

## Magic Quadrant for Data Warehouse Database Management Systems

Donald Feinberg, Mark A. Beyer

The data warehouse DBMS market is changing, with new entrants, new appliances and shifting demands. Here we outline vendors' capabilities to meet the new demands created by connected applications, exploding data volumes and the resulting mixed workload.

## WHAT YOU NEED TO KNOW

---

Today, most data warehouses are mission-critical (see Note 1), serving in an increasingly mixed workload capacity (see Note 2), including as a data source for online applications. "Deep mining" analysts and business analysts are running more ad hoc but equally complex queries and fast-running tactical queries, each with differing service-level expectations. These differing workloads are all competing for CPU, memory and disk access. At the same time, data latency continues to progress from batch to continuous loading demands.

In 2009, the latest wave of data warehouse adoption, which includes less mature organizations with little or no data warehouse management experience, continued to grow in size. Many of these organizations provide a new opportunity for mart-style deployments. At the same time, data warehouse appliances (see Note 3) are growing in popularity, with the large "mega vendors" (such as IBM, Oracle, HP and soon Microsoft) now offering an appliance solution of some type. End-user organizations should ignore marketing claims regarding the applicability and performance capabilities of solutions, and should instead base their decisions on customer references and proofs of concept (POCs) to ensure that claims made by vendors will hold true in a real-life environment — more specifically, their own environment.

Gartner clients increasingly report performance-constrained data warehouses during inquiries. Based on these inquiries, we estimate that nearly 70% of data warehouses experience performance-constrained issues of various types. These are data warehouses with high query counts, mixed query types, a high degree of connectivity and growing integration with both operational and business intelligence (BI) applications. Notably, most of these are data warehouses with multi-year track records of success in delivering to end users and supporting each of these performance areas. Importantly, constrained warehouses are difficult to define because frequently the enterprise has not established clear service-level expectations — making it impossible to actually determine if the warehouse is constrained relative to a service-level agreement (SLA).

The data warehouse database management system (DBMS) Magic Quadrant discusses the primary building blocks of a data warehouse infrastructure, and as such should be of interest to anyone involved in defining, purchasing, building and/or managing a data warehouse environment. This includes, but is not limited to, CIOs, chief technology officers (CTOs), the business intelligence competency center (BICC), infrastructure, database and data warehouse architects, database administrators (DBAs) and IT purchasing departments.

Organizations are advised to consider their own level of risk aversion when choosing their data warehouse DBMS, taking into account the following:

- In 2009, three separate strata of vendors emerged in the market, clustered primarily around strategies for addressing data warehouse needs. The strategies are based primarily on implementation styles. For discussion purposes, we refer to these three strata or clusters as: data warehouse veterans, recent contenders or experimentation. The top right of the quadrant shows data warehouse veterans IBM, Oracle and Teradata still battling for overall market leadership. A cluster of recent contenders are spread across the central portion of the data warehouse DBMS Magic Quadrant. This group of vendors is focusing on either vision or execution but not always with a good balance of both. The new entrants are trying to build their market presence with single-point or a limited number of innovations focused primarily on performance or delivery model niches.

- With most of the market's major vendors providing both software-only and appliance-based solutions (see Note 3), it is possible to select a solution offered by a vendor that is also an organization's preferred or standard infrastructure vendor. However, specific advantages in managing larger data volumes and mixed workloads should be considered, as should the increasing importance of managed support. Most of the leaders represent low-risk options, even if the selected platform is outside enterprise standards.
- Niche vendors are offering new technology in the market, such as data tokenization and the use of specialized compression and data storage techniques that reduce input/output (I/O) constraint issues for high performance. Many of these vendors include risk aspects such as small reference bases (and a commensurate lack of best practices) and vendor capitalization issues. Organizations that embrace a higher level of risk should seek compensation in the form of discounts, and even extended pilot implementations with below average prices for support during the pilot period.
- In 2009, low-price, entry-level offerings saw increased levels of interest (examples include Greenplum's Single-Node Edition, the Sun Oracle Database Machine basic system and Teradata's 551 platform). Some organizations are being very creative in using these "entry" solutions as dependent departmental platforms for data marts, as a way to extend their current platform standard. Demand for data mart support continues to be strong, and there has been a resurgence in independent data marts in 2009, a reversal on 2008. Mixed workloads continued to overwhelm enterprise warehouse performance and solution implementers responded with synchronized, dual-warehouse deployments.
- All the vendors continue to offer new products and innovations. In the case of the veterans, they have introduced low-cost entry solutions, as previously discussed. Niche vendors and the newest leaders are moving their positions on the quadrant toward a more complete vision or a greater ability to execute, expanding their existing delivery channels (Kognitio, for example, now mixes its hosted offering with a direct delivery channel), or introducing analytic capabilities (MapReduce and SAS embedded in the DBMS, as with Aster Data, Greenplum, Netezza, Teradata and others). However, over time market adoption becomes an indicator of the validity of these new offerings, and the market will determine if and how far these vendors progress toward leadership.

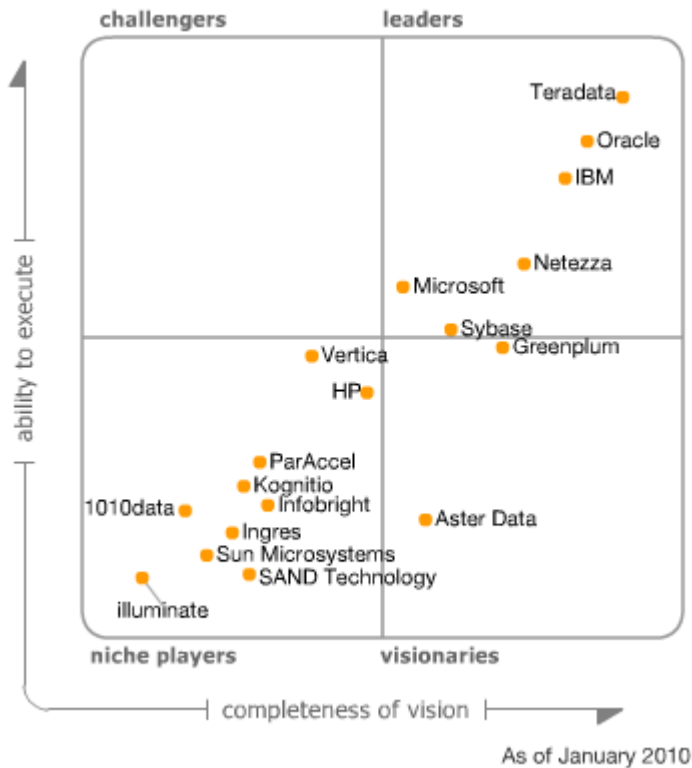
## **STRATEGIC PLANNING ASSUMPTION(S)**

---

Through 2012, mixed workload performance will remain the single most important performance issue in data warehousing.

## MAGIC QUADRANT

Figure 1. Magic Quadrant for Data Warehouse Database Management Systems



Source: Gartner (January 2010)

## Market Overview

Here we present a basic overview of the data warehouse DBMS market in 2009. After a basic description of the market, we discuss some of the key market influences and describe some of the overall vendor responses. The rest of this Magic Quadrant describes individual vendors' capabilities.

The data warehouse DBMS market has evolved from an information store supporting traditional BI platforms to a broader analytics infrastructure supporting operational analytics, corporate performance management and other new applications and uses (such as operational BI and performance management). Organizations are adding additional workloads with online transaction processing (OLTP) access, and data loading is moving to intra-day — approaching continuous loading. Gartner does not measure data warehouse DBMS market revenues separately (as most DBMS vendors cannot determine the percentage of their database revenue that comes from data warehouses). However, the overall DBMS software market continued to grow in 2009 at just over 10% annually (see "Market Share: Database Management System Software, Worldwide, 2008").

In 2009, cost control and performance optimization became the critical evaluation criteria (see "Critical Factors in Calculating the Data Warehouse Total Cost of Ownership"). These two criteria have pushed vendors to expand the offerings they have in their product portfolios, and are

contributing to the revenue increases in the data warehouse DBMS market. Some organizations have shown renewed interest in virtual data warehouses that use federation technology that defers cost or shifts it to another category (for example, the warehouse is really running on the operational systems which keep the history and "hide" the cost). Others maintain a large central warehouse, but deploy optimization structures like data marts on smaller platforms instead of on the main warehouse platform. While cost is driving alternative architectures, performance optimization is driving multi-tiered data architectures, including a strong interest in in-memory data mart deployments. At the high end, data warehousing is now mission-critical (see Note 1). In the past, buyers believed that the vendor with the largest database was the leader. In 2009 size was little more than a classification scheme. Today, smaller data warehouses (those less than 5 terabytes [TB]) are commonly solving organizations' analytic needs.

New entrants introduce technology, then the dominant vendors pick it up after the market shows an interest or they buy the smaller firms. At the same time, the veterans are adding functionality via their own research and development efforts and investments.

In 2009 we saw new vendors entering the market (such as Infobright and ParAccel), and mat0075re vendors offering new solutions (such as IBM with Smart Analytics and Oracle with Exadata). In 2010 we will be watching several new vendors of DBMS engines that currently meet some, but not yet all, of the criteria required to be included here (for example, Algebraix Data and EnterpriseDB's GridSQL); Microsoft has also announced that it will release its SQL Server 2008 R2 Parallel Data Warehouse (from the DATAlegro acquisition). The so-called mega-vendors have placed significant and appropriate emphasis on the formalization of professional services to support data warehouse delivery in 2009. Some have purchased consultancy organizations, others have introduced formal approaches for identifying best practices from their existing field delivery teams and are creating standards of delivery based on those experiences. At the same time, smaller vendors are developing technology and implementation partnerships to create alternative delivery channels for their customers. With a large population of data warehouse novices entering the market now, Gartner will be focusing on the quality of professional service offerings in 2010, and this could be a significant differentiator among vendors (for example, the quality of their partner networks and internal services).

Another trend that surfaced in 2009 was the power of purchasing argument. Gartner estimates that, for all vendors, 60%+ of most data warehouse accounts also have a competitor in the same account performing data warehouse duties, often for the same end-user community. For larger vendors this is a standard operation, for smaller vendors it is a threat. Organizations considering products that compete with their incumbent DBMS vendor should be aware of these larger vendor offerings, but should not accept purchasing arguments as the only justification for eliminating a specialized solution. Simplifying procurement strategies is only one aspect of vendor selection and, candidly, could make the procurement office's job easier, consequently making the BI and data management team's job harder. Be sure to include technical qualifications that consider query count and complexity, usage of the platform, ease of implementation, stability of the support model and other aspects of vendor evaluation.

In 2010 all current and new vendors will begin to establish their positions as they prepare for a major battle over data warehouse DBMS market share. What this means is that each vendor will refine its competitive position with specific peripheral defensive strategies, differentiation based on vertical or horizontal expertise, channel partner strategies, scale-out support in existing clients and the acquisition of new named clients. Additionally, alternative delivery strategies, such as the cloud or the use of open-source software, present niche opportunities, and in some cases rise to become requirements for the leaders. Vendor solutions will focus even more on the ability to isolate and prioritize workload types. Vendors that fail to differentiate their offerings will drop out of the market by choice, or will be forced out by economics. After vendors establish their positions during the next few years, this "title fight" will start, near the end of 2013. Organizations should

increase their emphasis on financial viability and closely align their analytics strategies and their vendor road maps when choosing vendors.

Specific market forces, end-user expectations and resulting vendor solution approaches in 2009 included the following:

- Increased demand for optimization techniques and performance enhancement.
- Prepackaged, pre-balanced warehouse environments delivered via data warehouse appliances.
- Expectations for the delivery of on-site POCs.
- Demands for delivering a fully mixed workload.
- Demands for departmental analytics delivered quickly via data marts.
- Wider indexing and fast performance within clusters of data is being delivered via column-based solutions.
- A wave of new data warehouse implementers is seeking fast-track, low-risk delivery.
- Global organizations are seeking distributed solutions as a potential architecture.

### **Optimization and Performance**

One particularly interesting development in 2009 was the market's acceptance of a standard for implementing two copies of the data warehouse to resolve service-level conflicts in the mixed workload demand. Many organizations are assigning load duties and operational analytics duties to one copy of the warehouse and executing strategic mining, tactical queries and static reports on the other copy of the warehouse. This is in part due to the demand for detailed data in more queries of all types, but also to the data volumes under management. Issues that accompany this trend include fast replication between the copies and the management of near-term data "partitions" for periodic updates from one copy of the warehouse to another. In 2009 and 2010 this was and will be responsible for increased growth and vendor differentiation in the market.

Sophisticated data warehouse platforms built or deployed on a data warehouse DBMS are now the rule, rather than the exception. Such platforms include hardware management of I/O, disk storage and CPU/memory balancing almost as a matter of course. IBM, Netezza, Oracle, Sybase and Teradata have all introduced various new styles of optimization in the past 15 months. Their long experience in data warehousing has placed them in an ideal position to identify "grassroots" issues and deploy solutions designed to meet real-world situations. The new entrants are focusing on optimization as a differentiator; for example, column stores (as opposed to traditional row stores), tokenization of data, and hardware parallelization. Finally, nearly every data warehouse vendor is now addressing the issue of optimized storage for the warehouse. Some are using the concepts of "hot" and "cold" data; others are using different sizes of storage devices to reduce cost (larger drives with lower performance) or increase performance (smaller drives at a higher cost per TB). In addition, many vendors are moving DBMS code into, or closer to, the storage devices, gaining a higher degree of parallelism and more computer power (by using the processors in the storage devices). Several vendors are now offering the option of using and managing solid-state storage as a replacement for some or all of the traditional spinning disc solutions, yielding higher performance, albeit at a much higher price tag.

### **Data Warehouse Appliances**

Data warehouse appliances are not a new concept (see Note 3). Teradata began life, 28 years ago, as a database machine and became a data warehouse appliance when the term "data

warehouse" began to be used. Netezza made the term "appliance" popular when it began as a company about nine years ago. Due to the marketing of Netezza, the DBMS market began to embrace the concept (see "Data Warehouse Appliances Are More Than Just Plug-And-Play"). Today, many vendors have developed a data warehouse appliance offering either as their only product or as a product combining their DBMS with a hardware offering. Roughly 50% of the vendors on the Magic Quadrant offer an appliance, or will in the very near future. Over the past two years, interest in using appliances has grown rapidly, as evidenced by the number of inquiries to Gartner about appliances and the number of POCs that either include or are exclusively appliance-based. Although there are many reasons why organizations consider buying an appliance, the main reason is simplicity. The vendor performs the configuration, balancing the hardware, software and services for a predictable performance. The appliance is delivered complete ("no assembly required") and installs rapidly. Finally, if there are any problems, the appliance requires complex analysis — but only a single call to the appliance vendor for a solution. Appliance offerings are taken into consideration when evaluating the vendors' positioning on the Magic Quadrant, as part of their overall offering.

### **The Intensive POC**

Most organizations heeded Gartner's advice in 2009 to perform a POC with a "shortlist" of vendors during the selection phase of the data warehouse DBMS, which has since become a best practice. A "better match" of solution capabilities to requirements is the result of these POCs. This is especially important when considering one or more of the newer entrants to this market. We recommend that POCs use as much real source-system extracted data (SSED) from the operational systems as possible. We also recommend performing the POC with as many users as possible, creating a data warehouse workload that approaches the environment to be used in production. For a POC, always withhold some of the more complex queries from the vendors in advance of the POC, to be certain that the DBMS has not been "pre-tuned" for your queries. We recommend data loading as part of the POC, even if that is not one of the important requirements of the system. If continuous loading is a requirement, or batch loading is desired, they must be part of the POC. Understanding a solution's capabilities and/or restrictions is important, as the size of the window for loading in most organizations is diminishing as the data warehouse becomes a 24/7 operation.

### **Data Warehouse Mixed Workloads**

The traditional data warehouse workload of queries and reporting is well along the path toward a data warehouse with a mix of several distinct workloads (see Note 2). These six workload types are creating more issues for vendors than the actual size of the data warehouse, even manifesting in databases smaller than 1TB. In addition to service-level expectations (see "Data Warehouse Architecture Best Practices and Guiding Principles"), the size and duration of "useful" data for each community often differs significantly, forcing every aspect of the data warehouse environment to become involved — from I/O channel balancing, through disk management, memory and processor allocation and beyond, into data life cycle management. Through 2012, mixed workload performance will remain the single most important performance issue in data warehousing. As a direct effect of the complex mixed workload, with continuous loading and the increase in automated transactions from the functional analytics in OLTP, the transactional OLTP DBMSs may be able to erode the performance edge that was formerly attributed to specialized data warehouse DBMS solutions.

### **The Resurgence of Data Marts**

Because of the wider acceptance and increasing variety of data-warehouse-driven applications and workload variations, data mart proliferation has become a primary concern. A data mart is defined as an application-specific analytic repository of any size, normally with a specific, smaller group of users than an enterprise data warehouse (EDW) (see "Of Data Warehouses,

Operational Data Stores, Data Marts and Data 'Outhouses'"). The use of data marts specifically for analytics continues with unabated growth (see "Understanding Market Drivers for Independent Data Marts"). This is not only because of the workload that analytics can place on the EDW, but also because some of the data warehouse DBMS engines excel in analytic applications. Data marts can be used to optimize the EDW by off-loading part of the workload to the data mart, returning greater performance to the warehousing environment (for example, column-oriented DBMS engines — such as ParAccel, SAND/DNA Analytics, Sybase IQ and Vertica). Organizations should consider specialized platforms such as column-oriented databases or in-memory capability when deploying data marts.

### **Column-Store DBMSs vs. Column Compression**

Over the years, we have seen an increase in the number of column-store DBMSs available in the data warehousing space. Sybase IQ, now over 14 years old, is the original commercially available Structured Query Language (SQL)-based column-store. Over the past two or three years, we have seen many new column-store entrants into the data warehouse DBMS market: ParAccel and Vertica, for example. As Gartner has stated in the past, the concept of columnization is not new or modern — it was originally conceived as an optimization technique in the 1970s on record-oriented files (called inverted files), indexing all the fields and discarding the records. Gartner stated in 2008 that column-level compression will become available in most, if not all, the traditional DBMS engines over time — this has now happened. Today, Greenplum and Oracle support this compression, with solutions in production, and we expect the others to follow suit. Notably, column compression is not the same as a column-store DBMS; however, it does have similar compression characteristics for both the reduction of storage and increased performance through lower I/O. This will increase the pressure on column-store DBMS engines to distinguish themselves in other ways or face a declining customer base. Some column-store DBMSs include tokenization as a means of further compression. Others create a metadata mapping of the tokens for better performance. Some have unique ways of creating the column-store structure. Finally, we see several changing the pricing model for the software from a more traditional per-user or per-core model to a price based on the volume of SSED loaded into the database. It is becoming clear, however, that creating a product from a single feature is not a long-term prospect.

### **Distributed Data Warehouses**

Distributed warehouses first began to emerge around 2007 (see "Emerging Trends: Introducing the Distributed Data Warehouse"). In 2009, some of the original pioneers of this approach have begun to report performance and data management issues (such as bandwidth for supporting federation). Items such as bandwidth combined with data volumes, globally disparate system administration hours and poorly planned data integration to resolve data discrepancies have presented challenges to this approach. DBMS vendors are beginning to understand these issues and we believe that we will see, in 2010, new features such as optimizer enhancements.

### **Alternative Delivery Models**

- *Managed service warehouse.* The use of data warehouses as a managed service has been an option in this market for more than 10 years. In 2009 two vendors continued to focus on managed data warehouses as a business model (1010data and Kognitio). As we stated in the 2008 Data Warehouse DBMS Magic Quadrant, the megavendors have begun to offer managed data warehouse services — this relegates managed warehouses to an offering and begins to erode this approach as a stand-alone business model. The managed data warehouse will develop into a software as a service (SaaS) model in the next few years for organizations in the small or midsize business category that lack the expertise and funds to support their own data warehouse.



- *Open-source data warehouses.* Open-source DBMSs are still being used in both experimental and more formalized approaches. At this point, open-source warehouses are rare and usually smaller than traditional ones (see "Open-Source DBMS 2009; Gaining in Maturity and Use"). They also generally require a more manual level of support. However, some solutions, such as Greenplum — commercial software using an open-source DBMS (PostgreSQL) — are optimized specifically for data warehousing.
- *Cloud computing and data warehousing.* The use of DBMSs in the cloud has been available since the inception of cloud computing. Many DBMS vendors have offerings that run "in the cloud" (see "Hype Cycle for Data Management, 2009"). Although there are many possible uses of a DBMS in the cloud, data warehousing is only recently emerging as a use case, with few cloud implementations to date. Several data warehouse vendors, such as Vertica, leverage the cloud for POCs due to the rapid setup possible. The issues around the security, multi-tenancy and latency of use of the Internet as a transport for data loading are among the reasons given for the reluctance to locate the data warehouse in the cloud. Although the cloud does contain some smaller data warehouse implementations, it is not rapidly gaining a position as a platform for data warehousing, and we believe that it will be two to five years before there is mainstream adoption.

## Market Definition/Description

The data warehouse DBMS market consists of the vendors that supply DBMS products that provide the database infrastructure of the data warehouse.

For the purposes of this document, a DBMS is defined as a complete software system that supports and manages a logical database or databases in storage. Data warehouse DBMSs are those systems that, in addition to supporting the relational data model (extended to support new structures and data types such as materialized views and XML), also support data availability to independent front-end application software, and include mechanisms to isolate workload requirements and control various parameters of end-user access within a single instance of the data. This market is specific to DBMSs that are used as a platform for a data warehouse. It is important to note that a DBMS cannot "be" a data warehouse. It is the platform on which a data warehouse (solution/data architecture) is deployed.

A data warehouse is a database in which two or more disparate data sources are brought together in an integrated, time-variant repository. Its logical design includes the flexibility to introduce additional disparate data without significant modification of its existing entity design. A data warehouse can be of any size, although Gartner defines a *small* data warehouse as less than 5TB, a *medium* data warehouse as 5TB to 20TB, and a *large* data warehouse as greater than 20TB. For the purposes of measuring the size of a data warehouse database, we define data as *SSED*, excluding all data warehouse design-specific structures (such as indexes, cubes, stars and summary tables). *SSED* is the actual row/byte count of data extracted from all sources.

Finally, there is some debate regarding the definition of a data warehouse appliance. Specific vendors insist that a software-only solution (for example, Greenplum and ParAccel) can be considered an appliance, while others promote the concept of software combined with hardware and support services (for example, HP, IBM, Oracle and Teradata). Gartner's definition is software combined with hardware and support services (see Note 3).

## Inclusion and Exclusion Criteria

- Vendors in this market must have DBMS software that has been generally available for at least a year. We use the most recent release of the software for our evaluation. We do not consider beta releases.
- Vendors must have generated revenue from a minimum of 10 verifiable distinct organizations with data warehouse DBMSs in production.
- Customers in production must have deployed enterprise-scale data warehouses that integrate data from at least two operational source systems for more than one end-user community (such as separate business lines or differing levels of analytics).
- Support for these data warehouse DBMS products must be available from the vendor — we consider open-source DBMS products from vendors that control or participate in the engineering of the DBMS (see "Open-Source DBMS 2009; Gaining in Maturity and Use").
- Data warehouse DBMSs or DBMS products that support an integrated front-end tool, but which can also open their DBMS to competing applications, are included if access is achieved via open-access technology, as opposed to custom-built application programming interfaces.
- Vendors participating in the data warehouse DBMS market must demonstrate their ability to deliver the necessary infrastructure and services to support an enterprise data warehouse.
- Products that include unique file management systems embedded in the front-end tools, or that exclusively support an integrated front-end tool, do not qualify for this Magic Quadrant.

### Vendors Added

- Aster Data.
- Infobright.
- ParAccel.

### Vendors Dropped

None.

## Evaluation Criteria

### Ability to Execute

Ability to execute is primarily concerned with the ability and maturity of the product and the organization. These criteria also consider the portability of the product and its ability to run and scale in different operating environments, giving the customer a range of options. This also includes the differentiation between data warehouse DBMS solutions and data warehouse appliances. The ability to execute criteria are critical to the level of satisfaction and success the customer has attained with the product, and so customer references are weighted heavily throughout these criteria.

### Specific Criteria

- **Product and service** includes the technical attributes of the DBMS. We include scalability, manageability, security, high availability/disaster recovery, support of mixed workloads, support of additional data structures (such as XML) and data loading. These attributes are measured across a variety of database sizes and workloads. We also consider the resources necessary to manage the data warehouse, especially as the data warehouse scales to larger sizes and more complex workloads.
- **Overall viability** includes the corporate aspects of ability to execute, such as the skill level of the personnel, financial stability, R&D investment, and merger and acquisition activity. This also includes management's ability to be responsive to market changes and, therefore, the ability of the company to survive through market difficulties (critical to the long-term survival of the vendor).
- Under **sales execution and pricing**, we examine the price and different pricing models of the DBMS, the ability of the sales force to manage accounts, and whether the sales team is compensated appropriately in line with corporate marketing initiatives. We also include the channel partnerships here, and the ability of the vendor to create and use the partner model.
- **Market responsiveness and track record** covers the issue of references (for example, how many, what size of companies, what configurations and workload mix). Also included are the vendor's ability to adapt to market changes, and its history of being flexible when it comes to market dynamics.
- **Marketing execution** explores how well the vendor understands and builds its products in response to customers' needs (across the spectrum of novice to advanced implementers), in addition to targeting offerings to these needs and to the needs of the market in general. This criterion also includes the completeness of the vendor's offering.
- Customer support and professional services are evaluated as part of the **customer experience** criterion, together with input from customer references, as described earlier. Also included is the track record of POCs and customers' perceptions of the product, as well as aspects of customer loyalty to a given vendor. This demonstrates customers' tolerance of vendor practices, and may indicate satisfaction.
- **Operations** covers the alignment of the company's operations, as well as whether and how they enhance the company's ability to deliver.

**Table 1. Ability to Execute Evaluation Criteria**

<b>Evaluation Criteria</b>	<b>Weighting</b>
Product/Service	high
Overall Viability (Business Unit, Financial, Strategy, Organization)	high
Sales Execution/Pricing	standard
Market Responsiveness and Track Record	high
Marketing Execution	high
Customer Experience	standard
Operations	low

Source: Gartner

## Completeness of Vision

Completeness of vision encompasses the vendor's ability to understand the functionality necessary to support the data warehouse workload design, the product strategy designed to meet market requirements, and the ability to understand overall market trends and influence or lead the market when necessary. A visionary leadership role is necessary for the long-term viability of the product and the company. A vendor's vision is enhanced by its willingness to extend its influence throughout the market by working with independent, third-party application software vendors that deliver data-warehouse-driven solutions (such as BI). A successful vendor will be able not only to understand the competitive landscape of data warehouses, but also to shape the future of this field. However, Gartner's clients are cautioned to be wary of vendors with extremely good vision (including the communication of that vision) but low execution capability. Data warehouses are mission-critical — and poor execution will begin to hurt the overall viability of the organization.

### Specific Criteria

- **Market understanding** covers the vendor's ability to understand and shape the data warehouse DBMS market and show leadership in it. In addition to examining the core competencies of the vendor in this market, we also consider the vendor's awareness of new trends in the market.
- **Marketing strategy** refers to the vendor's marketing messages and its ability to choose appropriate target markets and third-party software vendor partnerships to enhance the marketability of its products. For example, does the vendor encourage and support independent software vendors (ISVs) in its effort to support the DBMS in native mode?
- An important criterion for vision is the **sales strategy**. This encompasses all the channels and partnerships developed to assist with sales. This is especially important for younger organizations, allowing them to greatly increase their presence in the market while maintaining a *lower cost of sales*. This criterion also includes the company's ability to communicate its vision to its field organization and, therefore, to clients and prospects.
- **Offering (product) strategy** covers the areas of portability and packaging of the products. Vendors must demonstrate a strategy that enables customers to choose what they need to build a complete data warehouse solution. We also consider the availability of the vendor's DBMS as a data warehouse appliance.
- The **business model** covers how the vendor's model of a target market combines with product offerings and pricing, and whether it has the ability to produce profits with this model based on the packaging and offerings.
- We do not believe that **vertical/industry strategy** is a major focus of the data warehouse DBMS market, but it does affect the vendor's ability to understand its clients. Specific models for the data warehouse, however, belong in a discussion of applications.
- **Innovation** is a major criterion for evaluating the vision of data warehouse DBMS vendors in developing new functionality, R&D spending, pushing the market in new directions and "pushing the envelope" in the market. This also includes the vendor's ability to innovate and develop new functionality in the DBMS, specifically for the data warehouse. Increasingly, users are expecting the DBMS to become more self-managing and self-tuning, reducing the resources involved in optimizing the data warehouse, especially as the mixed workload increases. This also includes addressing the maturation of alternative delivery methods such as SaaS and cloud infrastructures.

- The organization's worldwide reach and **geographic strategy** are evaluated by considering its ability to leverage resources in geographic regions, as well as subsidiaries and partners in other regions. This is becoming increasingly important as the number of regionally distributed data warehouses increases (as discussed in the Market Overview section). A vendor's success increasingly depends on its ability to market and support its data warehouse DBMS in a geographically dispersed area, using subsidiaries or distributors. This criterion also includes the vendor's ability to support clients throughout the world, around the clock, in many languages.

**Table 2. Completeness of Vision Evaluation Criteria**

<b>Evaluation Criteria</b>	<b>Weighting</b>
Market Understanding	High
Marketing Strategy	Standard
Sales Strategy	Standard
Offering (Product) Strategy	High
Business Model	Standard
Vertical/Industry Strategy	Standard
Innovation	High
Geographic Strategy	Standard

Source: Gartner

## Leaders

The Leaders' quadrant for data warehouse DBMSs contains those vendors that demonstrate the greatest degree of support for data warehouses of all sizes, with large numbers of concurrent users and the management of mixed data warehousing workloads. These vendors lead the market in data warehousing by consistently demonstrating customer satisfaction and strong support, as well as longevity in the data warehouse DBMS market, with strong hardware alliances. Because of this track record, leaders also represent the lowest risk for successful data warehouse implementations, such as lower performance with increasing mixed workloads and database sizes and complexity. Additionally, the maturity of this market demands that leaders maintain a strong vision regarding the key points emerging during the past year: mixed workload management for end-user service-level satisfaction and data volume management.

## Challengers

In the past, the Challengers' quadrant has included vendors with strong offerings for the client base. In 2009, Gartner clients reported that vendor stability coupled with an established offering constitutes a challenger. These vendors have market presence in the data warehouse DBMS space and a proven product, and they have also demonstrated corporate stability. Challengers generally have a highly capable execution model. Ease of implementation, clarity of message and end-client engagement all contribute to making these vendors successful. Challengers show a wide variety of data warehousing implementations across different sizes of data warehouses with mixed workloads. Organizations often purchase challengers' products to deploy in a limited fashion, such as a departmental warehouse or large data mart, with the intention of scaling the solution to enterprise-class deployments. There were no Challengers in 2009 primarily due to new implementers seeking balanced execution and vision, as well as the maturity of the veteran vendors delivering against both axes.

## Visionaries

Visionaries take a forward-thinking approach to managing the hardware, software and end-user aspects of the data warehouse. Visionaries frequently suffer from a lack of global, or even a strong regional, presence. They normally exhibit a smaller market share than leaders and challengers. New entrants with exceptional technology may appear in this quadrant very early after their products have become generally available but, more typically, vendors with unique or exceptional technology will appear in this quadrant when their products have been generally available for several quarters. The Visionaries' quadrant is often populated by new entrants that have new architectures and functionality that are unproven in the market. Vendors must demonstrate that they have customers in production proving the value of the new functionality and architecture. The requirement for the existence of production customers and a general availability of at least a year indicates that visionaries must be more than just startups with a good idea. Frequently, visionaries will drive the others vendors and products in this market toward new concepts and engineering enhancements. In 2009, the Visionary quadrant was thinly populated with vendors answering demands from segments of the market for aggressive strategies in specific functional requirements of the market (for example, the use of MapReduce for large data analytics, massive process scaling in heterogeneous hardware environments, and so on).

## Niche Players

A niche player has low market share or low market appeal. Frequently, a niche player provides an exceptional data warehouse DBMS product, but it is isolated or limited to a specific end-user community, a specific region or a specific industry. Although the solution itself may be without limitations, market adoption is limited. This quadrant contains vendors in several categories: 1) vendors with data warehouse DBMS products that lack a strong or large customer base; 2) vendors with a data warehouse DBMS that lacks the functionality of those of the leaders; and 3) vendors with new data warehouse DBMS products that lack general customer acceptance or the proven functionality to move beyond niche status. Niche players typically offer smaller, specialized solutions that are used for specific data warehouse applications, depending on the needs of the client.

## Vendor Strengths and Cautions

### 1010data

1010data (New York, NY) is a 10 year-old managed service data warehouse provider with an integrated DBMS and BI solution targeted at the business side of organizations, primarily in the financial sector. 1010data can host its solution for its customers in the traditional SaaS model or support a managed solution at the customer site.

### Strengths

- 1010data offers a solution including a DBMS to provide high-speed analytics for businesses. It is a fast-to-market solution for organizations needing a BI application, or lacking BI and data warehousing expertise. Its DBMS is fully compliant with SQL and has an Open Database Connectivity interface that can be used for other applications, in addition to its own.
- As 1010data is a SaaS vendor with a complete solution, the business unit or even the IT organization needs little experience in data warehousing or BI. Also, as a managed service solution, it can complement the internal IT department with fast-to-market solutions for business units, alleviating the resource consumption within IT.

- The managed service model allows 1010data to leverage software solutions across multiple customers. As new applications are created for a client, they become available to all clients, increasing the availability of applications to businesses.
- According to our reference checks, 1010 data is beginning to move from the financial sector (where it began) to a broader market for its offerings. 1010data now reports over 100 customers, and its references support our belief that it is one of the stronger small data warehouse DBMS vendors. In addition, 1010data has seen a growing number of customers install the system on-premises as a managed solution, with several using 1010data as the enterprise data warehouse solution.

## Cautions

- With only a fully managed service model, 1010data is susceptible to push-back from IT departments that wish to have all data warehouses in-house with governance of the data assets of the organization. A big challenge for data warehouse SaaS solutions are the issues surrounding remote locations, security and data transfer performance (perceived or real). 1010data will and does install its system on-premises; however, it is still managed on-premises by 1010data, raising the issue of governance and control for some potential customers.
- 1010data is sold as a fully integrated DBMS and BI solution, limiting potential customers to those wanting a full solution. 1010data is a compliant relational DBMS (RDBMS), and customers can use the DBMS as a stand-alone system if desired, although few currently use 1010data this way.
- As a solution vendor, 1010data has a different competitive model than pure-play DBMS solutions. In addition to competing in the data warehouse DBMS market, it competes with system integration vendors that offer outsourced solutions, such as Cognizant and HP (via EDS). Additionally, IBM, Oracle and other large vendors with professional service organizations compete with 1010data in two markets, both for the data warehouse DBMS and the services.

## Aster Data

Aster Data (San Carlos, CA) is one of the new entries in this year's Magic Quadrant, with a massively parallel processing (MPP) DBMS for data warehousing and analytics (see "Hype Cycle for Data Management, 2009"). Aster Data offers in-DBMS analytics and MapReduce applications.

## Strengths

- Aster Data's *nCluster* is an MPP DBMS implementation with up to four distinct tiers of nodes: 1) queen tier for optimizing and coordinating the queries; 2) loader tier for loading or exporting data; 3) worker tier for performing the parallel queries; and 4) backup tier for online backup and recovery. This division allows resources to be balanced during periods of different workloads. There is also a dynamic workload manager that controls the execution of the workload, balancing it with the use of rule-based management.
- Aster Data also allows applications, such as analytics and MapReduce, to run in parallel execution on the worker servers. This enables in-DBMS analytics in a far less complex manner and with better performance than with other DBMS engines. Further, because these applications are running in the *nCluster* product, they are subject to control by the workload manager.

- Aster Data offers a cloud-enabled version of *nCluster* for both Amazon and AppNexus platforms. Although the cloud version is only beginning to see traction, it is a differentiator from many other data warehouse DBMS engines which only allow their licenses to be used in this environment. Aster Data also offers an appliance version of *nCluster* available on Dell hardware, combined with data integration software from Informatica and MicroStrategy for BI, allowing Aster Data to compete with the appliance-only vendors.
- Aster Data's references report very high performance with *nCluster* in workloads with and without the use of MapReduce, verifying the capabilities of the dynamic workload management.

### Cautions

- As the newest entrant to the DBMS data warehouse world, Aster Data carries more risk than the larger DBMS vendors. We recommend a thorough POC with Aster Data and a minimum of two other vendors. We also recommend, if MapReduce is to be used, that it be part of the POC.
- As with many new DBMS engines, there are some restrictions on the features available for deployment and administration. For example, references reported missing functionality such as database views, stored procedures and some types of database schema, although our understanding is that most of this is available in the next release.
- As with other small vendors with a solid architecture that is different from the traditional DBMS, Aster Data will be a candidate for acquisition by another vendor wanting to develop, adopt and implement Aster Data's architecture within its own DBMS infrastructure. Also, Aster Data will encounter the same issues with incumbent vendors, as described in the Market Overview section.

### Greenplum

Greenplum (San Mateo, CA) is an MPP data warehouse DBMS based on open-source DBMS PostgreSQL running on Linux and Unix. It can be sold as an appliance or as a stand-alone DBMS and has just over 100 customers worldwide.

### Strengths

- Greenplum is a stand-alone DBMS engineered for data warehousing. It has demonstrated scalability in production to hundreds of terabytes. It has also demonstrated the ability to run and manage the mixed workload in a number of references. Through its software architecture, Greenplum is able to move DBMS code to the storage device, thereby increasing performance. Greenplum is one of the first data warehouse DBMSs to implement MapReduce internally for high-scale analytics.
- Greenplum was the first data warehouse DBMS vendor to deliver a DBMS solution for use in a private cloud infrastructure (Enterprise Data Cloud Initiative), followed by Aster Data and Teradata. It allows for the creation of a data warehouse environment with self-service provisioning and elastic scale, using the Internet.
- Although Greenplum began by selling through a close partnership with Sun Microsystems (saving valuable cash for use in development rather than hiring a sales force), over the past two years it has added its own sales force. In addition, it has created partnerships with Dell, HP and recently Cisco, giving Greenplum additional options for hardware platforms. In hindsight, this was a solid strategy, in light of Oracle



acquiring Sun. This strategy averted the issue of having only one partner, as stated in the Greenplum Cautions section of the 2008 Magic Quadrant (see "Magic Quadrant for Data Warehouse Database Management Systems"). Today, the majority of new sales are on platforms other than Sun.

- The company's use of an open-source DBMS as the core work engine also helps to reduce costs, while it concentrates on the management software surrounding the data warehouse and the optimization features necessary for a complex, mixed workload environment. Greenplum has also been responsible for many enhancements to the community version of PostgreSQL.

## Cautions

- With only 100 customers, Greenplum is still a relatively small vendor, especially compared to the large mature vendors in the Leader quadrant. In POCs it finds itself competing with IBM, Oracle and Teradata. We note that Greenplum does win its share of these POCs. This remains a concern, though, as the larger, mature vendors have bigger R&D and marketing budgets and continue to add functionality, allowing them to compete with an innovative vendor such as Greenplum.
- As described in the Market Overview section, Greenplum, like other data warehouse-only vendors, will also begin to get push-back from prospects in situations where it is now possible to use the data warehouse from the incumbent vendor (such as IBM, Microsoft and Oracle).
- As competition in the data warehouse DBMS market grows and matures, Greenplum will find it much harder to differentiate itself. Within the next several years, Greenplum may be acquired by a vendor in the market to complement its technology, much as Microsoft acquired DATAlegro in 2008.

## HP

HP (Palo Alto, CA) Neoview came on to the market in April 2007 as an MPP data warehouse appliance based on the legacy of HP NonStop and HP Integrity hardware. It is a highly scalable clustered DBMS solution. It is an offering within the Business Intelligence Solutions (BIS) group of HP's software division.

## Strengths

- HP Neoview continues to have a strong technology base, combining HP server technology with the HP NonStop product line. This gives Neoview a strong fault-tolerant capability. The system is engineered to be highly scalable, handling databases ranging in size from relatively small to very large. Neoview also has many enhancements for data warehousing performance, such as its use of memory, a strong optimizer, compression and partitioning.
- HP Neoview optimizes all available memory when executing large queries. This is very effective when very large tables need to be joined to many small tables, which is the usual configuration for a data mart or star schema. As a result, HP Neoview can reduce the amount of optimization work required in larger data warehouses. The functionality is similar to "broadcasting" smaller tables as copies — but does it in-memory to avoid actually making the copies on disk, which would increase storage requirements. A recent announcement of support for SAP Business Warehouse (BW) leverages these capabilities.

- With the acquisition of EDS earlier in 2009, HP continues to amass data management service capabilities, adding data management expertise and a vast customer base in ERP and out-sourcing. This includes a large SAP BW customer base that relies on vendor-managed solutions, creating significant potential for increasing Neoview's penetration into the SAP BW customer base.
- In 2009, HP BIS announced several very strong partnerships to supplement the Neoview data warehouse appliance and BIS' go-to-market strategy, including Informatica, SAP and SAS (with in-DBMS analytics).

## Cautions

- HP Neoview references verified being in production in adequate volume to meet the inclusion criteria. Although some customers are happy with their experience, several references reported that there have been issues with Neoview meeting the contractual acceptance criteria for performance and stability, as well as taking longer to move into production than expected.
- HP's original stated go-to-market strategy for Neoview was to sell to a few large, strategic accounts, building a reference base, after which it planned to increase its market share by leveraging HP's formidable client base. HP Neoview debuted strongly in the 2008 Magic Quadrant as a Challenger, based primarily on its technical capabilities and the view that HP had developed a solid reference base to support its marketing efforts in 2009. Since the announcement of Neoview in April 2007, HP has not delivered on this strategy, continuing to struggle for wider market acceptance. Admittedly, the original strategy had a slow start; however, there has been little growth since, with customers estimated at fewer than 30 after nearly three years of availability.
- HP has not been able to leverage its breadth and depth of resources to reduce the risks associated with a relatively new product, especially during the current economic crisis. In several situations Gartner clients report that the HP Neoview deployment organization has struggled with implementation, as well as delivering the expected performance and stability.

## IBM

IBM (Armonk, NY) offers stand-alone DBMS solutions as well as data warehouse appliances, recently re-branded as the IBM Smart Analytics System. IBM's data warehouse software, InfoSphere Warehouse, is available on Unix, Linux, Windows and z/OS.

## Strengths

- IBM can engage most of the spectrum of data warehouse implementation approaches, from custom-built to pre-loaded data warehouse appliance users. IBM's InfoSphere Warehouse (a data warehouse offering based on IBM DB2) is a software-only solution. Its data warehouse appliance solution, the IBM InfoSphere Balanced Warehouse, is a combined server and storage hardware solution (using the IBM Power Systems server with AIX, or the System x server with Linux or Windows and the IBM InfoSphere Warehouse) complete with service and support. Recently, IBM introduced the IBM Smart Analytics System — IBM InfoSphere Balanced Warehouse integrated with the recently announced Cognos 8.
- Enhancements to the performance optimization, workload management and BI support in DB2 9.7 (specifically within the DB2 Optimization Feature) show significant promise. Using a combination of statistical performance metrics and estimation techniques,

queries can be classified for priority management. In addition, Optim Database Administrator can propagate schema changes from test to production environments. Another performance optimization feature is partitioned updates to cubes for real-time analytics support.

- IBM is the only DBMS vendor that can offer an information architecture across the organization (the Information Agenda), covering information across all systems, including OLTP, through data warehousing, to retirement of the data (with the Optim products). This is very compelling to the organizations where IBM is the incumbent vendor, and IBM is effective at leveraging the Information Agenda for data warehousing. IBM maintains a strong following among its existing, very large, customer base.
- IBM InfoSphere Warehouse includes many data warehousing features, such as embedded analytics, data visualization and transformation capabilities, integration with SAS and SPSS (recently acquired by IBM), logical and physical data partitioning, compression, workload and performance management, and multi-temperature warehouse support.

## Cautions

- IBM offers a complete data management solution that includes middleware, data architecture tools and DBMS-based warehouse solutions, including data warehouse appliances. However, IBM also has a long history of co-opetition — where IBM competes internally for customer solutions. IBM will sell hardware solutions with competitors' software and software solutions on competitors' hardware. Although this issue is primarily in the sales channel, Gartner clients have reported that they are confused as a result, and customers report differing responses: examining alternatives to IBM products, longer sales cycles and, infrequently, choosing another vendor.
- Some IBM clients continue to omit IBM's data warehouse solutions when they create their own shortlists of vendors under consideration. This is a symptom of IBM's continuing battle for wider name recognition — even though it has been a Magic Quadrant leader for years. As a result, in terms of its ability to execute, in this Magic Quadrant IBM continues to lose ground to Oracle year-on-year. Part of the issue actually arises from IBM's strength in delivering across the entire data warehouse horizon with professional services, product, appliance-based or combination solutions. IBM needs to provide clear guidance to the market (and even its own customer base) as to which of the many strategies are best applied to different customer situations.
- Overall, IBM has continued to grow its DBMS revenue, for 2007 to 2008, at 12.5%, or 0.6% greater than overall market growth for the same period (see "Market Share: Relational Database Management System Software by Operating System, Worldwide, 2008"). Although IBM's growth rate was higher than Oracle by 1.7% in the relational DBMS market for 2008, Oracle still enjoys 48.9% of the market over IBM's 21.9%. To accelerate this growth, IBM must acquire net new customers for data warehousing, or it will become increasingly difficult to gain or hold its position with respect to Oracle or Microsoft (with a share of 16.6% and a growth rate of 16.4%).

## illuminate

A small software vendor, illuminate (Barcelona, Spain) has an integrated data warehouse DBMS (correlation) and BI tools. The focus of the system is to store all the potential relationships between any data element in the database with any other data element. The company has just short of 100 customers, mainly in Spain and Europe, with a few in the U.S. and Latin America.

## Strengths

- Storage size is small and query performance speed is fast. The solution stores abstracted data values as a metadata master set in the database. This enhances data quality (along with some tools from illuminate) by ensuring single storage of each value. A purely column-vectorized approach reduces the volume of the database, as repeated values within a column are addressed, but repeated values in the overall database are still possible. The correlation theory and its use of metadata eliminate even those remaining multi-use redundancies.
- Proprietary technology is shielded via traditionally understood query language and system-level semantics. This structure is automatically built and maintained via illuminate's DBMS intellectual property as data is loaded — DBAs that are used to row- or column-vectorized, hierarchical files, or any other data file management system, do not have to develop custom-load processing.
- Query processing is enhanced, as this solution effectively creates pre-joins for all existing data relationships in the data model. The process is repeated when new data sets are added, with an almost "spider web" effect that stores every correlation that can be inherited from the data already stored in the database, as well as from any of the newly added data.
- Illuminate does have a small but loyal reference base of customers, primarily in Europe. The new go-to-market strategy of working exclusively with partners (now 48 strong) will help, especially in other parts of the world away from Europe.

## Cautions

- In 2008, the company created its partner program for channels and third-party software. Gartner noted in 2008 that illuminate would have to leverage partners and channels if it wanted to gain mind share in the market. This has not happened. In 2009, illuminate grew its partner channel to about 48 partners and two OEMs worldwide. This new strategy is beginning to help by increasing the exposure gained through partners.
- The vendor has little customer presence outside Europe, with most of its business in Spain. It has seen little success in North America since opening its first office in the U.S. over two years ago. This lack of market presence overwhelms its technical capabilities. Again, partnering in North America will have an effect during 2010.
- As with other small vendors, increased market pressure from the mature vendors and the availability of incumbent vendor solutions make it increasingly difficult to compete. This is the main reason that illuminate has shifted its go-to-market strategy to use the partner channel; however, even with the new marketing model, the pressure from mature vendors increases.

## Infobright

Infobright (Toronto, Ontario, Canada) makes its debut on the Magic Quadrant this year (see "Cool Vendors in Data Management and Integration, 2009"). With offices in Canada, Europe and the U.S., Infobright is a combination of a column-vectorized and fully compressed DBMS. Infobright has both open-source (Infobright Community Edition [ICE]) and commercial (Infobright Enterprise Edition [IEE]) versions.

## Strengths

- In 2008, Infobright began offering its ICE and IEE versions on the market, giving it commercially licensed and open-source licensed offerings. There is considerable differentiation between the two products, with IEE including additional features for performance, warranty indemnification and services. In addition, Infobright is part of an open-source reference architecture for BI and data warehousing (which includes Pentaho, JasperSoft, Talend and Infobright). Infobright is the only open-source column-store DBMS on the market, giving it a unique position.
- Since Infobright integrates MySQL's interfaces with the DBMS, customers can leverage existing tools (both data integration tools, including the MySQL loader, and BI tools), and use the Infobright high-speed loader. This allows Infobright to replace the existing DBMS infrastructure more easily. Also, the Knowledge Grid sits above the data packs, adding an additional set of metadata and allowing even greater performance, as reported by Gartner's reference checks.
- Queries are analyzed by the Knowledge Grid to minimize the number of data packs that have to be decompressed for the result. Data packs are the compressed domains/regions of data in Infobright. Decompressing data in-memory is already faster compared with reading full-volume data on disk, so this further enhances performance by limiting the decompression requirements to the data needed.
- Because Infobright has an open-source pricing model for ICE (no license fees) and a low-cost model for IEE (based on the amount of SSED), its cost model makes it very interesting for organizations that want to reduce costs.

## Cautions

- As Infobright is a small, relatively new vendor in the market, it will be challenged by the issues described in the Market Overview section. Infobright must continue to demonstrate differentiation from the mature vendors, offering column compression as well as the mature column-store DBMS, and moving quickly to manage complex workloads.
- Infobright makes extensive use of portions of MySQL using the OEM version of MySQL under the General Public License license. Although Infobright has a long-term (five-year) contract with Sun for MySQL, there remain risks around the uncertain future of MySQL and the Oracle acquisition. We believe that it would be a major effort to change the MySQL portions to another open-source DBMS, such as PostgreSQL, causing some delays in adoption by concerned prospects.
- Infobright will also have to demonstrate revenue growth from its commercial product, and that the effect of the open-source version is not to reduce revenue. If the open-source version is good enough, customers may opt for that in place of the commercial version. Additionally, Infobright may be able to license some of its technology (as EnterpriseDB has done) to increase revenue. Its distinct technology and low revenue also position it as an acquisition target.

## Ingres

Ingres' (Redwood City, CA) solution is now an open-source, general-purpose DBMS with a 30-year history as one of the original RDBMS engines. The company has many customers running mission-critical applications, including data warehouses.

## Strengths

- Ingres, a mature vendor in this market, has more than 10,000 customers using its DBMS. Most customers have OLTP applications, but Ingres has its share of smaller data warehouses (up to about 2TB). Ingres has converted virtually all its pre-open-source customer base to open-source subscription support. It is the only open-source DBMS with a substantial data warehouse customer base, especially with database sizes greater than several hundred gigabytes.
- Ingres has been gaining many third-party software partners, specifically in the BI market. An example is the open-source BI vendor Jaspersoft, which offers a software appliance or Ingres bundle for BI. This is driving a larger number of installations in data warehousing, with both new and existing customers looking for an open-source stack that supports BI.
- Data warehousing is mission-critical. The company's solution is the only open-source DBMS with proven maturity in mission-critical applications, including data warehousing. Although other open-source DBMSs do support mission-critical environments, they will need extra resources to replace missing tools and functionality (see "Open-Source DBMS 2009; Gaining in Maturity and Use").
- Ingres contains many of the necessary features for data warehousing, such as partitioning, compression, parallel query and multidimensional structures.

## Cautions

- Data warehousing is not a strength for Ingres, with its solution used mostly in OLTP applications. Most of its development has focused on stronger OLTP functionality and not on data warehouse functionality. Ingres must also address the areas of enhanced data warehouse functionality, storage management and mixed workload management to compete in the data warehousing DBMS market.
- Ingres' history works against it, as it is 30 years old and has not regained major market traction. This is an issue with market perception, which is difficult to change. Although Ingres has gained new customers and new third-party relationships since becoming an open-source company, to become a serious competitor in this market it must continue to show increased growth in both revenue and the number of new customers it has.
- Although Ingres has professional services in data warehousing and a go-to-market strategy with partners (as stated in the Strengths section, second bullet), it lacks data models and the necessary marketing and sales expertise for data warehousing. Although today Ingres is the strongest open-source DBMS offering for data warehousing; with these extra factors it would be better equipped to compete with the larger, more mature vendors.

## Kognitio

Several years ago, Kognitio (Bracknell, Berkshire, U.K. and Chicago, IL) began by offering data warehousing as a hosted service. Today, it has a mix of customers using its DBMS (WX2) separately as a data warehouse DBMS engine, as well as using data warehousing as a managed service, hosted on hardware located at Kognitio's sites or those of its partners.

## Strengths

- Kognitio comes from a strong DBMS appliance background (following its merger with WhiteCross Systems) and has a track record of solid performance. References report high satisfaction with performance against large database sizes in an analytics and reporting environment for thousands of users. With the 2009 release of WX2 version 7.0 Kognitio has included in-memory analytics as a feature of the DBMS.
- Kognitio has pioneered DaaS, the first data warehouse DBMS to be used as a managed service provided by the DBMS vendor, with its clients buying their data warehousing services from Kognitio, while Kognitio hosts the database. Recently, we have seen more activity in the market with this model. There are two reasons for this. First, some business units are dissatisfied with their IT departments. Buying a managed service for their BI needs bypasses the IT department. Second, some small companies cannot afford their own data warehouse infrastructure but have large volumes of data to process for analytics. This is a growing segment of the data warehouse DBMS market.
- Kognitio continues to gain large clients that install on-site, rather than using Kognitio as a managed service. These installations are large, analytic data warehouses. This demonstrates Kognitio's ability to supply a data warehouse DBMS that is capable of competing with those of many of the market incumbents. In addition, Kognitio offers customers the added flexibility of moving to or from a hosted model as desired.
- Kognitio has opened offices in the U.S. and is actively developing partnerships to sell its product. This has started to produce results, with several new customers from these partnerships. Kognitio has also added several hosting partners in the U.S. and the U.K. that offer managed services on WX2. The U.S. presence and additional partners have allowed Kognitio to grow in 2009, despite the economic downturn.

## Cautions

- Kognitio remains primarily a European vendor, with most of its customers in the U.K. It still has a very limited market reach outside Europe, despite opening offices in the U.S. Kognitio has sold several systems in North America and continues to increase its presence outside Europe.
- As a stand-alone solution, Kognitio must compete with the incumbents (such as IBM, Netezza, Oracle and Teradata), which is becoming increasingly difficult in this market. Kognitio will need to develop additional capabilities to distinguish its products from its competitors. The managed solution does help to distinguish it, as the company can offer on-site systems or managed solutions, as customers prefer. Currently, Kognitio has about half of its customers on-site.
- Kognitio remains a small vendor with fewer than 50 customers worldwide. This makes it increasingly difficult to sell into organizations with incumbent vendors, in addition to competing with some of the lower-priced appliance offerings.

## Microsoft

Microsoft (Redmond, WA) continues to market its SQL Server 2008 DBMS for data warehousing needs that do not require an MPP DBMS. Following its 2008 acquisition of DATAlegro, the market awaits the promise of a Microsoft MPP data warehouse appliance (announced as the SQL Server 2008 R2 Parallel Data Warehouse) in the first half of 2010.

## Strengths

- As reported by Gartner clients during inquiries, Microsoft continues to offer value for the price paid, giving high value with a low total cost of ownership (TCO). The purchase of SQL Server 2008 Enterprise Edition includes SQL Server Analysis Services (SSAS), SQL Server Reporting Services (SSRS) and SQL Server Integration Services (SSIS), which means that online analytical processing (OLAP), reporting and data integration for extraction, transformation and loading (ETL) are included in the entry price, although they are normally deployed using separate servers. The license cost — currently at a list price of \$25,000 per chip or socket — is also lower than many other vendors,
- Microsoft's acquisition of DATAlegro in 2008 has added skilled personnel in data warehousing — particularly in marketing large data warehouse solutions. The first result of the acquisition was the release of SQL Server Fast Track Data Warehouse — a set of optimized and balanced reference architectures for data warehouse deployments in the 4TB to 48TB range.
- SQL Server 2008 scales from small warehouses to medium ones without a great deal of effort, adding many new features such as star join query optimization, data compression, resource governance and policy-based management for the data warehouse workload. The acquisition of DATAlegro is one example of Microsoft's aggressive stance on data warehouse scaling.
- Worldwide support from the vendor is extensive, including partners, value-added resellers, third-party software and tools, and the wide availability of the SQL Server skill base.

## Cautions

- There is significant anticipation regarding SQL Server 2008 R2 (due in 1H10), along with the Parallel Data Warehouse (formerly project "Madison" from the DATAlegro acquisition). Gartner clients report that they are using SQL Server 2008 more freely in data warehousing and data mart roles, but that the majority of implementations are small (under 5TB) and medium (5TB to 20TB) in size. Some organizations report that they are waiting for the SQL Server 2008 R2 Parallel Data Warehouse before they implement larger data warehouses. The time-to-market (just short of two years) of the revised DATAlegro product (using Windows and SQL Server) has caused a few organizations to seek other solutions.
- One of the advantages of the Microsoft solution is also a challenge. Data warehousing and BI are related, but they are not the same topic. BI often leverages data marts (warehouse-dependent or independent) and, as a result, the ability to build an isolated data mart solution using SSIS, SSAS and SQL Server 2008 results in a need for more governance. The problem with flexible solutions is that they demand careful governance.
- SQL Server runs only on Windows Server and, therefore, lacks the portability of most of its competitors. Although Microsoft considers this an advantage (due to tighter integration of SQL Server with the operating system), many IT organizations do not consider SQL Server an option, as they are not willing to run production DBMS infrastructure on Windows Server in the data center environment. Microsoft has announced a Datacenter Edition of SQL Server 2008 R2 to match with the Datacenter Edition of Windows Server, targeted specifically at larger data center infrastructure clients with large Windows servers.



## Netezza

Netezza (Marlborough, MA) now markets its new TwinFin platform, continuing to leverage its hardware acceleration strategy with multi-layered processing, but it has introduced complex and large data set processing beyond the warehouse. Specifically, its work with ISV partners leverages the use of the processors in its architecture.

### Strengths

- In 2009, Netezza's introduction of TwinFin created a much-needed physical separation of its multiple levels of processing technology. The move to a standard hardware architecture (using the IBM System x BladeCenter) allows Netezza to market a modular, upgradable and scalable appliance. Not only does this remove the competitive threat of a proprietary architecture, it also removes the necessity for capacity on demand that was needed with the older Netezza platform. This will also allow Netezza to move to new processor technology more quickly. The new TwinFin product continues to make use of the proprietary field-programmable gate array chip that contains the Netezza DBMS software. Netezza also claims that TwinFin will be able to operate in a multi-generational environment as new versions are released (as Teradata has done for many years). This has yet to be proven.
- Netezza continues to create partnerships with vendors that wish to run their application code on the Netezza product's processors. Due to Netezza's architecture, the effort needed to do this is relatively small, but the result, increasing the parallelism and performance of the application, is very impressive.
- In 2009 Netezza continued to evolve its product with additional system administration, workload management and data management enhancements. While individually these represent technology evolution, taken as a whole Netezza continues to mature this product by following a customer-driven road map. Specific features added over the past several years include: recovery from S-blade failures, data compression, auto-regeneration of disk-stored data after a failure, and system and query statistical metadata for active optimization, among others.
- Netezza actually operates in two segments of the market — it is an add-on for existing warehouses as an appliance-based data mart and an enterprise warehouse for implementations with complex workloads. Netezza continues to improve its complex workload management capabilities, and this is working out, according to our reference checks. Netezza has a strong track record of new customers, with over 300 customers at the end of 2009.

### Cautions

- In 2009, Netezza held its own against the mega-vendors that have entered the space. In 2010, Microsoft will come alongside Oracle and IBM with its own appliance. At this point, Netezza will face its most critical challenges to date. It remains vulnerable to competition from large vendors (or incumbent vendors) in most of its accounts. Organizations considering Netezza should be aware of these larger vendor offerings, but should not accept purchasing arguments as the only justification for eliminating Netezza.
- Netezza is very good at isolating POC constraints when competing head-to-head with other vendors. Prospects are advised that POC results, while excellent and valid, are often based on isolated workload situations (single workload type), or leverage Netezza's massive hardware strategy. With TwinFin this is reported as less of an issue;

however, customers are advised to also perform complex workload testing as part of any POC.

- Netezza's prices are no longer a disruptive force. Other vendors have responded with published pricing and discounting, and have introduced entry-level solutions. Netezza's prospects should no longer assume that pricing is an automatic "win" with this vendor. TwinFin, with its new architecture based on standard hardware, will help here.

## Oracle

Oracle (Redwood Shores, CA) remains a leader in data warehousing, enjoying more than 48% of the RDBMS market (see "Market Share: Relational Database Management System Software by Operating System, Worldwide, 2008"). In 2008, Oracle added its first data warehouse appliance offering — the HP Oracle Database Machine with the HP Oracle Exadata Storage Server. In 2009, Oracle changed the platform from HP to Sun Microsystems and now offers Oracle Exadata V2 with Sun hardware. Also in 2009, Oracle made a bid to acquire Sun.

## Strengths

- Giving customers a wide variety of choices, Oracle now has three distinct data warehouse solutions: Oracle Database 11g (the DBMS stand-alone offering); Oracle Reference Configurations (certified server and storage configurations); and the Sun Oracle Database Machine (Exadata V2) — a data warehouse appliance with storage optimized for data warehouses (Oracle Exadata Storage Server) based on the Oracle Database 11g Release 2 of Real Application Clusters (RAC), Automatic Storage Management (ASM) and Sun hardware, sold and serviced by Oracle. Oracle continues to extend the stack, now to hardware, giving customers a single vendor for support. Since the release of the HP Oracle Database Machine (September 2008), Oracle has evolved its data warehouse go-to-market strategy so that it leads with the Exadata database machine.
- Oracle Database 11g has added enhanced materialized view and cube management (notably transparent SQL access and incremental update). This increases Oracle's capability to deploy end-user optimization layers with features not found in other DBMSs. It also includes enhancements to Oracle's partitioning option, including Partition Advisor, which suggests types of partitioning to enhance performance based on the database schema. Finally, with Oracle Database 11g Release 2, Oracle has added in-memory parallel execution and columnar compression (reducing storage requirements and increasing performance) to the DBMS (columnar compression is available when used with Exadata V2).
- Oracle's solution is the most portable data warehouse platform on the market — running on most hardware with Linux, Unix or Windows.
- Oracle's RAC with ASM has become accepted as an enterprise-level DBMS platform for data warehousing, capable of supporting large data warehouses (defined in the Market Definition section as those bigger than 50TB) (see "Oracle RAC Moved to Mainstream Use"). The scale-out configuration allows for flexibility (adding servers and storage without downtime), while providing a base for the high availability required by the new data warehouse SLAs being implemented.

## Cautions

- Gartner's client inquiries indicate that the DBA team supporting Oracle has a higher full-time equivalent (FTE) commitment — primarily from a higher degree of manual

administration — than some other DBMS solutions for data warehousing. Oracle continues to make progress in this area with automated management tools (for example, the Automatic Database Diagnostic Monitor [ADDM] and ASM), the release of Exadata V2, and by providing auditing tools, advisor tools and other metadata-driven system analysis capability. As Exadata achieves market penetration, the sheer speed of the machine precludes the need for much of the optimization needed in the stand-alone DBMS (one reference reported reducing the number of indexes by a factor of 100 to fewer than five). We continue to hear from some of our clients that the DBA FTE commitment is approaching parity with other mixed OLTP/OLAP DBMS vendors. In these same inquiries, many Oracle DBAs report that this perception is no more than a reaction to the high degree of control provided by the Oracle DBMS.

- Gartner clients continue to cite Oracle's pricing and contract practices (for example, its high prices, uneven and wide-ranging discounts, the high cost of maintenance and its reluctance to negotiate on renewals) as issues that are disproportionate to other vendors. Recently, with price adjustments from Oracle and the current economic environment, we have seen an increase in Gartner inquiries related to Oracle's pricing issues (see "What Oracle's U.S. Price Increases Mean to Your Organization"). Another issue for customers is understanding which features of Oracle's solution are optional and priced separately (for example, Oracle Enterprise Manager Management Packs, partitioning and compression). Organizations are encouraged to remain diligent in assuring compliance with Oracle licensing (see "Prevent Oracle Technology License Compliance Issues").
- We have received reports from our clients that Oracle will not install an Exadata system on-site for POCs (while other vendors will). Oracle prefers to do the POC on its own site and in some cases this has become an issue. If pressed, Oracle will install a machine on-site and has done so. If customers want an on-site POC, we recommend pushing Oracle for this.

## ParAccel

ParAccel (Cupertino, CA) is a new entrant to the Magic Quadrant in 2010. As a software solution it includes the ParAccel column-vectored database and storage management interfacing/management.

### Strengths

- ParAccel's customers routinely perform table joins with millions of records in the tables in their query execution. This includes self-joins in analytics such as market basket analysis and drug interaction analysis. We believe that this is due to solid query optimization techniques for very large data sets, and ParAccel references report that the DBMS performs very well in these cases and faster than the competitive DBMSs tested.
- ParAccel easily combines disk utilization with memory utilization in query processing. Most column-vectored DBMSs can also accomplish this functionality — the point is that ParAccel differentiates with it and provides runtime capability to manage this form of optimization.
- With approximately 20 customers in the pharmaceutical, retail, financial and media/advertising analytics sectors, ParAccel has a good reference base. Its reference base has also demonstrated support for existing BI tools (such as BusinessObjects, Cognos and MicroStrategy). Additionally, reference customers include storage management integration offerings (such as EMC).

- ParAccel performs well in many POCs. Its references report POC testing against many of the high-performance vendors such as IBM, Netezza, Oracle and Teradata.

## Cautions

- ParAccel is a recent entrant in a very big market with many vendors. Data warehouses/marts are a principle part of the data management infrastructure in practically all modern IT organizations. As with any new entrant, an organization must be willing to augment its corporate standard with ParAccel or, for each enterprise-level win, ParAccel has to displace corporate standards first, then supplant current market-share leaders in an organization's mind share, then finish up by winning out on both price and technical POC to become the corporate analytics repository standard. In the near term, ParAccel is likely to continue to compete in the specialty category of recursive, very large data analysis for departmental users.
- ParAccel performs well in less diverse workload environments such as data marts and analytic platforms. It will need to demonstrate the ability to diversify the workload and increase the number of concurrent users considerably to compete for mixed-workload implementations.
- ParAccel has services and customer support commensurate with its size. Although working with partners on implementation and analytic application development, ParAccel is a startup. It will have to be ready with a resolution to scale customer support for the issues that will emerge. Success brings new issues to startups and this is one that must be addressed.

## SAND Technology

SAND Technology (Westmount, Quebec, Canada) is a small DBMS vendor that has had a column-store DBMS in the market for approximately eight years. SAND makes use of additional techniques, such as tokenization and compression, to strengthen its column-store design. It is used as an analytic engine and as an archive engine.

## Strengths

- SAND Technology has had a consistent marketing strategy over the past five or six years as a near-line, SQL-accessible archive, although it also continues to market the SAND product as an analytic DBMS. It has enjoyed success with the archive because of the high compression ratios achieved with column-storage DBMSs. In fact, because of its use of tokenization in addition to the column-store, SAND's solution achieves greater compression than other DBMSs. Near-line archiving experienced a resurgence in 2009 as one potential performance solution for constrained warehouses — off-loading "cool" data that is used in longer period queries in an SQL-accessible archive as an extension to the primary warehouse repository.
- SAND continues to have a partnership with SAP (and has done since 2004) as a near-line data store (as opposed to an offline archive) for the SAP BW in installations where the size of the BW has grown to a degree that it is affecting performance. By integrating with NetWeaver (SAP's middleware), an SQL query to the BW can be routed automatically to either the BW or the SAND near-line storage engine. The performance degradation is minimal and the transparency for end users is an excellent feature.
- SAND continues to have a loyal client base. With new clients being slowly added, not only from its partnerships (with Accenture, Open Text, SAP and TG-Energy) but also

from native SAND products, it can remain in the market as a viable vendor or be acquired for the technology — either situation being good for its customer base.

- SAND, because of the tokenization and column store, is also a good choice for analytic data marts in support of off-loading workload from the enterprise data warehouse. The company also has several customers using SAND as the enterprise data warehouse.

### **Cautions**

- Because of SAND's small size — it has approximately 100 customers — it will continue to struggle against the larger vendors and venture-funded startups that can invest more in R&D, marketing and sales. As we described in the Market Overview section, this is an issue for all column-store DBMSs.
- The long-term partnership between SAP and SAND makes SAND a candidate for acquisition by SAP in the next five years. This would position SAND as a near-line storage engine specifically for the SAP BW, as well as integrating the SAND DBMS as an analytic engine (because of the analytic engine capabilities of a column-based DBMS, as described in the Market Overview section).
- Although SAND's DBMS model is more than a pure column store, it still faces the challenge of proving performance in an enterprise data warehouse environment with a mixed workload. SAND does have several customers using it as a stand-alone data warehouse. It will need to add more mixed-workload management tools and gain more customers for this use case to be considered against the other vendors.

### **Sun Microsystems**

MySQL was acquired by Sun Microsystems (Santa Clara, CA) in January 2008 and is the most widely used open-source DBMS engine (see "Market Share: Relational Database Management System Software by Operating System, Worldwide, 2008"). Now that Oracle has bid to acquire Sun, and with the delays incurred by the EU's opposition, there are many open questions about the future of MySQL (see "The Oracle-Sun Entanglement With the European Commission Continues: An Updated Position").

### **Strengths**

- MySQL has continued to mature, especially since its acquisition by Sun Microsystems in 2008 and with the release of MySQL 5.1. This has brought new functionality, growth in professional services, the addition of Sun's sales force and the continued addition of many new third-party software vendors. MySQL Enterprise, which offers a complete installable version with tools to manage the installation and operating environment, continues to gain rapid market acceptance. Many clients are beginning to use MySQL as a data warehouse engine for small data warehouses of approximately 200GB to 500GB. Generally, many data warehouse implementations begin small and grow over time. MySQL can benefit from this and will see the same pattern, as its scalability is proven over time. MySQL does have several references with multi-terabyte data warehouses in production using a technique called "sharding." This technique splits the database into smaller pieces of less than a terabyte. Although this requires more resources to manage the database and associated storage, it does represent another step in the direction of large data warehouse capabilities.
- Even after its acquisition by Sun, the MySQL solution still maintains a low price point — a free license with support subscriptions ranging from \$599 per year per server to \$40,000 per year for the unlimited server license of MySQL Enterprise.

- Sun continues to enhance the management tools available as part of the MySQL Enterprise offering, recently adding the SQL Query Analyzer tool to assist in optimizing the performance of poorly executing SQL statements.
- MySQL does offer a Cluster edition, which has the ability to scale the DBMS through scale-out with a shared-nothing environment. This will help to increase MySQL in data warehousing as well as OLTP (for which clustering was originally developed).

### Cautions

- MySQL continues to lack references for data warehouses that break the 1TB barrier in a single instance of the DBMS (see the note on sharding in the Strengths section). It will need to demonstrate scaling above 10TB in a mixed workload to dispel the perception of a lack of scalability of the MySQL solution in a data warehouse. As stated in the Strengths section, sharding can be used to break the 1TB barrier; however, it is not a best practice and requires more support from technical staff.
- MySQL still lacks many of the special features necessary to be a serious contender for large data warehouses. Although MySQL has some basic functionality for workload management (such as storing query statistics) and, with MySQL 5.1, it has added partitioning, it will need to add more control and automatic management functionality to handle large data warehouses and the mixed workload.
- The low entry cost of using MySQL does not always equate to low TCO, as the cost to manage a large data warehouse without the broad availability of management tools (as with the larger, more mature data warehouse DBMSs) leads to the use of resources to perform these management tasks manually.

### Sybase

Sybase's (Dublin, CA) IQ Analytics Server was the first of the column-store DBMS systems. It is available as a stand-alone DBMS and as a data warehouse appliance. Sybase also correctly positions Sybase IQ as a performance-capable tool for data marts as well as data warehouses.

### Strengths

- Sybase IQ achieves data compression ranging from two to five times compression, depending on the structure of the data. Because analytics typically makes use of fewer columns but larger numbers of rows, Sybase IQ performs very well for analytic applications. The company has been consistently winning POCs with analytic applications, on occasion, with a performance of 100 times greater. This makes Sybase IQ an extremely desirable DBMS platform for an analytic data mart to optimize and enhance an organization's overall data warehouse architecture. Over the past two years, Sybase has increased its Sybase IQ engineering FTEs by more than 70%, as well as its marketing/sales staffing — demonstrating significant commitment.
- Over the past several years, Sybase has shown an increased ability to move from an analytic data mart to an enterprise data warehouse DBMS. It has added much-needed mixed workload management, faster loading capabilities and query parallelism across multiple processors. Our inquiries show that organizations that started with Sybase IQ in a single application data mart configuration are moving to a more comprehensive data warehouse environment using Sybase IQ.
- Sybase has a strong alliance with IBM's Power Systems division. This channel has resulted in the Sybase Analytic Appliance, offered in three different scale-factor

configurations, based on Sybase IQ and IBM's Power Systems platform, and sold and supported by several system integrators globally. This is part of Sybase's strategy to build complete reference architectures for its target markets in collaboration with infrastructure and tools vendors. In addition, Sybase has implemented this strategy by integrating its Sybase ETL tool for data loading with the Sybase Replication Server for real-time loading solutions, expanding its capabilities and leading to additional prospects with a more comprehensive offering.

- In 2009 Sybase introduced its real-time analytics solution, called Sybase RAP — The Trading Edition, which includes Sybase CEP for complex event processing and a built-in package for time series analytics to support market demand for complex event processing. In addition, Sybase has also rolled out its Small Business Edition package (Sybase IQ SBE), which includes integrated offerings of Sybase IQ, MicroStrategy, Sybase PowerDesigner and data integration tools. This is not a data warehouse in a box, but it is the toolbox for building a departmental data warehouse.

## Cautions

- Although Sybase IQ has a large installed base (greater than 1,700 customers), it faces strong competition from market-leading vendors that have already started to introduce column-based compression within their row-vectorized DBMSs. As stated in the Market Overview section, this poses a threat to all the column-store DBMS engines, although less of a threat to Sybase IQ as a mature product.
- As Sybase expands into the enterprise data warehouse space, it will face increased competition from incumbent vendors and more difficult POCs. While Sybase IQ remains ahead of the column-based newcomers in execution and has shown the ability to scale to EDW solutions, the vendor's challenge will be to respond to new market demands by offering a wider variety of data warehouse solutions, evolving its customers into a full-scale EDW solution.
- Although Sybase IQ can run on clustered servers using its shared disk architecture, it currently only allows for parallel queries within a single server across the CPUs. It does not have the parallel query feature across multiple servers that is found in an MPP architecture, and which some customers want.

## Teradata

Teradata (Dayton, OH) offers several data warehouse appliances combining hardware, operating system and DBMS. Its offerings include dedicated development boxes, entry-level-priced solutions, data marts and data warehouses.

## Strengths

- Teradata allows customers to engage in several market microtrends. First, with the publication of lower pricing and greater discounts, it has addressed increased pricing pressure in the market. The second trend is to offer products to address a broader market — as organizations seek entry-level, mature solutions, for example. With the introduction of the Data Mart Appliance 551, the Extreme Data Appliance 1555 and the Data Warehouse Appliance 2558, Teradata is promoting a solution that addresses the "learn as you grow" mentality of new data warehousing entrants. Teradata has a formalized strategy for combining older equipment with new generations ("investment protection"); the use of virtual work units can be distributed with more work units on newer generation nodes, thus relieving some of the performance pressure on older

equipment. At the high-end (Teradata 5555), Teradata permits the connection of up to 1,024 nodes.

- Teradata's management software (including Teradata Active System Management [TASM]) is a clear strength. The management software manages the entire data warehouse environment, from the operating system to the workload, with software to manage the mixed workload, including a priority scheduling manager to prioritize the workload by application and/or groups of users. Anecdotal information from Gartner inquiries indicates that DBA FTE hours are reduced relative to many competing platforms, effecting the balance of TCO against cost of acquisition (see "Critical Factors in Calculating the Data Warehouse Total Cost of Ownership").
- Teradata continues to show solid growth, with more than 22 quarters of revenue increases, specifically from data warehousing. Its broad customer base (approximately 1,000 customers) also helps to enhance its product and service offerings, including tools to track the development of warehouse design and deployment, developed by its services organization.
- Because of Teradata's architecture, it is well positioned to support the modern mixed workload. In addition to the Enterprise Active Data Warehouse for operational analytics support, features such as object access and query resource filtering, throttles that can be applied to named users, connections or the entire system and performance groups (priority is high, medium, low) comprise most of this management capability. Teradata's solutions, which include the data warehouse platform, data models and professional services dedicated to data warehousing, set it apart from the rest of the market (Gartner estimates that more than 90% of Teradata's business is generated in data warehousing). Additionally, Teradata's focus on analytics and data warehousing workloads has resulted in the introduction of "infrastructure servers," which are servers managed within the Teradata cabinet and made available primarily for analytics applications (for example, SAS and Viewpoint).

## Cautions

- The first concern for Teradata remains the fierce competition from mature DBMSs (IBM DB2, Microsoft and Oracle) in the sub-10TB range, as they become stronger in supporting a mixed workload. Many new end-user organizations deploying new or revised data warehouses will implement them with these competitors, based simply on enterprise standards, because they lack the experience base necessary to discern the more advanced requirements of mixed workloads, high availability and analytics optimization, and because they simply want to use their standard.
- The second concern is the new competition that Teradata is beginning to encounter in larger data warehouse implementations, a portion of the market where Teradata has historically encountered little opposition. New appliance offerings from IBM, Netezza and Oracle are showing the ability to manage large data warehouses well, and in some cases (such as Netezza and Oracle) at a very attractive (low) price point. This will force Teradata to examine its pricing structure further, and will put new pressure on the development organization to design lower-cost hardware with the same or greater performance, without eroding margins.
- Common data warehouse practices have seen a renewal in single-vendor appeal. Teradata offers only the data infrastructure for the BI environment. Other vendors that offer a complete stack — from extraction to reporting and OLAP — will renew claims that "Teradata is proprietary." This is a false claim, as the DBMS is no more proprietary



than DB2, Oracle, SQL Server or Sybase IQ (among others). Today, Teradata runs on commodity hardware and Linux, and is even replacing the proprietary BYNET with InfiniBand and other standard interconnects. Organizations should ignore these claims and instead focus on the decision criteria regarding mixed workload demands, balanced system management and data optimization advantages, which are persistent and pervasive needs in the data warehouse DBMS market.

## Vertica

Vertica (BillERICA, MA), one of the newer entrants to the data warehouse DBMS market, has a column-store analytic DBMS with a number of functional capabilities added for high performance and high availability. It is based on research that came originally from the Massachusetts Institute of Technology (MIT).

### Strengths

- Vertica's solution has had strong early adoption as an analytic data mart, with more than 100 customers in less than two years, about 20% of which are outside North America. The DBMS is inexpensive, with a pricing model based on the amount of SSED loaded into the DBMS, rather than on users, servers, chips or cores. Its fast adoption is also a result of its simple installation and broad portability across hardware systems. References report that they are able to set up Vertica data warehouses very rapidly, sometimes in hours. This is partly because of a feature in the Vertica solution — automatic database design — that requires less optimization of the model.
- The implementation of Vertica is in a cluster of commodity servers, giving it scalability and reliability while differentiating it from Sybase IQ.
- Vertica's DBMS has many features that help to set it apart from other DBMS engines, such as built-in high availability (including active replicas, auto-node recovery and no single point of failure shared-nothing architecture) and data compression (additional to and different from the automatic compression realized as a column-store DBMS). In 2009, Vertica added FlexStore, which is used to increase loading and query performance. This is substantiated by several references that report an increase in performance with FlexStore.
- Vertica's solution was the first DBMS with a "cloud" implementation, using Amazon Elastic Compute Cloud (EC2). This has been an advantage in POCs, as personnel from Vertica have access to the cloud rather than getting through customers' firewalls. Further, implementation and setup are very fast, sometimes taking as little as an hour. This is also a strong offering for development purposes, as deployment is rapid and it can be discarded easily when finished. In addition, Vertica has added connectivity to the Cloudera MapReduce product, allowing users to take advantage of MapReduce without implementing it inside the DBMS.

### Cautions

- There are many entrants in the column-store DBMS space, which makes differentiation difficult. This clearly weighs in favor of more mature products with an installed base and makes it more difficult for newcomers such as Vertica. Although Vertica does have a degree of differentiation, as described in the Strengths section, the difficulty is explaining this to prospective customers. Also, as with the other column-store DBMSs, Vertica will begin to see competition from the more mature DBMS vendors, as column-store compression is added to their DBMSs, as stated in the Market Overview section.

- Vertica has a few customers with very large data sizes. Because of the exceptional compression in a column-store DBMS, we measure the amount of SSED that is loaded into the database. Vertica does have a few customers with as much as 300TBs of SSED requiring considerably less storage in the database. It also has only a few customers with large numbers of users (more than 100). Vertica will need to continue to gain customers with large SSED sizes and greater numbers of concurrent users to compete well against established products, both column-store and traditional.
- Today, Vertica's solution, as a column-store DBMS with little mixed-workload management, is limited to data marts for analytic applications. Gartner believes that this will change in time, as new functionality is added to broaden its capabilities as a data warehouse DBMS.

## **RECOMMENDED READING**

---

"Magic Quadrants and MarketScopes: How Gartner Evaluates Vendors Within a Market"

"Magic Quadrant for Data Warehouse Database Management Systems"

"Cool Vendors in Data Management and Integration, 2009"

"Hype Cycle for Data Management, 2009"

"Market Share: Relational Database Management System Software by Operating System, Worldwide, 2008"

"Market Share: Database Management System Software, Worldwide, 2008"

"Critical Factors in Calculating the Data Warehouse Total Cost of Ownership"

"Operational Analytics and the Emerging Mission-Critical Data Warehouse"

"Of Data Warehouses, Operational Data Stores, Data Marts and Data 'Outhouses"

"Mission-Critical Data Warehouses Demand New SLAs"

"Architecture Issues for Real-Time Data Warehousing"

"Key Issues for Data Management and Integration Initiatives, 2009"

"Key Issues for Delivering a Data Warehouse Project, 2008"

"The Data Warehouse Model Framework"

"Data Warehouse Appliances Are More Than Just Plug-And-Play"

"Sun Oracle Database Machine Focuses on Optimizing Mixed Workloads"

"Open-Source DBMS 2009; Gaining in Maturity and Use"

"How Oracle's Acquisition of Sun Will Affect Database Management Systems"

"The Oracle-Sun Entanglement With the European Commission Continues: An Updated Position"

"Oracle RAC Moved to Mainstream Use"

## **Acronym Key and Glossary Terms**

**ASM** Automatic Storage Management

<b>BI</b>	business intelligence
<b>BW</b>	Business Warehouse
<b>DBA</b>	database administrator
<b>DBMS</b>	database management system
<b>EDW</b>	enterprise data warehouse
<b>ETL</b>	extraction, transformation and loading
<b>I/O</b>	input/output
<b>MPP</b>	massively parallel processing
<b>OLAP</b>	online analytical processing
<b>OLTP</b>	online transaction processing
<b>POC</b>	proof of concept
<b>RAC</b>	Real Application Clusters
<b>RDBMS</b>	relational database management system
<b>SaaS</b>	software as a service
<b>SQL</b>	Structured Query Language
<b>SSAS</b>	SQL Server Analysis Services
<b>SSIS</b>	SQL Server Integration Services
<b>TCO</b>	total cost of ownership

**Note 1**  
**Definition of Mission-Critical Systems**

Mission-critical systems are defined as systems that support business processes and the generation of revenue that, if absent for a period of time, determined by the organization and its service-level agreements, must be replaced by manual procedures to prevent loss of revenue or unacceptable increased business costs. Normally, mission-critical systems require high-availability systems and disaster recovery sites. We have included the use of a DBMS as a data warehouse engine in the mission-critical systems category, as we believe that many (if not most) data warehouses in use today fit the definition of mission-critical.

**Note 2**  
**Definition of Mixed Workload**

The modern complex mixed workload consists of:

- Continuous (near-real-time) data loading — similar to an OLTP workload (due to the updating of indexes and other optimization structures in the data warehouse) — that forces issues in summary and aggregate management to support dashboards and prebuilt reports.

- Batch data loading continues to persist as the market matures and begins to realize that not all data is required for "right time" latency, and that some information, being less volatile, does not need records refreshed as frequently as the more dynamic real-time data elements.
- Large numbers of standard reports ranging in the thousands per day requiring SQL tuning, index creation, new types of storage partitioning and other types of optimization structures in the data warehouse.
- Tactical business analytics in which business process professionals with limited query language experience use prebuilt analytic data objects with aggregated data (pre-joins) and designated dimensional drill downs (summary). They rely on a BI architect to develop commonly used cubes or tables.
- An increasing number of true ad hoc query users (data miners) with a random, unpredictable use of the data, implying a lack of ability to specifically tune for these queries.
- The use of analytics and BI-oriented functionality in OLTP applications, creating a highly tactical use of the data warehouse as a source of information for the OLTP applications requiring high-performance queries. This is one force driving the requirement for high availability in the data warehouse.

### Note 3

#### Definition of a Data Warehouse Appliance

A prepackaged or pre-configured balanced set of hardware (servers, memory, storage and I/O channels), software (operating system, DBMS and management software), service and support, sold as a unit with built-in redundancy for high availability positioned as a platform for data warehousing. Further, it must be sold by the amount of SSED ("raw data") to be stored in the data warehouse and not by the configuration (for example, the number of servers or the number of storage spindles). We allow some flexibility with performance criteria to facilitate the vendors having several variations based on the desired performance SLAs and the type of workload targeted for the appliance. The primary concern is that the user (buyer) cannot change the configuration due to budget issues, therefore adversely affecting the performance of the appliance.

#### Vendors Added or Dropped

We review and adjust our inclusion criteria for Magic Quadrants and MarketScopes as markets change. As a result of these adjustments, the mix of vendors in any Magic Quadrant or MarketScope may change over time. A vendor appearing in a Magic Quadrant or MarketScope one year and not the next does not necessarily indicate that we have changed our opinion of that vendor. This may be a reflection of a change in the market and, therefore, changed evaluation criteria, or a change of focus by a vendor.

#### Evaluation Criteria Definitions

##### Ability to Execute

**Product/Service:** Core goods and services offered by the vendor that compete in/serve the defined market. This includes current product/service capabilities, quality, feature sets, skills, etc., whether offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria.

**Overall Viability (Business Unit, Financial, Strategy, Organization):** Viability includes an assessment of the overall organization's financial health, the financial and practical success of the business unit, and the likelihood that the individual business unit will continue to invest in the product, to continue offering the product and to advance the state of the art within the organization's portfolio of products.

**Sales Execution/Pricing:** The vendor's capabilities in all pre-sales activities and the structure that supports them. This includes deal management, pricing and negotiation, pre-sales support and the overall effectiveness of the sales channel.

**Market Responsiveness and Track Record:** Ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer needs evolve and market dynamics change. This criterion also considers the vendor's history of responsiveness.

**Marketing Execution:** The clarity, quality, creativity and efficacy of programs designed to deliver the organization's message in order to influence the market, promote the brand and business, increase awareness of the products, and establish a positive identification with the product/brand and organization in the minds of buyers. This "mind share" can be driven by a combination of publicity, promotional, thought leadership, word-of-mouth and sales activities.

**Customer Experience:** Relationships, products and services/programs that enable clients to be successful with the products evaluated. Specifically, this includes the ways that customers receive technical support or account support. This can also include ancillary tools, customer support programs (and the quality thereof), availability of user groups, service-level agreements, etc.

**Operations:** The ability of the organization to meet its goals and commitments. Factors include the quality of the organizational structure including skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently on an ongoing basis.

### **Completeness of Vision**

**Market Understanding:** Ability of the vendor to understand buyers' wants and needs and to translate those into products and services. Vendors that show the highest degree of vision listen and understand buyers' wants and needs, and can shape or enhance those with their added vision.

**Marketing Strategy:** A clear, differentiated set of messages consistently communicated throughout the organization and externalized through the website, advertising, customer programs and positioning statements.

**Sales Strategy:** The strategy for selling products that uses the appropriate network of direct and indirect sales, marketing, service and communication affiliates that extend the scope and depth of market reach, skills, expertise, technologies, services and the customer base.

**Offering (Product) Strategy:** The vendor's approach to product development and delivery that emphasizes differentiation, functionality, methodology and feature set as they map to current and future requirements.

**Business Model:** The soundness and logic of the vendor's underlying business proposition.

**Vertical/Industry Strategy:** The vendor's strategy to direct resources, skills and offerings to meet the specific needs of individual market segments, including verticals.

**Innovation:** Direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or pre-emptive purposes.

**Geographic Strategy:** The vendor's strategy to direct resources, skills and offerings to meet the specific needs of geographies outside the "home" or native geography, either directly or through partners, channels and subsidiaries as appropriate for that geography and market.

## REGIONAL HEADQUARTERS

---

### **Corporate Headquarters**

56 Top Gallant Road  
Stamford, CT 06902-7700  
U.S.A.  
+1 203 964 0096

### **European Headquarters**

Tamesis  
The Glanty  
Egham  
Surrey, TW20 9AW  
UNITED KINGDOM  
+44 1784 431611

### **Asia/Pacific Headquarters**

Gartner Australasia Pty. Ltd.  
Level 9, 141 Walker Street  
North Sydney  
New South Wales 2060  
AUSTRALIA  
+61 2 9459 4600

### **Japan Headquarters**

Gartner Japan Ltd.  
Aobadai Hills, 6F  
7-7, Aobadai, 4-chome  
Meguro-ku, Tokyo 153-0042  
JAPAN  
+81 3 3481 3670

### **Latin America Headquarters**

Gartner do Brazil  
Av. das Nações Unidas, 12551  
9º andar—World Trade Center  
04578-903—São Paulo SP  
BRAZIL  
+55 11 3443 1509