PROFESSIONAL Postgres

Что такое PostgreSQL ?

Иван Панченко
i.panchenko@postgrespro.ru

Что такое PostgreSQL?

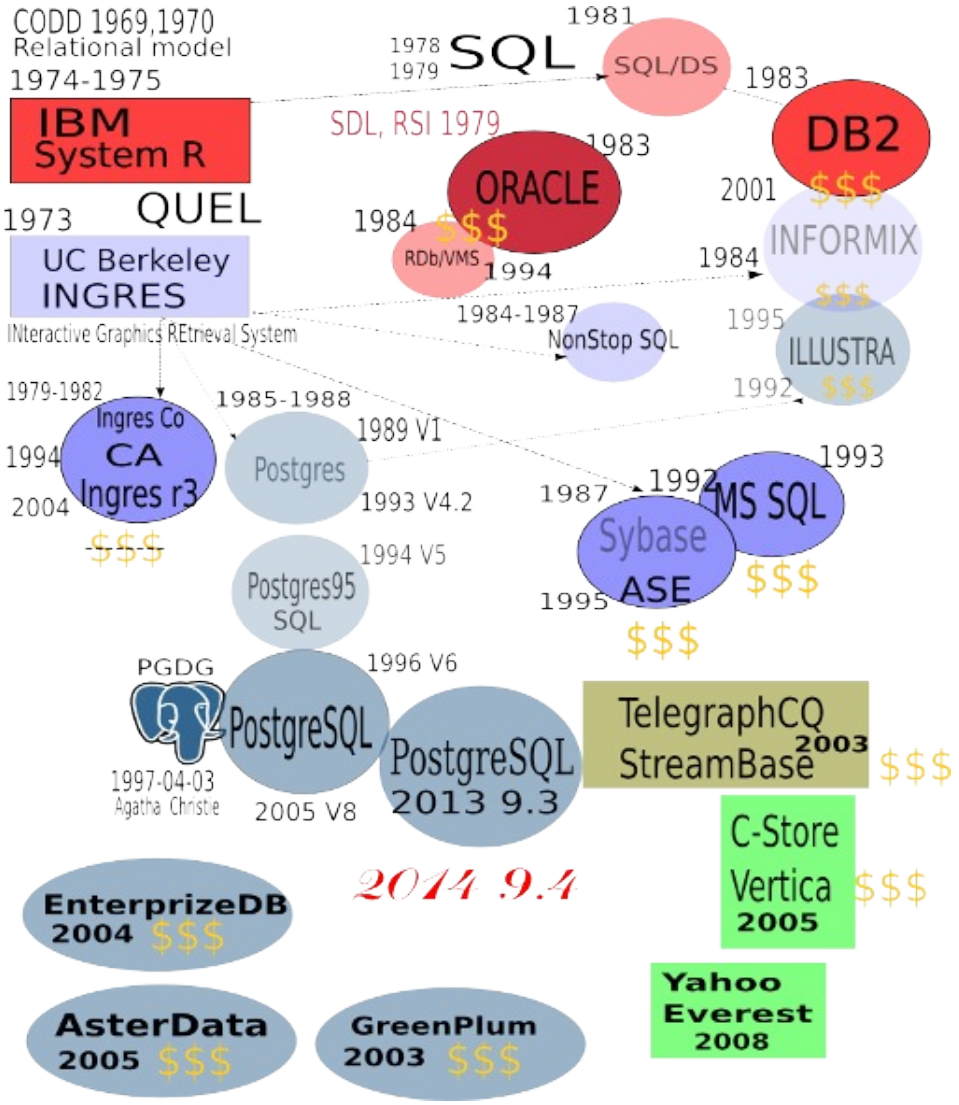
- **PostgreSQL** - это свободно распространяемая объектно-реляционная СУБД (ORDBMS)
- Поддержка ANSI SQL (1992, 1999, 2003, 2011), а также NoSQL (key-value, JSON, JSONB)
- Произношение: **post-gress-Q-L**, **post-gres**, **пост-грес**, **pgsql** (**пэ-жэ-эс-ку-эль**)
- Web: **http://www.postgresql.org**
- Лицензия: BSD, MIT - like

Пользователи PostgreSQL

История PostgreSQL

Michael Stonebreaker
Turing Award, 2015

PostgresPlus (EnterpriseDB)
Bisgres (GreenPlum)
Everest (Yahoo)
AsterData (Teradata)
JustOneDB,
HadoopDB (Hadapt)



PROFESSIONAL Postgres

Развитие PostgreSQL

Базовая функциональность
- JDBC
- MVCC
- Optimizer Stats
- PL/pgSQL

Стабилизация
- Исправление сбоев в работе
- Очистка кода
- Культура совершенства

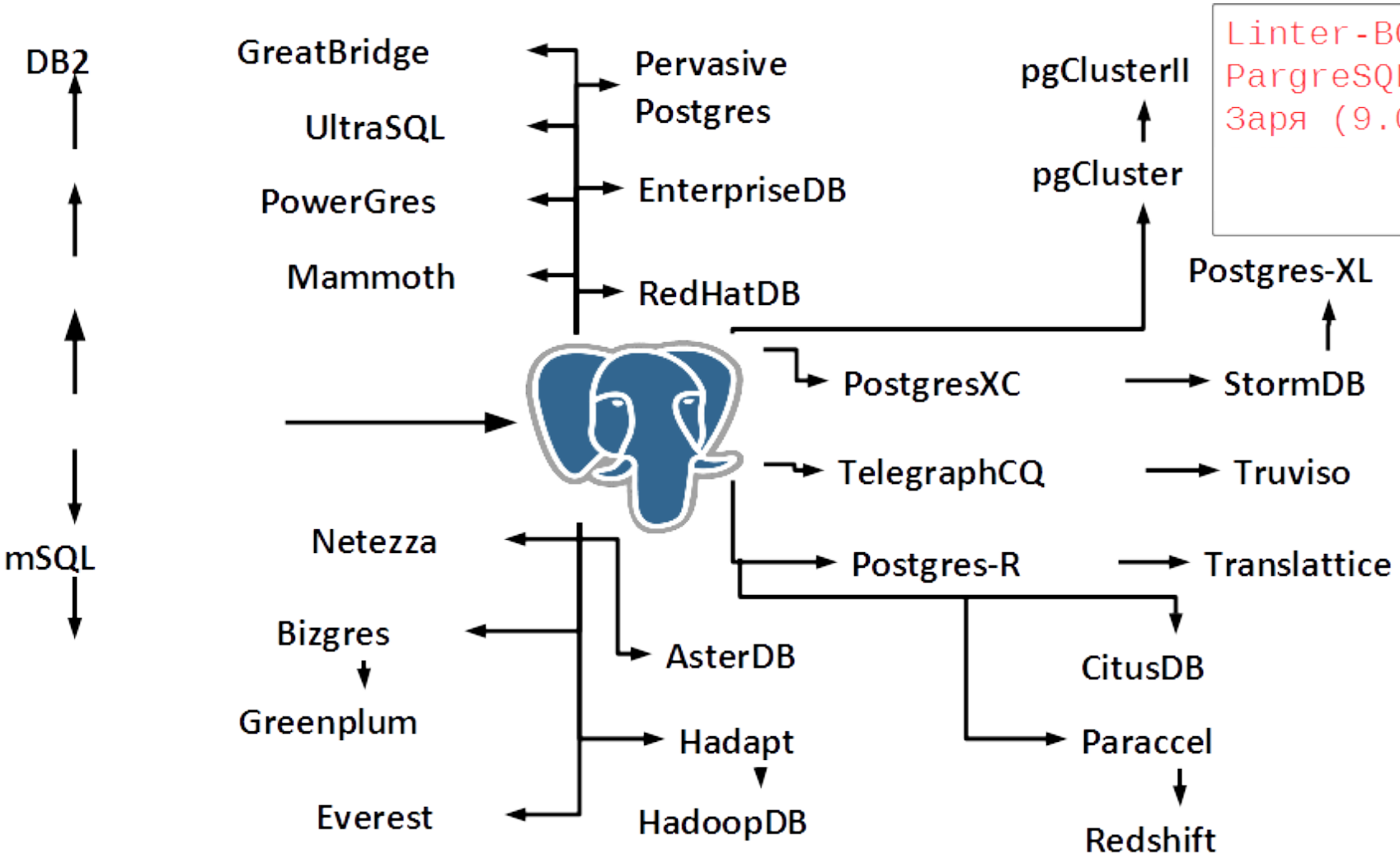
Стандарты
- SQL 92 Joins
- Prepared queries
- Foreign Keys

Функциональность ядра
- Write Ahead Log
- Prepared Queries
- Info. Schema
- Auto Vacuum

Возможности уровня Enterprise
- Потоковая репликация
- Производительность
- Вертикальное масштабирование
- PITR
- pg_upgrade
- NoSQL
- BDR
- Параллелизм

Простота использования
- Портирование на Windows
- pg_basebackup
- Различные инструменты

Основные форки PostgreSQL



PROFESSIONAL Postgres

Важнейшие свойства PostgreSQL

Надежность и устойчивость PostgreSQL

Надежность PostgreSQL является известным и доказанным фактом на примере многих проектов, в которых PostgreSQL работает без единого сбоя и при больших нагрузках на протяжении нескольких лет.

Кроссплатформенность

PostgreSQL поддерживает все виды Unix, включая Linux, FreeBSD, Solaris, HPUX, Mac OS X, а также MS Windows.

Конкурентная работа при большой нагрузке

PostgreSQL использует многоверсионность (MVCC) для обеспечения надежной и быстрой работы в конкурентных условиях под большой нагрузкой.

Масштабируемость

PostgreSQL отлично использует современную архитектуру многоядерных процессоров - его производительность растет линейно до 64-х ядер. Кластерные решения на базе Postgres XL обеспечивают горизонтальную масштабируемость.

Расширяемость

Расширяемость PostgreSQL позволяет добавлять новую функциональность, в том числе и новые типы данных, без остановки сервера и своими силами.

Доступность

PostgreSQL распространяется под лицензией BSD, которая не накладывает никаких ограничений на коммерческое использование и не требует лицензионных выплат. Вы можете даже продавать PostgreSQL под своим именем !

Независимость

PostgreSQL не принадлежит ни одной компании, он развивается международным сообществом, в том числе и российскими разработчиками. Независимость PostgreSQL означает независимость вашего бизнеса от вендора и сохранность инвестиций.

Превосходная поддержка

Сообщество PostgreSQL предоставляет квалифицированную и быструю помощь. Коммерческие компании предлагают свои услуги по всему миру.

Технические детали PostgreSQL

- Интеграция защиты данных с операционной системой (SE-Linux)
- View (materialized), sequences, inheritance, outer joins, subselects, referential integrity, window functions, CTE (WITH queries)
- Продвинутый планировщик выполнения запросов позволяет оптимизировать сложные запросы
- User defined functions, stored procedures, triggers
- Процедурные языки pl/PgSQL, pl/Perl, pl/Python, pl/V8, pl/Java и другие.
- Расширяемый набор типов данных с поддержкой индексов (GiST, GIN, SP-GiST)
- Встроенная гибкая система полнотекстового поиска с поддержкой русского и всех европейских языков
- Встроенная поддержка NoSQL: слабо-структурированные данные (xml, json,jsonb)

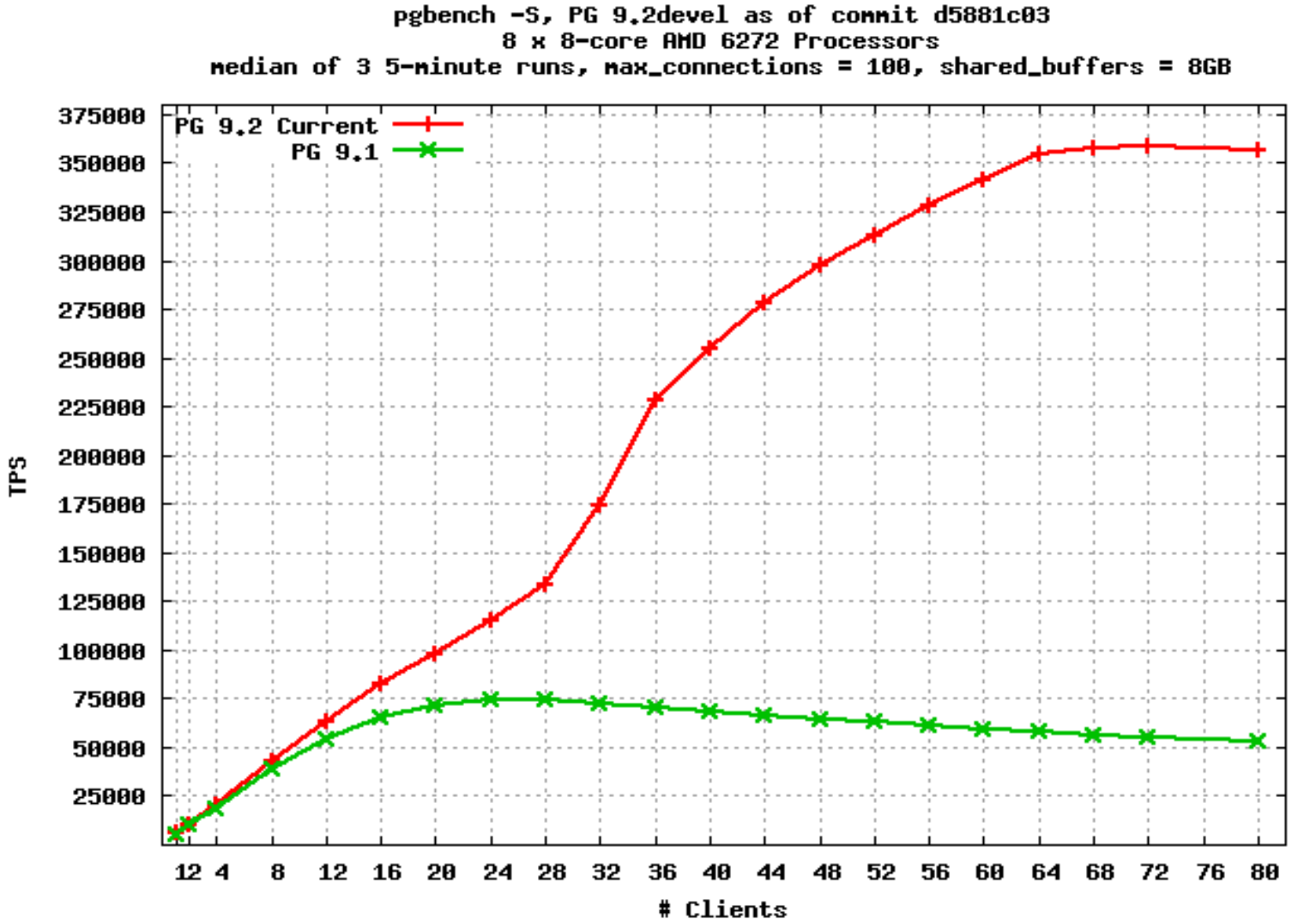
- Горячее резервирование и репликация (синхронная, асинхронная, каскадная), PITR
- Полная поддержка ACID и эффективной сериализации транзакций
- Композитные, функциональные и частичные индексы
- Интернационализация, поддержка Unicode и locale
- Загружаемые расширения, например, поддержка геоинформационных данных POSTGIS, нечеткий поиск с помощью триграм, эффективная работа с массивами
- Поддержка SSL и Kerberos аутентификации
- Foreign Data Wrappers (writable), поддержка всех основных баз данных
- Высокий уровень соответствия **стандартам** ANSI SQL 92, ANSI SQL 99 и ANSI SQL 2003, 2011
- Интерфейсы для Tcl, Perl, C, C++, PHP, Json, ODBC, JDBC, Embedded SQL in C, Python, Ruby, Java, ...

PROFESSIONAL Postgres

Ограничения

Максимальный размер БД	Не ограничено
Максимальный размер таблицы	32 Тб
Максимальная длина записи	400 Гб
Максимальная длина атрибута	1 Гб
Максимальное количество записей	Не ограничено
Максимальное количество атрибутов	250 - 1600
Максимальное количество индексов	Не ограничено

Масштабируемость, pgbench (TCP-B)



По данным Robert Haas

PROFESSIONAL Postgres

Российские разработчики

Вадим Михеев

- MVCC
- WAL
- Subselects
- Vacuum
- Triggers

Расширяемость PostgreSQL

Лежит в основе архитектуры

“It is imperative that a user be able to construct new access methods to provide efficient access to instances of nontraditional base types”

Michael Stonebraker, Jeff Anton, Michael Hirohama.
Extendability in POSTGRES , IEEE Data Eng. Bull. 10 (2) pp.16-23, 1987

Возможности для расширения

- Функции, типы данных, операторы
- Языки (sql, pl/pgsql, pl/perl, pl/python, pl/tcl, pl/R, pl/java, …,pl/v8)
- Индексный доступ (Btree, Hash, GiST, GIN, SP-GiST)
- Foreign Data Wrappers (практически ко всем субд)

Популярные расширения

В стандартной поставке PostgreSQL >60 расширений.
Некоторые популярные расширения:

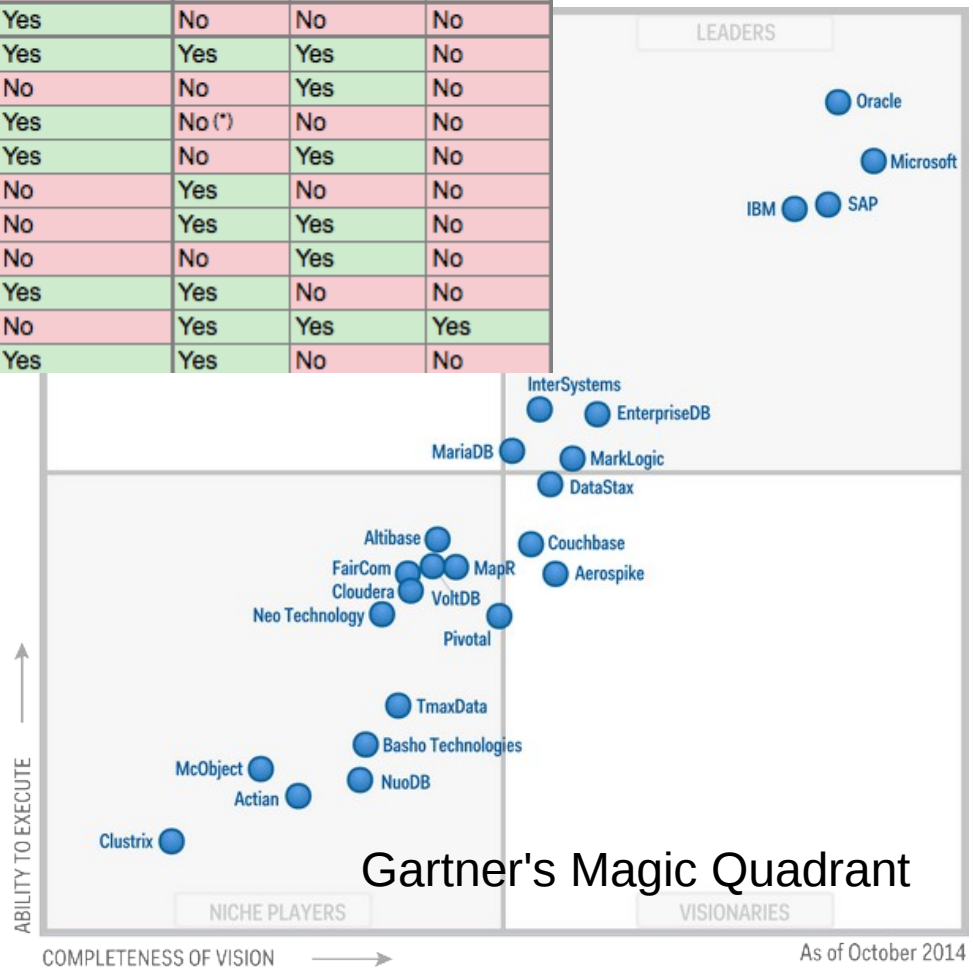
PostGIS	Поддержка пространственных объектов в PostgreSQL и всех стандартов ГИС !
PLV8	Разработка хранимых функций на языке V8 JavaScript
PL/proxy	Удаленный вызов процедур и партиционирование данных между разными базами
oracle_fdw	Доступ к СУБД Oracle. Запросы в PostgreSQL могут обращаться к данным Oracle как к обычным таблицам.
pg_partman	Управление партиционированными таблицами

PROFESSIONAL Postgres

http://sql-workbench.net/dbms_comparison.html

Feature	Oracle	Postgres	SQL Server	MySQL	IBM DB2	Firebird
Queries						
Window functions	Yes	Yes	Yes	No	Yes	No (*)
Common Table Expressions	Yes	Yes	Yes	No	Yes	Yes
Recursive Queries	Yes	Yes	Yes	No	Yes	Yes
Row constructor (*)	No	Yes	Yes (*)	Yes	Yes	No
Filtered aggregates (*)	No	Yes (*)	No	No	No	No
PIVOT Support	Yes	No (*)	Yes	No	No	No
GROUP BY .. ROLLUP	Yes	No	Yes	Yes	Yes	No
Temporal queries (*)	Yes	No	No	No	Yes	No
SELECT without a FROM clause	No	Yes	Yes	No (*)	No	No
Parallel queries (*)	Yes	No (*)	Yes	No	Yes	No
Aggregates for strings	Yes (*)	Yes	No	Yes	No	No
Tuple comparison	Yes	Yes	No	Yes	Yes	No
Tuple updates	Yes	Yes	No	No	Yes	No
UPDATE with a join	No	Yes	Yes	Yes	No	No
ANSI date literals (*)	Yes	Yes	No	Yes	Yes	Yes
Query variables (*)	No	No	Yes	Yes	No	No

Oracle	72
PostgreSQL	88
Firebird	40
MSSQL	64



Gartner's Magic Quadrant

PROFESSIONAL Postgres

Российские разработчики

Вадим Михеев

- MVCC
- WAL
- Subselects
- Vacuum
- Triggers

Российские разработчики

Олег Бартунов, Федор Сигаев

- Locale support
- Extendability (indexing)
- GiST(KNN), GIN, SP-GiST
- Full Text Search (FTS)
- Jsonb, VODKA

Расширения:

- Intarray
- Pg_trgm
- Ltree
- Hstore
- plantuner

Александр Коротков

- Indexed regexp search
- GIN compression & fast scan
- Range types indexing
- Split for GiST

Политика выпуска версий

- Major.minor – например, 9.4.3
  - Рекомендуется к использованию последняя минорная версия
- Major
  - изменения в системном каталоге, файлов данных. Dump/reload, pg_upgrade
  - Поддержка 5 лет
- Minor
  - Мелкие баги. Стоп сервер, установить бинарники, запуск сервера

Расширяемость PostgreSQL

Лежит в основе архитектуры

“It is imperative that a user be able to construct new access methods to provide efficient access to instances of nontraditional base types”

Michael Stonebraker, Jeff Anton, Michael Hirohama.
Extendability in POSTGRES , IEEE Data Eng. Bull. 10 (2) pp.16-23, 1987

Возможности для расширения

- Функции, типы данных, операторы
- Языки (sql, pl/pgsql, pl/perl, pl/python, pl/tcl, pl/R, pl/java, …,pl/v8)
- Индексный доступ (Btree, Hash, GiST, GIN, SP-GiST)
- Foreign Data Wrappers (практически ко всем субд)

Популярные расширения

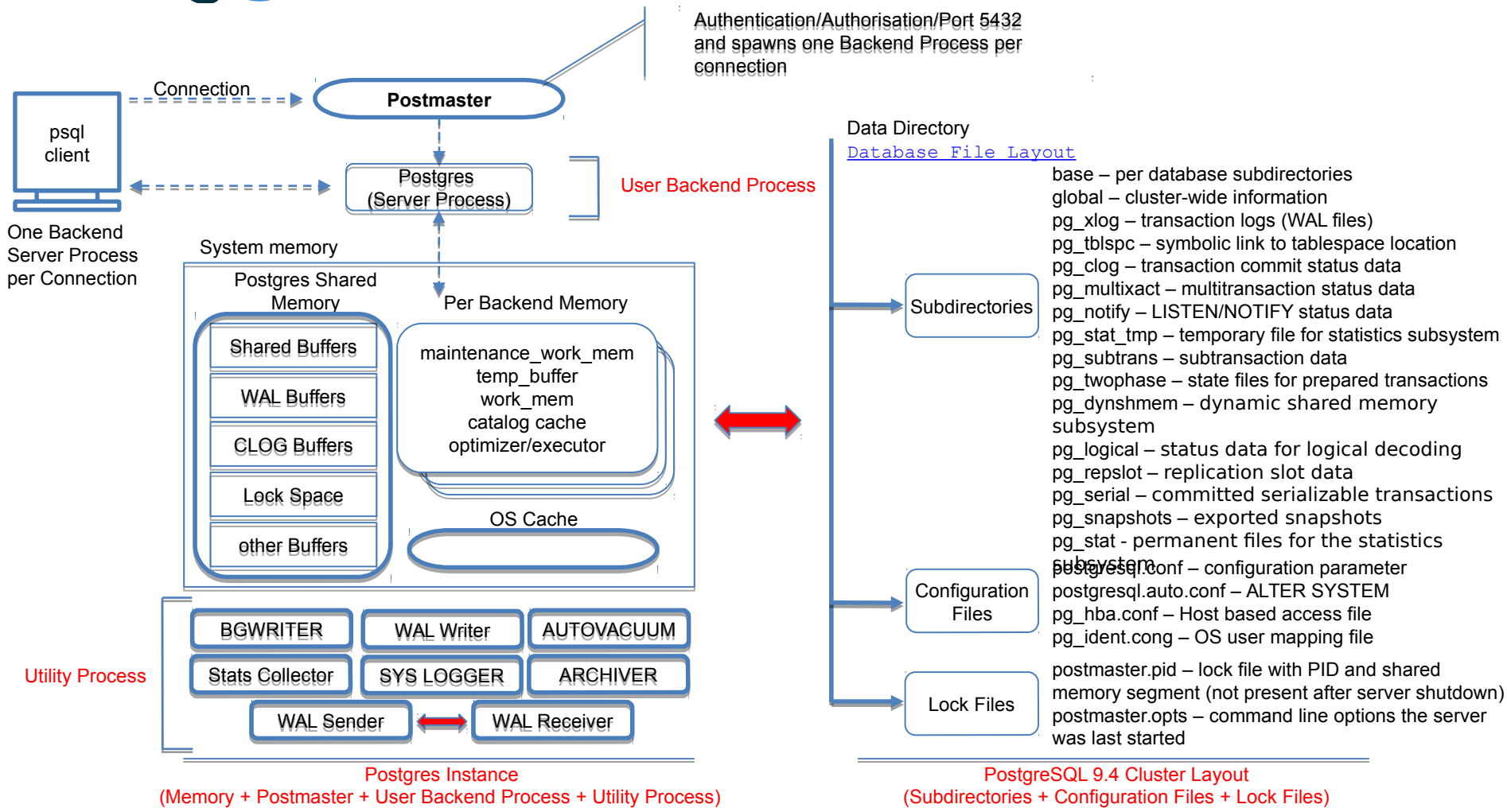
В стандартной поставке PostgreSQL >60 расширений.
Некоторые популярные расширения:

PostGIS	Поддержка пространственных объектов в PostgreSQL и всех стандартов ГИС !
PLV8	Разработка хранимых функций на языке V8 JavaScript
PL/proxy	Удаленный вызов процедур и партиционирование данных между разными базами
oracle_fdw	Доступ к СУБД Oracle. Запросы в PostgreSQL могут обращаться к данным Oracle как к обычным таблицам.
pg_partman	Управление партиционированными таблицами

Расписание разработки

- Мэйлинг листы
- Commitfest
  - Июнь 14, 2013 - branch 9.3 && CF1
  - Сентябрь 2013 – CF2
  - Ноябрь 2013 - CF3
  - Январь 2014 – CF4
  - Beta
    - 15 Мая 2014 - Beta 1 (PGCon conference, Ottawa, Canada)
    - 24 июля 2014 - Beta 2
    - 9 октября 2014 – Beta 3 (PGConf conference, Madrid, Spain)
  - Release Candidate
    - 20 ноября 2014 – RC1
  - 18 Декабря 2014– Release 9.4
- Development information, PostgreSQL Todo, Developer FAQ

Архитектура PostgreSQL 9.4



Keys: Independent Processes

attached to backend process

one time hit process

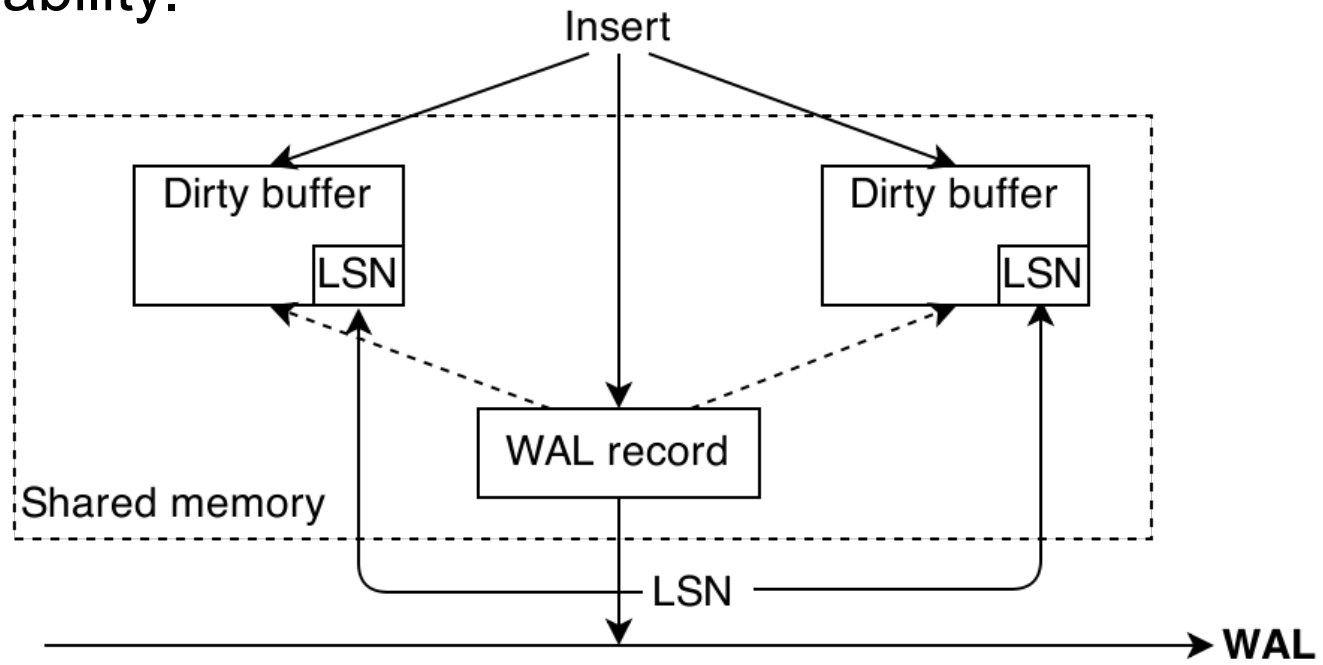
Postgres PROFESSIONAL

MVCC

- MVCC – механизм, позволяющий каждой транзакции видеть свой «слепок» (snapshot) базы данных на определенный момент времени, хотя данные на текущий момент уже могли измениться.
- В PostgreSQL MVCC обеспечивается тем, что данные не удаляются, старые версии строк остаются с отметками об окончании их актуальности, параллельно заводятся новые версии строк. Специальный процесс VACUUM удаляет старые версии строк.

Write-Ahead Log (WAL)

Любое изменение на дисковых страницах вначале записывает в WAL. Прежде, чем PostgreSQL сообщает об успешном commit'е транзакции для WAL делается fsync. Этим обеспечивается durability.



Логический backup – на уровне SQL команд

- pg_dump/pg_restore
- pg_restore в несколько потоков – «-j n» (8.4+)
- pg_dump в несколько потоков – «-j n», благодаря «EXPORT SNAPSHOT» (9.3+)
- Степень сжатия опциональна

Физический backup – на уровне файлов

Варианты реализации:

- Offline backup: остановить PostgreSQL, скопировать данные, запустить PostgreSQL
- С помощью возможностей FS/SAN
- С помощью inconsistent копии файлов и WAL
  - pg_start_backup('label'), копирование вручную, pg_stop_backup()
  - Утилита pg_base_backup

Continuous archiving

Файловый backup + непрерывное сохранение WAL даёт continuous archiving.

Возможны различные способы доставки WAL в архив:

- archive_command, restore_command – кусками по 16 MB
- pg_receivexlog – непрерывно, используя streaming replication protocol

Postgres PROFESSIONAL

PITR

На файловый backup можно накатывать произвольное количество WAL'ов, тем самым обеспечивая Point-In-Time Recovery.

- recovery_target = 'immediate' – восстановление до первого consistent состояния
- recovery_target_name – восстановление до заранее созданной именованной точки
- recovery_target_time – восстановление до конкретного момента времени
- recovery_target_xid – восстановление до конкретного id транзакции

Streaming репликация

Встроенная master-slave репликация, основанная на потоковой передаче WAL с master на slave.

- Синхронная/асинхронная – можно настраивать индивидуально для каждой транзакции
- Отставшие реплики могут получать WAL-файлы из архива. Master может держать достаточно WAL-файлов для восстановления отставших реплик (replication slots).
- Утилита pg_rewind позволяет вернуть подключить старый master как slave без его полной перезаливки.

Postgres PROFESSIONAL

Полезные утилиты

- repmgr – утилита для автоматизации streming репликации: упрощение настройки, мониторинг, автоматический failover.
- pglookout – утилита для мониторинга репликации и автоматического failover.
- wal-e – автоматизация файлового backup, continuous archiving и PITR.
- barman – автоматизация файлового backup, continuous archiving, PITR, incremental backup на уровне файлов.

Postgres PROFESSIONAL

Другие способы репликации

- PgPool II – statement-based репликация
- Slony – trigger-based master-slave
- Londiste – trigger-based master-slave
- BDR – 2ndquadrant fork: двунаправленный асинхронный multi-master
  - UDR – подмножество BDR, доступно как расширение к 9.4, master-slave

Postgres PROFESSIONAL

Connection pooling

- pgBouncer – легковесный и простой в настройке connection pool
- PgPool II – connection pool с поддержкой балансировки нагрузки в режимах master-master и master-slave

Postgres PROFESSIONAL

Отказоустойчивость

- repmgr
- pglookout
- corosync/pacemaker

Postgres PROFESSIONAL

Параметры для мониторинга производительности

- Скорость исполнения и количество запросов
- Объемы БД, таблиц, индексов
- Планы запросов и их изменение
- Эффективность буферов
- Дисковая активность
- Количество блокировок, время ожидания
- Работа отдельных подсистем (bgwriter, vacuum, репликация)

Средства мониторинга производительности

- Внутренние средства PostgreSQL
- Плагины для распространенных систем мониторинга
- Интегрированные системы мониторинга, ориентированные на PostgreSQL

Внутренние средства мониторинга

Statistics Collector
pg_stat_* : 27 различных таблиц, 18 функций для управления

Extensions

- pg_stat_statements: статистика по отдельным запросам
- pg_stat_plans: статистика по планам запросов
- pg_buffercache: статистика по буферам
- pg_stat_qcache: статистика на уровне кеша ОС
- и др.

Нужно использовать совместно со средствами мониторинга ОС

Пример: pg_stat_statements

```
\d pg_stat_statements
         View "public.pg_stat_statements"
      Column         |       Type        | Modifiers
---------------------+-------------------+-----------
 userid              | oid               |
 dbid                | oid               |
 query               | text              |
 calls               | bigint            |
 total_time          | double precision  |
 rows                | bigint            |
 shared_blks_hit     | bigint            |
 shared_blks_read    | bigint            |
 shared_blks_dirtied | bigint            |
 shared_blks_written | bigint            |
 local_blks_hit      | bigint            |
 local_blks_read     | bigint            |
 local_blks_dirtied  | bigint            |
 local_blks_written  | bigint            |
 temp_blks_read      | bigint            |
 temp_blks_written   | bigint            |
 blk_read_time       | double precision  |
 blk_write_time      | double precision  |
```

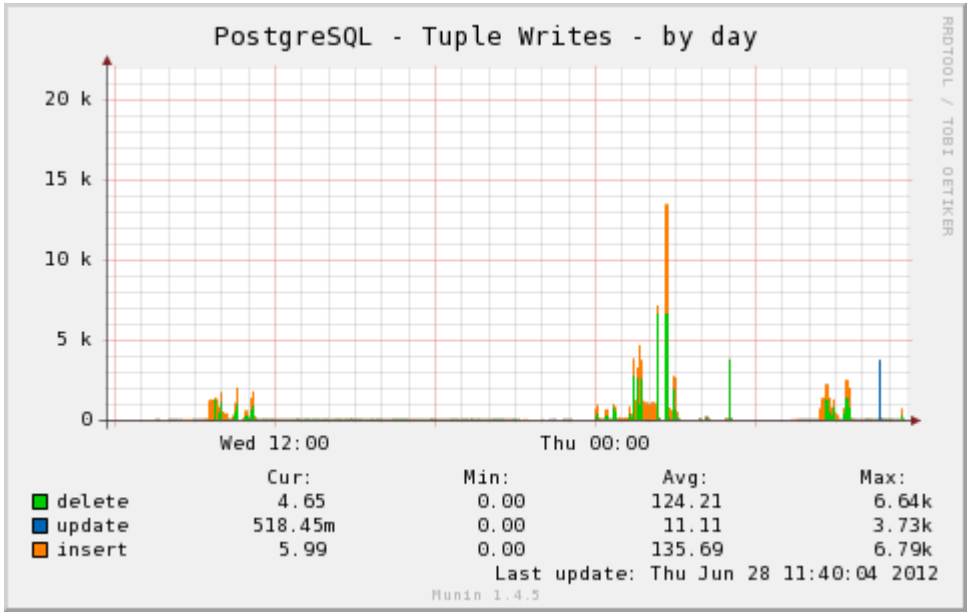
Пример: план запроса

```
explain analyze select * from grant_application a join grant_programs p on a.program = p.id
                                and a.create_time between '2015-01-01' and '2015-02-01';

 Hash Join  (cost=6.19..693.71 rows=878 width=545) (actual time=0.201..8.243 rows=889 loops=1)
   Hash Cond: (a.program = p.id)
   ->  Index Scan using grant_application_create_time on grant_application a
                 (cost=0.29..675.73 rows=878 width=178) (actual time=0.029..2.728 rows=889 loops=1)
         Index Cond: ((create_time >= '2015-01-01 00:00:00'::timestamp without time zone)
                   AND (create_time <= '2015-02-01 00:00:00'::timestamp without time zone))
   ->  Hash  (cost=5.40..5.40 rows=40 width=367) (actual time=0.142..0.142 rows=24 loops=1)
         Buckets: 1024  Batches: 1  Memory Usage: 17kB
         ->  Seq Scan on grant_programs p  (cost=0.00..5.40 rows=40 width=367)
                   (actual time=0.007..0.067 rows=24 loops=1)
 Total runtime: 9.851 ms
```

pg_stat_plans позволяет накапливать статистику по планам исполнения запросов.

Плагины для систем мониторинга

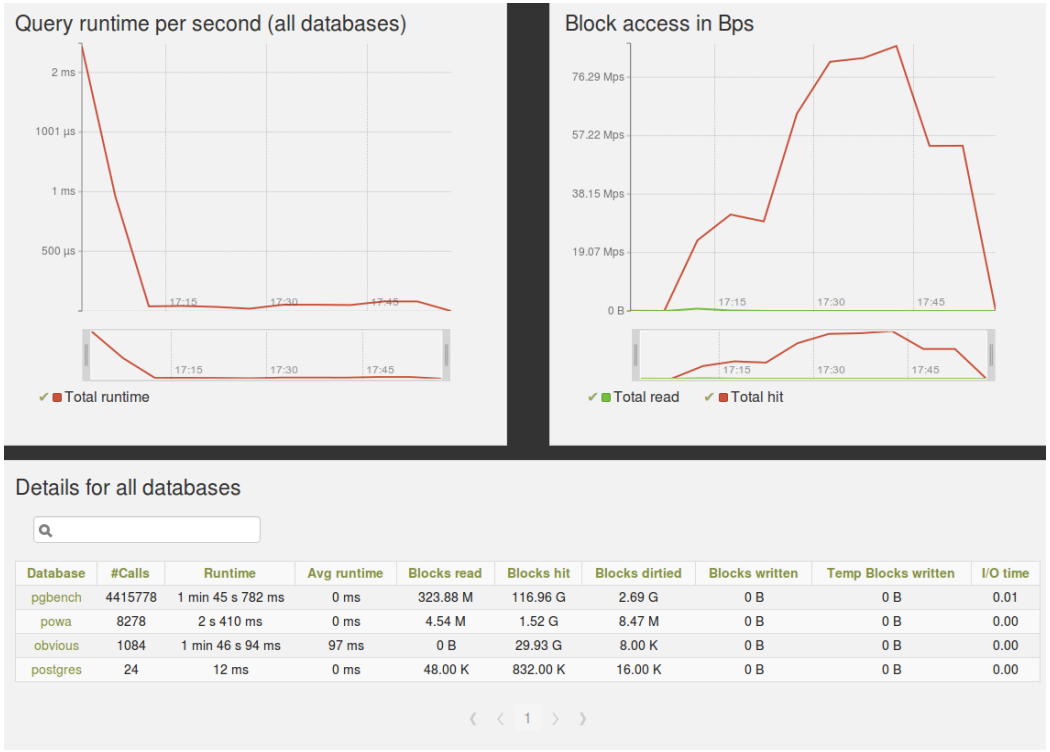


	графики	алерты
Nagios	нет	есть
Munin	есть	есть
Zabbix	есть	есть
Cacti	есть	нет

Интегрированные системы мониторинга

PoWA (PostgreSQL Workload Analyzer)
OPM (Open PostgreSQL monitoring)

- Данные хранятся в БД PostgreSQL
- Мониторинг параметров PostgreSQL и ОС
- Расширения pg_qual_stats, pg_stat_kcache



Database	#Calls	Runtime	Avg runtime	Blocks read	Blocks hit	Blocks dirtied	Blocks written	Temp Blocks written	I/O time
pgbench	4415778	1 min 45 s 782 ms	0 ms	323.88 M	116.96 G	2.69 G	0 B	0 B	0.01
powa	8278	2 s 410 ms	0 ms	4.54 M	1.52 G	8.47 M	0 B	0 B	0.00
obvious	1084	1 min 46 s 94 ms	97 ms	0 B	29.93 G	8.00 K	0 B	0 B	0.00
postgres	24	12 ms	0 ms	48.00 K	832.00 K	16.00 K	0 B	0 B	0.00

Postgres PROFESSIONAL

Партиционирование

Виды:

- Range Partitioning

  Например, по годам, месяцам
- List Partitioning

  Определенные списки значений (‘Москва’, ‘Пенза’)

Реализация:

- Партиции – это дочерние таблицы, наследуемые (inherits) от родительской таблицы
- Триггер на родительскую таблицу для разнесения DML операций по детальным таблицам
- Параметр CONSTRAINT_EXCLUSION=ON

Расширение pg_partman (блог) позволяет автоматически управлять партиционированием

Управление доступом и пользователями

Поддерживаются следующие методы аутентификации:

trust, password, GSSAPI, SSPI, Ident, Peer, LDAP, RADIUS, Certificate, PAM

Управление пользователями и доступом к объектам БД

- Пользователи и роли, роли могут быть вложенными
- Доступ к объектам БД (grant/revoke) как напрямую пользователям, так и косвенно через роли
- Разделение доступа на уровне столбцов и строк (Row Level Security, в 9.5)
- Поддержка SELinux через встроенную функциональность SE-PostgreSQL (мандатный доступ)

Разработка на стороне сервера

- Доступные (помимо C) языки программирования для разработки хранимых функций:

  sql, pl/pgsql, pl/perl, pl/python, pl/tcl, pl/R, pl/java, pl/v8…
- Триггеры на таблицы, представления, системные события

PROFESSIONAL Postgres

Миграция Oracle to PostgreSQL

- PostgreSQL for Oracle DBA (wiki)
- Oracle to Postgres Conversion (wiki)
- PL-SQL code rules to write Oracle and Postgresql code
- Comparison between Oracle & Postgis in terms of Spatial Queries
- Compare SQL Server 2008 R2, Oracle 11G R2, PostgreSQL/PostGIS 1.5 Spatial Features
- Features Compatibility with Oracle
- Oracle to PostgreSQL tips
- Orafce - Oracle functions in PostgreSQL
- Ora2pg - migration tool !
- Oracle to PostgreSQL migration
- Porting Oracle applications to PostgreSQL
- PostgreSQL Foreign Data Wrapper for Oracle

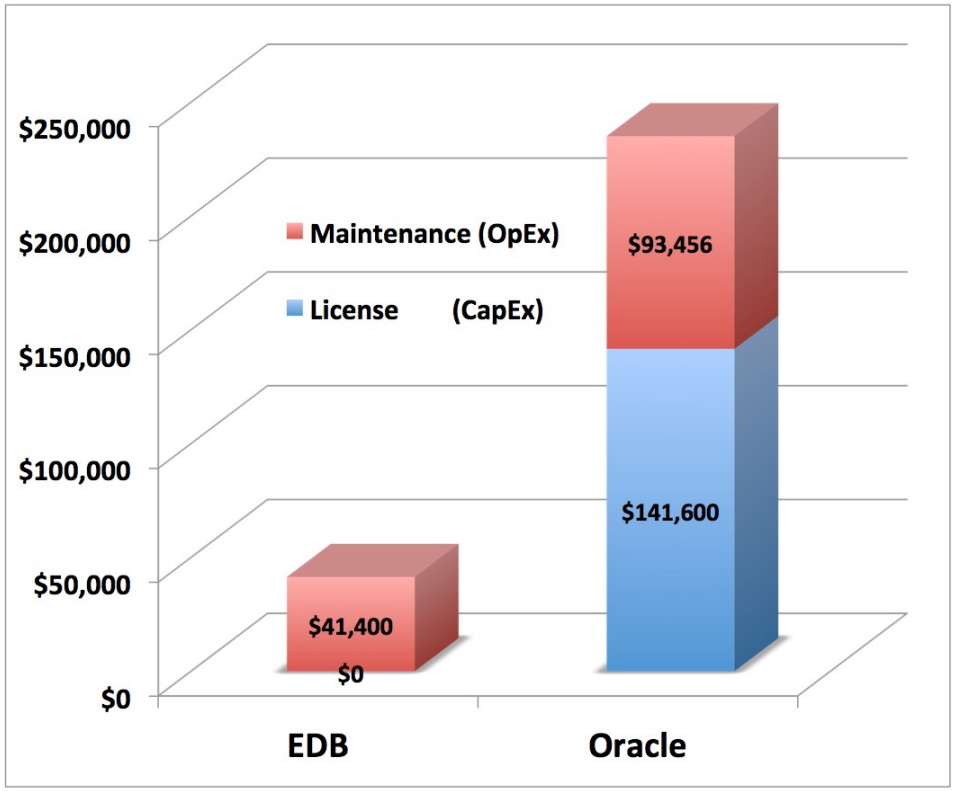
Примеры миграции из Oracle на PostgreSQL

- Национальный фонд семейных пособий (CNAF) во Франции (данные по 30 млн человек; миллиард запросов в день)
- Национальная метеослужба Франции (Размер самой крупной БД - 3,5Тб)
- leboncoin.fr (250 млн просмотров страниц в день; уникальных посетителей: в день – 5 млн, в месяц – 18 млн; 600000+ новых объявлений в день; 25млн актуальных объявлений)
- PostgreSQL в Яндекс:
  - История успеха 1
  - История успеха 2

PROFESSIONAL Postgres

TCO

3 Year TCO Comparison
x86 2 Sockets by 4 Cores



Prices include Partitioning?	YES
Prices include Active Data Guard?	NO
Price include Spatial?	NO
EDB price includes all three?	YES

Savings with EDB: $ 193,656

Postgres PROFESSIONAL

Почему PostgreSQL ?

- Одна из наиболее распространённых в мире СУБД. Большая экосистема.

  Skype, Instagram, Yandex, AVITO, Sony, Huawei, Caixa Econômica Federal, ЕЭТП, OpenStreetMap…
  Сообщество, расширения, разработчики.
  Импортозамещение без изоляции.

- Открытая лицензия
- Широкий круг решаемых задач
  OLTP, OLAP, GIS, NoSQL, полнотекстовый поиск,…
  Лидер среди РСУБД в области GIS и NoSQL!
- Существенный российский вклад и задел

  MVCC, GIS, GiST, GIN, полнотекстовый поиск, расширяемость, NoSQL, …Сообщество ~ неск.тыс.чел.

Формула импортозамещения

Свободно-распространяемый продукт
+ Высокое качество
+ Открытая лицензия
+ Универсальность
+ Существенный российский вклад
+ Увеличение вклада и рост компетенции
+ Отечественная экосистема
=
Импортозамещающий продукт
+
Отечественная отрасль СУБД-строения
=
Технологическая независимость
+
Конкурентоспособность на мировом рынке

Инициатива Минкомсвязи
Импортозамещение в области СУБД

Приказ Минкомсвязи России №96 от 01.04.2015
«Об утверждении плана импортозамещения программного обеспечения»

- Поддержка разработки
- Софинансирование
- Альянсы
- Задел + План

- 7 паспортов проектов по СУБД,
  в т.ч. проект от компании “Постгрес Профессиональный”

Постгрес Профессиональный

- Все российские ключевые международно признанные разработчики PostgreSQL работают в нашей компании
- Вклад наших сотрудников существенен и признан мировым сообществом
  - В направлениях, где мы ведем разработку, PostgreSQL является лидером*** среди РСУБД:
    - геоинформационные системы
    - слабоструктурированные данные,
    - полнотекстовый поиск
    - расширяемость
  - Более 20 докладов на международных конференциях
  - В нашей команде 4 кандидата наук, из них 3 – по PostgreSQL и технологиям БД
- Тему импортозамещения СУБД пропагандируем с 2011 г.
- Организованная нами PgConf.Russia 2015 была крупнейшей в мире конференцией по PostgreSQL (более 450 участников)

Основное содержание проекта

- Создаем самую совершенную СУБД

Кластер, подключаемые хранилища (в памяти, колоночное), оптимизация, расширение функциональности, промышленные средства разработки, администрирования и мониторинга

- Локализация

Документация, сертификация, лингвистика, поддержка отечественных продуктов.

- Создание экосистемы

СУБД-строение как отрасль экономики. Широкий круг системных и прикладных разработчиков. Учебные курсы. Наука и образование, конференции. Центр тестирования.

Ожидаемый результат

- Продукт

Полнофункциональная промышленная СУБД мирового класса,
- полностью контролируемая российскими разработчиками,
- поддерживаемая мировым сообществом,
- обладающая большим количеством вспомогательного ПО,
- конкурентоспособная на мировом рынке,
- активно используемая в отечественной ИТ-отрасли.

- Отрасль

Конкурентная на международном рынке отечественная индустрия СУБД-строения. Самовоспроизводящаяся экосистема, включающая разработчиков СУБД, разработчиков прикладных систем на базе СУБД, систему образования и научных исследований в области СУБД. Рынок труда разработчиков. Международная кооперация.

Спасибо за внимание!

Иван Евгеньевич Панченко

i.panchenko@postgrespro.ru
ООО “Постгрес Профессиональный”
http://postgrespro.ru/