



# Что нового в PostgreSQL 8.3?

Николай Самохвалов,  
«Постгресмен»

Первая встреча Российского PostgreSQL-сообщества  
Москва, февраль 2008

# PostgreSQL – это...

- ...самая развитая СУБД с открытым кодом
- ...надежность и устойчивость на любых нагрузках
- ...кроссплатформенность (FreeBSD, Linux, Solaris, Windows, OS X)
- ...высокий уровень соответствия ISO/ANSI SQL 92, 99, 2003
- ...интерфейсы для Tcl, Perl, C, C++, PHP, ODBC, JDBC, Embedded SQL in C, Python, Ruby, ...
- ...развитые административные утилиты (pgAdmin, phpPgAdmin, ...)

# Кто использует

- **Skype**

*Skype has been using PostgreSQL as the main DB for most of our business needs right from the start.*

*Hannu Krosing, Skype*

- **Hi5.com** (~60 млн. пользователей; Alexa's traffic rank: 8)

- **MyYearBook.com**

*We process over 18,000 queries per second on over 300GB of user data on our PostgreSQL servers, and those numbers go up every month.*

*Gavin Roy, CTO of MyYearbook.com*

# Кто использует: Россия

- Рамблер
- 1С:Предприятие (наряду с MS SQL Server)
- MirTesen.ru — *уже 8.3!*
- MoiKrug.ru (Яндекс)
- IRR.ru («Из рук в руки»)
- работа.ru, price.ru, webalta.ru, РБК, Мастерхост, Информ-мобил, ...

# Кто использует: Россия

- **udaff.com**

# 11 лет развития PostgreSQL

- ...1995 – «прошлая жизнь» (Postgres)
- 1996...1998 – стабильность
- 1998...2001 – стандарт SQL
- 2001...н. в. – выход на уровень Enterprise
  - производительность
  - простота администрирования
  - 24/7
  - множество проектов «вокруг»



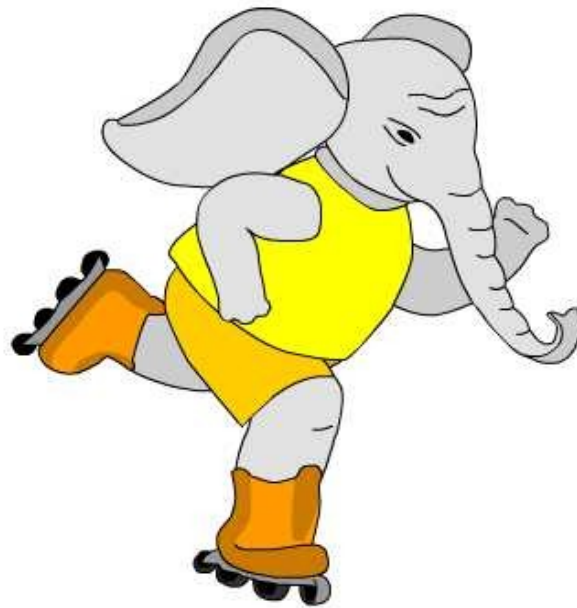
# PostgreSQL 8.3

- Более **280** патчей
- Десятки разработчиков из **18** стран
- **15** месяцев разработки и тестирования

# Новое в PostgreSQL 8.3

- Производительность
- Разработчикам баз данных
- Администраторам баз данных
- Проекты сторонних разработчиков

# Производительность



# Summer 2007: The 1<sup>st</sup> PostgreSQL Enterprise-level Benchmark (SPEC\*)

*Josh Berkus: «...a good day for Open Source»*

## PostgreSQL 8.2 – 813.73 JOPS

- SPECjAppServer2004 2x Sun Fire X4200 appservers (8 cores, 4 chips) and 1 Sun Fire T2000 DB server (8 cores, 1 chips) with PostgreSQL 8.2.4
- HW: ~\$60,000; SW: \$0

## Oracle 10g – 874.17 JOPS

- SPECjAppServer2004 1 HP rx2660 appserver (4 cores, 2 chips) and 1 rx2660 DB server (4 cores, 2 chips) with Oracle Database 10g Enterprise Edition Release 10.2.0.2
- HW: ~\$74,000; SW: ~\$110,000

## MySQL 5.0 – 720.56 JOPS

- SPECjAppServer2004 3x Sun Fire X4100 appservers (12 cores, 6 chips) and 1 Sun Fire X4100 DB server (4 cores, 2 chips) with MySQL 5.0.27 64-bit

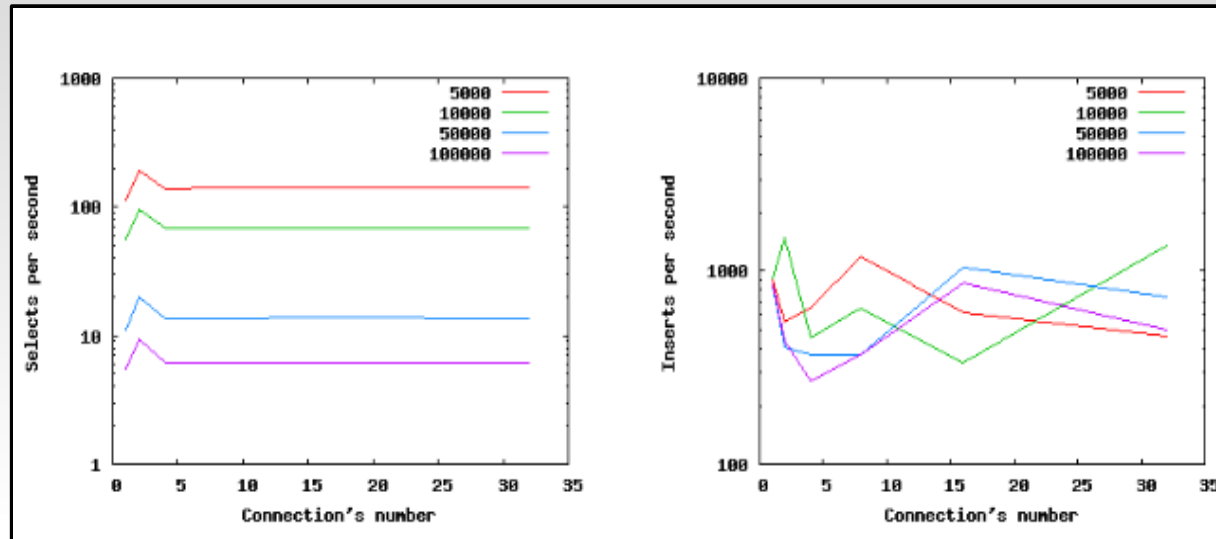
---

\*) SPEC – Standard Performance Evaluation Corporation, <http://spec.org>

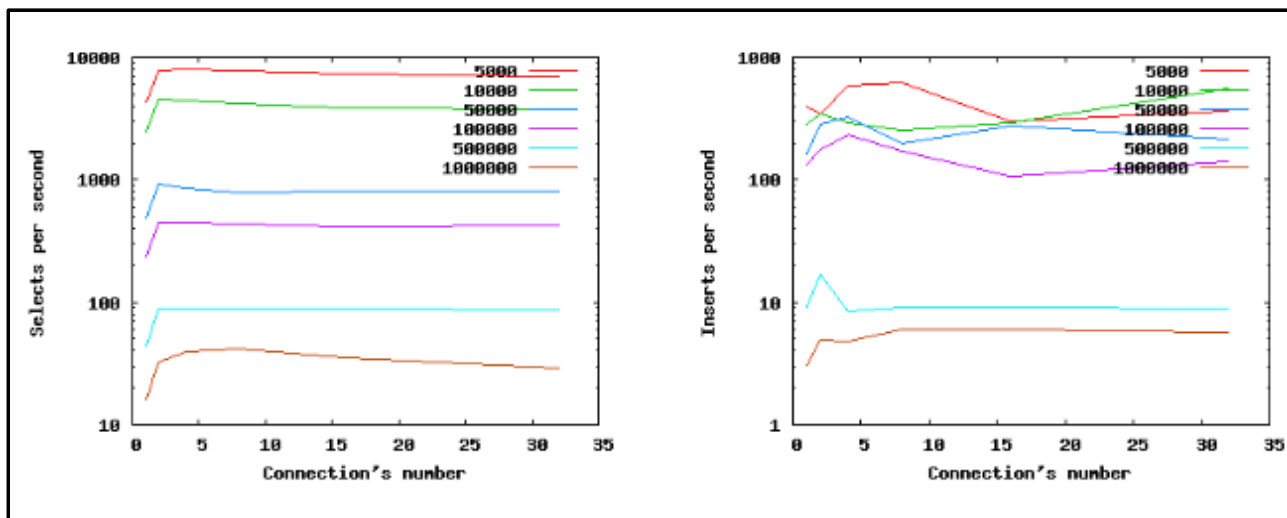
# Индексы: GiST и GIN

(Фёдор Сигаев, Олег Бартунов)

GiST



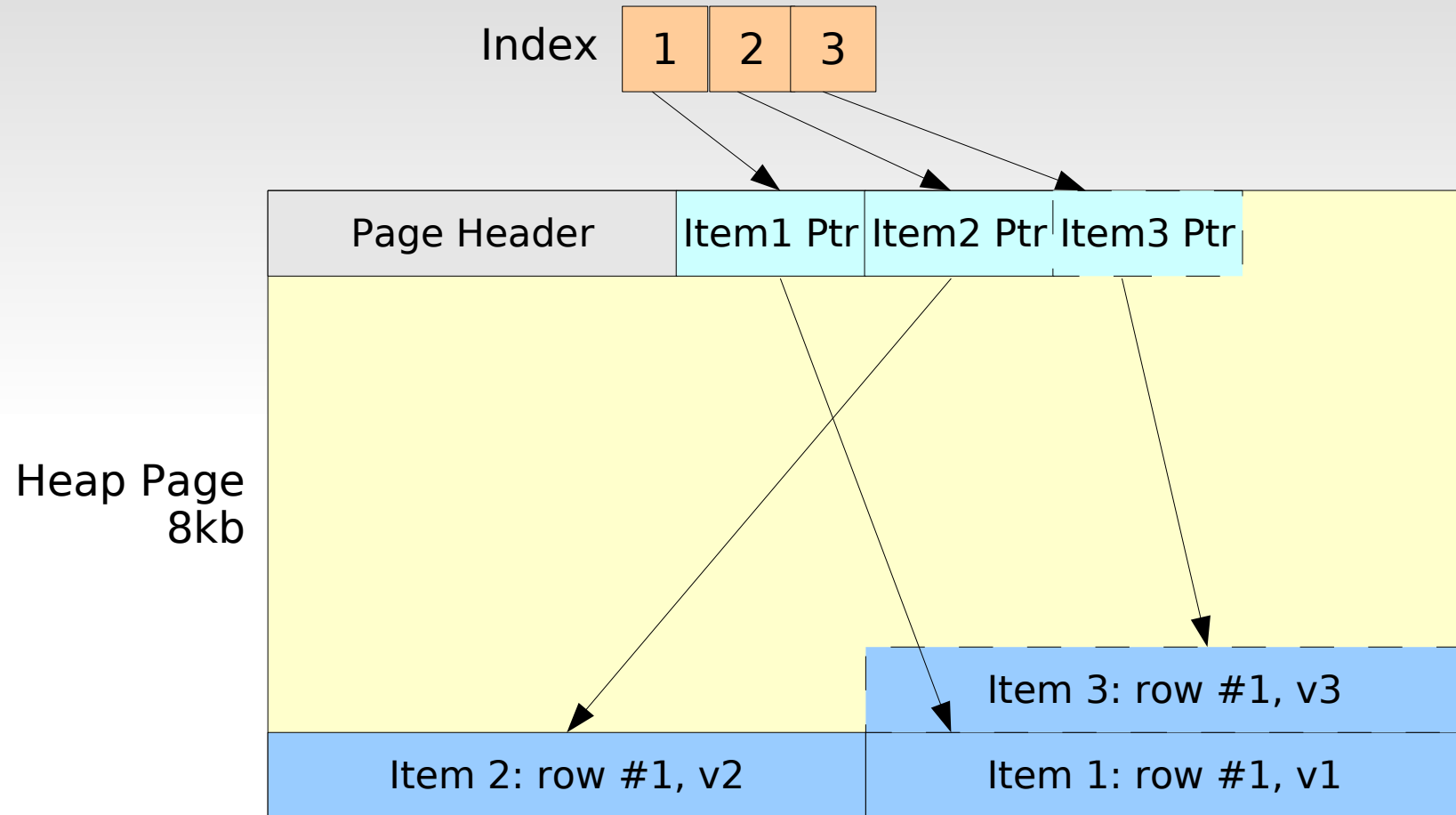
GIN



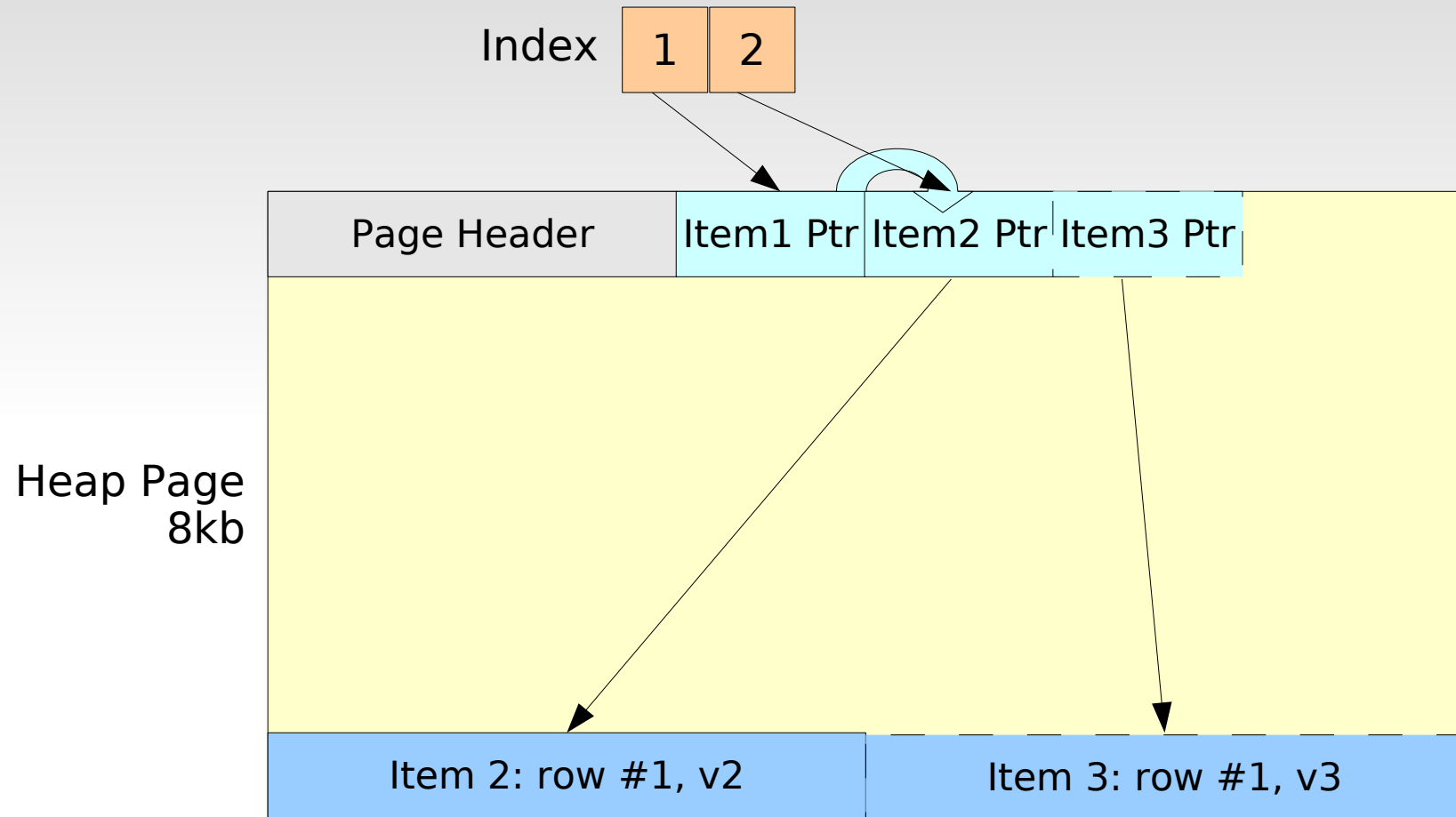
# HOT (Heap Only Tuples)

- При обновлении (UPDATE) индекс **не** меняется, если:
  - UPDATE не затрагивает проиндексированные столбцы,
  - новая версия строки записывается на той же странице данных (heap).
- «Моментальное» повторное использование места в heap без ожидания процесса VACUUM.

# Обновление без HOT



# НОТ-обновление



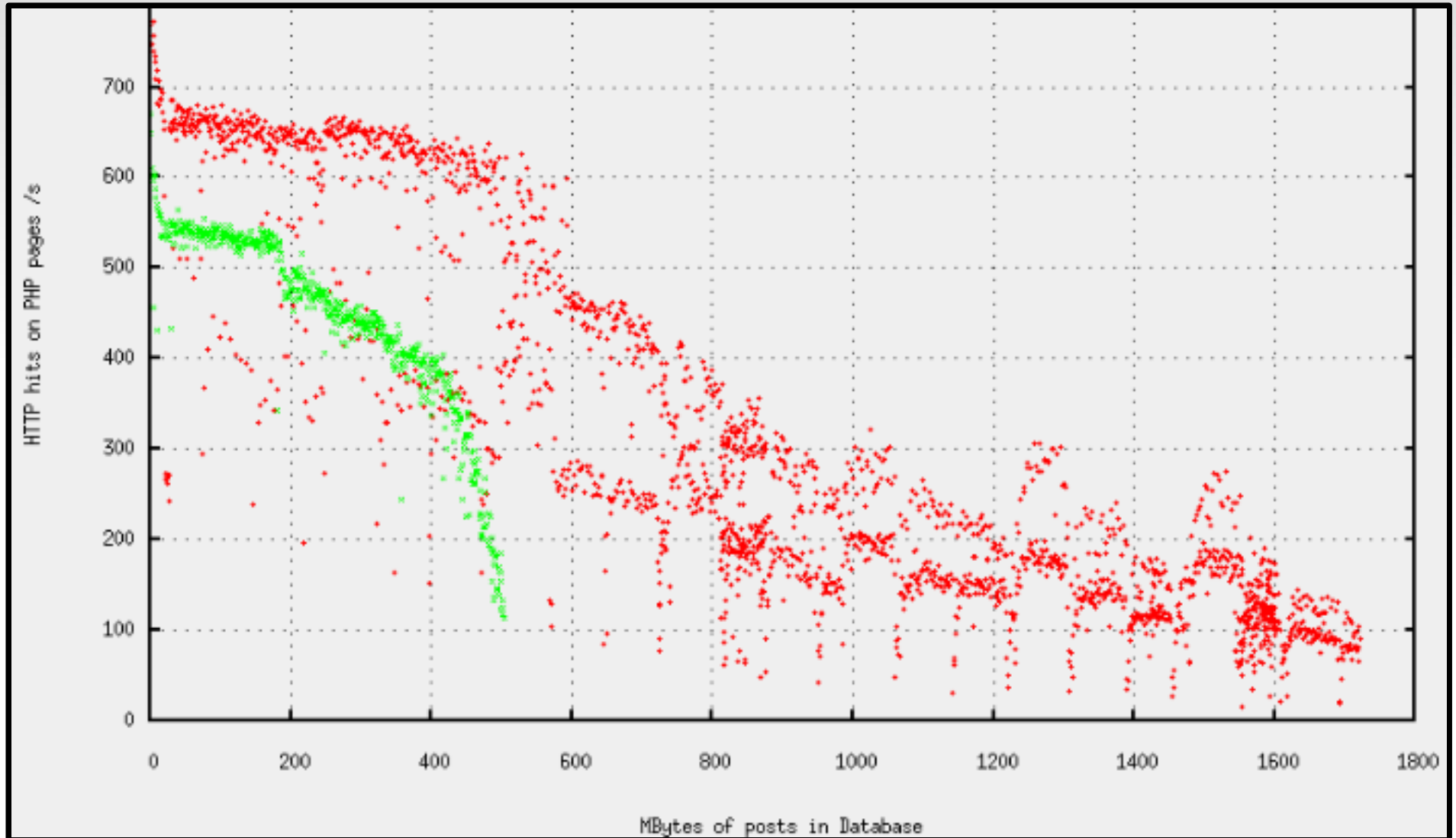
# Asynchronous Commit

```
SET synchronous_commit TO OFF;
```

Позволяет отложить fsync на небольшое время (COMMIT без ожидания fsync)

- возможна потеря небольшой порции данных (порядка 0.2...1.0 сек)
- нарушение целостности данных **невозможно!** (в отличие от fsync=off)

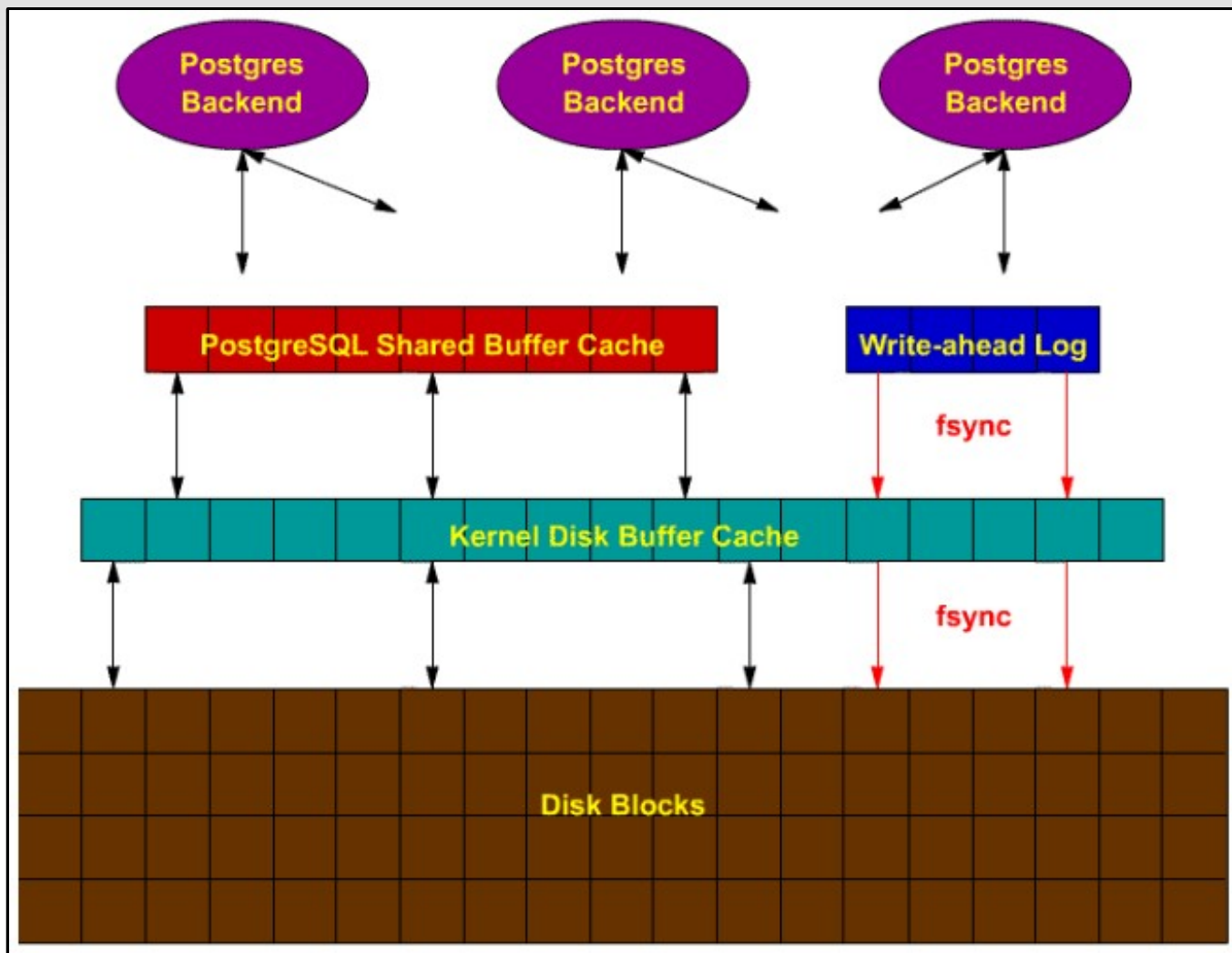
# Надёжность производительности



# Сглаживание контрольных точек (`checkpoint smoothing`)

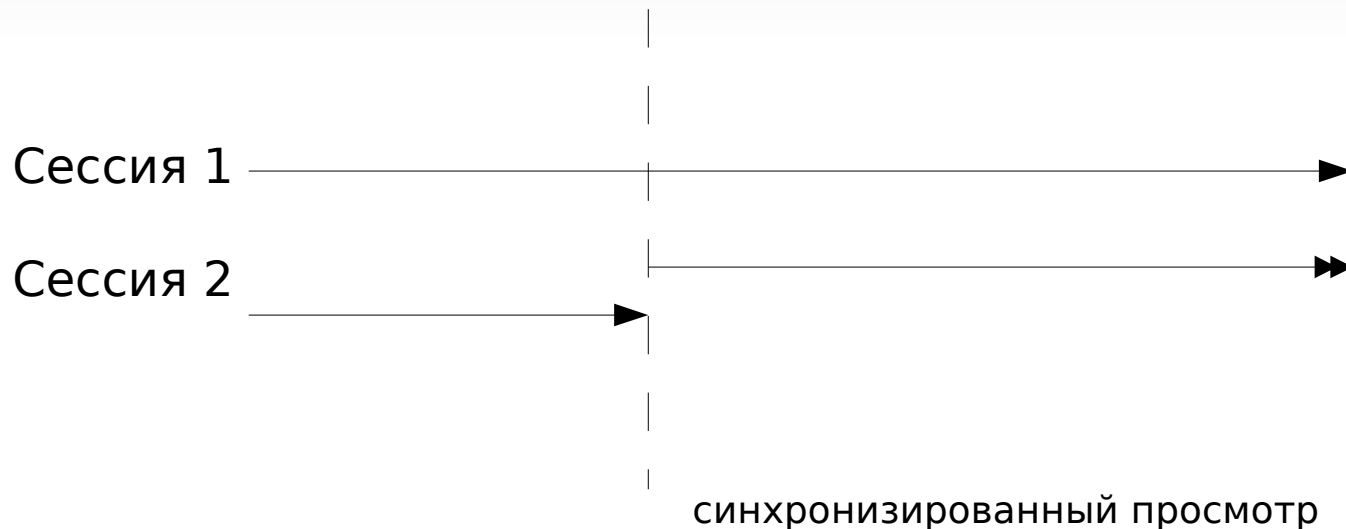
- «Размазывание» процесса `checkpoint` во времени, уменьшение интенсивности I/O-операций.
- Работает только для `background checkpoints` (при выключении сервера и явном выполнении команды `CHECKPOINT` – полная скорость I/O-операций).

# Устройство PostgreSQL



# Synchronized Scans

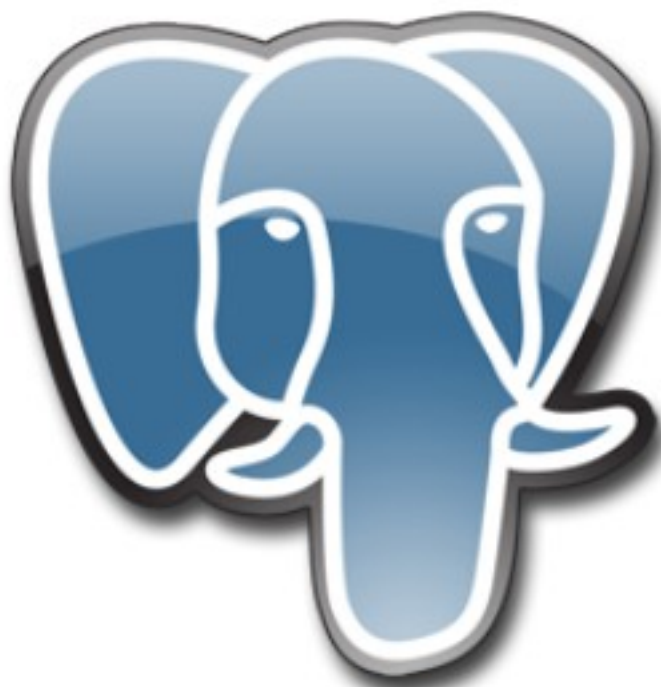
- Позволяет процессу «подсесть» к уже выполняющейся операции SeqScan другого процесса



# Кроме того

- Autovacuum включён по умолчанию
- Несколько параллельных процессов autovacuum
- Экономия пространства (varvarlena)
  - VARLENA-значения менее 128 байт: -6 байт!
  - заголовок строки: -4 байта!
- При «читающих» транзакциях xid не меняется
- Автонастройка bgwriter
- ORDER BY ... LIMIT w/o sorting («Top N» sorting)

# Разработчикам баз данных



# Полнотекстовый поиск

- contrib/tsearch2 --> ядро системы
- Новый синтаксис
- Упрощённое создание индекса (без триггера)
- Простая миграция и апгрейд версии PostgreSQL с (8.3 на следующие версии)



# Полнотекстовый поиск

```
CREATE TEXT SEARCH DICTIONARY russian_ispell (  
    TEMPLATE = ispell,  
    DictFile = russian, -- расширение (.dict) добавится автоматически  
    AffFile = russian,  
    StopWords = russian  
);
```

```
ALTER TEXT SEARCH CONFIGURATION utf8_russian  
    ADD MAPPING FOR hword, word  
        WITH russian_ispell, russian_stop;
```

```
CREATE INDEX i_person__tsvector ON person  
    USING gin(obj_tsvector);
```

```
CREATE INDEX i_person__tsvector ON person  
    USING gin(to_tsvector('russian', person_name ||  
        person_description));
```

# Полнотекстовый поиск

```
SELECT
    ts_rank_cd(obj_tsvector, q) AS rank,
    person_name
FROM
    person,
    plainto_tsquery('футбол ясенево') q
WHERE
    q @@ obj_tsvector
ORDER BY
    rank DESC LIMIT 50;
```

# Полнотекстовый поиск: миграция на 8.3 (Иван Золотухин для MirTesen.ru)

- Шаг 1: дампит базу, разделяя на три части:
  - схема без ограничений целостности и индексов
  - данные
  - ограничения целостности и индексы

При этом «вырезаем» подстроку «.tsearch2»:

```
pg_dump --exclude-schema=tsearch2 -U mirtesen mirtesen |  
sed -e "s|tsearch2\.||g" -e "s| gist_tsvector_ops||g" >  
mirtesen-schema-exclude-8.3.sql
```

# Полнотекстовый поиск: миграция 8.2->8.3 (Иван Золотухин для MirTesen.ru)

- Шаг 2: создаём скрипт для конфигурации словаря, используя новый синтаксис (ручная работа, увы):
- Шаг 3: «всасываем» всё в соответствующем порядке

# Поддержка XML

- Новый тип данных: `xml`
- Стандартные (ISO/ANSI SQL) функции публикации SQL/XML
- DTD-валидация
- XPath: функция `xpath()` (с поддержкой Namespaces)
- Альтернативные функции экспорта объектов БД в XML
- B-tree и GIN-индексы для XML-данных, полнотекстовый поиск

# Поддержка XML

SQL:

```
SELECT XMLROOT (
  XMLELEMENT (
    NAME 'some',
    XMLATTRIBUTES (
      'val' AS 'name',
      1 + 1 AS 'num'
    ),
    XMLELEMENT (
      NAME 'more',
      'foo'
    )
  ),
  VERSION '1.0',
  STANDALONE YES
);
```

Результат:

```
<?xml version='1.0'
      standalone='yes' ?>
<some name='val'
      num='2'>
  <more>foo</more>
</some>
```

# Поддержка XML

```
SELECT *
FROM table1
WHERE ((xpath('//person/@smp1:pid', xdata,
  ARRAY[ARRAY['smp1', 'http://example.com']]))[1])::text = '111';

CREATE INDEX i_table1_xdata ON table1 USING btree(
  (((xpath('//person/@smp1:pid', xdata,
    ARRAY[ARRAY['smp1', 'http://example.com']]))[1])::text)
);
```

---

```
SELECT *
FROM table1
WHERE ARRAY['111'] <@ xpath('//person/@smp1:pid', xdata,
  ARRAY[ARRAY['smp1', 'http://example.com']])::text[];

CREATE INDEX i_table1_xdata ON table
  USING gin(xpath('//person/@smp1:pid', xdata,
  ARRAY[ARRAY['smp1', 'http://example.com']])::text[]);
```

# Типы данных

- GUID/UUID
- enum
- МАССИВЫ СОСТАВНЫХ ТИПОВ

# Кроме того

- Автоматическая инвалидация плана («чистка» кэша функций)
- Table Function (RETURN QUERY)
- Updatable Cursors
- ORDER BY NULLS FIRST/LAST (ISO/ANSI SQL)
- Поддержка NULL-значений в GiST-индексах

# Администраторам баз данных



# EXPLAIN: больше информации

## QUERY PLAN

---

```
Sort (cost=34.38..34.42 rows=13 width=176) (actual time=0.946..0.948 rows=6
loops=1)
  Sort Key: obj2tag.o2t_tag_name
  Sort Method: quicksort Memory: 18kB
  -> Hash Join (cost=19.19..34.14 rows=13 width=176) (actual time=0.812..0.835
rows=6 loops=1)
    Hash Cond: (obj2tag.o2t_tag_obj_id = tag.obj_id)
    -> Bitmap Heap Scan on obj2tag (cost=4.43..19.21 rows=13 width=172)
(actual time=0.079..0.093 rows=6 loops=1)
      Recheck Cond: (o2t_obj_obj_id = 179313440)
      Filter: (obj_status_did = 1)
      -> Bitmap Index Scan on u_obj2tag_ids (cost=0.00..4.43 rows=24
width=0) (actual time=0.037..0.037 rows=24 loops=1)
        Index Cond: (o2t_obj_obj_id = 179313440)
    -> Hash (cost=9.89..9.89 rows=389 width=12) (actual time=0.700..0.700
rows=389 loops=1)
      -> Seq Scan on tag (cost=0.00..9.89 rows=389 width=12) (actual
time=0.012..0.396 rows=389 loops=1)
    Total runtime: 1.064 ms
(13 rows)
```

# Кроме того

- contrib/pg\_standby – создание Warm Standby (Linux, Windows)
- Удаление пользователя вместе со всеми принадлежащими ему объектами
- GSSAPI
- Улучшенная сборка под Win32 (MS VC++)

# Кроме того

- Оценки стоимости вызовов функций:

```
ALTER FUNCTION log_data(text)
    COST 100000 ROWS 1;
```

- Per-function GUC settings:

```
ALTER FUNCTION log_data(text)
    SET synchronous_commit TO OFF;
```

- Создание таблиц по «подобию»:

```
CREATE TABLE dict2 (LIKE dictionary INCLUDING INDEXES);
```

# Проекты сторонних разработчиков



# PL/pgSQL Debugger

The screenshot shows the PL/pgSQL Debugger window titled "Debugger - adm.grant\_on\_functions". The main editor displays the following SQL code:

```
num integer;  
BEGIN  
num := 0;  
/* select all function's names in given schema  
FOR func IN  
    SELECT  
        proname, proargtypes  
    FROM  
        pg_proc AS pro  
    JOIN pg_namespace AS ns ON pro.proname:  
    WHERE  
        ns.nspname = schema_name  
        AND proname LIKE mask
```

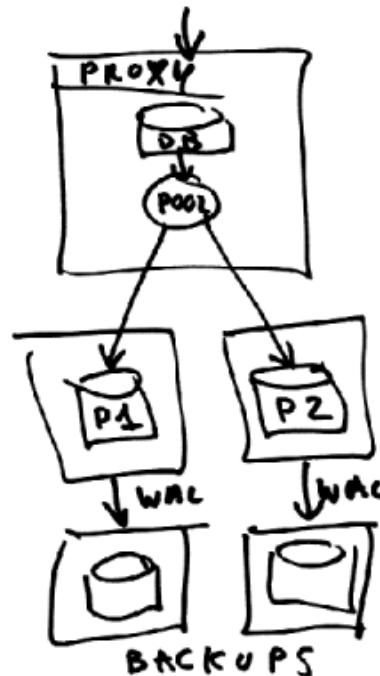
The line `num := 0;` is highlighted in green, and a red dot on the left margin indicates the current execution point. The "Stack pane" on the right shows the function `grant_on_func`. The "Output pane" at the bottom displays a table with the following data:

Name	Type	Value
role_name	text	nikolay
permission	text	read
mask	text	*
schema_nar	text	public

At the bottom of the window, it is noted "Paused at line 11" and the current cursor position is "Ln 16 Col 32 Ch 511".

# Продукты Skure

- PL/Proxy – прокси-язык для удаленного вызова процедур и партицирования данных между разными базами.
- PgBouncer – простой и быстрый менеджер соединений.
- SkyTools
  - WalMgr
  - Londiste
  - PgQ



# Кроме того

- pgSNMP – SNMP-агент для PostgreSQL (мониторинг)
- SEPostgres – расширение, основанное на модели SELinux (безопасность)
- phrrgadmin: настройка Slony и полнотекстового поиска, автовакуума
- Index Advisor



## Вопросы?

<http://postgresql.org>

<http://postgresmen.ru>

[http://developer.postgresql.org/index.php/Feature\\_Matrix](http://developer.postgresql.org/index.php/Feature_Matrix)

[nikolay@postgresmen.ru](mailto:nikolay@postgresmen.ru),

Skype: postgresmen

*Special thanks to: Bruce Momjian, Josh Berkus, Teodor Sigaev, «PFC»*

# Ежемесячные встречи сообщества

- Второй четверг каждого месяца?
  - 13 марта, 10 апреля, 8 мая, 12 июня...
  
- Только Москва?

# Активность сообщества

- Доклады на конференциях
  - РИТ-2008
  - Корпоративные базы данных-2008?
  - PHPConf
  - PostgreSQL Conference East ([postgresqlconference.org](http://postgresqlconference.org))?
  - PGCon-2008?
  - Highload-2008
- Статьи
  - «Системный администратор», LinuxFormat, ...?

# Активность сообщества

- ПЕРЕВОДЫ
  - ...статей
  - ...документации?

# Активность сообщества

- Курс для студентов ВМиК «**PostgreSQL: промышленная разработка баз данных**»
- **ОБЩЕНИЕ:**
  - `pgsql-ru-general` - ?
  - `sql.ru`???

**Слайды закончились :-)**