

СУБД Postgres Pro — сделано в России

Компания Postgres Professional создана российскими разработчиками свободно распространяемой системы управления базами данных PostgreSQL. Среди основателей — три ведущих разработчика и активных участника международного сообщества. В 2016 г. компания выпустила собственную СУБД Postgres Pro, представляющую собой улучшенную версию PostgreSQL. СУБД Postgres Pro включена в единый реестр отечественного ПО в числе первых продуктов. Об особенностях нового продукта рассказывает Иван Панченко — заместитель генерального директора Postgres Professional.



Иван Панченко — заместитель генерального директора Postgres Professional.

SN. История открытого проекта PostgreSQL начиналась в Калифорнийском университете Беркли. В какой мере разработку, которой занимается ваша компания, можно считать российской и какая часть функционала контролируется/развивается вашей компанией?

И.П. Примерно 30 лет назад в Калифорнии были начаты первые работы, которые привели к возникновению этой базы данных. Фактически СУБД PostgreSQL, какой мы её знаем, начинается свою историю с 1996 г., когда она начала разрабатываться международным сообществом. В числе первых к нему присоединились Вадим Михеев из Красноярска — сейчас он является вице-президентом Goldman Sachs и Олег Бартунов — основатель нашей компании. За 20 лет в число основных разработчиков Постгреса вошли еще два россиянина — Александр Коротков и Федор Сигаев — тоже наши сотрудники. Российский вклад в программный код PostgreSQL составляет до 30%, а российское сообщество пользователей и разработчиков является одним из крупнейших в мире. Это подтверждается хотя бы тем фактом, что российские конференции PgConf.Russia 2015 и 2016 г. били мировые рекорды по численности участников.

SN. PostgreSQL предлагается как альтернатива западным базам данных. Когда корпоративный клиент переходит на другую СУБД, он, прежде всего, думает о гарантированной поддержке приложений в случае возникновения сбоев или отказов. Какими силами организуется поддержка клиентов на территории России, сколько человек непосредственно занимается разработкой?

И.П. Техподдержка программного продукта — необходимое условие для его использования в серьезных задачах. Техподдержка — это не некоторый, заранее определенный комплекс работ и мероприя-

тий, а своего рода страховка на случай происшествия, с которым самостоятельно не справиться, это возможность оперативно подключить грамотный профессиональный ресурс, который решит любую возникшую проблему. Наличие профессиональной техподдержки — всегда сложный вопрос, когда прикладные информационные системы строятся на базе СПО. Разработчики таких систем вынуждены брать на себя обязательства по поддержке разработанного не ими СПО, а это несет дополнительные риски.

Наша компания берет на себя техподдержку PostgreSQL. Непосредственно этим занимается около 20 сотрудников. Еще 30 разработчиков Постгреса готовы при необходимости принять участие в решении задач по поддержке. Также в разных регионах работают специалисты на удаленном доступе. Кроме того, Postgres Professional активно развивает сеть партнеров по всей стране, в неё входит уже больше 30 компаний.

Таким образом, мы предлагаем клиентам круглосуточную техподдержку с часовым временем реакции на инцидент. Это очень существенно для тех, кому важно непрерывное функционирование систем. Ведь база данных — ключевой компонент, её поломка — проблема для всего ПО компании, из-за неё могут возникнуть проблемы в бизнесе.

SN. Поддерживается ли колоночное хранение таблиц?

И.П. В Постгресе существует расширение `columnstore`, которое реализует колоночное хранение данных. Кроме того, есть такие продукты, как CitusDB и GreenPlum, представляющие собой массивно-параллельные СУБД, созданные на основе Постгреса, они тоже поддерживают такой вид хранения.

SN. Поддерживается ли хранение и обработка таблиц in-мемори?

И.П. Постгрес устроен так, что, если свободная память имеется, при правильной настройке она активно используется всегда. Однако для максимальной производительности часто требуются специализированные алгоритмы, оптимизированные под работу в оперативной памяти. Тут основная сложность — обеспечить целостность данных, надежность и изоляцию транзакций, присущие классическим СУБД. Кроме того, чтобы in-мемори решение стало полноценным компонентом СУБД, оно должно быть готово к совместному использованию в одном запросе in-мемори и обычных данных. И мы планируем реализовать это. В команде Postgres Professional есть опыт создания

высокопроизводительных in-мемори решений, мы ведем такую разработку для PostgreSQL.

SN. Развиваются ли параллельные методы обработки таблиц на внешних хранилищах (для поддержки OLAP-приложений), например, по типу Oracle Exadata или за счет возможности динамической организации субкластеров по примеру решений EMC+SAS? Или что-то еще?

И.П. Да, в последние годы в PostgreSQL активно развивается направление, связанное с параллельной обработкой запросов. В списке новшеств, ожидаемых в версии 9.6, которая должна выйти в конце 2016 года, — параллельные последовательные сканы, объединения и агрегаты. Это позволит и в кластерной среде также достигать лучших результатов. В нашей компании ведется работа над параллелизацией создания индексов.

SN. Что лицензируется, на чем строится ваш бизнес?

И.П. В основном, бизнес строится на оказании различных услуг, связанных с нашей СУБД — это типично для компаний, чей продукт основывается на Open Source. Однако мы пошли дальше. На базе PostgreSQL создали собственный продукт — российскую СУБД Postgres Pro, которая вошла в единый реестр отечественного ПО. Её можно использовать как в свободно распространяемом, так и в лицензируемом вариантах. Кроме того, мы разрабатываем Enterprise-версию, в которую войдут функциональность, востребованная в крупных высокопроизводительных инсталляциях, адаптация к современным облачным платформам и многое другое.

SN. Могут ли работать OLTP- и OLAP-приложения с одной и той же СУБД или с одной и той же лицензией, если она нужна?

И.П. Все зависит от нагрузки. Обычно разный характер нагрузки в OLTP и OLAP-приложениях требует различной конфигурации СУБД и аппаратной платформы. Поэтому в случае, если используется PostgreSQL, организуют репликацию между узлом, принимающим OLTP-нагрузку, и узлами, на которых выполняются OLAP-запросы. Можно строить и более сложные системы, комбинируя PostgreSQL со специальными аналитическими продуктами.

SN. Что предлагается для поддержки больших данных?

И.П. Для работы с большими данными требуются кластерные решения. Примерами являются ранее упомянутые CitusDB и GreenPlum, созданные на основе PostgreSQL. Эти решения сейчас дос-

тупны как open source. PostgreSQL умеет работать и с Hadoop по технологии Foreign data wrapper (FDW). В рамках основной функциональности PostgreSQL как РСУБД тоже разрабатывается ряд возможностей для работы с большими данными, к ним относят, в частности, появившиеся в версии 9.5 BRIN-индексы, многоуровневые агрегаты (GROUPING SETS) и TABLESAMPLE.

SN. Развивается ли поддержка специализированных команд, интегрированных в ряде процессоров для поддержки OLAP-запросов?

И.П. Существует расширение pgstrom, использующее мощности графических процессоров (GPU) для решения различных задач СУБД, в частности, агрегации в OLAP. По нашему мнению, у него есть большие перспективы, особенно если будет повышаться скорость работы GPU с оперативной памятью.

SN. Предлагаются ли методы маскирования/криптографии строк/колонок в целях защиты персональных данных?

И.П. Задача шифрования отдельных колонок или строк решается средствами расширения PostgreSQL pgcrypto. Мы взаимодействуем с российскими производителями криптографических систем, чтобы совместно представить решение, сертифицированное по отечественным стандартам.

SN. На основе каких разработок происходит управление ключами?

И.П. Это, действительно, важный вопрос. Pgcrypto, например, не содержит средств для управления ключами, оставляя тут свободу прикладному разработчику. Отечественные программисты будут предлагать свои системы управления ключами.

SN. Что делается для сертификации PostgreSQL по российским требованиям?

И.П. Наша СУБД Postgres Pro проходит сейчас сертификацию ФСТЭК по 5-му уровню СВТ и 4-му уровню отсутствия НДВ. Это облегчит построение на PostgreSQL систем работы с персональными данными и с конфиденциальной информацией. Некоторые предыдущие версии PostgreSQL также сертифицировались ФСТЭК вплоть до уровня, требующегося для работы с секретными данными.

SN. Используются ли какие-либо методы компрессии/дедупликации? Сравнивались ли

(продолжение, начало – стр. 21)

производительных вычислительных систем, аналитики больших данных, приложений новых поколений и рабочих сред с открытым исходным кодом, например, Hadoop и Apache Spark.

И.П. Если говорить о миграции с Oracle, то средства автоматизации есть. Больше всего баз были перенесены с помощью инструмента oga2pg, разработанного на-

производительных вычислительных систем, аналитики больших данных, приложений новых поколений и рабочих сред с открытым исходным кодом, например, Hadoop и Apache Spark.

И.П. В PostgreSQL есть встроенный алгоритм сжатия, применяемый к полям большого размера. Он подключается автоматически. Поскольку время от времени мы получаем запросы от клиентов на большую степень сжатия, то начали вести разработки по новым алгоритмам компрессии в таблицах и индексах. Уже выполнена работа по дедупликации в B-Tree индексах.

SN. Проводилось сравнительное тестирование PostgreSQL по производительности?

И.П. Тесты производятся постоянно. Но в большинстве случаев их целью является не сравнение, а поиск оптимальных алгоритмов, конфигураций, изучение особенности работы в тех или иных условиях, под разной нагрузкой. Что касается сравнительного тестирования разных СУБД, то, во-первых, сложно поставить равные СУБД в равные условия, во-вторых, например, лицензионное соглашение Oracle запрещает публиковать результаты. Приведу два таких примера. Долгое время считалось, что в задачах IC PostgreSQL заведомо уступает Microsoft SQL Server (что неудивительно, т.к. именно на эту СУБД изначально была рассчитана структура данных в IC). Однако к нашему времени, к 2015 году, разрыв сократился (в частности, об этом тогда же говорили на конференции INFOSTART в 2015 году), и результаты практически равны.

Еще один пример – случай миграции системы документооборота Правительства Московской области с Oracle на PostgreSQL. По свидетельствам заказчика (об этом говорилось недавно на конференции Web Scale IT), скорость работы системы возросла, зависания прекратились, притом, что количество задействованных процессорных ядер стало меньше. При этом три сервера под управлением Oracle RAC были заменены на конфигурацию PostgreSQL из мастера и двух реплик с балансировщиком.

SN. Предлагаются ли какие-либо методы для автоматизированной миграции на PostgreSQL или все делается вручную?

И.П. Если говорить о миграции с Oracle, то средства автоматизации есть. Больше всего баз были перенесены с помощью инструмента oga2pg, разработанного на-

производительных вычислительных систем, аналитики больших данных, приложений новых поколений и рабочих сред с открытым исходным кодом, например, Hadoop и Apache Spark.

И.П. Если говорить о миграции с Oracle, то средства автоматизации есть. Больше всего баз были перенесены с помощью инструмента oga2pg, разработанного на-

И.П. Тесты производятся постоянно. Но в большинстве случаев их целью является не сравнение, а поиск оптимальных алгоритмов, конфигураций, изучение особенности работы в тех или иных условиях, под разной нагрузкой. Что касается сравнительного тестирования разных СУБД, то, во-первых, сложно поставить равные СУБД в равные условия, во-вторых, например, лицензионное соглашение Oracle запрещает публиковать результаты. Приведу два таких примера. Долгое время считалось, что в задачах IC PostgreSQL заведомо уступает Microsoft SQL Server (что неудивительно, т.к. именно на эту СУБД изначально была рассчитана структура данных в IC). Однако к нашему времени, к 2015 году, разрыв сократился (в частности, об этом тогда же говорили на конференции INFOSTART в 2015 году), и результаты практически равны.

И.П. Тесты производятся постоянно. Но в большинстве случаев их целью является не сравнение, а поиск оптимальных алгоритмов, конфигураций, изучение особенности работы в тех или иных условиях, под разной нагрузкой. Что касается сравнительного тестирования разных СУБД, то, во-первых, сложно поставить равные СУБД в равные условия, во-вторых, например, лицензионное соглашение Oracle запрещает публиковать результаты. Приведу два таких примера. Долгое время считалось, что в задачах IC PostgreSQL заведомо уступает Microsoft SQL Server (что неудивительно, т.к. именно на эту СУБД изначально была рассчитана структура данных в IC). Однако к нашему времени, к 2015 году, разрыв сократился (в частности, об этом тогда же говорили на конференции INFOSTART в 2015 году), и результаты практически равны.

SN. PostgreSQL поддерживает интеграцию с CXД BAUM для создания консистентных снимков. Какая еще интеграция развивается и для каких CXД?

И.П. Есть несколько проектов на эту тему, говорить о них пока преждевременно, мы обязательно будем выпускать новости, когда придет время рассказывать об этих проектах.

SN. Какие важные особенности/преимущества Вы бы отметили при интеграции с IBS СКАЛА?

И.П. Это надежная система, построенная из качественных компонент, с удобной системой виртуализации, уже давно проверена и подтверждена её хорошая совместимость с PostgreSQL. IBS – надёжный партнер, с которым комфортно работать и которой обладает всей нужной инфраструктурой для создания, испытания и обеспечения эксплуатации программно-аппаратных комплексов.

SN. Кто использует ваши разработки?

И.П. Среди крупных российских пользователей PostgreSQL – система межведомственного электронного взаимодействия, Правительство Московской области, силовые структуры. Эта СУБД используется в таких компаниях и организациях, как Яндекс, Авито, Рамблер, Манго Телеком, МИА “Россия Сегодня”, Skype, Instagram, MasterCard, Huawei, Sony, Hitachi, Tripadvisor, Федеральная резервная система США, ВМФ США, государственные органы Франции, структуры Европейского союза. Подробнее: <https://postgrespro.ru/> ■

– IBM Spectrum Conductor with Spark – интегрирующее Apache Spark ПО. Оно упрощает внедрение Apache Spark – платформы для аналитики больших данных с открытым исходным кодом – и позволяет на 60% ускорить получение результатов обработки информации (March 28, 2016 STAC Securities Technology Analysis Center, Spark Resource Manager– Phase 1);

– IBM Spectrum LSF – является универсальным ПО управления нагрузками. Гибкий и удобный интерфейс помогает ускорить исследования и проектирование в 150 раз (внутр. тестирование IBM: 16 физич. серверов x86_64; CPU – Intel Xeon CPU E5-2670 @ 2.60GHz, 16 ядер на каждом сервере; ОП 132 GB; OS: RHELS 6.4), контролируя затраты путем продвинутого распределения ресурсов и их рационального использования.