

# Construcción de la factoría de información corporativa

José Samos Jiménez  
Josep Curto Díaz

PID\_00203542



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

# Índice

<b>Introducción</b> .....	7
<b>Objetivos</b> .....	8
<b>1. Punto de partida: apoyo a la toma de decisiones en las organizaciones</b> .....	9
1.1. Apoyo a la toma de decisiones en el entorno operacional .....	9
1.2. Problemas del entorno operacional de telaraña .....	11
1.3. Transformación del entorno operacional para satisfacer las necesidades de información .....	14
1.4. La problemática de la calidad de los datos de la organización ...	16
<b>2. Variantes de la FIC</b> .....	18
2.1. Construcción de almacenes de datos departamentales .....	18
2.2. Construcción del almacén de datos corporativo <i>a posteriori</i> .....	21
2.3. Combinación del almacén de datos operacional y el almacén de datos corporativo .....	23
2.4. La FIC sin el almacén de datos operacional .....	25
<b>3. Estrategias para la construcción de la FIC</b> .....	28
3.1. Construcción de la FIC mediante un solo proyecto .....	29
3.2. Enfoques parciales para la construcción de la FIC .....	29
3.2.1. Enfoque centrado en el almacenamiento y acceso a los datos .....	30
3.2.2. Enfoque centrado en la obtención y el almacenamiento de los datos .....	33
3.3. Construcción de la FIC mediante proyectos autónomos .....	34
3.3.1. El primer proyecto: proyecto global de desarrollo .....	36
3.3.2. Proyectos de desarrollo de infraestructura .....	37
3.3.3. Desarrollo de proyectos autónomos .....	40
3.4. Evolución del entorno operacional .....	43
3.4.1. Evolución del entorno operacional en telaraña .....	43
3.4.2. Evolución de las aplicaciones operacionales .....	45
3.4.3. Otros cambios en la organización .....	46
3.5. Perfiles en el equipo de gestión y desarrollo de la FIC .....	47
3.5.1. El administrador de la FIC .....	47
3.5.2. Los analistas de requerimientos de negocio .....	47
3.5.3. El arquitecto de la FIC .....	48
3.5.4. El patrocinador de la FIC en la organización .....	48
3.5.5. El gestor de cambios organizacionales .....	49
3.5.6. El gestor de metadatos .....	50

3.5.7.	Los analistas de la calidad de los datos .....	50
3.5.8.	El administrador de bases de datos .....	51
3.5.9.	Especialistas en obtener y acceder a los datos .....	51
3.6.	Razones por las cuales los proyectos fallan .....	52
<b>4.</b>	<b>Desarrollo del componente de integración y transformación.....</b>	<b>54</b>
4.1.	Obtener los datos de las fuentes de datos .....	54
4.1.1.	Obtener la imagen inicial de los datos .....	54
4.1.2.	Métodos para obtener las actualizaciones de los datos .....	55
4.1.3.	Criterios de selección del método para obtener las actualizaciones de los datos .....	58
4.2.	Actualización de los datos del almacén de datos .....	59
4.2.1.	Métodos de actualización de los almacenes de datos ....	59
4.2.2.	Selección del método de actualización de los almacenes de datos .....	60
4.3.	Frecuencia y ventana de actualización .....	61
4.3.1.	Frecuencia de actualización de un almacén de datos ....	61
4.3.2.	Ventana de actualización de un almacén de datos .....	62
4.4.	Herramientas de apoyo para el desarrollo .....	63
4.4.1.	Funcionamiento de las herramientas .....	63
4.4.2.	Ventajas e inconvenientes de las herramientas .....	65
4.4.3.	Otras herramientas de apoyo .....	66
4.5.	Uso de metadatos .....	67
<b>5.</b>	<b>Desarrollo del almacén de datos corporativo.....</b>	<b>68</b>
5.1.	Revisión del proceso de desarrollo .....	68
5.2.	El modelo de datos del almacén de datos corporativo .....	69
5.3.	Transformaciones para construir el esquema del almacén de datos corporativo .....	70
5.3.1.	Adición de un elemento de tiempo .....	71
5.3.2.	Organización de los datos según su estabilidad .....	72
5.3.3.	Adición de datos derivados .....	73
5.3.4.	Cambio en la granularidad de los datos .....	74
5.3.5.	Fusión de entidades .....	75
<b>6.</b>	<b>Incorporación del almacén de datos operacional.....</b>	<b>76</b>
6.1.	Paquetes de aplicaciones y el almacén de datos operacional .....	76
6.2.	Velocidad de refresco de los datos .....	77
6.3.	Planificación de la incorporación del almacén de datos operacional .....	78
<b>Resumen.....</b>		<b>80</b>
<b>Actividades.....</b>		<b>81</b>

---

<b>Ejercicios de autoevaluación.....</b>	<b>81</b>
<b>Solucionario.....</b>	<b>82</b>
<b>Glosario.....</b>	<b>84</b>
<b>Bibliografía.....</b>	<b>85</b>



## Introducción

La factoría de información corporativa (FIC) permite a los analistas disponer de la información que necesitan como apoyo a la toma de decisiones. Aun así, generalmente, la FIC no se puede considerar como un producto o una aplicación empaquetada que se pueda adquirir y, una vez instalada en nuestras organizaciones, empiece a "fabricar" información a partir de los datos de las fuentes de datos operacionales.

Generalmente, la FIC no se puede adquirir, se tiene que construir en las diferentes organizaciones.

Podemos plantear la construcción de la FIC según distintos enfoques, y también podemos considerar variantes en la arquitectura de la FIC. En cualquier caso, implementar esta arquitectura o sus variantes no es una operación trivial.

En este módulo estudiaremos diferentes enfoques alternativos para la construcción de la FIC o sus variantes. Una posibilidad es llevarla a cabo aplicando la metodología de desarrollo de proyectos de nuestra organización, del mismo modo que implementaríamos cualquier otro proyecto que nos hubiera sido asignado. De manera alternativa, podríamos pensar en implementar la FIC según una estrategia específica de descomposición de la arquitectura en forma de proyectos. Estudiaremos las ventajas e inconvenientes de los diferentes enfoques. Asimismo, estudiaremos soluciones que hay que tener en cuenta en el desarrollo de algunos de los componentes de la FIC.

### Herramientas de apoyo

Se pueden adquirir herramientas de apoyo para construir sus componentes.

## Objetivos

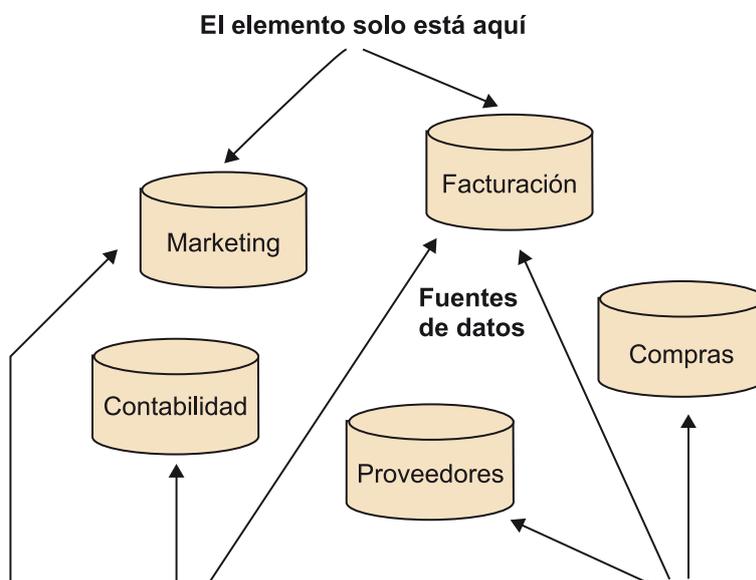
En este módulo se pretende ofrecer una visión global del proceso de construcción de la FIC, así como de la construcción de sus componentes. Mediante el estudio, se conseguirán los objetivos siguientes:

- 1.** Entender cómo se ofrece apoyo a la toma de decisiones en organizaciones que no disponen de la FIC.
- 2.** Conocer los problemas que surgen al tratar de implementar variantes de la FIC en organizaciones.
- 3.** Determinar cómo se puede estructurar el proceso de construcción de la FIC en forma de proyectos.
- 4.** Tener presente la evolución del entorno operacional para el desarrollo de la FIC.
- 5.** Comprender los nuevos roles que aparecen en los equipos de desarrollo que intervienen en la construcción de la FIC.
- 6.** Conocer soluciones específicas para la construcción de diferentes componentes de la FIC.

## 1. Punto de partida: apoyo a la toma de decisiones en las organizaciones

### 1.1. Apoyo a la toma de decisiones en el entorno operacional

En los módulos anteriores, hemos estudiado las características del entorno operacional. Sabemos que los datos en un entorno operacional generalmente están orientados a las aplicaciones o a la funcionalidad y desintegrados, además de ser volátiles y no históricos.



**El mismo elemento, diferente nombre**    **Diferente elemento, el mismo nombre**

Datos en el entorno operacional orientados a la funcionalidad y no integrados.

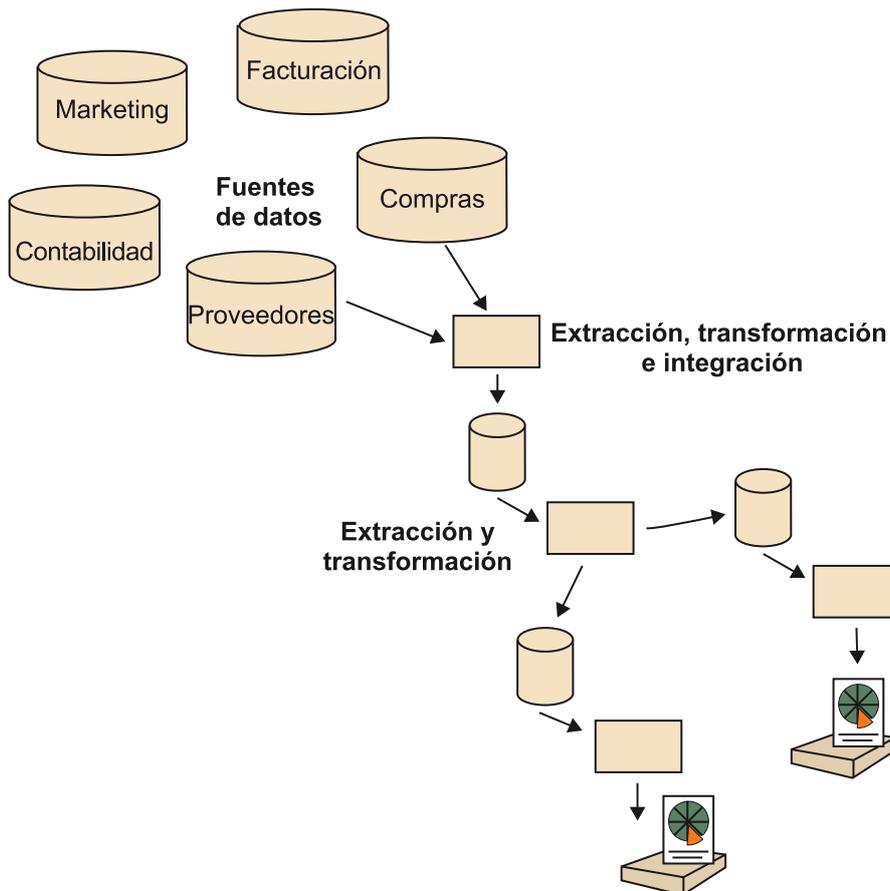
Cada aplicación operacional ofrece las funcionalidades operacionales para las que ha sido diseñada. Además de estas funcionalidades, o como parte de estas, generalmente también permite hacer consultas sobre sus datos y generar informes a partir de estos, habitualmente informes preestablecidos. Estos informes son los que pueden usar los analistas como apoyo en el proceso en que se toman decisiones.

En algunos casos, con la información que se obtiene en los informes preestablecidos no basta y se requiere la generación de nuevos informes, posiblemente incluyendo datos de más de una aplicación. Entonces, la solución consiste en desarrollar un conjunto de programas de extracción, transformación e integración de datos cuyo resultado sea el informe deseado. Generalmente, estos programas son desarrollados por el Departamento de Informática de la organización a petición de los analistas que usarán la información obtenida.

#### Ejemplo de aplicación de contabilidad

La aplicación de contabilidad permite, por ejemplo, llevar la contabilidad de la organización y hacer una lista de los movimientos de un determinado periodo ordenados por diferentes conceptos.

Los programas desarrollados transforman e integran los datos obtenidos de las bases de datos de las aplicaciones y generan bases de datos intermedias, o bien exclusivamente transforman los datos de las aplicaciones para adaptarlos a la estructura requerida por los analistas. Para generar un informe, pueden ser necesarios distintos pasos de transformación e integración de datos. En la medida en que se pueda, se trata de reutilizar el trabajo hecho en el desarrollo de informes previos, por ejemplo, accediendo a alguna de las bases de datos intermedias generadas. De este modo, se ahorra tiempo de desarrollo del informe y de ejecución para obtener los datos requeridos.



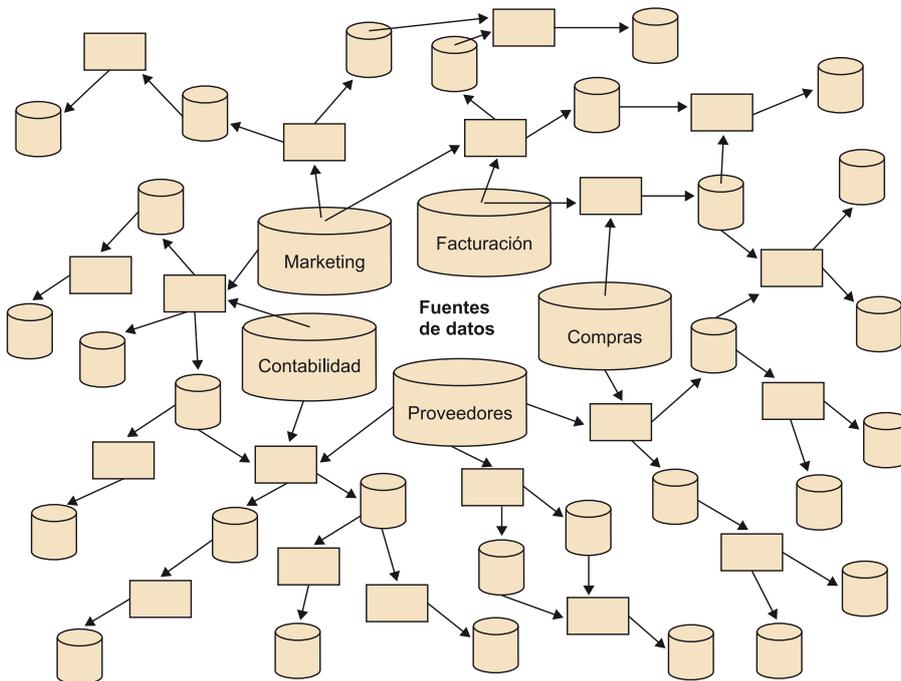
Generación de informes a medida a partir de las fuentes de datos del entorno operacional.

Los informes a medida que piden los analistas se generan mediante programas desarrollados con esta finalidad.

La necesidad de informes a medida por parte de los analistas no es un hecho aislado en las organizaciones, puesto que las posibilidades ofrecidas por los sistemas operacionales en este sentido suelen ser muy limitadas. Esto se produce porque el objetivo principal de estos sistemas es el apoyo a la operativa de la empresa: sus datos están estructurados y organizados con esta misión. Aun así, los analistas requieren datos estructurados para el apoyo al proceso de toma de decisiones que tienen que llevar a cabo, y estos datos se basan en

aquellos registrados por los sistemas operacionales. Por lo tanto, el desarrollo de programas específicos para la generación de informes a medida es bastante frecuente en las organizaciones.

Con el paso del tiempo, la situación a la que se llega en muchas organizaciones es similar a la mostrada en la figura siguiente. Es decir, partiendo de las fuentes de datos de los sistemas operacionales se construye una estructura parecida a una telaraña compuesta por programas de extracción, transformación e integración de datos, y también de bases de datos intermedias que son utilizadas por otros programas para producir los informes requeridos. Inmon denomina la estructura resultante como el entorno operacional de telaraña.



El entorno operacional de telaraña.

El entorno operacional de telaraña no es el resultado de un proceso de planificación, es decir, no estaba planificado así por adelantado, sino que surge de manera "natural" como resultado de las demandas de informes a medida por parte de los analistas.

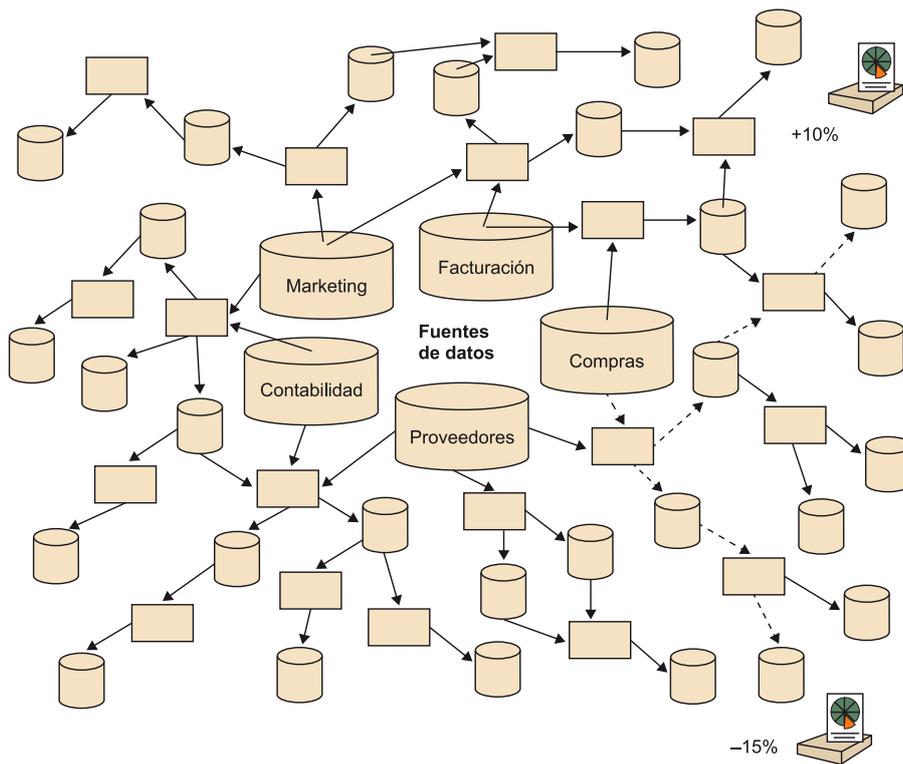
## 1.2. Problemas del entorno operacional de telaraña

El entorno operacional de telaraña resultado del desarrollo de informes a medida solicitados por parte de los analistas presenta graves inconvenientes. El primero de estos, que se ha intentado reflejar visualmente en la figura anterior, es un problema de **complejidad**: el resultado del desarrollo a petición de informes es un entorno complejo, no planificado ni estructurado.

En el desarrollo de cualquier aplicación informática, cumplir los plazos de entrega suele ser un objetivo muy importante, que a veces prevalece sobre otros, como por ejemplo, el de generar la documentación necesaria para su mantenimiento posterior. En el caso de las aplicaciones desarrolladas para generar un informe, el plazo de entrega es especialmente crítico, puesto que ha sido solicitado por un analista que lo necesita como soporte para tomar una decisión, por lo cual es frecuente que no se hagan las actividades de documentación que serían deseables. El objetivo de cada equipo de desarrollo es generar el informe requerido y para esto desarrolla los programas necesarios, basándose en la medida en que pueda en los programas o bases de datos intermedios previamente desarrollados. Aun así, esto último no siempre es posible porque no hay bastante documentación y control sobre estos o simplemente porque es más rápido y menos costoso en tiempo de desarrollo hacer algo nuevo que modificar programas que ya existían de manera previa y que, por lo general, no están suficientemente documentados.

En relación con esto último, tenemos un problema de **falta de productividad**: para cada informe solicitado, se tienen que localizar los datos necesarios (esta operación no es nada trivial, por la falta de integración de los datos en las fuentes de datos y en las bases de datos intermedias) y desarrollar los programas de extracción, integración y transformación de los mismos. Dado que para hacer estas operaciones es muy difícil reutilizar los esfuerzos dedicados previamente al desarrollo de otros informes, nos encontramos en una situación en la que, para desarrollar cada informe, prácticamente tenemos que partir de cero.

Uno de los problemas más graves de este entorno es el que se refleja en la figura siguiente, el de **falta de credibilidad**: debido a la complejidad del entorno, no es improbable que un mismo informe (por ejemplo, los resultados del departamento en el último mes) se haya obtenido de dos maneras distintas, posiblemente como parte de otro informe que contiene información adicional. Aunque se ha partido de los mismos datos, se han recorrido caminos diferentes para cada uno de los informes (los marcados en negrita en la figura). El problema surge cuando la información presentada en los dos informes no coincide y se genera desconcierto en los usuarios y la correspondiente falta de confianza y credibilidad respecto al Departamento de Informática como responsable de los informes generados.



Falta de credibilidad generada en el entorno operacional de telaraña.

No resulta extraño que se produzca esta situación, puesto que los datos de las fuentes suelen ser no integrados. Basta con que el equipo de desarrollo de uno de los informes interprete algún dato de las fuentes de datos o de las bases de datos intermedios de manera diferente para que se produzcan resultados como los que se han descrito antes. La pérdida de la confianza de los analistas en los datos con los que trabajan y en el departamento que los ha generado es un problema muy grave, puesto que esta es muy difícil de recuperar posteriormente.

El entorno operacional de telaraña presenta problemas de complejidad, falta de productividad y falta de credibilidad.

Aparte de los problemas descritos, la pregunta que nos tenemos que hacer es la siguiente: ¿puede satisfacer las necesidades de información de los analistas el entorno operacional de telaraña?

Respecto al tiempo requerido para obtener la información, el plazo que transcurre desde el momento en que el analista hace la petición del informe hasta que lo recibe seguramente supera lo que sería de esperar en un entorno tan cambiante como en el que se mueven las organizaciones actualmente.

Por otro lado, en lo que respecta al contenido de los informes, los analistas suelen necesitar estudiar la evolución de los diferentes datos, y para esto, necesitan analizar la información histórica de la organización. Aun así, la información de los sistemas operacionales generalmente es información no histórica, o contiene una historia muy limitada, puesto que no es necesario que lo sea para soportar las operaciones de la organización.

**Ved también**

Podéis ver el apartado de "Características del entorno operacional" del módulo "La factoría de información corporativa" de esta asignatura.

El entorno operacional de telaraña no satisface los requerimientos de información de los analistas.

### **1.3. Transformación del entorno operacional para satisfacer las necesidades de información**

Si el entorno operacional presenta graves problemas y, además, no satisface las necesidades de información por parte de los analistas, ¿lo podemos transformar para corregir estas deficiencias?

Respecto a la historia de la información almacenada en los sistemas operacionales, una solución podría ser almacenar la información histórica requerida por los analistas. Esto representaría un aumento en la complejidad de estos sistemas. Por otro lado, si las aplicaciones que hay no almacenan la historia de los datos, deberían modificarse para que lo hicieran. Esta operación puede resultar muy costosa, sobre todo en aplicaciones antiguas que hayan sufrido cambios a lo largo de su historia, puesto que, generalmente, los cambios suelen estar poco documentados y las modificaciones requeridas no son sencillas.

Para generar informes fiables más rápidamente y con menos coste, se tendría que reducir la complejidad del entorno operacional de telaraña. Para esto, principalmente necesitaríamos disponer de un entorno que tuviera los datos integrados. El problema es que los datos del entorno operacional, sobre todo si este se ha desarrollado de manera gradual a lo largo de la historia o si se han adquirido aplicaciones empaquetadas, suelen ser no integrados. En este caso, también necesitaríamos modificar todas las aplicaciones para integrar los datos en las mismas.

Por lo tanto, la situación es la siguiente: disponemos de un conjunto de aplicaciones, que forman el entorno operacional y que satisfacen las necesidades operacionales de la organización, pero que no satisfacen las necesidades de información de los analistas. Para intentar corregir esta situación, necesitaríamos modificar todas las aplicaciones que hay para integrar los datos y almacenar su historia. Esta operación puede resultar inviable por su complejidad y coste.

Por otro lado, si nos fijamos en las necesidades de los diferentes tipos de usuarios, los analistas necesitan hacer consultas muy complejas (por ejemplo, evolución de los resultados del departamento en los últimos años) y obtener los

resultados de manera inmediata. Aun así, los usuarios de los sistemas operacionales necesitan conocer de la manera más fiel posible la situación actual del sistema modelado (por ejemplo, situación actual del departamento). Por lo tanto, los sistemas operacionales se tienen que estructurar para reflejar todos los cambios que se produzcan y estos cambios pueden ser muy frecuentes (por ejemplo, interesa que el saldo de una cuenta esté actualizado en el momento en que se produzca cualquier cambio).

Las necesidades de los analistas y de los usuarios de los sistemas operacionales son contrapuestas: difícilmente un sistema se puede diseñar y configurar para ser óptimo tanto en la respuesta a consultas complejas requerida por los analistas como en la ejecución de modificaciones sobre los datos en los que estas se basan, requerida por los usuarios de los sistemas operacionales.

- Un mismo sistema no puede satisfacer al mismo tiempo las necesidades operacionales de la organización y las de información de los analistas.
- Aunque fuera posible hacerlo, los sistemas operacionales son muy costosos de modificar para integrar todos los datos y almacenar su historia.
- Por lo tanto, los analistas de las organizaciones requieren sistemas específicos como apoyo en el proceso de toma de decisiones.

Los nuevos sistemas requeridos por los analistas necesitan obtener los datos a partir de los sistemas operacionales que hay, y evitar modificarlos tanto como sea posible. La solución consiste en construir sistemas con esta finalidad específica, que estén diseñados para optimizar el tipo de operaciones que necesitan hacer los analistas y que funcionen en plataformas especialmente configuradas para esto.

Por otro lado, sería conveniente que los analistas dispusieran de sistemas que les permitieran obtener directamente la información que requieran, en lugar de lograrla mediante peticiones al Departamento de Informática. Esto se puede conseguir diseñando los nuevos sistemas mediante un modelo de datos que esté especialmente orientado a esta finalidad, que es el modelo de datos multidimensional.

Para ofrecer soporte a las necesidades de información de los analistas, la solución consiste en incorporar el concepto de almacén de datos a la organización mediante la implementación de la FIC.

**Ved también**

Estudiaremos este modelo de datos en el módulo "Diseño multidimensional" de esta asignatura.

**Ved también**

Podéis ver el módulo "La factoría de información corporativa" de esta asignatura.

## 1.4. La problemática de la calidad de los datos de la organización

El valor potencial del almacén de datos depende de la calidad de los datos de la organización. Con frecuencia, el crecimiento precipitado de muchas organizaciones que no tienen una estrategia TIC lo bastante clara genera un problema en la calidad de la información, que se agrava con el paso del tiempo. Se trata de un fenómeno que afecta a muchos de los sistemas de información presentes en una organización, desde la inteligencia de negocio (BI) a la gestión de clientes (CRM), los ERP o la gestión del ciclo de vida de productos (PLM), entre otros.

Una posible solución a este problema consiste en aplicar una estrategia de gestión de datos maestros (MDM<sup>1</sup>) combinada con una solución tecnológica que permita a las organizaciones gestionar y consolidar los datos maestros que se encuentran dispersos, con el objetivo de ofrecer una visión única de las entidades principales de la organización: clientes, productos, etc. El objetivo es eliminar las islas informacionales existentes y crear una estructura de datos maestra basada en una visión estratégica y corporativa:

- Identificar las fuentes de origen de los datos.
- Identificar a los productores y consumidores de datos maestros.
- Recopilar y analizar metadatos sobre los datos maestros recopilados en el primer paso.
- Determinar a los responsables (administradores) de los datos maestros.
- Implementar un programa de gobierno de datos (y por lo tanto, tener un grupo responsable de este programa).
- Desarrollar el modelo de metadatos maestros.
- Elegir una solución o un conjunto de soluciones como medio para mejorar la calidad de los datos.
- Diseñar la infraestructura necesaria.
- Generar y verificar los datos maestros.
- Modificar los sistemas consumidores y productores de información.
- Implementar un proceso de mantenimiento.

Una de las tareas que hay que hacer en la estrategia MDM es el despliegue de un programa de gobernanza de datos (DM<sup>2</sup>). Este programa une a personas, procesos y tecnología para cambiar la manera en que los datos son adquiridos, gestionados, mantenidos, transformados en información, compartidos en el contexto de la organización como conocimiento común y, de manera sistemática, obtenidos para mejorar la rentabilidad de la empresa. Es decir, hablamos de una disciplina en la que convergen conceptos como calidad de datos<sup>3</sup>, gestión de datos<sup>4</sup>, gestión de procesos de negocio<sup>5</sup> y gestión del riesgo<sup>6</sup>. Con frecuencia, estas iniciativas están motivadas por el cumplimiento de regulaciones que buscan mitigar el riesgo en lo que respecta tanto al ámbito español como al europeo (por ejemplo, Sarbanes-Oxley, Basilea I y II, Health Insurance

<sup>(1)</sup>Del inglés, *master data management*.

<sup>(2)</sup>Del inglés *data governance*.

<sup>(3)</sup>En inglés, *data quality*.

<sup>(4)</sup>En inglés, *data management*.

<sup>(5)</sup>En inglés, *business process management*.

<sup>(6)</sup>En inglés, *risk management*.

Portability and Accountability Act –HIPAA–). Las entidades financieras tienen experiencia en las mismas. También en este contexto confluyen prácticas de despliegue basadas en COBIT<sup>7</sup>, ISO/IEC, etc.

<sup>(7)</sup>Del inglés *control objectives for information and related technology*.

La gobernanza de datos busca responder a las preguntas siguientes.

- Exactitud: ¿los datos reflejan la realidad de la organización, o la de una fuente verificable?
- Integridad: ¿las relaciones entre los datos son consistentes?
- Consistencia: ¿hay una sola representación del dato?
- Completitud: ¿hay información clave perdida o desaparecida?
- Validez: ¿los datos están registrados en el formato correcto y son válidos?
- Accesibilidad: ¿los datos son accesibles de manera sencilla y comprensible y se utilizan consistentemente?
- Atemporalidad: ¿la información está registrada y es accesible a los sistemas de manera suficientemente rápida?

## 2. Variantes de la FIC

El concepto de almacén de datos es nuevo para muchas organizaciones. Normalmente, la primera toma de contacto de una organización con este suele ser mediante la presentación o bien de algún caso desarrollado en otras organizaciones, o bien de alguna herramienta de consulta o apoyo al desarrollo de almacenes de datos. Generalmente, en esta primera toma de contacto se ofrece una visión parcial de la arquitectura de la FIC<sup>8</sup>. A veces, esta visión inicial es la que predomina en la organización y la que se trata de implementar. En otros casos, se trata de simplificar la arquitectura de la FIC con el objetivo de conseguir de manera más rápida resultados tangibles.

<sup>(8)</sup>La que se ha estudiado en el módulo "La factoría de información corporativa" de esta asignatura.

Es frecuente que en las organizaciones se considere el concepto de almacén de datos bajo arquitecturas diferentes a la arquitectura de la FIC.

La implementación de estas arquitecturas produce resultados distintos, que revisaremos en este apartado.

### 2.1. Construcción de almacenes de datos departamentales

La mayoría de los analistas trabajan directamente sobre los almacenes de datos departamentales, es decir, estos son la "fachada" de la FIC para los analistas de información, lo que los usuarios ven. El resto de los componentes proporcionan los datos a los almacenes de datos departamentales y frecuentemente resultan invisibles para sus usuarios. Por lo tanto, es habitual que la percepción que los usuarios tienen de la FIC sea de manera exclusiva la ofrecida por los almacenes de datos departamentales.

#### Almacenes de datos operacional y corporativo

Algunos analistas también trabajan sobre el almacén de datos operacional y una minoría lo hace directamente sobre el almacén de datos corporativo. Recordad los tipos de usuarios presentados en el módulo "La factoría de información corporativa".

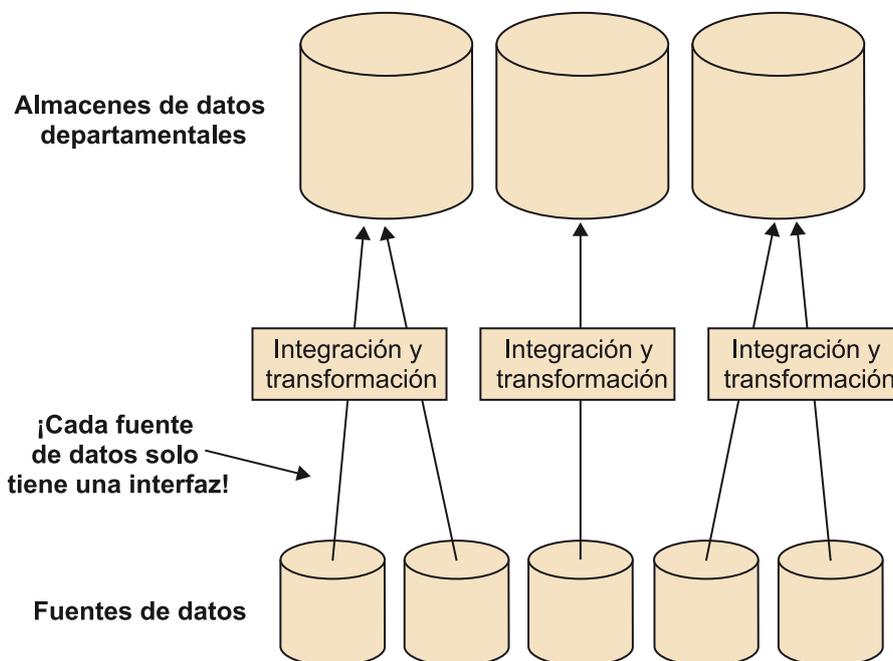
Esta percepción a veces es compartida por los desarrolladores, en algunos casos por desconocimiento. En muchos libros y otros materiales de formación, así como en presentaciones de vendedores de diferentes herramientas, el único componente que se estudia cuando se habla de almacenes de datos es el que corresponde a los almacenes de datos departamentales, y se ignora el resto de los componentes de la FIC.

En otros casos, aunque se conozca la arquitectura de la FIC, se ignora de manera consciente para perseguir resultados inmediatos y, aparentemente, construir exclusivamente almacenes de datos departamentales resulta más barato, fácil y rápido que construir la FIC.

Esto último es cierto cuando solo necesitamos un almacén de datos departamental o bien un reducido número de estos. Generalmente, esta es la manera de incorporar el concepto de almacén de datos a las organizaciones. Para empezar, se plantea la construcción de un almacén de datos departamental porque así podemos obtener resultados de manera inmediata. Partiendo de los datos de las fuentes de datos, los transformamos e integramos en un almacén de datos departamental diseñado según algún modelo multidimensional, de modo que sea fácil hacer consultas sobre este. Es decir, sin duda, la manera más rápida y barata de construir un primer almacén de datos departamental consiste en centrarse de manera específica en su construcción, en lugar de construir previamente la FIC. Además del almacén de datos departamental, también habremos construido un componente de integración y transformación específico para obtener los datos que almacenamos.

**Ved también**

En el módulo "Diseño multidimensional" de esta asignatura estudiaremos los detalles del diseño de los almacenes de datos departamentales.



Arquitectura resultado de la construcción de manera independiente de almacenes de datos departamentales independientes.

Una vez tenemos un almacén de datos departamental, no es extraño que usuarios de este u otros departamentos requieran la construcción de nuevos almacenes de datos para satisfacer sus necesidades específicas de información. Dado que los diferentes departamentos generalmente trabajan con datos distintos, aunque puedan compartir algunos, es frecuente que los proyectos siguientes se hagan de manera independiente del anterior: diseñamos el almacén de datos departamental y construimos un componente de integración y transformación según sus requerimientos en cada proyecto.

La construcción de almacenes de datos de manera independiente es el enfoque "natural" o intuitivo que se hace en muchas organizaciones, por desconocimiento de la FIC o, conociéndola, para tratar de ahorrar costes al comienzo. Es un enfoque válido cuando se trata de construir almacenes de datos departamentales totalmente independientes.

El problema surge cuando los almacenes de datos que se tienen que construir no son totalmente independientes, es decir, cuando hay datos comunes que deben incluirse en diferentes almacenes de datos departamentales. En estas situaciones, nos encontramos lo siguiente:

- 1) Los diferentes componentes de integración y transformación trabajan sobre las fuentes de datos comunes: generamos múltiples interfaces para los mismos datos. Esto representa un problema de coste en tiempo de desarrollo, mantenimiento y ejecución. Esto último se produce porque se accede diferentes veces a los mismos datos, una vez para cada almacén de datos departamental que los utiliza.
- 2) Cada almacén de datos ha podido hacer su interpretación de los mismos datos. Tenemos una falta de integración de los datos comunes en los diferentes almacenes de datos departamentales.

#### **Ejemplo de falta de integración de los datos**

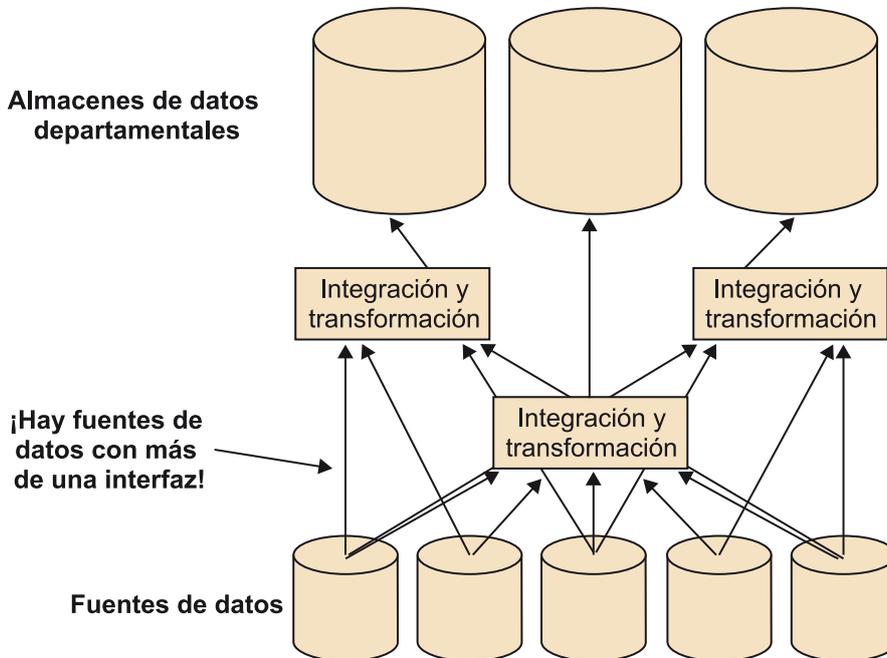
Un dato denominado beneficio puede interpretarse de manera distinta y tener significado diferente en cada uno de los almacenes de datos departamentales en los que se defina. En uno se puede tratar del beneficio antes de impuestos y en otro, una vez considerados los impuestos. Si comparamos el beneficio conseguido en cada uno de los departamentos de los respectivos almacenes de datos, podemos obtener resultados no reales o erróneos.

El problema de la falta de integración se acentúa cuando hay datos replicados entre las fuentes de datos, situación que se produce con frecuencia, y cada almacén de datos departamental hace su integración. En este caso, además de las diferentes posibilidades de interpretación de los datos de las fuentes, se tienen que añadir las diferencias que se puedan producir cuando se integran los datos de manera distinta en cada almacén de datos departamental.

#### **Diferentes interpretaciones en la integración de los datos de las fuentes**

En una organización disponemos de una aplicación de marketing con datos de clientes potenciales (clientes del pasado, clientes actuales y posibles clientes) y otra de facturación con los datos de los clientes actuales. En un almacén de datos departamental que necesita trabajar con los datos de los clientes, se ha partido de los datos de clientes en la aplicación de marketing y se han completado con los datos disponibles en la aplicación de facturación. Aun así, en otro almacén de datos departamental se ha procedido de manera inversa: se han tomado como base los datos de los clientes en la aplicación de facturación y se han completado con los datos que hay en la aplicación de marketing. El resultado global en los dos casos no tiene que coincidir; de hecho, es más probable que las bases de datos de clientes resultantes en los dos almacenes de datos departamentales sean muy distintas.

Por las razones mencionadas anteriormente, en muchas organizaciones el problema de construcción de la FIC se reduce al de construcción de manera independiente de un conjunto de almacenes de datos departamentales, a pesar de que haya dependencias entre estos. En estos casos, la arquitectura que se obtiene como resultado es la que se muestra en la figura siguiente.



Arquitectura resultado de la construcción, de manera independiente, de almacenes de datos departamentales con datos comunes.

Esta arquitectura puede resultar válida cuando se empieza a implantar el concepto de almacén de datos en la organización, cuando solo se dispone de unos cuantos almacenes de datos departamentales y estos son realmente independientes entre sí. Aun así, si hay dependencias entre los distintos almacenes de datos y estos son numerosos, por las razones que hemos comentado anteriormente, lo deja de ser. ¿Qué podemos hacer en esta situación?

Los problemas de integración y de multiplicidad de interfaces sobre las fuentes de datos se solucionan mediante la construcción del almacén de datos corporativo.

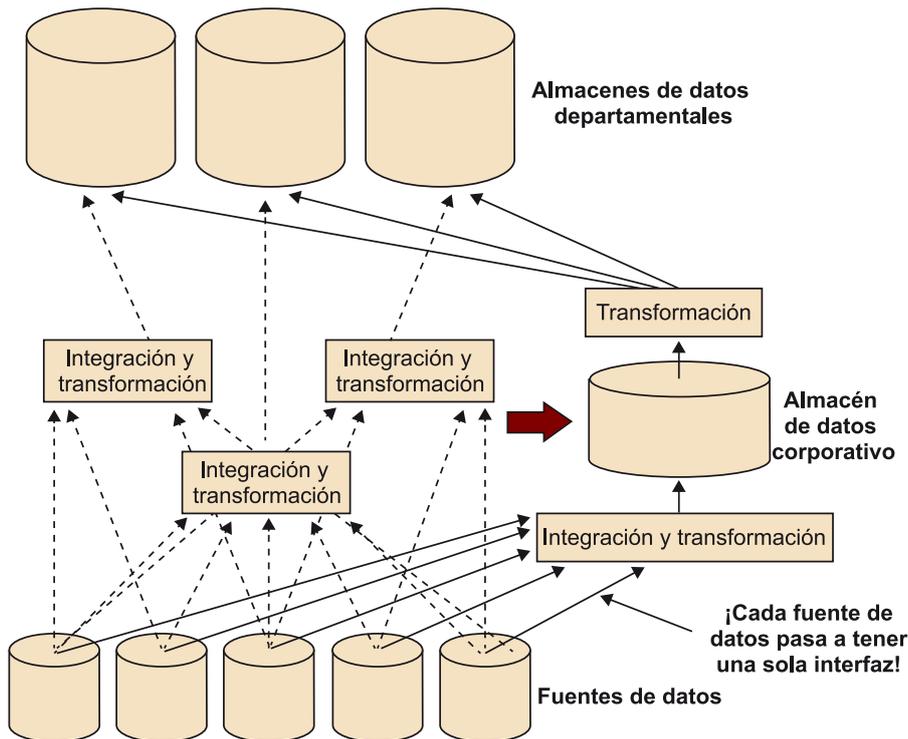
## 2.2. Construcción del almacén de datos corporativo *a posteriori*

Mediante la construcción del almacén de datos corporativo, orientado al almacenamiento de los datos más que a su consulta, se hace lo siguiente:

1) Se reduce el número de interfaces en las fuentes de datos: solamente es necesaria una interfaz para cada fuente para llevar los datos desde la fuente hasta el almacén de datos corporativo.

2) Los datos están integrados en el almacén de datos corporativo, y se evita tener múltiples interpretaciones de los mismos.

Si partimos de cero, la solución consiste en construir el almacén de datos corporativo de manera previa o al mismo tiempo que construimos los diferentes almacenes de datos departamentales. Aun así, si ya tenemos diferentes almacenes de datos departamentales construidos, tenemos que transformar la arquitectura basada exclusivamente en estos para incluir el almacén de datos corporativo.



Incorporación del almacén de datos corporativo a la arquitectura basada en almacenes de datos departamentales.

El problema para incluir el almacén de datos corporativo cuando ya tenemos diferentes almacenes de datos departamentales construidos radica en que tenemos que hacer un proceso de reingeniería sobre los diferentes componentes de integración y transformación, así como sobre los almacenes de datos departamentales construidos de manera previa, para adaptarlos a la nueva arquitectura. Esta transformación puede resultar muy compleja y costosa.

Particularmente, en lo que respecta a la transformación de los almacenes de datos departamentales, el principal problema que se presenta es el de integración de los datos: si se han construido de una manera no integrada, se les tiene que transformar y, en algunos casos, esta transformación en el ámbito técni-

co no es nada trivial. Además, tenemos el problema añadido de que los cambios hechos sobre los almacenes de datos departamentales son visibles para los usuarios finales.

La incorporación del almacén de datos corporativo en la arquitectura compuesta de manera exclusiva por almacenes de datos departamentales presenta problemas técnicos y de gestión, estos últimos a causa de que los cambios afectan a datos sobre los que trabajan directamente los usuarios finales.

#### **Corrección de la falta de integración de los datos en los almacenes de datos departamentales**

Si el dato "beneficio" del ejemplo utilizado anteriormente, en uno o diferentes almacenes de datos departamentales, se interpreta como "beneficio antes de impuestos" y la interpretación aceptada en el ámbito corporativo es del "beneficio una vez considerados los impuestos", la nueva interpretación se tiene que incorporar a los almacenes de datos departamentales que no la considerasen de este modo. Por lo tanto, los usuarios deben adaptarse al nuevo significado a la hora de generar los informes que necesiten.

Debido a los problemas que surgen al construir el almacén de datos corporativo *a posteriori* respecto a los almacenes de datos departamentales, resulta más conveniente construirlo previamente o de manera conjunta con estos. Es decir, es más adecuado planificar desde un principio la construcción de la FIC.

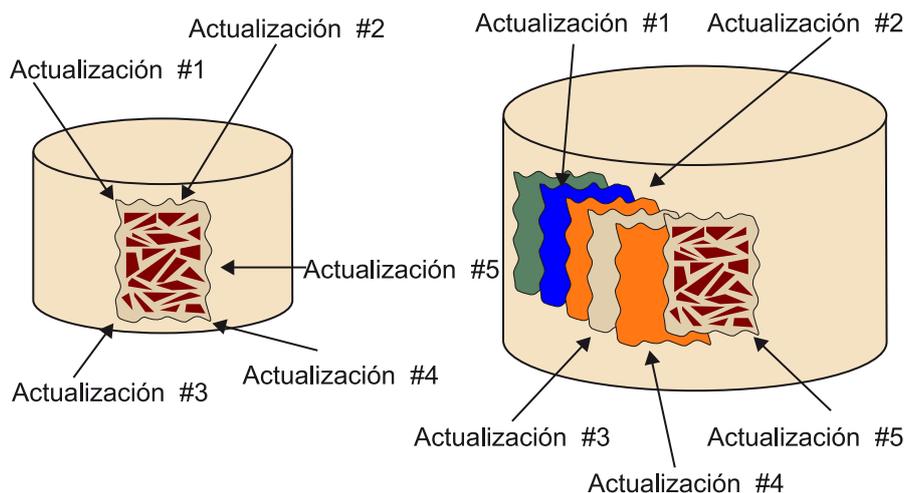
### **2.3. Combinación del almacén de datos operacional y el almacén de datos corporativo**

El almacén de datos operacional y el almacén de datos corporativo tienen características similares: los dos son orientados al tema e integrados. Se diferencian en el hecho de que los datos del almacén de datos operacional son volátiles y no históricos, mientras que los datos del almacén de datos corporativo son no volátiles e históricos.

El almacén de datos operacional almacena una imagen actualizada de los datos integrados de la organización. El almacén de datos corporativo almacena una película formada a partir de las diferentes imágenes de los datos de la organización.

#### **Ved también**

Podéis encontrar más detalles sobre las estructuras de los dos en el módulo "La factoría de información corporativa" de esta asignatura.



**Almacén de datos operacional**

**Almacén de datos corporativo**

Representación de los datos del almacén de datos operacional y del almacén de datos corporativo.

¿Podemos combinar la construcción de los dos almacenes de datos en una única estructura? En teoría sí que lo podríamos hacer, pero en la práctica no es conveniente. La razón principal para no combinarlos es que tienen que estar diseñados para soportar diferentes tipos de operaciones, hechas por tipos de usuarios también distintos.

El objetivo principal del almacén de datos operacional es mantener una visión integrada y totalmente actualizada de los datos operacionales. Sobre estos, se ejecutarán muchas operaciones de actualización y consultas simples llevadas a cabo por oficinistas principalmente. Su diseño y configuración están orientados para que este tipo de operaciones se hagan de una manera óptima.

Por otro lado, el almacén de datos corporativo no requiere que los datos que almacena estén totalmente actualizados: basta con que estén actualizados según las necesidades de los analistas que son sus usuarios (en algunos casos, bastará con que se actualicen de manera semanal o mensual). Está especialmente diseñado para que las actualizaciones se almacenen como imágenes nuevas para ir formando una película de los datos. Está configurado para optimizar las consultas de las imágenes de los datos, no para hacer modificaciones sobre estos.

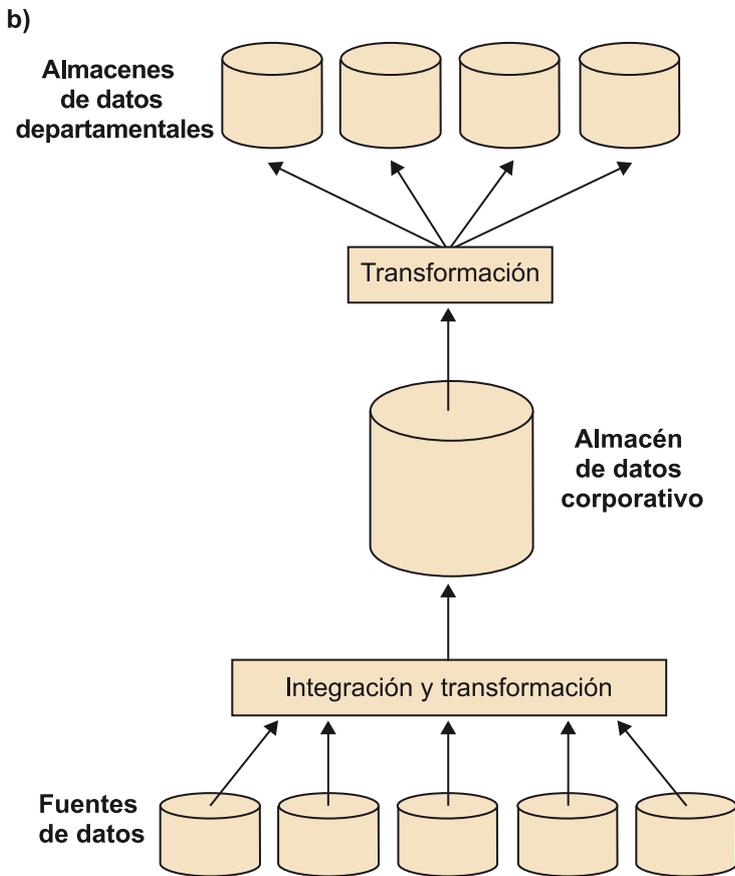
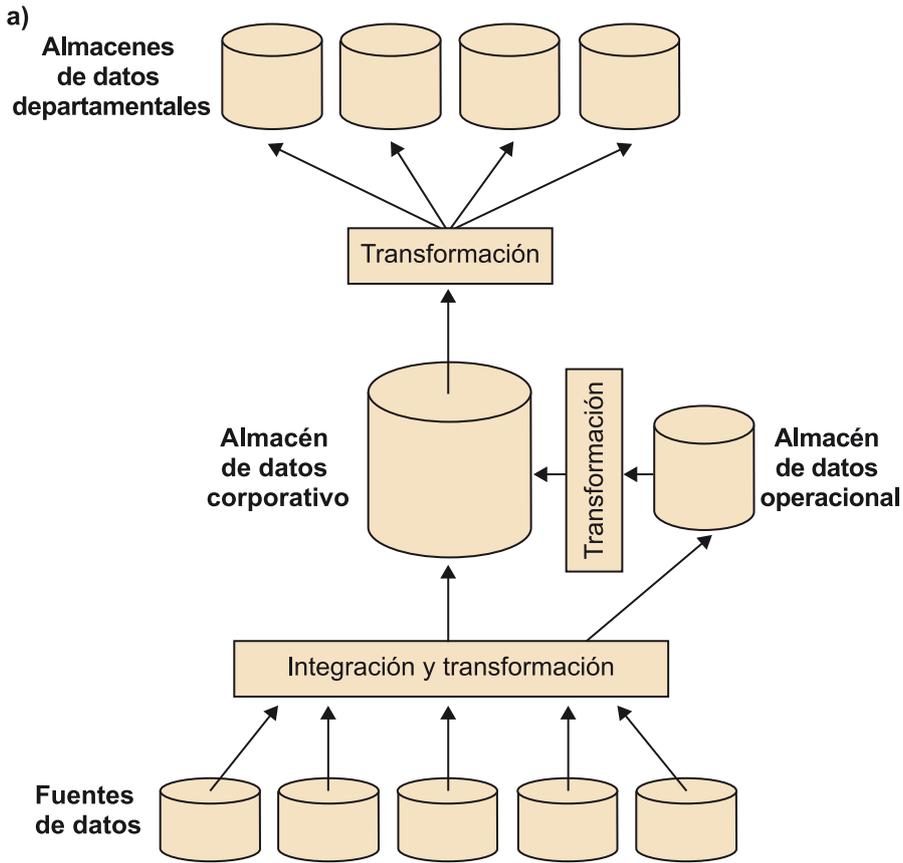
**Almacén de datos corporativo**  
 Se reconfigurará especialmente para la carga de datos cuando haya nuevas imágenes disponibles para almacenar.

Si combinamos las dos estructuras, el resultado será una base de datos con muchos registros (tenemos que guardar la historia almacenada en el almacén de datos corporativo), que debe estar configurada para hacer modificaciones sobre los datos (algo requerido por el almacén de datos operacional). De este modo, cualquier transacción será muy costosa, puesto que se pueden combinar consultas complejas con actualizaciones, y el volumen de datos que deben tratar las diferentes transacciones será muy grande porque tendrá que considerar los datos del almacén de datos corporativo.

El almacén de datos operacional y el almacén de datos corporativo tienen objetivos diferentes y están diseñados de manera óptima para conseguirlos. Si los combinamos en una estructura común, se degradará el tiempo de respuesta para las dos funcionalidades.

#### **2.4. La FIC sin el almacén de datos operacional**

El almacén de datos corporativo contiene todos los datos que necesitan los analistas, que se obtienen directamente a partir de las fuentes de datos operacionales y también a partir del almacén de datos operacional, que a su vez los obtiene de las fuentes de datos operacionales (podéis ver la figura *a*). El proceso de integración y transformación de los datos de las fuentes operacionales para incluirlos en el almacén de datos corporativo o en el almacén de datos operacional es común a los dos, y por este motivo es suficiente con transformar los datos del almacén de datos operacional para adaptarlos a las estructuras del almacén de datos corporativo.



Separación del almacén de datos operacional y el almacén de datos corporativo.

A diferencia del almacén de datos corporativo, el almacén de datos operacional no es estrictamente necesario en la FIC. Es decir, aunque conviene disponer del mismo por las funcionalidades que proporciona a los usuarios, y también para permitir construir más fácilmente el almacén de datos corporativo, no se necesita de manera estricta para construir el almacén de datos corporativo ni tampoco para construir los almacenes de datos departamentales, que son los que proporcionan la funcionalidad principal a los analistas (podéis ver la figura b).

Las dos arquitecturas representadas en la figura anterior son igualmente válidas, aunque la arquitectura de la figura *b* ofrece menos funcionalidades que la de la figura *a*.

El almacén de datos operacional es una estructura opcional en la FIC para proporcionar información a los analistas.

A pesar de que el almacén de datos operacional es opcional en la FIC, en algunas organizaciones será más necesario que en otras dependiendo de distintos factores, entre los cuales destacan los siguientes:

- El tamaño de la organización: cuanto más grande sea la organización, más probable es que necesite el almacén de datos operacional. En las organizaciones pequeñas, los problemas debidos a la falta de integración de los datos operacionales suelen ser menores.
- La naturaleza de los negocios: si la organización necesita acceder de manera inmediata a información integrada y actualizada -por ejemplo, porque interacciona directamente con los clientes o porque la necesita en el proceso de fabricación-, es muy probable que necesite el almacén de datos operacional.
- Finalmente, depende de la medida de las aplicaciones operacionales y del grado de integración que hay entre estas. En una organización con un conjunto pequeño de aplicaciones operacionales que están muy integradas, el almacén de datos operacional será menos necesario que en otra con gran cantidad de aplicaciones poco integradas.

#### Ved también

Podéis ver el módulo "La factoría de información corporativa" de esta asignatura para más detalles sobre este tema.

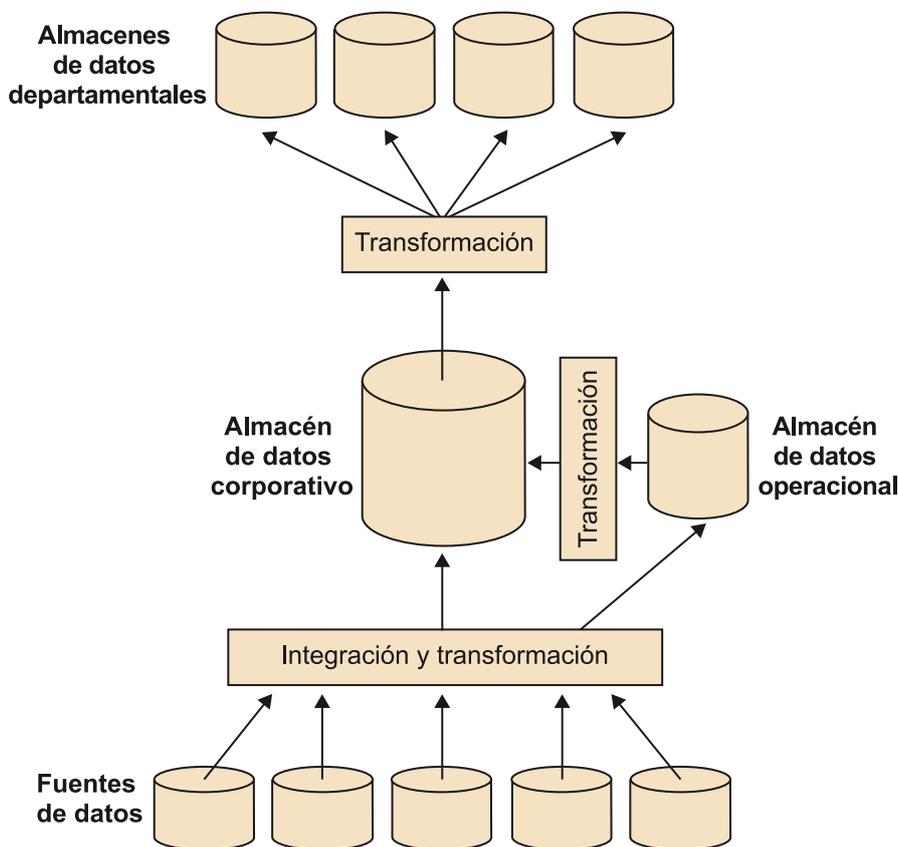
#### Lectura recomendada

Para más detalles, podéis ver **W. H. Inmon** (1999). *Building the Operational Data Store*. EE. UU.: John Wiley & Sons, Inc.

### 3. Estrategias para la construcción de la FIC

Una vez estudiados los apartados anteriores de este módulo, podemos llegar a la conclusión de que la organización necesita implementar la FIC para ofrecer a los analistas el apoyo de información que necesitan.

Conocer la arquitectura de la FIC representada en la figura siguiente representa un avance considerable respecto a la situación en la que solo se percibe la existencia de almacenes de datos departamentales, o bien se plantean arquitecturas alternativas. Aun así, no basta con conocer esta arquitectura, sino que se tiene que disponer de una estrategia adecuada para implementarla.



**Los metadatos en la figura**

En esta figura, aunque no se muestran de manera explícita, recordad que los metadatos aparecen en todos los componentes de la FIC. Recordad también que el almacén de datos operacional es un elemento opcional de la arquitectura.

Arquitectura de la factoría de información corporativa.

En este apartado, estudiaremos diferentes estrategias que se suelen aplicar en las organizaciones para la construcción de la FIC. Algunas de estas solo ofrecen una solución parcial, y otras abordan todos los problemas de manera conjunta.

### 3.1. Construcción de la FIC mediante un solo proyecto

Si una organización reconoce la necesidad de la FIC, generalmente considera que necesita todos sus componentes y, además, que los necesita de manera inmediata. En esta situación, la tendencia habitual es plantear la construcción de la FIC mediante un solo proyecto. El problema principal que plantea este proyecto es la complejidad.

Generalmente, en este entorno los analistas todavía no conocen claramente las funcionalidades que les pueden ofrecer los almacenes de datos departamentales. La idea inicial que tienen sobre sus necesidades de información cambia cuando descubren las posibilidades que les ofrecen los almacenes de datos. En estas condiciones, difícilmente pueden transmitir sus requerimientos a los desarrolladores de la FIC. Por otro lado, localizar los datos que se necesitan en las aplicaciones operacionales no es una tarea fácil. Si además se pretende desarrollar todo el conjunto de la FIC en un solo proyecto, el resultado es un proyecto de dimensiones y complejidad demasiado grandes en comparación con los que se suelen desarrollar en las organizaciones.

Por otro lado, precisamente debido a la complejidad del proyecto, se hace muy difícil justificar el coste de desarrollo de la FIC según el beneficio que aporta a la organización, puesto que los costes son muy grandes y el beneficio, aunque está claro, no se ha evaluado lo suficiente.

El resultado de plantear el desarrollo de la FIC como un solo proyecto es un proyecto muy grande y complejo, cuyos objetivos no están totalmente claros y cuyos costes son difícilmente justificables en lo que respecta a la organización. En esta situación, es más probable que sea un fracaso.

### 3.2. Enfoques parciales para la construcción de la FIC

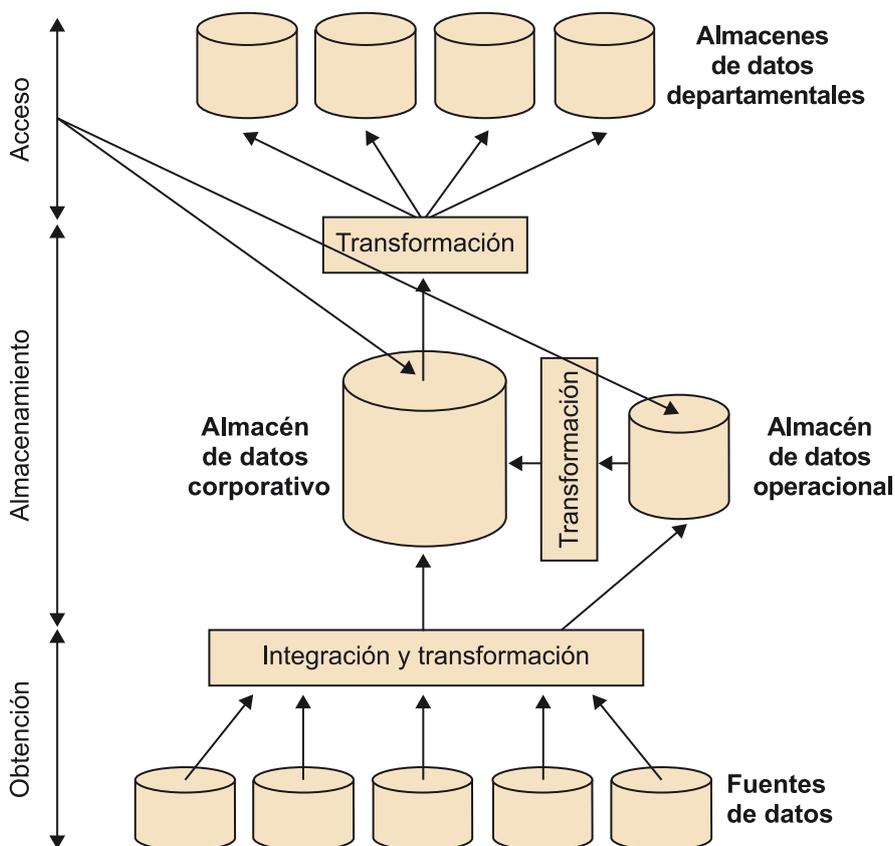
Si construir la FIC en un solo proyecto es demasiado complejo, para manejar la complejidad, una solución es dividir la arquitectura de la FIC en secciones según su funcionalidad y tratar de construirlas por separado.

En cuanto a la funcionalidad, la arquitectura consta de tres partes.

- Obtener los datos: el componente principal de esta parte es el elemento de integración y transformación. El objetivo principal consiste en obtener los datos que circulan a lo largo de la FIC.
- Almacenar los datos: los datos se almacenan principalmente en el almacén de datos corporativo y en el almacén de datos operacional. Los almacenes de datos departamentales también almacenan, pero todos estos se

obtienen a partir del almacén de datos corporativo y esta no es su función principal.

- Acceder a los datos: los analistas acceden a los datos principalmente desde los almacenes de datos departamentales. Los analistas también pueden acceder directamente a los datos desde el almacén de datos corporativo y desde el almacén de datos operacional, pero un mayor número de analistas acceden a los mismos exclusivamente a partir de los almacenes de datos departamentales. Los almacenes de datos departamentales almacenan datos, pero su único objetivo es facilitar su acceso, puesto que todos estos se obtienen a partir del almacén de datos corporativo.



División de la arquitectura de la FIC según su funcionalidad.

Es frecuente que por diferentes motivos se hagan intentos de construir solo partes de la arquitectura de la FIC. En este apartado, estudiaremos distintos enfoques en este sentido.

### 3.2.1. Enfoque centrado en el almacenamiento y acceso a los datos

Si analizamos los diferentes componentes de la FIC, los que resultan realmente nuevos para los equipos desarrolladores de las organizaciones son los que hacen referencia a los diferentes almacenes de datos (departamentales, corporativo y operacional). El componente de integración y transformación no deja

de ser un conjunto de programas que leen los datos de los sistemas operacionales que hay, que están bajo el control de nuestra organización, y los mueven al almacén de datos corporativo y/o al almacén de datos operacional.

Por lo tanto, si nuestra organización considera que necesita ayuda, es posible que alguna empresa consultora se ofrezca para ayudarnos, especialmente en los componentes que nos resulten más nuevos, los que hacen referencia a los distintos almacenes de datos. De este modo, es frecuente dividir la construcción de la FIC en dos partes y asignar su desarrollo a nuestra organización o a alguna empresa consultora según la experiencia en cada una de las partes.

- Almacenar y acceder a los datos: construcción de los diferentes almacenes de datos, generalmente a cargo de una empresa consultora. Aunque no necesariamente tiene que ser así, es frecuente.
- Obtener los datos: a cargo de nuestra organización, ya que es la especialista en las fuentes de datos operacionales a partir de las que se han de obtener los datos para los diferentes almacenes de datos.

Es decir, por un lado, construimos la FIC como si se tratara de un sistema nuevo (creamos un conjunto de bases de datos vacías y un conjunto de aplicaciones que trabajan con las mismas) y, por el otro, construimos las aplicaciones que se encargan de llenar las bases de datos con los datos que hay en los sistemas operacionales.

De este modo, podemos intentar desarrollar este proyecto como un proyecto más en la organización y aplicar, para esto, el ciclo de vida tradicional o en cascada<sup>9</sup>, también denominado SDLC<sup>10</sup>.

<sup>(9)</sup>Se denomina ciclo de vida en cascada porque cada fase es continuación de la anterior y se basa en esta.

Las fases habituales del ciclo de vida en cascada son las siguientes:

<sup>(10)</sup>Del inglés *systems development life cycle*.

- Definición de requerimientos.
- Análisis.
- Diseño.
- Programación.
- Prueba.
- Puesta en producción.

#### **Ved también**

Podéis encontrar más información sobre el ciclo de vida en cascada o SDLC en las asignaturas *Ingeniería del software I y II*.

Generalmente, no se emplea mucho tiempo en la fase de definición de requerimientos. Es evidente que necesitamos construir la FIC, aunque no tenemos claro ni el coste ni el beneficio que representará para la organización.

Del mismo modo que se llevaría a cabo para cualquier otro proyecto, en la fase de análisis se tienen reuniones con los analistas de los diferentes departamentos para determinar las necesidades de información. En estas reuniones,

es frecuente definir las pantallas y los campos que necesitan los analistas. Como resultado de las mismas, es frecuente que se creen expectativas en los departamentos sobre lo que la FIC les ofrecerá.

Partiendo de las necesidades recopiladas, se diseña el almacén de datos corporativo y los diferentes almacenes de datos departamentales. Se construyen los programas para transformar los datos del almacén de datos corporativo y adaptarlos a la estructura de los almacenes de datos departamentales.

Los almacenes de datos departamentales generalmente se implementan con una herramienta de consulta según algún modelo multidimensional. Para completar esta parte del proyecto, basta con dar a los analistas la herramienta de consulta elegida y formarlos en la misma.

**Ved también**

Podéis ver el módulo titulado "Diseño multidimensional" de esta asignatura.

Para acabar el proyecto, "solo" faltaría la parte de obtención de los datos, que generalmente está a cargo de nuestra organización. Cuando nos enfrentamos a esta parte podemos encontrar, entre otros, los problemas siguientes para obtener los datos de las fuentes:

- Los datos tienen muchos significados en las fuentes.
- Los datos no tienen integridad.
- Hay muchas fuentes potenciales para los datos requeridos.
- Algunos datos requeridos por los usuarios no están en el sistema.
- No está almacenada toda la historia que se quiere.
- No se pueden obtener las modificaciones periódicas para refrescar los datos en los almacenes de datos.

En este punto del proyecto, tenemos los diferentes almacenes de datos contruidos y solo nos falta incluirles los datos. Aun así, esto no lo podemos solucionar de una manera inmediata. El resultado es que si no podemos obtener los datos, el proyecto resulta un fracaso para la organización.

Además del fracaso global del proyecto, esta situación tiene la agravante de la imagen que se ofrece a los departamentos de nuestra organización por parte de los equipos de desarrollo internos. Si el proyecto se ha planteado como se ha pensado, para ser desarrollado por nuestra organización y una empresa consultora que nos ayude, en lo que respecta a los analistas de los departamentos la parte desarrollada por la empresa consultora habrá sido un éxito. Han hecho tareas idénticas a las que se hacen para cualquier otro proyecto con toda la normalidad y, además, han podido ver cómo funcionan las herramientas de consulta. Para ellos, lo que ha fallado en el desarrollo del proyecto es el equipo de desarrollo encargado de obtener los datos, es decir, el equipo de desarrollo interno.

Además, si hemos hablado con los analistas de diferentes departamentos y se les ha creado expectativas respecto a las funcionalidades del nuevo sistema, el fracaso del proyecto puede tener una gran repercusión en toda la organización.

Uno de los problemas más importantes que hay en la construcción de la FIC es el de la obtención de los datos, es decir, la localización de los datos necesarios y la construcción del componente de integración y transformación. Este es uno de los motivos más frecuentes de fracaso de los proyectos de construcción de la FIC.

En esta solución, aunque tenemos definida una posible estructura de los diferentes almacenes de datos, estos no pueden ser utilizados por los usuarios porque no disponen de los datos necesarios.

### **3.2.2. Enfoque centrado en la obtención y el almacenamiento de los datos**

Este enfoque se produce generalmente como reacción al fracaso producido en el momento en que se aplica el enfoque anterior, aunque no tiene que ser necesariamente así. Si hay problemas para construir lo que los usuarios necesitan, la respuesta más frecuente es la de construir lo que se pueda y a partir de aquí, que los analistas extraigan la información que sea posible.

En este caso, el objetivo es diseñar y crear el almacén de datos corporativo (y el almacén de datos operacional si así se considera), y también el componente de integración y transformación que lo cargue con todos los datos que se puedan obtener desde las fuentes de datos operacionales. A partir de los datos del almacén de datos corporativo, se construirán los almacenes de datos departamentales que los analistas consideran necesarios.

Este enfoque es muy complejo, puesto que muchos de los datos de los sistemas operacionales no serán necesarios, al menos de manera inminente, para construir los almacenes de datos departamentales. Por otro lado, hay muchas posibilidades para obtener las imágenes de los datos; según este planteamiento, siempre deberíamos obtener los datos tan detallados como resulte posible, aunque nunca se necesiten con este nivel de detalle.

Los problemas principales de centrarse en la obtención y el almacenamiento de los datos son el ámbito y la complejidad del proyecto: se trabaja con todos los datos de la organización con el máximo nivel de detalle.

Se necesita una gran inversión y mucho tiempo para disponer del almacén de datos a partir del cual podamos crear almacenes de datos departamentales directamente utilizables por los analistas. El coste de este proyecto resulta difícilmente justificable pensando en la organización. Por otro lado, es posible que el resultado obtenido mediante este enfoque no coincida con las necesidades de los analistas.

### 3.3. Construcción de la FIC mediante proyectos autónomos

En apartados anteriores, hemos visto que intentar construir la FIC como un solo proyecto es demasiado complejo. Por otro lado, si dividimos la arquitectura de la FIC horizontalmente, según la funcionalidad de los componentes, no se reduce la complejidad tanto como sería deseable y, además, es difícil satisfacer los requerimientos de los analistas de información.

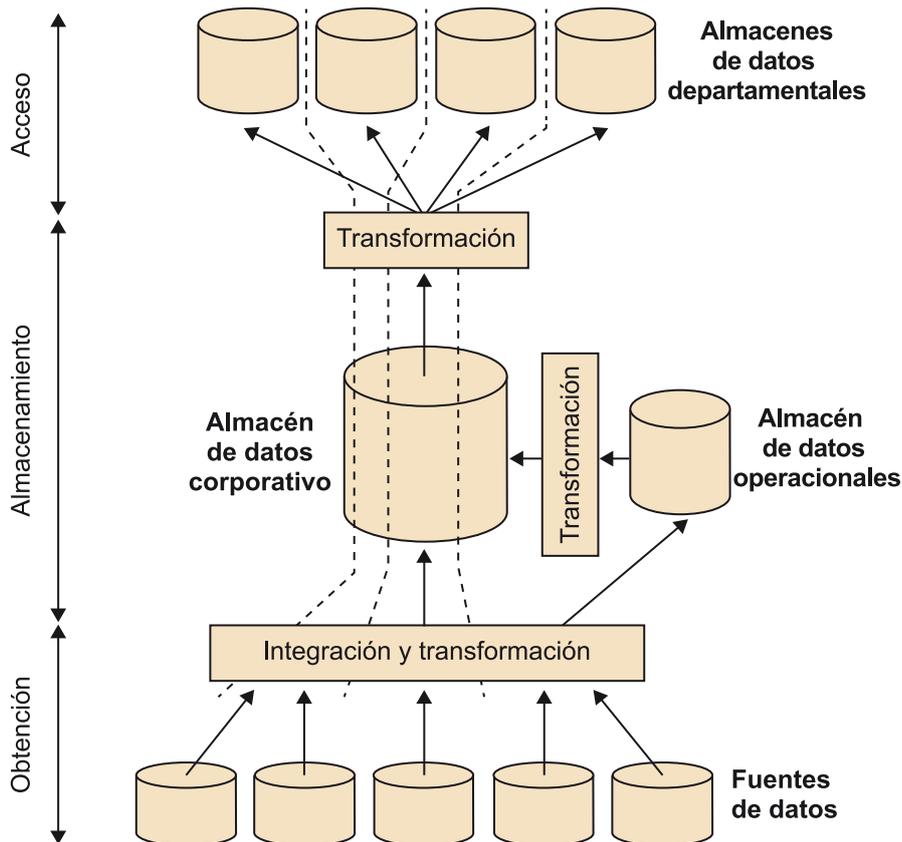
Por lo tanto, otra posibilidad consiste en dividir la arquitectura de la FIC de manera vertical, es decir, dividir la construcción de la FIC en forma de proyectos de modo que hagan lo siguiente.

- Que proporcionen un valor por sí mismos a la organización para que su coste sea justificable.
- Que sean completos y autónomos en la medida en que se pueda: que no necesiten otros proyectos para entrar en producción.

#### Lectura recomendada

Podéis encontrar más detalles en la obra siguiente:

**R. Mattison** (1996). *Data Warehousing: Strategies, Technologies and Techniques*. EE. UU.: Computing McGraw-Hill.



División de la arquitectura de la FIC en forma de proyectos completos y autónomos.

En la figura anterior, se muestra una representación de la estructura de los mencionados proyectos. Es decir, en cada proyecto se tienen que construir los componentes siguientes.

- El componente de acceso: o bien el almacén de datos departamental con los datos que necesitan los analistas, o bien la herramienta de acceso al almacén de datos corporativo o al almacén de datos operacional.
- El componente de almacenamiento: la parte del almacén de datos corporativo y/o almacén de datos operacional que contiene los datos a los que se necesita acceder, y que todavía no ha sido construida por ningún proyecto anterior.
- El componente de obtención de los datos: el componente de integración y transformación que obtienen los datos necesarios a partir de las fuentes de datos.

Cada proyecto consta de los componentes de acceso, almacenamiento y obtención de los datos de las fuentes de datos, es decir, se trata de un proyecto completo que puede entrar en funcionamiento de manera independiente del resto de los proyectos existentes.

De este modo, cada proyecto se intenta satisfacer los requerimientos de un grupo de analistas de manera independiente, de modo que sea justificable el coste del proyecto según los beneficios que aporta y se reduzca la complejidad del desarrollo de la FIC.

### **3.3.1. El primer proyecto: proyecto global de desarrollo**

En un primer momento, la visión que se tiene de la FIC es dispersa. Conocemos su arquitectura y, además, que su aportación como soporte de información para los analistas es positiva para la organización. Aun así, no conocemos todavía con detalle los almacenes de datos departamentales que se necesitan en la organización, ni lo que aportarán a esta.

El objetivo del primer proyecto es dividir la construcción de la FIC en proyectos, de modo que cada proyecto haga lo siguiente.

- Que corresponda a un problema concreto.
- Que tenga a un responsable en la organización: el analista que utilizará el sistema resultado del proyecto.
- Que ofrezca un beneficio tangible a la organización: podemos conocer cuál será el coste del proyecto y el beneficio que este aportará a la organización.

Generalmente, cada proyecto definido se corresponderá con un almacén de datos departamental que se tiene que desarrollar, su responsable será el analista (o uno de los analistas) que utilizará el almacén de datos departamental, y el beneficio se calculará según los objetivos que se pretendan conseguir con el almacén de datos desarrollado y su coste estimado de desarrollo.

Para definir los objetivos del proyecto, deberemos mantener reuniones con los analistas de información, que serán sus usuarios y responsables del proyecto. Aun así, no bastará con las reuniones con los usuarios. De manera adicional, para que la definición del proyecto sea realista, nos tendremos que asegurar de que los datos requeridos por los usuarios están en la organización, en las fuentes de datos operacionales, y de que tienen el grado de detalle y la calidad requeridos. Es decir, deberemos hacer una revisión global de todos los componentes de cada proyecto: obtener, almacenar y acceder a los datos requeridos por los usuarios.

De manera conjunta con la definición de los proyectos que se tienen que desarrollar, deberemos determinar las necesidades de infraestructura de estos y un plan para llevarlos a cabo.

Los resultados del proyecto global de desarrollo son el conjunto de proyectos que se tienen que desarrollar, los requerimientos de infraestructura de los proyectos y un plan de desarrollo que incluya el presupuesto, la planificación temporal y el enfoque para desarrollar los proyectos.

Es decir, el objetivo del primer proyecto es transformar la visión dispersa que se tiene de la FIC en una visión concreta en forma de proyectos, de modo que cada uno de estos satisfaga unos objetivos determinados y aporte un valor concreto a la organización.

Este primer proyecto es el más importante en el desarrollo de la FIC, puesto que el resto de los proyectos dependen del mismo. Aun así, generalmente resulta muy difícil justificar su ejecución en la organización: es preciso plantear un proyecto de estudio que tiene que analizar toda la organización y cuyo beneficio será reflejado exclusivamente por el éxito de los futuros proyectos que se desarrollen. Disponer de un patrocinador con suficiente nivel en la organización facilita el desarrollo del primer proyecto y de la FIC.

Los beneficios del proyecto global de desarrollo no son inmediatos, y por este motivo es más fácil de justificar y llevar a cabo si disponemos de un patrocinador en la organización que ofrezca soporte al desarrollo de la FIC.

### 3.3.2. Proyectos de desarrollo de infraestructura

Junto a los proyectos de desarrollo de almacenes de datos departamentales (compuestos por la parte de obtención y almacenamiento de los datos y acceso a estos), en el proyecto global de desarrollo se ha definido la infraestructura común en la que se basan. Parte de la infraestructura se puede desarrollar como un componente de cada uno de los proyectos. Aun así, por su envergadura e importancia, interesa desarrollar otras partes como proyectos independientes.

Se distinguen dos tipos de infraestructura: infraestructura física e infraestructura operacional.

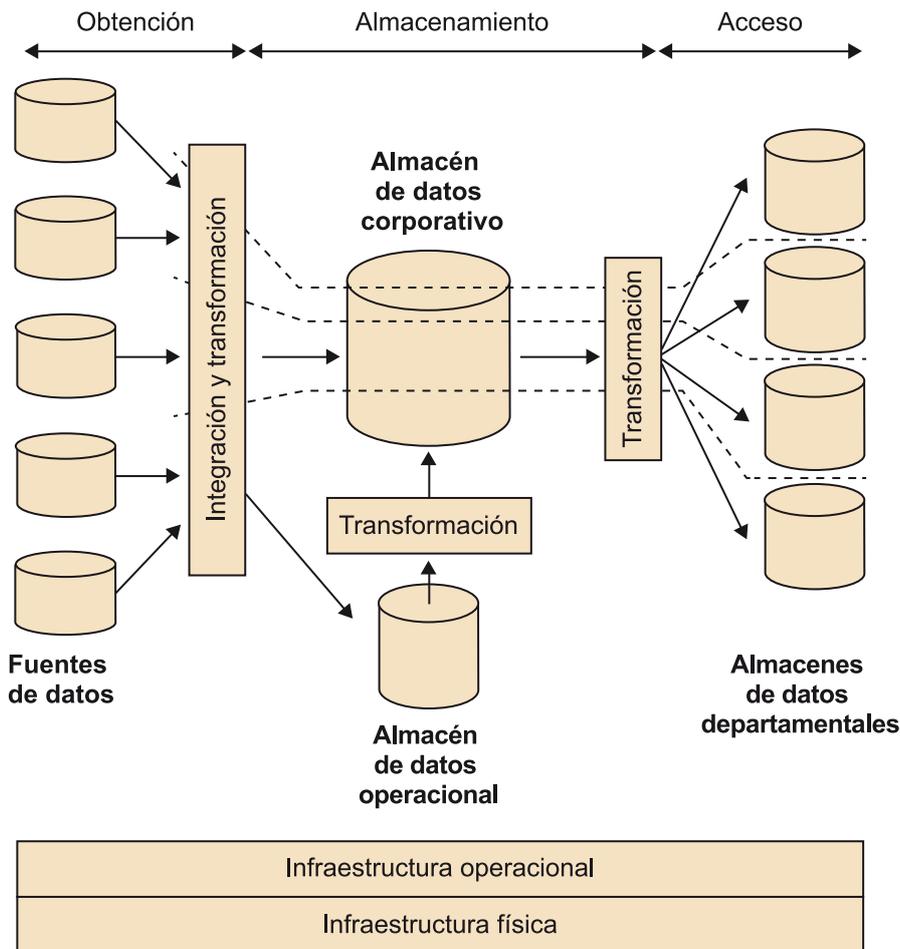
- La infraestructura física se refiere al conjunto de máquinas, redes que las interconectan, sistemas operativos, software de red, utilidades y personal necesario para que todo esto funcione, que necesitarán los proyectos para su funcionamiento.
- La infraestructura operacional está formada por el conjunto de procedimientos, software de gestión, herramientas de apoyo y personal necesario sobre los que se basan los diferentes proyectos.

El objetivo de los proyectos de desarrollo de infraestructura consiste en simplificar, dar coherencia y acortar la duración del resto de los proyectos.

**Infraestructura del componente de integración y transformación**

El componente de integración y transformación se puede basar en alguna herramienta de apoyo; de hecho, es conveniente que sea así para simplificar su desarrollo y estructura. Durante el proyecto de desarrollo de infraestructura como parte de la infraestructura operacional se determinará qué herramienta se debe utilizar y se instalará y configurará para que sea directamente utilizable durante el desarrollo del resto de los proyectos.

**Ved también**  
 Podéis ver el apartado "Desarrollo del componente de integración y transformación" de este módulo para más detalles.



Infraestructura de la FIC.

La infraestructura, tanto la operacional como la física, es el fundamento sobre el cual se basan el resto de los proyectos de construcción de la FIC.

La clave en los proyectos de construcción de infraestructura es la **escalabilidad**, es decir, la posibilidad de que, en caso de que se necesite, se pueda ampliar la infraestructura sin problemas. En los proyectos de construcción de infraestructura, se crea una infraestructura base que posteriormente se ampliará con el desarrollo de proyectos posteriores en cuanto aumenten las necesidades en este sentido.

La escalabilidad de la infraestructura física y operacional es muy importante, puesto que, por un lado, permite que la FIC crezca más allá de las previsiones iniciales que pudiéramos tener y, por otro, nos permite hacer las inversiones necesarias para la construcción de la FIC de manera progresiva. De este modo, inicialmente bastaría con disponer de la infraestructura mínima para las previsiones iniciales de funcionamiento. En cuanto se desarrollan nuevos proyectos, la infraestructura inicial puede crecer según las nuevas necesidades.

En lo que respecta a aspectos concretos de la infraestructura física, como son las máquinas que soportarán los diferentes almacenes de datos, el uso de máquinas de arquitecturas paralelas facilita que estas sean escalables y que puedan aumentar la cantidad de datos almacenados o el número de usuarios sin que se resienta su tiempo de respuesta, eso sí, aumentando las prestaciones de la máquina (pero sin necesidad de cambiarla).

Hay dos arquitecturas representativas.

- *SMP (symmetric multi processing)*: en esta arquitectura, un conjunto de procesadores comparten memoria común como medio de comunicación entre estos.
- *MPP (massively parallel processors)*: en esta arquitectura, distintas plataformas compuestas por procesador y memoria se interconectan mediante una línea de alta velocidad.

Actualmente, la frontera entre las dos arquitecturas se está disipando, puesto que la tecnología de interconexión permite compartir memoria entre distintas plataformas. De todos modos, se continúa hablando de mejores prestaciones de una arquitectura u otra en determinadas condiciones. Particularmente, la arquitectura MPP se considera más escalable y adecuada para trabajar con volúmenes de datos más grandes que la arquitectura SMP. Según esto, el almacén de datos corporativo será candidato a ser construido sobre una máquina con arquitectura MPP. Teniendo en cuenta la velocidad de evolución del hardware, conviene hacer un estudio en profundidad de las prestaciones de las plataformas disponibles en cada momento antes de tomar una decisión sobre ello. En cualquier caso, sea cual sea el tipo de plataforma seleccionada, uno de los aspectos principales que hay que tener en cuenta es que esta deberá ser escalable.

Cuando seleccionamos las plataformas para los diferentes almacenes de datos, tendremos que estudiar las prestaciones de las máquinas del mercado. El criterio principal que es necesario tener en cuenta a la hora de adquirir una plataforma es su escalabilidad.

En cuanto a los sistemas de gestión de bases de datos (SGBD), también podemos encontrar sistemas especialmente orientados para cada una de las arquitecturas mencionadas, que intentan aprovecharlas al máximo, así como sistemas que funcionan en las dos arquitecturas. Por otro lado, particularmente para los almacenes de datos departamentales, en el mercado SGBD algunos implementan directamente el modelo multidimensional y también hay SGBD relacionales que ofrecen apoyo multidimensional. Generalmente, tanto para el almacén de datos corporativo como para los almacenes de datos departamentales podremos usar un SGBD que configuraremos especialmente para que sea óptimo en la ejecución de las consultas, puesto que estas serán las operaciones que se harán principalmente. En el caso del almacén de datos operacional, el SGBD estará configurado como para cualquier otro sistema operacional para que soporte operaciones de modificación.

**Ved también**

Podéis encontrar más detalles en el módulo "Diseño multidimensional" de esta asignatura.

Podemos utilizar un SGBD de ámbito general para implementar los diferentes almacenes de datos. Para implementar el almacén de datos corporativo y los almacenes de datos departamentales, deberemos configurar sus respectivos SGBD para que sean óptimos en la ejecución de consultas, puesto que no se harán operaciones de modificación sobre estos.

### 3.3.3. Desarrollo de proyectos autónomos

Mediante el desarrollo del primer proyecto, el proyecto global de desarrollo, hemos dividido el desarrollo de la FIC en forma de proyectos con las características siguientes.

- Solucionan un problema concreto.
- Son autónomos: son proyectos independientes del resto de los proyectos pendientes de desarrollo, se pueden basar en proyectos desarrollados de manera previa y utilizan la infraestructura construida mediante los proyectos de desarrollo de infraestructura.
- Están compuestos por los elementos de obtención y almacenamiento de los datos y acceso a los mismos.
- Tienen a un responsable dentro de la organización.
- Proporcionan un valor real: se ha analizado su coste y el beneficio que aportan a la organización.
- Tienen una duración controlada: no son proyectos de una medida desmesurada. Generalmente, el objetivo de cada uno es construir un almacén de datos departamental como componente de acceso a los datos, y los componentes de obtención y almacenamiento de los datos correspondientes.

Habitualmente, son proyectos que se completan en unos seis meses, aunque posteriormente es posible hacer ampliaciones sobre estos.

Aunque no es fácil de conseguir, es muy importante que los proyectos sean autónomos, es decir, que cada proyecto tenga perfectamente delimitados sus objetivos y no dependa del desarrollo de otros proyectos en curso. Esto es así porque se trata de proyectos cuyos requerimientos son cambiantes y los usuarios no tienen totalmente claro qué necesitan ni tampoco qué les puede ofrecer la FIC. Si un proyecto no es autónomo y, por ejemplo, necesita los datos de otro proyecto que está en desarrollo, es muy probable que surjan conflictos entre estos, puesto que seguramente las necesidades de datos del proyecto no autónomo cambiarán durante el desarrollo, y para el otro proyecto adaptarse a estas necesidades cambiantes será más costoso de lo que se esperaba inicialmente.

Resulta especialmente importante que los proyectos sean autónomos cuando se planifica que sean desarrollados por diferentes empresas de servicios, sobre todo si estas lo llevan a cabo con un presupuesto cerrado.

Generalmente, muchos proyectos estarán interrelacionados, trabajarán sobre los mismos datos. Por lo tanto, si queremos que dos proyectos que trabajan sobre los mismos datos sean autónomos, deberán desarrollarse de manera secuencial.

Esta última condición frecuentemente no es fácil de imponer, puesto que se necesita disponer de los almacenes de datos departamentales lo antes posible. Aun así, es razonable si se quiere evitar un conflicto entre los proyectos. Si se desarrollan de manera paralela proyectos no autónomos, será crítico que el desarrollo del componente de integración y transformación se haga de manera que no se repitan elementos entre proyectos y tampoco surjan conflictos.

El conflicto principal entre los proyectos desarrollados de manera paralela aparece al obtener los datos, en el componente de integración y transformación.

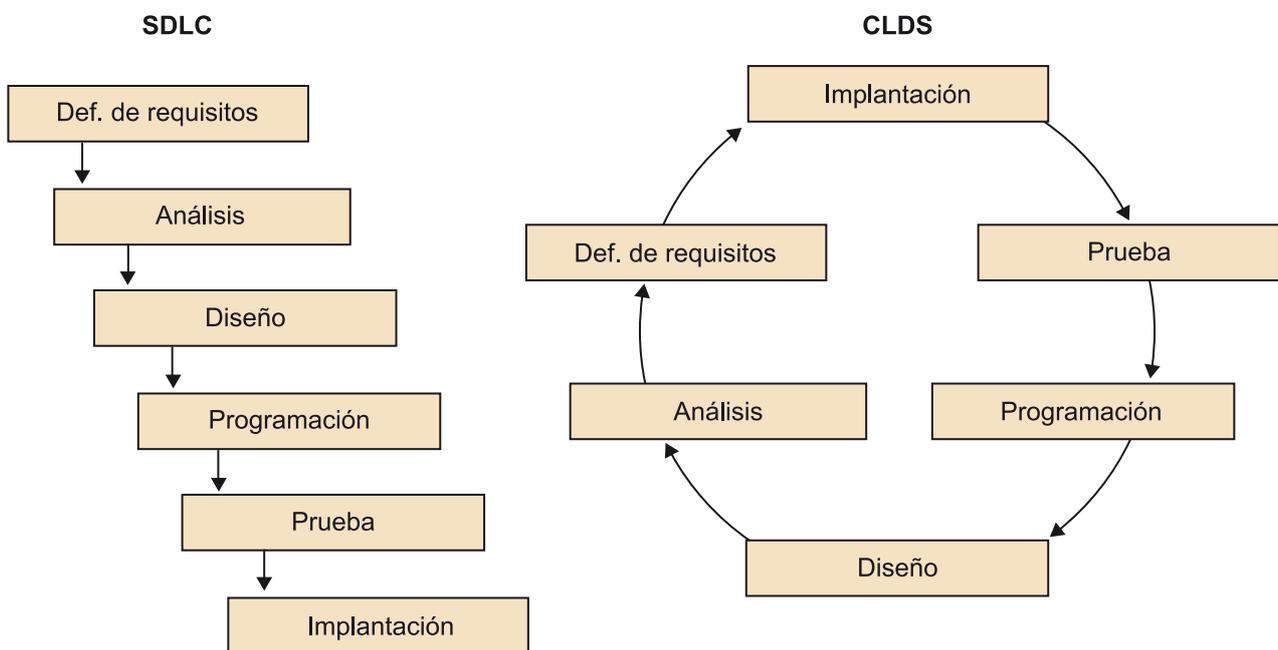
Todos los proyectos tienen que considerar el desarrollo de los componentes de obtención y almacenamiento de los datos y acceso a estos. Aun así, en algunos casos no se tienen que desarrollar todos los componentes porque ya están desarrollados. Es decir, se pueden plantear proyectos que se dediquen a explotar datos que ya hay en el almacén de datos corporativo (proyectos que solo tienen componente de acceso). También puede haber proyectos cuya misión sea ampliar el conjunto de datos que hay en el almacén de datos corporativo (proyectos que solo tienen componentes de obtención y almacenamiento). En estos casos, la situación es diferente de aquella en la que se llevaba a cabo un

desarrollo parcial de la arquitectura de la FIC, puesto que ahora se estudian todos los componentes de cada proyecto, pero algunos de estos ya están desarrollados.

Para desarrollar cada uno de los proyectos, podemos aplicar la metodología de desarrollo en cascada o SDLC, o la metodología en espiral, también denominada CLDS<sup>11</sup>.

<sup>(11)</sup>Corresponde a las siglas del ciclo de vida en cascada (SDLC) puestas a la inversa.

Las fases de la metodología de desarrollo en espiral son las mismas que las de la metodología en cascada. La diferencia entre las dos es que, en este caso, las fases se hacen en orden distinto y se llevan a cabo de manera iterativa en lugar de hacerse de manera secuencial.



Fases en las metodologías de desarrollo SDLC y CLDS.

Es conveniente aplicar la metodología de desarrollo CLDS cuando no se tienen claros los requerimientos del sistema que hay que desarrollar y estos no se pueden descubrir de manera inmediata por medio de entrevistas convencionales con los usuarios.

Las fases de la metodología de desarrollo CLDS son las siguientes:

- Se empieza por implantar una primera versión del sistema a los usuarios: esta podría ser un modelo en papel o un prototipo de sistema.
- Los usuarios prueban esta versión.
- Se hace el desarrollo necesario para obtener, almacenar y analizar los datos de la versión de prueba.

#### Lectura recomendada

Para más detalles, podéis ver: **W. H. Inmon; C. Imhoff; R. Sousa (1998). *Corporate Information Factory*. EE. UU.: John Wiley & Sons, Inc.**

- Una vez se han desarrollado los programas necesarios, se hace un diseño formal del sistema.
- Analizar los resultados del diseño, reformular y reprogramar si es necesario.
- Como último paso, entender los requerimientos del sistema.

Estos pasos se repiten hasta tener el sistema desarrollado según las necesidades de los usuarios. Los usuarios van descubriendo sus necesidades mediante el uso de versiones preliminares del sistema que tienen refinamientos sucesivos, hasta conseguir una versión final.

En el entorno de construcción de un almacén de datos departamental, es frecuente que los analistas no tengan una idea clara de las características del sistema que necesitan hasta que lo ven funcionando. En este caso, la metodología de desarrollo CLDS es más adecuada, puesto que permite hacer refinamientos sucesivos del sistema hasta definir claramente los requerimientos del sistema que se quería.

### **3.4. Evolución del entorno operacional**

En cuanto se construye la FIC, algunos elementos del entorno operacional dejan de ser útiles y se pueden dejar de usar.

#### **3.4.1. Evolución del entorno operacional en telaraña**

Concretamente, tal y como habíamos estudiado previamente en este módulo, el entorno operacional había evolucionado a partir del desarrollo de programas de extracción y almacenamientos temporales de datos en lo que habíamos denominado *entorno operacional en telaraña*. Cuando se construía cada uno de los proyectos autónomos en los que se ha dividido la construcción de la FIC, parte de los programas de extracción y de los almacenamientos temporales de datos –es decir, parte de la "telaraña"– deja de tener utilidad, puesto que su función pasa a ser ejercida por el nuevo almacén de datos departamental y, por lo tanto, se puede proceder a su desmantelamiento.

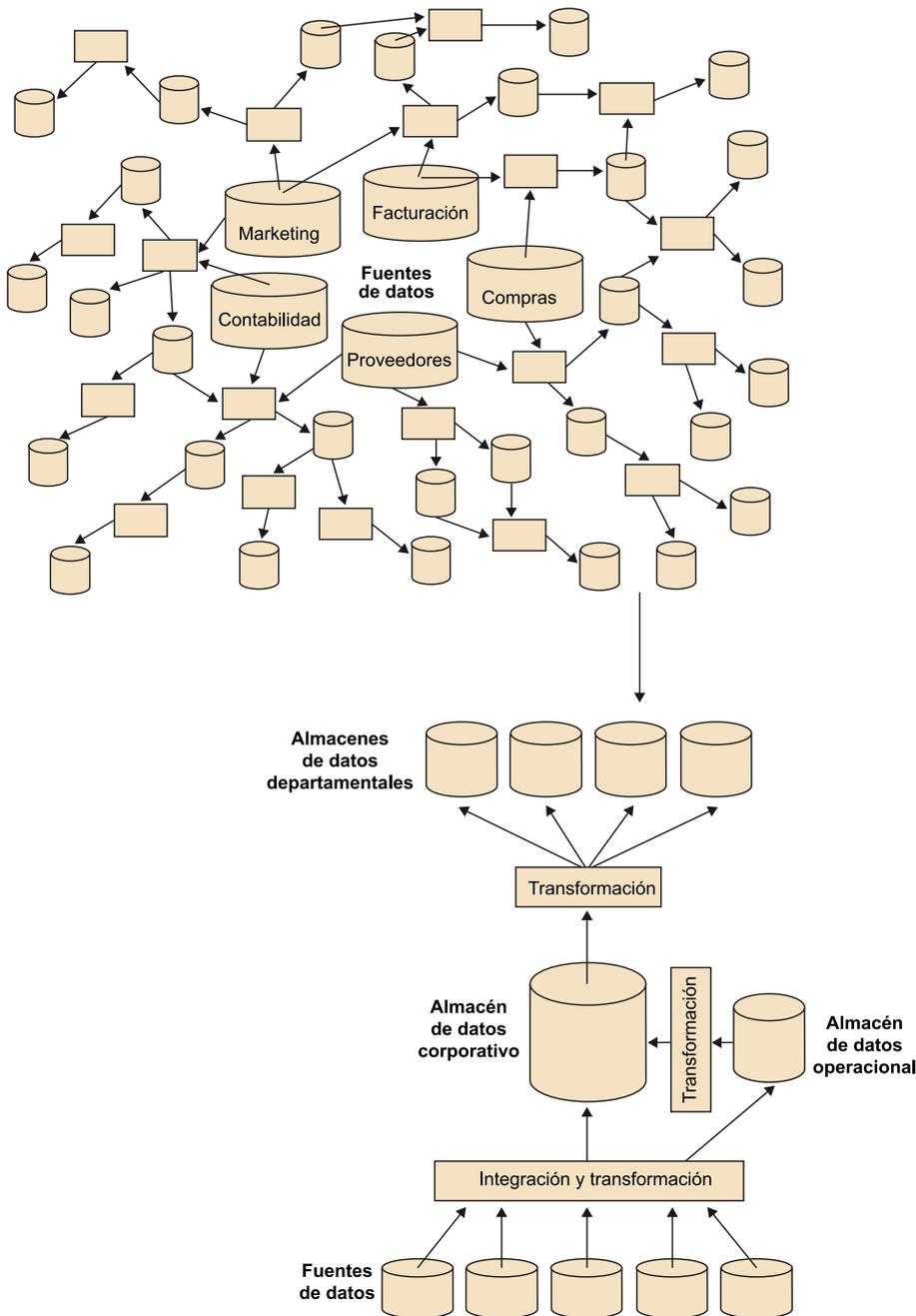
El desmantelamiento de la telaraña significa localizar los programas de extracción y los almacenamientos temporales de datos que dejan de ser útiles y evitar que se vuelvan a ejecutar o crear.

Es conveniente hacer las operaciones de desmantelamiento de manera progresiva dentro de cada proyecto de desarrollo de la FIC. Para desarrollar el componente de integración y transformación, deberían analizarse las diferentes

fuentes de datos: este es el momento de determinar y documentar la parte de la telaraña que queda cubierta y pasa a ser redundante con el desarrollo del nuevo proyecto.

El desmantelamiento de la telaraña se tiene que hacer de manera progresiva. Al construir un almacén de datos departamental, podemos desmantelar la parte de la telaraña del entorno operacional que ejercía la función del nuevo sistema construido.

El desmantelamiento de la telaraña del entorno operacional debe tenerse en cuenta cuando se planifica la construcción de la FIC. Si no se desmantela, la organización estará malgastando los recursos que consumen los programas de extracción y los almacenamientos temporales de datos.



Desmantelamiento del entorno operacional de telaraña.

Podemos ver el resultado de desmantelar el entorno operacional en la figura anterior. Pasamos de tener el entorno operacional en telaraña a tener la FIC, en la que las aplicaciones operacionales actúan como fuentes de datos y la extracción de los datos se hace de manera controlada y sistemática mediante el componente de integración y transformación.

### 3.4.2. Evolución de las aplicaciones operacionales

Además del entorno operacional en telaraña, al construir la FIC partes de las aplicaciones operacionales dedicadas a la generación de informes dejarán de ser útiles para los analistas, puesto que su funcionalidad estará cubierta por los diferentes almacenes de datos departamentales. Aun así, su desmantelamiento

es más complejo, puesto que implica modificar las aplicaciones operacionales. En este caso, simplemente se dejarán de usar estas partes de las aplicaciones sin eliminar físicamente su funcionalidad del código de las aplicaciones. En el supuesto de que haya componentes que se ejecuten de manera automática y sus resultados dejen de ser útiles, dependiendo de su coste, sí que podrá interesar modificar el código de la aplicación para que se dejen de ejecutar.

La parte de generación de informes de las aplicaciones operacionales generalmente no se eliminará por la creación de la FIC.

### 3.4.3. Otros cambios en la organización

Con la FIC se produce una evolución del tipo de informes que utilizan los analistas de la organización. La FIC, principalmente mediante los almacenes de datos departamentales, permite a los analistas generar los informes a medida. Las aplicaciones operacionales, y el entorno en telaraña creado a partir de estas, permitía a los usuarios obtener informes estructurados de formato fijo.

La FIC permite a los analistas incrementar de manera considerable el uso de informes a medida como apoyo al proceso de toma de decisiones, en detrimento del número de informes de formato fijo ofrecidos por las aplicaciones operacionales.

Por un lado, los equipos de desarrollo dedicados al desarrollo y mantenimiento de programas para la generación de informes (los programas de extracción que formaban la telaraña del entorno operacional) se tendrán que adaptar a la nueva situación y reubicar en el Departamento de Informática. Asimismo, en el supuesto de que alguno de los informes se generase de manera manual (situación bastante frecuente en algunas organizaciones) y ahora pasara a generarse dentro de la FIC, las personas encargadas de esta tarea se tendrán que reubicar para llevar a cabo otras tareas en la organización.

La FIC, además de afectar a la manera de trabajar de los analistas, afectará a los equipos de desarrollo que se dedicaban a generar los informes, los cuales pasan a generarse dentro de la FIC.

### 3.5. Perfiles en el equipo de gestión y desarrollo de la FIC

Además de los perfiles habituales en cualquier equipo de desarrollo de proyectos, en el equipo de desarrollo de la FIC aparecen un conjunto de perfiles que no son habituales en otros proyectos y que estudiaremos a continuación junto a los perfiles de gestión característicos de la FIC.

#### 3.5.1. El administrador de la FIC

El administrador de la FIC es su responsable ante la organización. Su misión es que la FIC se adapte a las necesidades de la organización. Por este motivo, debe conocer las posibilidades que esta ofrece, así como las necesidades de información de la organización, y controlar que las satisfaga lo suficiente.

Para cada proyecto que se desarrolle, se tiene que asegurar que cumpla los requerimientos de ámbito y funcionalidad, así como los que hacen referencia a coste y tiempo de desarrollo.

El administrador de la FIC es el máximo responsable ante la organización de que la FIC cumpla a lo largo del tiempo sus requerimientos de información.

#### 3.5.2. Los analistas de requerimientos de negocio

El analista de requerimientos de negocio es el responsable del primer proyecto: el proyecto global de desarrollo. Contará con un equipo de analistas para su desarrollo. Su misión es identificar los requerimientos de información por parte de la organización y planificar el desarrollo de la FIC de modo que estos se cumplan.

El administrador de la FIC es el máximo responsable del entorno de la misma, mientras que los analistas de requerimientos de negocio son los interlocutores entre los usuarios de la FIC y el equipo que la tiene que construir y mantener.

Los analistas de requerimientos de negocio recogen las necesidades de información de los usuarios de la FIC y las transmiten al resto del equipo de desarrollo.

Además de ser responsables del proyecto global de desarrollo, también lo son de los diferentes proyectos autónomos en los que se divide la construcción de la FIC y que se desarrollan posteriormente.

#### Lectura recomendada

Para más detalles sobre este tema, podéis consultar: **W. H. Inmon; J. D. Welch; K. L. Glassey (1997). *Managing the Data Warehouse*. EE. UU.: John Wiley & Sons, Inc.**

#### Ved también

Podéis ver el apartado "Construcción de la FIC mediante proyectos autónomos" de este módulo.

### 3.5.3. El arquitecto de la FIC

El arquitecto es el responsable del diseño de la FIC, diseña su arquitectura y es responsable de los proyectos de desarrollo de infraestructura.

El arquitecto de la FIC analiza las fuentes de datos de la organización y diseña los esquemas de datos y la estructura de la FIC según los requerimientos de información de los usuarios y las posibilidades ofrecidas por las fuentes de datos.

El arquitecto de la FIC trabaja de manera conjunta con el analista de requerimientos de negocio en el primer proyecto, el proyecto global de desarrollo de la FIC, especialmente para definir los requerimientos de infraestructura de los proyectos.

### 3.5.4. El patrocinador de la FIC en la organización

El patrocinador de la FIC es una figura política, cuyo objetivo es conseguir el apoyo necesario en la organización para desarrollarla salvando los obstáculos internos que surjan.

El ámbito de los proyectos de desarrollo de la FIC es toda la organización. Requiere integrar los datos de los diferentes departamentos y, una vez la FIC ya funciona, puede cambiar la manera de trabajar de equipos importantes de personal. Por todo esto, es frecuente que los responsables de algunos departamentos se muestren poco acostumbrados a compartir los datos o, generalmente, que algunos gestores de alto nivel en la organización muestren su oposición a la FIC por los cambios que representa. El patrocinador de la FIC tiene que interaccionar con todos ellos para intentar conseguir su apoyo o, al menos, reducir su oposición.

Por otro lado, según el planteamiento de construcción de la FIC mediante el desarrollo de proyectos autónomos presentado en este módulo, los diferentes proyectos de construcción de almacenes de datos departamentales están justificados, puesto que aportan un beneficio a la organización mayor que el coste que representa su desarrollo. Aun así, otros proyectos muy importantes, como por ejemplo el proyecto global de desarrollo y el proyecto de desarrollo de la infraestructura, no tienen una justificación inmediata de su coste. El patrocinador de la FIC en la organización tendrá como misión conseguir los fondos necesarios para el desarrollo de estos proyectos.

#### Ved también

Podéis ver el apartado "Construcción de la FIC mediante proyectos autónomos" de este módulo.

#### Ved también

Podéis ver el apartado "Otros cambios en la organización" de este módulo para tener más detalles sobre los cambios producidos por la FIC.

Teniendo en cuenta el tipo de personas con las que tiene que interactuar y la tarea que debe llevar a cabo, interesa que el patrocinador de la FIC en la organización sea un directivo de suficiente nivel para que su tarea resulte más fácil. Además, tiene que estar convencido de las ventajas que la FIC puede proporcionar a la organización.

El patrocinador de la FIC en la organización es una figura política. Generalmente, se trata de un directivo de alto nivel que conoce las ventajas que la FIC aporta a la organización y que se encarga de obtener los fondos necesarios y de resolver los problemas internos que surjan en la organización, con el objetivo de conseguir el éxito en el desarrollo de la FIC.

### 3.5.5. El gestor de cambios organizacionales

La implantación de cada uno de los proyectos de desarrollo de la FIC representa un cambio en la manera de trabajar en parte de la organización. El gestor de cambios organizacionales tiene como responsabilidad principal gestionar el impacto de la FIC en la organización.

#### Ved también

Podéis ver el apartado "Otros cambios en la organización" de este módulo.

Por un lado, cada proyecto de desarrollo de la FIC representa un cambio en la manera de trabajar para los usuarios del nuevo proyecto. Anteriormente obtenían los informes del Departamento de Informática o mediante personas que se los elaboraban manualmente, y con frecuencia los informes tenían forma de listados. Ahora, con la FIC, pueden generar ellos mismos los informes que necesitan y obtenerlos directamente por pantalla en tiempo real. Dado que a muchas personas no les gusta ningún tipo de cambio, aunque este sea beneficioso a corto plazo, se tiene que hacer una tarea de marketing para que los cambios se acepten más fácilmente. Esta tarea es responsabilidad del gestor de cambios organizacionales.

Por otro lado, el cambio en la manera de generar los informes puede desplazar a distintos trabajadores del Departamento de Informática u otros departamentos, que anteriormente eran los encargados de realizar la tarea que ahora pasa a estar cubierta por la FIC. Estos trabajadores serán muy útiles como miembros del equipo de desarrollo de la FIC o en otras actividades dentro de la organización. El gestor de cambios organizacionales tiene que intentar reducir la incertidumbre de estos trabajadores y colaborar para determinar cuáles serán sus nuevas funciones, así como para minimizar los posibles conflictos que se puedan producir.

El gestor de cambios organizacionales tiene como responsabilidad que los miembros de la organización acepten los cambios que implica la FIC.

### 3.5.6. El gestor de metadatos

Los metadatos son un componente muy importante de la FIC, puesto que describen su estructura y contenido y, además, permiten interconectar el resto de los componentes entre sí.

Teniendo en cuenta su importancia, interesa disponer de una persona responsable de los metadatos de la FIC. Su misión es asegurar que estos reflejen la situación actual de la FIC y sean accesibles tanto por los analistas de información como por los equipos de desarrollo que lo requieran. Además, tiene que asegurar que los metadatos puedan ser entendidos por los diferentes usuarios que acceden a los mismos. Teniendo en cuenta la variedad de herramientas que generan y utilizan metadatos, no resulta una tarea sencilla.

El gestor de metadatos es responsable de la situación y del acceso a los metadatos en la FIC.

#### Ved también

Podéis encontrar más detalles en la sección correspondiente a los metadatos del módulo titulado "La factoría de información corporativa".

### 3.5.7. Los analistas de la calidad de los datos

Los analistas de la calidad de los datos tienen como responsabilidad asegurar que aquellos que se han obtenido de las fuentes de datos operacionales satisfagan los requerimientos de información de la organización.

Se encargan de verificar que los datos se han obtenido, transformado y cargado de la manera requerida y el resultado es de la calidad esperada, es decir, que los datos en los distintos almacenes de datos de la FIC son los apropiados y son correctos.

En caso de detectar datos incorrectos, la solución no es corregirlos donde se han detectado, sino encontrar la fuente de esta incorrección y solucionar allí el problema que se haya producido.

Los analistas de la calidad de los datos tienen que detectar aquellos que no se adaptan al grado de calidad requerido por la FIC. Su misión no consiste en corregir estos datos, sino identificar los motivos de la baja calidad y recomendar acciones que conduzcan a la solución de este problema.

La cantidad de analistas de la calidad de los datos necesaria dependerá del grado de calidad de los datos operacionales y el grado requerido. Durante las fases de desarrollo de la FIC, seguramente serán necesarios distintos analistas de calidad. Cuando la FIC esté en funcionamiento, debería bastar con una sola persona, incluso a tiempo parcial.

### 3.5.8. El administrador de bases de datos

Aunque en la organización debe haber personal de administración de bases de datos para el entorno operacional, conviene que exista personal específico para el entorno de la FIC.

Aunque el sistema de gestión de bases de datos utilizado en el entorno operacional puede ser el mismo que el que se ha usado en los distintos almacenes de datos, sus características de configuración varían de manera radical entre los dos entornos. Por este motivo, interesa disponer de administradores especializados en el entorno de los almacenes de datos que sean capaces de obtener el máximo rendimiento de los sistemas y que no tengan que cambiar continuamente la manera de pensar al administrar tanto el entorno operacional como el de los almacenes de datos.

Asimismo, puesto que la FIC obtiene sus datos a partir del entorno operacional, es adecuado que los responsables técnicos de los dos entornos sean personas diferentes, de modo que la integridad de los datos de la FIC no se vea afectada por posibles problemas del entorno operacional. Si el administrador de los dos entornos fuera único, ante cualquier problema se puede producir un conflicto de prioridades; teniendo un administrador de bases de datos propio, su máxima prioridad ha de ser la FIC.

El administrador de bases de datos tiene que monitorizar de manera continua los procesos de obtención y acceso a los datos, y configurar los sistemas para que estos se hagan de manera óptima.

#### Ved también

En cuanto al almacenamiento, ya hemos estudiado en los apartados anteriores las figuras del arquitecto de la FIC y del administrador de bases de datos.

### 3.5.9. Especialistas en obtener y acceder a los datos

Además de las figuras estudiadas en los apartados anteriores, el equipo de desarrollo de la FIC estará formado por especialistas en las operaciones específicas en las que se descomponen los proyectos de desarrollo de la FIC: obtención y almacenamiento de los datos y acceso a los mismos.

En cuanto a la obtención de los datos, particularmente si esta se hace mediante el uso de alguna herramienta de soporte, el equipo de desarrollo contará con los especialistas desarrolladores del componente de integración y transformación. Estos conocen el contenido y las posibilidades de las fuentes de datos y también los procedimientos y las herramientas para obtener los datos a partir de estos.

#### Ved también

Podéis ver el subapartado "Herramientas de apoyo para el desarrollo" en el apartado "Desarrollo del componente de integración y transformación" de este módulo.

En cuanto al componente de acceso, el equipo de desarrollo contará con especialistas en las herramientas de acceso a los datos y metadatos que usarán los diferentes tipos de analistas de información de la organización. Estos especialistas construirán los métodos de acceso necesarios para satisfacer las necesidades de los usuarios en este sentido.

### 3.6. Razones por las cuales los proyectos fallan

En todos los ámbitos de la vida es interesante aprender de los errores cometidos, especialmente de aquellos que han cometido otros, aunque esto último es realmente difícil.

En lo que respecta a errores en los proyectos de construcción de la FIC, con el objetivo de que nos puedan resultar instructivos, hay dos categorías de razones de error -técnicas y no técnicas-, como se muestra a continuación.

Entre las razones técnicas, destacamos las siguientes:

- Falta de escalabilidad de la FIC: escalabilidad referida al aumento en el volumen de datos o en el número de usuarios que necesitan trabajar sobre la FIC, y también en los problemas para incorporar un nuevo proyecto autónomo o un nuevo componente a la FIC. De aquí viene la importancia de que tanto la infraestructura operacional como la física sean escalables, de modo que también podamos construir proyectos escalables.
- Problemas en la gestión y el intercambio de metadatos entre las diferentes herramientas que intervienen en la construcción y el funcionamiento de la FIC: si no podemos intercambiar metadatos entre herramientas, el resultado será un producto más difícil de mantener.
- Problemas para proporcionar a los usuarios aplicaciones reales que aporten un valor a la organización: se debe conseguir justificar el coste del desarrollo de la FIC mediante cada proyecto autónomo que se haga, y estos se tienen que poder llevar a cabo en un plazo realista.
- El error al diseñar una arquitectura de la FIC que sea lo bastante robusta y flexible como para permitir llevar a cabo los proyectos autónomos planteados.
- La imposibilidad de aprovechar los descubrimientos hechos mediante el uso de la FIC porque no hay estabilidad o flexibilidad de las fuentes de datos operacionales.

Las razones de error no técnicas son las siguientes:

#### Lectura recomendada

Podéis encontrar más detalles sobre este tema en la obra siguiente:

S. Kelly (1997). *Data Warehousing inAction*. EE. UU.: John Wiley & Sons Inc.

#### Proyectos autónomos

Recordad que en la definición de los proyectos autónomos, habíamos mencionado que estos deberían tener un plazo de ejecución de unos seis meses como máximo.

- No tener el apoyo de un responsable de alto nivel de la organización en el proyecto.
- Gestionar de manera incorrecta las expectativas de los usuarios y la dirección.
- No alinear el despliegue de la FIC con un cambio cultural en la organización de cómo se toman las decisiones.
- No definir correctamente las necesidades de la organización que ha de cubrir la FIC.
- No coordinar las diferentes iniciativas departamentales dentro de la organización.
- No tener el personal o los socios (*partners*) adecuados para llevar a cabo el desarrollo de la FIC.
- No ser capaces de implicar lo suficiente a los usuarios de la FIC en el desarrollo.
- No poder justificar el coste de los proyectos planteados para desarrollar o no conseguir los fondos necesarios para los proyectos de difícil justificación.
- No poder expresar los requerimientos de la organización en forma de proyectos desarrollables en nuestro entorno.
- Imposibilidad de cambiar la cultura de la organización para que utilice las posibilidades de información que ofrece la FIC.
- Imposibilidad para aprovechar los descubrimientos hechos mediante la FIC debido a la inercia que tiene la organización.

## 4. Desarrollo del componente de integración y transformación

En el módulo "La factoría de información corporativa" hemos visto cuáles son las funciones principales del componente de integración y transformación:

- Obtener los datos de las fuentes de datos para la imagen inicial y las actualizaciones.
- Transformar, integrar y depurar los datos.
- Transportar y cargar los datos en los almacenes de datos.

En este apartado, estudiaremos las principales soluciones que se suelen aplicar en el desarrollo de algunas de estas actividades (concretamente, por su complejidad, nos centraremos en la obtención de las actualizaciones de los datos y la actualización de los datos del almacén de datos mientras se carguen), así como el soporte ofrecido por algunas herramientas del mercado para implementarlas.

### 4.1. Obtener los datos de las fuentes de datos

El proceso para obtener los datos a partir de las fuentes de datos se lleva a cabo mediante un conjunto de aplicaciones que se ejecutan con esta finalidad. Este proceso se divide en dos fases:

- Obtener la imagen inicial.
- Obtener las actualizaciones.

#### 4.1.1. Obtener la imagen inicial de los datos

La imagen inicial se obtiene con un conjunto de aplicaciones que generalmente se ejecuta una sola vez. El resultado de esta fase es una imagen de la situación actual de los sistemas operacionales obtenida mediante un vaciado de sus respectivas bases de datos. Normalmente, la imagen inicial se obtiene sin dificultad, aunque, según las características de los sistemas operacionales, esto no siempre es así.

##### **Ejemplo de obtención de la imagen inicial de los datos:**

En un sistema operacional basado en una base de datos relacional, es posible obtener los datos de manera inmediata, haciendo un vaciado mediante las herramientas que el sistema de gestión de base de datos ofrece para esto. Aun así, en una aplicación empaquetada desarrollada sobre el sistema de ficheros, es posible que solo podamos acceder a los datos mediante las pantallas de la aplicación y se requiera, en este caso, un desarrollo complejo y costoso en el que se tengan que ir capturando los datos presentados en las diferentes pantallas.

Para reflejar en el almacén de datos la evolución que tienen los datos, partiendo de su imagen inicial debemos ir obteniendo las actualizaciones que se van produciendo.

Actualmente, los datos no estructurados o semiestructurados, es decir, encapsulados en ficheros XML, SGML o incluso en objetos sin una estructura predefinida (PDF o Word), son también fuentes de alto valor potencial para las organizaciones. Permiten desde conocer información de los competidores, hasta conocer en profundidad a los clientes. En el primer caso, por ejemplo, la información relativa a los precios del competidor se encuentra en su página web, pero donde realmente se encuentra la ventaja competitiva es en la automatización de este proceso.

En esta situación los procedimientos de extracción de información estándar, como ETL, con frecuencia no son suficientes y hay que utilizar técnicas de recuperación de información que hace uso de métodos de reconocimiento de patrones, muestreo estadístico, probabilidad, etc. Esto complica la obtención de la imagen inicial de los datos, que no siempre está focalizada en los datos internos de la organización.

#### 4.1.2. Métodos para obtener las actualizaciones de los datos

La manera de obtener los datos para las actualizaciones dependerá de los requerimientos de los analistas sobre los almacenes de datos, y también de las posibilidades ofrecidas por las fuentes de datos. En la figura siguiente, se presentan los diferentes métodos para obtener las actualizaciones:

1) Comparación de imágenes (figura 1): algunas fuentes ofrecen la posibilidad de obtener un vaciado de sus datos de manera masiva. Un ejemplo de este tipo de fuentes son los ficheros ordinarios. Si trabajamos con bases de datos relacionales, el vaciado se puede obtener mediante consultas que seleccionan los datos que nos interesen. Para este tipo de fuentes de datos, podemos obtener las actualizaciones que se hayan producido comparando imágenes sucesivas que se hayan obtenido.

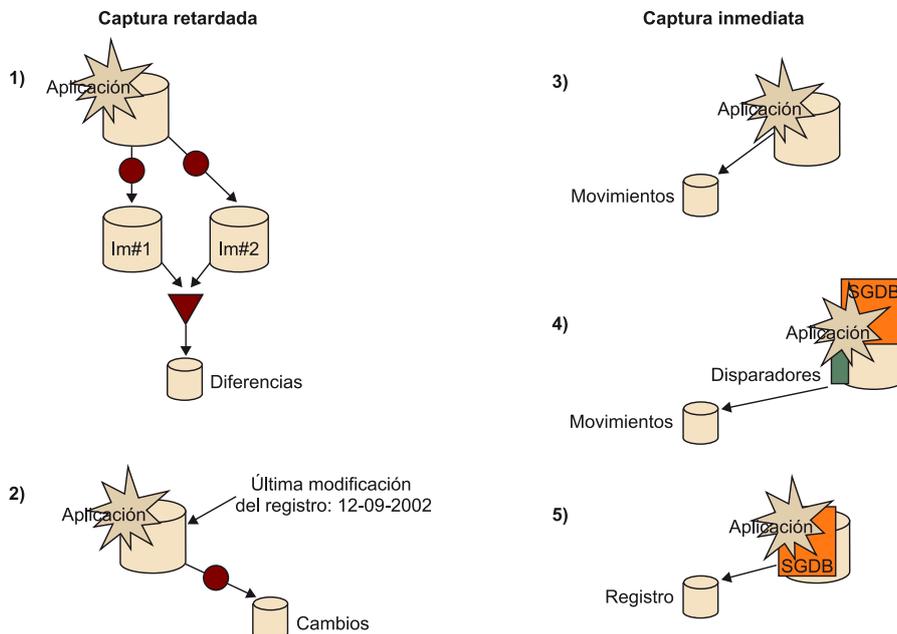
2) Fuentes con una huella de tiempo (figura 2): en este caso, las fuentes almacenan para cada registro modificado el momento en que se llevó a cabo la modificación. De este modo, obtener las últimas modificaciones es inmediato: basta con lograr los registros marcados con una huella de tiempo posterior a la del último conjunto de modificaciones obtenido.

Estos dos métodos se denominan de **captura retardada**, puesto que no se intenta capturar las modificaciones en el momento en que se producen, sino posteriormente. Tienen el inconveniente de que pueden perder movimientos: si se hacen diferentes modificaciones sobre un registro y estas no se obtienen

#### Lecturas recomendadas

Podéis encontrar información complementaria sobre este tema en: M. Jarque; M. Lenzerini; Y. Vassiliou; P. Vassiliadis (2000). *Fundamentals of Data Warehouses*. Berlín: Springer-Verlag. También la podéis encontrar en: B. Devlin (1997). *Data Warehouse from Architecture to Implementation*. EE. UU.: Addison Wesley Longman, Inc.

de manera inmediata después de producirse cada una, por medio de los métodos anteriores obtendremos un resumen de todas las operaciones hechas en una única operación de modificación.



Métodos de obtención de las actualizaciones.

### Ejemplo de pérdida de movimientos de los datos

En la imagen inicial del saldo de una cuenta obtenemos un valor de 1.000, y posteriormente se produce un ingreso de 500 y un reintegro de 700. Si en aquel momento volvemos a tomar la imagen del saldo de la cuenta por cualquiera de los dos métodos anteriores, obtenemos un valor de 800; si comparamos los saldos, obtenemos que se ha producido un reintegro de 200 y hemos perdido el detalle de los movimientos hechos.

3) Análisis del fichero delta de movimientos (caso 3 de la figura anterior): algunas aplicaciones almacenan las modificaciones hechas sobre sus datos en un fichero de movimientos denominado fichero delta. Generalmente, se utiliza para auditar las operaciones. Podemos utilizar este fichero para obtener las modificaciones hechas sobre los datos que nos interesen.

4) Detección de movimientos mediante disparadores (figura 4): si el sistema en el que se basa la fuente de datos ofrece la posibilidad de definir disparadores u otras funcionalidades de bases de datos activas, los podemos usar para obtener las modificaciones sobre los datos que nos interesen, activándolos cuando se produzcan acontecimientos de modificación sobre estos. Estos disparadores pueden comunicar los cambios producidos a las partes interesadas para que actúen de manera adecuada. Una manera habitual de actuar, en este caso, es la creación de un fichero delta de movimientos que se puede analizar posteriormente.

5) Análisis del fichero de registro (figura 5): algunas fuentes de datos, basadas en SGBD que lo permiten, almacenan las operaciones que se producen sobre estas en un fichero de registro, generalmente por motivos de recuperación ante errores, o bien para que sea posible deshacer modificaciones. Si analizamos este

fichero, se pueden obtener las modificaciones hechas sobre los datos que nos interesen. Esta solución presenta el inconveniente de que, aunque se genere en la fuente de datos, el fichero de registro no es de fácil acceso, puesto que está pensado para ser utilizado por el SGBD y su formato no siempre es de dominio público.

Estos tres últimos métodos de obtención de modificaciones sobre los datos, a diferencia de los dos primeros, tienen como característica principal que no pierden ninguno de los movimientos que se producen. Se les denomina métodos de **captura inmediata**, puesto que obtienen los movimientos en el preciso instante en el que se producen.

6) Métodos de recuperación de información semiestructurada: la información estructurada se encapsula en formato XML y SGML y, por lo tanto, las técnicas de acceso y consulta son las adecuadas para la extracción de esta información. A pesar de que encontramos distintas aproximaciones, no se recomienda el uso de lenguaje de programación ni la creación de rutinas *ad hoc*. Los métodos recomendados son los siguientes.

a) Utilización de herramientas ETL especializadas que incluyen los mecanismos de manipulación de ficheros XML y también la posibilidad de analizar la diferencia entre las versiones de los documentos.

b) Uso de motores especializados de indexación de información semiestructurada como Apache Lucene.

7) Métodos de recuperación de información no estructurada. En el caso de la información no estructurada, también hay diferentes aproximaciones. Los diferentes modelos que encontramos son estos.

a) Booleano: método de recuperación simple fundamentado en la teoría algebraica booleana. Establece una relevancia binaria de modo que un documento es relevante o no lo es. Su simplicidad ha motivado que caiga en desuso.

b) Vectorial: compara la consulta con el texto en un documento y genera un vector. Vectores que apuntan a la misma dirección son similares y el grado de verosimilitud es función del ángulo entre los vectores. Es uno de los métodos actuales más populares y fiables. En particular, motores Java como Apache Lucene lo utilizan.

c) Modelo LSI<sup>12</sup>: el objetivo es calcular la verosimilitud entre la consulta y el documento mediante la similitud entre conceptos, y no palabras.

<sup>(12)</sup>Del inglés, *latent semantic indexing*.

d) Métodos mejorados: los métodos anteriores se han mejorado con el uso de motores semánticos, redes semánticas o el análisis de la regresión.

### 4.1.3. Criterios de selección del método para obtener las actualizaciones de los datos

¿Qué método utilizaremos de entre los métodos que hemos presentado para obtener modificaciones? La respuesta a esta pregunta dependerá en gran medida de las funcionalidades ofrecidas por la fuente de datos, y también de los requerimientos por parte de los usuarios de los almacenes de datos. Los criterios siguientes no son estrictos, pero nos pueden servir como orientación:

- Si la fuente de datos dispone de un fichero delta de movimientos, generalmente esta será la opción que elegiremos, puesto que este fichero suele ser directamente accesible y fácil de analizar.
- La opción siguiente, si lo genera la fuente de datos, es analizar el fichero de registro. En este caso, la accesibilidad es menor y la dificultad para analizarlo, mayor.
- Si la fuente de datos está basada en una base de datos que permite crear disparadores, podremos utilizar este método. La definición de disparadores hará que las transacciones sean más costosas en tiempos de ejecución; además, el coste de desarrollo de esta solución suele ser mayor que el de las anteriores. Aun así, podemos obtener las modificaciones en el mismo momento en que se han producido, sin necesidad de hacer un análisis posterior.
- Si la fuente de datos almacena una huella de tiempo para cada modificación hecha, se preferirá esta solución a la de comparación de imágenes, puesto que es más inmediata y el coste para obtener las modificaciones es mucho menor. Aunque el coste también puede ser menor en este caso que en casos anteriores, el hecho de que perdamos movimientos hará que esta solución sea menos atractiva que aquellas que no presentan este inconveniente.
- La opción que se suele elegir como último recurso es la de comparar imágenes. Presenta los inconvenientes de que con esta podemos perder movimientos y además el coste en tiempo de ejecución puede ser bastante más elevado que el de las otras soluciones, dependiendo de la medida de la fuente de datos. Aun así, esta opción es inmediata, no exige ninguna funcionalidad especial a las fuentes de datos (siempre la podemos implementar) y, por lo tanto, frecuentemente es la elegida.

### Ejemplo de problema de obtención de las modificaciones de los datos

Queremos obtener los datos de una fuente que no crea un fichero delta, que está basada en una base de datos a cuyo fichero de registro no se puede acceder y que no permite la definición de disparadores, además de que tampoco almacena una huella de tiempo con las modificaciones. En esta situación, la alternativa más inmediata es la de comparar imágenes. Otra posibilidad sería la de modificar la fuente de datos para que genere un fichero delta de modificaciones; aun así, esto no siempre es factible (por ejemplo, si la fuente de datos es una aplicación empaquetada) o puede resultar demasiado costoso dependiendo del tamaño de la aplicación.

#### Ved también

Podéis encontrar más información sobre estas operaciones en el apartado "Transformación, depuración e integración de los datos" del módulo "La factoría de información corporativa" de esta asignatura.

## 4.2. Actualización de los datos del almacén de datos

Cuando ya hemos obtenido los datos a partir de las fuentes de datos, se transforman, depuran e integran si es necesario y, finalmente, se tienen que transportar al almacén de datos para proceder a cargar o actualizar los datos que hay.

#### Lectura recomendada

Podéis encontrar más información en: **B. Devlin** (1997). *Data Warehouse from Architecture to Implementation*. EE.UU.: Addison Wesley Longman, Inc.

### 4.2.1. Métodos de actualización de los almacenes de datos

La actualización de los datos de los diferentes almacenes de datos se puede llevar a cabo de cuatro maneras distintas, que comentamos a continuación:

1) Carga<sup>13</sup>: mediante la operación de carga, el almacén de datos de destino pasa a contener exclusivamente los datos que se indican en esta operación. Si previamente contenía otros datos, estos son reemplazados por los nuevos.

<sup>(13)</sup>En inglés, *load*.

2) Adición<sup>14</sup>: la operación de adición permite añadir los datos indicados a los que ya había antes en el almacén de datos. Se puede dar el caso de que algunos de los datos que se tienen que añadir ya estén en el almacén de datos. Si se produce esta situación, se puede elegir una de las dos alternativas siguientes: duplicar los datos que hay o rechazar la adición de los datos nuevos (exclusivamente los que ya hay).

<sup>(14)</sup>En inglés, *append*.

3) Fusión destructiva: la fusión destructiva permite añadir los datos indicados en la operación a los que ya había previamente de otro modo. Si la clave de un registro de los datos indicados en la operación coincide con la clave de alguno de los datos existentes, el registro que había previamente es reemplazado por el registro nuevo. Los nuevos registros cuya clave no coincide con otros que ya había se añaden directamente al almacén de datos.

4) Fusión constructiva: la fusión constructiva funciona de manera similar a la destructiva. La única diferencia consiste en el hecho de que, cuando coincide la clave de alguno de los registros que se tiene que añadir con la de un registro existente, marca estos registros pero no los reemplaza con los nuevos valores. De este modo, posteriormente se puede actuar sobre los registros de manera conveniente.

#### 4.2.2. Selección del método de actualización de los almacenes de datos

Para cargar un almacén de datos a partir de los datos obtenidos de los sistemas operacionales, utilizaremos una combinación de las operaciones que hemos estudiado en el apartado anterior. Utilizaremos una u otra dependiendo de la situación de los datos que se tienen que cargar y del almacén.

La carga inicial de los datos se hará mediante una operación de carga, puesto que partiremos de un almacén de datos vacío. Esta operación, incluso, nos permitirá crear de manera automática las tablas del almacén de datos si no estaban creadas anteriormente.

La actualización de los datos se hará con una combinación de las otras tres operaciones (adición, fusión destructiva y fusión constructiva). Con frecuencia bastará con hacer operaciones de adición, puesto que iremos almacenando nuevas imágenes de los datos. Por lo tanto, no encontraremos datos duplicados.

En algunos casos, interesa actualizar algunos campos de los datos existentes a partir de los nuevos datos obtenidos. Entonces, se utilizará la fusión constructiva.

##### Ejemplo de fusión constructiva

En el almacén de datos, para cada registro se indica mediante dos campos de tipo fecha (fecha de inicio, fecha final) el periodo de validez del resto de los campos del mismo. Al añadir un registro nuevo, el campo fecha final permanece no definido. Si se produce una modificación sobre cualquier campo del registro en la fuente de datos operacionales, lo que haremos será añadir un nuevo registro en el almacén de datos con los valores de los campos actualizados (el campo fecha final no estará definido). Sin embargo, además, tendremos que actualizar el campo fecha final de la versión previa del registro para indicar que estos valores ya no son válidos. Esta operación de actualización se llevará a cabo mediante la operación de fusión constructiva.

En otros casos, especialmente para corregir los errores que se hayan detectado en datos ya cargados en el almacén de datos (esta situación tendría que ser poco frecuente, pero es posible que se dé), o cuando tengamos datos en el almacén de datos que por alguna razón no han de ser historizados, utilizaremos la operación de fusión destructiva.

De manera adicional a la operación de carga inicial, se distinguen dos modos de actualización de los distintos tipos de almacenes de datos:

- **Refresco total:** todos los datos del almacén de datos se obtienen de nuevo y se cargan mediante la operación de carga. Esta situación se puede dar para cargar los datos de los almacenes de datos departamentales a partir del almacén de datos corporativo.
- **Mantenimiento incremental:** se trata de minimizar el tiempo requerido para hacer la actualización, para lo cual se parte de los datos que ya hay y

##### Refresco total

En el caso del almacén de datos corporativo, es infrecuente que se haga refresco total, aunque en algunas situaciones es posible si se puede obtener toda la historia de los datos que almacena.

se buscan procedimientos para actualizarlos según los nuevos datos. Para hacer el mantenimiento incremental, utilizaremos el resto de las operaciones estudiadas. Generalmente, esta será la manera de actualizar los datos en los diferentes almacenes de datos.

#### **Mantenimiento incremental de los datos agregados**

Los almacenes de datos departamentales suelen almacenar datos agregados (por ejemplo, suma de las ventas mensuales); hacer un mantenimiento incremental de los datos agregados es una operación muy compleja que, generalmente, requiere almacenar datos adicionales para que sea posible hacerlo.

### **4.3. Frecuencia y ventana de actualización**

#### **4.3.1. Frecuencia de actualización de un almacén de datos**

Cada almacén de datos tiene unos requerimientos de actualización propios y, además, estos no necesariamente han de ser homogéneos para todos los datos dentro de un almacén de datos.

##### **Requerimientos de frecuencia de actualización de diferentes campos**

En un almacén de datos operacional de un banco, disponemos de las direcciones de los clientes y de los datos del saldo de las cuentas. En el supuesto de que se produzca algún cambio en la dirección de un cliente en un sistema operacional, nos interesa que esta se actualice en el almacén de datos operacional al final del día, puesto que es exclusivamente entonces cuando utilizamos las direcciones para generar las etiquetas de la correspondencia que se envía a los clientes. Aun así, cualquier cambio en el saldo de alguna de sus cuentas registrada en una aplicación operacional interesa que se actualice lo más rápido posible en el almacén de datos operacional, puesto que se utilizará este campo para determinar si se autorizan o no al cliente las operaciones de crédito hechas con sus tarjetas.

#### **Un ejemplo**

Hay que definir la frecuencia con la que se produce una modificación en las fuentes de datos: cada hora, a diario, semanalmente, etc.  
Será más costoso actualizar el almacén de datos cada hora que una vez a la semana.

Por lo tanto, para cada almacén de datos y cada uno de sus datos se tendrá que definir la frecuencia de actualización. Cuanto más fuertes sean estos requerimientos, el coste de actualización será más elevado.

Generalmente, el almacén de datos operacional es el que tiene unos requerimientos más estrictos en este aspecto. En muchos casos, lo ideal es que la actualización sea inmediata, pero esto no siempre es factible. En este sentido, depende de las posibilidades que ofrezcan las fuentes de datos.

El almacén de datos corporativo suministra los datos a los almacenes de datos departamentales; por lo tanto, los requerimientos en este sentido de los distintos almacenes departamentales quedarán reflejados en el almacén de datos corporativo.

Para cada dato de los diferentes almacenes de datos se tiene que definir la frecuencia de actualización: cada cuánto tiempo se tiene que actualizar a partir de las fuentes de datos.

### 4.3.2. Ventana de actualización de un almacén de datos

El objetivo principal de los almacenes de datos es apoyar el proceso de toma de decisiones. Por lo tanto, las operaciones que se suelen realizar sobre estos son de consulta. Para optimizar la ejecución de este tipo de operaciones, los SGBD sobre los cuales están implementados los almacenes de datos están configurados de modo que solo permiten hacer operaciones de consulta.

#### Operaciones de modificación

El almacén de datos operacionales también soporta operaciones de modificación.

Si se quiere hacer otro tipo de operaciones distintas de las de consulta –por ejemplo, las de actualización del almacén de datos–, hay que parar los sistemas, cambiar la configuración, hacer las operaciones necesarias y restaurar la configuración para que las consultas continúen siendo óptimas. Durante este tiempo, el almacén de datos no está disponible para sus usuarios.

#### Disponibilidad del almacén de datos corporativo

En el caso extremo de requerir una disponibilidad total (veinticuatro horas al día todos los días del año), una posible solución es tener la base de datos del almacén de datos replicada y actualizar una de las copias mientras se utiliza la otra. Una vez actualizada esta, se puede usar mientras se actualiza la otra.

La ventana de actualización de un almacén de datos es el tiempo necesario para hacer las operaciones de actualización sobre este. Durante la ventana de actualización, el almacén de datos no es operativo para hacer consultas.

Relacionado con la ventana de actualización, tenemos el concepto de la ventana de extracción; esta se refiere al tiempo necesario para obtener las modificaciones de los datos a partir de las fuentes de datos.

Según el método de obtención de las modificaciones, la ventana de extracción será más o menos amplia. Generalmente, el método de obtención de modificaciones mediante comparación de imágenes es el que requiere una ventana de extracción más amplia, puesto que las operaciones que tiene que hacer consumen mucho tiempo. Además, junto con el método de obtención de modificaciones basado en fuentes con huella de tiempo, requiere que se haga un proceso en las plataformas de las fuentes de datos. Esto último puede resultar problemático si estas plataformas están muy sobrecargadas de trabajo.

En el caso del resto de los métodos, es suficiente con obtener el fichero de registro o los respectivos ficheros de movimientos y analizarlos en una plataforma distinta. En el método de comparación de imágenes, la comparación también se puede llevar a cabo en otra plataforma, pero la extracción de cada una de las imágenes (solo se tiene que extraer una cada vez) se tiene que hacer en la plataforma de la fuente de datos. Finalmente, si se usa el método basado en la huella de tiempo, generalmente el análisis de los datos se tiene que hacer en la plataforma que los contiene.

#### Situaciones extremas

En situaciones extremas en las que la plataforma está sobrecargada de trabajo, y la frecuencia de actualización lo permite, se puede utilizar una copia anterior de los datos en otra plataforma.

La ventana de extracción es el tiempo necesario para obtener los movimientos a partir de las fuentes de datos.

El componente de integración y transformación debe tener presentes los requerimientos de disponibilidad de las fuentes de datos, así como los de los almacenes de datos. Por este motivo, resulta muy importante minimizar el tiempo de proceso de los diferentes pasos, como se ha ido señalando cuando se han presentado, de modo que se puedan ejecutar dentro de las ventanas de extracción y actualización disponibles en cada caso.

#### 4.4. Herramientas de apoyo para el desarrollo

Las necesidades de los diferentes usuarios cambian con el tiempo, especialmente las de los analistas; por lo tanto, es habitual que se produzcan cambios tanto en las fuentes de datos como en los almacenes de datos. Por sus características de elemento intermediario entre el resto de los elementos de la FIC, el componente de integración y transformación se verá afectado por los cambios en cualquiera de los otros elementos. Por lo tanto, además de cumplir con las funciones que hemos explicado antes, este componente debe ser lo bastante flexible como para que se pueda adaptar a los cambios que se produzcan en cualquiera de los componentes con los que interacciona y, además, lo tiene que hacer de manera inmediata.

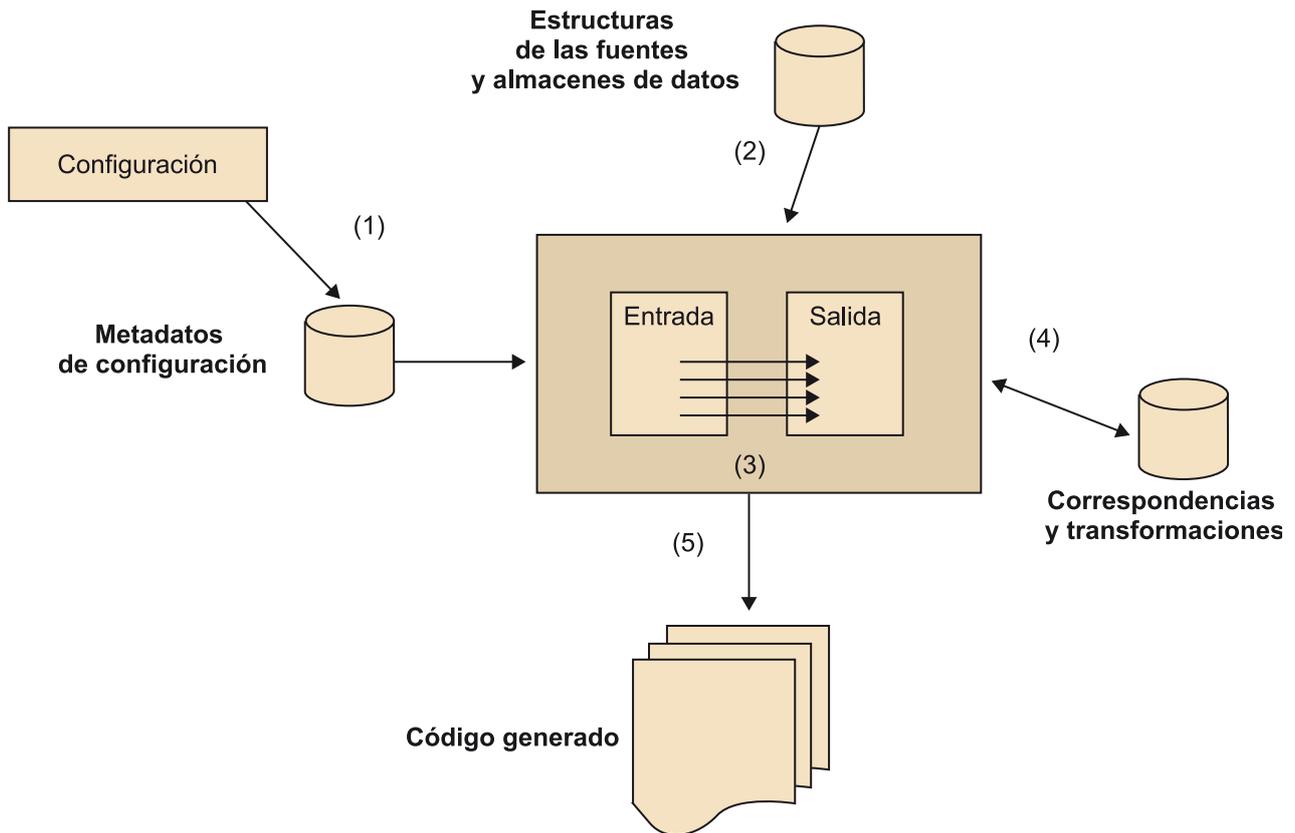
El componente de integración y transformación está formado principalmente por software. Para desarrollar cualquier componente de software, se presentan dos alternativas:

- Desarrollo manual.
- Desarrollo automático o con apoyo automático.

##### 4.4.1. Funcionamiento de las herramientas

Si lo que necesitamos fundamentalmente en el componente de integración y transformación es reaccionar a los cambios de manera inmediata, la solución más adecuada será llevar a cabo un desarrollo con soporte automático. Puesto que esta necesidad es ampliamente reconocida, en el mercado hay herramientas orientadas de manera especial para el desarrollo de este componente. Estas herramientas ofrecen un conjunto de transformaciones tipo de los datos, así como otras funcionalidades específicas para soportar las operaciones requeridas. Funcionan como cualquier herramienta CASE<sup>15</sup>, con la diferencia de que en este caso están orientadas a una función específica: la generación del componente de integración y transformación.

<sup>(15)</sup>Del inglés *computer aided system engineering*.



Funcionamiento de una herramienta de generación del componente de integración y transformación.

En la figura superior, se muestra una representación del funcionamiento de una herramienta dedicada a la generación del componente de integración y transformación. Inicialmente, se debe configurar la herramienta (fase 1 de la figura) y definir las características técnicas de los entornos de entrada y salida de datos, las fuentes de datos operacionales y los diferentes almacenes de datos. Asimismo, como parte de la configuración de la herramienta, se definen las plantillas de los programas tipo que se podrán generar: todos los programas de transformación tienen una estructura similar dentro de una casuística reducida. Cuando ya se ha configurado la herramienta, esta puede importar estructuras de datos de las fuentes de datos y también de los almacenes de datos ordenados en los correspondientes repositorios de metadatos (fase 2). La herramienta ofrece una interfaz gráfica para establecer la correspondencia y las transformaciones entre las estructuras de entrada y de salida de datos (fase 3). Las correspondencias y transformaciones definidas se almacenan en un repositorio, también como metadatos (fase 4) y, a partir de estos, junto con las plantillas y las características de los entornos definidas en la configuración, se genera el código que implementa las transformaciones (fase 5). De este modo, si se necesita hacer cualquier cambio, basta con acceder al repositorio de definición de correspondencias y transformaciones, hacer las modificaciones necesarias sobre los metadatos y regenerar de manera automática el código que los implementa.

#### 4.4.2. Ventajas e inconvenientes de las herramientas

Tanto el hecho de usar una herramienta como de llevar a cabo un desarrollo manual presentan una serie de ventajas e inconvenientes. En la situación de nuestra organización, deberemos determinar qué solución es más conveniente.

Las ventajas de hacer el desarrollo manual son las siguientes:

- Podemos empezar el desarrollo inmediatamente.
- Sea cual sea nuestro entorno, los desarrollos manuales se pueden adaptar a cualquier tipo de situación.

Los inconvenientes del desarrollo manual son estos:

- Hay muchos programas que es necesario construir, y todos tienen una estructura similar; no es motivador para los desarrolladores estar involucrados en este tipo de desarrollo.
- Los metadatos asociados a los programas deben desarrollarse de manera explícita, como una tarea adicional, por lo que presentan el mismo problema que la mayoría de los desarrollos manuales. Directamente no se desarrollan los metadatos para reducir el coste o el plazo de ejecución o, si se desarrollan, no se actualizan con los cambios que se van produciendo.
- Los programas requieren bastante tiempo para desarrollarse.
- Los requerimientos cambian constantemente y los programas deben adaptarse a los mismos. Por lo tanto, las modificaciones son muy frecuentes y suelen ser costosas.
- El coste total de desarrollo es muy alto.

Las ventajas principales del uso de herramientas son las siguientes:

- Los programas de extracción y transformación se pueden construir rápidamente.
- Gracias al hecho de que almacenan las definiciones hechas en forma de metadatos, los programas generados se pueden mantener rápida y fácilmente.
- Los metadatos asociados se producen y se mantienen de manera automática.
- Los desarrollos tienen más portabilidad. En caso de cambiar de plataforma de fuente o de destino de los datos, los metadatos que definen las co-

#### Lectura complementaria

Podéis encontrar más detalles sobre este tema en la obra siguiente:

**W. H. Inmon; J. D. Welch; K. L. Glassey (1997).** *Managing the Data Warehouse*. EE. UU.: John Wiley & Sons, Inc.

rrespondencias y transformaciones continúan siendo válidos. Es suficiente con regenerar el código teniendo en cuenta los nuevos parámetros de las plataformas de destino.

- Los costes de desarrollo se reducen de manera significativa.

Las herramientas del mercado presentan distintos inconvenientes:

- Están preparadas para que se utilicen en entornos muy generales; si nuestro entorno tiene características particulares (como es habitual que suceda), es necesario adaptarlas. En caso de soportarla, esta operación puede ser muy compleja y costosa.
- Es preciso dedicar un periodo a la formación de los desarrolladores que las tienen que usar.
- Su precio es muy alto. No están muy extendidas. Aunque potencialmente tienen un mercado amplio, actualmente no lo es tanto, en parte por su alto precio a causa de que todas las empresas desarrolladoras siguen una política de precios similar. Esto resulta problemático al hacer el desembolso inicial. Aun así, su coste se compensa con el esfuerzo de desarrollo que ahorran.

Por lo tanto, para empezar a desarrollar un almacén de datos, la solución que suelen adoptar muchas organizaciones es desarrollar el software del componente de integración y transformación de manera manual (o con el apoyo automático de cualquier otro desarrollo). Esta solución puede resultar adecuada al principio, para construir algún almacén de datos departamental. Sin embargo, en ámbitos más amplios o para diferentes almacenes de datos, los costes de desarrollo y las limitaciones en cuanto al tiempo de respuesta ante cambios se disparan. Por este motivo, es recomendable adquirir una herramienta de apoyo desde el principio, y planificar el tiempo necesario para adaptarla a nuestro entorno.

Es conveniente utilizar una herramienta como apoyo al desarrollo del componente de integración y transformación, puesto que se requiere que sea muy flexible y se adapte rápidamente a los cambios producidos en el resto de los componentes de la FIC.

#### **4.4.3. Otras herramientas de apoyo**

Además de las herramientas anteriores, hay otras herramientas especializadas para ofrecer soporte a las operaciones de depuración e integración de algunos tipos de datos, como por ejemplo las que hacen referencia a nombres y domicilios. Estas herramientas están basadas en diccionarios específicos para

cada idioma o entorno en el que almacenan nombres de personas, apellidos, nombres de empresas, nombres de calles, etc., así como distintas maneras de representarlos y abreviarlos o maneras erróneas que a veces se utilizan. Generalmente, usan patrones para reconocer los diferentes tipos de ocurrencia que se producen y de este modo identificar los sinónimos.

#### **4.5. Uso de metadatos**

Los metadatos son el componente de la FIC que permite mantener unidos el resto de los componentes. Tenemos metadatos asociados a cada uno de los elementos de la FIC; aun así, el componente de integración y transformación es el que depende de los metadatos en mayor medida, puesto que interacciona con el resto de los componentes de la FIC.

El componente de integración y transformación emplea los metadatos definidos en el resto de los componentes para obtener las estructuras de entrada y salida con las que tiene que trabajar. Las correspondencias definidas en forma de transformaciones entre las estructuras de entrada y salida también se pueden definir en forma de metadatos, a partir de las cuales se genera el código que hace las operaciones requeridas. Los metadatos proporcionan flexibilidad, puesto que para hacer cualquier cambio basta con modificar los metadatos correspondientes y generar de nuevo el código que implementa las operaciones (podéis ver la figura anterior).

En este contexto, la definición de metadatos se relaciona con el hecho de que haya alguna herramienta que lo soporte; aun así, los metadatos también se pueden generar manualmente. En el caso del componente de integración y transformación, aunque sería laborioso, se podría definir los metadatos asociados a las transformaciones de manera manual y, a partir de estos, generar el código que los implementa. A pesar de esto, es más habitual que, si no se dispone de una herramienta que haga esta tarea, el desarrollo del componente de integración y transformación se lleve a cabo mediante la codificación directa de las transformaciones necesarias.

Aunque se desarrolle manualmente el código que hace las operaciones del componente de integración y transformación, este puede tener asociados, y debería tenerlos, metadatos en forma de documentación que facilite su comprensión posterior.

## 5. Desarrollo del almacén de datos corporativo

El almacén de datos corporativo ofrece una visión integrada e historizada de los datos de la organización. Su misión es almacenar los datos para suministrarlos a los distintos almacenes de datos departamentales.

En este apartado, estudiaremos diferentes aspectos relacionados con el desarrollo del almacén de datos corporativo. Repasaremos brevemente lo que ya hemos estudiado relacionado con este punto y profundizaremos en un elemento que aún no hemos estudiado: el modelo de datos del almacén de datos corporativo.

### 5.1. Revisión del proceso de desarrollo

Según la estrategia presentada en este módulo, el almacén de datos corporativo se desarrollará de manera gradual. En un primer momento, mediante el proyecto global de desarrollo, tendremos una visión general de los datos que contendrá el almacén de datos corporativo. Por lo tanto, diseñaremos su esquema a alto nivel y definiremos las entidades que aparecerán, sin entrar en el detalle de los atributos concretos. Asimismo, podremos hacer una previsión del volumen de datos que llegará a tener y determinar qué máquina y qué SGBD lo podrán soportar.

Mediante el proyecto de desarrollo de infraestructura obtendremos la máquina, e instalaremos y configuraremos la SGBD.

El diseño inicial del esquema del almacén de datos corporativo guiará su desarrollo. Lo refinaremos con el desarrollo de los sucesivos proyectos autónomos en que dividimos la construcción de la FIC. El trabajo hecho en estos proyectos estará supervisado por el arquitecto de la FIC, junto con el administrador de bases de datos responsable del entorno. A medida que vamos poniendo en marcha estos proyectos, el almacén de datos corporativo se irá poblando de datos.

Construiremos el almacén de datos corporativo de manera gradual. Partimos de un diseño inicial que refinaremos con los diferentes proyectos de desarrollo de almacenes de datos departamentales que también lo poblarán de datos.

## 5.2. El modelo de datos del almacén de datos corporativo

Los almacenes de datos departamentales están orientados al acceso de sus datos por parte de los analistas, y por este motivo están diseñados según el modelo de datos multidimensional, que ha sido definido para permitir hacer consultas complejas de manera simple. Aun así, la principal misión del almacén de datos corporativo es almacenar y suministrar los datos necesarios a los almacenes de datos departamentales, es decir, está orientado al almacenamiento de los datos. ¿Según qué modelo de datos lo diseñaremos?

Una posibilidad consiste en utilizar el modelo de datos multidimensional del mismo modo que lo hacemos para los almacenes de datos departamentales. Este modelo se ha pensado de manera exclusiva para hacer consultas centradas en los hechos (los hechos son el foco de atención y el modelo permite almacenar datos históricos sobre estos muy fácilmente); aun así, no presenta esta facilidad para almacenar datos históricos sobre las dimensiones que califican los hechos. Aunque podemos hacer consultas sobre el almacén de datos corporativo, este no es su principal objetivo. Tendremos que utilizar un modelo que sea más adecuado para el almacenamiento de datos, particularmente de datos históricos.

Otra posibilidad es utilizar un modelo conceptual de los que usamos para construir sistemas operacionales: entidad/relación, orientado a objetos, etc. Estos modelos nos permitirán definir las entidades que aparecen y las relaciones que hay entre estas. Con estos modelos, llevaremos a cabo diseños similares a los que hacemos para los sistemas operacionales.

La diferencia sustancial aparecerá a la hora de implementar los esquemas diseñados, es decir, al situarnos en el ámbito lógico. Por ejemplo, si empleamos el modelo relacional, para construir un sistema operacional aplicaríamos las reglas de normalización para obtener un esquema normalizado. Uno de los objetivos de normalizar es evitar tener redundancia de datos. De este modo, si tenemos que modificar los datos correspondientes a una entidad, bastará con hacerlo en un solo lugar (puesto que solo están almacenados en una tabla y referenciados por otros). En el almacén de datos corporativo no haremos modificaciones sobre los datos almacenados: solo almacenaremos imágenes sucesivas de los datos. Por lo tanto, normalizar para evitar problemas con las modificaciones no tiene razón de ser. De este modo, podremos hacer diseños no normalizados si se adaptan mejor a la estructura de los datos que hay que almacenar.

### Almacén de datos corporativo

También hay analistas que hacen consultas sobre el almacén de datos corporativo. Para más información, recordad el módulo "La factoría de información corporativa".

### Ved también

Podéis encontrar más información en el módulo titulado "Diseño multidimensional" de esta asignatura.

Para implementar el almacén de datos corporativo podemos utilizar el modelo relacional. Aun así, no será necesario normalizar el diseño del esquema, puesto que no haremos modificaciones sobre los datos almacenados.

También podremos hacer la implementación utilizando el modelo orientado a objetos; en este caso, la implementación estará más cerca del diseño conceptual que hayamos hecho. Una de las ventajas principales de las bases de datos orientadas a objetos es que permiten asociar directamente los datos con el código que las trata, particularmente para hacer modificaciones sobre estas. Por otro lado, ofrecen más facilidad que las bases de datos relacionales para implementar esquemas muy complejos. En el almacén de datos corporativo, los esquemas diseñados no serán especialmente complejos. Además, no tendremos operaciones de modificación asociadas a los datos. Como conclusión, al implementar el almacén de datos corporativo en una base de datos orientada a objetos, no aprovechamos las ventajas que ofrece este tipo de base de datos.

No hay nada en contra de implementar el almacén de datos corporativo sobre una base de datos orientada a objetos. Aun así, por las características del sistema que debe construirse, no podemos aprovechar las ventajas principales que ofrece este tipo de base de datos.

### 5.3. Transformaciones para construir el esquema del almacén de datos corporativo

En este apartado, estudiaremos el conjunto de transformaciones que aplicaremos a los esquemas de las fuentes de datos operacionales para implementar el esquema del almacén de datos corporativo. Construiremos el esquema en cualquiera de los modelos lógicos mencionados en el apartado anterior.

El objetivo de estas transformaciones es doble:

- Por un lado, necesitamos que el esquema diseñado permita reflejar la evolución producida en los datos y sus relaciones, es decir, que almacene la historia de los datos.
- Por otro, puesto que solo haremos consultas sobre los datos, el objetivo es que el diseño esté optimizado para que estos consuman el menor tiempo posible. Todo esto teniendo presente que el diseño del esquema puede estar desnormalizado.

#### Lectura complementaria

Podéis encontrar más información sobre este tema en: **L. Silverston; W. H. Inmon; K. Graziano (1997). *The Data Model Resource Book*. EE. UU.: John Wiley & Sons, Inc.**

#### El esquema

No habrá modificaciones en los datos, sino solo una carga masiva de los datos y las consultas.

## Modelo del almacén de datos corporativo

En Silverston, Inmon y Graziano (1997), *The Data Model Resource Book*, la primera tarea que se propone para definir el modelo del almacén de datos corporativo a partir de un modelo de datos corporativo operacional es la de eliminar los datos operacionales que se crea que no serán utilizados por los analistas. Es así porque proponen construir el modelo del almacén de datos corporativo en un solo paso, centrándose en el almacenamiento de los datos. Nosotros lo construiremos en diferentes pasos, uno para cada proyecto autónomo que desarrollemos, y en cada paso definiremos en el almacén de datos corporativo de manera exclusiva los datos que necesiten los analistas del almacén de datos departamental asociado al proyecto autónomo. Por lo tanto, no debemos eliminar datos operacionales de manera explícita, puesto que incluiremos exclusivamente los datos que necesitamos para el proyecto.

### 5.3.1. Adición de un elemento de tiempo

Siempre tendremos que añadir al menos un elemento de tiempo a cada unidad de datos en el almacén de datos corporativo. El objetivo principal es expresar de alguna manera el periodo de validez de los datos en el entorno operacional, puesto que este es el tiempo que requieren los analistas para estudiar la evolución de los datos.

Además del periodo de validez, podremos disponer de otros datos relacionados con el tiempo; por ejemplo, el momento de la extracción de los datos, de su carga, de la disponibilidad para el analista, etc.

En algunos casos podremos obtener este tiempo del entorno operacional, puesto que habrá quedado registrado el momento en el que se han producido los cambios (es así en todos los métodos de extracción de datos inmediatos; en el de huella de tiempo también, a pesar de que en este caso podemos haber perdido movimientos). En otros, lo tendremos que añadir con el componente de integración y transformación.

Podemos expresar el periodo de validez de los datos de diferentes maneras, entre las cuales las más frecuentes son las siguientes:

- Utilizando dos campos, uno para indicar el momento de inicio y otro para el momento final.
- Mediante un solo campo que indique el momento de inicio. El momento final de validez corresponderá al momento de inicio de otro valor para los mismos datos, en caso de que lo haya.
- Utilizando un solo campo que indique el periodo de validez (por ejemplo, los datos del mes de enero de este año).

Tendremos que añadir elementos de tiempo a diferentes grados de granularidad de los datos, según si lo requieren.

#### Comparación de imágenes

Si obtenemos las modificaciones de los datos mediante el método de comparación de imágenes, tendremos que añadir el tiempo. Generalmente, será aquel en el que se hizo la extracción de los datos.

- Desde el punto de vista del fichero: todos los datos de un fichero corresponden al periodo de validez indicado. Por ejemplo, almacenamos los datos de cada mes en un fichero diferente.
- Desde el punto de vista del registro: para cada registro de un fichero podemos indicar el periodo de validez como parte de la clave del registro.
- Desde el punto de vista del campo: cada campo de un registro tiene asociados los campos necesarios para expresar el periodo de validez.

#### Modelo relacional

Si utilizamos el modelo relacional, por la limitación en sus prestaciones, podremos definir la granularidad en el ámbito de campo solo cuando sepamos el número máximo de ocurrencias que podemos tener. Esto no es así para el modelo orientado a objetos.

Asimismo, los campos que añadiremos para indicar el tiempo tendrán granularidades distintas, según se requiera. En algunos casos bastará con indicar el año, y en otros se tendrá que indicar el tiempo en segundos.

Cada dato del almacén de datos corporativo que lo requiera ha de tener asociado un periodo de validez. De este modo, los analistas pueden estudiar su evolución.

### 5.3.2. Organización de los datos según su estabilidad

Debemos tener presente que almacenaremos la evolución de todos los datos de la organización en una base de datos, el almacén de datos corporativo. Si no prestamos especial atención al espacio requerido por las soluciones que proponemos, el tamaño total del resultado se puede disparar. Concretamente, en relación con el apartado anterior, tenemos que prestar especial atención a qué grado de granularidad añadimos a un elemento de tiempo para almacenar la historia.

#### Estabilidad y grado de granularidad de los datos

En un banco, hemos diseñado la entidad cliente en el almacén de datos corporativo mediante una tabla en la que tenemos los datos personales del cliente (dirección, teléfono, etc.) junto a la suma del saldo total de sus cuentas y dos campos de tiempo que indican el periodo de validez de los datos (granularidad de los datos desde el punto de vista del registro). De este modo, cada vez que cambie alguno de los datos del cliente, tenemos que almacenar un nuevo registro con los nuevos datos. Si el saldo cambia de media una vez al día para cada cliente, cada día almacenaremos una nueva versión de cada registro completo, incluyendo los datos personales que cambian una vez cada cinco años de media. Es decir, malgastamos mucho espacio, pues hemos almacenado 1.825 veces ( $365 \times 5$ ) la misma dirección.

Este mismo diseño en un sistema operacional puede presentar otros problemas, pero no presentaría problemas de espacio. En este caso, si un atributo de un registro cambia, se pierde el valor anterior y no se almacena una versión completa del registro.

En el ejemplo anterior, los datos del registro de cliente tienen una estabilidad distinta. Aun así, hemos definido la granularidad de almacenamiento de cambios desde el punto de vista del registro, sin tener este hecho en cuenta. Para optimizar la cantidad de datos almacenados, deberíamos haber considerado, por un lado, el campo saldo y, por otro, los datos personales, puesto que tienen estabilidades distintas. Si implementásemos este ejemplo en una base de

datos relacional, deberíamos tener en una tabla los datos personales y en otra el saldo, y definir en los dos casos la granularidad desde el punto de vista del registro, porque no sabemos el número máximo de cambios que puede tener el saldo y, por este motivo, no podríamos tener la granularidad desde el punto de vista del campo.

Para optimizar el espacio ocupado en el almacén de datos corporativo, dentro de cada entidad tenemos que definir la granularidad de los datos a los que añadimos un elemento de tiempo según su grado de estabilidad ante los cambios.

### 5.3.3. Adición de datos derivados

Los modelos operacionales generalmente no incluyen datos derivados, es decir, aquellos obtenidos a partir de otros datos almacenados en la base de datos. No tiene mucho sentido almacenar datos derivados en los casos en que los valores base de los cálculos pueden cambiar con frecuencia, por lo que se tendrían que recalcularse cada vez que se produjera un cambio.

Las principales ventajas de almacenar datos derivados en la base de datos son las siguientes:

- Se tiene más velocidad de acceso a estos datos, puesto que no es preciso calcularlos cada vez que se necesitan.
- Se evitan errores. Si directamente proporcionamos los datos, evitamos que alguien los calcule de manera errónea.

Los inconvenientes son los siguientes:

- Si los datos base cambian, deben calcularse los datos derivados a partir de estos.
- Se ocupa más espacio de almacenamiento.

Por definición, los datos almacenados en el almacén de datos corporativo no cambian. Por este motivo, si incluimos datos derivados, salvo situaciones de error, no se tendrán que recalcularse y no tendremos este inconveniente. De este modo, el principal inconveniente de incluirlos es el problema del espacio de almacenamiento que requieren.

Las ventajas que aporta incluir datos derivados a los que acceden los analistas son más elevadas que el coste de almacenamiento, especialmente si calcularlos es complejo y costoso.

#### Modelos operacionales

A veces los modelos operacionales sí incluyen datos derivados para optimizar las consultas, pero es más frecuente que no sea así, aunque muchas veces sería conveniente que lo hicieran.

#### Algoritmo de cálculo

Facilitando el algoritmo de cálculo, también evitamos que alguien calcule los datos de manera errónea.

Por otro lado, también podemos almacenar estos datos derivados en los diferentes almacenes de datos departamentales, lo cual será el medio de acceso para la mayoría de los usuarios de la FIC.

Calcularemos los datos derivados que los analistas necesiten y los almacenaremos en el almacén de datos corporativo, o bien directamente en los almacenes de datos departamentales.

El concepto denominado por Inmon, índice creativo, está asociado a los datos derivados. El planteamiento es el siguiente: dado que para pasar los datos desde las fuentes de datos operacionales al almacén de datos corporativo tenemos que trabajar con los datos al más bajo nivel, con un poco de trabajo adicional podríamos calcular a partir de estos una serie de datos interesantes para los analistas. Es decir, precalculamos requerimientos de los analistas que podemos anticipar y los almacenamos en el almacén de datos corporativo.

#### Ejemplo de índice creativo

Por ejemplo, en un banco, las cuentas que menos actividad tienen, los ingresos más altos, etc.

A diferencia de los datos derivados comentados anteriormente, los datos calculados para el índice creativo son aquellos que no se consideran en el entorno de las aplicaciones operacionales: los obtenidos de manera exclusiva para los analistas.

#### 5.3.4. Cambio en la granularidad de los datos

A la hora de diseñar el almacén de datos corporativo, podemos tener tendencia a incluir todos los datos disponibles en el nivel de detalle más bajo que podamos obtener de las fuentes de datos operacionales. De este modo, intentamos prever necesidades futuras que pueden ser muy improbables. El resultado es que el espacio requerido se puede disparar.

El almacén de datos corporativo debe cubrir las necesidades actuales de los usuarios. Antes de añadir cualquier dato que no se necesite, debemos valorar las repercusiones de coste que esto puede tener.

Concretamente, en cuanto a la granularidad, hemos de adaptar el nivel de detalle de los datos que obtenemos de las fuentes de datos operacionales al requerido por los usuarios.

#### Adaptación de la granularidad de los datos de los clientes

En el ejemplo del banco, para cada cliente los analistas requieren estudiar la evolución de su saldo mensual. Particularmente, utilizan para cada mes los valores del saldo máximo, mínimo y medio. A partir del sistema operacional, podemos obtener todos los valores que toma el saldo a lo largo de cada día. Si los analistas conocen la posibilidad de tener los datos con un nivel más bajo, pero consideran que solo necesitan los datos de ámbito mensual<sup>16</sup>, para almacenar el saldo de cada cliente obtendríamos los valores requeridos por los analistas agregando todos los valores obtenidos para cada mes, y serían exclusivamente estos valores calculados los que almacenaríamos en el almacén de datos corporativo.

<sup>(16)</sup> Es poco probable que sea así, pero puede servir como ejemplo.

Con el objetivo de minimizar el espacio ocupado, el almacén de datos corporativo tiene que almacenar los datos con la granularidad que necesiten los analistas, no con la granularidad más baja que podemos conseguir de las fuentes de datos operacionales.

### 5.3.5. Fusión de entidades

Una de las transformaciones más frecuentes que se hacen sobre los datos en los sistemas operacionales para incluirlos en el almacén de datos corporativo es la de fusionarlos en una sola entidad. Es decir, diferentes entidades de los sistemas operacionales pasan a estar representadas por una sola entidad en el almacén de datos corporativo.

Esta fusión se puede hacer por diferentes motivos:

- Las entidades provienen de diferentes sistemas operacionales que se integran en una entidad en el almacén de datos corporativo, puesto que conceptualmente son la misma.
- En los sistemas operacionales, tenemos los datos organizados para que estos soporten de manera óptima las operaciones de modificación, generalmente dividiéndolos en entidades más pequeñas. Al pasar al almacén de datos corporativo podemos agrupar estas entidades, puesto que no tendremos problemas con las modificaciones. Concretamente, dado que no se producen cambios en los datos, sabemos el número de ocurrencias que tenemos y así podemos definir atributos para contener múltiples ocurrencias. De este modo, se facilita el acceso a los datos.
- Puede resultar más fácil representar la evolución de las relaciones entre dos entidades almacenándolas juntas, y definir una sola entidad a partir de estas. En este caso, fusionar entidades puede representar replicar datos.

#### Fusión de datos

Haremos la fusión de datos siempre que tengan niveles de estabilidad similares.

En el almacén de datos corporativo podemos tener los datos desnormalizados y replicados, puesto que no se producirán modificaciones. Mediante la fusión de entidades operacionales se facilita el acceso a los datos, dado que se puede acceder directamente a los mismos en lugar de tener que seguir las relaciones definidas entre los datos para obtenerlos. En caso de replicar datos, tenemos que evaluar el coste que representa en espacio adicional requerido.

## 6. Incorporación del almacén de datos operacional

De manera tradicional, el entorno operacional se ha estructurado en forma de aplicaciones independientes. Para adaptarse a los requerimientos del entorno, las organizaciones necesitan tener una visión integrada de sus datos.

### Aplicaciones no integradas en un banco

Un banco dispone de aplicaciones independientes para gestionar las tarjetas de crédito, créditos hipotecarios, créditos personales, cuentas de ahorro, etc. Cada aplicación funciona en su plataforma, aunque los trabajadores pueden acceder a la misma desde su terminal. Si un usuario necesita conocer toda la información asociada a un cliente (por ejemplo, el director de una oficina o el empleado de atención telefónica a los clientes), tiene que acceder a cada una de las aplicaciones para hacer la consulta correspondiente. Si las aplicaciones estuvieran integradas, bastaría con una sola consulta para obtener los datos personales y comerciales del cliente.

El almacén de datos operacional ofrece una visión integrada de los datos operacionales de la organización; a diferencia del almacén de datos corporativo, no almacena la historia de los datos, es volátil y está actualizado. Su misión principal consiste en ofrecer apoyo operacional a la organización. Lo utilizarán principalmente usuarios operacionales (oficinistas), aunque también lo hacen los analistas.

En este apartado, estudiaremos diferentes tipos de almacenes de datos operacionales, lo que aporta este concepto a la organización y también alternativas que tiene la organización para disponer de los mismos.

### 6.1. Paquetes de aplicaciones y el almacén de datos operacional

En el mercado, encontramos paquetes de aplicaciones que ofrecen una visión integrada del entorno operacional. Dado que la mayoría de las organizaciones en un sector tienen necesidades similares, la idea general de estos productos es ofrecer un conjunto de aplicaciones integradas que cubren las necesidades generales de una organización prototipo dentro de un determinado sector. En muchos casos, las aplicaciones se pueden adaptar a las características particulares del cliente que las adquiere, generalmente a un coste alto. Con frecuencia, son las organizaciones las que se tienen que adaptar a la manera de trabajar impuesta por las aplicaciones.

Algunas organizaciones intentan "liberarse" de sus antiguas aplicaciones mediante la adquisición de paquetes de aplicaciones integradas. Aun así, es frecuente que organizaciones que adquieren uno de estos paquetes deban conservar parte de sus aplicaciones tradicionales porque las nuevas no cubren totalmente sus necesidades, o bien se ha decidido adquirir solo una parte del

#### Ved también

Hemos estudiado más detalles sobre este tema al principio de este módulo.

#### Paquetes más conocidos

Entre los paquetes más conocidos, encontramos los ofrecidos por SAP y Oracle, pero son muchos más.

#### El euro

Muchas organizaciones han adquirido un paquete para solucionar el problema del año 2000 y adaptar el funcionamiento al euro.

paquete de aplicaciones. En estos casos, tienen que convivir los dos entornos y generalmente las aplicaciones del paquete adquirido han de obtener datos de las aplicaciones que había previamente.

¿Qué relación hay entre el almacén de datos operacional y los paquetes de aplicaciones integradas?

Los dos ofrecen una visión integrada de los datos del entorno operacional. Las características de estos también son idénticas para los dos y en los dos casos el objetivo principal es ofrecer apoyo operacional a la organización.

Los paquetes de aplicaciones integradas cuyo ámbito es toda la organización son implementaciones comerciales del almacén de datos operacional.

Algunos de los paquetes comerciales que hay en el mercado están diseñados teniendo en cuenta la arquitectura de la FIC y ofrecen medios para integrarse en la misma, particularmente para obtener los datos para construir el almacén de datos corporativo. En otros casos, van más lejos y ofrecen como ampliación del paquete básico una versión genérica de la FIC para las organizaciones del sector. Aun así, en lo que respecta al apoyo a la toma de decisiones de los analistas, ofrecer un apoyo estándar que sea adecuado para todos los analistas es mucho más complejo que lo que se refiere al apoyo operacional de la organización.

Si construimos el almacén de datos operacional, una vez diseñada la base de datos que lo soporta y los métodos de obtención y refresco de sus datos a partir de las aplicaciones operacionales, construiremos aplicaciones sobre este. En algunos casos, las nuevas aplicaciones cubrirán necesidades nuevas, y en otros sustituirán aplicaciones antiguas como ocurre con los paquetes de aplicaciones integradas.

El almacén de datos operacional puede ser adquirido en forma de paquete de aplicaciones integradas o construido en la organización.

## 6.2. Velocidad de refresco de los datos

En el caso de que conviva el almacén de datos operacional con las aplicaciones operacionales, caso frecuente en las organizaciones, este tiene que obtener parte de sus datos a partir de los datos de las aplicaciones. Inmon distingue cuatro clases de almacenes de datos operacionales según la velocidad de refresco de los datos.

- Clase I: se actualiza unos segundos después de que se produzcan las modificaciones en las aplicaciones operacionales.
- Clase II: se actualiza unas horas después de que se produzcan las modificaciones.
- Clase III: el periodo de actualización es superior a las veinticuatro horas.
- Clase IV: se actualiza de manera no planificada.

En los almacenes de datos operacionales de clase I, no se pueden acumular los movimientos obtenidos en las fuentes de datos operacionales para aplicarlos de manera masiva. Se tiene que disponer de un medio de trasladar las modificaciones directamente (mediante disparadores, RPC<sup>17</sup> u otros medios). Generalmente, no se aplican demasiadas transformaciones a los datos en los almacenes de esta clase.

<sup>(17)</sup>Del inglés *remote procedure call*, 'llamada a procedimientos remotos'.

Para actualizar los almacenes de datos operacionales de clase II, III y IV podemos aplicar los métodos de obtención de actualizaciones a partir de las fuentes de datos operacionales estudiadas en este módulo. Para los de clase II, en la mayoría de los casos necesitaremos un método de captura inmediata. Para los de clase III y IV, podremos utilizar cualquiera de los métodos estudiados; además, puesto que no guardamos la historia de los datos, no hay ningún problema para utilizar un método de captura retardada aunque perdamos detalle de los movimientos.

Cuanto más restrictivas sean las condiciones de actualización, más costará implementarlas. Por este motivo, un almacén de datos de clase I será mucho más caro que otro de clase IV.

### **6.3. Planificación de la incorporación del almacén de datos operacional**

Plantear la adquisición o la construcción del almacén de datos operacional como un solo proyecto presenta el grave inconveniente de la dificultad de justificar su coste ante la organización. En algunas situaciones especiales como las que hemos vivido en los últimos años, con el efecto 2000 y la adaptación al euro, puede resultar más fácil de justificar.

En situaciones normales, es más adecuado hacer un desarrollo iterativo similar al planteado para desarrollar el resto de la FIC. Frecuentemente, se puede plantear el almacén de datos operacional como una estructura de apoyo para actualizar los datos del almacén de datos corporativo. En estos casos se construirá, como parte de los proyectos atómicos en los que se ha dividido la construcción de la FIC, la parte correspondiente al almacenamiento de los datos, de manera conjunta con el almacén de datos corporativo. Asimismo, dentro de la construcción de la FIC, podemos plantear proyectos de desarrollo o de ampliación del almacén de datos operacional, con las mismas premisas con las que definimos el resto de los proyectos.

Si se planifica su construcción antes de disponer del almacén de datos corporativo, puede servir para construirlo. Aun así, se corre el riesgo de pretender incorporar al almacén de datos operacional funcionalidades propias del almacén de datos corporativo.

Dado que el coste resulta difícilmente justificable, es recomendable adquirir o desarrollar el almacén de datos operacional de manera iterativa, mediante proyectos autónomos.

## Resumen

Conocer la arquitectura de la FIC representa un avance para las organizaciones que quieren ofrecer herramientas de apoyo a la toma de decisiones para los analistas. Aunque se pueden considerar arquitecturas alternativas, estas presentan graves deficiencias que las hacen inviables en la mayoría de las situaciones.

La arquitectura de la FIC se tiene que trazar desde un nivel alto, teniendo en cuenta la perspectiva de toda la empresa. Aun así, es un error plantear el desarrollo de la FIC como un solo proyecto: este se tiene que hacer de manera iterativa, por medio de proyectos independientes con objetivos y beneficios claros. Es decir, los sucesivos proyectos de desarrollo de la FIC hacen que su funcionalidad aumente con el tiempo y, con esta, el beneficio que aportan a la organización.

En este módulo, además de estudiar diferentes alternativas en la arquitectura de la FIC, así como distintas maneras de planificar su construcción, hemos estudiado aspectos concretos de la construcción de sus componentes. Particularmente, hemos prestado especial atención al componente de integración y transformación por su dificultad; asimismo, hemos estudiado las peculiaridades de la construcción del almacén de datos corporativo y del almacén de datos operacional.

## Actividades

En el módulo de diseño multidimensional, se hará un diseño a partir de uno o varios sistemas operacionales; para este mismo caso:

1. Diseñad el modelo del almacén de datos corporativo según las transformaciones descritas en el punto 5.
2. Implementad este modelo sobre una base de datos relacional.
3. Implementad el componente de integración y transformación empleando cualquier lenguaje de programación.
  - a) Carga inicial del almacén de datos corporativo y departamental.
  - b) Actualización de los datos.

## Ejercicios de autoevaluación

1. ¿Podemos adquirir la FIC?
2. ¿En qué consiste el entorno operacional de telaraña?
3. ¿Podemos construir en primer lugar distintos almacenes de datos departamentales, y posteriormente el corporativo?
4. ¿Es adecuado combinar el almacén de datos operacional y el corporativo en una sola estructura?
5. ¿Tiene que incluir la FIC todos los componentes estudiados en el módulo "La factoría de información corporativa"?
6. ¿Podemos construir la FIC con un solo proyecto?
7. ¿Qué características deben tener los proyectos de construcción de la FIC?
8. ¿Qué estructura deben tener los proyectos de construcción de la FIC?
9. ¿Según qué metodología desarrollaremos los proyectos de construcción de la FIC?
10. ¿Qué ocurre con el entorno operacional cuando desarrollamos la FIC?
11. ¿Quién es el patrocinador de la FIC?
12. ¿Cuál es el principal inconveniente de los métodos de captura retardada para obtener las modificaciones producidas en los datos de las fuentes de datos operacionales?
13. ¿Qué características tiene el modelo de datos del almacén de datos corporativo?
14. ¿Qué relación hay entre el almacén de datos operacional y los paquetes de aplicaciones integradas?
15. ¿Qué modelo de datos tiene el almacén de datos operacional?

## Solucionario

### Ejercicios de autoevaluación

1. Generalmente, no. La FIC se tiene que construir en cada organización. Algunos paquetes de aplicaciones tienen como ámbito toda la organización, incluyen funcionalidades de la FIC y están pensados para cubrir las necesidades de una organización estándar. Si el sistema de información de la organización está implementado mediante un paquete estándar que incluye la FIC, en este caso la FIC habría sido adquirida. Esta es una situación límite; generalmente, la FIC se tiene que construir en las diferentes organizaciones.
2. Es el resultado de la evolución incontrolada del entorno operacional. Mediante la construcción de programas de extracción de información y bases de datos temporales para elaborar informes puntuales, llegamos a un entorno de alta complejidad con una estructura de relaciones entre los diferentes componentes que se asemeja a una telaraña.
3. La construcción de almacenes de datos departamentales debe de estar basada en el almacén de datos corporativo. Si intentamos construir el almacén de datos corporativo a partir de los diferentes almacenes de datos departamentales que había previamente, la integración de los datos de estos será muy compleja y seguramente también requerirá la modificación posterior de los almacenes de datos departamentales.
4. Aunque tienen algunas características similares, sus objetivos son distintos y el tipo de operaciones a las que ofrecer soporte son incompatibles. El almacén de datos operacional tiene que estar configurado para hacer operaciones de modificación. Aun así, el almacén de datos corporativo necesita estar optimizado para hacer de manera exclusiva operaciones de consulta.
5. No necesariamente. El almacén de datos operacional es una estructura opcional, aunque conviene incluirla por la funcionalidad que aporta. El resto de los componentes sí son necesarios.
6. Si planteamos la construcción de la FIC mediante un solo proyecto, este será demasiado complejo y tendremos dificultades para justificar su coste. Por lo tanto, esta estrategia de construcción no será adecuada.
7. El primer proyecto debe tener como objetivo planificar la construcción de la FIC. Posteriormente, tendremos proyectos de desarrollo de infraestructura. Dividiremos la construcción de la FIC en proyectos autónomos que aporten un valor claro a la organización, que tengan un responsable dentro de esta y se desarrollen en un plazo razonable.
8. Los proyectos han de ser completos, es decir, tienen que prever la obtención y el almacenamiento de los datos y el acceso a los mismos.
9. La metodología de desarrollo más adecuada es una metodología iterativa (CLDS), puesto que generalmente no se conocen con todo detalle los requerimientos de los analistas de información y esta metodología nos permite descubrirlos mediante refinamientos sucesivos.
10. Cuando planificamos la construcción de la FIC, también tenemos que planificar el desmantelamiento del entorno operacional en telaraña, de modo que a medida que construimos los proyectos que implementan la FIC también eliminamos los programas de extracción de datos y las bases de datos temporales que se dejan de utilizar.
11. Es una figura política. Suele ser un directivo de alto nivel de la organización que está convencido de los beneficios que puede aportar la FIC y su misión es obtener los recursos para los proyectos cuyo beneficio no es directamente justificable y tratar de solucionar la oposición que pueda surgir en la organización a la construcción de la FIC.
12. El problema principal que presentan estos métodos es que pueden perder el detalle de los movimientos producidos. Mediante estos métodos, obtenemos un resumen de todas las operaciones hechas en una única operación de modificación.
13. Se trata de un modelo de datos orientado a almacenarlos. Los esquemas desde el punto de vista lógico están diseñados de modo que el almacenamiento y las consultas se hagan de manera óptima.
14. Los paquetes de aplicaciones integradas cuyo ámbito es toda la organización son implementaciones comerciales del almacén de datos operacional.

15. El almacén de datos operacional está construido utilizando un modelo de datos como el de las aplicaciones operacionales. Los esquemas desde el punto de vista lógico están diseñados de modo que las operaciones de modificación se hagan de manera óptima.

## Glosario

**CASE** *Computer aided system engineering.*

**CLDS** *m* Metodología iterativa de desarrollo de proyectos. Corresponde a las siglas de SDLC puestas a la inversa, como contraposición a los planteamientos de esta metodología.

**factoría de información corporativa** *f* Conjunto de elementos de software y hardware que ayudan a analizar datos para tomar decisiones.  
sigla: **FIC**

**FIC** *f* Véase **factoría de información corporativa**.

**índice creativo** *m* Conjunto de datos interesantes para los analistas calculados en el momento de pasar los datos de las fuentes de datos operacionales hasta el almacén de datos corporativo.

**llamada a procedimientos remotos** *f*  
sigla: **RPC**  
*en* remote procedure call

**main frame** *m* Ordenador central o corporativo.

**SDLC** *m* Metodología de desarrollo de proyectos según un ciclo de vida en cascada.  
*en* systems development life cycle

**sistema de gestión de bases de datos** *m* Software que gestiona y controla bases de datos. Sus principales funciones son las de facilitar el uso de las bases de datos de manera simultánea a muchos usuarios de tipos distintos, independizar al usuario del mundo físico y mantener la integridad de los datos.  
sigla: **SGBD**

**sistema decisional** *m* Aquel que apoya los procesos de toma de decisiones por parte de los analistas en la organización.

**sistema informacional** *m* Véase **sistema decisional**.

**sistema operacional** *m* Sistema que ayuda en las operaciones diarias del negocio de una organización.

## Bibliografía

- Devlin, B.** (1997). *Data Warehouse from Architecture to Implementation*. EE. UU.: Addison Wesley Longman, Inc.
- Inmon, W. H.** (1996). *Building the Data Warehouse* (2.ª ed.). EE. UU.: John Wiley & Sons, Inc.
- Inmon, W. H.** (1999). *Building the Operational Data Store*. EE. UU.: John Wiley & Sons, Inc.
- Inmon, W. H.; Imhoff, C.; Sousa, R.** (1998). *Corporate Information Factory*. EE. UU.: John Wiley & Sons, Inc.
- Inmon, W. H.; Welch, J. D.; Glassey, K. L.** (1997). *Managing the Data Warehouse*. EE. UU.: John Wiley & Sons, Inc.
- Jarque, M.; Lenzerini, M.; Vassiliou, Y.; Vassiliadis, P.** (2000). *Fundamentals of Data Warehouses*. Berlín: Springer-Verlag.
- Kelly, S.** (1997). *Data Warehousing in Action*. EE. UU.: John Wiley & Sons, Inc.
- Mattison, R.** (1996). *Data Warehousing: Strategies, Technologies and Techniques*. EE. UU.: Computing McGraw-Hill.
- Silverston, L.; Inmon, W. H.; Graziano, K.** (1997). *The Data Model Resource Book*. EE. UU.: John Wiley & Sons, Inc.

