Bases de Datos Federadas, Almacenes de Datos y Análisis Multidimensional

Alberto Abelló [a], Francisco Araque [b], José Samos [b], and Fèlix Saltor [a]

[a] U. Politècnica de Catalunya (UPC), Dept. de Llenguatges i Sistemes Informàtics
{aabello,saltor}@lsi.upc.es
[b] U. de Granada (UGR), Dept. de Lenguajes y Sistemas Informáticos
{faraque,jsamos}@ugr.es

Resumen En estas páginas presentamos nuestro trabajo, dentro del proyecto BLOOM sobre Bases de Datos Federadas, en referencia al almacenamiento de datos y análisis multidimensional.

Palabras clave: Base de Datos Federada, Base de Datos de Tiempo Real, Modelado multidimensional

1 Introducción

Los datos de interés para una empresa se encuentran, cada vez con mayor frecuencia, desperdigados en múltiples fuentes de información. Para que un usuario pueda acceder a esas fuentes de un modo integrado, es decir, como si se tratase de una única fuente, hace falta instalar un sistema que produzca un acoplamiento entre esas fuentes. Sin acoplamiento, el acceso no es integrado y hace falta acceder separadamente a cada una de ellas y luego combinar las respectivas respuestas.

En [SAO+99] se presentó un proyecto de investigación sobre *sistemas cooperativos* del que deriva todo nuestro trabajo. El acceso integrado puede hacerse de tres formas diferentes:

1. A través de una consulta que accede directamente a las fuentes preexistentes, que interoperan formando un *sistema cooperativo*.
2. Consultando un *Data Warehouse* o bien un *Operational Data Store*, en los que se vuelcan y consolidan los datos de las fuentes; en ambos casos, los datos están materializados y son gestionados por el *sistema cooperativo*.
3. Mediante una consulta que puede acceder tanto a las fuentes preexistentes como a datos materializados vistos de forma uniforme mediante un esquema común y que están bajo el control del *sistema cooperativo*.

2 Una arquitectura común

El primer paso para empezar a trabajar en el área del almacenamiento de datos se dio en [SSSB98]. Aquí se presentaron las similitudes entre las Bases de Datos

Federadas y los *Data Warehouses*, de manera que se pudieran aprovechar los avances en las primeras para favorecer los segundos (por ejemplo en materia de integración de los datos). Es decir, se pasó de trabajar en un *sistema cooperativo* del tipo 1 mencionado anteriormente, a otro del tipo 3.

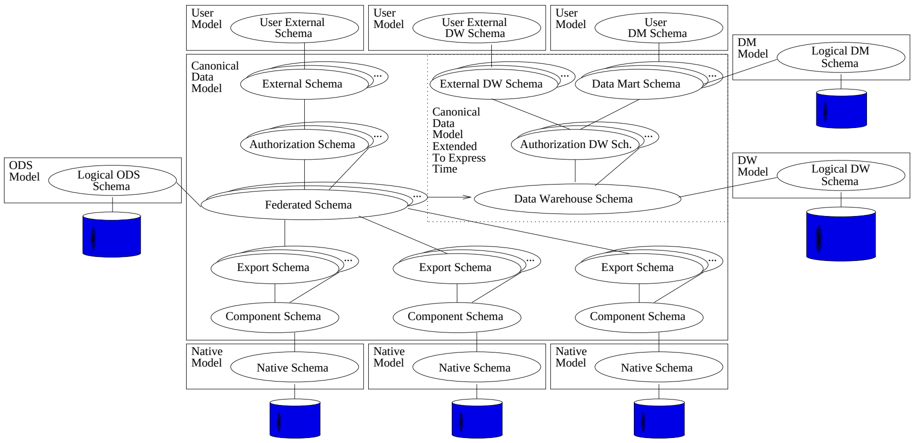

Figura 1. Arquitectura de esquemas que integra la federación y el *Data Warehouse*

En [AOSS00], se presentó la arquitectura de esquemas de la figura 1, que contempla diferentes niveles de esquemas para el almacenamiento de datos en paralelo con los niveles de esquemas necesarios en una federación. En esa arquitectura se apunta también la posibilidad de estudiar paralelamente el tratamiento de la seguridad en ambas áreas. Además, se situan los diferentes repositorios de información que aparecen en la literatura (*Operational Data Store*, *Data Warehouse* y *Data Mart*).

En este marco, el *Operational Data Store* no sería más que la materialización de los datos de un esquema federado. A partir de ese u otro esquema federado más apropiado y añadiendo elementos temporales para representar los datos históricos necesarios, se definiría el *Data Warehouse*. Finalmente, a semejanza de los esquemas externos se definirían esquemas multidimensionales para los *Data Marts*.

3 Extracción de datos

La extracción de datos de las fuentes de información de una federación se hace bajo demanda. Es decir, cada consulta de los usuarios federados se traduce en las consultas correspondientes sobre las fuentes de información componentes. La incorporación del *Data Warehouse* dentro de la arquitectura hace que esto ya no pueda ser siempre así. En este caso, la extracción de información hay que hacerla por adelantado, sin esperar las consultas de los usuarios. De esta manera, se hace necesario el estudio de la representación de los elementos que permitan expresar las características de obtención de los datos de las distintas fuentes.

Las Bases de Datos de Tiempo Real y las Bases de Datos Temporales utilizan conceptos que pueden ser aplicados en esta tarea de actualización del *Data Warehouse*. Características como temporalidad de los datos (datos temporales y datos no temporales), frecuencia de acceso a los datos (muy frecuentes, poco frecuentes), periodicidad de los datos (tareas periódicas, aperiódicas), consistencia temporal de los datos (relativa, absoluta), tiempos (usuario, transacción, válido), etc pueden utilizarse para definir las características de una fuente añadiéndole información temporal. Esto facilitaría tanto la definición del modelo de datos del *Data Warehouse* con funcionalidades de tiempo, como la carga de los datos en el mismo. En [AS99] y [Ara00] se presentó nuestro trabajo en esta línea.

4 Un modelo multidimensional Orientado a Objetos

En estudios realizados con anterioridad a la aparición de los almacenes de datos, se vieron las ventajas de los modelos Orientados a Objetos (O-O) como modelos de datos canónicos para una federación. Esto llevó a pensar que probablemente también darían buenos resultados a la hora de representar la información del almacenamiento de datos. Así, en [ASS00], se estudió la aportación que podría hacer un modelo O-O al análisis multidimensional.

Más adelante, en [ASS01a], se presentó un marco que permitía la comparacion de los diferentes elementos en los modelos multidimensionales existentes. De esta manera, se planteó la posibilidad de definir un modelo multidimensional que permitiera representar mucha más información en los esquemas. Se optó por basarse en la riqueza semántica de la orientación a objetos, con sus diferentes conceptos y relaciones. Como paso previo a la definición de un modelo multidimensional, se han estudiado por separado las dimensiones de análisis y los hechos, en [ASS01b] y [ASS01c] respectivamente. Además de definir claramente lo que son los hechos y las dimensiones, se ha estudiado su estructura y la influencia de los diferentes tipos de relaciones que podemos encontrar.

5 Conclusiones y trabajo futuro

El objetivo final es definir los modelos de datos necesarios (para el *Data Warehouse* y los *Data Marts*) que faciliten el almacenamiento de datos y su posterior análisis. Hasta ahora se ha trabajado en la definición de un modelo de datos multidimensional orientado a objetos. Esta nueva propuesta de modelo lleva a la necesidad de estudiar nuevos mecanismos para la definición de los niveles de esquemas que han aparecido en la arquitectura. Habrá que ver cómo se obtienen los datos a ser almacenados a partir de la federación, y cómo se definen y alimentan los *Data Marts* desde el *Data Warehouse*. Por otra parte, también será necesario definir los esquemas componentes enriquecidos con las características temporales de las fuentes, así como los mecanismos de definición del esquema de datos del *Data Warehouse* en base a la información que se pueda extraer de dichas fuentes.

Referencias

[AOSS00] A. Abelló, M. Oliva, J. Samos, and F. Saltor. Information System Architecture for Secure Data Warehousing. In *Proc. of the 3rd Int. Workshop on Engineering Federated Information Systems (EFIS)*, pages 33–40. IOS Press, Junio 2000.

[Ara00] F. Araque. Extracción de información de fuentes de datos heterogéneas con características temporales. In *IV Jornadas Científicas en Tecnologías de la Información (IV CINTE)*, Noviembre 2000.

[AS99] F. Araque and J. Samos. External Schemas in Real-Time Object-Oriented Databases. In *Proc. of the 20th IEEE Real-Time Systems Symposium (WIP)*. IEEE Computer Society, Diciembre 1999. ISBN 0-769-0475-2.

[ASS00] A. Abelló, J. Samos, and F. Saltor. Benefits of an Object-Oriented Multidimensional Data Model. In *Objects and Databases - Int. Symposium -*, volume 1944 of *LNCS*, pages 141–152. Springer, 2000. ECOOP, Sophia Antipolis (France).

[ASS01a] A. Abelló, J. Samos, and F. Saltor. A Framework for the Classification and Description of Multidimensional Data Models. In *Proc. of the 12th Int. Conf. on Database and Expert Systems Applications (DEXA)*, volume 2113 of *LNCS*, pages 668–677. Springer, 2001.

[ASS01b] A. Abelló, J. Samos, and F. Saltor. Understanding Analysis Dimensions in a Multidimensional Object-Oriented Model. In *3rd International Workshop on Design and Management of Data Warehouses (DMDW)*. SwissLife, 2001.

[ASS01c] A. Abelló, J. Samos, and F. Saltor. Understanding Facts in a Multidimensional Object-Oriented Model. In *Fourth International Workshop on Data Warehousing and OLAP*. ACM, 2001. Pendiente de publicación.

[SAO+99] J. Samos, A. Abelló, M. Oliva, E. Rodríguez, F. Saltor, J. Sistac, F. Araque, C. Delgado, and E. Garví. Sistema cooperativo para la integración de fuentes heterogéneas de información y almacenes de datos. *Novatica*, 142:44–49, Nov-Dic 1999.

[SSSB98] J. Samos, F. Saltor, J. Sistac, and A. Bardés. Database Architecture for Data Warehousing: An Evolutionary Approach. In *Proc. of 9th Int. Conf. on Database and Expert Systems Applications (DEXA)*, volume 1460 of *LNCS*, pages 746–756. Springer, 1998.