre datos provenientes de fuentes de datos heterogéneas, ofreciendo una vista unificada, coherente y consistente de los mismos. Finalmente quien se debe ocupar de la procedencia de los datos es la herramienta y no el usuario final.
4. **Rendimiento.** Aunque aumente el número de dimensiones o el tamaño de la base de datos, el usuario no debe percibir degradaciones importantes en el tiempo de ejecución.
5. **Arquitectura cliente - servidor.** Debido a que la mayor parte del procesamiento OLAP se lleva a cabo en sistemas centralizados y se accede a éste desde computadoras personales es imperativo que la herramienta OLAP trabaje bajo esta arquitectura.
6. **Dimensiones simétricas.** Todas las dimensiones definidas en el modelo de datos deben de ser accesibles para cada uno de los datos.
7. **Manejo óptimo de matrices poco densas.** El almacenamiento físico de los datos debe ajustarse a la distribución de sus valores.
8. **Soporte multiusuario.** Frecuentemente ocurre que varios usuarios deben trabajar concurrentemente con un modelo analítico, la herramienta OLAP debe proveer concurrencia para obtener los datos, manteniendo integridad y seguridad.
9. **Consistencia.** Todos los datos almacenados o calculados en el cubo deben ser accesibles y las reglas de gestión deben aplicarse siempre. Todos los cortes del cubo deben ser visualizables.
10. **Facilidad de uso.** Navegación cómoda a través de los datos para los usuarios de manera intuitiva.
11. **Generación de reportes flexible.** Tanto filas como columnas deben poder incluir cualquier número de dimensiones en cualquier orden, mostrando para cada dimensión cualquier subconjunto de datos en cualquier orden.
12. **Numero ilimitado de dimensiones y de niveles de agregación.** Una herramienta OLAP debería permitir definir modelos multidimensionales

con sin restricciones en el número de dimensiones y éstas con un ilimitado número de niveles de agregación.

La creación de esta norma ha servido de base para la fundación en los EE.UU. del OLAP Council, asociación para la promoción de la informática dimensional, que agrupa a los principales fabricantes de software que respetan las 12 reglas enumeradas anteriormente.

Más tarde Nigel Pendse toma lo que considera adecuado de las 12 reglas y desarrolla las reglas **FASMI**³² (FAST ANALYSIS SHARED MULTIDIMENSIONAL INFORMATION), que significa Análisis Rápido de Información Multidimensional Compartida.

Fast: La herramienta debe responder a las peticiones en cerca de cinco segundos. Los análisis sencillos deben tomar aproximadamente un segundo, y los más complejos no más de veinte.

Analysis: La herramienta debe ser capaz de trabajar con vistas del negocio y brindar los análisis estadísticos que requiere el usuario, y aun mantener facilidad de uso.

Shared: La herramienta debe ser segura para mantener la confidencialidad necesaria de la información, soportar concurrencia.

Multidimensional: Es el requerimiento clave, la herramienta debe proveer una vista multidimensional de los datos, incluyendo soporte de jerarquías. La vista por dimensiones es la manera más lógica de analizar una organización.

Information: La herramienta debe ofrecer la información y los datos calculados sin importar si provienen de fuentes heterogéneas. Se refiere a la capacidad de la herramienta de saber cuánta información puede manejar y no cuantos Gigabytes le toma almacenarla.

³² <http://www.olapinfo.de/whitepapers.php>

La importancia del examen FASMI radica en que señala las metas que debe alcanzar la tecnología OLAP, y no en cómo llegar a ellas.

Beneficios de la tecnología OLAP.

Las herramientas OLAP ayudan a incrementar la productividad de administradores, desarrolladores, operativos y en general a toda la organización. La flexibilidad y facilidad que ofrecen vuelve a los usuarios autosuficientes. El personal con pocos conocimientos técnicos ya no solicita al departamento de informática ayuda para crear joins, o para obtener reportes. Los tomadores de decisiones pueden localizar más fácilmente los problemas existentes o potenciales. En pocas palabras mayor control y acceso rápido a información estratégica equivale a un proceso de toma de decisiones más efectivo.

El área de informática también se beneficia del uso de herramientas OLAP debido a que la entrega de aplicaciones y modelos a usuarios se hace más rápido, reduciendo al mismo tiempo la acumulación de solicitudes de información.

Las herramientas OLAP también ayudan a los sistemas operacionales reduciendo las consultas y tráfico de red sobre ellos.

Finalmente al proveer la habilidad de analizar mejor la información y a un mejor uso de recursos humanos, la organización debería responder mejor a los cambios.

Visualización de los datos mediante un cubo.

La vista de los datos como un cubo es una extensión natural de como la mayoría de los usuarios de negocios interactúan con los datos

Ellos ven a un problema de la organización en términos de un cierto número de componentes (dimensiones) tales como productos, tiempo, regiones,

servicios, fabricantes, o artículos. Los usuarios desean poder analizar un conjunto de números usando cualquier par de estos componentes, como así también poder intercambiarlos para lograr distintas vistas.

	A	B	C	D	E
1		Camping Equipment	Mountaineering Equipment	Personal Accessories	Golf Equipment
2	Australia	3417635	524789	892556	533782
3	Brazil	4264667	132324	1088958	490678
4	Canada	15466549	2147662	2894066	2640205
5	England	6194416	1290988	2776420	1982892
6	France	5585582	1303342	2141912	1877678
7	Germany	9827518	1851665	3362213	1699034
8	Italy	3753653	1397729	1662668	1300104
9	Japan	330018	192425	1906067	1540894
10	Mexico	1790189	163956	619489	629210
11	Spain	1501396	1257417	540099	769878
12	Switzerland	3286840	1076180	1073308	468313
13	United States	15728344	2422470	4331996	3462584

Tabla 8. Vista tabular tradicional

Sin embargo, la mayoría de los usuarios también desearía ver como se desarrollan las ventas en el tiempo. Para hacer esto, se necesitarían varias hojas de la planilla de cálculo como se muestra en la figura

	A	B	C	D	E
1		Camping Equipment	Mountaineering Equipment	Personal Accessories	Golf Equipment
2	Australia	0	0	0	0
3	Brazil	1252978	0	361188	147329
4	Canada	3632986	0	921608	778419
5	England	2046000	0	981614	613980
6	France	1802389	0	469540	507955
7	Germany	2011323	0	925965	498174
8	Italy	1190338	0	374615	213016
9	Japan	99798	0	212868	248425
10	Mexico	538489	0	123029	179843
11	Spain	521349	0	121402	162540
12	Switzerland	0	0	0	0
13	United States	5012728	0	1097998	872463

Tabla 9. Tabla con información de 1999.

	A	B	C	D	E
1		Camping Equipment	Mountaineering Equipment	Personal Accessories	Golf Equipment
2	Australia	1557708	245641	392902	248802
3	Brazil	1409641	61202	322824	179416
4	Canada	5025695	904524	889093	809170
5	England	1918367	607018	807577	673778
6	France	1730836	542810	746233	705830
7	Germany	3739396	899331	1118553	586747
8	Italy	1128110	637501	550217	519598
9	Japan	126139	96125	750798	525690
10	Mexico	523378	71400	197712	257982
11	Spain	469555	630417	196963	296040
12	Switzerland	1504930	506362	488614	256147
13	United States	4817222	1081153	1477088	1182482

Tabla 10. Tabla con información de 2000.

Las mismas celdas de datos se visualizan mediante un cubo

	1999				2000			
	Camping Equipment	Mountaineering Equipment	Personal Accessories	Golf Equipment	Camping Equipment	Mountaineering Equipment	Personal Accessories	Golf Equipment
Australia	0	0	0	0	18526	3472	4604	1122
Brazil	13744	0	4794	596	16472	1282	4554	702
Canada	28308	0	10790	2736	42612	12988	10676	2836
England	19542	0	11374	2316	20438	9024	9634	2614
France	19256	0	5586	1854	19890	8256	9448	3328
Germany	19698	0	11662	1596	40014	13532	14436	2388
Italy	11402	0	4608	756	11640	9064	6890	2128
Japan	898	0	2414	808	1186	1608	8628	2458
Mexico	5120	0	1460	520	5334	1134	2472	824
Spain	5246	0	1544	508	5108	8848	2398	1152
Switzerland	0	0	0	0	14122	7144	5742	1068
United States	35030	0	13024	2766	36008	15384	17700	3784

Tabla 11. Información de 1999 y 2000 en una sola vista.

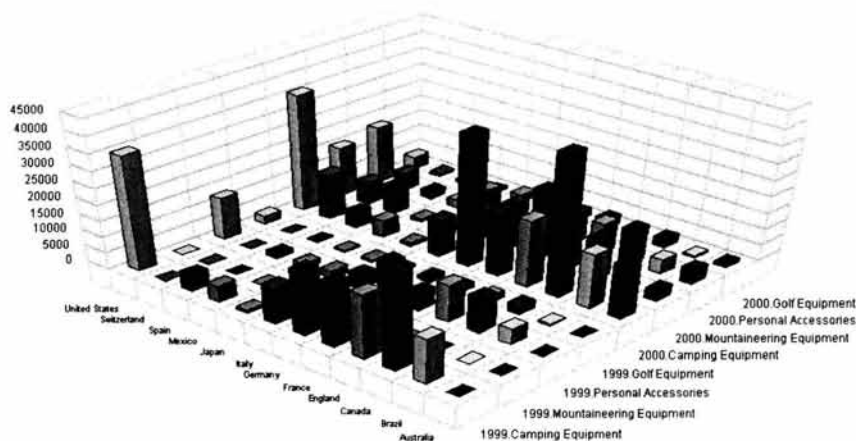


Tabla 12. La graficación resulta de gran utilidad, por ejemplo para identificar máximos.

Dado que las celdas de datos pueden ser fácilmente representadas en un cubo, se pueden tomar fracciones del mismo para responder preguntas como:

¿Cómo se venden los productos en cada región en un mes dado? Esto es equivalente a ver Producto por Región en un mes dado. Figura 25.

¿Qué regiones han mejorado las ventas de un producto dado a través del tiempo? Esto es equivalente a Región por Tiempo de un producto dado. Figura 26.

¿Cómo se venden los productos a través del tiempo en una región dada? Esto es equivalente a Producto por Tiempo en una región dada. Figura 27.

Revenue	1997/Jan			1997/Feb			1997/Mar			1997 Q1
como valores	Europe	Far East	Americas	Europe	Far East	Americas	Europe	Far East	Americas	Locations
Outdoor Products	14,325	0	19,035	35,888	5,689	4,378	20,450	31,795	34,827	166,387
Environmental Line	22,512	3,900	24,822	21,669	12,532	3,330	14,107	8,932	41,040	152,844
GO Sport Line	4,864	0	5,248	8,208	1,200	752	6,704	1,392	4,688	33,856
Products	41,701	3,900	49,105	65,765	19,421	8,460	41,261	42,119	80,555	352,287

Figura 29 Vista anidada proporcionada por un cubo OLAP.

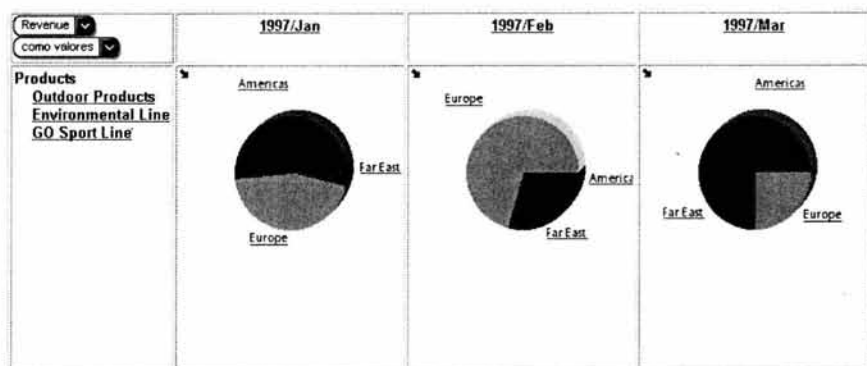


Figura 30 Gráficas proporcionadas por un cubo OLAP.

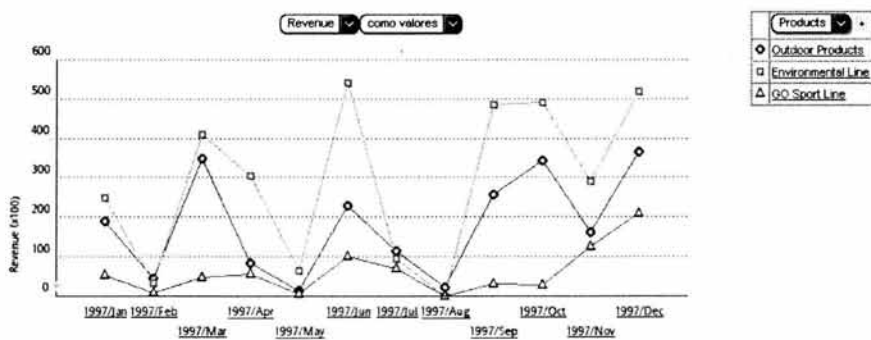


Figura 31. Gráficas proporcionadas por un cubo OLAP

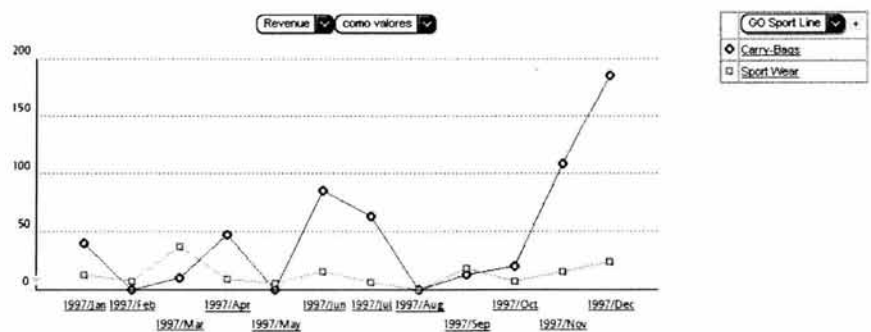


Figura 32. Gráficas proporcionadas por un cubo OLAP

6.3.2 Modelos de almacenamiento

MOLAP, ROLAP y HOLAP³³

MOLAP

En un sistema MOLAP (OLAP multidimensional) los datos se encuentran almacenados en una estructura multidimensional.

Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados o agregaciones son la base de las ganancias de desempeño de este sistema.

Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores precalculados.

ROLAP

ROLAP (OLAP Relacional) es un sistema en el cual los datos se encuentran almacenados en una base de datos relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas. Los esquemas más comunes sobre los que se trabaja son estrella o copo de nieve, aunque es posible trabajar sobre cualquier base de datos relacional

HOLAP

Un sistema HOLAP (OLAP Híbrido) mantiene los registros detallados en la base de datos relacional, mientras que los datos resumidos o agregados se almacenan en una base de datos multidimensional separada. Este método de almacenamiento es una combinación de los dos anteriores e intenta rescatar lo mejor de cada uno.

³³ Holger Frietsch, Rolap, Molap, Holap: How To Determine Which Technology Is Appropriate.

Diferencias

El primer enfoque corresponde a MOLAP. Con este método, los datos son precalculados y luego son almacenados en cubos de datos multidimensionales. El resultado se traduce en una mejor performance en los tiempos de respuesta debido a que los datos se encuentran disponibles sin necesidad de calcularlos en cada nueva consulta.

La desventaja de este enfoque se debe a que almacenar datos multidimensionalmente ocupa mucho más espacio que con ROLAP.

El segundo enfoque es ROLAP donde los datos son accedidos directamente del Data Warehouse (u otro tipo de fuente relacional) y no son almacenados por separado. Estos datos se calculan en tiempo de consulta (al vuelo). Por lo tanto el tiempo de respuesta sería mayor pero sin usar enormes cantidades de almacenamiento en disco.

Las implementaciones MOLAP normalmente se desempeñan mejor que la tecnología ROLAP, pero tienen problemas de escalabilidad.

Las implementaciones ROLAP son más escalables y son frecuentemente atractivas a los clientes debido a que aprovechan las inversiones en tecnologías de bases de datos relacionales preexistentes

La mejor solución probablemente esté entre los dos extremos.

En muchos casos se utilizan HOLAP que resultan de una combinación entre ROLAP y MOLAP. HOLAP mantiene los volúmenes de datos más grandes en la base de datos relacional y las agregaciones en una base de datos MOLAP separada, logrando con esto un balance entre tiempo y espacio.

6.4 Minería de datos

*Estamos en un mar de datos y morimos de sed de información*³⁴.

El volumen de información en las bases de datos de las organizaciones ha crecido de una manera sin precedente, la mayor parte de esta representa las transacciones registradas. Una sola organización multinacional genera en una semana más información de la que cualquier persona podría leer en toda su vida. Debido al gran volumen de datos, su proceso de análisis, ya no puede ser manual, se necesita una herramienta que analice los datos de la manera más automática posible. Es en este punto que la minería de datos, como una etapa del proceso de extracción de conocimiento a partir de datos se presenta y se diferencia de las herramientas vistas previamente ya que a diferencia de éstas, que transforman y facilitan el acceso a la información para que el usuario la analice, la minería de datos analiza los datos, descubre hechos y reglas desconocidas, relaciones y patrones entre ellos, patrones, tendencias y regularidades que describen y ayudan a una mejor comprensión de los datos.

La información histórica no sólo es el registro de cada movimiento o transacción de la organización, es la base para predecir información ya que es la referencia más cercana del comportamiento en la realidad. Buena parte de las decisiones de empresas, organizaciones e instituciones tienen sustento en información de experiencias pasadas provenientes de diversas fuentes, sin embargo los volúmenes de datos a analizar actualmente rebasan por mucho a la capacidad de procesamiento humano.

Hoy en día el almacenamiento de la información es más sencillo y barato, debido a esto se guardan datos de funcionamiento de los procesos, sistemas de venta, de clientes, etc. Al almacenar esta información generalmente se

³⁴ Director de DSS México, 2003.

tiene la intención de analizarlos posteriormente, sin embargo cuando se llega a la etapa del análisis, se suele realizar de manera superficial y guiado por los resultados que se espera encontrar. Sin embargo, hay información que no puede ser encontrada con los métodos tradicionales, la minería de datos ayuda al análisis detectando relaciones ocultas entre los datos.

Una manera de definir la minería de datos es como el proceso de extracción de información implícita no trivial, previamente desconocida a partir de los datos³⁵. Para conseguirlo hace uso de diferentes tecnologías que resuelven problemas típicos de agrupamiento automático, clasificación, asociación de atributos y detección de patrones secuenciales.

Es importante mencionar que la minería de datos, es una fase dentro del proceso de Descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases o KDD), sin embargo en la mayor parte de las fuentes se menciona a la minería como el proceso completo y no como la fase de extracción de conocimiento.

Por otro lado, en ocasiones se confunde a la minería de datos con el análisis estadístico, la diferencia es que en la estadística es necesario formular hipótesis de las relaciones posibles entre datos y cuantificarlas, la minería de datos en cambio, opera de manera muy distinta, ya que por sí misma busca relaciones entre parejas de datos, la técnica empleada deduce las relaciones existentes, realizando un acto de descubrimiento de conocimiento que tal vez ni siquiera se sospechaba.

³⁵ Daedalus, Minería de datos: Conceptos y objetivos-

La clave que separa la minería de datos respecto de otras técnicas es que el análisis que efectúa es exploratorio, no corroborativo. Ya que trata de encontrar conocimiento nuevo, no de confirmar o desmentir supuestos, encuentra hipótesis y comprueba su validez. Con las otras técnicas hace falta

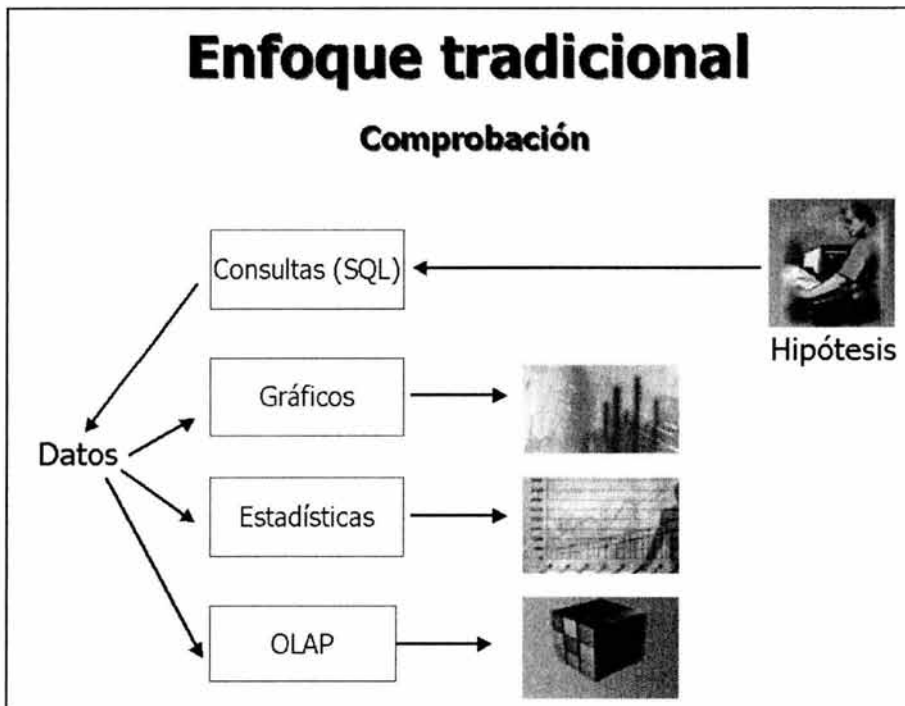
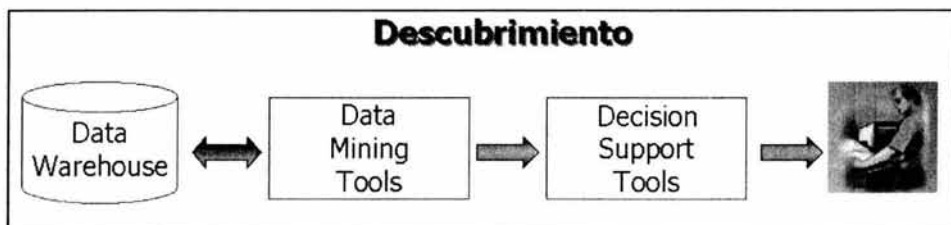


Figura 33 Enfoque dirigido tradicional.

tener una idea de lo que se busca, la información que se obtiene está condicionada al supuesto con que se plantea el problema.



El proceso general de extracción de conocimiento a partir de datos consta de las siguientes fases:

- Preparación de datos
 - Selección
 - Limpieza

Figura 34 Enfoque automático.

- Transformación
- Minería de datos
- Evaluación
- Difusión
- Uso de modelos.

El proceso de minado de datos

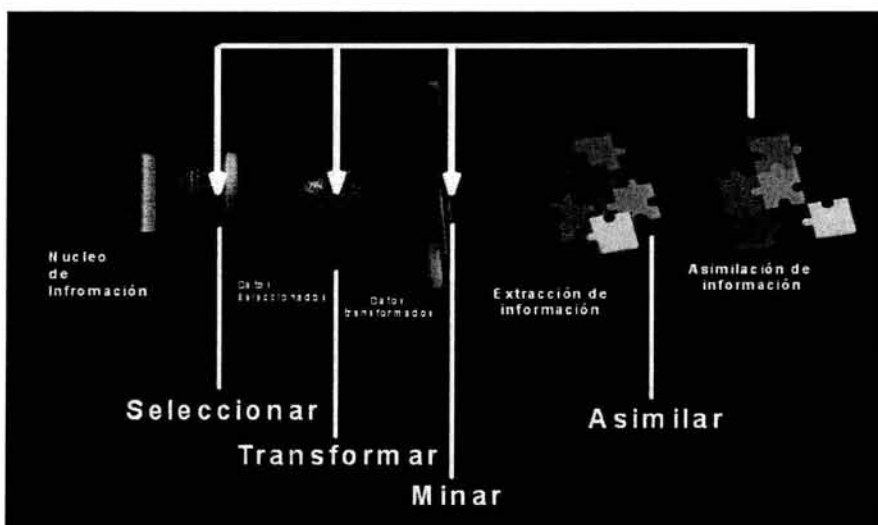


Figura 35. Proceso de minería de datos. Administración de Transformación y Consistencia en Base de Datos, SAT. 2002.

Distribución del esfuerzo

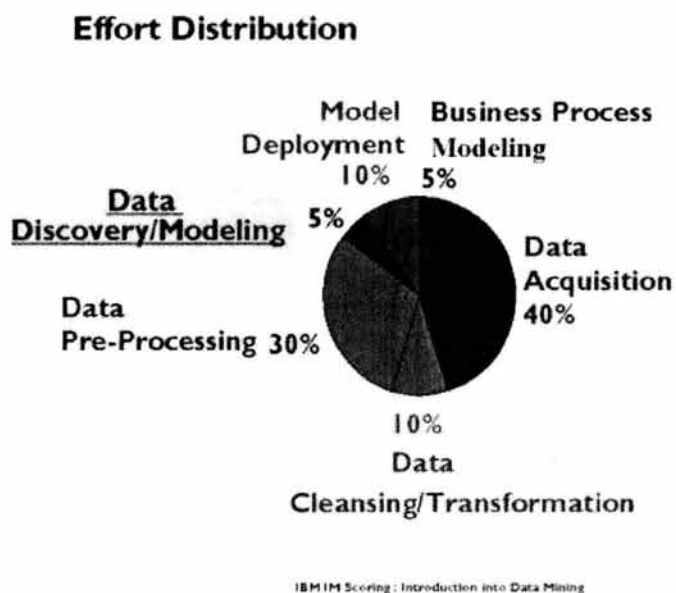


Figura 36. Gráfica sobre la distribución del esfuerzo en un modelo de Minería de datos. IBM.

Este proceso emplea técnicas como son:

- Árboles de decisión.
- Regresión lineal.
- Redes neuronales.
- Técnicas bayesianas.
- Maquinas de soporte vectorial.

Provenientes los campos de: aprendizaje automática e Inteligencia Artificial, Estadística, Bases de datos y aborda las problemáticas de clasificación, categorización, estimación / regresión, agrupamiento.

Para hacer minería de datos no es necesario contar con un Data Warehouse, basta con un archivo de datos, sin embargo al estar presente éste, se tienen fuertes ventajas por contar con grandes volúmenes de datos, por haber integrado y consolidado previamente fuentes de datos heterogéneas. El proceso de adquisición de datos se simplifica y se hace más efectivo.

Minería de datos vs OLAP.

Las herramientas OLAP permiten navegar a través de los datos y analizarlos dinámicamente desde una perspectiva multidimensional. Sin embargo el análisis OLAP depende de un usuario que plantee una consulta o supuesto, es él quien dirige el análisis y por tanto lo limita por sus ideas a priori.

La minería de datos no sustituye al análisis OLAP, lo complementa efectuando análisis más avanzado sobre los datos, obteniendo más información a partir de ellos. La minería de datos descubre nuevas hipótesis y relaciones, no se limita por la perspectiva del usuario sobre la información.

El análisis que realizan las herramientas OLAP es deductivo, ya que es dirigido por el usuario, parte de un supuesto y los datos se analizan para resolver o comprobar el supuesto. La minería por su parte permite razonar de forma inductiva a partir de los datos para llegar a una hipótesis general que modele el problema.

Otra diferencia fundamental es que las herramientas OLAP usualmente trabajan con datos resumizados y brindan una visión global de la organización, en contraste la minería de datos trabaja con datos individuales, descubriendo las regularidades y patrones que se presentan y más tarde generalizando a partir de ellos.

6.4.1 Relación de la minería de datos con otras disciplinas

La minería se nutre del desarrollo de diferentes disciplinas³⁶:

- Estadística.
- Sistemas de información.
- Bases de datos.
- Aprendizaje automático.
- Inteligencia artificial.
- Visualización de datos.
- Computación paralela.
- Computación distribuida.

La minería como se mencionó previamente sólo es una etapa de KDD, en la que se integran los métodos estadísticos y de aprendizaje para obtener hipótesis de patrones y modelos.

En relación con la tecnología OLAP y los sistemas de reporte la minería destaca por tener objetivos más ambiciosos.

La minería, por ejemplo podría responder preguntas como: ¿cuáles son las principales variables involucradas en la venta de un producto en particular?

Otras herramientas	Minería de datos
El usuario construye el modelo o hipótesis. Comienza con una suposición, pregunta o intuición y explora los datos que soporten su modelo.	Además de que el usuario mantiene la posibilidad de proponer modelos, la herramienta encuentra y sugiere modelos.

Tabla 13 Comparativa entre los modos de operación de la Minería y de otras herramientas.

³⁶ www.crm-forum.com

6.4.2 Ventajas de la minería de datos.

A continuación se encuentran algunas de las ventajas del uso de la minería de datos en las Organizaciones.

- La posibilidad de encontrar un buen modelo aumenta ya que se prueban múltiples modelos propuestos automáticamente.
- Con la ayuda de la herramienta de minería no se necesita un especialista en cada técnica y campo para poder construir modelos.
- Contribuye al proceso de toma de decisiones tácticas y estratégicas proporcionando información clave a pesar de trabajar con altos volúmenes de datos.
- Permite a los usuarios asignar prioridades a la toma de decisiones y a la de acciones al identificar los factores de mayor importancia en un objetivo dado.
- Ofrece modelos descriptivos de manera implícita que permiten apreciar el panorama general del estado actual de los datos.
- Ofrece modelos predictivos, ya que construye modelos probabilísticos que ayudan en la realización de actividades de investigación. Por ejemplo:
 - investigación de operaciones de importación / exportación sensibles.
 - Otorgamiento de créditos a personas con las mejores posibilidades de cubrirlos.

2D Monday, January 3, 2000 The Journal News ■■■ RK

DILBERT By Scott Adams



Figura 37 Caricatura de Scott Adams bromeando un poco sobre los alcances de la minería.

6.4.3 Áreas de aplicación:

KDD para toma de decisiones.

Predicción; A partir de datos históricos se pueden elaborar modelos que estimen la evolución de una variable en el futuro.

Detección de fraudes; Estos modelos descubren fraudes potenciales, detectan comportamientos anormales.

Reducción de riesgos; Se pueden elaborar modelos de evaluación de riesgos basados en casos reales, resultan de gran ayuda cuando la cantidad de datos es excesiva para su procesamiento manual.

La mercadotecnia es uno de los campos de aplicación más conocidos.

Se puede incrementar el índice de respuesta a campañas publicitarias si se enfoca al segmento de mercado objetivo adecuado. En este rubro la minería de datos ayuda a identificar quién comprará qué, cuándo y dónde.

Comercio:

- Identificación de patrones de compra de clientes.

- Identificar relaciones entre los clientes y las características demográficas.
- Predicción de impacto de campañas de correo.
- Análisis de compras.

Bancos:

- Detección de fraude.
- Identificación de clientes.
- Detección de clientes con posibilidades de cambiar de banco.

Seguros

- Análisis de cobertura de servicios médicos solicitados conjuntamente.
- Predicción de los clientes potenciales para adquirir seguros.
- Identificación de patrones de comportamiento para clientes de alto riesgo.
- Identificar prácticas fraudulentas.

Transporte:

- Analizar patrones de carga.
- Identificar pedidos por zona geográfica.

Medicina:

- Identificación de variables involucradas y su peso en las terapias médicas satisfactorias en el tratamiento de enfermedades.
- Asociación de síntomas y clasificación de patologías.
- Estudio de factores de riesgo tales como genéticos, hábitos alimenticios, consumo de alcohol, etc. en el desarrollo de distintas patologías.
- Clasificación de pacientes según características en común para aprovechar mejor los recursos humanos y materiales.

- Estudio epidemiológicos, análisis de rendimientos de campañas de información, preventivas, etc.

En la industria

- Predicción de fallos.
- Evaluación de calidad.
- Estimación de compuestos con mezclas óptimas.

6.4.4. Modelos predictivos y modelos descriptivos

Por su uso los modelos resultantes de la minería de datos se dividen en predictivos y en descriptivos.

Los modelos predictivos nos ayudan a estimar el posible resultado de un evento, basado en datos históricos.

Por ejemplo puede ayudar en la reducción de riesgos al evaluar el otorgamiento de créditos, cuáles son las cifras de ventas para la siguiente temporada, identificar clientes con actividades anómalas con posibilidades de estar haciendo fraude, etc.

Ejemplo de Modelo Predictivo³⁷:

- Se desea saber si se va a jugar Tenis en la tarde.
- Se han recogido datos de experiencias anteriores:

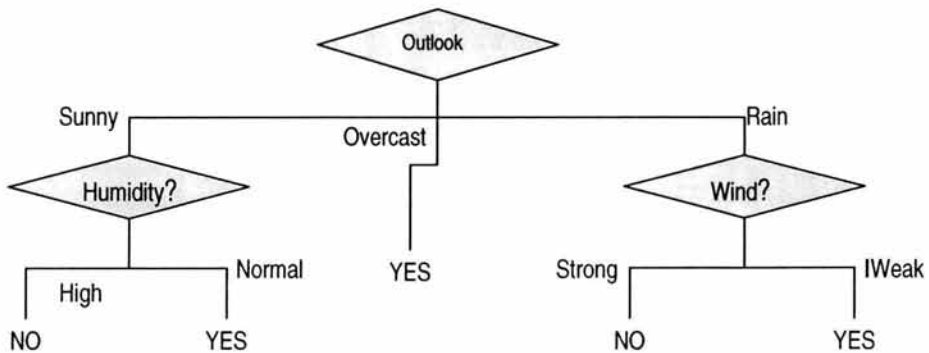
Example	Sky	Temperature	Humidity	Wind	PlayTennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes

³⁷ Silvia M Aranguren, Silvia L. Muzachiodi, Implicancias del Data Mining, 2000.

Example	Sky	Temperature	Humidity	Wind	PlayTennis
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Tabla 14 Información sobre las condiciones climáticas y la decisión de jugar.

- Pasamos estos datos a un algoritmo de aprendizaje de árboles de decisión, señalando el atributo "PlayTennis" como la clase (output).
- El resultado del algoritmo es el siguiente modelo:



- Ahora se puede utilizar este modelo para predecir si esta tarde se jugará o no al tenis.

Se toma el siguiente caso:

Outlook = sunny
 Temperature = hot
 Humidity = high
 Wind = strong

La respuesta es que no se jugará. El modelo descriptivo proporciona información sobre las relaciones entre los datos y cómo se presentan.

Ejemplos:

- Los clientes que compran pañales suelen comprar cerveza.
- El tabaco y el alcohol son los factores más importantes en la enfermedad Y.
- Los clientes sin televisión y con bicicleta tienen características muy diferenciadas del resto.

Ejemplo de Modelo Descriptivo:

Se desea categorizar a los empleados.

- Se tiene la siguiente tabla con sus datos:

#Ej	Sueldo	Casado	Coche	Hijos	Alq/Prop	Sindic.	Bajas/Año	Antigüedad	Sexo
1	10000	Sí	No	0	Alquiler	No	7	15	H
2	20000	No	Sí	1	Alquiler	Sí	3	3	M
3	15000	Sí	Sí	2	Prop	Sí	5	10	H
4	30000	Sí	Sí	1	Alquiler	No	15	7	M
5	10000	Sí	Sí	0	Prop	Sí	1	6	H
6	40000	No	Sí	0	Alquiler	Sí	3	16	M
7	25000	No	No	0	Alquiler	Sí	0	8	H
8	20000	No	Sí	0	Prop	Sí	2	6	M
9	20000	Sí	Sí	3	Prop	No	7	5	H
10	30000	Sí	Sí	2	Prop	No	1	20	H
11	50000	No	No	0	Alquiler	No	2	12	M
12	8000	Sí	Sí	2	Prop	No	3	1	H
13	20000	No	No	0	Alquiler	No	27	5	M
14	10000	No	Sí	0	Alquiler	Sí	0	7	H
15	8000	No	Sí	0	Alquiler	No	3	2	H

Tabla 15. Datos de los empleados,

Se pasan estos datos a un algoritmo descriptivo (clustering K-means).

- Se crean tres clusters, con la siguiente descripción:

Cluster 1		Cluster 2	Cluster 3
Tamaño	5	4	6
Sueldo	22,600	22,500	18,833
Casado	0.8 No	1 No	1 Si
	0.2 Si		
Coche	0.8 No	1 Si	1 Si
	0.2 Si		
Hijos	0	0	2
Vivienda	1	0.75 Alquiler	0.17 Alquiler
	Alquiler		
		0.25 Propia	0.83 Propia
Sindicalizado	0.8 No	1 Si	0.67 No
	0.2 Si		0.33 Si
Bajas/Año	8	2	5
Antigüedad	8	8	8
Sexo	0.6 H	0.25 H	0.83 H
	0.4 M	0.75 M	0.17 M

Tabla 16. Grupos obtenidos a partir de los datos.

De estos grupos se pueden destacar las siguientes características:

GRUPO 1:

- Personal sin hijos.
- Alquiler su vivienda.
- Poco sindicalizados.
- Muchas bajas.

GRUPO 2:

- Personal sin hijos
- Con coche.

- Muy sindicalizados
- Pocas bajas.
- Normalmente de alquiler .
- La mayoría son mujeres.

GRUPO 3:

- Con hijos.
- Casados.
- Con coche.
- Propietarios de vivienda.
- Poco sindicalizados.
- Predominantemente hombres

6.4.5 Tipos de técnicas:

Asociaciones: Una asociación entre dos atributos ocurre cuando la frecuencia de que se den dos valores determinados de cada uno conjuntamente es relativamente alta.

Ejemplo, en un supermercado se analiza si los pañales y los potitos de bebé se compran conjuntamente.

Dependencias: Una dependencia funcional (aproximada o absoluta) es un patrón en el que se establece que uno o más atributos determinan el valor de otro. Ojo! Existen muchas dependencias nada interesantes (causalidades inversas).

Ejemplo: que un paciente haya sido ingresado en maternidad determina su sexo.

La búsqueda de asociaciones y dependencias se conoce a veces como análisis exploratorio.

Clasificación: Una clasificación se puede ver como el esclarecimiento de una dependencia, en la que el atributo dependiente puede tomar un valor entre varias clases, ya conocidas.

Ejemplo: se sabe (por un estudio de dependencias) que los atributos edad, número de miopías y astigmatismo han determinado los pacientes para los que su operación de cirugía ocular ha sido satisfactoria.

Podemos intentar determinar las reglas exactas que clasifican un caso como positivo o negativo a partir de esos atributos.

Agrupamiento / Segmentación: El agrupamiento (o clustering) es la detección de grupos de individuos. Se diferencia de la clasificación en el que no se conocen ni las clases ni su número (aprendizaje no supervisado), con lo que el objetivo es determinar grupos o racimos (clusters) diferenciados del resto.

Tendencias/Regresión: El objetivo es predecir los valores de una variable continua a partir de la evolución sobre otra variable continua, generalmente el tiempo.

Ejemplo, se intenta predecir el número de clientes o pacientes, los ingresos, llamadas, ganancias, costes, etc. a partir de los resultados de semanas, meses o años anteriores.

Información del Esquema: (descubrir claves primarias alternativas, R.I.).

Reglas Generales: patrones no se ajustan a los tipos anteriores. Recientemente los sistemas incorporan capacidad para establecer otros patrones más generales.

6.4.6 Visualización³⁸

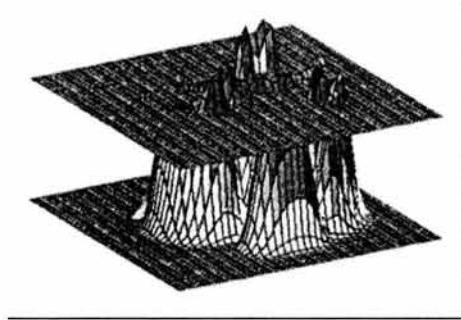


Figura 38. Visualización en 3D de grupos de datos.

Las técnicas de visualización de datos se utilizan fundamentalmente con dos objetivos:

- aprovechar la gran capacidad humana de extraer patrones a partir de imágenes.
- ayudar al usuario a comprender más rápidamente patrones descubiertos automáticamente por un sistema de KDD.

Estos dos objetivos marcan dos momentos diferentes del uso de la visualización de los datos):

- visualización *previa*: se utiliza para entender mejor los datos y sugerir posibles patrones.
- visualización *posterior* al proceso de minería de datos: se utiliza para mostrar los patrones y entenderlos mejor. •La visualización *previa* se utiliza frecuentemente por picapedreros, para ver tendencias y resúmenes de los datos, y por exploradores, para ver 'filones' que investigar.
- La visualización *posterior* se utiliza frecuentemente para validar y mostrar a los expertos los resultados del KDD.

Enfoques en la minería de datos.

³⁸ Visual Data Mining, Wong 1999

Discovery Mining - Encuentra patrones en los datos que pueden ser utilizados para guiar las decisiones.

Mínima dirección sobre el proceso de minado.

"Minado Fácil", En muchas ocasiones el más valioso.

Predictive Mining - Utiliza resultados conocidos para crear modelos y predecir valores futuros.

Mayor dirección sobre el proceso de minado.

Aprende de ejemplos para aplicar posteriormente a una población.

"Minado Sofisticado", requiere de habilidades estadísticas formales.

La minería de datos descubre relaciones ocultas en los datos, predice comportamientos, disminuye riesgos, el análisis OLAP brinda vistas de datos integrando varios aspectos de la organización, los reportes ofrecen la información puntual necesaria para saber el estado detallado de un hecho en particular, sin embargo al final son las personas quienes toman decisiones y no los sistemas de información, los resultados obtenidos con éstas herramientas debe potencializar a estas personas a tomas mejores decisiones y a lograr mejores resultados.

CAPÍTULO VII

Capítulo Siete. Aduanas

Las Aduanas tienen un papel clave en el comercio exterior, tanto desde la perspectiva fiscalizadora, como desde el punto de vista del desarrollo comercial y el posicionamiento del sector exportador nacional en los mercados internacionales.

En este marco, el propósito de las Aduanas es facilitar y agilizar las operaciones de comercio exterior, sin embargo el número de operaciones es cada vez más grande e impacta en diversos rubros del país, como son la economía, la ecología, la cultura, la seguridad nacional, sectores industriales, agrícolas, etc. En buena parte debido a la firma de tratados comerciales y convenios de cooperación económica.

7.1 Antecedentes

La necesidad de contar con información útil de manera oportuna juega un papel muy importante en nuestro desarrollo comercial, las decisiones que se toman pueden hacer prosperar o desaparecer a todo un ramo productivo, por lo que es necesario dar información que fundamente correctamente la toma de decisiones, sin embargo los sistemas tradicionales se ven limitados a responder a las necesidades operativas de la organización, los análisis que brindan estos sistemas resultan de corto alcance, por otra parte intentar consultar exhaustivamente estos datos impactaría negativamente a la operación, resultando en demoras excesivas en los tiempos de respuesta. Tradicionalmente las solicitudes especiales de información demoran en ser atendidas de 1 a 20 días y frecuentemente no responden adecuadamente a las preguntas que las generaron. Es por ello que se desea contar con una herramienta que tenga la capacidad de procesar los grandes volúmenes de operación con que trabajan las Aduanas, que proporcione flexibilidad de

consulta, que mejore el análisis estadístico de los datos, que se disminuyan los tiempos ocasionados por la burocracia, y desde luego que la información sea fácil de consultar, se acceda rápidamente y se encuentre de manera oportuna y confiable.

7.1.1 Reseña histórica³⁹

Históricamente y a nivel mundial, las Aduanas nacieron para impedir la entrada de productos a un determinado territorio, tarea en la que, posteriormente, el interés económico pasó a constituir una preocupación relevante.

A través del tiempo, los países latinoamericanos han definido las Aduanas como el "organismo encargado de aplicar la legislación relativa a la importación y exportación de mercancías y a los otros regímenes aduaneros, percibir y hacer percibir los gravámenes que les sean aplicables y cumplir las demás funciones que se le encomienden".

Una versión etimológica del vocablo permite establecer que "Aduana" es una palabra de origen árabe, que corresponde a "adayuan", traducido como registro o libro de cuentas. No obstante, si nos atenemos al concepto que de ella se tenía en la antigüedad, por Aduana se entendía "el paraje desde donde colocada una persona ve, fisga o registra todo cuanto pasa".

Resumen histórico de la aduana mexicana

Época prehispánica.

En la época prehispánica existía una intensa actividad comercial entre los pueblos mesoamericanos, tenían, además del trueque, sistemas de valor y medida que se empleaban para el intercambio de mercancías. Tal y como

³⁹ <http://www.sat.gob.mx>

ahora utilizamos billetes, monedas y cheques, entonces se usaban pequeños carrizos rellenos de polvo de oro, plumas de aves preciosas y semillas de una cierta especie de cacao, los cuales eran reconocidos y aceptados en toda Mesoamérica. Los tianguis o plazas de mercado estaban sujetos a reglamentos estrictos, cuyo cumplimiento era vigilado por inspectores especiales. Las culturas dominantes, como la mexicana, exigían a los pueblos sojuzgados el pago de tributos e imponían una organización del comercio a grandes distancias, organización que resultó vital para la sociedad precortesiana.

El comerciante que recorría largas distancias era llamado entre los mexicanos "pochteca", él era quien intercambiaba los productos de la región con los de otros pueblos situados más allá de las fronteras del estado mexicano. Así, en México Tenochtitlan podían adquirirse pescados y mariscos frescos de la zona del Golfo o textiles y plumajes del área del Petén. El esplendor del mercado de Tlatelolco, fue registrado incluso en las crónicas de los conquistadores hispanos.

1500

La formalización del comercio de España con sus colonias comenzó con la expedición de las reales cédulas de 1509, 1514, 1531 y 1535, que legitimaban el monopolio mercantil de ésta con los territorios recientemente conquistados. Para tal efecto se instalaron en la Nueva España las Casas de Contratación, las cuales fueron instituciones creadas desde 1503 con el propósito de controlar y fiscalizar el comercio y la navegación entre España y las Indias. En 1551, en Veracruz, se inició la edificación de las primeras instalaciones portuarias.

El comercio con Oriente se inició en la segunda mitad del siglo XVI, cuando se instituyó la ruta mercantil entre Acapulco y Manila. Cabe recordar que las islas Filipinas formaban parte también del Virreinato de la Nueva España. En

1593, la Real Cédula de Felipe II ordenó restringir el volumen de la carga comercial para limitar las mercancías no filipinas, los llamados "productos de la China", procurando así evitar daños al comercio español. Dado que en el puerto de Manila no se efectuaba ningún control de los embarques, la revisión aduanal se realizaba en Acapulco.

En el Virreinato de la Nueva España se estableció el llamado derecho de almojarifazgo (impuesto a la importación). La relevancia de este tributo fue de tal magnitud que la corona española dictó sobre la materia numerosas cédulas reales, decretos y ordenanzas, desde 1532 a 1817, mismas que regulaban la entrada y salida de mercancías; incluso ya se hablaba de franquicias diplomáticas por la introducción de mercancías.

1600

Debido a los constantes ataques de los barcos piratas ingleses y franceses a los puertos de Veracruz, Acapulco y Campeche, en 1597 se dispuso que los oficiales a cargo de la Casa de Contratación se trasladaran a la Banda de Buitrón, lugar situado frente a San Juan de Ulúa, Veracruz, hecho que permitió que en 1601, por órdenes de Felipe II, se instalara la aduana en tierra firme. En 1647, el rey Felipe IV vio la necesidad de que la ciudad contara con una aduana cerrada donde fueran captados todos los productos que entraban al puerto.

1700

En el puerto de Acapulco, el tráfico comercial se llevaba a cabo mediante esporádicas ordenanzas y cédulas reales; en 1702 se elaboró el primer reglamento para el tráfico comercial entre Filipinas y la Nueva España.

El virrey Casafuerte expidió en 1728 una cédula a fin de que ninguna de las mercancías que ingresaban al puerto de Veracruz pudiera bajarse a tierra sin

el consentimiento del oficial de justicia o regidor. La pena por incumplimiento era el decomiso de los productos.

El establecimiento de la Real Aduana del puerto de Acapulco se efectuó alrededor del año 1776.

En 1795 fue constituido el Consulado de Comerciantes de Veracruz, primera asociación gremial de este ramo, cuyos miembros contribuyeron a realizar mejoras al puerto y a la ciudad.

1800

El primer documento legal del México independiente fue el Arancel General Interno para los Gobiernos de las Aduanas Marítimas en el Comercio Libre del Imperio, publicado el 15 de diciembre de 1821. En este documento se designaron los puertos habilitados para el comercio, se especificó el trabajo que debían realizar los administradores de las aduanas, los resguardos y los vistas; además se plantearon las bases para la operación del arancel, estableciendo que los géneros, las mercancías de importación prohibida y las libres de gravamen quedaban a criterio de los administradores de las aduanas.

En 1821, la Sección de Aduanas se encontraba adscrita a la Secretaría de Estado y del Despacho de Hacienda.

En el año de 1831, México firmó un tratado de amistad con los Estados Unidos, uno de sus primeros convenios en materia de comercio internacional.

Por disposición presidencial se creó la Aduana de México en 1884 y se instaló el 8 de mayo en el antiguo edificio de la Casa de Contratación y de la Real Aduana en la plaza de Santo Domingo.

El 1 de marzo de 1887 se expidió una nueva Ordenanza General de Aduanas Marítimas y Fronterizas con dos anexos: en el primero apareció en forma separada la tarifa general; el segundo contenía la aplicación de la tarifa.

1900

El 19 de febrero de 1900 se constituyó por decreto presidencial la Dirección General de Aduanas, conformada por seis secciones.

El 17 de octubre de 1913 se firmó el decreto por el cual se crearon ocho secretarías de Estado para el despacho de los negocios administrativos, entre ellas la Secretaría de Hacienda y Crédito Público y Comercio, con atribuciones sobre aranceles de aduanas marítimas y fronterizas, vigilancia, impuestos federales y otros conceptos.

En 1916 se publicó la nueva tarifa de aranceles, en la cual el único producto prohibido era el opio; se redujeron los gravámenes a los artículos de primera necesidad y se aumentaron para los artículos de lujo.

En 1929 la Ley vigente intentó modernizar y simplificar los procedimientos para el despacho de mercancías en las aduanas, junto con ella se pretendió la unificación en una sola tarifa de los diferentes impuestos. Esta ley fue abrogada por la nueva Ley Aduanera de 1935, en la que se incluyó un nuevo régimen jurídico para las actividades de los agentes aduanales.

El 18 de noviembre de 1931 se expidió el Reglamento Interior de la Dirección General de Aduanas y el Reglamento Interior de la Secretaría de Hacienda. En su artículo 3 se ubicó a la Dirección de Aduanas dentro de las Oficinas de Servicios Generales, con la función de administrar, coordinar y controlar los impuestos, derechos y aprovechamientos aduanales.

En 1951 se publicó el Código Aduanero, vigente hasta 1982, durante el periodo de sustitución de importaciones. Señalaba los lugares para realizar la introducción o extracción de mercancías; los casos de excepción para comerciar con los países, los requisitos especiales, las prohibiciones y la documentación para la operación; así como los productos sujetos a contribuciones aduaneras.

Con la publicación de la Ley Orgánica de la Administración Pública Federal, el 29 de diciembre de 1976, se publicó un nuevo Reglamento Interior de la Secretaría de Hacienda y Crédito Público el 23 de mayo de 1977, quedando integrada por la Subsecretaría de Hacienda y Crédito Público; la Subsecretaría de Ingresos; la Subsecretaría de Inspección Fiscal, a la que se encontraba adscrita la Dirección General de Aduanas; la Oficialía Mayor; la Procuraduría Fiscal de la Federación y la Tesorería de la Federación.

La Ley Aduanera, publicada el 30 de diciembre de 1981, contenía la terminología utilizada internacionalmente; se simplificó la estructura de los recursos administrativos, remitiéndose a los previstos en el Código Fiscal de la Federación y se regía por un nuevo principio de confianza en el contribuyente, a través de la autodeterminación del impuesto; se recopilaban las normas de valoración de mercancías; se promovía la industria maquiladora y las empresas Pitex, y se definieron los regímenes aduaneros que se conocen actualmente.

México se incorporó en 1986 al Acuerdo General sobre Aranceles Aduaneros y Comercio (GATT) antecesor de la OMC, y en mayo de 1988 al Consejo de Cooperación Aduanera (CCA), que tienen por objeto armonizar y facilitar el comercio internacional.

En 1989, la Dirección General de Aduanas quedó asignada a la Subsecretaría de Ingresos, mediante la reforma del Reglamento Interior de la Secretaría de Hacienda y Crédito Público.

Mediante el Decreto por el que se reforman, adicionan y abrogan disposiciones del Reglamento Interior de la Secretaría de Hacienda y Crédito Público, publicado el 4 de enero de 1990, gran parte de las funciones de la Dirección General de Aduanas fue distribuida a unidades administrativas adscritas a la Subsecretaría de Ingresos, con funciones meramente fiscales.

El 25 de enero de 1993 se publicó la reforma al Reglamento Interior de la Secretaría de Hacienda y Crédito Público, en la que se modificó el nombre de la Dirección General de Aduanas por el de Administración General de Aduanas y se estableció la jurisdicción de 45 aduanas en el país.

En 1994 se separaron de la Resolución que establecía reglas fiscales de carácter general (Resolución Miscelánea Fiscal) las reglas de Comercio Exterior.

Posteriormente se vio la necesidad de hacer una reforma integral a la ley que venía regulando la operación aduanera, publicándose una nueva Ley Aduanera el 15 de diciembre de 1995, la cual entró en vigor el 1 de abril de 1996, reformada mediante publicación del 30 de diciembre del mismo año. Los cambios consistieron en la introducción de mecanismos que permitirían valorar la mercancía de acuerdo con lo establecido por el artículo VII del GATT, así como el cambio del sistema aleatorio por un sistema automatizado (inteligente); se reforzaron los métodos para el control de los agentes y apoderados aduanales, así como de sus representantes.

A partir del 1 de julio de 1997 se creó el Servicio de Administración Tributaria (SAT), al cual quedó adscrita la Administración General de Aduanas. El

Reglamento Interior del Servicio de Administración Tributaria se publicó el 30 de junio de 1997.

En 1998 nuevamente se reformó la Ley Aduanera, en el sentido de revisar y fortalecer los mecanismos de control que permitieran combatir la evasión en el pago de contribuciones, el cumplimiento de las regulaciones y restricciones no arancelarias y en general el fraude aduanero, que representa una competencia desleal para la industria nacional, el comercio formalmente establecido y el erario público.

2000

Se hicieron algunas modificaciones, principalmente en el manejo, almacenaje y custodia de mercancías de comercio exterior; el ingreso o extracción de mercancías por vía postal; el reconocimiento aduanero de mercancías; la valoración aduanera; la garantía de contribuciones para el régimen de tránsito de mercancías, los procedimientos administrativos y las infracciones aduaneras, entre otras.

Ante la apertura del comercio exterior emprendida en la década pasada, se pretende prestar el servicio aduanero donde la demanda comercial lo requiera, para lo cual se establece contacto con las cámaras industriales, para conocer las necesidades de importación y exportación de la planta productiva. Con el fin de facilitar la operación mediante métodos de control, se instrumentó el Sistema Automatizado Aduanero Integral (SAAI), así como el sistema de selección automatizado en los reconocimientos, lo que conocemos como semáforo fiscal. Se delegó la responsabilidad de clasificar, determinar el valor en aduana y el origen de las mercancías a los agentes aduanales, quedando a cargo de la autoridad sólo la facultad de verificación del cumplimiento de esta obligación, así se redujo la discrecionalidad de los empleados en la aduana.

La instalación de equipo de alta tecnología ha permeado todas las áreas y procedimientos que se realizan, de tal forma que la red informática se actualiza en forma permanente y puede utilizarse en el ámbito nacional con reportes automatizados del quehacer aduanero. Por otra parte, los sistemas de control (videos, aforos y rayos X) son continuamente renovados.

En cuanto a la facilitación en salas internacionales de pasajeros, se sustituyó el semáforo fiscal por un mecanismo que automáticamente determina si procede o no la revisión del equipaje, mejorando la detección de mercancías no declaradas. Con el rediseñamiento del área de aduanas se está agilizando la revisión de los flujos de pasajeros.

Actualmente el marco legal que rige a las Aduanas de México consiste en⁴⁰:

Constitución Política de los Estados Unidos Mexicanos

Artículo 73, fracciones VII y XXIX

Artículo 89, fracciones X y XIII

Leyes y Reglamentos⁴¹

Compilación de Legislación Fiscal 2003, Tomo I

Código Fiscal de la Federación y su Reglamento

Ley del I.S.R. y su Reglamento

Ley del Impuesto al Activo y su Reglamento

Ley del Impuesto al Valor Agregado y su Reglamento

Compilación de Legislación Fiscal 2003, Tomo II

Ley de Ingresos de la Federación

Ley del Impuesto Especial Sobre Producción y Servicios y su Reglamento

Ley del Impuesto Sobre Automóviles Nuevos

Ley del Impuesto Sobre Tenencia o Uso de Vehículos

Ley Aduanera y Su Reglamento

⁴⁰ www.conggro.gob.mx/Federal/fed00.htm

⁴¹ www.shcp.gob.mx/servs/normativ/index.html

Ley de Coordinación Fiscal

Ley de Contribuciones de Mejoras por Obras Públicas Federales de Infraestructura Hidráulica

Ley Orgánica del Tribunal Federal de Justicia Fiscal y Administrativa

Ley Federal de Derechos

Ley del Servicio de Administración Tributaria

Reglamento Interior del Servicio de Administración Tributaria

Ley de los Impuestos Generales de Importación y de Exportación

Ley de Comercio Exterior

Reglamento de la Ley de Comercio Exterior

Ley Federal de Armas de Fuego y Explosivos

Ley General de Salud

Ley Federal de Responsabilidades de los Servidores Públicos

Ley Orgánica de la Administración Pública Federal

Reglamento Interior de la Secretaría de Hacienda y Crédito Público

Reglas de Carácter General en Materia de Comercio Exterior

7.1.2 La Administración General de Aduanas⁴²

La AGA es una entidad del gobierno federal dependiente del Servicio de Administración Tributaria (SAT, órgano desconcentrado de la Secretaría de Hacienda y Crédito Público), cuya principal función es la de fiscalizar, vigilar y controlar la entrada y salida de mercancías, así como los medios en que son transportadas, asegurando el cumplimiento de las disposiciones que en materia de comercio exterior haya expedido la Secretaría de Hacienda y Crédito Público, así como otras secretarías del Ejecutivo Federal con competencia para ello; coadyuvar a garantizar la seguridad nacional; proteger la economía del país, la salud pública y el medio ambiente,

⁴²[http:// www.aduanas.sat.gob.mx](http://www.aduanas.sat.gob.mx)

impidiendo el flujo de mercancías peligrosas o ilegales hacia nuestro territorio, además de fomentar el cumplimiento voluntario de esas disposiciones por parte de los usuarios.

La aduana es una institución estratégica para el país, por las siguientes razones:

- Vigila y fiscaliza el paso de mercancías por las fronteras y aeropuertos del país.
- Previene y reprime los ilícitos aduaneros.
- Administra los regímenes y operaciones aduaneras.
- Genera estadísticas de las operaciones de comercio exterior.
- Ofrece servicios al comercio exterior.
- Atiende el cumplimiento de obligaciones fiscales de miles de pasajeros y residentes fronterizos, diariamente.
- Es una vía importante de recaudación fiscal.
- Asegura el cumplimiento de las disposiciones que en materia de comercio exterior se emitan.
- Hace cumplir las leyes aplicables y las que se relacionan con su actividad, como las de seguridad nacional, economía, salubridad, comunicaciones, migratorias, fitosanitarias, entre otras.

Sin embargo, diversos factores limitan el cumplimiento de la misión de la AGA, tales como:

- Falta de personal.
- Subcultura de incumplimiento de obligaciones fiscales.
- Corrupción.
- Complejidad del marco regulatorio.
- Amplia gama de servicios y alto volumen de operaciones.
- Elevado número de procesos operativos internos.

- Falta de control sobre la operación cotidiana de los procesos estratégicos de la aduana.

Objetivos principales de la AGA.

Modernizar el sistema aduanero.

Integrar procesos que permitan fortalecer el servicio, con infraestructura para mejorar las instalaciones y la introducción de tecnología de punta para competir a nivel mundial.

Combatir el contrabando.

Mediante la óptima detección y solución de irregularidades, al aplicar controles más estrictos en el sistema aduanero, apoyados con la colaboración nacional e internacional.

Transparentar y mejorar la imagen del servicio aduanero.

Con la continua profesionalización del personal y la difusión de sus procesos para ofrecer al usuario un servicio íntegro.

7.1.2.2 La misión de la Administración General de Aduanas

Controlar la entrada y salida de mercancías del país mediante un servicio aduanero íntegro, transparente, justo y de calidad que facilite a los usuarios el cumplimiento de las disposiciones legales.

7.1.2.3 Procesos realizados en las Aduanas

La aduana efectúa más de 30 procesos. A algunos de ellos son:

- 1) Modular pedimentos de importación.
- 2) Modular pedimentos de exportación.
- 3) Importar temporalmente remolques.

- 4) Importar temporalmente vehículos.
- 5) Internar temporalmente vehículos.
- 6) Cancelar permiso de importación de vehículos.
- 7) Cancelar permiso de internación de vehículos.
- 8) Cancelar permiso remolques.
- 9) Elaborar informe único.
- 10) Elaborar cuenta comprobada.
- 11) Realizar glosa de documentos.
- 12) Administrar oficialía de partes.
- 13) Reconocer mercancías de importación.
- 14) Reconocer mercancías de exportación.
- 15) Reconocer mercancías de importación de ferrocarriles.
- 16) Reconocer mercancías de exportación de ferrocarriles.
- 17) Practicar órdenes de verificación.
- 18) Verificar mercancías y autodeclaración en centros tácticos.
- 19) Administrar muestras.
- 20) Elaborar PAMA⁴³s.
- 21) Elaborar notificaciones.
- 22) Determinar multas de 1er reconocimiento.
- 23) Determinar multas del 2o reconocimiento.
- 24) Administrar incidencias de laboratorio.
- 25) Determinar retornos extemporáneos.
- 26) Determinar créditos fiscales de tránsitos internacionales.
- 27) Determinar créditos fiscales de tránsitos internos.
- 28) Poner mercancías a disposición.
- 29) Emitir resoluciones definitivas.
- 30) Realizar visitas domiciliarias.
- 31) Realizar revisiones de gabinete.

⁴³ Proceso Administrativo en Materia Aduanera.

De ellos, los siguientes seis tienen mayor importancia debido a su contribución en el cumplimiento de la misión de la AGA, por ser parte del valor agregado a los servicios de las aduanas, por su nivel de demanda, por su impacto directo sobre los usuarios y por su prioridad institucional.

Modular pedimentos de importación

Proceso de recibir la mercancía y el pedimento de importación, ingresar la información del mismo al SAAI, y certificar el resultado de la primera selección automatizada.

Constituye el primer filtro de la aduana, brindando la oportunidad de monitorear hasta el 100% de las operaciones y de focalizar los controles por tipo de mercancía, importador y patente aduanal.

Reconocer mercancías de importación

Proceso de identificar, examinar y cuantificar las mercancías, así como la documentación que ampara su importación al país, a fin de verificar que la mercancía corresponde a lo declarado y que se ha cumplido con las disposiciones legales aplicables. Garantiza la correcta aplicación de las disposiciones legales relativas a la importación de mercancías, detectando irregularidades en los despachos sujetos a este proceso.

Verificar mercancías y autodeclaración de mercancías en centros tácticos.

Proceso de inspeccionar aleatoriamente mercancías de pasajeros y residentes fronterizos en cruces internacionales, aeropuertos, centrales de autobuses y garitas, para verificar que cumplen con la franquicia aplicable, según el caso, y brindar el servicio que permita al usuario pagar impuestos por la mercancía que exceda su franquicia.

Esto garantiza el respeto a los derechos de los usuarios para importar, para consumo personal, las mercancías exentas del pago de impuestos, detectando cualquier abuso que se pretenda cometer.

Elaborar PAMA's

Proceso Administrativo en Materia Aduanera (PAMA) que se utiliza para realizar el embargo precautorio de las mercancías cuando se presentan incidencias graves, de conformidad con los supuestos establecidos en el Artículo 151 de la Ley Aduanera. Garantiza la correcta aplicación de las disposiciones legales relativas a las sanciones que se imponen a quienes cometen irregularidades graves en la importación y exportación de mercancías.

Importar temporalmente vehículos extranjeros (CITEV)

Proceso que se realiza para proporcionar un permiso de importación, por un periodo de 180 días, a los vehículos extranjeros, cuyos propietarios son también extranjeros o mexicanos residentes en el extranjero, con la obligación de retornarlo en el plazo señalado. Contribuye a ejercer un control eficaz sobre los vehículos extranjeros que se internan al territorio nacional, otorgando a extranjeros y, sobretodo, a los emigrantes mexicanos la facilidad de utilizar su vehículo extranjero en sus visitas a nuestro país.

Internar temporalmente vehículos fronterizos

Proceso que se realiza para proporcionar un permiso de internación, por un periodo de 120 días, a los vehículos previamente importados a región o franja fronteriza, cuyos propietarios son también residentes fronterizos, con la obligación de retornarlo en el plazo señalado. Contribuye a ejercer un control eficaz sobre los vehículos fronterizos que se internan al territorio nacional, otorgando a los mexicanos residentes de la frontera la facilidad de utilizar su vehículo fronterizo, en sus viajes al interior del país.

7.1.2.4 Comunicación con otras Secretarías del Ejecutivo Federal

Se ha establecido comunicación con diversas áreas externas de otras Secretarías del Ejecutivo Federal, con el propósito de verificar documentación expedida por las mismas, así como información o criterios de aplicación en el marco de su competencia. A continuación se citan algunas de estas dependencias.

Secretaría de Economía

A efecto de verificar diversa documentación e información relacionada con el comercio exterior, se solicita apoyo de esta Secretaría a través de sus diversas dependencias, como son

Dirección General de Normas (DGN)

Esta Dirección formula, revisa, expide, modifica, cancela y difunde las Normas Oficiales Mexicanas y Normas Mexicanas, por lo que se le solicita diversa información relacionada con las materias de su competencia, como criterios interpretativos respecto a diversas normas oficiales mexicanas y formalidades que deben cumplirse en la aplicación de las mismas.

Dirección General de Servicios al Comercio Exterior (DGSCE)

La Dirección General de Servicios al Comercio Exterior tiene una amplia competencia en materia de comercio exterior, relacionada con la emisión de autorizaciones de Programas de Fomento a la Exportación, así como los programas de Promoción Sectorial, aunado a la expedición de diversas regulaciones y restricciones no arancelarias, como son avisos automáticos de importación, constancias de productos nuevo, permisos previos y cupos; en tal virtud, se recurre con frecuencia a esa Dirección General con el fin de verificar la autenticidad y veracidad de los documentos relacionados con dichas obligaciones, o bien conocer si las mercancías introducidas a nuestro

país se encuentran amparadas con alguna de estas autorizaciones o programas, o se encuentran beneficiadas por las tasas aplicables en los diversos cupos asignados.

Unidad de Prácticas Comerciales Internacionales (UPCI)

Esta Unidad tiene como función conocer, tramitar y resolver procedimientos administrativos de investigación en lo referente a prácticas desleales de comercio internacional y medidas de salvaguarda, por lo que brinda su apoyo proporcionando diversa información, como aquella que se relaciona con el cumplimiento de Avisos de Certificados de Uso final en mercancías. Asimismo, nos apoya resolviendo consultas respecto a la correcta interpretación y/o aplicación, en su caso, de las Resoluciones que se dictan en materia de cuotas compensatorias publicadas en el Diario Oficial de la Federación.

Secretaría de Salud

Con objeto de conocer la autenticidad de diversos documentos expedidos por esa Secretaría, ha sido necesaria su colaboración a fin de conocer la veracidad de documentos presentados en el despacho aduanero, tales como los Avisos Sanitarios, o bien normatividad y criterios aplicables.

Secretaría de la Defensa Nacional

En virtud de que esta Unidad Administrativa del Ejecutivo Federal tiene competencia para vigilar el cumplimiento de la importación, venta y fabricación de diversas sustancias explosivas, se demanda su apoyo en aquellas importaciones en las cuales, al momento de la operación correspondiente, se requiere presentar los permisos generales y ordinarios otorgados por esta dependencia para la introducción a territorio nacional de mercancías como los fósforos y cerillos; en tal supuesto, se verifica el permiso correspondiente con dicha Unidad Administrativa y se solicita

diversa información relacionada con la competencia de la misma, a fin de verificar la autenticidad de otro tipo de permisos.

Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación

Se requiere el apoyo y coordinación con esta Secretaría a fin de conocer la autenticidad y veracidad de la información contenida en los certificados fitosanitarios y zoonosanitarios que bajo su respectiva competencia expide, para el caso de la introducción a territorio nacional de mercancías como mantecas y grasas, carnes de bovino y porcino, frijol, uvas, maíz, papa y animales en pie, entre otros.

Comisión Intersecretarial para el Control del Proceso y Uso de Plaguicidas, Fertilizantes y Sustancias Tóxicas (Cicoplafest)

Esta Comisión es de carácter intersecretarial y tiene por objeto proteger la ecología y el medio ambiente, que puede ser lesionado por diversas sustancias tóxicas, por lo que se requiere estar en constante comunicación con la misma a fin de verificar la legitimidad de los diversos permisos otorgados a los importadores de este tipo de mercancía, como son ciertos químicos nocivos que degeneran el medio ambiente.

7.2 Necesidad de información

La operación aduanera en México es cada vez más grande y tiene un impacto creciente en diversos aspectos del país, llámese Economía, industria, agricultura, ecología, cultura, entre otros, debido en gran medida a la apertura y creación de tratados comerciales internacionales y continentales.

La necesidad de información útil, de manera oportuna es muy grande, cabe decir que juega un papel estratégico en nuestra Economía, ya que las

medidas que se toman o se dejan tomar, se puede hacer prosperar o desaparecer todo un ramo productivo interno, saber como se comporta el flujo de mercancías, qué entra, que sale cómo lo hace y con que frecuencia lo hace son solo algunas de las interrogantes que deben ser contestadas, sin embargo los sistemas actuales se ven rebasados, por la cantidad de datos que no deja de aumentar y no siempre pueden contestar, o no lo hacen de manera oportuna, las solicitudes especiales de información se atienden lentamente, y son poco flexibles.

Se desea contar con una herramienta que tenga la capacidad de procesar los grandes volúmenes de operación con que trabajan las aduanas, que proporcione flexibilidad de consulta, que mejore el análisis estadístico de los datos, que disminuya los tiempos ocasionados por la burocracia, que la información sea fácil de consultar, se acceda rápidamente y se encuentre de manera oportuna y confiable

7.3 Descripción del Proyecto:

El proyecto de creación del Data Warehouse de Aduanas tiene como objetivo centralizar la información de la validación y el despacho aduanero que se tiene en las aduanas, en un esquema multidimensional, para permitir que las diferentes área de la Administración General de Aduanas, realicen consultas y análisis de esta información, de manera ágil y oportuna⁴⁴.

El proyecto va del diseño de la base de datos hasta la generación de cubos OLAP y reportes a detalle para análisis multidimensional y operativo para los usuarios.

⁴⁴ Documento del SAT. AIE-SADL-2002-43 Anexo A: Solicitud de atención de requerimientos de sistemas.

7.3.1 Requerimientos

- Establecer un repositorio central, donde se concentre toda la información de pedimentos (inicialmente validación SAAIM3) y con esto eliminar los repositorios redundantes e incompletos.
- Definir modelos de datos enfocados al análisis estadístico y operativo, que cubran los requerimientos actuales de información de aduanas.
- Fomentar el uso de herramientas que faciliten el análisis estadístico, multidimensional y ejecutivo de la información y con esto evitar desarrollos particulares en las áreas que actualmente demandan esto.
- Contar con información completa para el intercambio con usuarios internos (Administraciones de Regulación, Jurídica de ingresos, etc.) y externos (BANXICO, Secretaría de Economía, INEGI, etc.).
- Implementar con procesos automáticos y generales para la extracción de información de pedimentos tanto en las aduanas, como del repositorio, que puedan ser parametrizados a nivel central y controlados localmente (procesos ETL).

7.3.2 Beneficios

Al contar con el repositorio de datos de aduanas se espera:

- Brindar información del contribuyente, del flujo de mercancías.
- Dar seguimiento a las operaciones que involucran mercancía sensible y a los sujetos relacionadas a ellas.
- Disminuir las explotaciones especiales en los equipos de las aduanas, así como el exceso de manipulación.
- Efectuar el monitoreo de actividades sospechosas.
- Generación de Modelos Dimensionales que permitan hacer cruces de información entre diversos aspectos de aduanas.
- Generar independencia de acceso a los datos a través de herramientas de consultas no planeadas.

- Incrementar la recaudación mediante la disminución de fraude.
- Integrar y estandarizar la información de la operación con el fin de facilitar la toma de decisiones basadas en datos y hechos a través del tiempo.
- Poner a disposición de los usuarios la información de los registros de validación SAAI M3.
- Reducir el tiempo de obtención de información así como las solicitudes de extracción de datos.

Se elaboró el plan de trabajo de la primera Etapa, que contempla la información de los pedimentos validados.

ODS SAAIM3

Etapa1. Validador SAAI M3

- Pre-análisis
 - Propuesta de solución.
 - Información general del proyecto.
 - Plan de Trabajo preliminar.
 - Aprobación de la fase.
- Análisis
 - Análisis de la situación actual.
 - Establecer requerimientos a detalle.
 - Revisión de esquema de registros SAAI M3.
 - Revisión y ajustes a la propuesta de solución.
 - Plan de trabajo definitivo.
 - Aprobación de la fase.
 - Modelado
 - Modelo ER destino.
 - Pedimento normal.
 - Pedimento complementario.

- Industria automotriz.
- Previo de consolidado.
- Prueba conceptual del modelo.
 - Esquema de transformación del histórico.
 - Ambientación de los equipos.
 - Generación de datos representativos.
 - Pruebas de consistencia Origen-Destino.
 - Presentación del modelo con herramientas.
 - Prueba de funcionalidad del modelo.
 - Pruebas de desempeño.
 - Revisión y ajustes.
 - Aprobación de la fase.
- ETL Histórico.
 - Extracción de datos origen (histórico).
 - Descarga de datos del origen.
 - Recepción central.
 - Validador central de datos.
 - Carga central.
 - Calidad da datos.
 - Limpieza de datos.
 - Generación de estadísticas de control.
 - Revisión y ajustes.
 - Ambientación de los equipos.
 - Transformación del histórico.
 - Pruebas de consistencia Origen-destino.
 - Migración de estructura de BD a producción.
 - Carga inicial a producción.
 - Solicitud de carga y tranferencia de información.
 - Aprobación de la fase.

- Infraestructura.
 - Infraestructura actual.
 - Infraestructura actual de DW (producción).
 - Necesidades de infraestructura
 - Requerimientos para transformación (producción).
 - Estimación de espacio en BD.}
 - Aprobación de la fase.
- Publicación
 - Esquema de publicación de información.
 - Catálogo del ODS en herramientas BI.
 - Pedimento normal.
 - Pedimento complementario.
 - Industria automotriz.
 - Previo de consolidado.
- Capacitación.
 - Estrategia de capacitación.
 - Capacitación de usuarios.
 - Capacitación de usuarios de AGA.
 - Capacitación de usuarios de otras AG.
 - Instalación de productos a usuarios.
 - Instalación de clientes de consulta.
 - Aprobación de la fase.
- Diseño.
 - Diagramas de procesos.
 - Generación de ODS SAAI M3.
 - Extracción.
 - Esquema de extracción de información diaria.
 - Transformación.

- Esquema de transformación de información.
- Mapeo de información (origen-destino).
- Carga.
 - Esquema de carga de información histórica.
 - Esquema de carga de información diaria.
- Aprobación de la fase.
- Construcción.
 - Desarrollo de programas de extracción.
 - Desarrollo de demonios.
 - Desarrollo de extractor.
 - Desarrollo de transmisor.
 - Desarrollo de esquema de replica de catálogos.
 - Desarrollo de programas de transformación.
 - Desarrollo de integración en equipo central.
 - Desarrollo de transmisor en equipo central.
 - Desarrollo de programas de carga y actualización.
 - Desarrollo de verificador.
 - Desarrollo de cargador OLTP.
 - Desarrollo de transformador OLTP-ODS.
 - Desarrollo del cargador ODS.
 - Reportes de control.
 - Pruebas de consistencia.
 - Revisión y ajustes.

- Aprobación de la fase.
- Implantación.
 - Manual de operación.
 - Manual técnico.
 - Liberación de programas.
 - Liberación a laboratorio.
 - Liberación piloto.
 - Liberación nacional.
 - Aprobación de la fase.

7.3.3 Revisión de consistencia del ODS SAAI M3 – Validador.

Procedimiento:

Se revisó la consistencia de la información del histórico siguiendo estos pasos:

1. Se realizó la transformación histórica.
2. Se recuperaron de las aduanas, archivos de validación de pedimentos enviados por los agentes aduanales.
3. Se desarrolló un programa de revisión para comparar el contenido de los registros en el archivo, contra el contenido de la base de datos transformada.
4. Se ejecutó el programa de revisión con algunos de los archivos.
5. Se realizaron los ajustes necesarios a la transformación.
6. Se volvió a ejecutar el programa de revisión para validar las modificaciones.

7.3.4 Revisiones y ajustes.

Con base en las pruebas funcionales y de desempeño realizadas al modelo, se detectó que era necesario realizar cambios, algunos de ellos se enlistan a continuación:

Campos para el pago de pedimento

Se agregaron los campos en las tablas 301 Datos generales del pedimento complementario y 501 Datos generales:

- a) Clave de banco. Clave de banco en el que se realizó el pago.
- b) Número de caja. Caja en que fue pagado el pedimento en el banco.
- c) Bandera de turno de pago. Turno en que fue pagado el pedimento en el banco.

Campos para agilizar búsquedas.

Se agregaron a la tabla 551 Partidas:

Clave de documento: Se agregó para facilitar las consultas como requerimiento del usuario.

Clave de tipo de operación: Indica si la operación es de importación o exportación.

Campos para las rectificaciones.

Se agregaron a la tabla 701 Rectificaciones, los siguientes campos:

Fecha de balanza anterior. Referencias del pedimento anterior a la rectificación.

Fecha de balanza original. Referencias del pedimento original de la serie de rectificaciones.

Clave de sección aduanera de despacho original. Referencias del pedimento original de la serie de rectificaciones.

Número de pedimento original. Número de documento asignado por el agente o apoderado aduanal al pedimento original o a la última rectificación.

Patente aduanal original: Identifica al agente o apoderado aduanal que efectuó la operación original.

Campos de control.

Se agregaron a la tabla 601 Previos de Consolidado, los siguientes campos:

Bandera de consolidación. Permite saber si el pedimento es consolidado.

Se agregaron a las tablas 301 Datos generales del pedimento complementario y 501 datos generales, los siguientes campos:

Bandera de justificación. Indica si el pedimento fue cargado mediante el proceso de justificación.

7.3.5 Infraestructura.

Infraestructura del equipo de base de datos

Actualmente se cuenta con 4 equipos con las siguientes características:

HP Serie n-4000.

4 Procesadores 440Mhz por nodo.

2 GB en memoria RAM por nodo.

S.O. Unix HP-UX.

Arreglo de discos Symetrics 1TB.

Infraestructura del equipo de extracción, transformación y carga.

2 equipos con las siguientes características:

HP Serie n-4000.

4 Procesadores 440Mhz por nodo.

2 GB en memoria RAM por nodo.

S.O. Unix HP-UX.

Infraestructura del equipo de publicación de información con herramientas especializadas.

2 equipos con las siguientes características:

Compaq Proliant.

2 Procesadores PIII 900 Mhz por equipos.

2 GB en memoria RAM por nodo.

S.O. Windows 2000 Advanced Server.

Arreglo de discos RAID.

7.4 Áreas Usuarias.

Las áreas usuarias del Data Warehouse, en principio son las propias subadministraciones de la AGA, a continuación se enlistan algunas de las que se han visto beneficiadas por contar con la información.

Administración Central de Investigación Aduanera.

Apoyados con los Convenios Internacionales sobre Asistencia Mutua entre Administraciones de Aduanas celebrados por México con otros países, realiza cotejos de documentación que permiten comprobar la autenticidad de la información emitida por empresas extranjeras, la existencia jurídica de ellas y su ubicación. El Data Warehouse les proporciona información de los pedimentos desde que son validados hasta que son despachados, así como del propio Contribuyente, sus facturas, permisos, etc.

Administración Central de Fiscalización Aduanera

Esta Administración Central se encarga de realizar visitas domiciliarias a instalaciones de importadores, con objeto de comprobar el cumplimiento de Normas Oficiales Mexicanas de mercancías importadas, el correcto y oportuno pago de contribuciones en materia de comercio exterior al revisar su contabilidad interna y en la comprobación de la legal estancia y/o retorno físico de mercancías importadas temporalmente, entre otros rubros. El Data Warehouse les proporciona información sobre los permisos declarados en el SAAI, datos generales del contribuyente, mercancías y restricciones aplicables.

Administración Central de Informática

Esta Administración Central proporciona las estadísticas de operaciones de comercio exterior, facilita el análisis y estudio de operaciones bajo investigación. El Data Warehouse les ha permitido responder con mayor

rapidez a las solicitudes de información de las demás administraciones de la AGA, y de otras administraciones del SAT, como son AGAFF,

Administración Central de Visitaduría

Su función es verificar que el personal de las Aduanas realice los procedimientos conforme a las normas existentes a fin de detectar responsabilidad en los servidores públicos que practicaron el reconocimiento aduanero de mercancías sin detectar incidencias que la Administración de Glosa detectó. Consultan información de los pedimentos que se pagan y despachan en las aduanas y comparan la información electrónica del Data Warehouse con la documentada existente.

GLOSA

Su función es analizar la información de los pedimentos después de que éstos han sido pagados y despachados. Verifican documentalmente y por medio de visitas la información declarada. La información del Data Warehouse les ayuda a disminuir el tiempo de consulta de los pedimentos en papel, así como detectar si existen faltantes en los documentos que reciben de las aduanas.

Las áreas externas que solicitan información y se ven beneficiadas por el establecimiento del Data Warehouse de Aduanas son entre otras:

- Banco de México.
- Comisión Intersecretarial para el Control del Proceso y Uso de Plaguicidas, Fertilizantes y Substancias Tóxicas (Cicoplafest).
- INEGI.
- Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación.
- Secretaría de Economía.
- Secretaría de la Defensa Nacional.
- Secretaría de Salud.

7.5 Modelo de base de datos resultante para el Data Warehouse

La etapa inicial del repositorio consiste en incluir la información de validación y despacho de mercancías.

El modelo operacional del SAAIM3 consiste en más de 230 tablas, Fue necesario establecer un único diccionario de datos con cada una de estas tablas⁴⁵.

Ejemplo de un elemento contenido en el diccionario de datos origen:

Tabla ci_aduana

Descripción Catálogo de Aduanas a nivel nacional

Columna	Tipo	Opción Null	Descripción de la Columna
c_aduana	Smallint	Not null	Llave de la aduana
N_cve_aduana	Smallint	Not null	Clave de la aduana
D_descripcion	Varchar(40)	Null	Descripción o nombre de la aduana
F_ini_vig	Date	Not null	Fecha de inicio de vigencia
F_fin_vig	Date	Null	Fecha de fin de vigencia
B_estado	Char(1)	Not null	Bandera que indica el estado de la clave (0: inactivo, 1: Activo).

El modelo ODS para funcionar como repositorio, por su parte consiste en 41 tablas que agrupa a la información en tres temáticas que son

- Pedimento normal.
- Pedimento Complementario.
- Industria automotriz.

Ejemplo de documentación de una tabla de Data Warehouse, en primer lugar con nombres de negocio y después el nombre dentro del manejador de base de datos.

⁴⁵ Diccionario de Datos SAAIM3 vb15.2

Tabla Datos Generales r5dg0ae1⁴⁶

Nombre de negocio	Tabla columna
Patente aduanal	"r5dg0ae1""paadtue1")
Número de pedimento	"r5dg0ae1""npuemde1")
Clave de sección aduanera de despacho	"r5dg0ae1""csadled1")
Clave de tipo de operación	"r5dg0ae1""ctolipa1")
Clave de documento	"r5dg0ae1""cdloacv1")
Clave de sección aduanera de entrada	"r5dg0ae1""csaeled1")
Clave Unica de Registro de Población del contribuyente	"r5dg0ae1""curpcln1")
Registro Federal de Contribuyentes	"r5dg0ae1""rfceeg1")
Clave Unica de Registro de Población del agente o apoderado aduanal	"r5dg0ae1""curpaaa1")
Tipo de cambio	"r5dg0ae1""tciapmo1")
Total de fletes	"r5dg0ae1""tfoltea1")
Total de seguros	"r5dg0ae1""tsoetga1")
Total de embalajes	"r5dg0ae1""teomtba1")
Total de otros incrementales	"r5dg0ae1""toioint1"
Total de otros deducibles	"r5dg0ae1""todotet1"
Peso bruto de la mercancía	"r5dg0ae1""pbmeres1"
Clave de medio de transporte de salida	"r5dg0ae1""cmtsler1"
Clave de medio de transporte de arribo	"r5dg0ae1""cmtaler1"
Clave de medio de transporte de entrada o salida del país	"r5dg0ae1""cmtespl1"
Clave de destino de la mercancía	"r5dg0ae1""cdmleea1"
Nombre del contribuyente	"r5dg0ae1""ncoomb1"
Calle del domicilio del contribuyente	"r5dg0ae1""cdcaool1"
Número interior del domicilio del contribuyente	"r5dg0ae1""nidcuno1"
Número exterior del domicilio del contribuyente	"r5dg0ae1""nedcuxo1"
Código postal del domicilio del contribuyente	"r5dg0ae1""cpdcoo1"
Municipio/Ciudad del domicilio del contribuyente	"r5dg0ae1""mcdcuio1"

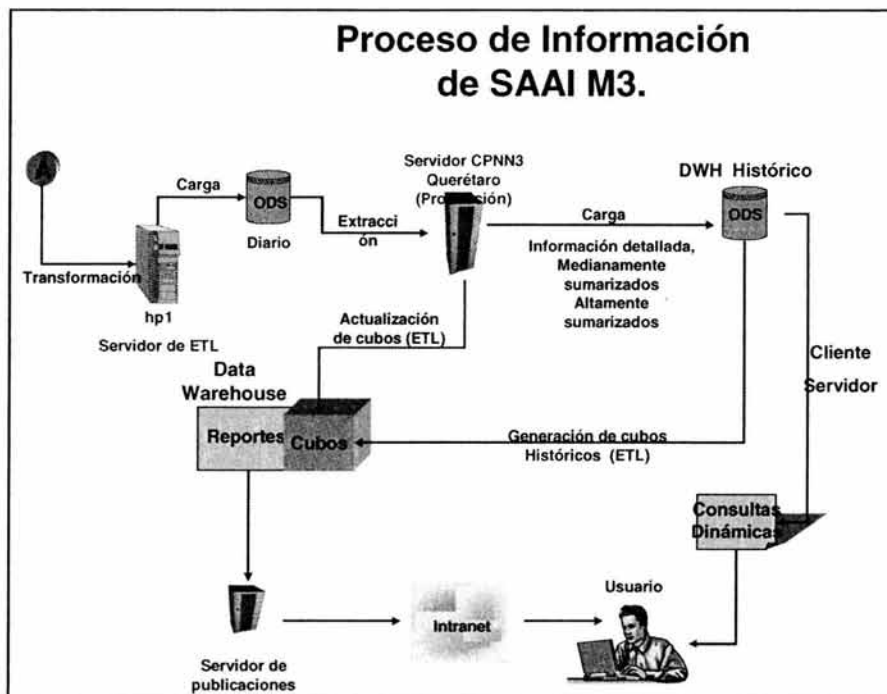
⁴⁶ Documentación del ODS SAIM3, de la Administración de Desarrollo para Aduanas, del SAT, 2002.

Nombre de negocio	Tabla columna
Clave de entidad federativa del domicilio del contribuyente	"r5dg0ae1""cefdcln1"
Clave de país del domicilio del contribuyente	"r5dg0ae1""cpdclao1"
Clave de prevalidador	"r5dg0ae1""cplraev1"
Bandera de justificación	"r5dg0ae1""bjaunsd1"
Clave de banco	"r5dg0ae1""cblaavn1"
Número de caja	"r5dg0ae1""ncuamje1"
Bandera de turno de pago	"r5dg0ae1""btpauan1"
Clave de tipo de pedimento	"r5dg0ae1""ctpliea1"
Fecha de recepción de pedimento	"r5dg0ae1""fveaclh1"
Fecha de balanza	"r5dg0ae1""fbeatclh1"
Fecha de pago real	"r5dg0ae1""fpreaec1"

Las tablas que conforman el Data Warehouse de Aduanas son:

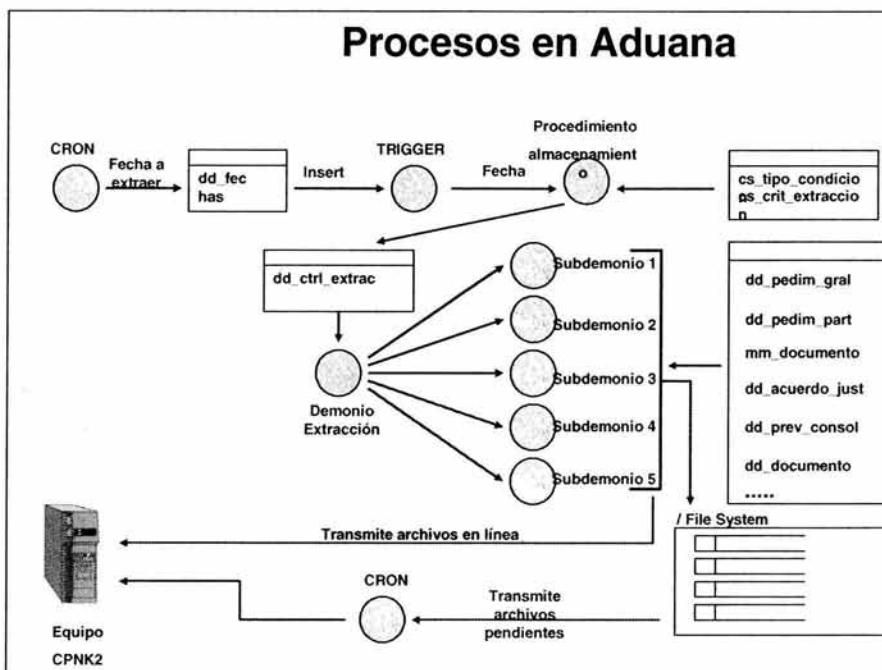
Tablas del ODS SAAIM3
R 301 Datos generales del pedimento complementario
R 302 Prueba suficiente del pedimento complementario
R 351 Información de la partida del pedimento complementario
R 352 Detalle de importaciones complementarias a EUA/Canadá
R 353 Información complementaria de la partida
R 355 Contribuciones complementarias de la partida
R 358 Observaciones complementarias de la partida
R 501 Datos Generales
R 502 Transportes de las mercancías
R 503 Guías
R 504 Contenedores
R 505 Facturas
R 506 Fechas del pedimento
R 507 Casos del Pedimento
R 508 Cuentas aduaneras de garantía del pedimento
R 509 Tasas del pedimento

Tablas del ODS SAAIM3
R 510 Contribuciones del pedimento
R 511 Observaciones del pedimento
R 512 Descargos de mercancías
R 514 Movimientos en cuenta aduanera
R 515 Informe de la industria automotriz
R 520 Destinatarios de la mercancía
R 551 Partidas
R 552 Mercancías
R 553 Permisos de la partida
R 554 Casos de la partida
R 555 Cuentas aduaneras de garantía de la partida
R 556 Tasas de las contribuciones de la partida
R 557 Contribuciones de la partida
R 558 Observaciones de la partida
R 559 Normas oficiales mexicanas
R 560 Partidas del informe de la industria automotriz
R 601 Previos de consolidado
R 701 Rectificaciones
R 702 Diferencias de contribuciones del pedimento



A partir de este sistema OLTP con información de un solo día se ejecutan los procesos de transformación de datos que permiten operar con menos de 41 tablas, esta información se carga en un ODS que también contiene información de un solo día y el cual envía esta información al Data Warehouse ubicado en instalaciones en Querétaro. La información del Data Warehouse se explota mediante herramientas de consulta no planeada (dinámicas), reportes predefinidos, y cubos multidimensionales.

7.6.1 Proceso implementado en las aduanas



En cada aduana se implementó el proceso de extracción que se indica en la figura, el cual indica los datos y la fecha a extraer (un día antes del actual), insertando un registro en la tabla dd_fechas, este evento activa la extracción de datos mediante un demonio en el equipo, la información se almacena en archivos y se envía al servidor de transmisión.

7.6.2 Factores críticos de éxito

- Dar a conocer el modelo.
- Implementar mesas de trabajo para la definición de formas de explotación del modelo propuesto.
- Participación activa de los usuarios.

- Hardware y Software
- Marco de integración de datos

7.7 Experiencia

A lo largo de poco más de dos años en el proyecto se pueden destacar los siguientes puntos:

El Data Warehouse, ha permitido facilitar el despegue de los usuarios en el uso de herramientas de consulta AD-HOC, para la generación de sus propias consultas de información en forma ágil y oportuna, el uso y explotación de cubos multidimensionales comienza cuando los usuarios han cubierto sus necesidades más básicas de información.

Los usuarios han podido atender muchas de sus necesidades de información sin hacer una solicitud a la administración central de informática y de explotaciones especiales.

Se ha conseguido ofrecer la información a los usuarios con un solo día de retraso, previamente solo en atender su solicitud de información pasaban entre 3 y 20 días.

En lo posible se debe conservar al personal asignado al proyecto, no hacerlo ocasiona retrasos, hacerlo contribuye a la especialización de cada integrante.

La preparación de los datos es indispensable para el manejo adecuado de estos y de los procesos que van a operarlos.

Se deben establecer procesos para identificar y manejar los registros duplicados. Aun existen procesos manuales para el manejo de registros duplicados

Las reglas de depuración de datos y reglas de rechazo deben documentarse ampliamente. Se desea establecer mecanismos de limpieza y conversión de datos desde las fuentes que agilicen los procesos y para disminuir los registros rechazados.

Falta automatizar completamente la transferencia de datos entre equipos de operación, de integración y transformación. Aun existe dependencia del factor humano durante la migración de datos, aumenta el costo de integración de datos y disminuye la confiabilidad de la integración.

Uno de los mayores retos es la carga Inicial de datos históricos, debido a los altos volúmenes de datos. Los procesos de carga y transformación de datos consumen gran cantidad de tiempo y recursos de los equipos de cómputo.

Se deben acordar políticas y horarios para efectuar las cargas de datos para interferir lo menos posible con el desempeño de las consultas que realizan los usuarios.

Se debe tener un plan consistente y amigable de capacitación a los usuarios. Es importante considerar que generalmente no cuentan con conocimientos técnicos o de bases de datos.

Finalmente es poco preciso considerar que, una vez que el Data Warehouse está funcionando, los problemas terminarán, tanto para el equipo de desarrollo como para los usuarios. Para los primeros implica un cambio en la manera en que suelen adquirir información, los obliga a capacitarse y manifestar sus inquietudes al equipo de desarrollo, y a éste le obliga a rediseñar parte de la estrategia y estructura de Data Warehouse debido a cambios en los requerimientos.

7.8 Proyección

Ampliar la cobertura de usuarios del Data Warehouse a todas las administraciones de la Aduanas.

Ampliar el alcance de las herramientas de explotación de información.

Automatizar completamente los procesos de extracción transformación y carga.

Introducir herramientas que ayuden a mejorar la calidad de los datos.

Iniciar el proceso de minería de datos.

Generar indicadores y tableros de control.

Hacer cruces de información entre el Data Warehouse de Aduanas y la información de Recaudación, Auditoría Fiscal, Grandes Contribuyentes, que permita conocer integralmente las operaciones del contribuyente.

El proyecto de crear un Data Warehouse implica llevar a cabo una serie de actividades en muchas ocasiones bastante complejas, la participación de los usuarios enriquece directamente su contenido, éxito y evolución, asimismo es un proyecto que requiere de todo el apoyo por parte de los directivos tanto por los costos e inversiones que necesita, como para dar a conocer los resultados y evaluarlos. Es importante buscar prioridades, y comenzar con uso específico a fin de entregar resultados palpables, esto no implica que el proyecto sea finito de hecho se considera que no termina nunca, por la misma razón el almacenamiento masivo se vuelve crucial. En relación a los usuarios es importante definir esquemas de seguridad para controlar el acceso a información sensible y a su correcto aprovechamiento. En cuanto al equipo de desarrollo es positivo no solo apoyarle en su capacitación técnica, sino en mejorar su conocimiento del negocio, finalmente se sabe que es un proyecto difícil, pero bien manejado trae resultados para lo cual se debe estar preparado para invertir.

Conclusiones

El enfoque de la creación de un Data Warehouse, para el desarrollo e implementación de un esquema de Inteligencia de Negocio en una Institución Gubernamental es una filosofía que se ha presentado a lo largo de este trabajo, pese a que prosigue el debate acerca del paradigma correcto y los métodos más efectivos para el desarrollo de un proyecto de Data Warehouse, ¿Por qué entonces a últimos años se esta viendo una adopción masiva de este concepto?

La respuesta según veo se encuentra en que el ámbito empresarial se desarrolla en un entorno altamente competitivo, demandante de acciones precisas y oportunas que conlleven un cambio inaplazable hacia el manejo adecuado de la información, aunque la dificultad de la transición tecnológica y del cambio cultural que lo acompaña representen el gran obstáculo a vencer, esto es, que aún cuando la mayor parte de nosotros apreciamos la necesidad de una disciplina de tecnología de la información, seguimos luchando contra la inercia de los viejos métodos.

Para facilitar la transición se necesitan, como se ha presentado en los capítulos que nos preceden, procesos de modelado de datos, de extracción, transformación y carga más maduros y efectivos, potentes herramientas de análisis OLAP y de reporte, aceptación por parte de las áreas gerenciales involucradas, a compartir la información inherente a sus áreas y el apoyo total por parte de los Ejecutivos tomadores de decisiones que rigen el rumbo de la institución, así como una fuerte difusión al respecto. Un proyecto de Data Warehouse raramente dispone de los beneficios de la publicidad masiva, pero a medida que pasa el tiempo, el concepto se vende por si mismo, en cierta manera este trabajo es un testigo al respecto.

Tal vez no se esté de acuerdo con todos los enfoques descritos en este trabajo, algunos conceptos habrán de ser ajustados para que funcionen bien en entornos distintos de negocio, sin embargo espero sinceramente que este trabajo haya esbozado la problemática a la que nos hemos estado enfrentando y que asimismo, haya demostrado las ventajas de la implementación de Data Warehouse y de la Inteligencia de Negocio; que sirva también como referencia acerca de los métodos y las herramientas a utilizar, sin embargo aun quedan demasiados aspectos por investigar y desarrollar, ojalá que también sirva como punto de partida para hacerlo.

Este es el momento en que comenzamos un nuevo milenio, la información ha pasado a ser el producto más importante y más valioso, su adecuada explotación se ha convertido en toda una industria a nivel mundial, su impacto e importancia han avanzado a pasos agigantados y sin embargo una nueva generación de desarrolladores de inteligencia de negocios deberá de enfrentarse a muchos de los mismos desafíos con los que se enfrentaron generaciones anteriores, esperemos que quienes hoy día enfrentan este desafío lo hagan con la plena conciencia de crear un mundo mejor.

Glosario

BI, business intelligence Inteligencia de negocio.

Celda Se trata de los datos concretos que aparecen en la intersección que queda definida al seleccionar un miembro de cada dimensión en un cubo.

Cluster Grupo de registros con características similares.

DBMS Manejador de base de datos

DIMENSIÓN Una dimensión es el atributo estructural de un cubo que consiste en una lista de miembros, los cuales son todos de un tipo parecido según la percepción que el usuario tiene de los datos.

DWH data warehouse

E/S Entrada/salida

ETL Proceso de extracción, transformación y carga.

GB gigabyte (1024 MB)

HOLAP OLAP híbrido

IA, AI inteligencia artificial*

IBM International Business Machines Corporation

KB kilobyte (1024 bytes)

KDD Extracción de conocimiento a partir de datos

LAN Red de área local

MB megabyte (1024 KB)

ODS almacén de datos operacionales.

OLAP OnLine Analytical Processing procesamiento analítico en línea.

OLTP Online Transactional Processing procesamiento de transacciones en línea.

Procesamiento por lotes Se refiere a una forma de procesamiento que el usuario programa, pero que el ordenador completa de forma automática. Fuera de las horas punta, se puede programar el procesamiento por lotes y a menudo se completa de forma más eficaz.

RDBMS Manejador de base de datos relacional.

Replicación: Es el proceso de crear y administrar versiones duplicadas de una base de datos, sincroniza el conjunto de replicas de tal manera que los cambios hechos en una replica se vean en las demás.

ROI return on investment, Retorno de inversión.

ROLAP Olap relacional

Roll-up, consolidación Por lo general, dentro de cada dimensión, las bases de datos multidimensionales establecen jerarquías o relaciones de datos basadas en fórmulas. La consolidación consiste en calcular todas estas relaciones de datos para una o más dimensiones. Aunque por lo general, estas relaciones son sumas, se puede definir cualquier tipo de relación que implique un cálculo o una fórmula.

Sistema legado generalmente son aplicaciones que pueden estar en uso o desuso o están en proceso de desecharse. Proveen los datos más antiguos de la operación de la organización.

Slice and Dice; Se trata del proceso de navegación iniciado por el usuario al solicitar la visualización de una página de forma interactiva mediante la especificación de fragmentos.

Slice, fragmento Un fragmento es un subgrupo de una matriz multidimensional que corresponde a un sólo valor que comparten uno o más miembros de las dimensiones fuera del subgrupo.

Socket en Unix y en algunos otros sistemas operativos es un elemento de software que conecta una aplicación a un protocolo de red.

SQL Structured Query Language

Tabla Se trata de una estructura en la cual se puede almacenar la información en la base de datos. Las tablas de datos se almacenan en filas y en columnas. Ver también code table; tabla de códigos

TB terabyte 1024 GB)

Transformation A re-expression of the data such as aggregating it, reformatting it, and changing its unit of measure.

Triggers es un procedimiento SQL que inicia una acción cuando ocurre un evento Insert, delete o update.

Tuberías, pipes Conexión temporal entre dos programas o comandos. Normalmente el sistema operativo acepta entradas del teclado y envía la salida a la pantalla. A veces es útil tomar la salida de un comando y usarla como entrada para un segundo comando sin pasar los datos a través del teclado o de la pantalla.

Glosario sobre Aduanas⁴⁷.

Aduanas. Son los lugares autorizados para la entrada o la salida del territorio nacional de mercancías y de los medios en que se transportan o conducen.

Agente Aduanal. Es la persona física autorizada por la Secretaría de Hacienda y Crédito Público, mediante una patente, para promover por cuenta ajena el despacho de las mercancías en los diferentes regímenes aduaneros.

Arancel. Impuesto sobre los bienes importados. Relacionado con la lista de gravámenes.

Arancel Ad-Valorem. Gravamen expresado como porcentaje fijo del valor del bien importado.

Apoderado Aduanal. Es la persona física que haya sido designada por otra persona física o moral para que en su nombre o representación se encargue del despacho de mercancías.

Arancel específico. Es el gravamen expresado en unidades monetarias sobre o por cada unidad de medida de un bien importado.

Arancel mixto. Es aquel compuesto por un arancel ad valorem y uno específico, aplicados simultáneamente para gravar la importación de un bien.

Autoridad aduanera. Es la autoridad competente que, conforme a la legislación interna de una parte, es responsable de la administración de sus leyes y reglamentaciones aduaneras.

Balanza de pagos. Se denomina así al sumario de transacciones económicas de un país con el resto del mundo durante un lapso de tiempo determinado.

Base Gravable del Impuesto General de Importación. Es el valor en aduana de las mercancías.

Bienes de una Parte. Son los productos nacionales como se definen en el GATT y aquellos que las partes convengan en un tratado de libre comercio.

Bien no originario. Es aquel bien que no califica como originario, de conformidad con las reglas correspondientes.

Bien originario. Significa que cumple con las reglas de origen establecidas en el capítulo correspondiente de un tratado de libre comercio.

⁴⁷ <http://www.aduanas.gob.mx>

Clasificación arancelaria. Es la clasificación de las mercancías objeto de la operación de comercio exterior que deben presentar los importadores, exportadores y agentes o apoderados aduanales, previamente a la operación de comercio exterior que pretendan realizar.

Código de Valoración Aduanera. Es el Acuerdo para la aplicación del artículo VII del GATT, contenido en los artículos 64 a 79 de la Ley Aduanera de 1998.

Comprador final. Es la última persona que, en territorio de la parte importadora, adquiere los bienes en la misma forma en que serán importados. Este comprador podría no ser necesariamente el usuario final del bien.

Contribuciones. Son los créditos fiscales, impuestos, derechos y otras obligaciones que se pagan por la exportación e importación de productos y servicios.

Contribuyentes. Persona física o moral sujeto de impuestos por la actividad que realiza, con la finalidad de contribuir al gasto público.

Costo neto. Se refiere a todos los costos menos los de promoción de ventas, comercialización y de servicio posterior a la venta, regalías, embarque y empaque, así como de los costos financieros

Cuotas compensatorias. Son los derechos aplicables a ciertos productos originarios de determinado o determinados países para compensar el monto de la subvención concedida a la exportación de estos productos.

Medida de regulación o restricción no arancelaria que se impone a aquellos productos que se importan en condiciones de prácticas desleales de comercio internacional y se aplican independientemente del arancel que corresponda a las mercancías

Son los derechos antidumping y cuotas o derechos compensatorios, según la legislación de cada parte del tratado.

Cupos de importación o exportación. Capacidad permitida tanto al régimen de importación como de exportación en cuanto a cantidad y tiempo.

Declaración. Es la obligación que tienen los importadores, exportadores y agentes o apoderados aduanales de manifestar a las autoridades aduanales las mercancías objeto de comercio exterior.

Despacho aduanero. Conjunto de actos y formalidades relativos a la entrada y salida de mercancías del territorio nacional.

Dictamen anticipado o resolución. Se refiere a la resolución favorable al particular que emite la autoridad aduanera a petición del importador, productor o exportador,

previamente a la importación, certificando que la determinación de valor de una mercancía o de su origen por cambio de clasificación arancelaria o valor de contenido regional o marcado de país de origen son correctos, en términos del tratado aplicable y sus reglamentaciones uniformes.

Empresa. Cualquier entidad constituida u organizada conforme al derecho aplicable, tenga o no fines de lucro, sea de propiedad privada o gubernamental, incluidas todas las sociedades, fundaciones, compañías, sucursales, fideicomisos, participaciones, empresas de propietario único, coinversiones u otras asociaciones.

Embargo precautorio. Retención de mercancías por no acatarse a lo establecido en la Ley Aduanera.

Fracción arancelaria. Es la descripción numérica o desglose de un código de clasificación que otorga el Sistema Armonizado.

Franja fronteriza. Es el territorio comprendido entre la línea divisoria internacional y la línea paralela ubicada a una distancia de veinte kilómetros hacia el interior del país.

Franquicia. Libertad y exención que se concede a una persona para no pagar derechos e impuestos por las mercancías que introduce o extrae del país.

Gastos de corretaje. Retribuciones pagadas a un tercero por los servicios prestados como intermediario en la operación de compraventa de las mercancías objeto de la operación.

Impuestos. Son las contribuciones establecidas en las leyes que deben pagar las personas físicas y morales que se encuentren en la situación jurídica o de hecho prevista por la misma y que sean distintas de aportaciones, contribuciones y derechos.

Impuestos al comercio exterior. Son los gravámenes que se tienen que cubrir de acuerdo con las tarifas que establecen las leyes por las operaciones de internación y externación en el territorio nacional de bienes y servicios.

Infracción. Es el hecho de violación por fallar en el cumplimiento de la ley o de no hacer lo que la ley indica.

LAB. Significa libre a bordo, independientemente del medio en que se transporte, desde el punto de embarque directo del vendedor al comprador.

Ley. Norma jurídica obligatoria y general dictada por un órgano legítimo para poder regular la conducta de los hombres o para establecer otros órganos necesarios para el cumplimiento de sus fines.

Mercado común. Significa un grado de integración económica tal entre países y territorios en el que se eliminan todas las barreras arancelarias y comerciales entre los países miembros y que permite el libre movimiento de personas y capitales a lo largo del territorio nacional de los países miembros.

Mercancías. Son todos los bienes y servicios que cruzan nuestra frontera nacional aun cuando las leyes las consideren como no sujetas a una operación comercial. De acuerdo con la Ley Aduanera, se consideran mercancías los productos, artículos, efectos y cualesquiera otros bienes, aun cuando las leyes los consideren inalienables o irreductibles a propiedad particular.

Mermas. Son consideradas como aquellas que se consumen durante el proceso de producción o bien se pierden por desperdicio y no es posible comprobar.

Normas oficiales mexicanas (NOM) Es una regulación técnica de observancia obligatoria, expedida por las dependencias competentes, con una multiplicidad de finalidades, cuyo contenido debe reunir ciertos requisitos y seguir el procedimiento legal.

Parte. Significa todo estado respecto del cual haya entrado en vigor un tratado de libre comercio.

Partida. Significa clasificación arancelaria de cuatro dígitos.

Patente. Documento expedido por la Secretaría de Hacienda y Crédito Público para el ejercicio de las actividades relacionadas con la importación y exportación de mercancías

Pedimento. Solicitud oficial que expide la Secretaría de Hacienda y Crédito Público para permitir la importación y exportación de mercancías a las personas físicas y morales que cumplan con la Ley Aduanera.

Pedimento consolidado. Documento de solicitud oficial que ampara diversas operaciones de un solo exportador.

Precios estimados. Valor presuntivo que se fija a un determinado producto (prohibidos según la regla 7 del GATT) sólo para efectos de determinar el valor en aduana.

Procedimiento Administrativo en Materia Aduanera (PAMA). Es el conjunto de actos previstos en la Ley Aduanera, ligados en forma sucesiva, con la finalidad de

determinar las contribuciones omitidas y en su caso, imponer las sanciones que correspondan en materia de comercio exterior, respetando al particular su derecho de audiencia al considerarse las probanzas y argumentaciones que pretendan justificar la legalidad de sus actos.

Recintos fiscales. Son aquellos lugares donde se encuentran las mercancías de comercio exterior controladas directamente por las autoridades aduaneras y en donde se da el manejo, almacenaje y custodia de dichas mercancías por las autoridades.

Reconocimiento aduanero. Procedimiento mediante el cual se realiza una revisión de documentos y mercancías para precisar la veracidad de lo declarado en el pedimento, con la finalidad de determinar la cantidad, características y la plena identificación de las mercancías.

Región fronteriza. Es el territorio que determine el Ejecutivo Federal, en cualquier parte del país, incluyendo la franja fronteriza.

Regla de origen. Es el criterio general o específico pactado expresamente en un tratado de libre comercio para definir cuándo un bien deberá considerarse como no originario y cuándo un bien podrá calificar como originario.

Reglamento. Conjunto de normas que rigen un determinado acto, conducta o actividad.

Reglamentaciones uniformes. Son las normas jurídicas que, a la fecha de entrada en vigor de los tratados, establecen y ejecutan las partes, mediante sus respectivas leyes y reglamentaciones, referentes a la aplicación, interpretación y administración de los asuntos que convengan a las partes.

Restricciones no arancelarias. Acto administrativo por medio del cual se imponen determinadas obligaciones o requisitos a la importación, exportación y circulación o tránsito de las mercancías, distintos a aquellos de carácter fiscal.

Resolución de determinación de origen. Significa una resolución de autoridad aduanera que establece si un bien califica como originario de conformidad con las reglas de origen.

Responsables solidarios. Serán los mandatarios, los agentes aduanales, el propietario o tenedor de las mercancías, del pago de impuestos al comercio exterior y de las demás contribuciones, así como de las cuotas compensatorias que se

causen con motivo de la introducción de mercancías al territorio nacional o de su extracción del mismo.

Secretaría de Hacienda y Crédito Público. Considerado como el órgano máximo para desarrollar diversas actividades en materia fiscal.

Sistema armonizado. Es un código de clasificación arancelaria de seis dígitos. Su nombre completo es Sistema Armonizado de Designación y Codificación de Mercancías y sus notas interpretativas, que en México ha sido publicado bajo la forma de las Leyes del Impuesto General de Importación y Exportación.

Territorio aduanero. Territorio de un estado en el cual las disposiciones de su legislación aduanera son aplicables.

Tiendas libres de impuestos. Tiendas bajo control aduanero generalmente situadas en los puertos marítimos y aeropuertos, situadas donde los viajeros salen al extranjero pudiendo adquirir mercancías con exoneración de derechos de aduana y otros impuestos.

Tipo de divisa. Unidad monetaria utilizada en una transacción u operación de comercio.

Tráfico fronterizo. Los desplazamientos efectuados de una parte a otra de la frontera aduanera por personas residentes en una de las zonas fronterizas adyacentes.

Tráfico interno. Transporte de personas embarcadas o de mercancías cargadas en un lugar del territorio aduanero para ser desembarcadas o descargadas en otro lugar del mismo territorio aduanero.

Transacción en materia de infracción aduanera. Acuerdo por el cual las autoridades aduaneras, actuando dentro de los límites de su competencia, renuncian a perseguir una infracción aduanera, siempre y cuando las personas implicadas cumplan con ciertas condiciones.

Transbordo. Traslado de mercancías efectuado bajo control aduanero de una misma aduana, desde una unidad de transporte a otra, o a la misma en distinto viaje, incluida su descarga a tierra, conl objeto de que continúe hasta su lugar de destino.

Tránsito. Paso de mercancías extranjeras a través del país cuando éste forma parte de un trayecto total comenzado en el extranjero y que debe ser terminado fuera de sus fronteras. Igualmente se considera como tránsito de mercancías el envío de mercancías extranjeras al exterior que se hubieren descargado por error u otras

causas calificadas en las zonas primarias o lugares habilitados, con la condición de que no hayan salido de dichos recintos y que su llegada al país y su posterior envío al exterior se efectúe por vía marítima o aérea.

Tránsito aduanero. Régimen aduanero bajo el cual las mercancías sujetas a control aduanero son transportadas de una aduana a otra.

Transporte internacional. El tráfico de naves o aeronaves, nacionales o extranjeras, de carga o de pasajeros hacia o desde el exterior.

Transporte interno. Transporte de personas embarcadas o de mercancías cargadas en un lugar situado dentro del territorio nacional para ser desembarcadas o descargadas en un lugar situado dentro del mismo territorio nacional.

Transportista. Persona que transporta efectivamente las mercancías o que tienen el mando o la responsabilidad del medio de transporte.

Turista. Dentro de este término se comprende: los extranjeros que ingresan al país con fines de recreo, deportivos, de salud, de estudios, de gestiones de negocios, familiares, religiosos u otros similares, sin propósito de actividades remuneradas y sin propósito de inmigración. El término turista designa a toda persona, sin distinción de raza, sexo, lengua o religión, que entre en el territorio de un Estado distinto de aquel en que dicha persona tiene su residencia habitual y permanezca en él 24 horas cuando menos y no más de seis meses.

Unión aduanera. Entidad constituida por un territorio aduanero que sustituye a dos o más territorios aduaneros y que posee en su última fase las características siguientes:

Un arancel aduanero común o armonizado para la aplicación de este arancel.

La ausencia de percepción de los derechos de aduana y de tasas de efecto equivalente en los intercambios que conforman la unión aduanera de los productos enteramente originarios de estos países o de productos de países terceros cuyas formalidades de importación han sido cumplidas y los derechos de aduanas y tasas de efecto equivalentes percibidos o garantizados y que no se han beneficiado de una bonificación total o parcial de estos derechos o tasas.

Eliminación de reglamentaciones restrictivas a los intercambios comerciales en el interior de la unión aduanera.

Usuario. Persona natural o jurídica que haya convenido con la Sociedad Administradora el derecho a desarrollar actividades instalándose en la Zona Franca.

Valor en aduana. Es el valor de un bien para efectos de cobro de aranceles sobre un bien importado.

Valor reconstruido. Es la suma del costo o valor de los materiales y de la fabricación u otras operaciones efectuadas para producir las mercancías importadas, más la cantidad global por concepto de beneficios y gastos generales, más los gastos de transporte, seguros y otros en los que se incurra con motivo del transporte de las mercancías.

Valor de transacción. Significa el precio efectivamente pagado o por pagar por un bien o material relacionado con una transacción del productor de ese bien, ajustado conforme al Código de Valoración Aduanera o bien de la Ley Aduanera, sin considerar si el bien o material se vende para exportación.

Valor CIF. Cláusula de compraventa que incluye el valor de las mercancías en el país de origen, el flete y seguro hasta el punto de destino.

Valor declarado. Valor con fines aduaneros de las mercaderías contenidas en un envío que están sometidas a un mismo régimen aduanero y clasificadas en una misma posición arancelaria.

Valor FOB. Cláusula de compraventa que considera el valor de la mercancía puesta a bordo del vehículo en el país de procedencia, excluyendo seguro y flete.

Vehículo. Cualquier medio de transporte de carga o de personas..

Verificación previa. Revisión o inspección de mercadería antes de someterla a un régimen aduanero determinado.

Visita de recepción y control a bordo. Operaciones por las cuales la nave, aeronave u otro vehículo de transporte es visitado por el personal de la aduana, a su llegada o durante su permanencia en puertos, aeropuertos o terminales, con objeto de recibir y examinar los documentos del medio de transporte y proceder a su registro y vigilancia.

Vista de aduanas. Funcionario técnico aduanero encargado de aforar las mercaderías, ahora denominados verificadores.

Zona de libre comercio. Entidad constituida por los territorios aduaneros de una asociación de estados que posee en su última fase las características siguientes: Eliminación de los derechos de aduana para los productos originarios de un país de la zona.

Cada Estado conserva su arancel de aduana y su legislación aduanera.

Cada estado de la zona conserva su autonomía en materia de aduana y de política económica.

Los intercambios se basan en la aplicación de reglas de origen para tener en cuenta los diferentes aranceles aduaneros y evitar los desvíos de tráfico.

Eliminación de las reglamentaciones restrictivas en los intercambios comerciales en el territorio de la zona.

Zona de vigilancia aduanera. Parte determinada del territorio aduanero donde la aduana ejerce poderes especiales en virtud de los cuales aplica o puede aplicar medidas especiales de control aduanero. Parte del territorio aduanero en que la posesión y la circulación de mercancías pueden someterse a medidas especiales de control por la aduana.

Zonas libres Son determinadas regiones ubicadas fuera de los centros de distribución y abasto de insumos y bienes básicos nacionales, por lo cual se les exenta de impuestos en la compra de mercancías de procedencia extranjera necesarias para la producción y el consumo. Actualmente en México ya no existen zonas libres. La firma del Tratado de Libre Comercio permite, sin embargo, el establecimiento de zonas fronterizas para aplicar políticas de tratamiento fiscal y comercial especiales.

Bibliografía

Date, C. J.

“Introducción a los sistemas de bases de datos”

Prentice Hall.

México.

2001.

“Ascential DataStage Designer Guide”

Ascential Software.

E.E.U.U.

2001.

Chanona Carmona, Miguel Ángel.

“IBM DB2 Intelligent Miner for Data”.

IBM de México.

México.

2002.

“Introducción a la Administración y Consistencia en Base de Datos”

Servicio de Administración Tributaria, Sistema Único de Normatividad.

México.

2003.

Inmon, Bill.

“Building the Data Warehouse Getting Started”

Ed. John Wiley & Sons.

E.E.U.U.

2002.

Stevens, W. Richard

“Advanced Programming in the Unix Environment.”

Ed. Addison-Wesley.

E.E.U.U.

1992.

“Manejando y optimizando servidores de bases de datos dinámicos en Informix”

IBM Data Management Education.

E.E.U.U.

1999.

De Miguel, Adoración

“Diseño de bases de datos. Problemas resueltos”

Ed. Alfaomega, Ra-Ma

México.

2001.

Peña, Daniel

“Análisis de datos multivariantes”

Ed. McGraw Hill.

España.

2001.

F. Codd, S.B. Codd, & C.T. Salley,

“Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate”

1993.

Referencias en Internet.

- <http://www.daedalus.es/DocWhitepapers-E.php>
- <http://bbdd.escet.urjc.es/documentos/mineria%20de%20datos%20con%20ceptos%20y%20objetivos.pdf>
- <http://itmanagement.webopedia.com/>
- <http://www.aduanas.sat.gob.mx>
- <http://www.businessdecision.com/esp/Decouvrir/Glossaire.htm>
- <http://www.cognos.com/es/products/etl>
- <http://www.crm-forum.com>
- <http://www.olapcouncil.org/research/whtpapy.htm>
- <http://www.olapinfo.de/whitepapers.php>
- <http://www.sat.gob.mx>