

InDetail



Sybase IQ 15.3

An InDetail Paper by Bloor Research
Author : Philip Howard
Publish date : June 2011

... it is our view that Sybase IQ merits careful review by organisations investigating data warehousing, high-speed analytics and business intelligence options

Philip Howard

Executive summary

Fast facts

Sybase IQ is a column-based relational database that has been designed specifically for analytics and business intelligence applications. It can offer a number of very significant advantages within a data warehousing environment, including performance, scalability and cost of ownership benefits, when compared to conventional approaches.

Although the product has broad applicability, the company targets a number of primary markets. These include:

- Data warehouses for aggregators of data, typically those who offer multi-client data and analytics services.
- Advanced analytics—where there is a significant requirement to support complex and unpredictable queries, either as a data warehouse in its own right or as an analytics accelerator, where Sybase IQ is complementary to an existing warehouse.
- As a report accelerator, offloading high performance reports from operational databases or centralised data warehouses that perform too slowly to meet business needs.
- Applications where there are significant amounts of unstructured elements (for example, documents and images for insurance claims) and where very large quantities of data need to be kept on-line—that may be subject to querying—for a considerable period of time.

More generally, and bearing in mind that Sybase is an SAP company, Sybase IQ is targeted at environments where users have heterogeneous or non-SAP application environments. Where SAP is the dominant provider of applications these accounts will be targeted by SAP BW and SAP's in-memory technology product for data management, HANA.

The benefits associated with Sybase IQ are predicated upon its column-based approach, its scalable grid architecture, and the performance enhancements (often measured in orders of magnitude) that it can offer, when compared to row-based database solutions, while requiring fewer hardware resources. This is especially true where queries are complex or require large table scans and, in the latter case, this has the knock-on advantage

that you do not have to pre-aggregate data, which represents both a performance and a management saving when compared to traditional approaches to data warehousing. The reduced size of Sybase IQ data warehouses (along with other features of the product) also means that Sybase IQ has the potential to offer significant performance advantages when scaling for large numbers of users.

In other words, Sybase aims to provide better performance with a lower total cost of ownership. Moreover, apart from the fact that data is stored by column, in all other respects Sybase IQ acts exactly like a conventional relational database. For instance, you use standard SQL, hardware and operating systems: database schemas are (or may be) the same, as are applications; and training requirements are similar, in that you can add a column as easily as a row, and so on.

Parallelism in Sybase IQ

Parallelism is particularly important in supporting high performance query processing for large and complex environments. In the original 15.0 release of Sybase IQ, the company introduced multiple write nodes into its Multiplex architecture (it had previously only supported multiple read nodes) along with query parallelisation and pipeline parallelisation (which will speed up both query and load performance). Now, in this release, the company has extended its parallel capabilities still further with the introduction of what Sybase calls its PlexQ architecture, which extends Sybase IQ into supporting massively parallel processing (MPP). Previously, you could parallelise a query within a node and you could have queries running in parallel on different nodes but now you can also have a single query distributed across multiple nodes. This can have significant performance advantages for appropriate queries.

Key findings

In the opinion of Bloor Research, the following represent the key facts of which prospective users should be aware:

- In addition to its column-based storage, Sybase IQ delivers a number of specialised indexes in order to further accelerate ad hoc query performance. These include indexes for low cardinality data (which further reduces storage requirements, and improves

Executive summary

query performance, through the use of tokenisation), grouped data, range data, joined columns, real-time comparisons for Web applications, date, and time analysis. In addition, there are textual analysis indexes (providing analytics on unstructured data that may be combined with structured analysis), with capabilities that have been extended in this release through the introduction of in-database text search and analytics.

- Both multi-threading and 24/7 high availability features (including partnerships with relevant storage vendors for high availability and disaster recovery) are available with Sybase IQ. In particular, separate read and write nodes allow for procedures to be executed in parallel, without affecting one another. Read nodes are particularly useful for data aggregators offering multi-client analytics services because a node can be assigned to an individual account for later chargeback.
- The PlexQ architecture (see above) is augmented by the Sybase IQ optimiser, which has been enhanced in this release to recognise when a query will benefit from being distributed and to which nodes the query should be distributed.
- Sybase IQ offers significant performance advantages when compared to conventional approaches. Apart from the features already mentioned, it also supports Rcube flat schemas that can provide major benefits when compared to conventional star schemas. In particular, Rcubes can significantly speed up implementation, as well as improve run-time performance and provide increased flexibility. In addition, Sybase IQ allows on-the-fly changes to schema attributes (columns); that is, you can add/delete columns in a table while the Sybase IQ server is up and running.
- Sybase provides column-based encryption capabilities as well as database-level encryption. This is particularly important for data aggregators with multi-client services where you want to be able to encrypt different customer's data using different algorithms. Encryption is supported for both data-at-rest and data-in-flight.
- In-database analytics provides much better performance than traditional approaches to analytics. It is supported through the use of user-defined functions that are treated as part of Sybase IQ's relevant SQL functions. This means that the analytic processes can take full advantage of Sybase IQ's optimiser and parallel capabilities. The company has a partnership (and there will be others) with Fuzzy Logix to exploit these capabilities in Sybase IQ and with Visual Numerics for Sybase RAP (Sybase RAP embeds Sybase IQ).
- Sybase IQ provides standard ODBC/JDBC/OLE-DB connections to its query engine, thereby enabling access from any standards-based front-end BI tool. Sybase IQ is certified to work with most industry leading tools such as SAP BusinessObjects, Cognos, MicroStrategy, QlikView, iDashboards, SAS, SPSS, KXEN and others. As part of SAP, Sybase IQ is being optimised for use with SAP BusinessObjects.
- Sybase IQ also offers optimised ETL (and ELT) capabilities as a separate add-on product enabling developers to quickly build and deploy their data sets for analysis on Sybase IQ. Alternatively, Sybase IQ is also certified to work with leading third party ETL tools and, again, there are specific optimisations built for ETL functionality inside SAP BusinessObjects Data Services. **Sybase IQ also supports loading data directly from a client, bypassing the DBA and removing the need for him or her to intervene in the loading process.** This is important for environments where, for security or confidentiality reasons, the DBA should not be able to see the data.
- For situations where operational data that you want to query cannot (for compliance or other reasons) be loaded into the data warehousing environment, Sybase IQ supports query federation with data held in Sybase ASE, Oracle, MySQL, SQL Anywhere, and SQL Server environments.
- Sybase IQ can also be loaded on a continuous near-real time basis using an infrastructure comprised of Sybase Replication Server (which has real-time loading capabilities), and a set of scripts generated by Sybase PowerDesigner. Sybase IQ can also be loaded on a real-time basis with events data via Sybase ESP (Event Streaming Platform). Simultaneous loading and querying is provided via Sybase IQ's versioning capability—a new version is created for the load process while the queries run on the older version until the new load is committed.

Executive summary

- Information lifecycle management, which supports the archival of data from front-line to near-line to historic storage, was introduced with Sybase IQ 15. The ability to formally build data retention rules is supported by means of PowerDesigner, a leading modelling and metadata management tool, and WorkSpace Data Analytics, which is an Eclipse-based development environment that supports database, reporting and analytic development in conjunction with Sybase IQ.

The bottom line

Unlike standard row-based databases that were originally designed for online transaction processing, Sybase IQ has been engineered specifically for query processing and ad hoc analysis. As a result, Sybase IQ can offer significant performance and total cost of ownership advantages over traditional products for query-intensive computing requirements. In terms of performance, these benefits have been significantly extended in this release with the introduction of the PlexQ architecture.

These advantages will be most obvious in environments where the query loads are unpredictable, composed largely of ad hoc enquiries. In this scenario, traditional databases cannot be pre-tuned for unexpected queries, but the column-based approach used by Sybase IQ provides highly effective self-tuning capabilities. In addition, complex queries that involve multiple selection criteria across a variety of tables, and those involving large table scans, can be deployed much more efficiently within a column-based environment. Finally, Sybase IQ scales well for large data stores containing finely detailed transactions and sub-transactions, such as clickstream data. Sybase IQ does not require data to be pre-aggregated for analysis, allowing users to efficiently and quickly analyse atomic level data.

Sybase IQ's underlying columnar architecture is more efficient: servers have to do much less work to answer any particular query when data is organised by column. Of course, Sybase has augmented this basic design advantage in other ways but this is its key differentiator, along with its reduced cost of ownership. This is achieved in two ways: first through lower absolute price, which is both a function of this improved performance as well as the compression techniques that Sybase IQ applies against individual columns. This reduces disk requirements and, consequently, the

necessary investment in hardware. Secondly, this combines with the product's reduced administration and tuning requirements to produce significantly less management overhead when compared with that of the traditional enterprise data warehouse vendors.

The data warehouse landscape, however, is now no longer the sole domain of the traditional suppliers. Today, there is significant interest in data warehouse appliances. In one sense, the market entry of these specialists has helped Sybase IQ because it has called into question the dominance exerted by the major providers of row-based databases as organisations re-think their available options. It has also validated the column-based approach. The rise of purpose-built appliances has raised the interest in Sybase IQ for the report 'accelerator' market in particular. However, interest in data warehouse appliances also introduces a new class of competitors that claim similar performance improvements with an excellent ease of use profile.

Where Sybase has an advantage over these appliance vendors is that Sybase IQ offers more flexible tuning abilities. Most appliance vendors eschew the use of indexes completely or only use them in very limited circumstances. Sybase IQ provides a range of indexing options. While these increase the amount of administration required, they make the product far more adaptable to variations in data cardinality and datatypes. Also, a built-in index advisor removes some of the mystery around selection and creation of appropriate indexes. Further, we know of no data warehouse appliance that is currently able to support text analytics (as Sybase IQ does) and the ability to manage mixed query workloads is also typically limited when compared to what Sybase can offer. It is also worth noting that Sybase IQ can easily and linearly scale up to support a large data set and user workload. Additional disk space can be added to the shared disk pool to support growing data sets and read and write nodes may be independently added to the Sybase IQ grid in small increments to support an expanding user base.

To conclude, Sybase IQ is well positioned to compete with both the traditional and appliance vendors. While it has different advantages in different environments it is our view that Sybase IQ merits careful review by organisations investigating data warehousing, high-speed analytics and business intelligence options.

The product

The current version number of Sybase IQ is version 15.3. Supporting products, including Sybase IQ InfoPrimer (formerly Sybase ETL), Sybase PowerDesigner, Sybase Replication Server and the WorkSpace Database Analytics IDE, all leverage the features available within Sybase IQ. In the case of Sybase IQ InfoPrimer this is specifically optimised to be used in conjunction with Sybase IQ and only supports Sybase IQ as a target.

The product runs under Windows, Linux (Red Hat and SuSE) and the leading UNIX operating systems from HP, IBM and Sun Microsystems (Oracle). Language support for analytic developers includes Perl, Python, PHP, ADO.Net, OLE-DB, and Ruby on Rails amongst others.

Logical architecture

Sybase IQ has been designed specifically for data warehousing. That is to say, it is not optimised for transaction processing and, therefore, it does not include the sort of facilities you would need for transaction processing, as opposed to data warehousing. This is important because the leading merchant databases offer both sets of capabilities so that Sybase IQ has a smaller footprint and is less complex than these offerings.

Sybase IQ also differs from merchant databases in that it is a column-based relational database rather than a row-based relational database. The latter is required for transaction processing where individual records (rows) are constantly being inserted into the database and updated. Conversely, column-based databases have significant advantages when it comes to query processing because each column is, effectively, an index but without any of the overheads associated with defining and storing those indexes. Moreover, every column is indexed in this sense, something that would never normally be possible when using a row-based approach. That said, Sybase also supports a number of index types that you can optionally implement: these are discussed later.

Another major advantage of a column-based approach is simply the amount of data that needs to be read for each query. Whenever you access data for a query from a conventional database, you read each row in its entirety, regardless of the actual fields that you are interested in for that specific query. In practice, this might mean reading a 3000 byte record to retrieve just 20 characters of data but by

reading data on a columnar basis, you only have to read what is specifically needed for the query at hand. Of course the difference in performance when you are reading a single record will be negligible, but many queries require full table scans. Multiply that single read by a few million rows per table and the performance difference is very significant.

A further consequence of using a column-based approach is that you typically do not employ conventional horizontal partitioning, which is predicated upon a row-based approach. Instead, Sybase IQ implements vertical partitioning: partitioning by column rather than by row. One of the advantages of this approach is that partitions can never become unbalanced, since there will always be the same number of fields in each column of a table. This significantly reduces the maintenance requirement of managing partitions and should eliminate the database re-organisation that may become necessary when conventional partitions become unbalanced and start to impair performance.

In addition, columns are easy to compress because you can have different algorithms for different datatypes. As a result, Sybase has had an historic advantage over the merchant databases in this area. This has been whittled away now, though Sybase IQ still has advantages. In particular, Sybase claims that a Sybase IQ data warehouse will never exceed the size of the raw data: this is by no means the case with its merchant rivals.

Physical architecture

The architecture of Sybase IQ is illustrated in Figure 1. It is a shared everything massively parallel architecture, with every node connected to every other node in a full mesh provided by the interconnect. This reduces I/O issues and improves resilience. The only exception to the shared everything approach is that each node can have its own, local, temporary storage. The big advantage of offering shared everything, and shared disks in particular, is that you do not have to distribute your data across the various disks in use, thereby removing what can be a significant administrative headache.

Each node in the Sybase IQ environment is designated as either a read/write node or a read only node. In the case of the former, each node can be flexibly designated as a read or a write node, as required. Thus, if you are running a large overnight batch update you might

The product

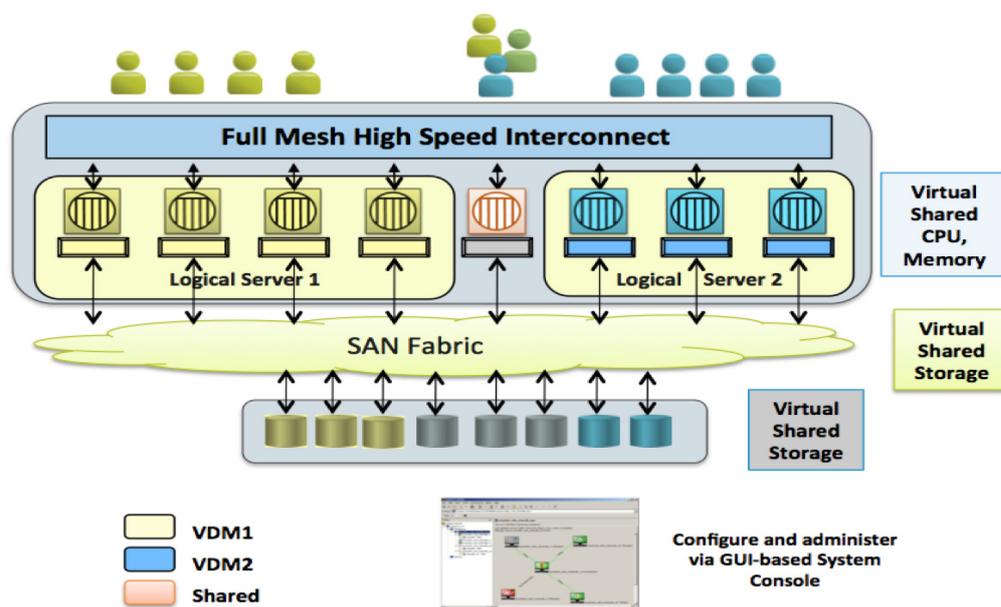


Figure 1: Sybase IQ 15.3 PlexQ architecture

want all of your read/write nodes to operate as write nodes, but have them running as read nodes during the day. In addition, you can add new nodes as needed, dynamically, so that you can scale up incrementally.

Nodes (servers) can be linked into a logical server, as shown. In addition, one logical server can “loan” nodes to other logical servers on a scheduled basis, for example to support overnight batch loading.

This approach to logical servers supports mixed query workloads because you can assign particular queries to a logical server and that query can then only use the resources available to that logical server. How many logical servers, and the number of nodes within each group, is entirely up to you. A graphical administration tool (see later) is provided to support the creation of these logical groupings, add or remove nodes, designate read only or read/write nodes and so on.

Each logical server can have its own logins so that specific users, departments or queries can always be directed to a designated logical server.

When a query is received the receiving node is designated as the “leader” node and, if the optimiser (see below) decides that this query can benefit from distribution, then the other nodes to be involved in processing this query are designated as “worker” nodes. Any node

can be a leader or worker but only one node can be a leader for any particular query.

There are a number of functions specifically to support high speed loading. The first is that the product supports pipeline parallelism so that, where indexes have separate data structures (which apply to Word and High Group indexes—see later), these can be updated at the same time as data is being loaded.

It was historically the case that, in many environments, the data was loaded to the server and then the database administrator pushed the data to the warehouse. However, this means that the DBA can see the data, which is not acceptable in many environments (for example, if the warehouse is outsourced), and many clients want to load the data directly. In order to support this, Sybase supports a “load from client” option that supports the loading of both data and LOBs (large Objects) via DBLib.

Another major feature is support for information lifecycle management (ILM). There are specific features to support different forms of storage for archival purposes: adding near-line and historical storage capabilities to active data. These can be designated as read-only stores for compliance purposes, if required, and, similarly, you can apply different security policies to each store. There are specific facilities provided to support time-based retention periods with data first being marked as read-only and then dropped.

The product

Another point to note about this architecture is the advantage that it offers to data aggregators and resellers (which represent a target market for Sybase), because it means that each subscriber can have its own read and read/write nodes, separate from anyone else, which obviously has beneficial security as well as chargeback implications; and it also allows you to define different service levels for different users. Further, Sybase IQ allows you to encrypt data on a column-by-column basis, which further reinforces this message. In fact, the product supports three levels of encryption: RSA and strong encryption (ECC) for data in flight plus RSA and RSA with FIPS 140-2 strong encryption for data at rest.

Should any node fail, you can switch users or responsibilities to another node. Hot standby, failover and load-balancing are possible across nodes. These functions are not automated but are under the DBA's control, which allows the DBA to dynamically allocate resources based upon business needs. In addition, there is an OpenSwitch load balancing application available, if required, that operates at the application server level. Sybase IQ InfoPrimer also has load-balancing capabilities for Sybase IQ data loading tasks. It is further worth noting the company's partnerships with a number of storage hardware vendors to further ensure high availability and disaster recovery. There is support for range partitioning and you can partition, re-partition, join, rename, split and drop table partitions as required.

There is also a NonStopIQ HA-DR methodology, which typically employs a local SAN and a remote one, with either synchronous or asynchronous communications between them. The big advantage of this is not just that it provides disaster recovery but also that it eliminates the need to take the system down—even for planned outages. Note that, as more and more companies adopt operational BI and embed query capability into operational applications, then the warehouse increasingly becomes as mission-critical as those applications, for which you need a solution such as this.

There is also support for query federation. This is intended for environments where operational data cannot be moved or copied from their source systems for compliance or other reasons but which you may want to include in queries or reports. In this sort of environment, the amount of data to be sourced from the operational system is typically small so near

real-time support can be achieved. The query federation technology supports Sybase ASE, Oracle and Microsoft SQL Server.

One element of the physical architecture that is not illustrated above is that the product now has a built-in web server to enable participation in web services.

Query performance

The Sybase IQ optimiser has been significantly enhanced in this release to take advantage of the newly introduced MPP-based capabilities of the product. Previously, the optimiser only had to consider the degree of parallelism that would be useful within a node. However, with a massively parallel architecture you have the ability to distribute queries across nodes (within a logical server if you have one). So the first thing that the optimiser does is to determine whether the query will benefit from this sort of parallelism. Not all queries do. For example, if the query is I/O bound then extra processing capacity may make an insignificant impact on query performance. Similarly, the overhead involved in distributing a query may be deleterious for, say, a short running query. As another example, if a query cannot make use of all the resources available on single server then, again, it will probably not make sense to distribute the query. The optimiser may decide that no part of a particular query should be distributed, that the whole query can usefully be distributed or that a part or parts (fragments) can be distributed.

Within each node that is executing a query, threads are allocated dynamically with threads added or removed as the query executes. Threads are scaled up or down according to workload and resource availability. As noted previously, physical servers can be dynamically allocated to a logical grouping.

Other notable features include the ability to determine how much parallelism to apply to which tasks and sub-query correlations, optimised use of temp space, operations that can run directly off compressed data without requiring decompression, and concurrent workload management.

Indexes

Although every column is, in effect, its own index, there are substantial advantages to using specific indexes in a number of situations.

The product

This is one area where Sybase has a major advantage over appliance vendors. Indeed, the secret of Sybase IQ is its indexing capabilities. As Sybase customers discover new needs for analysis, Sybase can simply create new index types to meet those needs. The beauty of this approach is that new indexes can be added to the data warehouse with little, if any, impact on the data warehouse architecture or the analytical applications using the warehouse. Sybase IQ offers a number of different indexing techniques:

- **Low Fast indexes:** these are low cardinality indexes (typically used for fields that have less than 1,500 unique values) that use a process known as tokenisation. Using this process, non-integer data is converted into a token (an integer; an existing integer becomes its own token) and then the tokens are stored rather than the data. This is particularly useful for reducing the quantity of redundant data and saving on disk space. Once the tokens are established (an automated process), a bitmapped index is created to reference these tokens.
- **Bit-Wise indexes:** for high cardinality fields, where the number of possible values exceeds 1,500 (for example, monetary values) Sybase IQ uses a patented technology known as Bit-Wise indexing. This is particularly useful where you want to combine calculations with range searches, for example to find the total revenue and number of units sold where the price was less than £50.
- **High Group indexes:** these are, in fact, B-trees. However, the principle here is that the user only defines these indexes when several columns are likely to be used in a group, in particular to combine low and high-cardinality searches. An example here might be an inquiry about product item sales and value (high cardinality) by store (low cardinality). High Group indices are now multi-threaded.
- **Fast Projection indexes:** the default "index" is simply the column store itself. If a user always plans to retrieve an entire column of data, then the fact that storage is columnar means the column can be projected into a report or inquiry without having to explicitly define any index at all. This is useful, for example, in WHERE clauses. The capabilities of this index type have been extended in the latest release through the support for

3-byte sizes that enable addressability for 16 million unique values, which will extend performance and compression for large data sets.

- **Text indexes:** these support full text search. The text index stores complete positional information for every instance of every term in the indexed columns. Some of the functions that are possible with the text index are discussed in the next section.
- **Compare indexes:** this indexing technique allows data column comparisons that are effectively equivalent to an "if ... then ... else" statement. For example, "if expenses are greater than revenue, then ...". This type of index is particularly useful for real-time comparisons in web applications.
- **Join indexes:** as the name implies, these are designed to obviate the need for table joins. Like a number of the supported indexes, these will be most useful when query requirements can be predicted in advance. Significant performance enhancements have been made to join indexes in this release with query parallelism allowing you to scan columns in parallel and then join in parallel.
- **Time Analytic indexes:** these offer the option to create indexes based on a date, time, or date and time. It should be noted that time-based queries tend to be particularly difficult for conventional relational databases to handle.

A number of extended facilities are supported to allow the use of these indexes in a variety of circumstances. These include index compression to reduce disk (or memory: bitmaps may be cached) requirements, the ability to use different types of index in combination, pipeline parallelism for GROUP BY and ORDER BY as well as hash and merge joins, and the facility to filter bit arrays using Boolean operators such as AND and OR. These features mean that the indexing in Sybase IQ overcomes a number of the traditional drawbacks of bitmapping, namely, that it is not suitable for joining tables or aggregating data. While on this topic it is also worth noting that while Sybase is fast enough that you do not need to pre-aggregate data for OLAP-based processing (which is a significant advantage in administrative terms) Sybase IQ does support OLAP capabilities with features such as rankings, partition windows,

The product

percentiles and averaging. It is also noteworthy that Sybase IQ includes an Index Advisor that will advise administrators as to when it would be useful to add a new index and of what type.

Other features provided to improve performance include predicate pushdown and sub-query optimisations (correlated sub-queries, sub-query disjunction and automatic query flattening).

In-database analytics

In conventional environments, data mining functions have to be performed outside the database: the data is extracted from the warehouse and then processed by the relevant software in a conventional manner. The problems with this approach are twofold. First, performance is degraded because there is an extraction process and because the application server will not have the same sort of parallel capabilities as provided in the analytics server. In order to offset this performance loss, analysts performing data mining will typically only sample the data. While this can offset performance consequences to some degree, it means that accuracy is sacrificed. In particular, population outliers can easily be missed.

In-database analytics resolves both of these problems: because mining algorithms are run inside the database you are not limited to sampling the data, thereby maximising accuracy and, for the same reason, you can take advantage of parallelism and other characteristics of the analytics server in order to optimise performance.

Sybase has implemented in-database analytics by means of user-defined functions (UDFs) that are executed as SQL functions and which can be designed using the WorkSpace Database Analytics IDE (see next). However, the company is not making this capability open to customer-created UDFs, but is, at least currently, limiting the ability to define these UDFs to certified partners. The first of these is Fuzzy Logix, whose DB Lytix for Sybase IQ is available (on Linux, Solaris, Windows and AIX), with support for algorithms including neural networks, k-means clustering, Monte Carlo simulation, linear and logistic regressions, and so on. Details can be found on the Sybase website at <http://www.sybase.com/detail?id=1065214>.

Going beyond structured analytics, Sybase supports significant text analytics capabilities, which also operate in-database. These include functions that include proximity search, support for Boolean operators, searching for phrases as well as terms, and scoring both within and across documents. There are also new plug-in libraries for multi-media analysis.

Workspace

The Sybase WorkSpace Database Analytic IDE supports the broad Sybase offering but in terms of Sybase IQ there are three specific components:

- PowerDesigner, which provides data modelling, reverse engineering to support migrations, data movement (ETL/ELT and replication), modelling, and lifecycle modelling to support the ILM capabilities that are in Sybase IQ. PowerDesigner integrates with WorkSpace via an Eclipse plug-in.
- DB Developer for Sybase IQ, which offers a customised development environment for Sybase IQ, including the UDFs already discussed.
- Open Source BIRT, which is a part of the Eclipse framework specifically intended for the development of reports, charts and so on.

Database operations

Sybase IQ includes a SQL API that allows SQL-based access. This is SQL-99 compliant and is the same SQL that is used in Sybase Adaptive Server Anywhere and (with a few exceptions) is also compatible with the syntax employed in Sybase ASE (that is, T-SQL) so that Sybase IQ can natively use most Sybase ASE stored procedures. In this context it is also worth noting that both Sybase IQ and Sybase ASE have the same look and feel. Within the product, Sybase IQ includes a graphical SQL Editor.

Sybase IQ also supports both ODBC and JDBC (2.0) call-level interfaces. Alternatively, Sybase IQ also provides Java 2 capability, and this language can be used for writing stored procedures and for creating user-defined functions. However, Java objects are not supported in the database.

The product

There is also support for XML, including the ability to store and retrieve XML documents as well as the ability to export query results in XML format (with an embedded DTD). In addition, it is important to appreciate the web services functionality that is available in Sybase IQ. There is an HTTP(S) web server built directly into the database, which supports the retrieval of data in XML format as well as via standards such as SOAP. There is also direct integration with Microsoft Visual Studio .NET via an ADO.NET provider.

Database administration

Sybase IQ supports conventional relational schemas, including the normalised schemas used for transaction processing as well as the star, snowflake, and constellation (a collection of stars) schemas that are used in data warehousing. In addition, the product also supports flat schemas (known as Rcubes) that have a number of advantages such as fewer tables (therefore reducing the number of joins and thereby improving performance), reduced complexity and greater manageability.

As far as the actual process of administration is concerned, this is provided through Sybase Central, illustrated in Figure 2, here showing the topology of the environment.

Sybase Central supports 'one-click' cluster management; in-flight maintenance operations (including adding columns on the fly); and graphical displays for CPU, thread utilisation and timing, in order to support problem resolution. This is also where you define and manage both resource management and security. A web-enabled companion to Sybase Central is the Sybase Control Center for Sybase IQ that helps monitor the Sybase IQ infrastructure remotely using web interfaces. An extensive library of system metrics along with historical metrics records can be displayed in a rich graphical format for ease of use. This is to be extended to incorporate the functionality of Sybase Central but for the time being both products remain available.

In terms of security the product provides support for Kerberos authentication, which enables common user IDs and passwords (user settable) to be used across the database and operating system environments. LDAP is not directly supported at this time. Further, role-based security not only applies to users but also administrators, so that you can have different roles for system monitoring, login and permission management, backup and restore administration, and multiplex grid management. This segregation of duties is important where you have multi-domain data warehouses.

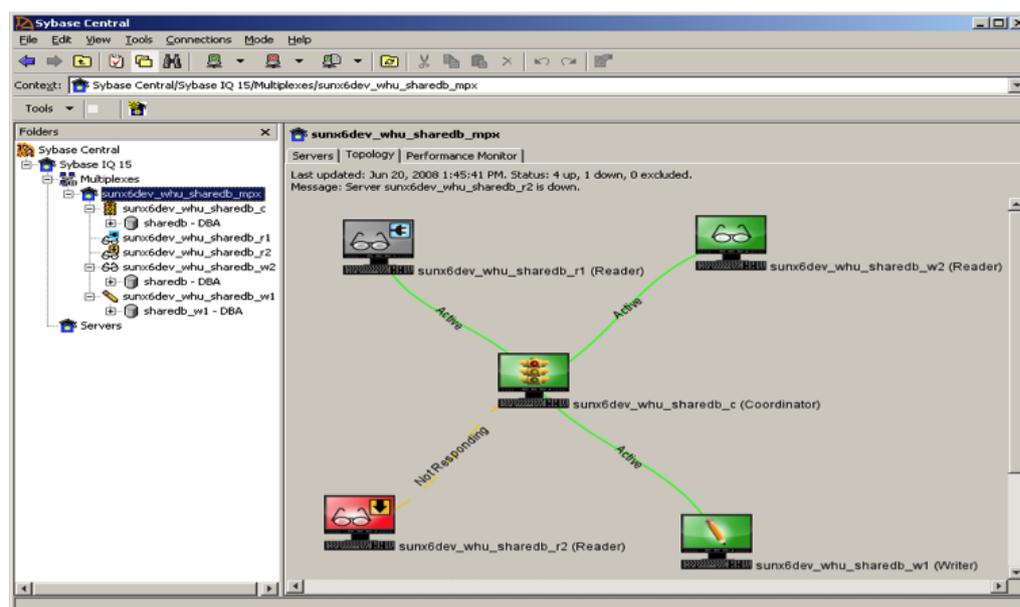


Figure 2: Sybase Central

The vendor

Sybase IQ is based on technology that Sybase (which is an SAP company) acquired when it purchased Expressway in 1995. It has always had an emphasis on scaling incrementally all the way up to very large data warehouses (VLDW). In this context it is worth noting Sybase's 2007 implementation (using Sybase IQ 12.x) of a reference one petabyte (1,000 TB) VLDW in conjunction with Sun (now Oracle) and BMMsoft, which was the first independently audited, petabyte scale warehouse.

Sybase now focuses on Sybase IQ as an analytic server especially for analytics services, advanced analytics and fast reporting, as discussed previously, and this approach has clearly paid off. There are now more than 3,500 Sybase IQ installations worldwide in more than 2,000 organisations. Further, Sybase IQ revenues have been rising steadily—so this strategy is clearly working. It is also particularly pertinent to note that a significant percentage of the product's sales have been to organisations that do not use Sybase ASE (the company's flagship transactional database).

In addition to the markets identified at the top of this paper, Sybase is now actively working in a number of interesting emerging areas where the company expects Sybase IQ's combination of features to offer significant value, including:

- **Capital Markets Risk Management:** Sybase RAP – The Trading Edition is a product that leverages Sybase IQ as a VLDW engine along with the company's complex event processing engine, which is capable of combining massive volumes of both real-time and historical data to provide a holistic view of the market needed by traders, portfolio managers and corporate risk officers.

- **Compliance:** with the retention and reporting of both structured and unstructured data. This often means the need to keep very large and increasing volumes of data on-line (which you might also want to do for analytic purposes) for which Sybase IQ is well suited. Note that for capital markets, compliance has a wider connotation, which can be met through the Risk Analytics Platform (RAP). Sybase IQ is itself compliant to various regulations pertaining to security (as previously discussed) and accessibility; such as 508 compliance—which relates to disabilities—for both the user interface and documentation. A further compliance capability is that you can lock down selected IQ databases (time-stamped) into read-only hardware.
- **Data/text mining:** Sybase is working with partners to provide solutions for the packaged data and text mining markets. In particular, you can cross-correlate relational data stored in IQ tables with non-relational data stored as a LOB. In this context, it is worth noting that Sybase IQ has extended facilities for supporting complex analytics that involve CLOBs (character large objects) as well as BLOBs (binary large objects) and XML. In particular, the product supports CLOB indexing capabilities to support text searching.

Sybase has entered into a number of partnerships focused on Sybase IQ with vendors that include specialists in hardware, storage, data quality, business intelligence and other areas, as well as various VARs and system integrators. Business Objects, as a fellow subsidiary of SAP, is a close partner.

Sybase IQ Web address: www.sybase.com/bi.

Summary

We used to believe that Sybase IQ faced two hurdles: its column-based technology (which was not well understood by the market) and disbelief that it could genuinely offer the performance advantages it claims. However, we do not think that either of these are obstacles anymore. In the first case, Sybase has proved the column-based concept with repeated success in many leading companies. Secondly, both Sybase, as well as the appliance vendors, have made it quite clear that better analytics performance is truly possible, whetting the appetites of companies who still struggle with traditional relational databases in query-intensive applications.

From Bloor Research's viewpoint the column-based approach advocated by Sybase will provide substantially better performance at lower cost than traditional approaches for analytical, reporting and data warehousing environments. Furthermore, it provides considerably more flexibility than can be provided by appliance vendors. We therefore see no obstacles to its continued success. In particular, it has features, such as text analytics, that are not in rival products and with the product's newly introduced massively parallel capabilities it should be able to match, if not surpass, competitors from a performance perspective.

We have previously advocated that the product should be targeted more widely and we are pleased to see that the company is now (since the takeover by SAP) following our advice. Certainly Sybase IQ has strengths in particular areas but its overall capabilities allow it to compete across the board.

Further Information

Further information about this subject is available from <http://www.BloorResearch.com/update/2092>

Bloor Research overview

Bloor Research is one of Europe's leading IT research, analysis and consultancy organisations. We explain how to bring greater Agility to corporate IT systems through the effective governance, management and leverage of Information. We have built a reputation for 'telling the right story' with independent, intelligent, well-articulated communications content and publications on all aspects of the ICT industry. We believe the objective of telling the right story is to:

- Describe the technology in context to its business value and the other systems and processes it interacts with.
- Understand how new and innovative technologies fit in with existing ICT investments.
- Look at the whole market and explain all the solutions available and how they can be more effectively evaluated.
- Filter "noise" and make it easier to find the additional information or news that supports both investment and implementation.
- Ensure all our content is available through the most appropriate channel.

Founded in 1989, we have spent over two decades distributing research and analysis to IT user and vendor organisations throughout the world via online subscriptions, tailored research services, events and consultancy projects. We are committed to turning our knowledge into business value for you.

About the author

Philip Howard

Research Director - Data

Philip started in the computer industry way back in 1973 and has variously worked as a systems analyst, programmer and salesperson, as well as in marketing and product management, for a variety of companies including GEC Marconi, GPT, Philips Data Systems, Raytheon and NCR.



After a quarter of a century of not being his own boss Philip set up what is now P3ST (Wordsmiths) Ltd in 1992 and his first client was Bloor Research (then ButlerBloor), with Philip working for the company as an associate analyst. His relationship with Bloor Research has continued since that time and he is now Research Director. His practice area encompasses anything to do with data and content and he has five further analysts working with him in this area. While maintaining an overview of the whole space Philip himself specialises in databases, data management, data integration, data quality, data federation, master data management, data governance and data warehousing. He also has an interest in event stream/complex event processing.

In addition to the numerous reports Philip has written on behalf of Bloor Research, Philip also contributes regularly to www.IT-Director.com and www.IT-Analysis.com and was previously the editor of both "Application Development News" and "Operating System News" on behalf of Cambridge Market Intelligence (CMI). He has also contributed to various magazines and published a number of reports published by companies such as CMI and The Financial Times.

Away from work, Philip's primary leisure activities are canal boats, skiing, playing Bridge (at which he is a Life Master) and walking the dog.

Copyright & disclaimer

This document is copyright © 2011 Bloor Research. No part of this publication may be reproduced by any method whatsoever without the prior consent of Bloor Research.

Due to the nature of this material, numerous hardware and software products have been mentioned by name. In the majority, if not all, of the cases, these product names are claimed as trademarks by the companies that manufacture the products. It is not Bloor Research's intent to claim these names or trademarks as our own. Likewise, company logos, graphics or screen shots have been reproduced with the consent of the owner and are subject to that owner's copyright.

Whilst every care has been taken in the preparation of this document to ensure that the information is correct, the publishers cannot accept responsibility for any errors or omissions.



2nd Floor,
145-157 St John Street
LONDON,
EC1V 4PY, United Kingdom

Tel: +44 (0)207 043 9750
Fax: +44 (0)207 043 9748
Web: www.BloorResearch.com
email: info@BloorResearch.com