Sybase IQ Analytics Server
Competitive Assessment

Seth Grimes

Alta Plana

Sponsored by

Introduction

Old approaches to data warehousing no longer deliver the scalability and performance that users require. Exploding data volumes, operational analytics, mixed analytical workloads, and the need for real-time responsiveness demand new approaches. This realization, now widely accepted by researchers, analysts, and solution providers, has spurred the adoption of a variety of innovative technologies by start-ups and established vendors alike.

These technologies – they include data compression, adaptive indexing methods, and parallelization of data loading and query processing – are not new. **These new technologies were introduced to the data-warehousing market in the 1990s by two companies, Sybase, the leader in high-performance, columnar database technology**, and Teradata, whose proprietary, MPP parallel computing hardware and software, designed for high-throughput transaction processing, was repurposed for data warehousing. Due to rapid advances in processor design – multi-core chips and large-scale clusters are now the norm, almost commodities – the scalability advantage formerly afforded by MPP systems such as Teradata's has been largely eliminated.

Sybase IQ and data warehousing rivals

Despite the emergence of rival column stores in the last several years, Sybase IQ remains a data-warehousing leader with over 1,500 customers (citing 96% satisfaction ratings) compared with around 900 for Teradata. Sybase is far ahead of other rivals in scalable, high-performance data warehousing, whether those rivals are industry giants that propose systems designed for transaction processing, MPP database appliance and software vendors, or columnar database management system (DBMS) vendors.

One emerging rival, Vertica, has attracted significant attention since its 2005 founding by database systems pioneer Dr. Michael Stonebraker. Vertica commercializes a column-oriented approach prototyped by Dr. Stonebraker and others in a university open-source system called C-Store. Vertica is a column store that relies on an MPP architecture for scalability and reliability. Vertica's arrival on the market has helped create a market category out of the already-thriving columnar approach, in large part due to Dr. Stonebraker's star power despite current sales dwarfed by those of Sybase IQ.

The need for an objective comparative evaluation

Vertica's emergence has created the need for an objective comparative evaluation of the two systems, Sybase IQ and Vertica. This paper is such an assessment. It introduces technology elements that comprise high-performance data-warehousing solutions and relates them to business considerations. It is intended to assist organizations that are looking to implement a high-performance data-warehousing system in the evaluation of options and the selection of the solution that offers the best performance, scalability, reliability, and value.

This paper concludes that Sybase IQ is positioned to maintain, if not to extend, its columnar data-warehousing market lead due to technical considerations, the completeness of the Sybase solution set, the availability of associated support and services, and the company's track record of many years of delivering performance and innovation accompanied by the highest levels of customer satisfaction.

Data Analysis for Better Enterprise Decision Making

Data analysis improves business decision making at levels that range from operational to strategic. Analytics embedded in line-of-business systems supports real-time risk assessment, credit decisions, fraud detection, and on-line commerce. At the other end of the spectrum, Business intelligence (BI) enables numbers-driven performance management through forecasting, budgeting, and planning from historical and operational data. Between the extremes, BI tools and dashboards empower managers and analysts to respond promptly and effectively to evolving business conditions.

Given exploding data volumes, intense competitive pressures, and an accelerating pace of business, comprehensive, accurate, useful, and timely data analyses are more important than ever.

Analysis processes

Analysis processes include steps to gather, model, and transform data under the data integration, business intelligence, and data mining rubrics. Users – business and IT analysts and the business stakeholders they support – rely on analytical databases, primarily data warehouses, to support those analyses. Responsive databases are essential, particularly when data volumes grow beyond desktop manageability, when users collaborate and share data, and as organizations adopt processes and software tools whose capabilities go far beyond those of spreadsheets. In effect, users rely on data warehouses in just about every modern, data-driven enterprise.

Analytical databases and data warehousing

Data warehousing emerged in the late 1980s as a means of collecting and organizing operational and transactional data for data analysis. Restructuring is needed because analytical demands differ markedly from a transactional or operational work-load. Further, a separate database is called for to avoid slowing transaction processing.

Online Transaction Processing (OLTP) involves rapid execution of high volumes of relatively simple, predictable read, update, and insert queries, each touching only a few data records. Data structures are highly normalized to ensure fast query processing and minimize locking, which may occur when multiple queries access the same data structures simultaneously. BI and data mining, by contrast, involve aggregations and other calculations that may touch many data records. E.F. Codd, who invented the relational database model, coined the term Online Analytical Processing (OLAP) in 1993 to distinguish this latter class of data usage from OLTP.

> *"Attempting to force one technology or tool to satisfy a particular need for which another tool is more effective and efficient is like attempting to drive a screw into a wall with a hammer when a screwdriver is at hand: the screw may eventually enter the wall but at what cost?"*
>
> -- E.F. Codd, 1993

OLAP took off, typically running against data-management back-end systems that relied on pre-computation of aggregates with accompanying difficulty supporting large data volumes. Relational OLAP, with uses a conventional relational DBMS for data management, boosted scalability, but performance was still limited by reliance on database systems that were designed for transaction processing. Many users still rely on those systems for data warehousing.

A new-old performance approach: columnar database systems

Sybase pioneered a new way of organizing data for analysis with the 1996 introduction of Sybase IQ, an RDBMS based on technology acquired in 1995. Sybase IQ was the first mainstream system to store data by column rather than in the conventional, row-wise fashion that is preferred for transaction processing.

> *"Column-store DBMSs should be considered as primary candidates for analytic data marts due to I/O performance gained from compression and retrieval of fewer columns (typical of analytic applications)... In the two specific areas where column-store DBMS excel (analytics and archiving solutions), there can be a large cost savings and, with analytic applications, a large performance gain."*
>
> -- Gartner Vice President and Analyst Donald Feinberg, 2008

While the columnar approach has steadily gained popularity, it has won broad-market visibility only with the emergence of competing systems. The competition affirms the power and the viability of the approach and its ability to solve real-world business problems. The new visibility in turn creates opportunity to reach new audiences – individuals tasked with delivering responsive analytical solutions and the users they support – to communicate the advantage of the columnar approach and the technical, architectural, and business considerations that differentiate leading products.

Architectural distinctions

Architectural differences among the competing column-stores include, most notably, the scalability model implemented. Sybase IQ uses symmetric multi-processing (SMP) for scale-out on clusters of commodity computing hardware running Windows, Linux, or Unix. A multiplexing component can distribute processing over multiple SMP machines. Leading competitors, by contrast, have adopted a Linux-reliant grid-computing model. There are arguments in favor of each model that an evaluator who considers this distinction important should weigh.

Another distinction is breadth of data management capabilities and the degree of standards support, for the SQL query language and standard interfaces. Sybase IQ is a full-featured DBMS whose capabilities extend to disparate database schemas, to SQL-99 OLAP functions, and to management of diverse data types including text and binary large objects. Rival columnar systems do not offer these capabilities.

Further, architectures should scale not only for growing data volumes but also to support diverse users and usage, complex, unpredictable queries, and, given demand for "front-line" data warehousing, operational analytics on near-real time data. Lastly, enterprise users demand reliability and manageability features, whether built into the server software or delivered via integrated software tools. Solution availability as part of an appliance platform or integrated with business-ready analytical applications is a final architectural distinction to consider.

Enterprise business concerns

The degree of market success achieved by competing vendors reflects non-technical concerns that go beyond product technical capabilities. These business concerns include solution suitability for the business problem, quality of support and professional services, cost of ownership, and confidence in the vendor. Vendor stability and experience and a record of customer satisfaction count, especially given a volatile economic climate.

Columnar Database Technology: Sybase IQ and Vertica

> *"Systems oriented toward ad-hoc querying of large amounts of data should be read-optimized. Data warehouses represent one class of read-optimized system, in which periodically a bulk load of new data is performed, followed by a relatively long period of ad-hoc queries... In such environments, a column store architecture, in which the values for each single column (or attribute) are stored contiguously, should be more efficient. This efficiency has been demonstrated in the warehouse marketplace by products like Sybase IQ."*
>
> -- Michael Stonebraker *et al.*, 2005

Column Store Basics

Sybase IQ and Vertica share an approach to managing data to support analysis. They physically align stored data, whether on disk or held in memory, by column, by the set of values of a particular field across all rows in a table. Otherwise, they look and act like other relational database management systems (RDBMSes), most of which store data by row.

Row storage works well if you will be accessing many or most of the fields in a row in typical queries and when queries touch (read, insert, update, or delete) a small number of table rows. By contrast, when queries touch few table fields, and especially when they calculate aggregate values such as sums, far fewer input/output operations are required when data fields are stored in a columnar fashion, hence the shared Sybase IQ-Vertica approach. There are many implementation differences however.

Comparing Sybase IQ and Vertica

Sybase created the column-store DBMS category and dominates the market it built. Over 1,500 organizations run Sybase IQ. The author estimates the overall column-store market at approximately 1,800 organizations. Sybase IQ is the standard against which challenges define themselves; witness the compliment (overloaded with an unfounded implication that Sybase has not continued to develop its technology) contained in Vertica CEO Ralph Breslauer's August 2008 statement, "Sybase IQ is a great column store, but it's now something like 15 years old."

The products differ in their approaches to optimizing database performance, capacity, and scalability; in the variety of data supported; in their query and user interfaces; in administration approach and manageability; and in deployment options. To understand the differences, product evaluators need an objective, point-by-point comparison of leading options to facilitate their evaluations. This paper's comparison groups comparison points into two, broad sets: technical and business. For each comparison point, the author –

- Defines the comparison point and explains its importance.
- Explains Vertica's approach.
- Explains Sybase IQ's approach.
- Provides an objective analysis.

Technology Comparison: Sybase IQ and Vertica

The following is a summary table of technology comparison points, which are then individually covered in detailed subsections.

Consideration	Vertica	Sybase IQ
SQL & Data management	Basic numerical, character, and date/time types; varchar up to 4,000 characters	Numerical and character, money, decimal, character & binary large object (LOB) types; OLAP extensions; stored procedures
Scale-up vs. scale-out	Uses MPP grids for scale-out	Uses SMP and grids for scale-up and scale-out; independent storage and compute scaling
Platform	Linux; virtual deployment; cloud deployment	Linux, Unix, Windows; appliance; virtual deployment
Partitioning and storage	Column sets are replicated across MPP nodes in multiple sort orders; record sets are horizontally partitioned	Shared-disk architecture relies on optional RAID5 for availability; horizontal partitioning possible
Indexing and compression	Non-indexed column compression	Five core, compressed index types; five specialized indexes: date, date-time, time, text and compare indexes
Schema support	Designed for Star and Snowflake, now generalized	Star and Snowflake; 3NF information factory; flat R-Cube
Data loading	Parallelized load: batch & trickle	Parallelized load: batch & trickle
Enterprise data management		Built-in encryption, ETL; domain vertical applications
Tools, "ecosystem"	Partner analysis & ETL tools	Software for modeling & development, data integration, server management; partner analysis & ETL tools

SQL and data management

Definition: Each DBMS field has a type that defines the content of the field. Is the field a number or a character string or does it have some special meaning and use? How big is it? The more types available, the greater the data-storage flexibility afforded the user.

Vertica: Supports basic numeric (integer, float), character, and date/time types and variable-length characters strings up to 4,000 characters. Support for text and binary large objects is planned.

Sybase IQ: Supports basic numeric (integer, float), character, date/time, money, and decimal types. A Large Object (LOB) management option enables management of character (CLOB) and binary (BLOB) large objects within the database. LOB access, including text search and string

manipulation, is handled through standard SQL statements. Supports stored T-SQL procedures and OLAP functions.

Analysis: Sybase IQ's support for stored procedures, OLAP functions, and extended types such will be a convenience for many users. In particular, large objects, which enable unified analysis of text and data, afford a possibility that is becoming increasingly attractive given the vast amount of "unstructured" information of concern to enterprise analytics users.

Scalability architecture: SMP and MPP

Definition: SMP stands for symmetric multi-processing, distribution of computing workload across a cluster of linked computers. Each node may have multiple, multi-core processors to further parallelize processing on that node. Nodes may have local disk storage and also share disk storage with other nodes. A database system that runs across nodes of an SMP cluster is termed "shared disk." It uses locks to regulate concurrent access to the database, although locking is typically not a bottle-neck for read-intensive data-analysis applications.

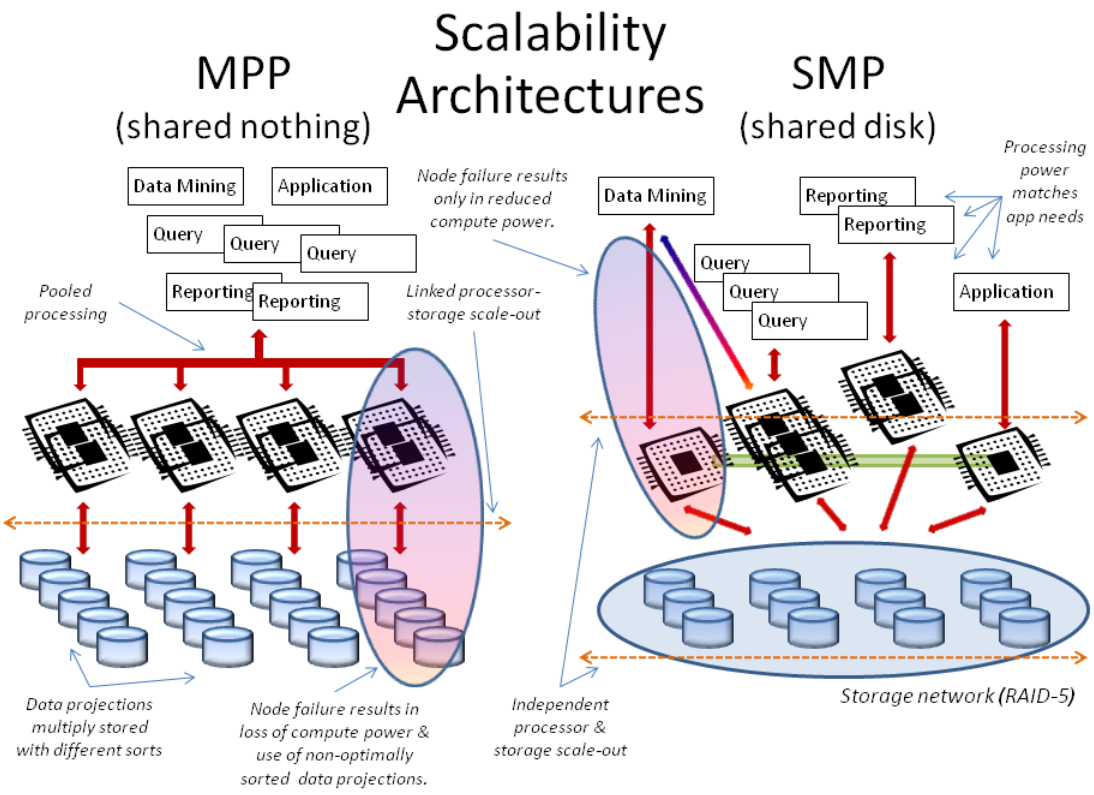
MPP stands for massively parallel processing. MPP enables a "shared nothing" database architecture where each computer-node has only local memory and disk storage. The database is partitioned across the nodes of the MPP cluster, which is termed a "grid" when the hardware is homogeneous.

Vertica: Vertica uses MPP for database scale-out, where one simply adds grid nodes to add both compute and storage capacity. The company states that the system scales to "100s of terabytes of data."

Sybase IQ: Sybase IQ uses SMP for database scale-up/out through the addition of processors or compute nodes with the additional option of multiplex grid technology to support node scale-out. It is possible to assign different tasks to compute nodes that are optimized for those tasks. Compute resources and storage scale independently.

Sun, Sybase, and BMMSoft have built what is claimed and independently benchmarked to be the "world's largest data warehouse," with a full petabyte of data including six trillion transactional records and two billion searchable documents. This news was announced in August, 2007. Other Sybase customers maintain production databases of over 100 TB in size.

Analysis: The author has seen no evidence that either vendor's hardware scalability (SMP or MPP) or reliability model has limited or adversely affected any customer deployment. Sybase IQ's independent compute and storage scaling will be attractive to prospective users, however, as will the Sybase IQ Multiplex Grid's ability to support many concurrent users and heterogeneous processing without performance degradation. Sybase IQ's long-standing ability to store large objects will also benefit some users.



Vertica implements an MPP scalability architecture compared to Sybase IQ's use of SMP for independent processor and storage scalability.

Platform

Definition: Most data warehouses are built on DBMS software that is installed locally on servers that are under the DW user's control. Configuration and maintenance of the DBMS, hardware, operating system, and other system software are handled by in-house IT staff.

Some organizations find it attractive to run a DBMS-hardware pairing where the database software, operating system, and hardware have been specially designed and tuned to optimize data warehousing performance. Data warehouse appliances, in particular, have gained in popularity in recent years.

A third deployment option that has emerged recently is hosted, "as a service" data warehousing, where data is loaded across the network for management on remote servers.

Use of a compute grid is possible with any of these three options. In grid computing, uniform, modular hardware (processors and storage) elements are added or removed from task availability on an as-needed basis, often with pay-as-you-go pricing.

Vertica: Vertica is available in all three deployment modes: 1) For installation on grid hardware running Linux, 2) as an appliance with HP BladeSystem c3000 hardware and Red Hat Enterprise Linux, and 3) hosted on the Amazon Elastic Compute Cloud (EC2). Several of Vertica's

approximately 50 customers have chosen EC2 deployment for production systems. Vertica may be deployed as an embedded DBMS as part of a value-added solution. Vertica is also recently available for VMware's hypervisor virtualization platform.

Sybase IQ: The vast majority of Sybase IQ deployments are installed. The software runs on clusters of Linux, Unix, or Windows environments. The Sybase Analytic Appliance features Sybase IQ running on IBM Power Systems Unix hardware. It is bundled with Sybase PowerDesigner and business intelligence software from MicroStrategy, the leading pure-play BI provider. Sybase IQ may be deployed as an embedded DBMS as part of a value-added solution. Sybase is planning to offer telecommunications vertical specific analytics services for telecom operators via a hosted SaaS model powered by Sybase IQ. Sybase is also evaluating cloud deployment but has not yet announced an offering.

Analysis: Most customers will choose the installed option whether for Vertica and Sybase, and appliances will continue to be attractive for many. Cloud deployment will be attractive for some customers because it can lower entry costs and speed deployment and does not entail long-term commitment although it will not be suitable for higher data volumes and it will not be attractive for organizations that are creating heavily used enterprise data warehouses. Vertica clearly has a cloud-deployment head start although the number of adoptions is small and the advantage will not last given Sybase SaaS availability.

Partitioning and storage

Definition: Partitioning splits databases and individual data tables into parts, typically with the goal of boosting performance by distributing data across physical devices to enable parallel data access. Tables may be partitioned "horizontally," with sets of rows in different partitions. Column-stores are typically partitioned "vertically," with different columns in different partitions. And databases may be partitioned, with different tables in different partition spaces.

Sybase IQ: Sybase IQ newly allows optional range partitioning to split data objects but does not rely on partitioning for performance or recoverability. This option is meant primarily to address Information Lifecycle Management needs, i.e., to manage data in the data warehouse as it "ages" for both cost and regulatory purposes. The company notes that in some use cases, range partitioning also improves performance for both loads and queries.

Sybase IQ may be placed on RAID-5 storage to boost availability and recovery, and storage may be striped over disk volumes. These are established and very widely used approaches.

Vertica: Physically, Vertica stores data as "projections," defined as "collections of sorted columns similar to materialized views. Multiple projections stored on [MPP nodes] can contain overlapping subsets of columns with different sort orders to ensure high availability and enhance performance by executing queries against the projection(s) with the most appropriate columns and sort orders." Vertica uses an on-node Write-Optimized Store (WOS) for recording inserts, updates, and deletes and a Read-Optimized Store (ROS) for handling queries.

Vertica's redundant data storage in multiple, differently sorted partitions, supports the DBMS's k-safe availability and recovery model, albeit with reduced query performance when a node is lost. It is an appropriate reliability model for an MPP grid system although the design means that 1) the system does not handle significant volumes or DELETE or UPDATE queries well and 2) the system may not handle unforeseen workloads/queries well.

Analysis: The author has seen no evidence that either vendor's storage model has limited or adversely affected any customer deployment.

Indexing and compression

Definition: Conventional RDBMSes rely on b-tree indexes to speed random access to small numbers of records. Performance with columnar databases, with few updates and heavy use of aggregations, are better served by other approaches. Columnar storage also allows for aggressive compression, reducing space and disk-read requirements. Choice of indexing/storage and compression methods for a well-designed columnar system will match the characteristics of the data stored in a particular column.

Vertica: Vertica stores multiple, sorted, compressed data projections to enhance query speed and does not use indexes. Design of projections is DBA controlled, with projections generated automatically from a logical schema and sample queries. Compression is adaptive, depending on data characteristics and including run-length encoding and delta indexing

Sybase IQ: Supports five core indexes – fast projection, low cardinality, bit-wise, high group, and join – and five specialized indexes, text, date, date-time, time, and compare. Index usage is determined by data cardinality (number of unique values), data type, and column usage. With fast-projection indexes, the column is the index. Low cardinality indexes are used when there are relatively few unique values. They rely on tokens that represent actual values. Bit-wise indexes are for high-cardinality columns. High-group indexes are b-trees, used when several columns are included in a compound index. Join indexes optimize table joins, both linear (conventional) and star.

Application of indexing and compression is automatic or DBA controlled in certain cases. Sorting is not required.

Analysis: Vertica projections and Sybase indexing are appropriately optimized to the differing database architectures and supported schema, albeit at the cost of storage duplication in Vertica's case.

Schema support

Definition: Data warehouses typically follow one of two design patterns, 1) a star- or snowflake schema that implements a dimensional model or 2) a normalized, (3NF or third-normal-form) information-warehouse schema. Data marts, whether free-standing or extracted from the data warehouse, are often built as OLAP multi-dimensional cubes.

Vertica: Initially optimized for star and snowflake schemas, now claims support for any type of schema design. Stored procedure support is not clear.

Sybase IQ: Supports star and snowflake schemas, 3NF information-factory schemas, and flat, R-Cubes as well as stored procedures.

Analysis: Sybase IQs schema flexibility means that the data warehouse can better support diverse analyses including, in particular, data mining, which often requires a flat, R-Cube type schema. It also eases data and application migration from other data-warehouse solutions.

Data loading

Definition: Data loading is, of course, the necessary before you can use a database.

Vertica: Vertica has fast, parallelized data loading. The company has released excellent figures loading simulated TPC-H benchmark data although it has not run the benchmarks on the loaded data. Note that Vertica relies on multiple-storage of redundant projections for high-availability and optimal query performance. Company literature does state, "In general, it may take more time to load more projections. Typically, excellent performance can be achieved by a 2-projection design."

Sybase IQ: Sybase IQ parallelizes data loading over multiple grid nodes. The company states that version 15 offers "Improved large data loads, direct loads from client, and flexible multi-node loading capability across the shared-disk multiplex grid to add speed to data loading when needed." Sybase also offers Sybase ETL – a user-friendly grid based ETL tool that is specifically optimized for loading Sybase IQ.

Analysis: Both systems support both batch and trickle loading. The author knows of no organization that has implemented either system that is unable to load information to the data-warehouse fast enough to meet requirements. Both companies partner with leading ETL and data-integration vendors.Sybase also offers its own ETL tool for Sybase IQ.

Enterprise data-management needs

Definition: Enterprises need robust reliability, manageability, and security features. (Reliability is covered under "partitioning and storage" above and manageability under "ecosystem" below.)

Vertica: Vertica does not appear to have built-in security features.

Sybase IQ: Offers built-in strong encryption including Elliptic Curve Cryptography, FIPS certified algorithms, Kerberos based authentication, IPV6 support for secure client-server communication, and data encryption at a database and column levels, user-settable passwords. Sybase IQ is managed via a powerful Grid administration console, Sybase Central. The software includes a smart workload monitor.

Analysis: Sybase IQ offers a more robust response to enterprise security needs and also appears more able to handle mixed, unpredictable workloads than Vertica given Vertica's reliance on multiply-sorted, replicated partitions, whose physical implementation is based on predetermined queries, for performance.

Ecosystem

Definition: A database data-management system doesn't stand on its own. Users and

administrators need tools to use and manage databases and the system.

Vertica: Vertica's toolset includes a Database Designer, which a DBA uses to "recommend a physical database design (including projections, compression and partitioning) that will provide good performance for the queries that the user most commonly issues over his or her database and cover any ad-hoc queries that users may choose to run." DB designer may be used "to re-design the database incrementally to iteratively optimize it for the current workload."

Partner data analysis and data-movement software is available.

Sybase IQ: Sybase IQ is part of a unified, comprehensive set of data management products that includes the Data Integration Suite, the Sybase ASE RDBMS, and SQL Anywhere mobile/embedded data management in addition to analytical applications built on top of Sybase IQ. Sybase IQ uses the Adaptive Server Anywhere front-end, query parser, and catalog, which is also used by other Sybase DBMSes, and supports transactional T-SQL programming.

Sybase IQ is complemented by the PowerDesigner database modeling tool, which accommodates enterprise architecture considerations, and the Eclipse-based WorkSpace modeling tool.

Certified data analysis and data-movement software from partners is available.

Analysis: The completeness and maturity of the Sybase ecosystem offers definite advantages for enterprises with more-complex computing needs without of course creating extra demands on organizations that do not need tools beyond the core analytical database system.

Conclusions

The fundamental technical difference between the Sybase IQ and Vertica columnar DBMSes is the processing-storage model, SMP for Sybase IQ and MPP for Vertica. This difference manifests itself in ways that include:

- Vertica's use of multiply-stored "projections" to speed data access and boost availability versus Sybase's reliance on more typical, established storage methods, and also on the ability to add capacity.
- Sybase IQ's independent scaling of compute-power and storage, which will prove advantageous for certain users, versus Vertica's grid architecture that links compute and storage scaling.
- Sybase IQ's availability on most leading OS platforms (Windows, Linux, Unix)

Other significant differences include Vertica's availability for cloud, as-a-service use, which Sybase IQ is only now rolling out, and Sybase IQ's more extensive and robust set of enterprise features and tools.

Organizations considering either system will want to undertake a structured evaluation and selections process that relates these differences – the strengths of each system as well as business considerations – to their own, particular circumstances and requirements. For this reason, we now recommend a selection process.

How to Select a Column Store DBMS

The choice of column-store DBMS technology should be rooted in business drivers. You first identify the business problems to be solved and then design an evaluation process that will address both technical and business concerns linked to those needs. Evaluations should assess product architectural, functional gaps, vendor domain experience, time-to-implementation, cost, and return on investment. One shouldn't simply compare laundry lists of product capabilities. *Relevant* capabilities should be traced back to *prioritized* requirements in a structured evaluation process.

Best practices call for proof-of-concept prototyping. You will learn lessons in the course of prototyping that will help you shape a sensible, workable implementation approach. Implementation may involve reworking business processes to take advantage of the new technologies. At every stage of a phased introduction of the new technology, it helps to manage and measure results against expectations.

Define business drivers

Every business seeks to deliver quality products or services, efficiently, effectively, competitively, and sustainably. The search for profitability – or for public-sector concerns, mission performance – is the core driver. It is implemented in strategy that addresses factors such as revenue and market share, cost of production and delivery, quality, timeliness, customer satisfaction, and customer retention. But strategy sometimes falls short, creating risk/opportunity gaps between goals and capabilities.

Business gap analysis

A gap analysis will catalog unmet needs, which may stem from scenarios that include:

- Failure to use important data sources in decision making.
- Unacceptable delay reacting to significant business events.
- Outdated data-warehousing practices that limit data-analysis capabilities.
- Inability to get appropriate, useful analytics to stakeholders who need them.

A gap analysis would appraise current systems' ability to respond to business needs via a line of questioning that justifies exploring alternative approaches.

Map data, analytics, and business needs

Good, timely data is essential, but it is not enough. We need insights that can inform optimal business decision making. Column-store DBMS technology support this need by enabling complex and/or high-volume analyses – via business intelligence, data mining, predictive analytics – against large volumes of customer, transactional, operational, competitive, and demographic information. It is up to you relate these abstract capabilities to your own situation via questions that include:

What tactical steps will help you optimize business performance?	Sales focused organizations seek to boost customer capture and retention and revenue generation. Payment processors aim to reduce fraud. These are samples. What are your organization's goals?
What are data will help you with these steps?	Transactional systems are key source, linked to data from other enterprise operational databases.
What analyses will deliver the answers your users need?	Business drivers – competition – motivate many organizations to advance from reporting to interactive analyses to predictive modeling and embedded, operational analytics.

What levels of performance, usability, and timeliness are required?	Data and analytical *latency* – the lag between a business event and availability of analytical insight – is an increasingly important business factor.

Answers to questions of this nature link business drivers and analytics. They will prove useful when you relate technology options to your organization's own, particular circumstances via a structured solution evaluation.

Evaluate candidate solutions

A solution evaluation studies software capabilities and performance, and it addresses business concerns that start with cost and ROI and extend to support, vendor experience, and customer satisfaction.

Technical and business requirements

You will look at both *functional* and *non-functional* requirements. Functional requirements cover what a system does, how it works. Non-functional requirements describe a system's characteristics. They include capacity, manageability, performance, scalability, security, usability, and so on. They also include business considerations such as cost, support, and vendor experience and reliability.

Start with a couple of questions:

> *What technical capabilities will enable your organization to conduct data analyses that optimize operational and strategic decision making?*
>
> *How and by whom will they be used?*

You may find it useful to explore use cases, which document how *actors* (users, administrators, etc.) carry out particular tasks, looking at a variety of scenarios. Use cases will also surface non-functional requirements, qualitative aspects.

Study customer stories

Case studies explain how other organizations have harnessed a columnar DBMS. They will suggest how a candidate solution can be adapted for your own use.

Also think of customer testimonials as first-line vendor references. Their availability is an indicator of customer satisfaction and vendor performance. Their absence – a vendor's inability to document experience solving business and technical challenges similar to those your organization faces – may indicate significant risk. Do follow up with customer references – you may wish to research these independently – to learn more, directly about their experiences with the software and the vendor.

Design your evaluation

Create evaluation criteria from your prioritized business and technical requirements. Some may be answerable by studying product literature and documentation and others by questioning the vendor and current customers. The evaluation may include creating a proof-of-concept (POC) prototype if the response to evaluation criteria, customer stories and references, and other due-diligence steps do not cover all evaluation needs.

Decision and deployment

Decision and beyond

Your selection decision will weigh responses to evaluation criteria and the proof-of-concept experience if you were able to conduct a POC. Scoring should account for

implementation considerations: solution gaps and development of needed work-arounds, training and deployment-platform costs, time to deployment, and so on.

Implementation strategy

A successful implementation strategy typically involves phased introduction of new technologies, information sources, applications, and business processes. An organization that already has a strong BI/analytics program enjoys a head start. It's nonetheless prudent, when new technology is being introduced, to first target discrete, clearly defined business problems where impact can be measured. Learn from your POC and from early experiences as you deploy your new solution to its full user base.

Do consider that:

You will need to migrate data and applications from your old platform to your new one.

You will need to test the performance and correctness of the new system.

You may need to redesign work flows and work practices.

You may need to rework user interfaces and applications.

You may need to train staff.

A carefully planning technology introduction and early successes help pave the way for broad acceptance and adoption, turning stakeholders into advocates.

Sybase IQ competitive profile

Sybase publishes a comprehensive set of data management and analysis software products that include transactional, analytical, and embedded and mobile database management systems: Sybase Adaptive Server Enterprise (ASE), Sybase IQ, and Sybase SQL Anywhere. These systems are complemented by software for data movement and synchronization, data integration, and data modeling, by the Open Server applications platform, and a variety of other software tools.

Sybase IQ

Sybase announced Sybase IQ 15 in February, 2009, stating "Sybase IQ 15 builds upon our strong heritage in data analytics and represents the most mature & feature-rich column-oriented analytics offering in the marketplace today, as well as the fastest analytics server in the world."

Business and financial profile

Sybase shares are publicly traded with a market valuation of $2.4 billion and almost 300 institutional share holders. 2008 revenue of $1.13 billion represents 10% overall growth, driven by analytics. The company has stated 28% growth in database license revenue in 2008 with Sybase IQ related revenues growing at a higher rate. According to analyst firm IDC, Sybase's business analytics software revenue grew increased 41.7%, from $104 million in 2006 to $148 million in 2007. Growth substantially outpaced the 12.7% rate of increase for the business analytics software market as a whole, which grew to $22.1 billion worldwide.

Sybase's software solutions are used in a broad variety of business domains internationally. The company's industry-specific solutions target business domains that include financial services, healthcare, and government. The company has forged important technology, systems-integrator, ISV, and service-partner alliances. They include BI software leader MicroStrategy and IBM for the Sybase analytical appliance and notable BI and analytics vendors Business Object, Cognos, QlikView, IBM DataStage, Informatica, KXEN, SAS, and SPSS.

Key differentiators

Key points distinguish the Sybase IQ solution from column-store DBMS rivals:

Company experience: Sybase IQ serves a customer base of over 1,500. Analyst firm Gartner stated in December, 2008 that Sybase is one of several veteran vendors whose "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."

Sustained market dominance: The author estimates that Sybase IQ's 200-customer 2007-to-2008 growth represents almost as many customers as license all other column store DBMS systems combined.

Scalability architecture: Sybase IQ scales to handle query-workload and database-size growth via the *independent* addition of compute nodes and storage capacity. The system is adapted to use both symmetric multi-processing (SMP) and grid hardware.

Sybase IQ, given technical and company strengths, supported by customer loyalty and new-customer acquisition, is positioned to maintain and extend its technology and market edge in the columnar DBMS market and for high-end data warehousing.

Sybase IQ Case Studies

Sybase IQ is deployed by more than 1,500 current customers. They apply Sybase's columnar DBMS solution in a variety of industries. Insights drawn from their experience may help you in planning your own evaluation and deployment. Their stories will illustrate how organizations like yours – in your industry, facing business challenges similar to yours, or with a similar computing environment – have benefitted from the technology.

The following three customer stories are only a small sampling of those available, presented here to relate typical benefits gained via a Sybase IQ solution.

Panasonic Industry Technology

Business challenge: Panasonic Industry Technology (PIT) is a subsidiary of Panasonic Japan with $2 billion in annual revenue. In partnership with major high technology manufacturers, PIT provides facilities, electronic components, and sale and maintenance services for manufacturing automation. Operations depend on analysis of data from multiple brands, manufacturers, and business partners, yet PIT reported that even a simple operation such as sorting results by sales person, product category or sales team performance required help from the IT department, often with unacceptably slow response time.

Sybase IQ solution: PIT deployed MicroStrategy BI software with a Sybase IQ database management system. PIT developed Web-based inquiry applications in-house in order to provide end users the flexibility of working directly from the database. The solution is available to end users ranging from senior management to sales and finance staff.

Business benefits: According to Richard Shiau, PIT senior manager of information system department, "The benefits from Sybase IQ reflect our commitment to providing high-quality service for users. The faster we can respond to queries, the more competitive we can be in the market place." The Sybase IQ solution provides fast query responses against timely data. Charles Yu, assistant manager of information system, explains, "Details of each transaction can be displayed... including information of the involved institutions, contact sales person, product tracking numbers, delivery method, and even the warehouses from which they are shipped out." Richard Shiau says "Sybase offered us a solution that was a perfect fit."

Alvion Technologies

Business challenge: Alvion Technologies, Inc. provides outsourced data management services to some of the largest marketing data list owners in the world – names like Axiom, Equifax, Experian, and Dunn & Bradstreet.

Alvion manages outsourcing customers' datasets that range in size up to 190 million records and 200 attributes. Individual customers submit anywhere from 5,000 specialized records of information, to 120 gigabytes of data. Alvion then runs customer-specific data transformations and uploads it to their production servers for access by the end user – the customers of the original data owners. Alvion has 50,000 registered users and is supporting rapidly growing query loads. Data update cycles vary from daily to annual.

Sybase IQ solution: Alvion maintains roughly 550 gigabytes of data in Sybase IQ for query via Web-based storefronts Alvion hosts for the data owners.

The company reports that its Sybase IQ pilot installation was up and running in three days, with benefits immediately apparent. On a random set of previously submitted queries, Sybase IQ reduced large-list response times by a factor of six. One notable query, previously taking 53 minutes, now completed in 3 seconds.

Business benefits: According to Alvion database administrator Bojan Belovic, "Our fast query times are what defines us in this industry. We are building in more complexity with Sybase IQ as the backbone, while maintaining lightning-fast response times. That's going to take us a long way."

The company had been able to expand capacity by simply adding servers without a change to the Sybase IQ software environment, but undertook a hardware-platform upgrade in order to double the volume of customer data managed by Sybase IQ. The company reports that Sybase IQ provides such effective compression that the new data will not impact the performance. Rather, Alvion see this change as dramatically improving its competitive advantage, shortening data fulfillment times while maintaining trademark query speeds appreciated by customers.

Agencia Tributaria

Business challenge: The Agencia Estatal de Administración Tributaria (AEAT) is the Spanish national tax office, responsible for collection of taxes and customs duties. Identification and prevention of possible tax fraud is a key objective. To effectively manage this large-scale initiative, the AEAT handles enormous volumes of data on a daily basis, to which hundreds of users have access.

AEAT was looking to migrate from an internally developed system in order to provide rapid response times to its large quantities of complex inquiries and to accommodating data and usage growth. The tax office had a requirement, inherent in the existing information infrastructure, that the new solution needed to address. It would have to process, manage, and analyze volumes complex, detailed information consolidated in data tables containing over one hundred thousand columns.

Sybase IQ solution: Sybase IQ fulfilled all the demanding requirements of the Spanish tax office, providing the required efficient processing, faster query responses, and dramatic storage compression. AEAT's Sybase IQ database holds 14 terabytes of data, which equates to almost 2,000 million records. And, as the population grows and the tax office generates new legislation, these events create new demands for information. AEAT predicts that in the next few years its database will grow to around 25 terabytes representing around 13,000 million records.

Business benefits: AEAT found that Sybase IQ has the capability to access the information, conduct analysis, and provide answers to all types of inquiries, including unplanned queries, in real-time. This has enabled response times to be significantly reduced and has expanded the output of the organization. AEAT has been able to focus the organization's resources on deploying the analytical tools to improve tax processing and fraud investigations.

About

Seth Grimes

White paper author Seth Grimes is an information technology analyst and analytics strategy consultant. He is contributing editor at *Intelligent Enterprise* magazine, founding chair of the *Text Analytics Summit*, an instructor for *The Data Warehousing Institute (TDWI)*, and text analytics channel expert at the *Business Intelligence Network*.

Seth has worked with database, BI, and decision-support applications and users for over 25 years. He founded Washington DC-based Alta Plana Corporation in 1997. He consults, writes, and speaks on information-systems strategy, data management and analysis systems, industry trends, and emerging analytical technologies.

Seth can be reached at *grimes@altaplana.com*, +1 301-270-0795.

Sybase

For 25 years, Sybase has been a leader in developing and expanding innovative database technology. Since its founding in a Berkeley, Calif., home in 1984, Sybase has earned the trust of many of the world's leading companies for its ability to manage information and deliver unsurpassed levels of data reliability and security. Today, Sybase is the largest enterprise software and services company exclusively focused on managing and mobilizing information. With its global solutions, enterprises can extend their information securely and make it useful for people anywhere using any device.

Sybase

One Sybase Drive
Dublin, CA 94568
Phone: +1 925-236-5000

Visit *www.sybase.com*.