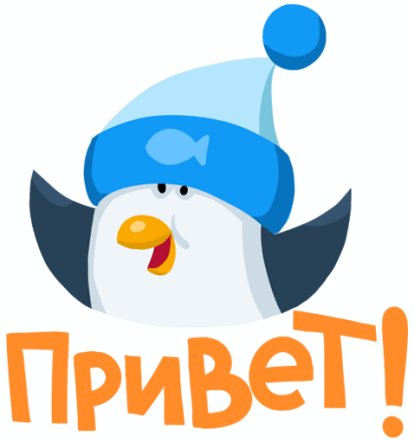
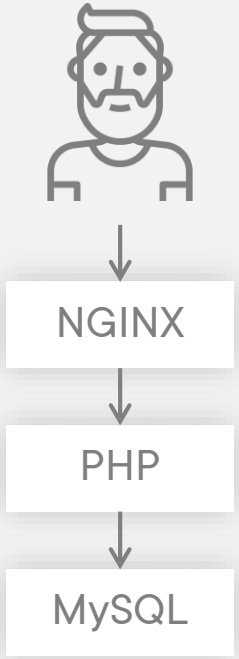


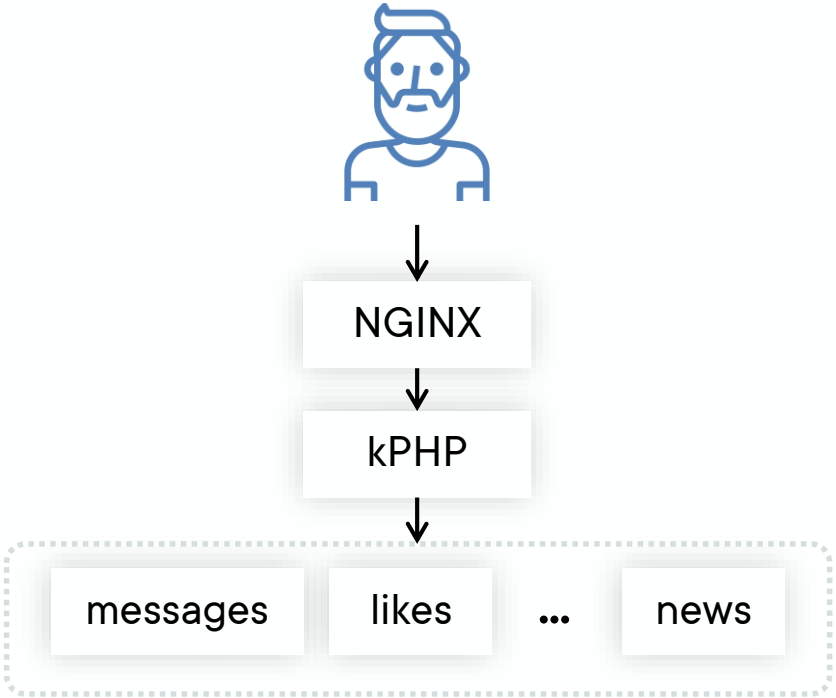
Как устроены базы данных ВКонтакте?

Борис Минаев

Типичный сайт



ВКонтакте



Что такое движок?

- Хранилище данных
- Бизнес логика
- C/C++
- Микросервис?

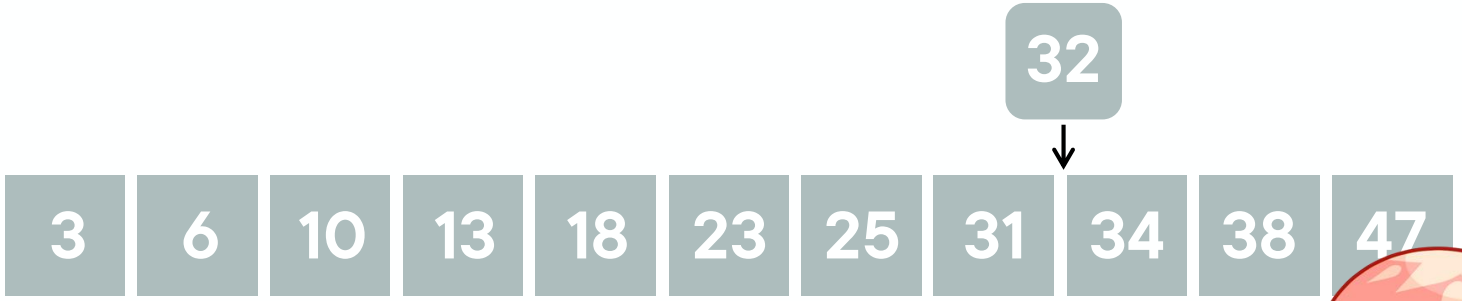

rpc-proxy
logs-db
persistent-longpoll
messages
antispam
blackkad
photo
filesys
memcached
queue
playlists
storage-redirect
notify
tlclient
cache
logs-stats
storage
kafka-proxy
money
smart-alerts
random
dns
sawnews
statsx
online
streaming-api
db-proxy
ipdb
views
likes
targ
copyfast
logs
friend
weights
logs-collector
ping
sandbox
places
seqmap
socket
tasks
graph
liked
magus
meowdb
watchcat
replicator
hints
expressivita
captcha
trees-storage
copyexec
mutual-friends
set
news
letters
search
pmemcached
image
password
lists
live-balancer
text
mc-proxy
nostradamus
isearch

Что получает ВКонтакте от использования движков?

- Скорость
- Эффективность хранение данных
- Атомарность
- Легкость добавления нового функционала

Мы очень любим оптимизировать

Как найти место числа в отсортированном массиве?

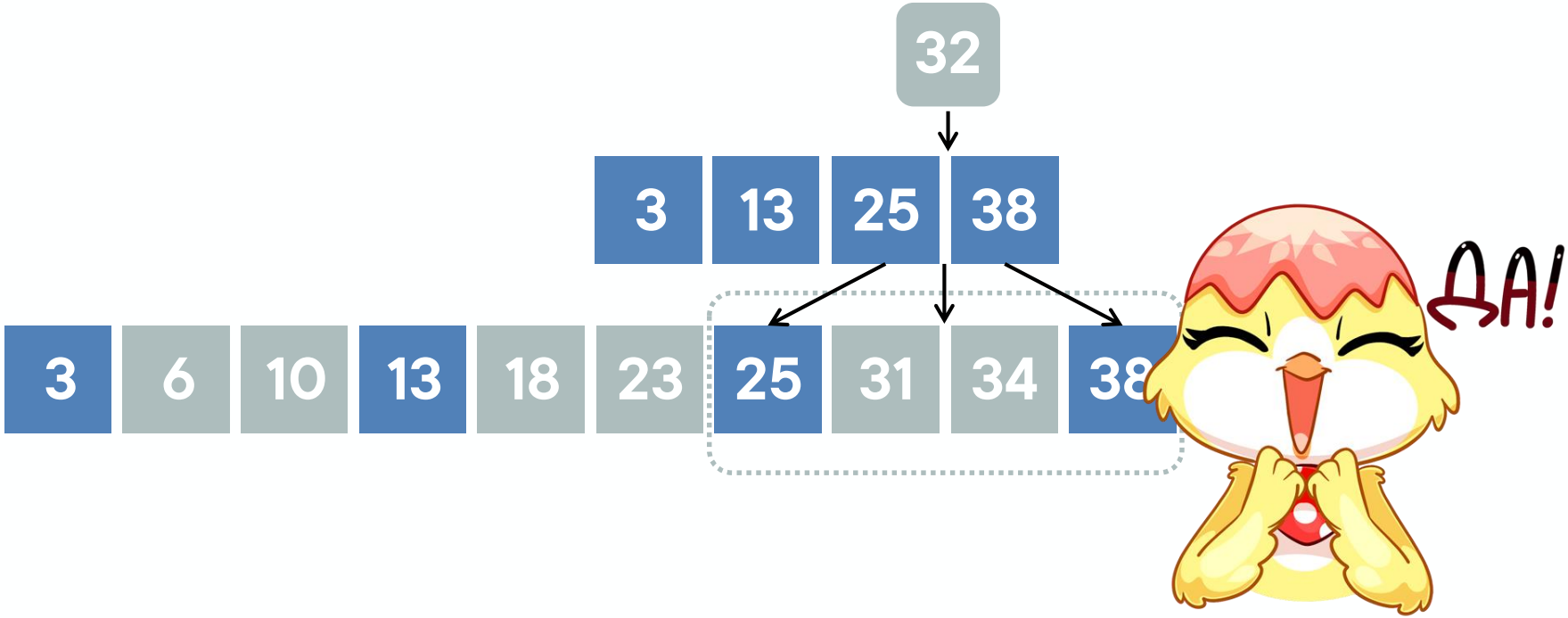


10 миллионов элементов

std::lower_bound?

Мы очень любим оптимизировать

1,5x быстрее



Зачем создавать свои БД в 2018?

Незачем*

А зачем тогда этот доклад?

- Узнать что-нибудь новое
- Перестать смотреть на базы данных как на черный ящик
- Интересно, как устроены высоконагруженные проекты

Пишем свою БД

- Общение с внешним миром
- Сохранение данных на диск
- Шардируем и реплицируем
- Оптимизации

Протокол общения (memcached-like)

```
printf "set mykey 0 60 4\r\ndata\r\n" | nc localhost 11211
STORED

printf "get mykey\r\n" | nc localhost 11211
VALUE mykey 0 4
data
END
```

read_inbound9822647,21524190,1441548

Протокол общения (RPC/TL)

```
messages.readMessages user_id:int peer_id:int up_to_local_id:int = Bool;
```

```
$query = ["_"                => "messages.readMessages",
          "user_id"          => 9822647,
          "peer_id"          => 21524190,
          "up_to_local_id"   => 1441548];
```

```
0xfbd90216 0x95e1b7 0x1486ede 0x15ff0c
```

Протокол общения

- Десятки тысяч движков
- Хотим гарантии tcp
- Не хотим держать коннекты ко всем
- Пишем свой!

Программируем, как умеем!

Бинарный лог событий

```
online.getUserStatus user_id : 169
online.setOnline user_id : 169 is_mobile : true
online.getUserStatus user_id : 169
online.setOnline user_id : 182 is_mobile : true
online.setOnline user_id : 119 is_mobile : false
online.getUserStatus user_id : 123
online.getUserStatus user_id : 149
online.getUserStatus user_id : 127
online.getUserStatus user_id : 115
online.setOnline user_id : 134 is_mobile : true
```

```
online.setOnline user_id : 169 is_mobile : true
online.setOnline user_id : 182 is_mobile : true
online.setOnline user_id : 119 is_mobile : false
online.setOnline user_id : 134 is_mobile : true
```

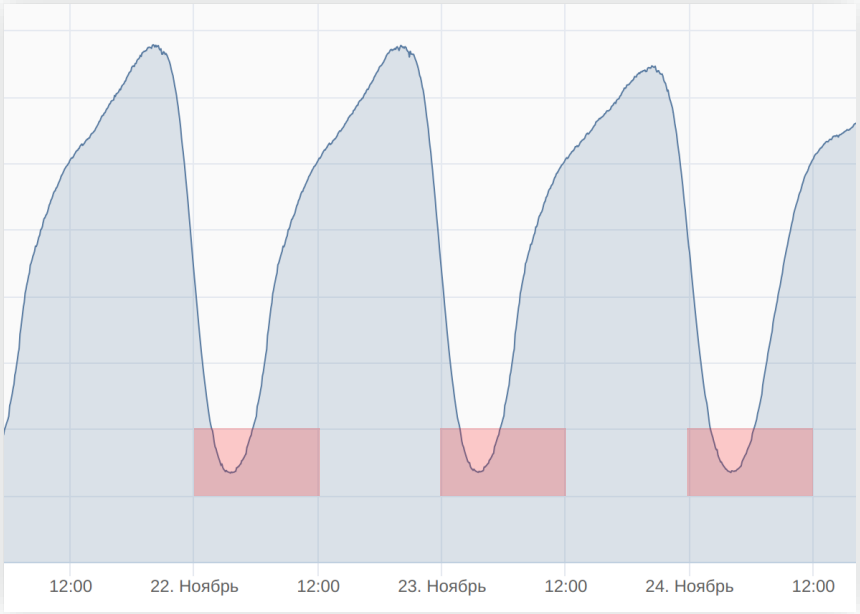
Снимок

```
online.setOnline user_id : 136 is_mobile : true
online.setOnline user_id : 142 is_mobile : true
online.setOnline user_id : 192 is_mobile : false
online.setOnline user_id : 191 is_mobile : true
online.setOnline user_id : 191 is_mobile : true
online.setOnline user_id : 142 is_mobile : true
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 142 is_mobile : true
online.setOnline user_id : 136 is_mobile : true
online.setOnline user_id : 192 is_mobile : false
online.setOnline user_id : 191 is_mobile : false
online.setOnline user_id : 192 is_mobile : false
online.setOnline user_id : 136 is_mobile : false
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 191 is_mobile : false
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 142 is_mobile : false
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 136 is_mobile : true
online.setOnline user_id : 191 is_mobile : false
online.setOnline user_id : 136 is_mobile : false
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 136 is_mobile : false
online.setOnline user_id : 192 is_mobile : true
online.setOnline user_id : 136 is_mobile : true
```

Оставим только нужное, отсортируем

```
online.setOnline user_id : 136 is_mobile : true
online.setOnline user_id : 142 is_mobile : false
online.setOnline user_id : 191 is_mobile : false
online.setOnline user_id : 192 is_mobile : true
```

Когда создавать снимки?



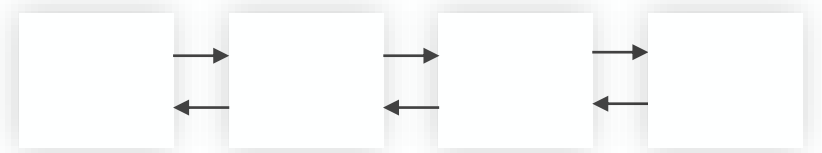
Метафайлы

Отсортированы лексикографически

```
"ad" -> "deafening"
"advertisement" -> "design"
"back" -> "obscene"
"battle" -> "horses"
"bubble" -> "continue"
"chance" -> "engine"
"changeable" -> "bad"
"check" -> "bruise"
"cloistered" -> "wealth"
"collect" -> "ritzy"
"cowardly" -> "abstracted"
"different" -> "children"
"division" -> "shiver"
"ear" -> "wide-eyed"
"free" -> "tongue"
"guard" -> "bright"
"hum" -> "soggy"
"ice" -> "pointless"
"loose" -> "hose"
"overrated" -> "hateful"
"reason" -> "rice"
"request" -> "statement"
"truthful" -> "youthful"
"wall" -> "floor"
"wire" -> "plucky"
```

метафайл
256kB

LRU

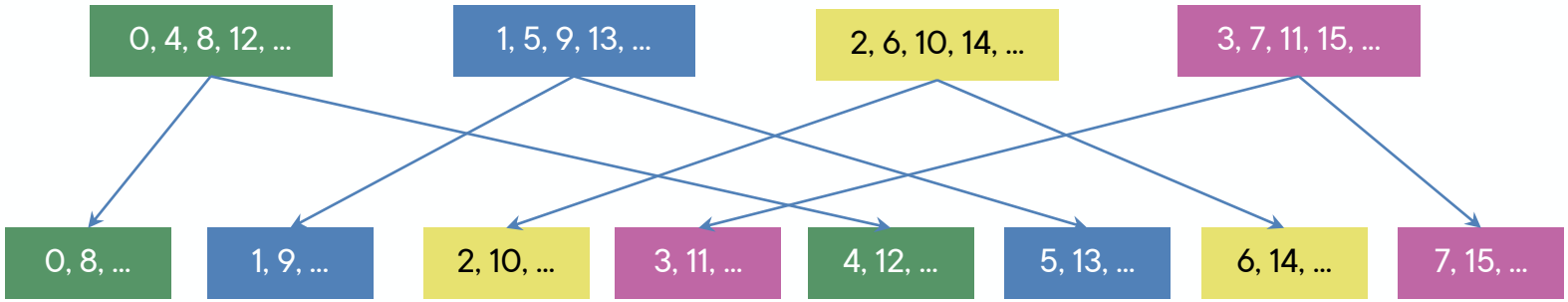


Использовали недавно

Использовали давно

Шардирование

- Чем меньше зависимости между данными на разных серверах, тем лучше
- User_id % shardes_num
- Увеличиваем кластера только в 2 раза



Оптимизации

- Вместо деревьев используем массивы
- Сжатие данных — простой способ получить выигрыш почти бесплатно
- Для каждого движка специфичны

Выводы

- Есть смысл писать свои БД, только если много серверов
- Нужно понимать, чем ваши данные специфичны
- Не нужно смотреть на БД как на черный ящик
- Полезно знать, как устроены высоконагруженные проекты