Integrating Information for On Demand Computing

Nelson Mendonça. Mattos

IBM Silicon Valley Laboratory
555 Bailey Avenue
San Jose - California, 95124
mattos@us.ibm.com

Abstract

Information integration provides a competitive advantage to businesses and is fundamental to on demand computing. It is strategic area of investment by software companies today whose goal is to provide a unified view of the data regardless of differences in data format, data location and access interfaces, dynamically manage data placement to match availability, currency and performance requirements, and provide autonomic features that reduce the burden on IT staffs for managing complex data architectures. This paper describes the motivation for integrating information for on demand computing, explains its requirements, and illustrates its value through usage scenarios. As shown in the paper, there is still a tremendous amount of research, engineering, and development work needed to make the full information integration vision a reality and it is expected that software companies will continue to heavily invest in aggressively pursing the information integration vision.

1. Motivation

An on demand business is an enterprise whose business processes are **integrated end-to-end** across the company and with key partners, suppliers and customers so it can respond with speed to any customer demand, market opportunity or external threat. To achieve that vision, enterprises need an IT infrastructure that enables their business to be flexible and responsive. Such an IT infrastructure is called the *on demand operating environment.* An on demand operating environment needs to be

- Integrated end-to-end across people, processes, and information, creating business flexibility to optimize operations across and beyond the enterprise.
- Automated by reducing the complexity and cost of IT management and improving availability and resiliency.
- Virtualized to give a single, consolidated view of and easy access to all available resources in a network to improve working capital and asset utilization.

Information integration technology plays a crucial role in creating such an on demand operating environment. It enables integrated, real-time access to traditional and emerging data sources, transforms information to meet the needs of the business analytics and manages data placement for performance, currency, and availability. It, therefore, enables the creation of an environment in which information is **integrated** across and beyond the enterprise and is **virtualized** to provide transparent read and write access to diverse and distributed information resources across the network (or on the Grid).

Providing such integration and virtualization of information is tremendously ambitious. However, the returns make pursuing it very worthwhile.

- Dataquest estimates that business integration represents a $10 Billion opportunity by 2006.
- Aberdeen estimates that the information integration sector will fuel a $7.5 Billion market in 2003.
- Meta group suggests an untapped opportunity of $3 Billion.

And it shouldn't be a surprise that the opportunity is so large. Customers are plagued with accessing and integrating disparate data. In fact, very few organizations find that all of the information they need is readily available. Information is typically scattered throughout the enterprise in a wide variety of locations, data stores, and representations. Once you have access to the information, using combinations of data in meaningful ways is itself a challenge. The results may themselves need to be analyzed and/or transformed into the final representation and delivered to the desired location.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment

Proceedings of the 29th VLDB Conference,
Berlin, Germany, 2003

2. Integration Market Shifts

Information integration has now reached the minds of key decision makers and has become the top IT spending priority for CIOs jointly with other integration technologies [Ci02, MS01]. This has driven software companies from different areas to extend their value propositions in an attempt to capture their "share of the pie": application vendors rapidly moving to integration infrastructure by leveraging their installed base; database providers fiercely extending their current portfolio with XML, data federation, and Web services capabilities; ETL, portal, content management vendors, and messaging as well as application server vendors all aggressively enhancing their still weak data management infrastructure. There is significant overlap between very different products all of which are trying to address a broad range of customer requirements.

However, it is just a matter of time (not a lot of time) until market boundaries become clearer based on true technology innovation instead of revenue-driven intent. The prediction is that the information integration market will quickly go through a significant transformation. In fact, not only the information integration market, but the whole business integration market will see dramatic consolidation and redefinition around 3 main platforms -- exactly those that will lead business transformation to the on demand era:

- Portal platforms that will continue to improve its standardized access to applications and information sources, expanding its capabilities to provide collaboration technologies, improved personalization, and dynamic and adaptive role-based workplaces.
- Process-oriented platforms that will remain the most effective way to integrate applications, through workflow and messaging technologies, improving connectivity across enterprises and means to monitor processes and business activities.
- Information integration platforms that have made federated databases a reality and are now combining it with data placement capabilities and XML in order to effectively deal with both structured data and content.

From a data management perspective, the focus and investment in information integration is likely to cause a major shift in that industry. Database systems as we know of today may lose their important role as a control point for software vendors. Because information integration platforms can integrate data in any database, such platforms will become the new control point as applications will become dependent of their interfaces and totally independent from the database that physically store the data. The shift in control point will in turn fuel an even larger technology investment into information integration platforms, investments that have traditionally gone into database systems of today.

3. Requirements on an Information Integration Infrastructure

But, what is really information integration? It is referred to a category of middleware which lets applications access data as though it were in a single database, whether or not it is, enabling the integration of structured and unstructured data by

- providing real-time read and write access,
- transforming data for business analysis and data interchange, managing data placement for performance, currency, and availability.There are two

primary approaches for integrating information:

- Consolidating data for local access
- Accessing data in place via federation technology

Consolidating data into a single physical store has traditionally been the most effective approach to provide fast, highly available, and integrated access to related information. Whether coalescing all the required data for a new e-business application for online transactions or enabling sophisticated data mining of warehoused historical data, creating a single physical copy enables businesses to meet access performance or availability requirements, deliver point-in-time consistent snapshots, and provide for sophisticated transformation to achieve semantic consistency. Typically managed by Extract, Transform, and Load (ETL) or replication processes, they still represent the workhorses of information integration today.

However, there are also drawbacks of consolidated stores. They introduce significant added expense due to server and storage costs as well as administration costs, there is a latency between the copy and the source of record which can be a problem for real-time access requirements, and rich content or mixed format data is typically not included. Furthermore, due to organizational or operational constraints, more and more data sources do not generally lend themselves to being fully replicated or consolidated within a single database. Hence, there is an increased demand for federated access to distributed sources.

Federation is middleware technology that lets applications access diverse and distributed data as if it were a single source, regardless of location, format, or access language. As expected, access performance will be slower than for consolidated stores, but at the same time federation can meet needs not well addressed by a consolidated approach:

- Reduced implementation and maintenance costs
- Access to fully current data from the source of record
- Ability to easily combine traditional data with mixed format data, e.g. customer operational data stores (ODS) with related contract documents or images
- Access to data whose copy is prohibited based on data security, licensing restrictions, or industry regulations that restrict data movement

Industry analysts agree that neither federation nor data consolidation alone have answers to all problems. However, when combined in effective ways, they provide significant value to a set of problems, some of which could not be addressed before. By combining these two technologies customers will have immediate read/write access to distributed data as if it were stored in a single place and the flexibility to consolidate the data overtime to achieve higher levels of performance and availability without impacting any applications. This is the true value of Information integration.

For this reason, a complete information integration infrastructure needs to provide both **data federation** and the means to manage **data placement** in order to allow transparent access to data anywhere and data consolidation whenever required. Additionally, such infrastructure needs to provide rich services for **transforming and analysing** the integrated information, for **metadata management**, and for enabling information integration systems to interact with other middleware systems (e.g. messaging and workflow systems and web services).

2.1 Data Federation

Data federation lets applications access information as if it were physically stored in a local, single database, regardless of the form and location. Therefore, federation lets applications view and manipulate a collection of data sources as if they were a single source, while retaining their autonomy and integrity. The resources may be uniform or diverse, collocated or distributed, depending on the implementation. To achieve that, it needs to provide [MKTZ99]:

- *Transparency*, which helps mask from the user the differences, idiosyncrasies and implementations of the underlying data sources, to make the set of federated sources appear like a single system.
- *Heterogeneity*, which implies the ability to federate highly diverse types of data, including structured data (e.g., relational databases), semi-structured data (e.g. XML documents) and unstructured data (e.g. free-form text).
- *Extensibility*, such that the federation can be extended to almost any data source. Specifically, extensibility has to be designed to minimize the effort required to integrate a new source, yet offer the flexibility to provide the necessary information to optimize query access.
- *Rich functionality*, which includes the functions available through the supported query languages, compensation for missing functions in the backend data sources, plus the ability to surface source-specific capabilities into the query language seamlessly.
- *Autonomy* for data sources such that data sources can be federated with little or no impact on existing applications or systems.
- *Performance* to make federated query a real-world option.

Access needs to be possible through a heterogeneous set of programming models, application programming interfaces, and query languages:

- *SQL* as the standard for querying structured relational data, including support for XML through SQL extensions as defined by the *SQL/XML* standard [ISO03b].
- *Native XML query* support which is currently being standardized as XML Query by W3C [CCF+02].
- *Content-specific programming models*, APIs, and query languages to better support applications working with unstructured data (content). This includes support for free-text search interfaces such as those provided by engines like Google or Altavista [SCK02].

Client access needs to be supported through various interfaces: standard call level interfaces such as SQL/CLI, ODBC [ODBC] or JDBC [JDBC], *web services, and asynchronous client* capabilities based on messaging.

Access to foreign data sources is supported through *federated* query capabilities, which employ wrapper technology [ISO01] to pull data from foreign sources into the integration engine in an optimized manner [TS97, MKTZ99, HLR02, MM+01, MM+02, Wi93] or by a flexible *crawler* available to get to the unstructured data to be indexed, regardless whether this data is stored in the information integration system, the web or an intranet.

Regardless of the client access and query language used, the application should be able to access all of the data connected through the information integration server. This combination of client access and query language flexibility makes it possible for existing development and analytical tools to take immediate advantage of the broader data access and integration features provided by the integration server. It also allows the infrastructure to plug into service-oriented architectures using Web services, to provide asynchronous clients for easy integration with workflows or scheduling long-running queries, and to extend customer investment in current and new application infrastructures.

2.2 Data Placement

An information integration infrastructure must support placing and managing data at multiple points in the data hierarchy to improve performance and availability. Any changes in the data location are however hidden from the application by the federated capabilities described in the previous section.

Beyond simple caching, this is policy-based data placement and management [LKM+02], a range of caching strategies is required to achieve adequate performance while respecting the characteristics of the requesting applications vis-à-vis currency, semantic correctness, and availability.

The cache should be able to store data in at least two formats: relational, which has recently been extended by complex object-relational structures and large objects [ISO99] and native XML for storing XML data and documents and, in particular, to efficiently support processing of XML Query [CCF+02, Ch02].

Heterogeneous replication (i.e., the ability to replicate data across different database systems) is required as a fundamental characteristic of an information integration infrastructure. It complements the distributed access features, enabling management of centralized data stores, and provides the necessary infrastructure for efficiently managing data caches. Replication should be complemented with ETL capabilities to improve the movement of large amounts of data and to provide complex transformations and data cleansing efficiently.

2.3 Data Transformation and other Information Integration Services

The infrastructure must provide rich transformation features to facilitate analysis, interchange or presentation. This includes offering advanced meta data management capabilities as well as a comprehensive set of services that enable information integration systems to interact with other middleware systems (e.g. messaging and workflow systems and web services).

Basic *transformations* that map data from one representation into a different one have multiple usages, such as view support (in combination with federated access) or ETL and replication. Other types of transformations can be used for realizing mappings between different data models. Through the definition of view and transformation capabilities, mappings between relational data and XML can be established which can then be exploited to, for example, query relational data using the XML Query language [SS+00, SK+01, FKSSW02].

As part of the transformation services, sophisticated *analysis* and *mining* technology, both for structured as well as semi-structured or unstructured data, should be provided as part of the query language. One example of this approach is defined in SQL MM, Part 6, Data Mining [ISO03], which describes the integration of data mining capabilities into the SQL engine from a language perspective and are already available in data base products [IBM01].

Moreover, tight integration with *workflow* and *messaging systems* needs to be provided. This support is twofold in the sense that (1) these services are available inside the information integration system to initiate and coordinate information integration tasks and operations (e.g., data exchange and transformation steps) in a robust and reliable manner, but also (2) to ensure that information integration tasks can participate in more general business processes and can interoperate with applications and system components using reliable messaging services [LR02].

Last, but by no means least, an information integration system must provide a robust infrastructure for meta-data management. The metadata will be used not only by tasks supported by the information integration system, but also by solutions across disparate capabilities and toolsets. This infrastructure will support various development phases, during which different types of meta-data are collected and created. The support should be provided through a set of tools that accompany the information integration management system.

There are three basic steps when using an information integration system: discover, design, and deploy [DMMW03]. Once a business problem has been defined, developers and system architects begin the integration work by first discovering what is available and how it relates to solving the problem. The design phase focuses on designing all of the elements of the solution that need to be crafted yielding a logical design that is then deployed onto one or more physical topologies. Information developed during this process should be maintained as meta-data to facilitate the implementation of future projects. In addition, these same meta-data services are highly important for upper layers of an overall business integration architecture. For example, a workflow modeler may need this meta data to better understand where to get which data, how to combine it, and how to exploit it during the integration of two business processes.

4. Usage Scenarios and Benefits of Information Integration

The usage scenarios for an information integration system are extensive given that many key initiatives currently being deployed, such as customer relationship management, supply chain management, consolidations following mergers or acquisitions, and business intelligence, are based on successfully integrating information from multiple data sources, both structured and unstructured. Let us review a few.

Adding value to existing data warehouses – Information integration can help businesses develop enterprise views of disparate data marts or warehouses, enrich reporting with relevant unstructured content, and combine historical warehouse views with real-time data. for improved decision making [De03].

Simplifying composite application development - A new generation of composite applications is emerging typified by the requirement to access and integrate distinct application and data domains. Portal infrastructures, customer data integration scenarios, and integrated supply chain applications are representative of these applications. Information integration opens up opportunities to incorporate more data, more easily. while reducing coding and skills requirements and speeding development of such applications. Experiments show that Information integration can reduce the hand-coding requirements by 40%-65%, reduce the skill requirements, and cut development time in half for projects that are integrating disparate data sources. This is achieved because federation takes on the burden of query decomposition (a source of accuracy issues for hand-coded tasks), query optimization for reasonable responsiveness, translation between incoming/outgoing SQL and native APIs, connection management as well as data type and platform mismatches (e.g., EBCDIC/Ascii conversions). The details of such an experiment can be found in [IBM04].

Facilitating migration and co-existence scenarios - Mergers and acquisitions present significant integration challenges. New corporate entities must grapple with understanding their customers and partners across multiple IT infrastructures while strategic plans are assessed and implemented. With the decentralization of computing in many large enterprises, an acquisition is not a prerequisite to disparate IT infrastructures. Information integration can help businesses gain a better understanding of their combined business and significantly facilitate migrations by providing transparent access across databases so that data migration becomes transparent to applications.

Speeding up drug discovery – To increase drug research efficiency and encourage interdisciplinary cooperation between chemists and biologists, scientific users require integrated view of chemical & biological information stored in distributed sources: relational and non-relational and internal as well as external. Information integration allows the creation of a single virtual database from the multiple data management platforms that house research data. Scientists can thus view the distributed data more effectively and cohesively and search it with a single query, allowing them to more quickly identify new drug candidates, develop more effective therapies and bring new biopharmaceutical products to market.

Technology does not provide benefits by itself. Its value is directly related to the customer pain points it addresses. Information integration addresses several pain points in businesses today and has, for this reason, become a major area of investment by software companies. Its benefits can be summarized as follows:

- It helps businesses extend their current IT investments. Using familiar programming model (e.g., SQL) and existing tools (e.g., reporting generation tools), enterprises can now combine data more easily and more quickly. This translates into more comprehensive and more timely analysis and reporting. In the life sciences industry, it translates into faster drug discovery. In the insurance industry, into faster customer service.
- It speeds time to value for composite application deployment. It simplifies complex integration tasks reducing coding requirements, skill requirements, and development time enabling businesses to more quickly deliver new function to customers, partners or employees.
- It helps senior IT control costs. It reduces payroll costs through its productivity enhancements, reduces the need to replace existing systems, and reduces the need to manage more redundant data.

5. Challenges

While Information integration is a new concept, many of its architectural components are not and its individual technologies have received a lot of attention from research and product development in the past. In particular, the field of federated databases [SL90, LMR90, MKTZ99], and related areas such as data integration [Ha03, Hä02], schema integration [Co02], and schema matching [RB01] have received a lot of attention over the last years. (These references are by no means exhaustive.)

However, there are a large number of remaining open issues still requiring additional research and the combined use of some of these technologies will challenge many software engineers worldwide.

The support of heterogeneous interfaces (query and search) has significant impacts to the federated engine and system architecture. While federated query over structured data usually involves optimized processing of possibly complex queries with joins or aggregations over tables residing in multiple data stores, federated search over unstructured data is usually a union-style search ranging over a more loosely coupled federation of search engines, comparable to a free-text search over web content as provided by search engines like Google or Altavista. In contrast to the precise SQL-type search, content-specific search for unstructured data is typically fuzzy and applications are often interested in “the best n” matching objects only [SCK02]. How will these two different paradigms be supported in a single information integration system?

Support of native XML, unstructured document, and in particular rich content in a consistent way also requires attention to arrive at a single, flexible data model and query language. Some standards efforts such as WebDAV [Web-DAV] are only beginning while others such as XQuery have not even started to seriously address the requirements of all forms of data (e.g., rich content). And even basic SQL attributes such as the notion of collections of XML documents or fragments, as well as support for updates has not been addressed in the XQuery specification yet, and research in this area is still fairly immature.

Research results in autonomic computing and self-tuning capabilities for relational databases have started to appear in commercial systems, but the applicability of those same research results to the heterogeneous environment of information integration is still an open question. Related to the heterogeneity of information integration are open questions about the transactional semantics of operations across federated data sources, synchronized backup and recovery, a uniform privacy and security model across a multitude of systems, as well as general query and operational performance aspects in the presence of huge data volumes and increasing numbers of data sources. Of particular interest are open questions related to the introduction of police-based data placement in an information integration system. These should be used not only to manage the contents of the cache, but also help determine when and how data from different back end systems should be moved from one to another to achieve higher levels of performance and availability given application workload and semantics.

6. Summary

Information integration represents a key component of an on demand operating environment – the IT infrastructure needed to integrate businesses end-to-end so they can respond with speed to any customer demand, market opportunity or external threat. It is strategic vision that when completely provided will serve up a unified view of the data regardless of differences in data format, data location and access interfaces, dynamically manage data placement to match availability, currency and performance requirements, and provide autonomic features that reduce the burden on IT staffs for managing complex data architectures.

The ideas presented in this paper are partially reflected in IBM's DB2 Information Integrator [IBM03], which addresses some of the challenges discussed in this paper and can provide the initial building blocks for a complete information integration system along the lines discussed above. However, there is still a tremendous amount of research, engineering, and development work to make the information integration vision a reality.

Given the significance of the integration challenges faced by customers across industries today, it is expected that software companies will continue to heavily invest in aggressively pursing the information integration vision. Businesses today need to integrate information to drive customer loyalty and satisfaction, improve operational efficiency, compete for online customers and trading partners, and identify and respond to emerging opportunities.

In short, information integration provides a competitive advantage and is fundamental to on demand computing.

7. References

[CCF+02]D. Chamberlin, J. Clark, D. Florescu, J. Robie, J. Simeon, and M. Stefanescu, XQuery 1.0: An XML Query Language, W3C Working Draft (April 2002). Available at http://www.w3.org/TR/xquery/.

[Ch02]D. Chamberlin, "XQuery: An XML Query Language," [IBM02], 597-615 (2002).

[Ci02] CIO Magazine, March 2002.

[Co02]S. Conrad: Schemaintegration: Integrationskonflikte, Lsungsanstze, aktuelle Herausforderungen, [Hä02], pp. 101-111, 2002.

[De03] Devlin, b.: 'Solving the Data Warehouse Puzzle', DB2 Magazine vol.. 8, no. 2, Quarter 2, 2003.

[DMMW03]: Dessloch, S., Maier, A., Mattos, N., Wolfson, D.: Information Integration – Goals and Challenges, Datenbank Spektrum, Vol. 6, No. 6, June 2003.

[FKSSW02]J. E. Funderburk, G. Kiernan, J. Shanmugasundaram, E. Shekita, and C. Wei, "XTABLES: Bridging Relational Technology and XML," [IBM02], 616-641 (2002).

[Ha03]A.Y. Halevy: Data Integration: A Status Report, Proc. BTW 2003, pp. 24-29, 2003.

[Hä02]T. Härder (ed.), Informatik Forschung und Entwicklung, Themenheft Datenintegration, vol. 17, no.3, Sept. 2002.

[HH02]T. Härder, K. Hergula, "Ankopplung heterogener Anwendungssysteme an Fderierte Datenbanksysteme durch Funktionsintegrati-on", [Hä02], 135-148, 2002.

[HLR02]L. M. Haas, E. T. Lin, and M. A. Roth, "Data Integration Through Database Federation," [IBM02], 578-596 (2002).

[IBM01] Intelligent Miner Scoring: Administration and Programming for DB2. International Business Machines Corp, 2001. http://pu-blib.boulder.ibm.com/epubs/pdf/idmr0a00.pdf

[IBM02]IBM Systems Journal 41, no.4, special issue on Information Integration, 2002.

[IBM03]IBM DB2 Information Integrator, see: http://www.ibm.com/software/data/integrati-on/iipreview.html

[IBM04] J2EE Development with DB2 Information Integrator: Digging into the Details, see http://www.ibm.com/developerworks/db2.

[ISO99] ANSI/ISO/IEC 9075-2 Information Technology - Database Languages - SQL –Part 2: Foundation (SQL/Foundation), 1999

[ISO01]ISO/IEC 9075-9:2001, Information technology - Database language - SQL - Part 9: Management of External Data (SQL/MED), International Organization for Standardization, June 2001

[ISO03]SQL/MM Part 6: Data Mining.

[ISO03b] XML-Related Specifications (SQL/XML) - Working Draft SQL:200n Part 14, H2-2001-149, WG3:YYJ-012, Jim Melton (Editor)

[JDBC]See http://java.sun.com/products/jdbc/.

[JMP02]A. D. Jhingran, N. Mattos, H. Pirahesh, "Information Integration: A research agenda", [IBM02], 555-562, 2002.

[LKM+02]Q. Luo, S. Krishnamurthy, C. Mohan, H. Pirahesh, H. Woo, B. Lindsay, and J. Naughton, "Middle-Tier Database Caching for e-Business," Proceedings,ACM SIGMOD International Conference on Management of Data, Madison, WI (June 3-6, 2002).

[LMR90]W. Litwin, L. Mark, N. Roussopoulos: Interoperability of Multiple Autonomous Da-tabases, ACM Computing Surveys 22 (3), pp. 267-293, 1990.
[LR02]F. Leymann and D. Roller, "Using Flows in Information Integration," [IBM02], 732-742 (2002).
[MKTZ99] Nelson Mendona Mattos, Jim Kleewein, Mary Tork Roth, Kathy Zeidenstein: From Object-Relational to Federated Databases. BTW 1999: 185-209
[MM+01] Jim Melton, Jan-Eike Michels, Vanja Josifovski, Krishna G. Kulkarni, Peter M. Schwarz, Kathy Zeidenstein: SQL and Management of External Data. SIGMOD Record 30(1): 70-77 (2001)
[MM+02] Jim Melton, Jan-Eike Michels, Vanja Josifovski, Krishna G. Kulkarni, Peter M. Schwarz: SQL MED - a Status report.. SIGMOD Record 31(3), (2001)
[MS01] Morgan Stanley CIO Survey, May 2001.
[ODBC]See http://www.microsoft.com/data/odbc/default.htm.
[RB01]E. Rahm, P. Bernstein: A Survey of Approaches to Automatic Schema Matching, VLDB Journal 10 (4), pp. 334-350, 2001.
[SCK02]A. Somani, D. Choy, and J. C. Kleewein, "Bringing Together Content and Data Management Systems: Challenges and Opportunities," [IBM02], 686-696 (2002).
[SL90]A.P. Sheth, J.A. Larson: Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, ACM Computing Surveys 22 (3), pp. 183-236, 1990.
[SS+00]J.Shanmugasundaram, E.J. Shekita, R. Barr, M.J. Carey, B.G. Lindsay, H. Pirahesh, B. Reinwald: Efficiently Publishing Relational Data as XML Documents. VLDB 2000: 65-76
[SK+01]J. Shanmugasundaram, J. Kiernan, E.J. Shekita, C. Fan, J. Funderburk: Querying XML Views of Relational Data. VLDB 2001: 261-270
[TS97]M. Tork Roth and P. Schwarz, "Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources," Proceedings, 23rd Conference on Very Large Data Bases, Athens, Greece (August 26-29, 1997).
[WebDAV]See http://www.webdav.org
[Wi93]G. Wiederhold: Intelligent Integration of Information. SIGMOD Conference 1993: 434-437.