

# InDetail



## Sybase IQ 15.3

Отчет InDetail компании Bloor Research  
Автор: Филип Говард  
Июнь 2011

...мы считаем, что Sybase IQ заслуживает самого пристального рассмотрения со стороны организаций, планирующих строить хранилища данных и быстродействующие аналитические системы.

Филип Говард

## Резюме для руководителя

### Некоторые факты

Sybase IQ — это специализированная реляционная СУБД с поколоночным хранением, рассчитанная на применение в приложениях бизнес-аналитики и других средствах аналитической обработки данных. Ее применение в качестве платформы для хранилища данных может дать ряд существенных преимуществ по сравнению с традиционными СУБД, в том числе в части производительности, масштабируемости и совокупной стоимости владения.

Хотя область применения продукта весьма широка, есть ряд задач, для решения которых он предназначен в первую очередь. К ним относятся:

- организация хранилищ данных для компаний-агрегаторов, типичным примером которых являются операторы услуг доступа к базам данных по большой совокупности клиентов, предлагающие также аналитические сервисы для работы с этими данными;
- реализация передовых методов аналитической обработки с повышенными требованиями к поддержке сложных и незапланированных запросов. Здесь Sybase IQ может служить как платформой для хранилища, так и аналитическим акселератором, функционально дополняющим уже существующее хранилище;
- ускорение подготовки ресурсоемких отчетов путем их переноса из оперативных систем или централизованных хранилищ данных в тех случаях, когда скорость генерации не соответствует требованиям бизнеса;
- обработка больших массивов неструктурированных данных (например, первичных документов и фотоснимков по страховым случаям), а также хранение очень больших массивов данных с обеспечением оперативного доступа к ним и выполнением запросов.

Учитывая тот факт, что Sybase входит в состав SAP, следует отметить, что в целом продукт рассчитан на применение в информационных средах с неоднородным составом программного обеспечения либо состоящих не из прикладных средств фирмы SAP. Для тех инфраструктур, прикладная часть которых строится в основном на ПО SAP, соответствующие задачи предлагается решать с помощью SAP BW, а также средства SAP по управлению данными в оперативной памяти под названием HANA.

Преимущества Sybase IQ обусловлены примененным в этой СУБД принципом организации данных по столбцам, масштабируемой кластерной архитектурой и на порядки более высокой производительностью, чем в базах данных со строковой организацией. При этом Sybase IQ требует меньших вычислительных мощностей. Все эти преимущества особенно ярко проявляются при обработке сложных запросов и сканировании больших таблиц. В последнем случае отсутствует необходимость в обязательной предварительной агрегации данных, что, по сравнению с традиционными подходами, дает увеличение производительности и экономию ресурсов на операциях

обслуживания базы. Сравнительно меньший размер хранилищ на Sybase IQ, наряду с другими преимуществами продукта, придает Sybase IQ огромный потенциал в отношении производительности при росте числа обслуживаемых пользователей.

Короче говоря, Sybase IQ обеспечивает значительно более высокое быстродействие при меньшей совокупной стоимости владения. При этом, если не брать в расчет поколоночный принцип хранения данных, в остальном Sybase IQ выглядит аналогично традиционным реляционным СУБД — используются стандартный язык SQL, стандартные аппаратные средства и операционные системы. Схемы баз данных, как и приложения, при внедрении IQ могут оставаться без изменений. В освоении продукт не сложнее обычных СУБД — новый столбец добавляется в таблицу столь же легко, как и новая строка и т. д.

### Параллельная обработка в Sybase IQ

Возможность распараллеливания вычислений особенно важна для скоростной обработки запросов в больших и сложных информационных системах. В 15-м выпуске Sybase IQ была реализована возможность работы нескольких записывающих узлов в конфигурации «мультиплекс» (ранее поддерживалась совместная работа лишь считывающих узлов) наряду с распараллеливанием запросов и параллелизацией конвейера — это ускоряет как обработку запросов, так и загрузку данных. В текущем выпуске возможности параллельной обработки расширены за счет реализации архитектуры массово-параллельных вычислений PlexQ. Ранее система позволяла распараллеливать обработку запроса в рамках узла; на разных узлах могли исполняться одновременно лишь разные запросы. Теперь же появилась возможность распределять обработку одного запроса по нескольким узлам. Это существенно ускоряет обработку определенных типов запросов.

### Основные выводы

По мнению Bloor Research, потенциальным пользователям продукта следует знать о следующих ключевых моментах.

Наряду с поколоночной организацией данных в Sybase IQ применен ряд специализированных алгоритмов индексирования, которые предназначены для дополнительного повышения производительности при обработке произвольных запросов. К ним относятся алгоритмы индексирования: данных с низкой мощностью связей (что позволяет еще снизить требования к памяти и ускорить обработку запросов благодаря хранению меток вместо данных); данных, по которым осуществляется группировка; данных, ограниченных диапазоном; без соединения таблиц; для сравнения в режиме реального времени в веб-приложениях; для анализа данных типа «дата» и «время». Предусмотрены индексы для текстового анализа (обеспечивающие обработку неструктурированных данных, при необходимости совместно со структурированными). В текущем выпуске

## Резюме для руководителя

эти возможности расширены благодаря реализации внутрибазового текстового поиска и внутрибазовой аналитической обработки.

- Sybase IQ обеспечивает возможности как многопоточной обработки, так и высокой готовности в круглосуточном режиме (для чего Sybase имеет партнерские соглашения с поставщиками систем хранения). Принцип разделения вычислительных узлов по функциям чтения и записи позволяет выполнять соответствующие операции параллельно, без взаимного влияния. Использование считывающих узлов особенно полезно для компаний — агрегаторов данных, предоставляющих доступ к аналитическим сервисам большому числу подписчиков, так как позволяет выделять клиентам отдельные узлы и легко учитывать используемые ресурсы.
- Возможности массово-параллельной архитектуры PlexQ дополняются встроенным оптимизатором. В данном выпуске Sybase IQ оптимизатор усовершенствован и позволяет определять, ускорится ли обработка запроса от распределения, и каким именно узлам лучше всего ее поручить.
- Sybase IQ значительно превосходит обычные СУБД в отношении производительности. Помимо перечисленных возможностей, он поддерживает плоские схемы (R-кубы), в ряде случаев обеспечивающие весомые преимущества в сравнении с традиционными схемами типа «звезда». Так, применение R-кубов может способствовать значительному ускорению процесса построения хранилища данных, повышению его производительности и гибкости. Кроме того, система позволяет на ходу вносить изменения в структуру таблиц: можно добавлять и удалять столбцы во время работы базы данных.
- Sybase IQ обеспечивает шифрование как на уровне столбцов, так и базы данных в целом. Это особенно важно для агрегаторов данных, когда данные разных подписчиков требуют шифрования по разным алгоритмам. Поддерживается шифрование как хранимых, так и обрабатываемых в текущий момент данных.
- Внутрибазовая аналитическая обработка обеспечивает гораздо более высокое быстродействие, нежели традиционные механизмы. Она реализована в форме определяемых пользователем функций (UDF), которые при обработке рассматриваются как подмножество SQL-функций Sybase IQ. При этом процессам аналитической обработки доступны все возможности оптимизатора Sybase IQ и средства распараллеливания задач. Для реализации этих функций в IQ компания Sybase заключила партнерское соглашение с Fuzzy Logix (запланированы альянсы и с другими производителями библиотек), а также, для расширения возможностей Sybase RAP (в состав которого входит IQ), с Visual Numerics.
- Взаимодействие с сервером Sybase IQ осуществляется

посредством стандартных соединений ODBC/JDBC/OLE-DB, что позволяет применять на прикладном уровне любые средства бизнес-аналитики со стандартными протоколами обмена. Sybase IQ сертифицирован для работы с большинством ведущих аналитических средств таких производителей, как SAP BusinessObjects, Cognos, MicroStrategy, QlikView, iDashboards, SAS, SPSS, KXEN и др. Так как Sybase входит в состав SAP, лучше всего Sybase IQ приспособлен для работы с SAP BusinessObjects.

- Для Sybase IQ в виде дополнительного продукта предлагаются оптимизированные средства извлечения, преобразования и загрузки данных (ETL и ELT), позволяющие разработчикам быстро создавать и развертывать подлежащие анализу наборы данных. Sybase IQ сертифицирован для работы с ведущими в отрасли ETL-инструментами и особо адаптирован к ETL-средствам SAP BusinessObjects Data Services. Продукт поддерживает загрузку данных непосредственно из клиента в обход администратора базы данных, что исключает необходимость участия последнего в процессе загрузки. Это важно для таких систем, где по соображениям безопасности администраторы не должны иметь доступа к информации.
- На случай, если данные оперативных систем не могут быть загружены в хранилище (согласно нормативным требованиям или по другим причинам), Sybase IQ поддерживает федерацию запросов с данными, хранимыми в системах Sybase ASE, Oracle, MySQL, SQL Anywhere и SQL Server. Загрузка в Sybase IQ также может осуществляться в режиме почти реального времени с использованием Sybase Replication Server (поддерживающего загрузку в реальном времени) и набора скриптов, генерируемых с помощью PowerDesigner. В систему можно загружать событийные данные в реальном времени через Sybase ESP (Event Streaming Platform). Возможность выполнять загрузку не прерывая обработки запросов реализуется благодаря существующей в Sybase IQ функции версионности — для загрузки создается новая версия данных, а запросы обрабатываются применительно к старой версии до момента завершения загрузки.
- Начиная с 15-го выпуска в Sybase IQ реализован механизм управления жизненным циклом информации, предусматривающий перемещение данных из хранилища с оперативным доступом в промежуточное хранилище, а затем в архив. Передовое средство моделирования метаданных и управления ими PowerDesigner позволяет определять формальные правила хранения данных, а среда разработки на базе Eclipse «WorkSpace Data Analytics» поддерживает создание баз данных, средств анализа и отчетности.

### Резюме

В отличие от стандартных СУБД с строчным хранением данных, спроектированных в расчете на

## Резюме для руководителя

онлайновую обработку транзакций, Sybase IQ изначально рассчитывался на исполнение аналитических запросов и параметризованный анализ. Благодаря этому он обеспечивает значительные преимущества в части производительности и совокупной стоимости владения в условиях высокой нагрузки по запросам по сравнению с традиционными системами. Быстродействие в данном выпуске значительно увеличено благодаря применению архитектуры PlexQ.

Наиболее ярко названные преимущества проявляются в системах, где нагрузки по обработке запросов непредсказуемы, а количество произвольных запросов велико. Традиционные СУБД в принципе невозможно оптимизировать для обработки запросов, тип которых заранее не известен. Реализованный же в IQ поколоночный принцип хранения обеспечивает эффективную «самонастройку». Кроме того, СУБД с поколоночным хранением намного более эффективны при обработке сложных запросов, требующих выборки из множества таблиц по многим критериям и сканирования больших таблиц. Наконец, Sybase IQ обеспечивает хорошую масштабируемость, позволяющую строить большие хранилища, содержащие детальные транзакционные или субтранзакционные данные, например протоколы навигации посетителей веб-сайтов. Sybase IQ не требует предварительной агрегации данных для анализа, благодаря чему пользователи могут эффективно и быстро анализировать данные на атомарном уровне.

Лежащий в основе Sybase IQ принцип поколоночного хранения обеспечивает более высокую эффективность работы СУБД, поскольку в этом случае серверу для ответа на запрос приходится выполнять гораздо меньше действий. Конечно, в продукте применены и другие механизмы оптимизации, но главные преимущества IQ заключаются именно в указанном принципе, а также в более низкой совокупной стоимости владения. Последнее достигается за счет двух вещей. Во-первых, за счет снижения абсолютной стоимости информационной системы, обусловленного повышенной производительностью и применением специализированных алгоритмов сжатия по столбцам — это сокращает требования к дисковой памяти и, соответственно, объем капиталовложений в технические средства. Во-вторых, за счет упрощения администрирования и настройки, что значительно снижает затраты на сопровождение по сравнению с традиционными средствами организации хранилищ данных.

Сегодня на рынке хранилищ данных присутствуют не только поставщики традиционных решений. Высокий интерес к специализированным машинам хранилищ данных — готовым аппаратно-программным комплексам. С одной стороны, это обстоятельство помогает Sybase, поскольку положило конец безраздельному господству крупнейших основных поставщиков СУБД с построчной организацией и дало возможность заказчикам рассматривать другие варианты. Появление специализированных машин способствовало повышению интереса к Sybase IQ, в частности, как к средству ускорения генерации отчетов. С

другой стороны, интерес к этой категории информационной техники породил новый класс конкурентов, заявляющих о сопоставимом росте производительности и чрезвычайной простоте применения своих изделий.

Преимущество Sybase IQ перед машинами хранилищ данных заключается в более гибких возможностях настройки. Большинство производителей машин полностью отказываются от индексов или используют их крайне редко. Sybase IQ предлагает широкий выбор алгоритмов индексирования. Хотя при этом возрастает объем работ по администрированию, в то же время система значительно легче адаптируется к изменениям в связности и типах данных. Встроенный мастер индексирования (Index Adviser) облегчает процесс выбора и создания необходимых индексов. Кроме того, нам не известна ни одна машина-хранилище данных, которая поддерживала бы анализ текстов (как это делает Sybase IQ), да и возможности управления смешанной рабочей нагрузкой у аппаратно-программных систем, как правило, более ограничены. Еще одним преимуществом Sybase IQ является возможность легкого и практически линейного масштабирования в отношении как объемов данных, так и количества пользователей. При возрастании объемов данных достаточно добавлять диски в общий дисковый пул, а рост количества пользователей поддерживается путем инкрементального включения в кластер новых считывающих или записывающих узлов.

Отметим в заключение, что Sybase IQ с большим успехом конкурирует как с традиционными, так и со специализированными системами. В разных областях применения продукт обеспечивает разные преимущества, и потому мы считаем, что Sybase IQ заслуживает самого пристального рассмотрения со стороны организаций, планирующих строить хранилища данных и быстродействующие аналитические системы.

## Продукт

Текущая версия Sybase IQ — 15.3. Все сопутствующие продукты, в том числе Sybase InfoPrimer (ранее Sybase ETL), Sybase PowerDesigner, Sybase Replication Server и WorkSpace Database Analytics IDE, в полной мере используют возможности этой версии. Что касается Sybase InfoPrimer, то это средство было оптимизировано специально для использования в тандеме с Sybase IQ и в качестве приемника данных поддерживает только его.

Sybase IQ работает под управлением операционных систем Windows, Linux (Red Hat и SuSE) и основных UNIX-систем производства HP, IBM и Sun Microsystems (Oracle). Для разработки аналитических приложений поддерживаются такие языки программирования, как Perl, Python, PHP, ADO, Net, OLE-DB, Ruby on Rails и др.

### Логическая архитектура

СУБД Sybase IQ разработана специально для хранилищ данных. Не будучи рассчитана на обработку транзакций, она не содержит соответствующих специализированных функций. Благодаря этому продукт имеет более компактный код и менее сложен, чем универсальные транзакционные СУБД.

Еще одно отличие Sybase IQ от средств обработки бизнес-транзакций заключается в том, что эта СУБД, имея реляционную архитектуру, организует данные по столбцам, а не по строкам. Для обработки транзакций, когда в базу данных постоянно добавляются новые записи, оптимальна строковая организация. В задачах же обработки запросов выигрывают базы с построчным хранением, поскольку каждый столбец в сущности является индексом, при этом на задание и хранение индексов системные ресурсы не расходуются. Кроме того, проиндексированными оказываются являются абсолютно все столбцы всех таблиц базы, что невозможно при построчном хранении. Вместе с тем Sybase поддерживает и традиционные типы индексов, которые можно использовать при необходимости; мы рассмотрим их ниже.

Другим важным преимуществом построчной организации является существенно более малый объем данных, подлежащих считыванию при запросе. При обращении к данным в обычной БД требуется считывать записи целиком, вне зависимости от количества интересующих полей. На практике это может означать, например, что для получения 20 байт данных требуется загрузить 3-килобайтную запись. При построчной организации загружаются лишь столбцы, необходимые для выполнения конкретного запроса. Если речь идет о единственной записи, разница в быстродействии незначительна, однако при обработке разных запросов нередко требуется сканирование таблицы целиком. Если таблица содержит несколько миллионов записей, разница в быстродействии будет огромна.

Еще одна положительная сторона построчной организации — отсутствие в большинстве случаев необходимости прибегать к традиционному горизонтальному разбиению на партиции. В Sybase IQ

реализовано вертикальное разбиение — по столбцам, а не по строкам. Благодаря этому сбалансированность партиций не нарушается, ведь все столбцы таблицы содержат одно и то же число полей. Это значительно упрощает управление партициями и исключает необходимость в реорганизации БД, возникающую при дисбалансе партиций, которому сопутствует снижение производительности.

Кроме того, столбцы эффективно сжимаются благодаря возможности использовать разные алгоритмы компрессии для разных типов данных. Эта возможность была реализована в продукте Sybase давно, благодаря чему он имел преимущество перед транзакционными СУБД. По мере освоения технологии конкурентами указанное различие стирается, однако Sybase IQ все еще превосходит конкурирующие СУБД. В частности, Sybase заявляет, что объем хранилища данных Sybase IQ никогда не превысит объема исходных данных. Этого не позволяет добиться ни одна СУБД для обработки бизнес-транзакций.

### Физическая архитектура

Архитектурная схема Sybase IQ приведена на рис. 1. Система имеет массово-параллельную архитектуру с разделением всех ресурсов, где каждый узел соединен с каждым всесторонним скоростным межсоединением. Это повышает эффективность ввода-вывода и устойчивость системы. Единственным отличием от классических систем с разделением всех ресурсов является наличие у каждого узла собственного локального временного хранилища. Большим преимуществом принципа разделения всех ресурсов, и в частности дисков, является отсутствие необходимости распределять данные по разным дискам, что существенно упрощает администрирование.

Каждый узел (сервер) в системе Sybase IQ может быть работать либо в режиме чтения-записи, либо только чтения. Режимы работы узлов чтения-записи легко переключаются при необходимости. Таким образом, например, в ходе крупного пакетного обновления в ночные часы все узлы чтения/записи могут работать только на запись, а в дневное время — только на чтение. По мере необходимости систему можно постепенно наращивать, добавляя в нее новые узлы.

## Продукт



Рис. 1. Архитектура PlexQ в Sybase IQ 15.3

Узлы можно объединять в логические серверы, как показано на рис. 1. Логические серверы могут на плановой основе временно передавать свои узлы другим логическим серверам, например для обеспечения необходимых мощностей для пакетной загрузки в ночное время.

Такой способ организации вычислений повышает эффективность обработки смешанных нагрузок благодаря возможности отводить определенным типам запросов отдельные логические сервера. При этом обработка запросов будет осуществляться только на мощностях, принадлежащих данному логическому серверу. Общее число логических серверов и количество входящих в них узлов определяется администратором. Для управления логическими серверами — их создания, добавления и удаления узлов, назначения типов узлов («чтение» или «чтение-запись») — предусмотрено графическое средство администрирования, рассмотренное ниже.

На каждом логическом сервере можно заводить свои учетные записи, что позволяет отводить серверы в распоряжение определенных пользователей, отделов или для обработки запросов определенного типа.

Узел, на который первоначально поступает запрос, получает статус ведущего узла, и если оптимизатор (см. ниже) решает, что обработку данного запроса можно ускорить за счет распределения, то в процесс обработки вовлекаются другие узлы, называемые рабочими узлами. Любой узел может играть роль как ведущего, так и рабочего, однако для каждого запроса ведущим может быть лишь один узел.

Для обеспечения скоростной загрузки данных в системе предусмотрен ряд специальных функций. Во-первых, Sybase IQ поддерживает конвейерное распараллеливание,

так что если для индексов имеются самостоятельные структуры данных (это относится к индексам типа «word» и «high group» — см. далее), они могут обновляться одновременно с загрузкой данных.

Исторически сложилось, что во многих информационных системах, содержащих хранилище, данные загружаются на сервер, а затем администратор базы данных переносит их в хранилище. Однако в некоторых случаях доступ администратора к данным должен быть исключен (например, если сопровождением хранилища занимается сторонняя организация), поэтому многие заказчики требуют загружать данные непосредственно в хранилище. Для этого в Sybase IQ реализована возможность загрузки с клиента — данные, в том числе большие объекты (LOB) могут загружаться через DBLib.

Еще одна важная функция — поддержка управления жизненным циклом информации (ILM). Предусмотрена работа с разными ярусами хранения: кроме яруса активных данных, можно организовать полуоперативное и архивное хранилища. При необходимости, в целях соблюдения нормативных требований, этим ярусам может быть назначен режим «только чтение» и индивидуальные политики безопасности. Предусмотрены специальные средства хранения в течение установленного срока, когда вначале данные помечаются как доступные только для чтения, а затем удаляются.

Важно подчеркнуть еще одно преимущество этой архитектуры, которое представляет интерес для агрегаторов данных и торговых посредников (целевые для Sybase рынки): каждому подписчику могут быть выделены собственные узлы чтения и чтения/записи, не связанные ни с какими другими. Это повышает уровень



## Продукт

защищенности данных и облегчает учет машинного времени. Кроме того, это позволяет задавать для разных пользователей разные уровни обслуживания. Имеющаяся в Sybase IQ возможность шифровать данные на уровне отдельных столбцов дополнительно усиливает позиции продукта на названных рынках. В настоящее время система поддерживает три уровня шифрования: для защиты данных при передаче по сети используется алгоритм RSA и особо стойкий алгоритм ECC, а для криптозащиты внутри базы применяется RSA, а также RSA совместно с особо стойким алгоритмом FIPS 140-2.

Если какой-либо узел выходит из строя, его функции можно передать другому узлу. Система поддерживает горячее резервирование узлов, аварийное переключение и балансировку нагрузки. Эти функции не являются автоматическими; они управляются администратором базы данных, который имеет возможность динамически выделять ресурсы в соответствии с задачами. Кроме того, можно использовать специализированное средство балансировки нагрузки под названием OpenSwitch, которое при необходимости может работать на уровне сервера приложений. В Sybase IQ InfoPrimer также имеются функции распределения нагрузки по загрузке данных в IQ. Сотрудничество Sybase с производителями систем хранения данных дополнительно повышает готовность создаваемых на базе IQ информационных систем и их способность к аварийному восстановлению. Обеспечивается удобное разбиение данных по диапазонам, изменение структуры табличных разделов, их слияние, переименование, разделение и удаление.

Для обеспечения высокой готовности и аварийного восстановления в Sybase IQ предусмотрена технология под названием NonStopIQ. При этом задействуются, как правило, две сети хранения данных SAN — локальная и удаленная, связанные между собой синхронным или асинхронным подключением. Большим преимуществом такого решения является не только возможность аварийного восстановления, но и исключение необходимости останова работы системы — даже на время планового обслуживания. Стоит отметить, что по мере все более широкого внедрения средств оперативного бизнес-анализа и встраивания в рабочие приложения функций обработки запросов хранилища данных приобретают столь же важное значение, что и эти приложения, и возникает необходимость обеспечения их высокой готовности.

Предусмотрена поддержка федерации запросов. Эта функция предназначена для систем, где по соображениям соответствия нормативным требованиям или иным причинам оперативные данные не могут быть перемещены или скопированы из исходных баз, однако должны быть включены в запросы или отчеты. Объем таких данных, как правило, столь мал, что обращение к ним вне хранилища незначительно сказывается на быстродействии и выполняется почти в реальном времени. Механизм федерации запросов работает с СУБД Sybase ASE, Oracle и Microsoft SQL Server.

Еще одним элементом физической архитектуры, не

показанным на иллюстрации, является встроенный веб-сервер, обеспечивающий работу с веб-сервисами.

### Скорость обработки запросов

В данном выпуске Sybase IQ оптимизатор был существенно усовершенствован, с тем чтобы в полной мере использовать возможности массового параллелизма. Раньше оптимизатор должен был определять только степень распараллеливания в рамках одного узла. Массово-параллельная архитектура позволяет распределять обработку запросов между несколькими узлами (в рамках логического сервера при наличии такового). Поэтому вначале оптимизатор должен определить, выиграет ли запрос от такого распараллеливания. Не все запросы поддаются ускорению обработки путем распараллеливания. К примеру, если скорость обработки зависит в первую очередь от скорости ввода-вывода, то выделение дополнительных процессорных мощностей мало что даст. Если по своей природе запрос обрабатывается быстро, то накладные расходы на распределение могут лишь замедлить вычисления. Если для обработки запроса мощности одного сервера избыточны, то, очевидно, не имеет смысла ее распределять. Таким образом, оптимизатор в одних случаях приходит к выводу, что никакая часть текущего запроса не подлежит распределенной обработке, в других — что целесообразно распределить обработку запроса в целом, в третьих — что следует распределить лишь один или несколько компонентов запроса.

На каждом обрабатывающем запрос узле процессорные потоки выделяются динамически: они могут добавляться и удаляться в ходе обработки. Количество выделяемых потоков определяется текущей загрузкой, а также количеством доступных ресурсов. Как было отмечено выше, в логические группы могут по мере надобности динамически включаться новые физические серверы.

Заслуживают упоминания и такие возможности, как определение требуемой меры параллелизма для конкретных задач и связанных подзапросов, оптимизированное использование временного пространства, обработка сжатых данных без их распаковки, а также управление одновременными рабочими нагрузками.

### Индексы

Хотя каждый столбец в Sybase IQ в сущности является индексом, использование специальных индексов в ряде случаев обеспечивает существенные преимущества. Это как раз та область, где Sybase IQ значительно превосходит специализированные машины хранилищ данных. Несомненно, одним из важнейших преимуществ продукта являются широкие возможности индексирования. По мере появления у заказчиков новых аналитических задач Sybase добавляет в продукт требуемые типы индексов. Эффективность подхода состоит в том, что добавление новых индексов практически не влияет ни на архитектуру хранилища, ни на использующие его аналитические приложения. Sybase IQ обеспечивает целый ряд различных



## Продукт

способов индексирования:

- Индексы «low fast». Это индекс для данных с низкой мощностью связей (обычно используется для полей с менее чем 1 500 разных значений). Для его построения нецелочисленные данные преобразуются в метки (метка имеет целочисленный тип; целочисленное значение само становится меткой), которые затем сохраняются вместо данных. Это особенно полезно для сокращения объема избыточных данных и экономии дисковой памяти. После того как метки созданы (этот процесс выполняется автоматически), создается индекс — битовый массив, ссылающийся на эти метки.
- Индексы «bit-wise». Это битовый индекс для полей с высокой мощностью связей, где количество возможных значений превышает 1500 (например, денежные суммы). Применяемая в Sybase IQ запатентованная технология Bit-Wise-индексирования особенно полезна, когда необходимо сочетать вычисления с поиском по диапазонам, например нужно вычислить общую выручку и количество проданных товаров по цене менее 50 фунтов стерлингов.
- Индексы «high group». Фактически это классические В-деревья, которые создаются лишь в том случае, если предполагается группировать столбцы — например, для сочетания поиска по данным с низкой и высокой мощностью связей. Примером может служить запрос о количестве проданных единиц товара и вырученной сумме (высокая мощность связей) с группировкой по магазинам (низкая мощность связей). В новом выпуске Sybase IQ индексы «high group» обрабатываются в многопоточном режиме.
- Индексы «fast projection». Это «индексы» по умолчанию, роль которых играют сами столбцы. Если пользователь планирует всегда запрашивать столбец целиком, то, поскольку данные хранятся по колонкам, содержимое столбца может быть прочитано («спроецировано») в отчет или запрос без явного задания каких-либо индексов. Это удобно, например, в операторе WHERE. В последнем выпуске возможности индексирования этого типа расширены: теперь поддерживается 3-байтовый индекс, что позволяет определить 16,7 млн. уникальных значений. Это повышает производительность обработки и эффективность сжатия больших массивов данных.
- Индексы «text». Эти индексы поддерживают полнотекстовый поиск. В текстовом индексе хранится адрес каждого вхождения каждого выражения в проиндексированных столбцах. Некоторые из возможностей, обеспечиваемых текстовым индексированием, рассмотрены ниже.
- Индексы «compare». Эта методика индексации позволяет выполнять сравнение столбцов данных, эквивалентное выражению «if-then-else». Например, «если расходы выше, чем выручка, то...». Такой тип индексов особенно полезен для выполнения операций сравнения в реальном времени в веб-приложениях.

- Индексы «join». Как следует из названия, эти индексы предназначены для устранения необходимости объединения таблиц. Как и многие другие индексы, данный тип наиболее полезен, когда требования к запросу известны заранее. В данном выпуске продукта скорость работы с индексами «join» существенно повышена благодаря параллельному выполнению запросов, что позволяет сканировать столбцы одновременно и затем параллельно проводить объединение.
- Индекс «time analytic» — на основе даты, времени или того и другого вместе. Следует подчеркнуть, что запросы с параметрами времени особенно трудно поддаются обработке средствами обычных реляционных СУБД.

Поддерживается ряд расширенных функций, позволяющих прибегать к индексированию в самых разных обстоятельствах. Это сжатие индексов для сокращения объема занимаемой дисковой памяти (и ОЗУ — битовые массивы могут кэшироваться в оперативной памяти), совместное использование индексов разных типов, конвейерное распараллеливание для операторов GROUP BY и ORDER BY, соединения хэшированием и слиянием, а также возможность фильтрации битовых массивов с помощью логических операторов, таких как AND и OR. При этом устраняется целый ряд недостатков, присущих индексированию с помощью битовых массивов, в частности появляется возможность его применения для операций объединения таблиц или агрегации данных. Стоит упомянуть также, что хотя Sybase IQ достаточно быстр, чтобы исключить предварительную агрегацию данных для OLAP (что является значительным преимуществом с точки зрения администрирования), вместе с тем продукт поддерживает OLAP-функции, такие как ранжирование, окна разбиения, процент или и вычисление среднего значения. Необходимо отметить также, что в Sybase IQ имеется встроенный мастер индексирования (Index Adviser), помогающий определять целесообразность добавления новых индексов и выбирать их типы.

К другим средствам повышения производительности относятся возможности спуска предикатов и оптимизации подзапросов (связывание, разделение подзапросов и автоматическое упрощение структуры запросов).

### Внутрибазовая аналитическая обработка

В традиционных системах функции информационной проходки (data mining) выполняются вне базы данных: данные извлекаются из хранилища, а затем обрабатываются соответствующими программами обычным способом. Такому подходу присущи два серьезных недостатка. Во-первых, уходит время на извлечение данных, во-вторых, сервер приложений значительно уступает аналитическому серверу в возможностях распараллеливания. Чтобы нивелировать потерю производительности, прибегают к выборке данных, из-за чего страдает точность. В частности, легко могут быть пропущены выбросы.

Внутрибазовая аналитическая обработка позволяет

## Продукт

решить обе проблемы. Поскольку алгоритмы информационной проходки выполняются внутри базы данных, нет необходимости ограничиваться выборкой, и точность не снижается. Кроме того, доступны все возможности параллельной работы и другие преимущества аналитического сервера, что повышает быстродействие.

Внутрибазовая аналитика в Sybase IQ реализуется в форме функций UDF, которые выполняются как функции SQL и могут создаваться в среде разработки WorkSpace Database Analytics (см. далее). Однако это, по крайней мере пока, не касается UDF, разработанных заказчиком — в настоящее время такие функции могут создаваться только сертифицированными партнерами. Первым из таких партнеров является компания Fuzzy Logic, чья СУБД Lytix для Sybase доступна для платформ Linux, Solaris, Windows и AIX. Lytix поддерживает различные алгоритмы, в том числе нейронные сети, кластеризацию k-средних, моделирование по методу Монте-Карло, линейные и логистические регрессии и т. д. Более подробные сведения см. на сайте Sybase по адресу <http://www.sybase.com/detail?id=1065214>.

### Рабочая среда WorkSpace

Интегрированная среда разработки Sybase WorkSpace Database Analytic IDE поддерживает широкий спектр продуктов Sybase, из которых для Sybase IQ первоочередную роль играют следующие:

- PowerDesigner — средство моделирования данных, обратного проектирования в целях обеспечения поддержки миграции, моделирования перемещения данных (ETL/ELT и репликации) и моделирования жизненного цикла для поддержки имеющихся в Sybase IQ возможностей ILM. Интеграция PowerDesigner с WorkSpace осуществляется через подключаемый модуль Eclipse.
- DB Developer для Sybase IQ — настраиваемая среда разработки, содержащая в том числе упомянутые функции UDF.
- Open Source BIRT — часть платформы Eclipse, автоматизирующая разработку отчетов, графиков и т. п.

### Работа с базой данных

В Sybase IQ входит программный интерфейс для SQL (SQL API). Применяемый в системе диалект языка SQL совместим со стандартом SQL-99. Он полностью аналогичен используемому в Sybase Adaptive Server Anywhere, а также совместим по синтаксису (за некоторыми исключениями) с языком Sybase ASE (T-SQL), благодаря чему Sybase IQ может естественным образом использовать большую часть хранимых процедур ASE. В связи с этим стоит упомянуть, что Sybase IQ и Sybase ASE имеют одинаковое оформление и сходные органы управления. В состав Sybase IQ входит графическая утилита SQL Editor.

Sybase IQ поддерживает также интерфейсы ODBC и JDBC

(2.0). В качестве альтернативы предлагается возможность использовать Java 2. На этом языке можно создавать хранимые процедуры и пользовательские функции. Однако размещение Java-объектов в базе данных не поддерживается.

Реализована также поддержка XML, в том числе возможность сохранять и извлекать XML-документы, а также экспортировать результаты запросов в формате XML (с внедренным определением DTD). Важно принять во внимание имеющуюся в Sybase IQ функциональность веб-сервисов. Непосредственно в СУБД встроены веб-сервер HTTP(S), позволяющий извлекать данные в формате XML, а также по таким протоколам, как SOAP. Обеспечивается прямая интеграция с Microsoft Visual Studio .NET через поставщик данных ADO.NET.

### Администрирование базы данных

Sybase IQ поддерживает традиционные реляционные схемы, в том числе нормализованные схемы, используемые для обработки транзакций, а также схемы типа «звезда», «снежинка» и «созвездие», применяемые в хранилищах данных. Кроме того, поддерживаются плоские схемы (известные как R-кубы), обладающие целым рядом преимуществ, к которым относится меньшее количество таблиц (в результате чего уменьшается число объединений и возрастает производительность), большая простота и лучшая управляемость.

Что касается собственно администрирования, то оно реализуется с помощью средства Sybase Central — см. рис. 2 (в программе открыта топологическая схема системы). Sybase Central поддерживает управление кластерами «одним щелчком мыши», операции обслуживания на ходу (в том числе добавление столбцов в работающую базу данных), а также графическое представление загрузки процессоров, отдельных потоков и их тайминга, что облегчает поддержку системы. Здесь же выполняется управление ресурсами и задаются параметры безопасности. Дополнением к Sybase Central служит Sybase Control Center для Sybase IQ, позволяющий удаленно контролировать работу инфраструктуры IQ через веб-интерфейс. Имеется богатая номенклатура метрик работы системы (в том числе в ретроспективе) с наглядным графическим представлением. В перспективе в Sybase Control Center планируется включить функциональность Sybase Central, пока же заказчикам предлагаются оба средства.

Что касается безопасности, то Sybase IQ поддерживает аутентификацию по протоколу Kerberos, позволяющую использовать в базе данных общие с операционной системой имена и пароли (задаваемые пользователями). Непосредственное использование LDAP в настоящее время не поддерживается. Следует отметить, что ролевые механизмы обеспечения безопасности применимы не только к пользователям, но и к администраторам: для задач системного мониторинга, управления учетными записями и полномочиями, резервного копирования

## Продукт

и восстановления, а также управления кластером (мультиплексом) существуют отдельные роли. Такое разделение обязанностей полезно, когда одно хранилище обслуживает несколько доменов.

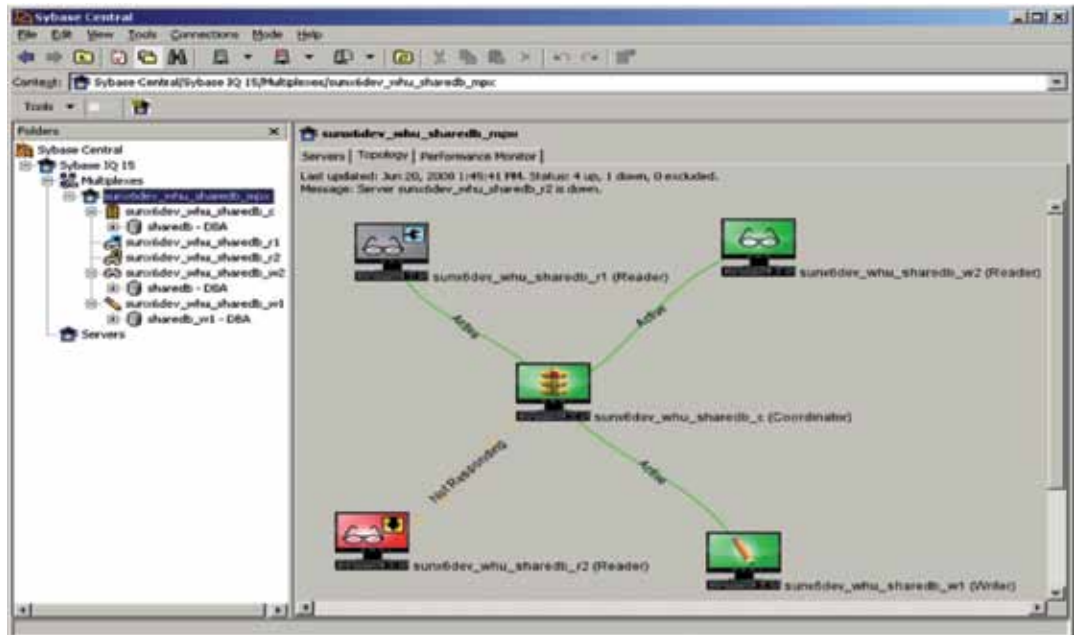


Рис. 2. Средство администрирования Sybase Central

## Заключение

Система Sybase IQ основана на технологии, которую компания Sybase (входящая ныне в состав SAP) получила в свое распоряжение в результате приобретения в 1995 году фирмы Expressway. Одним из основных принципов этой технологии является поддержка масштабирования путем постепенного наращивания мощности вплоть до очень крупных хранилищ данных. Следует отметить, что в 2007 году Sybase построила, при участии компаний Sun (ныне Oracle) и VMsoft особо крупное хранилище объемом 1 Пбайт (1 000 Тбайт); его основой послужил Sybase IQ 12.x. Это было первое хранилище такого размера, которое прошло независимый аудит.

Современная стратегия развития Sybase IQ предусматривает приоритетную специализацию на задачах предоставления аналитических услуг, передовой аналитической обработки и быстрого создания отчетов, и как показала практика, этот подход полностью оправдывается. Сейчас по всему миру установлено свыше 3 500 систем на базе Sybase IQ, применяемых более чем 2 000 организаций. Доходы, приносимые продажами Sybase IQ и оказанием сопутствующих услуг, постоянно растут. Уместно заметить, что значительная доля продаж продукта приходится на организации, не использующие Sybase ASE (транзакционная СУБД — основной продукт компании).

Помимо рынков, рассмотренных в начале этого документа, сейчас Sybase активно ведет деятельность в ряде новых перспективных областей, где, как ожидает компания, возможности Sybase IQ найдут особый спрос:

- **Управление рисками на рынке капитала.** Продукт Sybase RAP — The Trading Edition использует Sybase IQ для организации особо крупных хранилищ данных и содержит механизм обработки сложных событий. Он позволяет, обрабатывая большие массивы данных реального времени совместно с ранее накопленной информацией, формировать цельную картину рынка, необходимую для работы трейдеров, управляющих портфелями и менеджеров рисков в корпорациях.
- **Исполнение требований надзорных органов,** предусматривающее особые условия хранения информации (структурированной и неструктурированной) и создание отчетности. При этом зачастую бывает необходимо хранить очень большие, постоянно растущие массивы легко доступных данных (иногда такое хранение нужно и для решения аналитических задач). Sybase IQ прекрасно для этого подходит. Следует отметить, что для инвестиционных рынков соответствие требованиям надзорных органов понимается в более широком смысле и может быть достигнуто с помощью Risk Analytics Platform (RAP). Система Sybase IQ сама по себе отвечает различным требованиям в части безопасности (как уже было показано выше) и пригодности для эксплуатации лицами с ограниченными возможностями, соответствуя требованиям раздела 508 Акта о реабилитации — как в отношении интерфейса, так и документации. Еще одним средством обеспечения соответствия нормативным требованиям является возможность зафиксировать

выбранные базы IQ по состоянию на определенный момент времени на специальных носителях с доступом только для чтения.

- **Информационная проходка данных/текста.** В сотрудничестве с партнерами Sybase создает средства комбинированной аналитической обработки данных и текста. В частности, это позволяет сопоставлять реляционные данные, хранимые в таблицах IQ, с нереляционными данными, хранящимися в виде больших объектов (LOB). Здесь следует отметить, что Sybase IQ обладает расширенными возможностями сложной аналитической обработки таких типов данных, как CLOB (большие символьные объекты), BLOB (большие двоичные объекты) и XML. В частности, поддерживается индексирование CLOB для текстового поиска.

Компания Sybase имеет ряд договоров о сотрудничестве в отношении Sybase IQ с различными поставщиками вычислительного оборудования, систем хранения данных, средств обеспечения качества данных, бизнес-аналитики и других, а также создающими добавленную стоимость реселлерами и системными интеграторами. Компания Business Objects, также входя в состав SAP, является близким партнером Sybase.

### Адреса страниц Sybase IQ в сети Интернет:

[www.sybase.com/bi](http://www.sybase.com/bi) и [www.sybase.ru/iq](http://www.sybase.ru/iq)

## Заключение

---

Ранее мы полагали, что продажам Sybase IQ могут воспрепятствовать два обстоятельства: во-первых, недостаточное понимание рынком преимуществ поколоночной организации данных, во-вторых, недоверие к заявленным сведениям о разнице в производительности по сравнению с традиционными СУБД. Однако теперь мы не видим этих препятствий. С одной стороны, преимущество поколоночной организации было многократно подтверждено практическим ее использованием во многих ведущих компаниях. С другой стороны, как Sybase, так и поставщики специализированных машин хранения данных доказали, что ускорение аналитической обработки вполне реализуемо, и тем самым способствовали формированию спроса среди компаний, использующих для решения аналитических задач мало подходящие для этого традиционные СУБД.

По мнению Bloor Research, поколоночный принцип хранения, сторонником которого является Sybase, обеспечивает существенно большее быстродействие при более низкой стоимости сравнительно с традиционными подходами к созданию систем анализа, отчетности и хранилищ данных. Более того, он отличается гораздо большей гибкостью по сравнению с машинами хранения данных. Поэтому мы не видим никаких препятствий дальнейшему успешному продвижению системы. Sybase IQ имеет функции, отсутствующие у конкурентов (такие, как аналитическая обработка текстовой информации), а с реализацией в новом выпуске массово-параллельной обработки продукт должен достигнуть равной производительности с конкурирующими решениями, если не более высокой.

Ранее мы уже рекомендовали адресовать продукт более широким рынкам, и теперь рады отметить, что компания (с момента своего вхождения в состав SAP) последовала нашему совету. Разумеется, в определенных областях система обладает наиболее значительными преимуществами, и в то же время ее возможности должны позволить ей успешно конкурировать на профильном рынке целиком.

### Дополнительная информация

Более подробные сведения по данной теме опубликованы по адресу <http://www.BloorResearch.com/update/2092>

## Сведения о компании Bloor Research

Bloor Research — одна из ведущих европейских исследовательских, аналитических и консультативных фирм на рынке ИТ, сфера деятельности которой охватывает все сегменты индустрии. Мы помогаем заказчикам повышать гибкость информационных систем за счет эффективного руководства, управления и использования информации. Благодаря принципам независимости, взвешенности и ясности, которым мы неизменно следуем при предоставлении информации и подготовке публикаций, мы заслужили репутацию откровенной компании. Мы считаем, что предоставление полной информации позволяет:

- описать технологию с точки зрения ее ценности для бизнеса в контексте взаимодействующих с ней систем и процессов;
- понять, как новаторские средства и технологии могут быть адаптированы к имеющейся ИКТ-инфраструктуре;
- рассмотреть рынок в целом и представить всю совокупность существующих решений, предложив способы их эффективной оценки;
- исключить постороннюю информацию и упростить поиск сведений, полезных при приобретении и внедрении программно-технических средств;
- обеспечить предоставление всех имеющихся у нас сведений в наиболее подходящей форме.

Компания Bloor Research основана в 1989 году и уже более двадцати лет, действуя во всемирном масштабе, обеспечивает организации, потребляющие и производящие информационную технику и технологии, результатами исследований и аналитической информацией. Мы распространяем информацию через онлайн-подписку, предлагаем специализированные исследовательские услуги, организуем мероприятия и реализуем консультативные проекты. Мы стремимся превратить свои знания в ценный актив вашего бизнеса.

## Об авторе

Филип Ховард

Директор по исследованиям в области обработки данных

Филип начал свою карьеру в компьютерной индустрии в 1973 году и имеет богатый профессиональный опыт: в разное время он работал системным аналитиком, программистом, специалистом по продажам, маркетингу и разработке продуктов. Он трудился в разных компаниях, в том числе в GEC Marconi, GPT, Philips Data Systems, Raytheon и NCR.



В 1992 году, спустя четверть века после начала карьеры, Филипп организовал собственную компанию, ныне называющуюся P3ST (Wordsmiths) Ltd. Ее первым клиентом стала фирма Bloor Research (тогда ButlerBloor), где Филип стал партнером-аналитиком. Его взаимоотношения с Bloor Research с тех пор не прекращались; теперь он занимает в ней должность директора по исследованиям. Область компетенции Филипа — все, что касается обработки данных (в том числе контента); в его коллективе работают еще пять аналитиков. Обозревая всю область в целом, Филип специализируется на СУБД, вопросах управления данными, интеграции данных, обеспечения качества данных, федерации данных, управления мастер-данными, руководства процессами обработки данных и на хранилищах данных. Он также занимается проблемами обработки сложных событий и потоков событий.

Помимо многочисленных отчетов, которые Филип пишет для Bloor Research, он регулярно сотрудничает с электронными изданиями IT-Director.com и IT-Analysis.com; он был также редактором изданий «Application Development News» и «Operating System News», выпускаемых компанией Cambridge Market Intelligence (CMI). Кроме того, Филип сотрудничает с разными журналами и составил ряд отчетов, опубликованных такими организациями, как CMI и Financial Times.

В свободное от работы время Филип любит управлять узкой баржей, кататься на лыжах, играть в бридж (имеет звание лайфмастера) и гулять с собакой.



## Информация об авторских правах и условия использования

© 2011, Bloor Research. Никакую часть настоящей публикации не разрешается воспроизводить никаким способом, не получив предварительного согласия Bloor Research.

Природа изложенных здесь сведений требовала упоминания названий многочисленных аппаратных и программных продуктов. Как правило, все эти названия являются товарными знаками производящих эти продукты компаний. Bloor Research не претендует на присвоение этих названий и товарных знаков. Приведенные здесь логотипы компаний, графические изображения и снимки экрана воспроизведены с разрешения правообладателей, которые и регулируют их использование.

При подготовке документа компания приложила все усилия к обеспечению достоверности информации. Тем не менее, издатель не может взять на себя ответственность за возможные ошибки или упущения.



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

Тел.: +44 (0)207 043 9750

Факс: +44 (0)207 043 9748

Интернет: [www.BloorResearch.com](http://www.BloorResearch.com)  
Электронная почта: [info@BloorResearch.com](mailto:info@BloorResearch.com)