Practical Issues with Commercial use of Federated Databases

Jim Kleewein

Database Strategy, Architecture, and Technology, IBM Santa Teresa Lab, San Jose, CA 95141
kleewein@vnet.ibm.com

Abstract

Virtually all large enterprises have data stored in many databases. Often these databases are heterogeneous in terms of vendor, version, architecture, and capability. Federated databases, such as IBM's DataJoiner, are being used to provide rapid, cost effective access to this data across a variety of business types.

1 Introduction

This presentation discusses the who, why, and how of federated database exploitation with a focus on problems encountered and impediments to deployment.

2 Who

Users of federation range across industries from manufacturing, to finance, to retail, and more. The databases in their federations also vary from relational (such as DB2, Oracle, and Informix), to arguably relational (such as IDMS and Datacom), to decidedly non-relational (such as flat files, and OO-DBMS).

3 Why

Applications want their business data to appear as a coherent whole rather than as 'islands of data', They want to access that data with high functionality, in a uniform way that is supported by existing development tools, environments, and people. If that data is in a 'legacy' database they want that same set of high level functions, and they don't want to touch existing users of that data. Federation is used to provide that access and preserve the investment in 'legacy' systems.

4 How

Uses of federation can be divided into intelligent gateways, replication, and on-line access.

Intelligent gateway users want a uniform query language and a high level of abstraction. They want to learn one query language and use that to access all their data via the federation.

Federation provides replication support by enabling transparent access to the replication source or target, thus making it easier to create replication tools. It also enables new replication capability by enabling replication to or from a multiple database view.

On-line access provides applications and queries with access to data stored in multiple databases without requiring knowledge of location, multiple, query dialects, and multiple APIs.

5 Impediments

The impediments to federation fall into several areas: (1) non-technical, (2) Usability, (3) Performance, and (4) Data source support.

Some non-technical issues that impede federation are skepticism about the performance of the system or its capabilities.

Schema integration is one aspect of usability that impedes federation. There are often thousands of tables or views involved in a federation making maintenance of a global schema difficult. Another aspect in installation and configuration; establishing a federation requires lots of setup and expertise.

Performance, performance monitoring, turning, and problem determinate are areas of concern. Federations, and their associated communication systems, require considerable tuning.

The databases themselves are problematic. Differences in query dialects, functions supported, and data types are some factors that complicate federation. Other include support for multiple versions of databases, different behaviors based on configuration parameters, and concurrency control schemes.

6 Summary

Federation is where relational was ten years ago; the problems are increasingly well defined, the market is ready, some technology is ready, and commercial products are being gradually accepted. There continues to be, however, need for more research into the problems that prevent wide scale deployment of the technology.