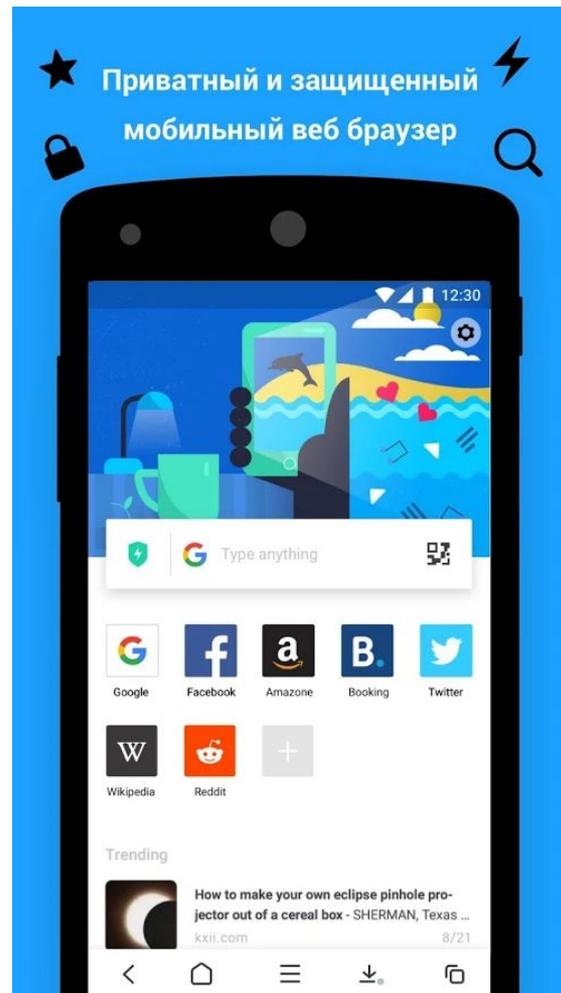


# ClickHouse

там где не ступала нога  
человека

# Обо мне

Владимир, backend-разработчик Aloha Browser.



# Мониторинг

- Если вы ничего не измеряете, то вы ничего не знаете
- Чем крупнее проект, тем больше нужно метрик
- Статус серверов, сервисов, приложений и т.д.
- Приложение на nginx/php

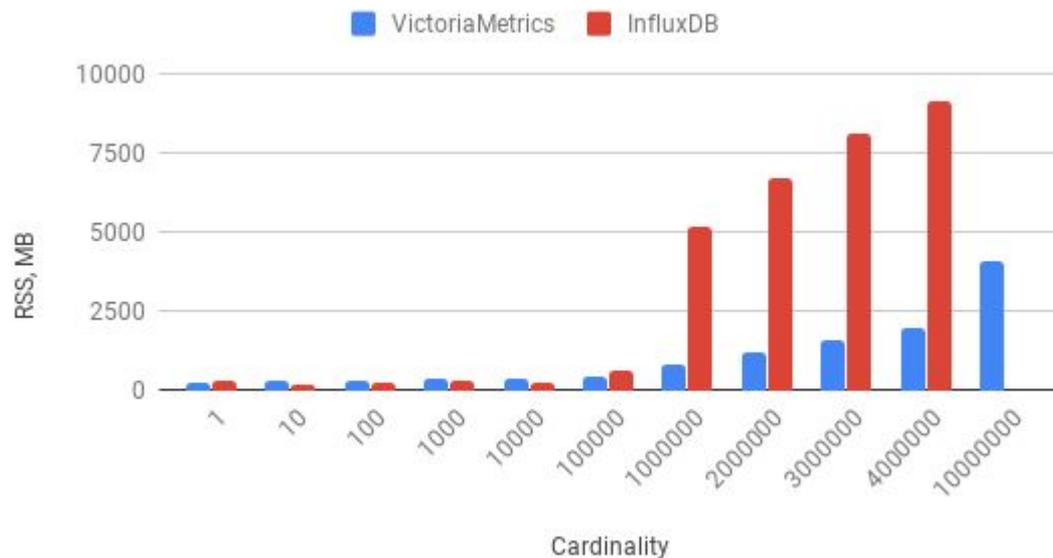
# Pinba

- Модуль для php/nginx
- cpu, memory, time..., protobuf, udp
- pinba\_engine, pinba2, pinboard
- pinba-server (go + OpenTSDB)
- pinba-influxdb (go + InfluxDB)
- proton (go + ClickHouse)
- pinba-server (php + ClickHouse)

# Influxdb

Адреса страниц сайта = теги

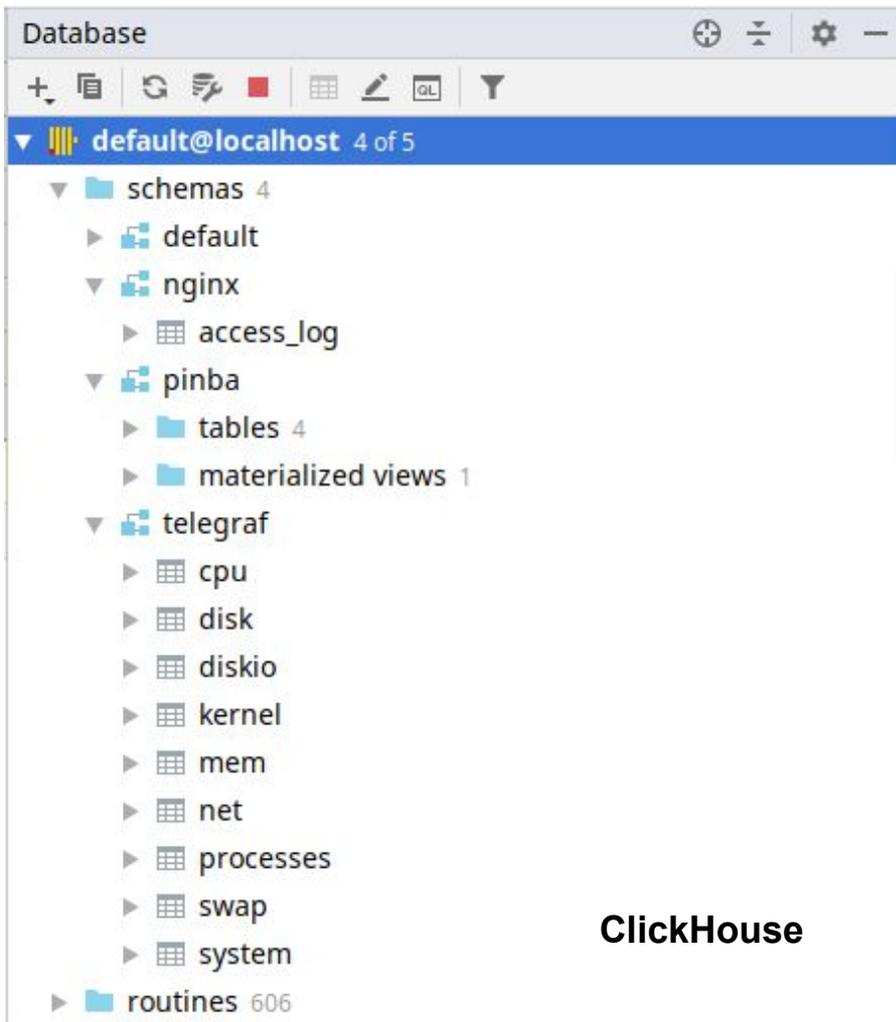
Большое количество тегов - большая проблема



# Influxdb

- Нет нормальных интерфейсов для написания запросов и просмотра данных
- Непривычная структура
  - SERIES
  - MEASUREMENTS
  - TAG KEYS
  - FIELD KEYS
- Проблема туннельного зрения

```
Connected to http://localhost:8086 version 1.7.5
InfluxDB shell version: 1.7.5
Enter an InfluxQL query
> █
```



**ClickHouse**

# ClickHouse спешит на помощь

- Полезные нам движки: MergeTree, SummingMergeTree, Null
- Поддержка массивов и вложенных массивов
- materialized view

```
tags Nested(  
  name String,  
  value String  
),  
timers Nested(  
  hit_count UInt32,  
  value Float32,  
  tag_name Array(String),  
  tag_value Array(String)  
),
```



```
tags.value[indexOf(tags.name, 'os')] as os,  
tags.value[indexOf(tags.name, 'country')] as country,
```

schema

web ▾

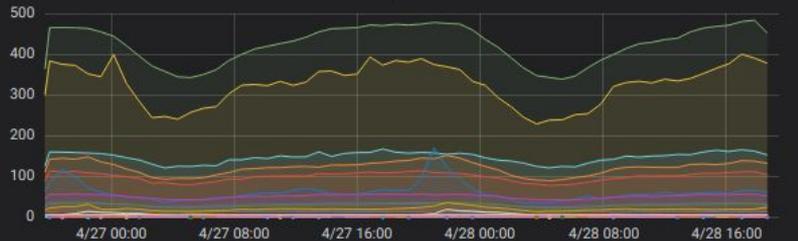
server name

All ▾

hostname

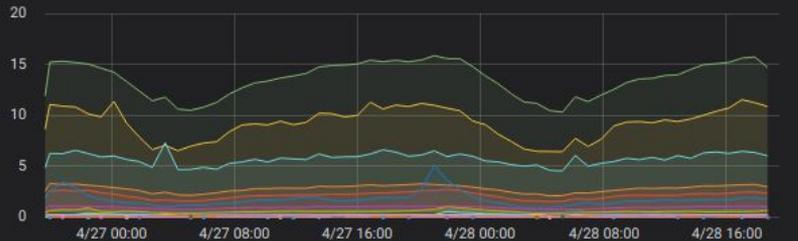
All ▾

req\_count



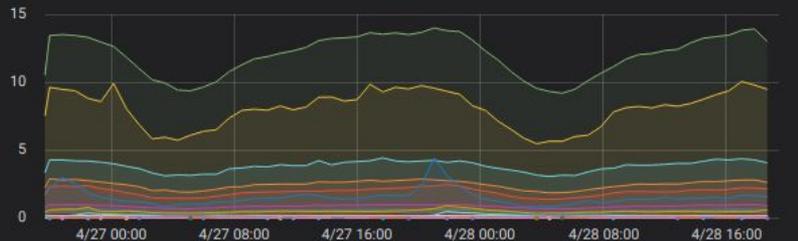
	max	avg	current
/api/status	483	421	452
/news/top	400	323	378
/api/postback	171	61	62

time



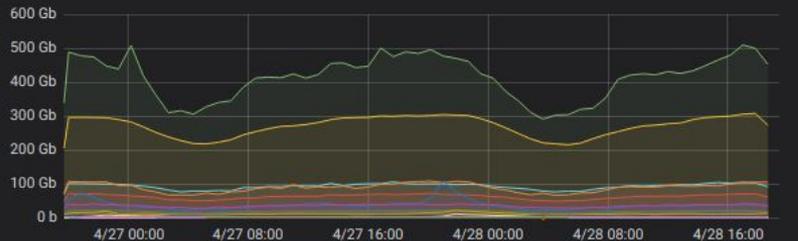
	max	avg	current
/admin/report/1/export	15.84	13.49	14.70
/user/login/facebook	11.53	9.15	10.86
/admin/push/send	7.30	5.70	6.02

cpu



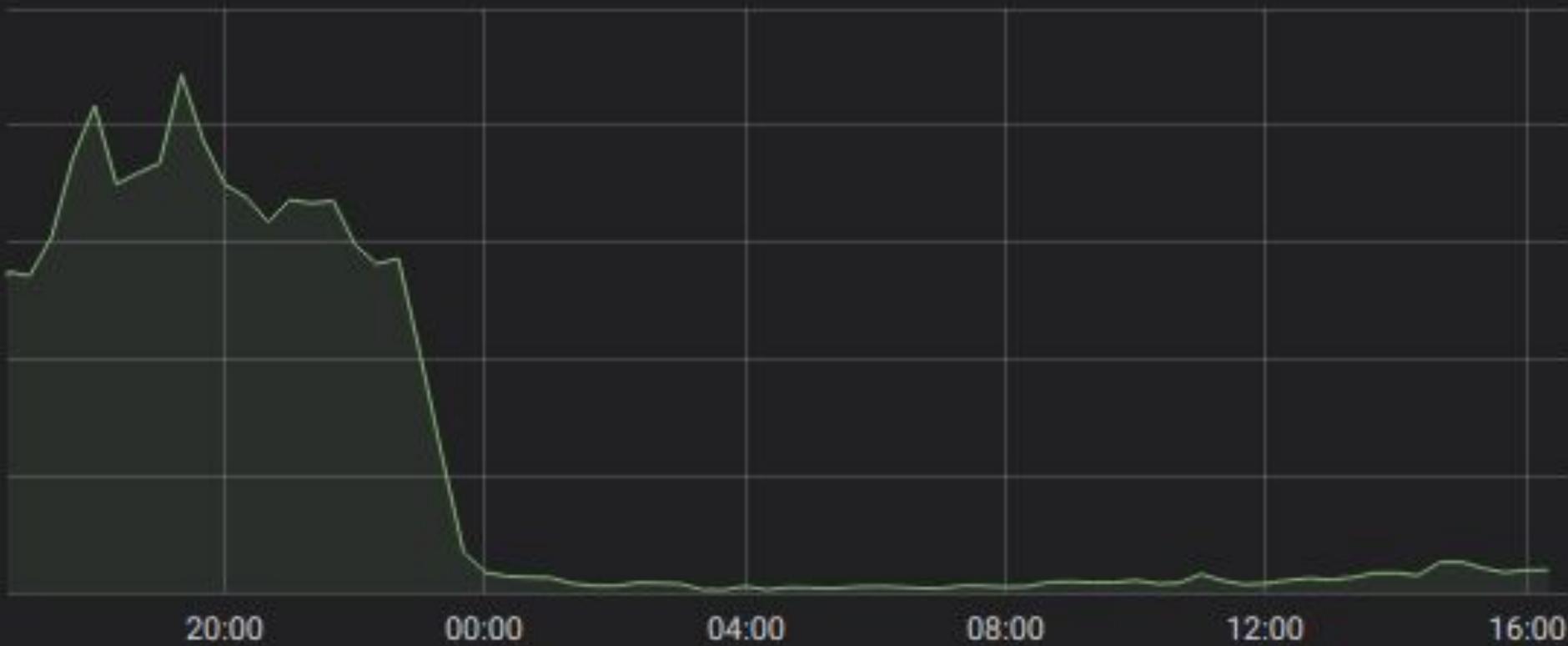
	max	avg	current
/admin/system/check_db	13.99	11.94	12.99
/admin/report/4/update	10.05	7.98	9.48
/api/abtest	4.43	3.81	4.06

memory



	max	avg	current
/admin/report/2/export	509 Gb	410 Gb	452 Gb
/admin/report/1/export	308 Gb	267 Gb	274 Gb
/admin/report/3/export	109 Gb	89 Gb	106 Gb

cpu



# nginxhouse

```
log_format json escape=json '{"@timestamp": "$time_iso8601", "remote_addr":  
"$remote_addr", "body_bytes_sent": "$body_bytes_sent", "status": $status, "request":  
"$request", "url": "$uri", "request_method": "$request_method", "response_time":  
$temppt, "http_referrer": "$http_referer", "http_user_agent": "$http_user_agent"}';
```

```
access_log syslog:server=127.0.0.1:1514 combined;
```

- исправление json
- изменение формата даты
- вставка пачками

# influxhouse

```
weather,location=us-midwest,season=summer temperature=82 1465839830100400200
```

- динамическое создание колонок
- вставка пачками

- ▼  telegraf
  - ▼  cpu
    -  time DateTime
    -  host String
    -  cpu String
    -  usage\_user Float64
    -  usage\_idle Float64
    -  usage\_nice Float64
    -  usage\_iowait Float64
    -  usage\_irq Float64
    -  usage\_guest\_nice Float64
    -  usage\_system Float64
    -  usage\_softirq Float64
    -  usage\_steal Float64
    -  usage\_guest Float64
  - ▶  disk
  - ▶  diskio
  - ▶  kernel
  - ▶  mem
  - ▶  net
  - ▶  processes
  - ▶  swap
  - ▶  system

# Логи в ClickHouse вместо Elasticsearch

- создание колонок на лету
- вставка пачками
- проблема вложенности
- проблемы json-функций

Делаются следующие допущения:

1. Имя поля (аргумент функции) должно быть константой;
2. Считается, что имя поля в JSON-е закодировано некоторым каноническим образом.  
Например, `visitParamHas('{ "abc": "def" }', 'abc')` = 1, но  
`visitParamHas('{ "\u0061\u0062\u0063": "def" }', 'abc')` = 0
3. Поля ищутся на любом уровне вложенности, без разбора. Если есть несколько подходящих полей - берётся первое.
4. В JSON-е нет пробельных символов вне строковых литералов.

# Вопросы?

3 автора лучших вопросов получают кепку в подарок ;)