

# Сердце бизнес-приложений



**Вадим ЯЦЕНКО,**  
генеральный директор компании «Тантор Лабс» (входит в ГК «Астра»)

## Место СУБД в корпоративной инфраструктуре

СУБД и базы данных не существуют «сами по себе», без каких-либо приложений, если говорить не только о корпоративном хранилище, хотя даже для работы с ним используются средства бизнес-аналитики и другие инструменты. СУБД – неотъемлемая часть любого бизнес-приложения, информационной системы, которая, в свою очередь, управляет производственными процессами, ресурсами и т. д. Таким образом, СУБД помогает бизнесу именно как составная часть той или иной инфраструктуры. В основе любой управляющей ресурсами системы должен лежать надежный фундамент – качественная и высокопроизводительная СУБД.

В первую очередь СУБД позволяют правильно организовать данные, чтобы можно было выстроить аналитический онлайн-процессинг. Схемы «Снежинка», «Звездочка» и другие строятся

Система управления базой данных (СУБД) является центром современных корпоративных бизнес-приложений. Сегодня иностранные производители этих базовых продуктов отказываются поддерживать их на территории России, и отечественным разработчикам приходится переходить на российские аналоги.

в базе данных на определенной СУБД. Сейчас все заказчики стараются трансформироваться в компанию, управляемую данными (Data Driven Company), т. е. принимать решения на основе не каких-то человеческих факторов, а объективных данных – статистики, аналитики и т. д.

Меняются и сами СУБД. Их разработчики все чаще фокусируются на задачах по обработке больших объемов данных, что необходимо для принятия правильных решений. Эти данные должны обрабатываться за приемлемое время. Период от их получения до момента, когда они уже обработаны и переданы бизнесу в удобном для него формате, у каждого бизнеса свой. Для одной компании сутки – это нормально, для другой требуемая скорость обработки данных исчисляется секундами или даже миллисекундами: за это время информация должна успеть пройти через аналитическую систему, в том числе СУБД, и быть представлена для бизнеса. Это уже так называемая потоковая обработка данных.

Что касается аналитической обработки данных средствами СУБД, то чаще всего речь идет о языке SQL. К нему все привыкли, на нем все работают. Возможности для потоковой обработки данных имеются в СУБД, что позволяет повысить их粒度 (например, это агрегативные функции, функции прореживания данных и многое другое).

## Выбор СУБД

Облачные СУБД бывают транзакционные, гетерогенные и чисто

аналитические. Кроме того, существуют документоориентированные, реляционные, с обработкой в памяти (in-memory), колумнары – одним словом, типов очень много. Практически все они реализованы и в виде облачных сервисов.

Есть универсальные базы данных, такие как PostgreSQL и проекты, построенные на ее коде. Почему они называются универсальными? Потому что предоставляют возможность работать и с транзакционной нагрузкой, и с аналитической. При этом имеются определенные пределы, после достижения которых необходимо применять специализированные СУБД для решения определенных задач.

При выборе конкретной СУБД рекомендуем руководствоваться следующим. Если профиль нагрузки неизвестен и не определено, какие данные будут храниться, то лучше использовать универсальную базу данных, чтобы в дальнейшем не столкнуться с ограничением специализированных решений. При наличии точной информации о том, какие данные будут обрабатываться, лучше отдать предпочтение специализированной базе данных. Например, требуется хранить данные временных рядов, и информации будет очень много: какие-то метрики, датчики IoT. В этом случае целесообразно выбрать специализированную базу данных для работы с временными рядами. Но если нужны данные временных рядов, однако в процессе их получения необходимо каким-то образом интегрировать транзакционные данные другого происхождения, то выбор инструмента уже не столь очевиден. Здесь можно

подумать и о применении транзакционной БД.

Таким образом, выбор зависит от конкретной бизнес-задачи, т. е. потребностей и способов обработки данных, их специфики и природы. Это касается любых СУБД, не только облачных, поскольку облачная СУБД представляет собой просто готовый сервис (Database as a Service), хотя и с расширенными возможностями масштабирования.

Если же компании нужно обрабатывать большие объемы данных, то придется использовать другой тип СУБД – ориентированный на Big Data. Сам по себе термин Big Data уже вышел из обихода, в свое время это просто был модный хайп. Никто не знает, что такое Big Data, с какого уровня начинается: 1, 10, 100 терабайт или петабайт, однако понятно, что для обработки больших объемов данных нужно использовать распределенные решения, такие как Hadoop. Это скорее способ организации параллельной обработки данных с помощью небольших вычислительных устройств, нежели что-то принципиально новое.

В целом для анализа больших объемов данных СУБД предоставляет широкий набор инструментов, в первую очередь параллелизм. Чтобы работать с большим объемом данных, необходимо максимально распараллелить операции. Чем больше параллельных потоков можно задействовать, тем шире возможности для анализа и обработки данных. В классических СУБД параллелизм может достигаться за счет различных факторов, например, разбиения базы на несколько фрагментов (шардинга) или разделения вычислительного узла и хранилища. В распределенных для этого используются методы map-reduce. Современные СУБД предоставляют и поддерживают множество разных механизмов и архитектурных подходов для обработки больших объемов данных. Для больших хранилищ разрабатываются специфические форматы хранения данных. Например,

для записи на диски больших объемов данных часто применяются хранилища формата S3.

## Безопасность

СУБД хранит данные и обеспечивает не только их безопасность, но и сохранность, и консистентность, выполняет множество различных функций. В частности, реализует такие механизмы, как ролевое разграничение прав доступа – это гранулированный допуск именно к той информации, которая предназначена для конкретного пользователя. Можно обеспечить доступ до конкретной ячейки данных, доступ к самой СУБД тоже регулируется – пользователи, которые не прошли аутентификацию, не смогут получить информацию. Следующий момент – шифрование данных. Они не хранятся в открытом виде, а шифруются на диске как средствами самой СУБД, так и сторонними инструментами.

В соответствии с требованиями законодательства СУБД должна журналировать все действия, которые происходят с объектами базы данных, пользователями и доступом к этим данным. Это необходимо для того, чтобы, например, специалисты службы информационной безопасности могли настраивать политики и проводить расследования. К информационной безопасности относятся и требования парольной политики и консистентности информации – защиты от подмены.

## Производительность

СУБД и база данных работают на определенном оборудовании. Если СУБД построена таким образом, что неэффективно использует его, то можно столкнуться с ситуацией, когда бизнес будет переплачивать за хранение данных.

Кроме того, бизнес-приложения, которые пишутся разработчиками, различаются по качеству. Особенно это проявляется в производительности запросов, поступающих в базу данных. Для их оптимизации полезной будет система

профилирования и мониторинга, а в самой СУБД – механизмы, которые позволяют выявлять некачественные показатели, предоставляя интерфейсы для оптимизации неэффективных запросов и код. Но оценка эффективности для бизнеса здесь вряд ли применима, речь скорее о снижении инфраструктурных затрат.

## Отказоустойчивость

Базы данных относятся к сервисам, которые должны безотказно работать в режиме 24/7/365. Типичные проблемы, как правило, связаны с отказоустойчивостью. Для того чтобы бизнес-приложения работали с тем временем отклика, которого ожидает пользователь, обычно используются программы резервного копирования, сохранности данных и оптимизации производительности БД. Стандартные задачи для администраторов БД (DBA) – построение отказоустойчивого решения с резервным копированием и восстановлением данных при необходимости, постоянный мониторинг, обслуживание, наблюдение за тем, что происходит со схемой данных, аудит.

В целях исключения аварий в арсенале DBA всегда должны присутствовать инструменты повышения отказоустойчивости.

## Особенности импортозамещения

Чаще всего DBA сталкиваются с нехваткой функциональности решений с открытыми кодами, предлагаемых сегодня на рынке. Прежде всего, это отсутствие графических интерфейсов и удобных инструментов управления, к которым все привыкли за годы пользования коммерческими корпоративными продуктами западных компаний. Однако сейчас появляются инструменты искусственного интеллекта, которые можно использовать, в частности, для автоматизации управления СУБД – возможно, с их помощью удастся сгладить ограничения командных интерфейсов. ■