

# Почему PostgreSQL не является аналогом СУБД Oracle

С 1 января 2016 года в России вступает в действие постановление РФ "Об установлении запрета на допуск программного обеспечения, происходящего из иностранных государств, для целей осуществления закупок для обеспечения государственных и муниципальных нужд". В соответствии с этим постановлением с 1 января 2016 года формируется реестр российского программного обеспечения (ПО) и госорганизации могут покупать ПО иностранных фирм (например, Oracle) только в случае, если в реестре нет российского аналога этому ПО, то есть, как написано в пункте б постановления, купить иностранное ПО можно только если:

программное обеспечение, сведения о котором включены в реестр и которое соответствует тому же классу программного обеспечения, что и программное обеспечение, планируемое к закупке, **по своим функциональным, техническим и (или) эксплуатационным характеристикам не соответствует установленным заказчиком требованиям** к планируемому к закупке программному обеспечению.

Что касается СУБД, то скорее всего в реестр будет внесена СУБД с открытым кодом PostgreSQL представляемая на рынке компанией Postgres Professional. Как же можно обосновать, что PostgreSQL не является аналогом СУБД Oracle и по функциональным, техническим и эксплуатационным характеристикам не соответствует требованиям заказчика?

При выборе СУБД специалисты анализируют следующие основные характеристики ПО:

1. Производительность: количество транзакций в секунду, время отклика, время выполнения сложных SQL запросов
2. Надежность: вероятность и продолжительность простоев, вероятность потери данных
3. Безопасность и защита данных
4. Управляемость: стоимость администрирования, эффективность использования ресурсов оборудования
5. Масштабируемость: обеспечение производительности при увеличении нагрузки и числа пользователей, возможность увеличения объема используемых аппаратных ресурсов
6. Работа с большими объемами данных и большим числом пользователей
7. Наличие и уровень технической поддержки
8. Совокупная стоимость владения
9. Зрелость СУБД и перспективы ее развития
10. Поддержка новых современных технологий

По каждому из этих параметров PostgreSQL значительно отстает от СУБД Oracle. Другими словами, если Вы выбираете СУБД для создания небольшой информационной системы, где простои допустимы, нет конфиденциальных данных, к скорости работы высоких требований не предъявляется, количество пользователей и объем данных невелики, то PostgreSQL вполне подходит для Вашего решения. Если же хоть одна из характеристик, перечисленных выше, важна для Вас, то следует рассмотреть использование СУБД Oracle.

Чтобы обосновать это сравним СУБД PostgreSQL и Oracle в соответствии со списком перечисленных характеристик. Опции и компоненты СУБД Oracle, использованные в этом сравнении, более подробно описаны в документации, а так же в каталоге программных продуктов Oracle.

## 1. Производительность: количество транзакций в секунду, время отклика, время выполнения сложных SQL запросов

Данный критерий является наиболее очевидным и традиционным при сравнении различных СУБД. Однако необходимо понимать, что производительность ИТ систем зависит от всех компонент архитектуры – от систем хранения до непосредственно бизнес-приложений. СУБД находится в середине этого стека, но может оказывать существенное влияние на все компоненты, например, за счет эффективного использования аппаратных ресурсов, или эффективного управления соединениями приложений и кешированием. Немаловажным фактором при выборе СУБД так же является наличие и возможности инструментов отладки и оптимизации производительности, управления распределением ресурсов.

Существует большое количество тестов - бенчмарков, позволяющих сравнить разные СУБД в условиях одинаковой нагрузки и окружения. Наиболее доверенными являются тесты, проводимые независимыми компаниями или консорциумами, результаты которых, наряду с условиями тестирования и нагрузочными сценариями, публикуются открыто.

В СУБД PostgreSQL отсутствуют аналоги технологий Oracle:

- Real Application Clusters – технология балансировки нагрузки по нескольким физическим узлам кластера, которые работают как единая СУБД и обеспечивают большую совокупную вычислительную мощность, чем максимальная мощность каждого отдельного узла;
- GRID – перераспределение вычислительных мощностей и узлов для обслуживания критичных в данный период сервисов и приложений;
- Развитые средства диагностики, оптимизации, проактивного мониторинга и самонастройки, включая средства управления планами выполнения запросов (декларативные подсказки оптимизатору SQL, профайлы, динамические планы SQL);
- Partitioning – технология секционирования больших реляционных таблиц и возможность интеллектуальной оптимизации запросов к этим таблицам с целью существенного сокращения объема операций ввода-вывода, и как следствие общего времени выполнения запросов. Данный функционал присутствует в PostgreSQL, но реализован он на основе триггеров настраиваемых вручную, то есть требует значительных ресурсов по сопровождению, в то время как в Oracle эта функциональность автоматизирована, вплоть до автоматического создания секций при необходимости;
- In-Memory – технологии обработки табличных данных в оперативной памяти включая поколоночные алгоритмы сжатия при хранении, низкоуровневую оптимизацию обработки данных с использованием SIMD инструкций процессора и специальных алгоритмов, таких как Bloom Filters и Vector GroupBy;
- Параллельная обработка запросов, в том числе DML инструкций;
- Real Application Testing – технологии автоматизированного нагрузочного тестирования на уровне СУБД с захватом и воспроизведением реальной прикладной нагрузки позволяющие проводить анализ деградации производительности при обновлениях, изменениях параметров, росте объемов БД и т.д.
- Resource Management – средства приоритезации и распределения вычислительных ресурсов между приложениями на основе декларативных политик;
- Oracle Exadata – специализированные программно-аппаратные комплексы, интегрирующие в себе все лучшие технологии Oracle обеспечения производительности и отказоустойчивости, такие как GRID, RAC, In-Memory и др. Данные системы отличаются оптимизированной архитектурой, позволяющей распределять нагрузку между компонентами и, за счет синергетического эффекта применяемых технологий, обеспечивать производительность, недоступную системам, построенным по традиционной архитектуре;

Из-за низкой производительности СУБД PostgreSQL никогда не участвует в независимых тестах производительности, таких как TPC ([www.tpc.org](http://www.tpc.org)).

## 2. Надежность: вероятность и продолжительность простоев, вероятность потери данных

Современные базы данных практически всегда являются критичным компонентом ИТ окружения. От их надежности, работоспособности и бесперебойности зачастую зависят целые компании и предприятия. При этом понятие надежности СУБД включает в себя не только защиту от программных и аппаратных сбоев, но и возможность восстановления БД после логических ошибок или намеренной порчи данных, обеспечение защиты БД при обновления приложений, защита и валидация клиентских подключений. Важным фактором является так же продолжительность работ по восстановлению и администрированию. Все эти факторы обычно описывают формальными показателями RTO (recovery time objective) – время требуемое на восстановление, и RPO (recovery point objective) – объем данных, который может быть потерян в ходе восстановления. Сочетание технологий Oracle позволяет свести оба этих показателя к нулю, обеспечив нулевую потерю данных за счет средств синхронной репликации и нулевое время восстановления сервиса за счет средств active-active кластеризации.

В PostgreSQL отсутствуют:

- Real Application Clusters – технология active-active кластеризации на уровне СУБД, обеспечивающая защиту от серверного сбоя, перевод новых транзакций на работающие узлы кластера, а так же автоматический повтор прерванных транзакций;
- Data Guard Far Sync – технология обеспечения катастрофоустойчивости путем синхронизации копии БД на удаленной площадке в режиме реального времени без влияния на производительность и обеспечением нулевой потери данных. Технология Oracle Data Guard так же включает в себя средства автоматического переключения приложений и пользователей на резервную СУБД в случае выхода из строя основной, средства автоматической проверки целостности БД и восстановления в случае обнаружения ошибок;
- Recovery Manager – богатый инструмент создания резервных копий СУБД;
- Multitenant – технология позволяющая консолидировать множество БД в едином контейнере, консолидируя так же задачи по управлению ими;
- Flashback – технологии, обеспечивающие защиту от человеческих ошибок со стороны администраторов СУБД, пользователей, программистов, а так же от логических ошибок приложения;
- Automatic Storage Management – технология обеспечивающая защиту от выхода из строя медиа носителей, балансировка нагрузки ввод/вывода на устройства хранения;
- Edition based redefinition – средства обеспечения обновления приложений без прерывания работы пользователей;
- Online redefinition – возможность администрирования СУБД без прерывания работы пользователей и приложений;
- Real Application Testing – технологии автоматизированного нагрузочного тестирования при обновлениях СУБД и/или платформы;
- Enterprise Manager и OEM Packs – развитые средства автоматизированной диагностики, управления жизненным циклом и конфигурациями СУБД, автоматизации рутинных задач администрирования;
- Transparent Application Failover, Fast Application Notification, Application Continuity – набор технологий, работающих на уровне клиентского драйвера СУБД, и обеспечивающих прозрачное переключение приложения на резервный узел в случае выхода из строя основного.

### 3. Безопасность и защита данных

Понятие защиты данных так же включает в себя множество аспектов: от защиты персональных данных в БД от доступа извне, до периодического мониторинга выделенных пользователям СУБД прав и ролей. Правильно организованная политика защиты данных требует журналирования операций доступа к данным, а так же средств предотвращения несанкционированного доступа, средств единого управления ключами шифрования и доступа и т.д.

У PostgreSQL отсутствуют:

- Сертификация по ФСТЭК как средства хранения персональных данных;
- Database Vault – средства защиты БД от привилегированных пользователей;
- Advanced Security – шифрование данных на средствах хранения, при передаче по сети, а так же в резервных копиях и выгрузках;
- Database Masking and Subsetting – интеллектуальное маскирование и усечение данных при создании копий БД, например для тестирования и разработки;
- Database Audit Vault – регулярный контроль доступа и выявление избыточных привилегий пользователей СУБД;
- Database Firewall – мониторинг и перехват опасного SQL трафика, например, для предотвращения SQL injection атак;
- Key Vault – единое средство управление ключами доступа СУБД и шифрования данных.

### 4. Управляемость: стоимость администрирования, эффективность использования ресурсов оборудования

Сопровождение систем баз данных является трудоемкой задачей, требующей развитых инструментов проактивного мониторинга производительности систем в условиях увеличивающихся объемов БД и роста нагрузки на них. Рутинные операции сопровождения, например установка пакетов обновлений, так же требуют автоматизации и инструментов верификации конфигураций.

Кроме того, в целях обеспечения бесперебойности работы критичных систем, работы по сопровождению обычно пытаются уложить в специально выделенный для этого интервал времени – технологическое окно. На длительность этого интервала имеют влияние и объемы БД, и возможность выполнения отдельных работ без прерывания работы систем, и требования обеспечения согласованности отдельных элементов инфраструктуры.

У PostgreSQL отсутствуют:

- Возможность выполнения работ по администрированию без прерывания работы приложений и пользователей как в режиме on-line, так и в режиме "rolling fashion";
- Enterprise Manager, Diagnostic and Tuning Pack – развитые средства администрирования, диагностики, настройки, самонастройки. Отсутствие подобных инструментов не позволит эффективно использовать оборудование, предупредить сбои, обеспечить высокую производительность.
- Enterprise Manager, Database Lifecycle Management Pack – средства автоматизации ежедневных задач по сопровождению и администрированию СУБД;
- Средства самоуправления и адаптации СУБД – специализированные алгоритмы и компоненты, позволяющие СУБД собирать диагностическую и статистическую информацию о ходе выполнения запросов, и на основе этой информации проводить оптимизацию своей работы.

## 5. Масштабируемость: обеспечение производительности при увеличении нагрузки и числа пользователей, возможность увеличения объема используемых аппаратных ресурсов

По мере роста объема данных, увеличения количества пользователей, добавления компонент и приложений возникает необходимость масштабирования систем, то есть увеличения объема программно-аппаратных ресурсов, выделенных им, с целью увеличения производительности. Очевидным, но при этом менее эффективным с точки зрения затрат решением в этом случае является замена серверного оборудования на более производительное. Но этот вариант вертикального масштабирования в большинстве случаев требует длительного периода тестирования и простоя систем, когда СУБД копируется на новый сервер, приложения и пользователи перекоммутируются на новую конфигурацию. Дополнительным недостатком этого решения является невозможность динамического масштабирования ресурсов СУБД, как в сторону увеличения, так и в сторону уменьшения, при периодических, например, сезонных, всплесках нагрузки. Именно для возможности обеспечения динамического горизонтального масштабирования ресурсов, то есть масштабирования путем добавления (или удаления) дополнительных узлов в кластер СУБД Oracle предлагает широкий выбор кластерных технологий, охватывающих весь стек серверного оборудования – от серверов СУБД до систем хранения. Эти технологии позволяют строить распределенные системы обслуживающие десятки тысяч пользователей, обеспечивая при этом высочайшие уровни надежности и защиты.

У PostgreSQL отсутствуют:

- Real Application Clusters – технология active-active кластеризации на уровне СУБД, обеспечивающая равномерную балансировку нагрузки на все узлы кластера, с возможностью динамического добавления и исключения узлов;
- Automatic Storage Management – технология кластеризации устройств хранения, с распределением нагрузки на них, автоматическим перераспределением данных в случае изменения конфигурации;
- Active Data Guard Far Sync – технология распределения нагрузки между основной и резервной базой данных, позволяющая вынести отчетность и аналитику, а так же бэкапирование с основной площадки. Технология обеспечивает синхронную передачу данных без влияния на производительность обработки пользовательских запросов;
- Exadata – специализированный программно-аппаратный комплекс, использующий все перечисленные технологии, и благодаря им допускающий плавное масштабирование в том числе аппаратной части без прерывания работы приложений и пользователей;
- Global Data Services - набор технологий для управления сложными IT ландшафтами и эффективной балансировки пользовательской нагрузки в них.

## 6. Работа с большими объемами данных и большим числом пользователей

Для обеспечения производительности при работе с большими БД (объемом в десятки и сотни терабайт) в СУБД обычно предусмотрены специальные механизмы и алгоритмы. При этом важной проблемой таких БД, помимо времени выполнения запросов на большом объеме данных, является время выполнения работ по администрированию, так как эти операции сильно влияют на производительность СУБД или даже требуют ее остановки, а в ограничения технологического окна они часто не укладываются.

PostgreSQL не поддерживает работу с очень большими БД, обеспечивая эффективную обработку баз данных с объемом не более нескольких терабайт.

У PostgreSQL отсутствуют:

- Технологии, обеспечивающие необходимую производительность не только OLTP нагрузки, но и OLAP – такие как материализованные представления, многомерные OLAP кубы, оптимизация пакетной загрузки, поддержка схем "звезда" или "снежинка" для хранилищ данных, специальные типы индексов, например, BitMap Index;

- Advanced Compression – сжатие данных алгоритмами, наиболее сочетающимися с характеристиками обращения к этим данным (по сочетанию критериев степень сжатия и скорость обращения);
- In-Memory – дополнительный уровень хранения табличных данных в памяти, позволяющий эффективно совмещать в одной системе OLTP и OLAP приложения;
- Heat Map и Information Lifecycle Management – средства автоматического сбора статистики обращения к данным и их модификации, а так же средства декларативного объявления политик перемещения данных по уровням сжатия и/или хранения;
- Multitenant – контейнерная архитектура СУБД, облегчающая процессы консолидации и миграции баз данных;
- Параллельная обработка запросов, в том числе на разных узлах кластера СУБД;
- Enterprise Manager, Diagnostic and Tuning Pack – развитые средства администрирования, диагностики, настройки, самонастройки;
- Возможность выполнения работ по администрированию без прерывания работы приложений и пользователей как в режиме on-line, так и в режиме "rolling fashion";
- Средства оптимизации клиентских подключений к СУБД – пулы соединений, различные виды клиентских драйверов и библиотек;
- Active Data Guard Far Sync – технология репликации данных с промышленной БД на резервную в режиме реального времени без влияния на производительность и нулевой потерей данных;
- Global Data Services – набор технологий для управления сложными ИТ ландшафтами, включающими в себя кластерные и катастрофоустойчивые конфигурации. Эти инструменты позволяют управлять доступом клиентских приложения к сервисам СУБД в зависимости от работоспособности и нагруженности баз данных, взаимного гео-расположения, политик доступа и т.д.

## 7. Наличие и уровень технической поддержки

Работоспособность СУБД гарантирована только при наличии надежной системы технической поддержки, которая работает круглосуточно и оперативно решает проблемы. Выстроить такую службу поддержки с анализом проблем, контролем ответственности, работой в режиме 24x7, приоритизацией работ, возможностью моделирования сбоев, отдельными подразделениями для поиска путей решения проблемы и разработки патчей – сложная и объемная промышленная задача. У PostgreSQL, как open source продукта такой службы нет. Компания Postgres Professional среди своих сервисов заявляет службу техподдержки и возможность внесения исправлений в СУБД PostgreSQL, однако на данном этапе сложно судить о эффективности этого сервиса в силу очевидной малочисленности компании, особенно в сравнении с промышленными масштабами Oracle.

## 8. Совокупная стоимость владения

PostgreSQL не имеет лицензионной стоимости и является абсолютно бесплатным программным обеспечением. Однако полная стоимость владения программным продуктом состоит не только из стоимости лицензии, но так же из стоимости внедрения, сопровождения, техподдержки, масштабирования и множества других затрат, включая расходы на заработную плату DBA и т.д. При этом необходимо корректно сравнивать конфигурации ПО, исходя из требуемой производительности и эффективности использования аппаратных ресурсов. Немаловажным фактором так же является разница в стоимости разработки и сопровождения ПО прикладного уровня, наличие сертификации приложений под конкретную СУБД и необходимость дополнительной оптимизации под ее специфику. Как правило, лицензионная стоимость составляет не самую большую, порядка 10-20%, часть общего ИТ бюджета, в то время как остальные 80-90% - это расходы на сопровождение, разработку, тестирование, внедрение и т.д.

Если сравнивать Oracle и PostgreSQL, необходимо отметить:

- Степень автоматизации задач администрирования у Oracle значительно выше, что снижает затраты на ежедневные задачи по сопровождению, такие как обновления, диагностика, контроль конфигураций и политик доступов и т.д.
- Стек продуктов Oracle включает не только СУБД, но и многое другое технологическое ПО, такое как сервера приложений, интеграционные продукты, операционные системы и т.д. Все эти продукты прединтегрированы между собой, в том числе на уровне конфигураций. Более того, Oracle предлагает уникальные программно-аппаратные комплексы, которые уже совмещают в себе все необходимое серверное оборудование и программное обеспечение, и не требуют продолжительной работы по установке ПО, конфигурации, тестированию и т.д. Все это позволяет значительно, в десятки раз, сократить время внедрения и ввода в эксплуатацию новых систем, что снижает затраты проектов внедрения.
- Эффективность использования аппаратных ресурсов у Oracle выше, а так же Oracle имеет множество технологий повышения производительности обработки данных. В итоге это приводит к тому, что для обеспечения заданной производительности, Oracle потребует меньше аппаратных ресурсов, а значит снижает и капитальные, и операционные затраты, связанные с аппаратной частью ИТ систем;
- В силу широчайшей распространенности, СУБД Oracle сертифицирована подавляющим большинством разработчиков прикладного программного обеспечения, а значит исключает затраты, связанные с адаптацией приложений под СУБД;
- Oracle имеет свои программы обучения и сертификации разработчиков и администраторов, что значительно облегчает поиск персонала необходимой квалификации. Количество таких специалистов на рынке труда всегда стабильно высоко. Кроме того, распространенность продукта и количество имеющихся специалистов снижает необходимость и затраты на консультационные сервисы при внедрении и сопровождении систем.

## 9. Зрелость СУБД и перспективы ее развития



Oracle является признанным лидером в сегменте производителей программного обеспечения СУБД, оказывая существенное влияние своим инновациям на общее направление развития индустрии и программных продуктов всех без исключения вендоров. Это подтверждается независимыми аналитическими отчетами, например Gartner Magic Quadrant for Operational Database Management Systems:

PostgreSQL не представлен в данном отчете, т.к. является open source продуктом. Однако его положение можно оценить исходя из позиций одного из клонов данной СУБД – EnterpriseDB, которая является коммерческим продуктом одноименной компании, имеющим над ядром PostgreSQL множество надстроек в части отказоустойчивости, распределенных вычислений, управляемости и т.д. Компания так же предоставляет сервисы технической поддержки и консалтинга. Но даже при наличии всех этих усовершенствований EnterpriseDB значительно уступает СУБД Oracle.

## 10. Поддержка новых современных технологий

Современные системы управления базами данных - это динамично развивающаяся ИТ область, в которой новые технологии появляются ежегодно. Компания Oracle зачастую является локомотивом таких разработок, предлагая рынку продукты, которые оказывают влияние на развитие индустрии в целом. Многие вендоры СУБД, не имея возможности вкладывать сравнимые ресурсы в исследования, оказываются в роли догоняющих, когда новые функции, реализованные в их продуктах, являются по сути адаптацией или просто копией функционала, уже реализованного в СУБД Oracle, протестированного и проверенного множеством внедрений.

PostgreSQL не имеет разработок, аналогичных:

- Облачной платформе баз данных (DBaaS – Database as a Service) - полнофункциональной СУБД, развернутой в собственных ЦОД-ах Oracle и доступной заказчикам компании по стандартным сетевым протоколам. Данный сервис позволяет значительно снизить затраты на развертывание и сопровождение ИТ окружений, сократить сроки вывода продуктов на рынок, повысить надежность своих систем. Эти технологии включают в себя средства построения гибридных (hybrid) облачных окружений, когда базы данных из частных (private) облачных инфраструктур могут быть легко мигрированы в общедоступное (public) облако;
- In-Memory технологиям, включающим не только возможность буферизации данных в оперативной памяти, но и механизмы сжатия данных при хранении в памяти, интеллектуальную оптимизацию запросов, низкоуровневые алгоритмы обработки ориентированные на оперативную память, включая SIMD инструкции процессоров и т.д.;
- Собственным разработкам серверных процессоров SPARC, имеющим уникальные технологии аппаратного ускорения операций СУБД, сжатия, шифрования и защиты данных;
- Специализированным программно-аппаратным комплексам, обеспечивающим высочайшую надежность и производительность обработки данных, за счет широкого использования кластерной GRID архитектуры;
- Polyglot Persistence – поддержке полуструктурированных (JSON, XML) и неструктурированных (Key-Value, NoSQL) данных как внутри СУБД, так и в других источниках, включая Big Data технологии;
- Полнофункциональной поддержке BigData технологий, включая как программное обеспечение (HDFS, Hive, Hadoop и т.д.), так и интегрированные программно-аппаратные комплексы, реализующие данную архитектуру и обеспечивающие высокопроизводительную обработку сверхбольших объемов данных.