This document describes several vendors that offer innovative approaches and technology for searching structured data assets, integrating structured and unstructured data, improving the scalability of data warehouses, and facilitating data governance activities. Each of these areas will grow in importance as organizations work toward managing information as a strategic corporate asset.

**Key Findings**
- Search technology will become increasingly important for structured data management.
- Data integration tools are evolving toward the unification of structured and unstructured data, and will begin to include semantic capabilities.
- Demand is increasing for appliance-oriented approaches to improve performance of data warehouses.
- Data governance is becoming “top of mind” for many organizations, which is creating demand for governance-oriented tools and technologies.

**Recommendations**
- Investigate data integration technologies that can unify both structured and unstructured data, and resolve semantic discrepancies across data silos – semantically-based approaches to integration and modeling hold significant promise for addressing a range of information management issues.
- Consider search-oriented technologies to be a pervasive component of your information infrastructure – organizations generate massive volumes of structured transactional data, and simplified and efficient access to specific records will be key to meeting regulatory reporting and e-discovery requirements.
- Explore appliance-oriented solutions for addressing scalability challenges in certain types of data warehouse – data warehouse environments that are constrained in terms of capacity might benefit from distributing portions of the workload to appliance solutions that promise lower cost and greater efficiency.
- Adopt a holistic viewpoint on information management and the development of a comprehensive information infrastructure – organizations must leverage and govern the full breadth of the content continuum, not only their structured data assets.
ANALYSIS
What You Need to Know

Advancements in technologies relating to managing and integrating data will help organizations to begin fulfilling the promise of a cohesive information infrastructure, spanning the full range of the content continuum (see “Cool Vendors in Content Management, 2007”). Gartner’s Cool Vendor selections for 2007 showcase innovation in structured data search, data integration, semantics, high-performance query processing and data governance. We advise clients to explore these concepts and the available technologies (including, but not limited to, the vendors and products described here), as they reflect important trends in data management and integration.


Why Cool: CopperEye seeks simply to make reports and single-record location faster and cheaper than could be performed using a relational database management system (RDBMS). Other tasks at which RDBMSs excel (such as transaction processing) are left aside in favor of swiftly locating and examining information for enterprises that need particularly effective search for data that resides in columns and rows. This task is facilitated through a layer where the data exists in an index created via an Open Database Connectivity (ODBC) connector.

Dealing with data in RDBMSs may require enterprises to spend substantially on hardware and software, as well as on database administrators. CopperEye’s index may be queried with less demand for such resources. For ad hoc queries that seek specific records or light reports, it may be enough to turn to a vendor that specializes in speed and simplicity, instead of the full menu of functions allowed by RDBMS powers. While CopperEye does not make it possible to ignore RDBMS software, it can reduce the load on such software and otherwise limit enterprise resources devoted to it.

Challenges: “Search” increasingly implies the free text capability inherent in the Google model, where users relish the privilege to treat numeric value and data structure as irrelevant. In fact, enterprise search vendors seek to replace the RDBMS from the perspective of providing greater flexibility and tolerance for ambiguity in index value and query design. CopperEye is an unconventional way of solving a conventional and familiar information location problem; search represents an unconventional framing of the problem and the solution. CopperEye therefore faces the challenge that it lacks the undisciplinary interface common in search, which enterprises find increasingly appealing.

Who Should Care: Information architects might recommend CopperEye as a way of accessing data more effectively, without the delays implied in the massive scale demanded by database queries. CIOs might seek to use CopperEye to reduce dependency on RDBMS vendors. Customer service managers might suggest CopperEye serves to allow the enterprise to make data look-up services available to its customer base.

Metatomix, Waltham, Massachusetts, U.S. (www.metatomix.com) Analysis by Ted Friedman

Why Cool: Metatomix provides technology for creating “semantic services” – runtime components which retrieve data from a range of data sources (databases, content repositories, Web services and more) and apply a common model to determine how the data relate to each other. The goal is to break down the “silos” that exist across a typical large organization – when departments and systems use different terminology and apply different meanings to the same data. Semantically based technologies represent an emerging and increasingly important area of innovation, particularly as they relate to data integration. Most organizations invest significant effort in dealing with semantic discrepancies across databases and applications. These discrepancies are either dealt with manually or by building complex data integration architectures, based on custom coding or tools, which deal only with the structure (format) of the data and must be augmented with complex business rules to align semantics.

Emerging semantic technologies promise to elevate data integration work to a higher level by automating the discovery and definition of relationships, and by driving out the rules necessary to translate data from one context to another. Metatomix’s core technology is, in effect, a workbench for automating the development of services that reconcile semantic discrepancies across related data. The vendor currently goes to market with a set of packaged applications, built on its semantic services technology and aimed at the legal/judicial segment of the market.

Challenges: Semantic technologies are early in their maturity and market adoption. Many Type B and C organizations (see Note 1) will struggle to understand semantic approaches and will view such technology as “bleeding edge,” avoiding it because they are risk averse. While virtually all organizations face challenges due to a lack of semantic consistency in their data, Metatomix will need to be selective in how it targets the market, continuing to seek out verticals and prospects that are ready to adopt this type of approach. It has begun to do this already with its focus on the legal/judicial vertical segment. Major vendors are already increasing investments in semantic technologies, which will increase the pressure on Metatomix to differentiate itself clearly and to “out innovate” larger competitors to survive.

© 2007 Gartner, Inc. and/or its Affiliates. All Rights Reserved. Reproduction and distribution of this publication in any form without prior written permission is forbidden. The information contained herein has been obtained from sources believed to be reliable. Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner’s research may discuss legal issues related to the information technology business, Gartner does not provide legal advice or services and its research should not be construed or used as such. Gartner shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The opinions expressed herein are subject to change without notice.
Who Should Care: Organizations seeking to solve data integration problems that extend across departmental silos and across the breadth of the content continuum must wrestle with complex semantic issues. Information architects, services architects, integration architects, and integration developers involved in data integration activities should begin to understand the capabilities and state of semantic technologies as they relate to integration work. Over time, as these technologies mature, they may alleviate some of the complexity and reduce the level of effort in comprehensive data integration initiatives.

Varonis, Saddle Brook, New Jersey, U.S.
(www.varonis.com)
Analysis by Ted Friedman

Why Cool: Data governance issues are increasing in importance for all organizations, driven by compliance and transparency pressures, and the general need to regain control over their most critical corporate asset – information. Most organizations struggle with seeing how all data across the organization is being accessed and used, as well as with auditing the integrity of that data. Varonis provides a suite of tools aimed at supporting organizations in their data governance efforts. The Varonis Intelligent Data Usage (IDU) technology gathers information about directory/file structures and their associated user permissions, and is able to transparently capture data access events (for example, reads, updates and deletes) against those structures. Varonis builds a model (stored in a relational database) of data access patterns based on this information. By applying sophisticated statistical and pattern-recognition approaches to this model, the Varonis technology computes groupings of users (what Varonis calls a “business group”) that behave in a relatively consistent manner with regard to data access and usage. Through this approach, Varonis, in effect, identifies which groups have ownership for which subsets of the data. Users exhibiting infrequent or inconsistent access behavior, compared with that of members of a business group, are identified as potential policy violators. Varonis, therefore, recommends that access privileges be revoked for those users. The vendor claims that its self-learning algorithm achieves very high levels of accuracy when exposed to data access events for a training period of two to four weeks in most environments.

Challenges: Data governance involves much more than control of access privileges, data access reporting and the identification of data owners, and it is an initiative that cannot be addressed by tools and technology alone. Processes to ensure adherence to standards in data model design and model management, policies for assurance of data quality levels and introduction of data quality controls, and governance-specific roles and responsibilities are also part of a holistic data governance framework.

Varonis will need to build expertise in these other areas and should partner with service providers to present a truly complete solution for data governance. In addition, the Varonis technology currently works only with data residing in standard operation system file structures, as opposed to data residing in relational (or other) DBMSs. Organizations certainly face challenges in governing the massive amounts of semi-structured data they hold but, to provide a comprehensive governance offering, Varonis will need to address DBMS-based data as well.

Who Should Care: As the pressure increases to control and safeguard information assets, business leaders, CIOs and information architects will be forced to widen their focus on data governance. People in these roles, as well as database managers, data stewards, security teams and compliance officers should seek technologies which could, potentially, simplify the challenge of establishing data usage controls, and of monitoring data access across the complex and diverse environments they manage.

Denodo Technologies, Palo Alto, California, U.S.
(www.denodo.com)
Analysis by David Newman

Why Cool: Denodo seeks to provide data “mashups” that combine content from more than one source into an integrated experience. While the term mashup is popular with the Internet crowd, it is not a preferred data integration method among database developers. In the database world, a mashup is similar to a virtual query or view of data that provides a temporary snapshot of information for a specific purpose or time span. Tools that perform this function have been marketed under the term enterprise information integration (EII). However, these products lack the semantic reconciliation, complex transformation and data quality functionality required in data warehousing, business intelligence or other data integration efforts.

What is different about the Denodo approach (compared with EII) is its ability to merge data across any digital source, whether it be structured (databases, applications and XML), unstructured (documents, static HTML and e-mail) or semi-structured (Web content). The Denodo platform goes one step beyond typical EII products, because it not only accesses diverse data, it lets you combine structured and unstructured content. It then indexes the content and pushes it to where it’s needed to support single customer views, data consolidation, competitive or business intelligence activities, business-to-business (B2B) or business-to-consumer (B2C) automations, and some of the basic data services required for service-oriented architectures (SOAs).
Challenges: Significant changes are taking place in the data integration market, as evidenced by consolidation, convergence and partnering among major extraction, transformation and loading (ETL), enterprise application integration (EAI) and EII players. Vendors are offering one-stop shopping to meet the growing demands for data integration to support service-orientation, master data management, information governance and data transparency. Denodo needs to do more than mash up and present data – other styles of data integration are required. With clients mostly in Europe, Denodo will be challenged to carve out a niche in a market dominated by larger players. Given that its current customers are carrying out departmental projects in industries such as telecom, financial services, e-commerce and government, Denodo may wish to maintain its competitive pricing model. However, Denodo will be forced to add more sophisticated capabilities in data transformation, semantic reconciliation and data services for SOA to compete in the growing data integration market.

Who Should Care: Organizations that maintain complex information architectures – in which important information about customers, products or markets is dispersed across multiple locations and channels (including Web content) – and that need a variety of data integration methods to sense, share and exchange data anywhere and on any device. Information architects, database specialists and composite applications developers, in support of delivering “information as a service,” that have to understand how to apply the range of integration technologies available to solve their emerging information infrastructure needs.

ParAccel, San Diego, California, U.S. (www.paraccel.com )
Analysis by Donald Feinberg

Why Cool: Although ParAccel offers two types of acceleration appliance – stand-alone SQL DBMS and SQL Server DBMS – we will concentrate on the SQL Server Edition, a software appliance to accelerate query processing in a Microsoft SQL Server environment. It uses a massively parallel processing (MPP) engine, based on the PostgreSQL open-source software (OSS) DBMS, and runs on any 64-bit (Intel or AMD) hardware platform and Red Hat Linux. It is a vertical-oriented DBMS – that is, a column-oriented DBMS, where all columns are indexed and the rows do not exist, yielding high levels of compression similar to DBMSs such as Sand, Sybase IQ and Vertica – which can achieve very good performance for analytical queries.

ParAccel SQL Server Edition offers an almost resource-free installation once the appliance is connected to the network. It will automatically create a duplicate schema and can be loaded simply with SQL Server Integration Services (SSIS). It is kept in “sync” with the master database through triggers and audit table entries, again with few or no resources used to set this up. ParAccel’s Q-Router, sitting in front of the master database, traps all queries to the master database, executing only those complex queries it can accelerate. It passes the other, simpler queries through to SQL Server. According to ParAccel, it is an almost “drop in” installation that will accelerate complex, analytical queries with little performance degradation of the primary database.

Challenges: In addition to the normal startup problems that affect any company, ParAccel’s biggest challenge will be to find customers that have large SQL Server databases or smaller, but more complex, databases. We believe this number of databases will grow rapidly over the next year or two, but it may slow adoption of ParAccel’s SQL Server edition. Another challenge will be to convince a Microsoft Windows “shop” to use Red Hat Linux alongside Windows and SQL Server. ParAccel does have plans to use Windows, but initially only Red Hat Linux. Finally, there will be some degree of performance degradation in the SQL Server DBMS, because of the use of triggers for synchronization. The sooner ParAccel has a Windows version with “log scrapping” and native SQL Server replication, the sooner these latter two issues will diminish.

Who Should Care: Microsoft SQL Server customers with large databases or complex workloads desiring greater performance with analytical queries and a simple almost drop-in installation.

Disclaimer
This research does not constitute an exhaustive list of vendors in any given technology area, but rather is designed to highlight interesting, new and innovative vendors, products and services. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Note 1
Type B and C Organizations

Type B organizations adopt technologies once they have been proven useful. Selections are made as a result of strategic planning and are based on the experiences of others. Type B organizations are mainstream adopters of technology that use IT to maintain parity or catch up with early adopters.

Type C organizations are slow adopters of technology. They use IT only when obvious or when required to meet mandated or market requirements. Type C organizations have strong financial motivation and are efficiency-focused. Technologies are only adopted when necessary and must be justified. Many opportunities are passed over to avoid risk or expense. As conservative adopters of IT and innovation, Type C organizations are characterized as cautious and late adopters.