

Яндекс

Яндекс Облако

Индекс

DIY

Андрей Бородин

Яндекс

Индекс

Do It Yourself

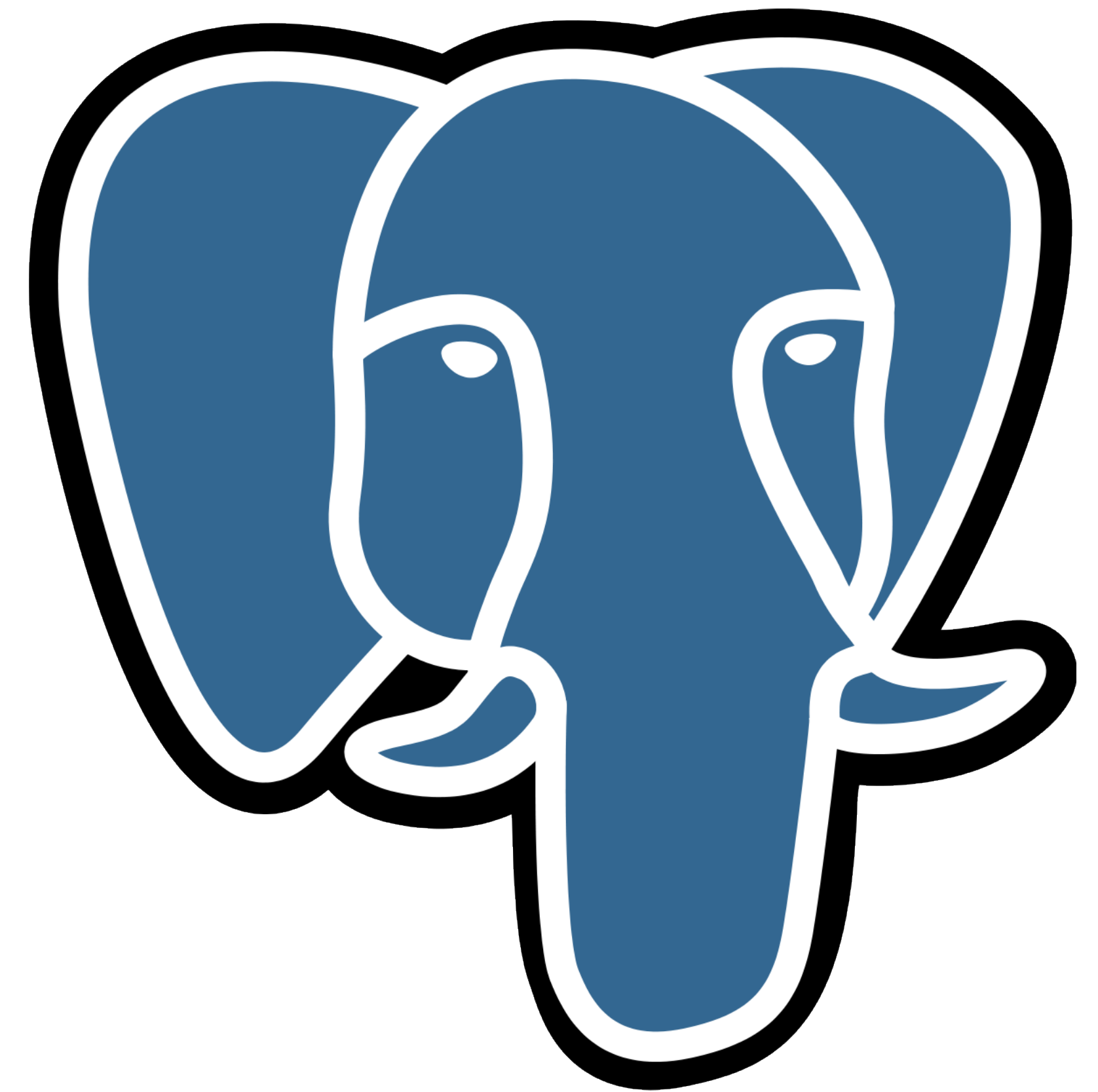
Андрей Бородин

Яндекс

Индекс PostgreSQL

Do It Yourself

Андрей Бородин





PostgreSQL extensibility: Origins and new horizons

Towards pluggable storage engines

Alexander Korotkov

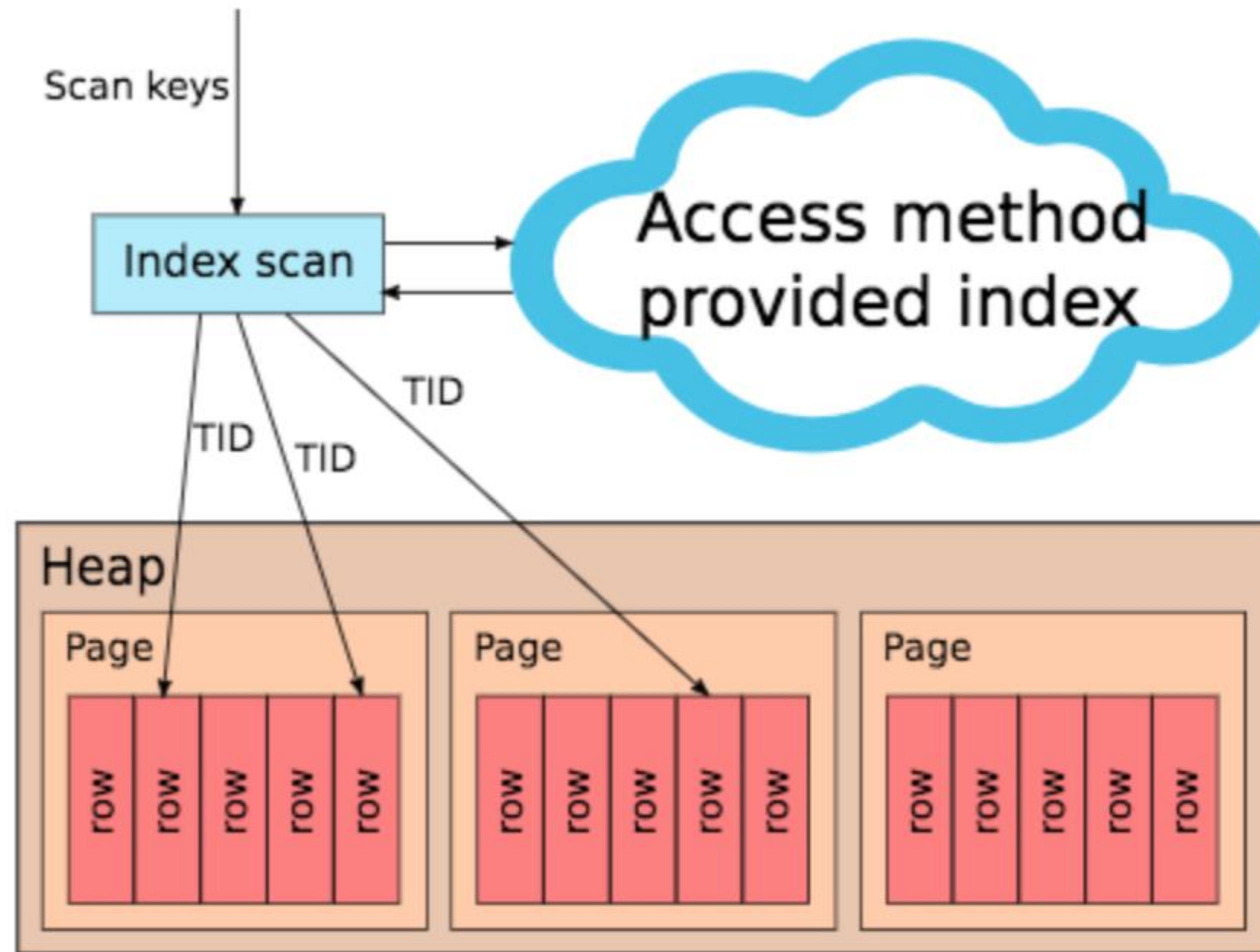
Postgres Professional

2016

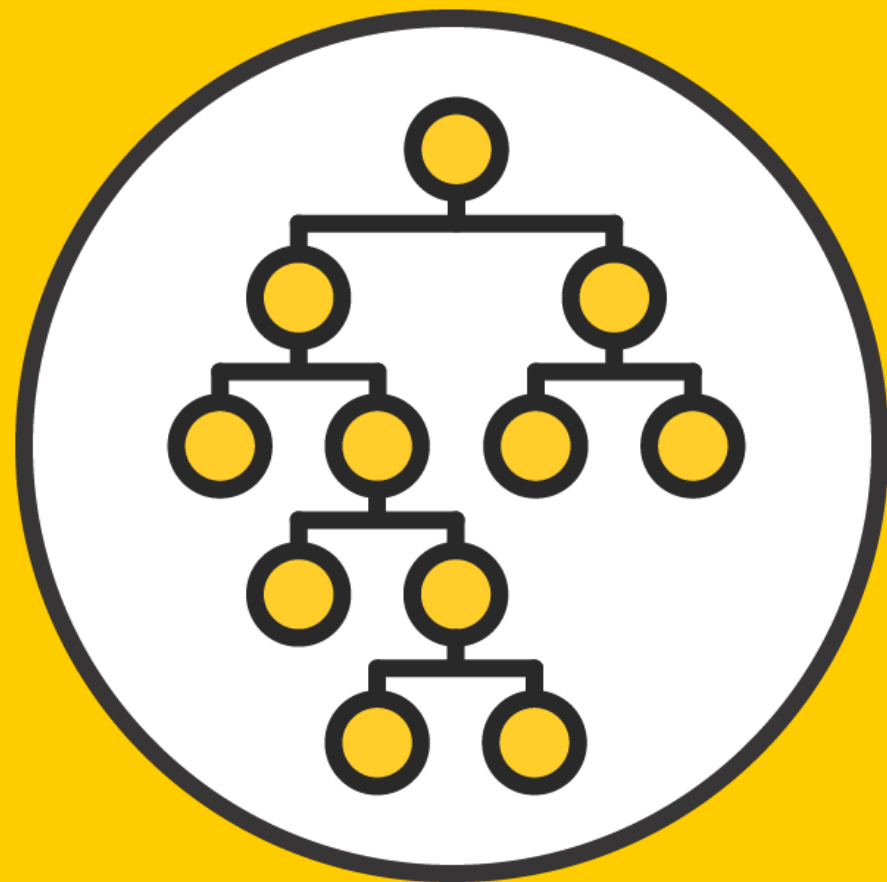
pg_am



PostgreSQL Index

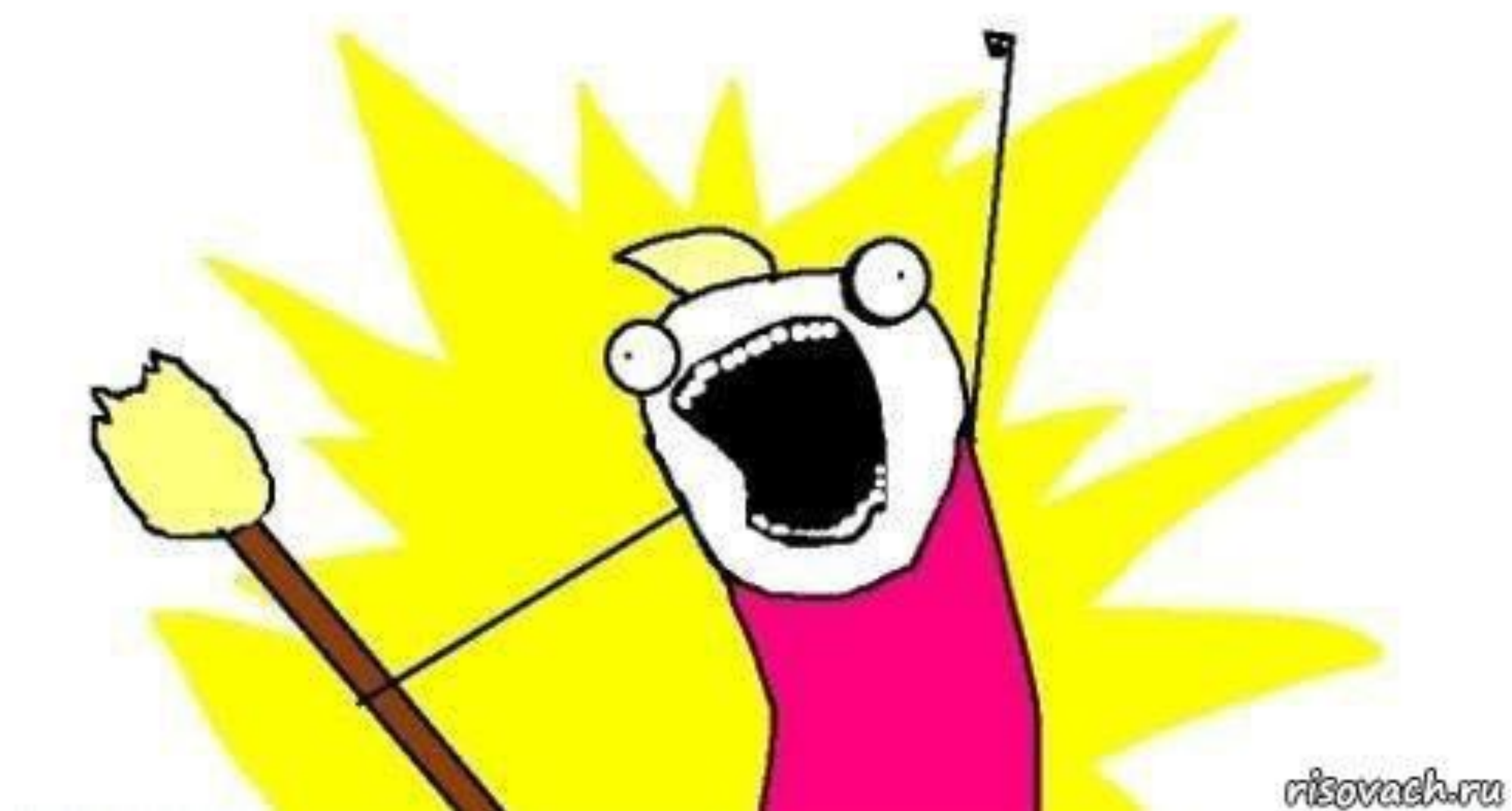


Как сделать индекс?



- › Создать extensions
- › Определить index handler
- › Вставка данных
- › Index Scan
- › VACUUM
- › WAL log
- › Что-то много всего?..

CODE!!



Что можно сделать,
например, с В-деревом?



BinSearch

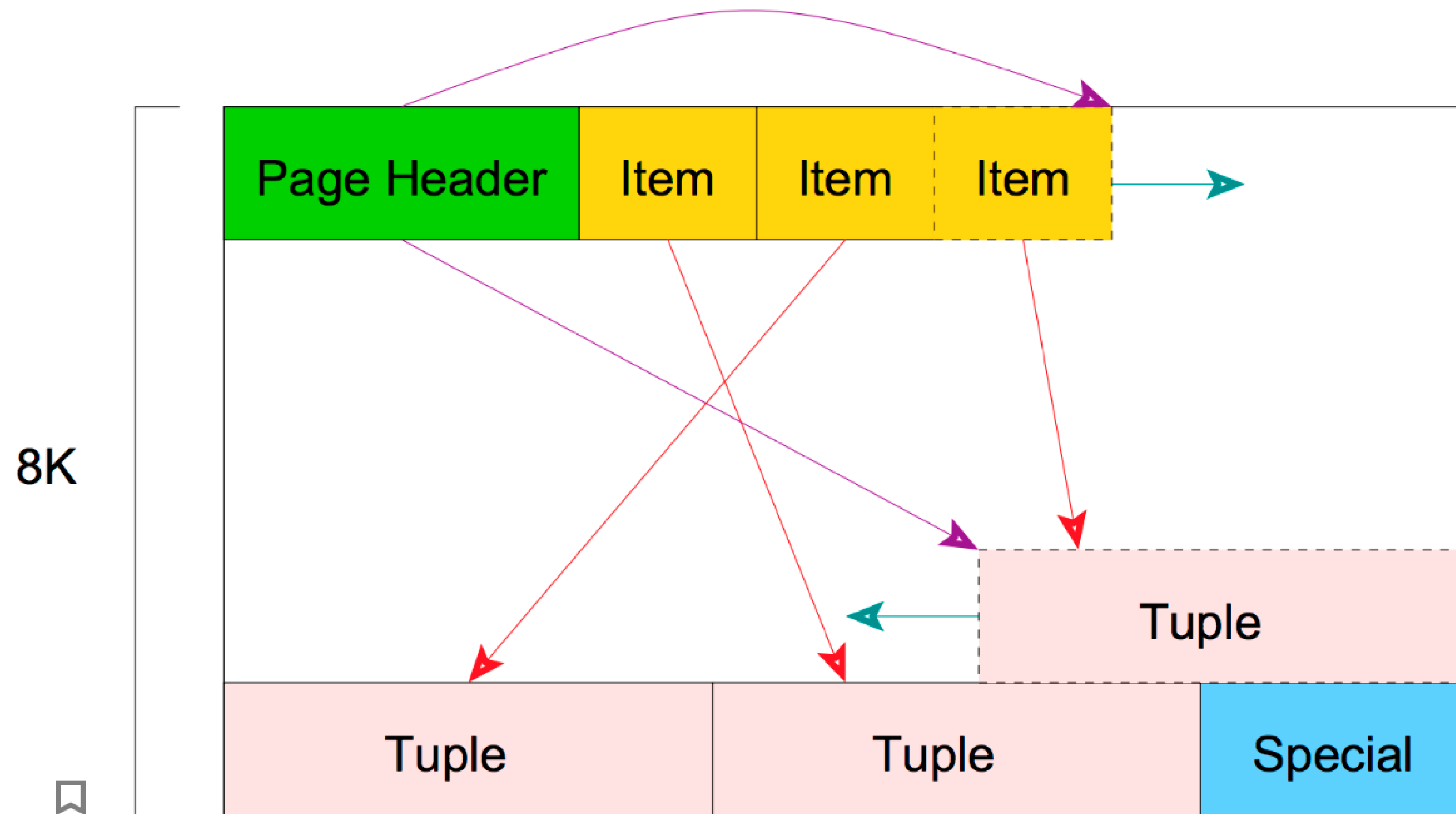
```
while (high > low)
{
    OffsetNumber mid = low + ((high - low) / 2);
    /* We have low <= mid < high, so mid points at a real slot */
    result = _bt_compare(rel, keysz, scankey, page, mid);
    if (result >= cmpval)
        low = mid + 1;
    else
        high = mid;
}
```

BinSearch

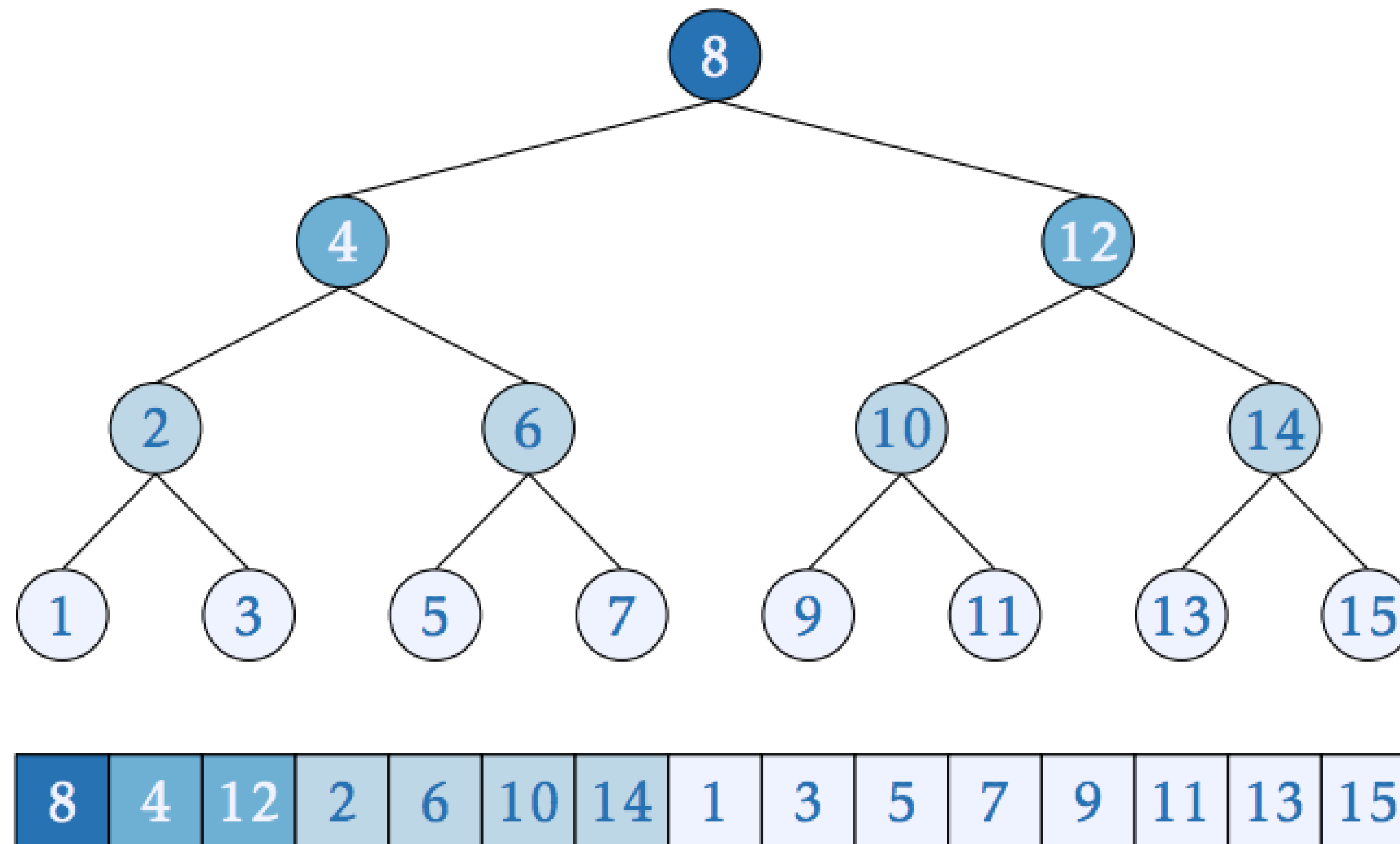
```
while (high > low)
{
    OffsetNumber mid = low + ((high - low) / 2);
    /* We have low <= mid < high, so mid points at a real slot */

    OffsetNumber x = mid + 1 + ((high - mid + 1) / 2);
    if (x < high)
        __builtin_prefetch (PageGetItem(page, PageGetItemId(page, x)), 0, 2);
    x = low + ((mid - low) / 2);
    if (x > low)
        __builtin_prefetch (PageGetItem(page, PageGetItemId(page, x)), 0, 2);

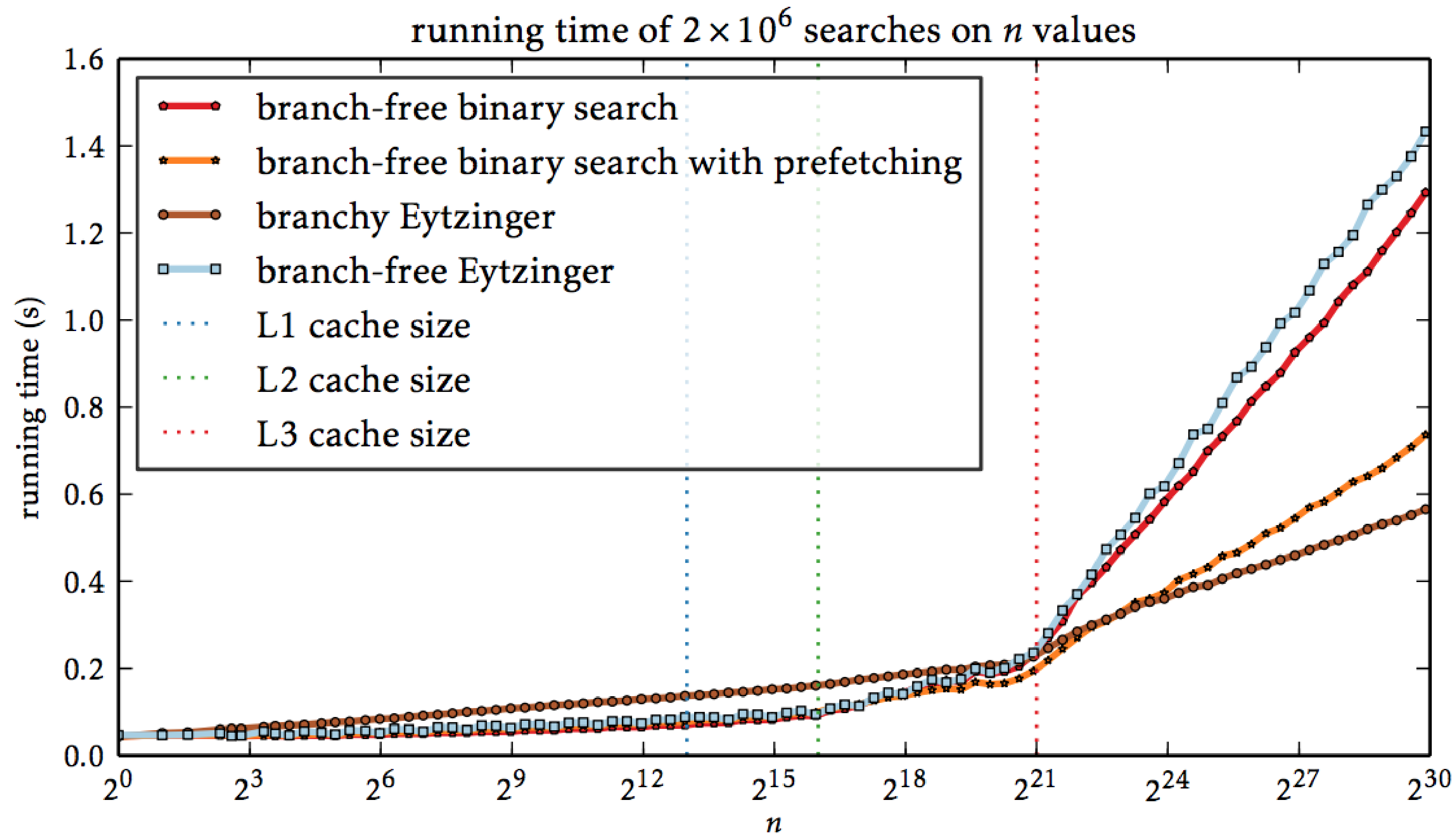
    result = _bt_compare(rel, keysz, scankey, page, mid);
    if (result >= cmpval)
        low = mid + 1;
    else
        high = mid;
}
```



The Eytzinger layout



The Eytzinger layout



Стратегии размещения

- › Sequential
- › B-tree
- › Van-Emde-Boas
- › Eytzinger

pgbench -i -s 25 && pgbench -T 100

	Prefetch OFF	Prefetch ON
baseline	1448.331199	1486.585895
sequential	1433.737204	1471.757042
Bt-order	1463.701527	1480.169967
Van-Emde-Boas	1457.586089	1464.834678
Eyzinger	1483.654765	1460.323392

PostgreSQL – это же open source

Может попробовать
сделать это для ванильного
PostgreSQL?



Covering + unique indexes

2015-12-01 11:33:45	Anastasia Lubennikova (lubennikovaav)	Created patch record
---------------------	---------------------------------------	----------------------

Covering + unique indexes

2018-04-07 20:35:37	Andres Freund (andresfreund)	Closed in commitfest 2018-03 with status: Committed
---------------------	------------------------------	---

Что ещё можно сделать с B-деревом?



ACM ICPC-style улучшение



Хардкод класса операторов

Suffix & Prefix Truncation

AAAAAA**A**AAAAA

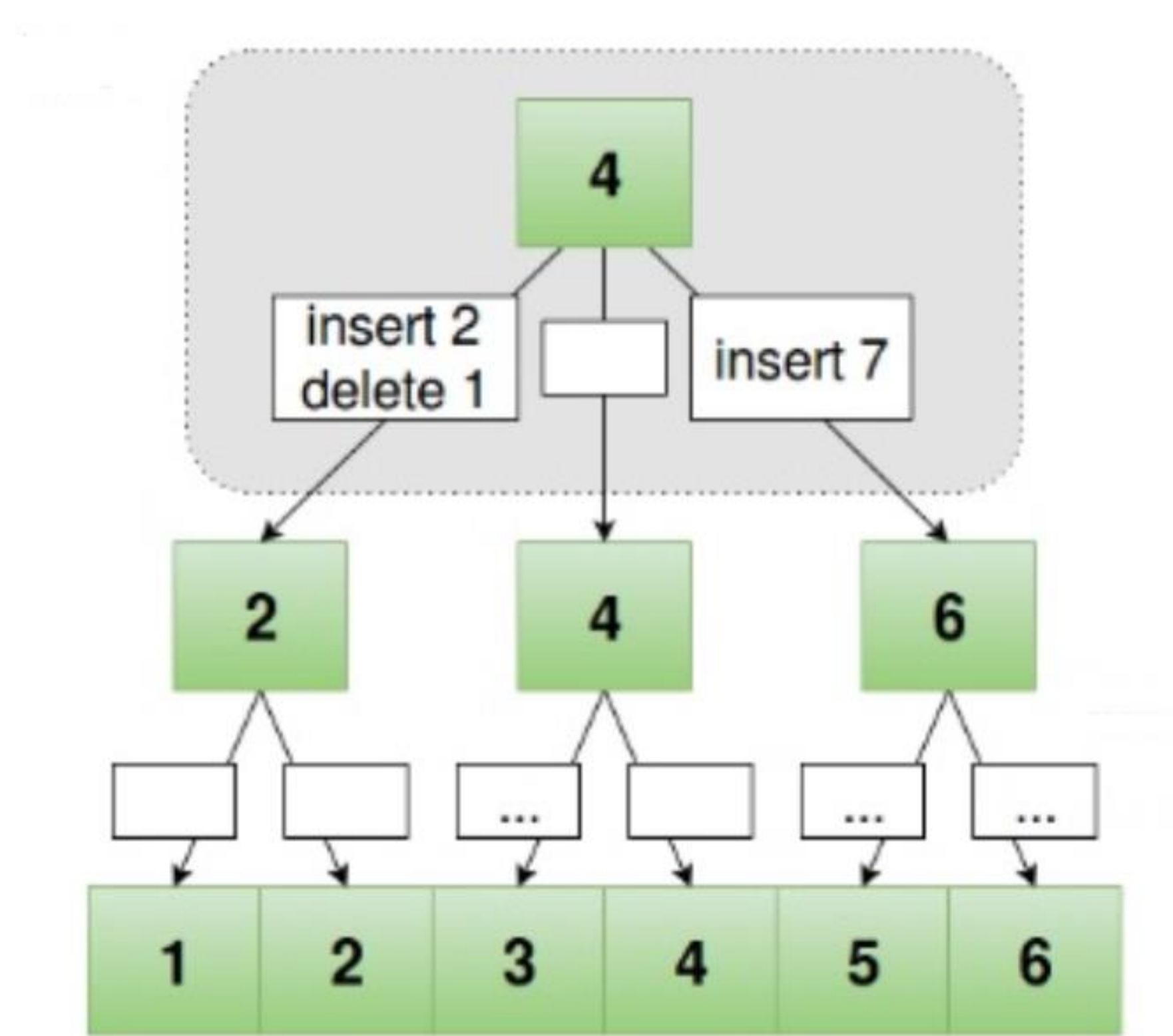
AAAAAA**B**BBBBBB

AAAAAA**C**CCCCC

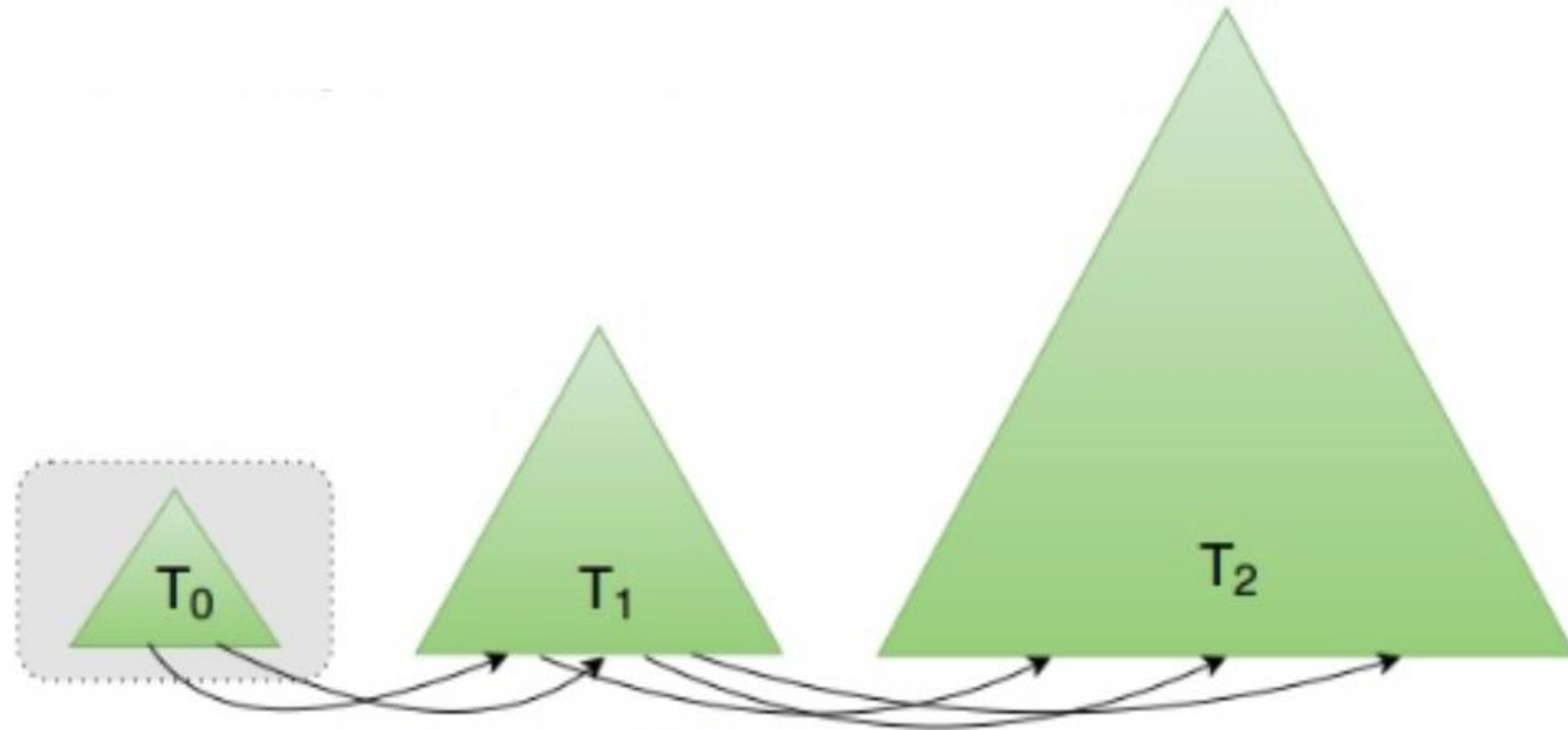
AAAAAA**D**DDDDDD

AAAAAA**E**EEEEEEEE

Fractal Tree



LSM-trees



И ещё много всего

- › COLA
- › Bw-trees
- › ...

 Но всё это уже совсем не DIY

- › Индексы – не всегда ускоряют базу
- › Самодельные индексы-прототипы – тем более

Яндекс.Станция в подарок за самую быструю конфигурацию PostgreSQL

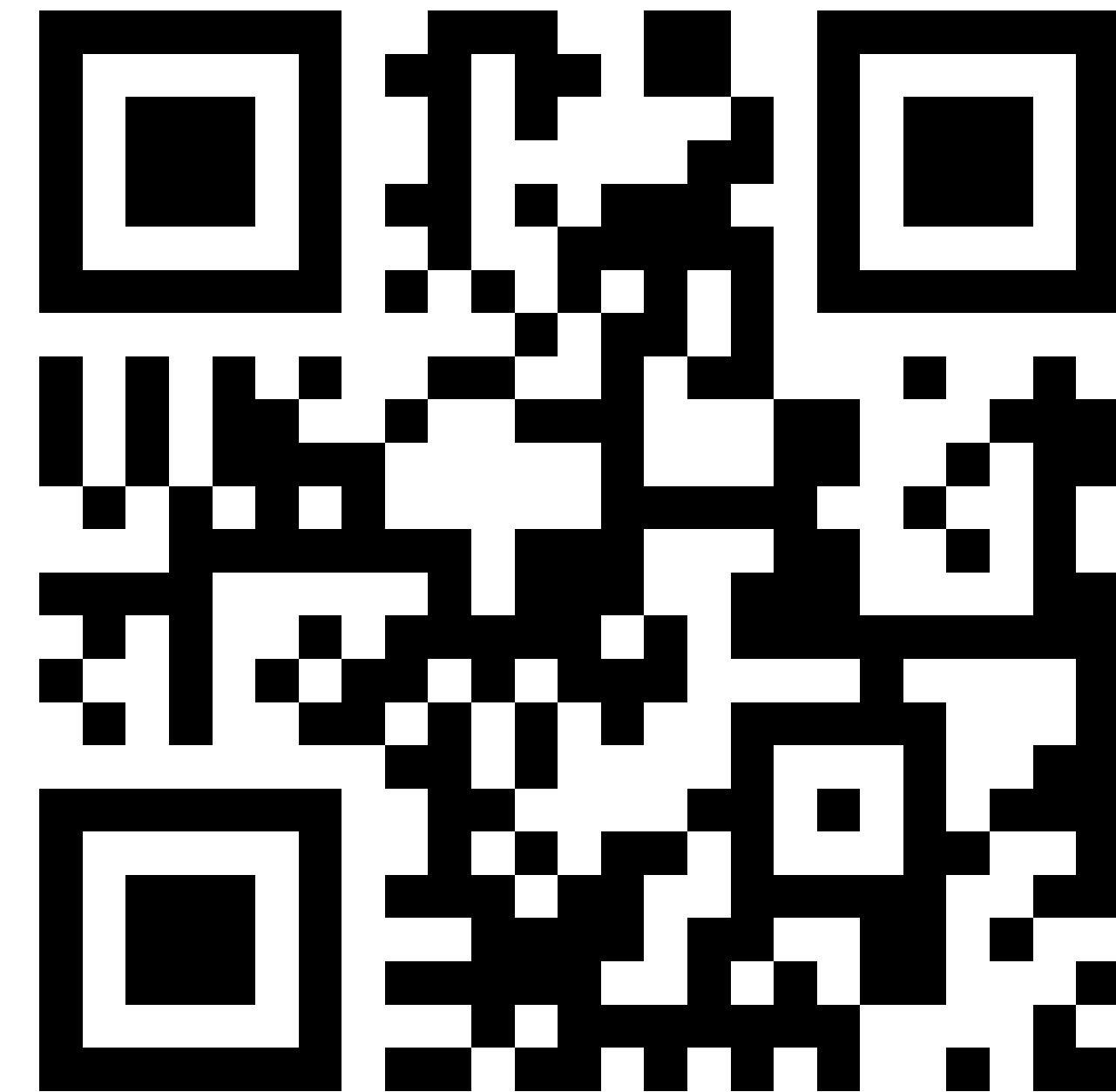
Настройте базу PostgreSQL для быстрой обработки запросов на чтение и запись. Если ваша база будет в десятке самых быстрых, вы получите приз, а победителю мы подарим **Яндекс.Станцию**.



**Чтобы начать, отсканируйте QR-код
или перейдите по ссылке**

cloud.yandex.ru/pg-game

Вручение призов на стенде **Я**ндекс.Облака
пройдёт в среду, 6 февраля, в 16:30.



Грант до 8 тысяч ₽ на тестирование платформы **Я**ндекс.Облако.

Стандартный грант — это 4 тысячи ₽.

На стенде **Я**ндекс.Облака можно увеличить его на сумму от 1 до 4 тысяч ₽.



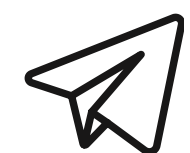
Жду вопросов 😊

Андрей Бородин

Разработчик



amborodin@acm.org



[x4mmm](https://t.me/x4mmm)