

На пути к автономным базам данных

В арсенале современных СУБД, управляемых квалифицированными администраторами, имеется богатый набор средств управления, масштабирования и обеспечения безопасности, однако рост сложности решаемых задач и необходимость моментально реагировать на изменения требуют автономных баз, снижающих влияние человеческого фактора и способных самостоятельно поддерживать работоспособность оптимальной конфигурации.

Ключевые слова: Облака, автономная база данных
Keywords: Cloud, Autonomous Database

Марк Ривкин

Системы хранения и обработки данных условно можно разделить на простые и сложные. К первым относятся, например, Microsoft Excel, почта mail.ru или популярная в 1980–1990-е годы СУБД dBASE. Для того чтобы их развернуть и начать применять, не требуется администратор, здесь никто не обеспечивает многопользовательский режим или настройку производительности, защита данных сводится к установке пароля доступа к файлу, а надежность обеспечивается путем его копирования. Конечно, администраторы mail.ru поддерживают систему, но вся их работа скрыта от конечного пользователя. Обратная сторона такой простоты в том, что сложные, бизнес-критичные системы хранения и обработки данных на таких продуктах не построить. Для этого используются СУБД класса Microsoft SQL Server, Oracle, IBM DB2, PostgreSQL и т. д., позволяющие обеспечить требуемую корпорациям высокую производительность, надежность, масштабируемость и безопасность, но в обмен на необходимость содержать администраторов, регулярно выполнять операции по установке, настройке, сопровождению, резервному копированию и восстановлению. Автономные базы данных призваны совместить простоту первых с мощностью вторых.

Идея автономных баз данных отчасти перекликается с популярной сегодня концепцией беспилотных автомобилей, для

которых выделяется пять уровней автономности (на пятом уровне уже не требуется вмешательство человека в процесс управления). Пользователь формулирует свои требования к надежности, безопасности, производительности и другим параметрам базы данных, определяет политики и уровень сервиса, а система автоматически развертывает и поддерживает эксплуатацию конфигурации нужной архитектуры. Однако следует

обратить внимание на то, что автономные базы — это концепция, а не конкретный продукт, решение или очередная СУБД, снабженная разнообразными средствами автоматизации.

К наиболее трудоемким задачам, выполняемым администратором базы данных, относятся следующие: развертывание и конфигурирование ИТ-инфраструктуры; установка СУБД, обновление версий; резервное копирование и восстановление;



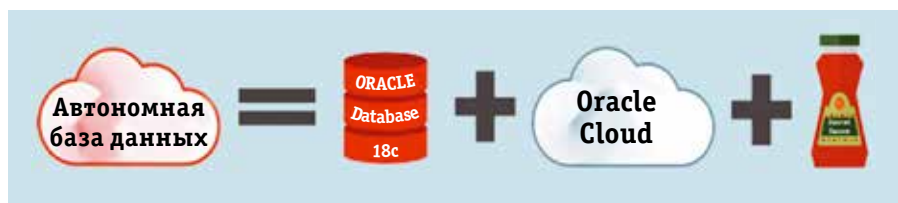


Рис. 1. Компоненты автономной базы данных

развертывание новых сред для тестирования, разработки и т. д.; построение отказоустойчивого решения; защита данных, устранение дыр в безопасности; оптимизация работы базы, настройка SQL; управление жизненным циклом данных; выявление и устранение ошибок в ПО; выделение дополнительных вычислительных ресурсов (поддержка эластичности системы). Все эти задачи и должна вместо администратора выполнять автономная база данных, на которую возлагаются три основные функции:

- **самоуправляемость** — автоматическое создание базы, выполнение обновлений и исправлений, резервное копирование, настройка производительности, обеспечение эластичности;
- **самозащита** — защита от внешних и внутренних атак, оперативное исправление дыр в безопасности, шифрование данных, маскирование;
- **самовосстановление** — автоматическое восстановление после сбоев, минимизация простоев, обеспечение заданного уровня обслуживания сервиса.

По версии компании Oracle, основными компонентами автономной базы данных (рис. 1) являются СУБД Oracle 18c, облачная инфраструктура Oracle Cloud и дополнительное программное обеспечение.

Автономная база данных Oracle может работать только в облачной инфраструктуре Oracle Cloud, предоставляющей все необходимые ресурсы. Конечно, пользователь в своем локальном ЦОД может установить СУБД Oracle 18c или заказать облачный сервис DBaaS, но это не будет автономной базой, которая по сути представляет собой специальный сервис, работающий в облаке Oracle или на Oracle Cloud@Customer (облачная машина Oracle, развернутая в локальном ЦОД и удаленно администрируемая сотрудниками Oracle). Для того чтобы автономная база данных эффективно поддерживала конкретные задачи, пользователь должен заказать конкретный специализированный сервис, и тогда база будет автоматически сконфигурирована и настроена для выполнения именно этой задачи. Например, сейчас

для создания витрин и небольших хранилищ данных можно заказать автономный сервис хранилищ данных (Autonomous Data Warehouse Service, ADWS), а вскоре будет доступен автономный сервис для систем оперативной обработки транзакций и смешанной нагрузки.

Еще одна особенность автономной базы — эластичность. Изменение количества требуемых процессоров и объема дискового пространства, необходимого для решения задач пользователя, производится через веб-интерфейс без остановки работы, а оплата осуществляется по факту использования конкретного сервиса. При увеличении числа задействованных процессоров автоматически увеличивается и объем оперативной памяти для автономной базы. Администрирование всего стека автономной базы берет на себя поставщик СУБД — он отвечает за поддержку оборудования и операционной системы, выпуск обновлений и мониторинг конфигурации.

Для самовосстановления автономной базы автоматически создается конфигурация, необходимая для поддержания заданного уровня обслуживания максимальной доступности и надежности. Максимальный уровень надежности для Oracle ADWS — это 99,995%. Он обеспечивается, в частности, благодаря Real Application Clusters (защита от сбоя узла), резервной (Standby) базе для защиты от катастрофических сбоев, а также механизму ASM (Automatic Storage Manager), который защищает от сбоя дисков и т. п.

Для обеспечения самозащиты автономная база обязательно шифрует все данные, автоматически применяет обновления безопасности, а также использует механизмы машинного обучения для выявления и предсказания сбоев и внешних атак.

МЕХАНИЗМЫ АУТОНОМНОЙ БАЗЫ

Сегодня в арсенале опытного администратора базы данных имеется множество инструментов, без которых невозможна автономная база.

- **Оптимизатор запросов.** Позволяет эффективно выполнять даже плохо написанные запросы, динамически менять планы их выполнения при изменении статистики. Оптимальный план выполнения запроса должен быть построен быстро, что невозможно сделать при проверке всех вариантов, поэтому для детального анализа и оптимизации запросов и настройки базы используется Oracle Tuning Advisor, применяющий все накопленные знания и наработки по настройке базы, позволяющие создать профиль, который, в частности, укажет оптимизатору, как надо изменить план запроса.
- **Советчику (Advisors).** В современных СУБД имеются советчики, генерирующие рекомендации по улучшению структуры и работы базы и помогающие в ее настройке. Например, Access advisor в Oracle позволяет улучшить структуру базы данных для повышения ее производительности, а In-memory advisor на основе анализа реальной нагрузки советует, какие таблицы и разделы следует поместить в колонный кэш в оперативной памяти. У Oracle сегодня есть более дюжины советчиков, результаты работы которых можно будет использовать в будущих механизмах управления автономными базами.
- **Механизмы автономного управления.** В любой СУБД есть свои элементы (кэши, файлы, алгоритмы сжатия и т. д.), которыми надо управлять, однако делать это вручную сложно и долго, и к тому же это создает большие риски. Поэтому чем совершеннее СУБД, тем выше у нее степень автоматизации механизмов автономного управления. Например, для Oracle реализованы: автоматическое управление кэшами оперативной памяти (Automatic Memory Management); автоматическое управление сегментами отката (Automatic Undo Tablespace); автоматическая система сбора, проверки и публикации статистики; автоматическое ведение температурной карты данных и автоматическая выгрузка и загрузка данных в кэш в оперативной памяти на ее основе; Automatic Data Optimization — автоматическая реализация механизмов поддержки ILM (information lifecycle management). К этой же группе можно отнести механизмы умного сжатия данных (Advanced Compression) и автоматического определения оптимальной степени параллели-

лизма (Automatic Parallel Degree), а также механизм автоматического создания и поддержки резервной базы (Automatic Standby Management), все вместе значительно упрощающие работу администратора. Здесь же следует упомянуть и об [I] адаптивных планах [S]. Суть их в том, что оптимизатор на основе собранной статистики строит оптимальный план выполнения запроса, но если статистика устарела и это обнаружится только во время выполнения запроса, то оптимизатор может на ходу перестроить план выполнения запроса и пометить для других запросов, что надо использовать другую статистику.

ASM позволяет не управлять непосредственно файлами и томами, а передать СУБД некоторый объем сырого дискового пространства. Этот механизм осуществляет балансировку ввода-вывода и обеспечивает зеркалирование кусков логических файлов на разные диски для защиты от сбоев.

- **Механизмы мониторинга.** Возможности мониторинга работы СУБД и инфраструктуры в той или иной степени реализованы у многих производителей. В автономных базах стоит задача автоматического применения таких средств, анализа результатов их работы и автоматического принятия корректирующих действий при выявлении деградации. Например, во время работы СУБД Oracle в специальном репозитории (Automatic Workload Repository, AWR) собирается и анализируется статистика работы экземпляра СУБД, на основе которой монитор ADDM (Automatic Data Diagnostic Monitor) предупреждает администратора о неполадках и негативных тенденциях, давая при этом рекомендации по их устранению. По сути, в алгоритмах ADDM заложен весь опыт Oracle по настройке баз данных. В дополнение к AWR в режиме онлайн собирается информация о всех выполняющихся сессиях (что делает сессия, время работы и ожидания и пр.), а механизм ASH Analytics (Active Session History) анализирует эту информацию для выявления проблем. Механизм Automatic SQL Monitor позволяет в режиме реального времени анализировать выполнение SQL-запросов, определять, на что и почему расходуется время, наблюдать статистику выполнения (например, степень параллелизма, количество обработанных строк на каждом этапе плана выполнения, план запроса и т. д.). Мониторингом

и оптимизацией работы кластеров занимается механизм AHF (Autonomous Health Framework), предоставляющий утилиты, которые дают возможность в режиме онлайн анализировать состояние аппаратуры кластеров базы данных, предпринимая заданные действия в случае ухудшения работы.

- **Прочие инструменты.** Прежде чем вносить изменения в базу или менять ИТ-инфраструктуру, обычно производят тестирование для оценки их влияния на производительность и работоспособность приложения. В случае Oracle инструменты DB Replay и SPA (SQL Performance Analyzer) позволяют захватить реальную нагрузку на работающей базе и в тестовой или (в случае SPA Quick Check) основной среде проанализировать и сравнить результаты, а затем принять решение о допустимости изменений.

Для облегчения управления и сопровождения множества автономных баз они могут реализоваться в виде подключаемых баз (Pluggable Database, PDB), а для управления множеством независимых PDB как единым целым предназначена мультиарендная (multitenant) архитектура. Иногда требуется переместить PDB из одной контейнерной базы в другую, что бывает нужно для изменения версии или развертывания копии PDB с целью независимого использования. Пока пользователь работает с базой, она перемещается или копируется в другую контейнерную базу данных, а поскольку за время перемещения исходная база изменялась, происходит автоматический «накат» этих изменений и пользователь всегда получает согласованную копию.

Во время работы базы данных и приложений создается огромное количество журналов, которые очень трудно все изучать при поиске ошибки, анализе работы базы или приложения, а добавление в систему новых узлов и программных компонентов требует корректировки системы анализа журналов. Компонент OMC (Oracle Management Cloud) Log Analytics автоматизирует эту работу, собирая множество журналов из разных систем, выбирая из них сообщения и позволяя быстро анализировать эти сообщения в разных разрезах: по продукту, узлу, времени возникновения, критичности сообщения и т. п.

Все это далеко не полный список «кирпичиков», необходимых для автономной базы, но главное, что все эти инструменты уже имеются в арсенале админист-

ратора, который пока использует их по своему усмотрению. В автономной базе эти инструменты применяются автоматически. В современном автомобиле также уже имеется масса средств автоматизации (круиз-контроль, GPS-навигатор, ABS, предупреждение об уходе с полосы, датчики давления в шинах и т. д.), которые водитель использует либо нет при движении, однако пассажир («водитель») беспилотной повозки о них может и не догадываться.

МАШИННОЕ ОБУЧЕНИЕ

Автономная база данных невозможна без машинного обучения, уже используемого в OMC, компонентах AHF, системах защиты базы от внешних и внутренних вторжений, в работе служб технической поддержки для выявления и исправления ошибок и т. п. Например, в компоненте Log Analytics использование алгоритмов машинного обучения позволяет объединять (кластеризовать) похожие сообщения, что упрощает анализ непрерывно собираемых журналов и позволяет быстро находить источник возникающих проблем.

В других компонентах применяется байесова сеть для поиска причин возникновения нештатной ситуации или появления негативных тенденций. Как правило, во всех таких компонентах автоматически строится базовая модель нормального поведения системы, затем при работе системы выявляются отклонения от базовой модели и производится их анализ, после чего выявляются либо тенденции к ухудшению работы, либо причины отклонений от модели, которая может быть затем скорректирована. Так, например, работает AHF, а в механизме Trace File Analyzer выявляются аномалии в работе системы, определяется список файлов, которые нужно отправить в службу технической поддержки для поиска причин возникновения ошибок. В составлении алгоритмов работы все этих компонентов участвуют эксперты по машинному обучению, а модели совершенствуются и обновляются.

Многие компоненты OMC построены на основе механизма контроля отклонений: система создает базовый профиль работы множества баз данных, компьютеров, приложений и сравнивает непрерывно собираемые текущие характеристики работы с характеристиками базовых профилей (базовой моделью поведения). Если выявляются аномальные отклонения, то система либо сообщает



Рис. 2. Архитектура ADWS

о них администратору, либо реагирует автоматически, выполняя корректирующие действия.

В частности, IT Analytic — это компонент ОМС, который позволяет собирать информацию о работе различных баз, автоматически выявляя базы с деградацией производительности или SQL-запросы и базы с нестабильным временем выполнения. В работе IT Analytics также используются модели, полученные при машинном обучении.

AUTONOMOUS DATAWAREHOUSE SERVICE

Специализированный сервис «Автономное хранилище данных» (Autonomous DataWarehouse Service, ADWS) представляет собой первую реализацию автономной базы данных Oracle. Сервис позволяет автоматически создать базу для витрины или хранилища данных (рис. 2). Работать с ADWS можно через SQL Developer, интегрируя с ним другие системы через различные интеграционные сервисы от Oracle или других производителей, причем эти интеграционные сервисы могут работать как в облаке, так и в локальном ЦОД. Прежде чем загружать данные в таблицы созданного хранилища, их предварительно помещают в облачную систему хранения Oracle Object Storage или Amazon S3. В базе данных имеется процедура для загрузки этих данных или для работы с ними как с внешними таблицами.

Для анализа загруженных в таблицы данных можно использовать любые ин-

струменты бизнес-аналитики, способные работать с SQL*Net как в облаке, так и в локальном ЦОД. Сервис снабжен веб-консолью и встроенным инструментом для анализа данных и решения задач машинного обучения.

НОВАЯ РОЛЬ АДМИНИСТРАТОРА

Автономная база данных позволяет автоматизировать рутинные операции, освобождая администратора для выполнения творческой работы, например поддержки большого числа баз данных. Сегодня интерес к автономным базам данных проявляют прежде всего разработчики приложений и аналитики, занятые созданием витрин и некритичных приложений: опытные администраторы все еще работают лучше — например, поддержку высоконагруженных бизнес-критичных приложений или загрузку данных в хранилище с миграцией данных из других систем никто, кроме администратора, выполнить не сможет.

Предстоит еще долгий путь совершенствования алгоритмов самоуправления баз данных, однако администраторам придется постепенно перестраивать свой подход к работе — они будут меньше внимания уделять поддержке работающих баз и больше времени тратить на освоение новых продуктов и технологий, а также на разработку архитектуры, планирование и моделирование данных, настройку кода приложений, управление жизненным циклом данных и обеспечение безопасности. Немаловажно то, что у адми-

нистратора появится больше времени на общение с разработчиками приложений и бизнесом для оперативного реагирования на запросы.

Автономная база данных облегчает работу администратора, снижает риски, повышает качество функционирования СУБД и приложений, позволяя сократить время развертывания систем и обеспечивая более высокую надежность и безопасность, чем традиционные базы. Пользователь при заказе базы данных ставит цель, а автономная база берет на себя достижение этой цели. При этом не требуется создавать индексы, разделы, материализованные представления, табличные пространства и т. д. Все это кардинально меняет подход пользователей к решению бизнес-задач и сокращает сроки создания приложений, позволяя сосредоточиться на развитии бизнеса и выводе новых услуг и продуктов на рынок.

Автономная база данных должна использовать механизмы облака и СУБД, чтобы автоматически выполнять большинство уже имеющихся в СУБД функций обеспечения самоуправления, самозащиты и самовосстановления. Однако следует еще раз подчеркнуть, что автономные базы данных — не продукт, а концепция, в соответствии с которой компания Oracle выпустила первый продукт данного класса, ADWS. ■

Марк Ривкин (mark.rivkin@oracle.com) — директор по технологическому консалтингу, Oracle CIS (Москва).