

Открытые системы. СУБД №05, 2012

**Аналитика без ограничений**

Большим Данным сегодня уделяется много внимания, и, как правило, работа с ними ассоциируется с большими проблемами, но так ли это? Решения Sybase IQ и СУБД HANA, интегрированные на платформе SAP Real-time Data Platform, позволят в реальном времени решать

*Артем Гришковский*

Безусловно, объемы данных увеличиваются — согласно исследованиям IDC, объем корпоративных данных к 2020 году возрастет в 35 раз, однако «большие данные» были всегда, но лишь недавно они привлекли к себе внимание.

Под Большими Данными понимаются большие объемы структурированной и неструктурированной информации, поступающей из разных источников, которую сложно анализировать традиционными методами. В этой связи фокус деятельности большинства компаний сегодня направлен на повышение скорости и частоты загрузки новых данных, ускорение выдачи результатов на запросы, снижение затрат на хранение таких данных. В России проблема Больших Данных актуальна для компаний банковской и телекоммуникационной отраслей, накапливающих большие объемы информации различного типа, к частоте обновления и скорости обработки которой предъявляются строгие требования. Крупные торговые сети также нуждаются сегодня в решениях для обработки Больших Данных. Однако не следует считать, что аналитика Больших Данных — это какая-то специализированная технология для определенных компаний, напротив, она становится экономически эффективным способом хранения и анализа больших массивов данных для предприятий всех отраслей. Например, западные компании, работающие с Большими Данными, уже вводят новую должность — Chief Data Officer (CDO), который отвечает не только за технологии, но и за управление информацией.

**Смутное время СУБД**

Под влиянием роста объемов данных в бывшей совсем недавно стабильной области СУБД начались подвижки, выразившиеся, в частности, в появлении двух движений — NoSQL и NewSQL.

*Леонид Черняк*

Большие Данные могут сделать компанию или организацию более подготовленной к изменениям на рынке, более конкурентоспособной и прибыльной, способны помочь в принятии более обоснованных решений. Однако, чтобы воспользоваться этими преимуществами, необходимо внедрять новые инструменты, а не пытаться расширить сферу применения уже существующих. Компания SAP предлагает несколько технологий для работы с Большими Данными: аналитический сервер SAP Sybase IQ, платформу для обработки сложных событий Sybase ESP и СУБД HANA.

**SAP Sybase IQ.** Специализированный сервер, предназначенный для сверхбыстрой работы в качестве хранилища данных, платформы для систем отчетности и интерактивной бизнес-аналитики. Сервер работает на стандартных аппаратных средствах и под управлением различных операционных систем. Система Sybase IQ была создана еще в 1994 году специально для решения задач аналитики, а не для обработки транзакций, поэтому изначально был применен поколоночный способ организации данных, который на момент создания продукта был еще только предметом обсуждения академического сообщества. Примерно 10–12 лет назад объемы обрабатываемых данных начали стремительно расти, и стали очевидны преимущества поколоночного хранения — скорость выполнения запросов в IQ оказалась на порядки выше, чем в реляционных СУБД. Система сжатия позволяет экономить до 70% дискового пространства. При этом возможно практически линейное масштабирование.

В 2009 году вышла версия Sybase IQ 15.0 (рис. 1), и с этого момента эта СУБД начала приобретать необходимый функционал для работы с большими объемами данных. Новшеством версии 15.0 стала платформа для работы с особо крупными базами данных, а также улучшенные алгоритмы загрузки данных и обработки запросов, учитывающие особенности многоярусных хранилищ. В следующих версиях 15.x была предусмотрена возможность внутрибазовой аналитической обработки, появились функции оптимизированного текстового поиска, федерации запросов и поддержки Web 2.0 на уровне программных интерфейсов. Кластерная архитектура массово-параллельной обработки PlexQ позволяет ускорить обработку запросов высокой сложности за счет распределения нагрузки между множеством машин.



Рис.1. Этапы развития Sybase IQ 15

Версия 15.4, вышедшая в 2011 году, была полностью посвящена Большим Данным: введен интерфейс для работы с технологией MapReduce, применяемой для анализа больших массивов структурированных, квазиструктурированных и неструктурированных данных; добавлен механизм интеграции с Hadoop, позволяющий находить интересующую информацию в больших массивах структурированных и неструктурированных данных, а затем интегрировать ее в Sybase IQ для совместного анализа с транзакционными данными, а также результатами, полученными из других источников данных; включена поддержка языка PMML (Predictive Model Markup Language) описания аналитических моделей. Кроме того, в состав пакета включена расширенная библиотека алгоритмов статистической обработки информации, в полной мере использующая возможности распределенной обработки запросов в кластерной конфигурации PlexQ.

**Sybase ESP.** Данная платформа смещает фокус обработки информации с анализа исторических данных на оперативную ситуацию. Технология обработки потоков событий обеспечивает прием данных в реальном времени и их представление в виде динамического потока событий: изменение курса ценных бумаг, установка цены исполнения, исполнение заявки либо сочетание каких-либо событий. Система может быть запрограммирована на генерацию оповещений при выявлении определенных условий. Sybase ESP в основном используется финансовыми компаниями, которые с ее помощью проверяют рыночные данные, осуществляют трейдинг, мониторинг и анализ (включая анализ рисков в реальном времени), выявляют соответствие нормативным требованиям, получают информацию о прибылях и убытках и выполняют другие задачи.

**SAP HANA.** Аналитический программно-аппаратный комплекс поддержки операционных и аналитических приложений. Изначально СУБД HANA с обработкой данных в оперативной памяти проектировалась как система с построчным хранением, однако позже было объявлено, что она поддерживает еще хранение по строкам. Согласно стратегии компании SAP, технология HANA станет основой платформы обработки данных в реальном времени, в которую также войдут и другие решения Sybase. На Sybase IQ будет лежать функция по управлению большими массивами данных, а Sybase ESP будет использоваться для захвата данных в реальном времени.

### MapReduce — будущее баз данных

СУБД — продукты своего времени, и они не могут вечно оставаться образцами совершенства. Есть все основания полагать, что в недалеком будущем свое место займут параллельные СУБД, использующие программную конструкцию MapReduce.

Леонид Черняк

В начале 2012 года компания SAP AG анонсировала платформу SAP Real-time Data Platform. Платформа (рис. 2) основана на флагманском продукте компании SAP HANA, включает в себя возможности по управлению данными средствами Sybase IQ, Sybase ESP, Sybase ASE и SAP Enterprise Information Management, а решения на ее основе позволяют работать с Большими Данными в режиме реального времени.

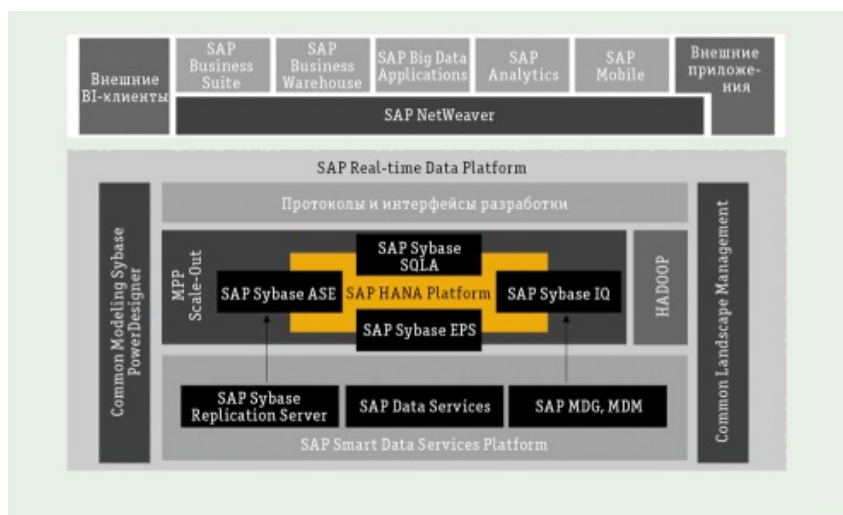


Рис. 2. Платформа SAP для обработки данных в реальном времени

SAP HANA и SAP Sybase IQ могут совместно использоваться для решения специфических аналитических задач. Первая лучше подходит для анализа оперативной информации, а вторую лучше использовать в качестве масштабируемого хранилища данных. Поддержка MapReduce и Hadoop в Sybase IQ, а также его оптимизатор и средства индексирования расширяют возможности HANA, благодаря чему пользователи получают средства более глубокого анализа, доступ к внешним и внутренним интеллектуальным библиотекам и поддержку массовой одновременной нагрузки. Помимо этого, совместное использование этих продуктов позволяет обеспечить мгновенный доступ ко всем массивам данных компании и основу для проведения анализа, позволяющие дать ответ на любой бизнес-запрос.

Поскольку с приобретением Sybase в портфеле решений SAP появилось несколько СУБД, заказчику предоставляется возможность выбора СУБД, отвечающей его задачам, требованиям и бюджету. Если для компании в первую очередь важна быстрая аналитика реального времени, то подходящим решением будет SAP HANA, а если требуется недорогое решение для высокопроизводительного анализа корпоративных данных — то SAP Sybase IQ.

\*\*\*

Технологии для обработки Больших Данных позволяют увеличивать глубину анализа данных, расширять спектр связанных массивов при анализе, уменьшать время отклика по сложному запросу, выделять связи между структурированными и неструктурированными данными и т. д. При работе с большими объемами корпоративных данных все это дает предприятиям и организациям новые возможности управления и ведения бизнеса, позволяющие получать постоянный доступ к нужным данным на требуемом уровне детальности.

**Артём Гришковский** ([Artem.Grishkovskiy@sybase.ru](mailto:Artem.Grishkovskiy@sybase.ru)) — заместитель генерального директора Sybase CIS (Москва).

Постоянный URL статьи: <http://www.osp.ru/os/2012/05/13016218/>

© «Открытые системы», 1992-2012. Все права защищены.