

Sybase для высокопроизводительной аналитики



Андрей Хромов – ведущий технический консультант Sybase CIS.

Sybase IQ – это быстродействующая аналитическая платформа, которая снаружи выглядит, как обычная реляционная СУБД. Она имеет открытые интерфейсы, поддерживает стандарты ODBC и JDBC, а также язык SQL и не требует использования специальной схемы, оптимизированной для аналитики. Sybase IQ работает на стандартных аппаратных средствах и под управлением разнообразных операционных систем. В то же время внутри SAP Sybase IQ устроена иначе, чем обычные СУБД. Это высокопроизводительное хранилище данных с организацией по столбцам, рассчитанное на скоростную загрузку данных, быструю обработку запросов и легкое масштабирование в отношении количества пользователей. Благодаря этому Sybase IQ выполняет запросы в 10-100 раз быстрее, чем традиционные СУБД. Эффективные алгоритмы сжатия данных позволяют экономить до 70% дискового пространства. При этом система масштабируется почти линейно.

Sybase IQ позволяет значительно упростить архитектуру хранилища данных, делая ненужным многочисленные агрегаты, витрины данных, OLAP-кубы. Все эти дополнительные конструкции были придуманы с одной лишь целью – добиться нужной скорости работы аналитики. Правда, ценой потери детальности, возможности анализа по всему спектру

данных. Способность Sybase IQ очень быстро работать даже на больших объемах данных позволяет обойтись лишь одной базой – центральным хранилищем данных, и получать аналитику по всему объему детальных данных с временем отклика в несколько секунд.

Начиная с 2009 года, было выпущено 5 новых версий Sybase IQ. Главными нововведениями версии 15.0 стали: новая платформа работы с особо крупными базами данных (VLDB) и модернизированные алгоритмы загрузки данных и обработки запросов, учитывающие особенности новых многоядерных процессорных архитектур и многоярусных систем хранения данных. В версии 15.1 была введена возможность внутрибазовой аналитической обработки – запуск внешних программ на C++ средствами механизма БД. В версии 15.2 были представлены новые функции поиска в текстовых массивах, федерации запросов и поддержки Web 2.0 на уровне программных интерфейсов (API).

Главным новшеством выпуска 15.3 стала кластерная архитектура массово-параллельной обработки PlexQ. Наконец, в версии 15.4 была введена поддержка нативного API MapReduce, обеспечивающего внутрибазовую аналитическую обработку больших объемов данных. Благодаря этому, а также поддержке языка предсказательного моделирования PMML (Predictive Model Markup Language), интеграции с Hadoop, наличию расширенной библиотеки алгоритмов статистических расчетов и информационной проходимости, использующих все возможности технологии PlexQ®, разработчики могут быстро и надежно программировать аналитическую обработку внутри базы данных, что дает более чем 10-кратный прирост производительности по сравнению с обычным способом.

С помощью Sybase IQ можно применительно к крупнейшим массивам данных решать такие задачи, как выявление тенденций, поиск первопричин событий, определение паттернов и моделирование по принципу "что, если", совместно обрабатывать структурированные и неструктурированные данные, получать более глубокую и точную картину эффективности деятельности компании и динамики рынка, а также строить прогнозы, полностью высвобождая потенциал больших информационных массивов.

Для организации хранилищ данных может использоваться и другой продукт SAP – программно-аппаратная платформа HANA (High-Performance Analytic Appliance), где также применена технология поколоночного хранения. Совместное использование Sybase IQ и HANA позволяет решать специфические аналитические задачи, где необходимы высокие скорости обработки больших объемов данных. При этом Sybase IQ может исполнять роль корпоративного хранилища, а HANA – высокопроизводительного аналитического кэша. Например, исторические данные могут находиться в дисковом хранилище под управлением Sybase IQ, а наиболее активно используемую их часть можно поместить в оперативную память HANA. Прочие данные по мере необходимости будут извлекаться из Sybase IQ с помощью интерфейса Near-Line Storage. Такой интерфейс имеется, в частности, в SAP BW.



BitLocker (впервые появилась в Windows Vista) в кластерной конфигурации для шифрования Cluster Shared Volumes версии 2.0. BitLocker позволяет защищать данные путем полного шифрования тома. В новой версии серверной ОС от Microsoft BitLocker доступен в виде опционального компонента. Кроме того, в Windows Server 2012 будут доступны еще два компонента: Enhanced Storage, предназначенный для поддержки устройств с аппаратным шифрованием, и BitLocker Network Unlock. Максимальную функциональность BitLocker обеспечивает, работая с Trusted Platform Module версии 1.2 и выше – криптопроцессором, предназначенным для хранения ключей и сертификатов. Рассмотрим изменения, появившиеся в BitLocker, для операционной системы Windows Server 8 Beta:

- шифрование только использованного пространства тома;
- активация BitLocker до начала установки ОС;
- возможность изменения пароля и стандартным пользователем;
- Network Unlock. Автоматическое разблокирование системного тома при перезагрузке системы при подключении к сети в доменной среде. Функция доступна только для систем, имеющих поддержку UEFI с модулем UEFI DHCP driver;
- возможности шифрования кластерных томов. Основная, на наш взгляд, функция добавленная в Windows Server следующего поколения;
- поддержка носителей с встроенным аппаратным шифрованием;

- управление BitLocker в распределенных средах в соответствии с групповыми политиками.

Выводы

Разработчики Microsoft проделали большую работу по развитию сервисов хранения в новых версиях своей операционной системы и гипервизора. Хотя мы и не видим здесь такого разнообразия функций, которое существует в продуктах компании VMware, все вновь добавленные или улучшенные функции действительно являются необходимыми для построения частных и публичных облаков и долгое время ожидалось пользователями, которые ориентируются на продукты компании Microsoft при построении своей инфраструктуры.

Сергей Платонов,
компания AVRORAID