Data Access and Security in Multiple Heterogeneous Databases

1. Introduction

In the present scenario, many of the organizations require to share and access data, which is usually spread over multiple heterogeneous databases. As the data is typically confidential and private, there should be a secure mean to access it. In our paper, we give a brief on two approaches. In both the approaches the autonomy of the database is preserved.

The first approach is called Federated Database Management System (FDBMS). A federation is a closed system of members who access data from one another. This approach shares both Data and Metadata with the participating entities [1]. Therefore the participating entities ‘trust’ each other. Section 2 discusses briefly the Data accessing in FDBMS.

On the other hand, sharing of metadata may not be acceptable for certain organizations due to privacy concerns. Mediation is the other approach that uses independent components called as Mediators between the participating entities. Section 3 gives an overview of data accessing in Mediated Systems.

We then discuss how Mediation is made secure and then contrast both the approaches. Finally we take a position on the basis of the comparison, relative merits and performance.

2. Federation Database Architecture

The first approach used for accessing data in multiple heterogeneous databases is by federation. A Federated database system is a distributed system that provides global interface to heterogeneous local database management systems. All databases in this system are autonomous and they have full control over the local data they manage. The data access across a distributed system in federated database systems is done through five levels. The lower level consists of autonomous source schemas which are translated into an accepted data model in the component level. The relevant database parts for federation are shown at the export level as export schemas. These export schemas are completely integrated at the federation level and the appropriate solution is given out to the user in the external level. Interoperability among heterogeneous databases is provided in this approach by constructing global schemas from the underlying source schemas [2].
The above figure shows how data is accessed and controlled in a federated database system. After connecting to the federation, the user may submit his queries/requests to access federated objects. Every user request is split into a set of requests since the requests may require accessing the underlying local systems. Then, an access request is sent by the federation to all the groups to which the user belongs along with the user's identity. The user identifies and authenticates himself at the local system in case of a local authentication. Each local system then decides whether to grant permission for access request to be allowed or not and returns the result to the federation. The federation then makes the decision on data access depending upon the number of replies from the different systems [3].

### 3. Mediation Database Architecture

The other approach to access multiple heterogeneous databases is using mediators. Mediators are the components which mediate between the client and the sources for the data access. Mediators use attributes of clients for eligibility of data unlike Federation which is Identity based.
In mediation architecture, the Data access can be divided into two phases, Request phase and Delivery phase [4]. In request phase, a client sends a global query to a mediator.

**Request Phase**

A client sends a global query to mediator, the mediator decomposes the query into set of sub queries and then it sends the sub query to the appropriate sources.

**Delivery Phase**

In the delivery phase a source evaluates its subquery, produces a sub answer and sends back to mediator. The mediator integrates all the sub answers into global answer using wrappers\(^1\) and sends it back to client.

**Secure Mediation**

In mediation, though there is no one-one connection between the client and sources, they reveal their data schema to the mediators. This means that mediators can misuse them. Therefore both the parties rely on ‘trust’ in the mediators. The data accessibility should not rely on ‘trust’. Thus the mediators should be very secure so that the privacy and confidentiality is not compromised. Secure mediation is possible by using the concept of credentials and encryption. This concept in brief is described here.

The client submits the personal authorization attributes which are encoded in credentials to mediators. The mediator then decomposes the query in to subquery and forwards it to the data sources. The data sources based on the credentials return the sub query answers in the encrypted form using client’s public key. Then the mediator integrates the sub answers to form a global answer which is still in an encrypted form and returns to the client [4].

\(^1\) “Wrappers can be used to present a simplified interface, to encapsulate different sources to present a common interface, to add functionality to the data source, or to expose some of the data source's internal interfaces “[5].
4. Federation Vs Mediation

- Since the authorization is based on attributes rather than identities, the anonymity of the participating entities is preserved [4].
- Unlike federation, the participants in Mediation need not be closed.
- Mediation uses top-down and hence any changes to local schema doesn’t affect the data integration [4].
- In federation, the Metadata is revealed to the participating entities [4]. There might be situations in which the source doesn’t want its schema to be viewed by the client. Thus anonymity and confidentiality is preserved in Mediation.
• The mediators integrate only those parts of information sources that are crucial to users query via wrappers [4].

• Performance degradation occurs in federated database systems due to complexity in processing as every global query must propagate through the five levels of architecture [2].

• In federation, mutual trust between the participants is high which is not good for confidentiality of data [4].

Our Position:

In our conclusion, we state that though both Federated database management systems and mediated systems preserve autonomy among multiple heterogeneous databases, there is a drastic need for maximum privacy and confidentiality of the data. This is achieved by the Mediator approach combined with concepts such as encryption and credentials, which evolved to be a clear winner as it deals with data access and security in a more austere way.

References


