zalando

NOSQL INSIDE SQL

strategy and tactics

Dmitry Dolgov

zalando

➔ Jsonb internals (a.k.a кишки)

zalando

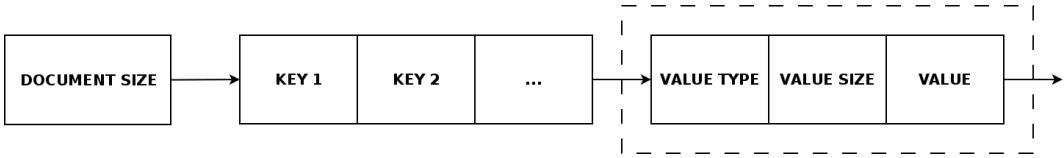
- ➔ Jsonb internals (a.k.a кишки)
- ➔ Performance-related factors

- Jsonb internals (a.k.a кишки)
- Performance-related factors
- How to shoot yourself in the foot

- Jsonb internals (a.k.a кишки)
- Performance-related factors
- How to shoot yourself in the foot
- Benchmarks

INTERNALS

Jsonb



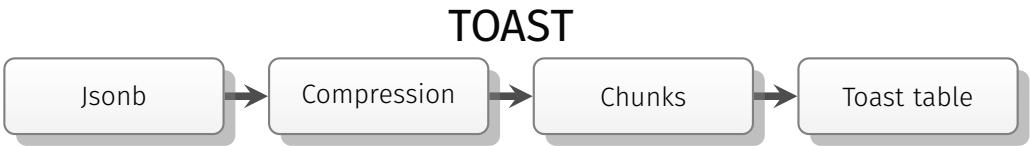
zalando

```
select pg_relation_filepath(oid),
relpages from pg_class
where relname = 'table_name';

 pg_relation_filepath | relpages
----------------------+----------
 base/40960/325477    |        0
(1 row)
```

```
\x10\x03\x00\x00\x00ab\x00\x00\x00\x00\x00\x00\x80\x03\x00xyz\x00\x00\x00\x00
```

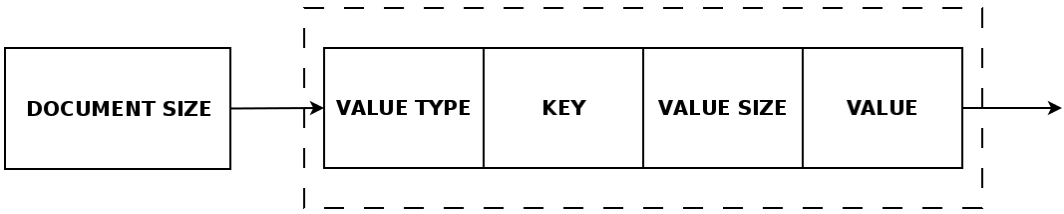

zalando



zalando

JB_OFFSET_STRIDE

Bson

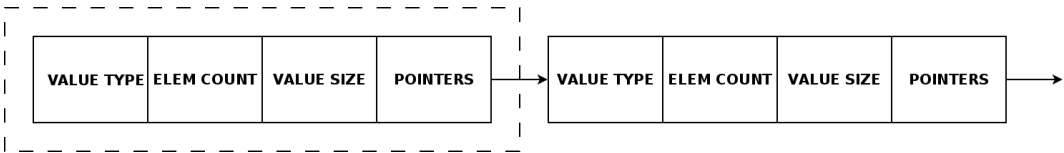


zalando

```
bson.dumps({"a": 3, "b": u"xyz"})
```

```
\x17\x00\x00\x00\x10a\x00\x03\x00\x00\x00\x02b\x00\x04\x00\x00\x00xyz\x00\x00
```

MySQL json



zalando

Performance-related factors

Performance-related factors

➔ Структура данных на диске

Performance-related factors

- ➔ Структура данных на диске
- ➔ Сериализация данных

Performance-related factors

- ➔ Структура данных на диске
- ➔ Сериализация данных
- ➔ Поддержка индексов

Сериализация данных

- MongoDB – дерево Document -> Elements
- Postgresql – JsonbValue со списком элементов
- MySQL – класс Value

Индексы

- Postgresql – общий индекс, индексы для полей
- MongoDB – индексы для полей
- MySQL – виртуальные колонки для индексирования

PG indexing details

- ➔ JGIN_MAXLENGTH
- ➔ Index rebuild
- ➔ Different types of index

ТЕСТИРОВАНИЕ

zalando


GREAT PERFORMANCE

PostgreSQL 9.5.4

MySQL 5.7.9

MongoDB 3.2.9

YCSB 0.9

10^6 rows and operations

AWS EC2

Воспроизводимость

erthalion/YCSB

erthalion/ansible-ycsb

AWS EC2

m4.xlarge instance
separate instance (database and generator)
16GB memory, 4 core 2.3GHz
Ubuntu 14.04
Same VPC and placement group
AMI that supports HVM virtualization type
at least 4 rounds of benchmark

Конфигурация

shared_buffers
effective_cache_size
innodb_buffer_pool_size
write concern level (journaled or transaction_sync)

Виды документов

“простой” документ
10 ключей и значений (100 символов)

“большой” документ
100 ключей и значений (200 символов)

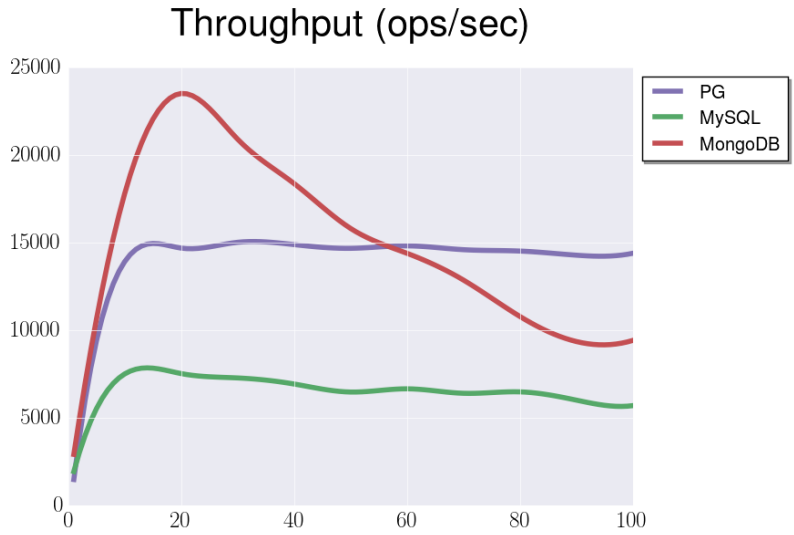
“сложный” документ
100 ключей, 3 уровня вложенности (100 символов)

Простая выборка по ключу с

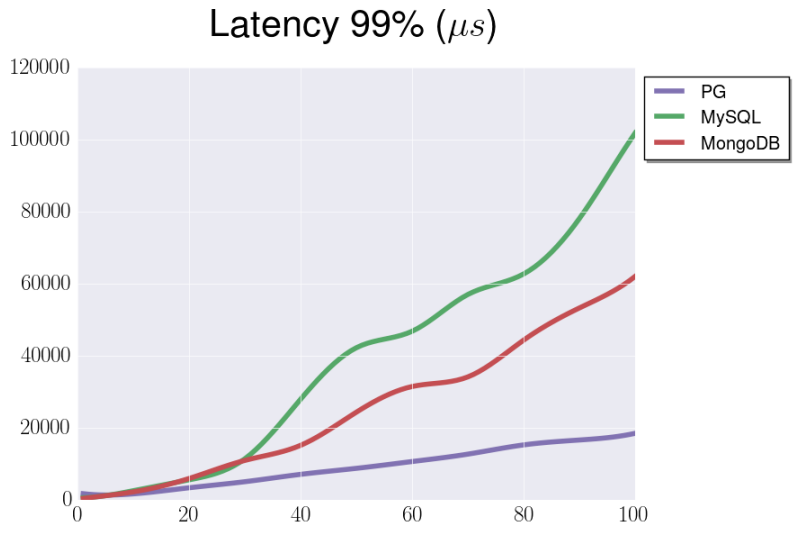
”Простой документ”

jsonb_path_ops

where data @> ’”key”: ”value”’::jsonb


Throughput (ops/sec)
25000
20000
15000
10000
5000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando


Latency 99% (μs)
120000
100000
80000
60000
40000
20000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

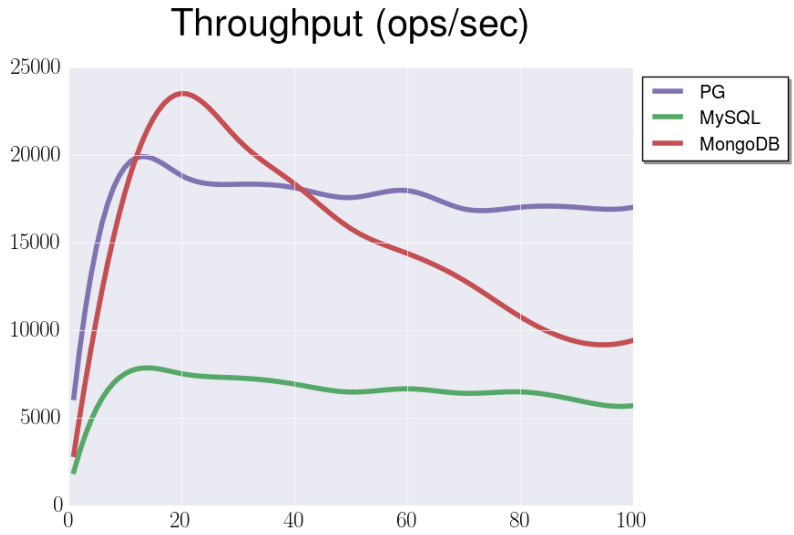
zalando

Простая выборка по ключу

”Простой документ”

jsonb_path_ops

where data @> jsonb_build_object(’key’, ’value’)

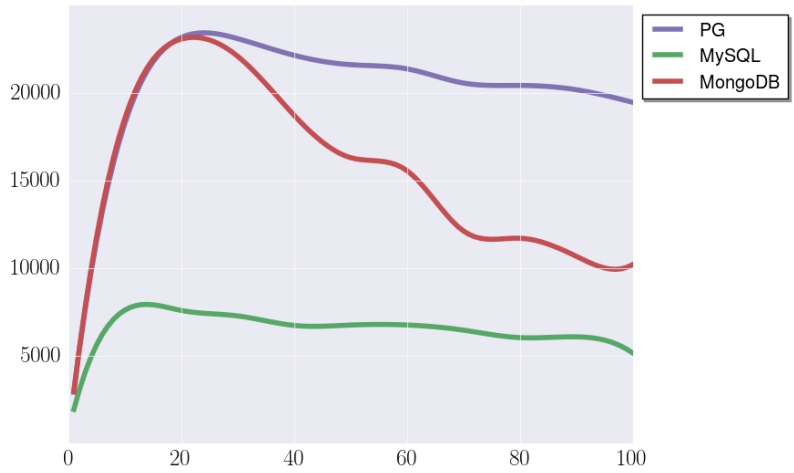

Throughput (ops/sec)
25000
20000
15000
10000
5000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando

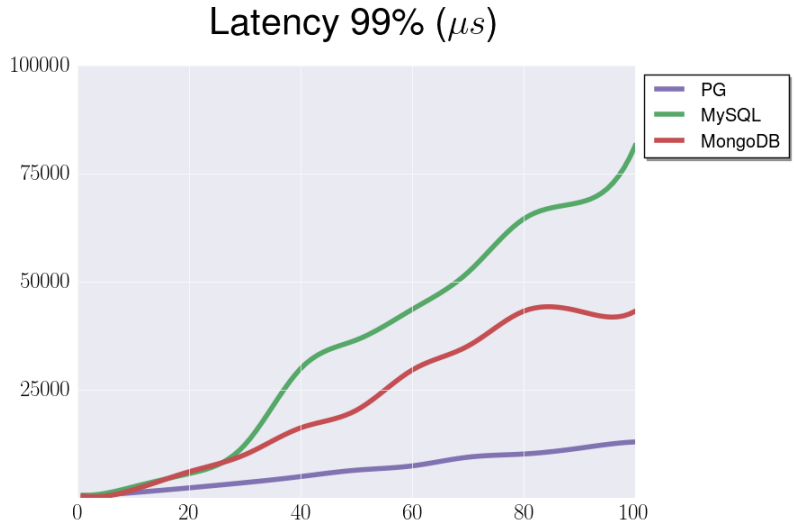
Простая выборка по ключу с Btree индексом

”Простой документ”

btree


Throughput (ops/sec)
PG
MySQL
MongoDB
20000
15000
10000
5000
0
20
40
60
80
100

zalando

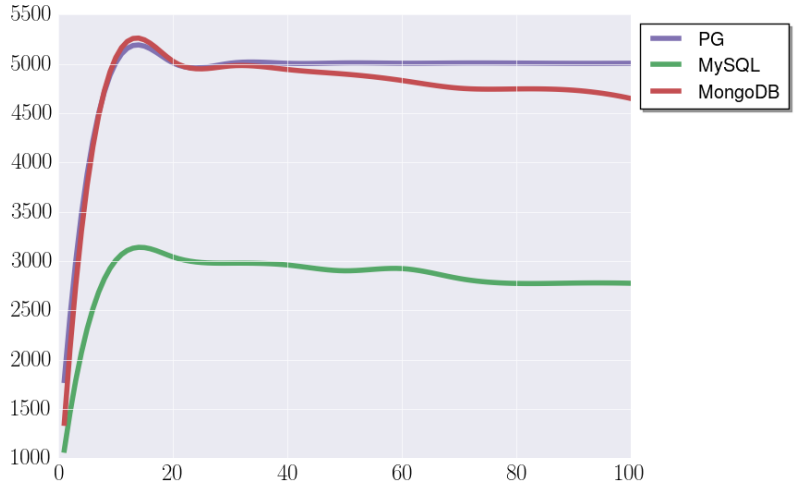

Latency 99% (μs)
100000
75000
50000
25000
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando

Простая выборка по ключу с Btree индексом

”Сложный документ”

btree


Throughput (ops/sec)
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
0
20
40
60
80
100
PG
MySQL
MongoDB

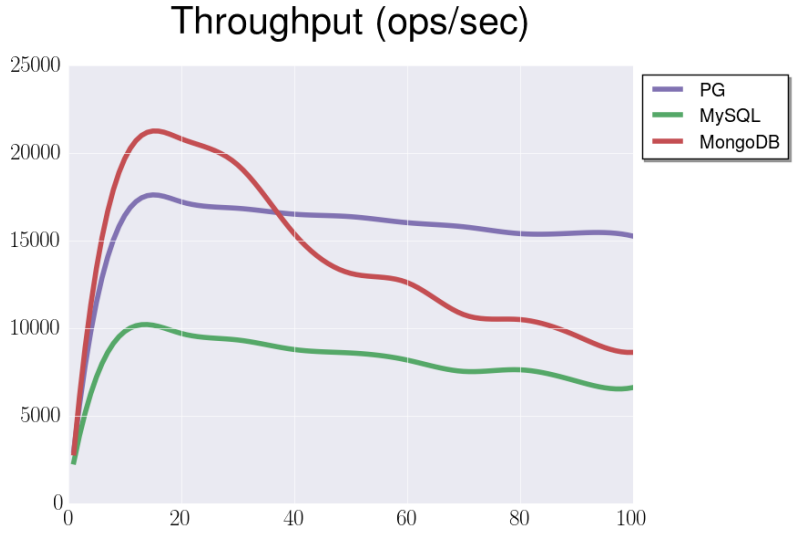
zalando

Срез по документу

”Большой документ”

Из документа выбирается одно поле

zalando


Throughput (ops/sec)
25000
20000
15000
10000
5000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

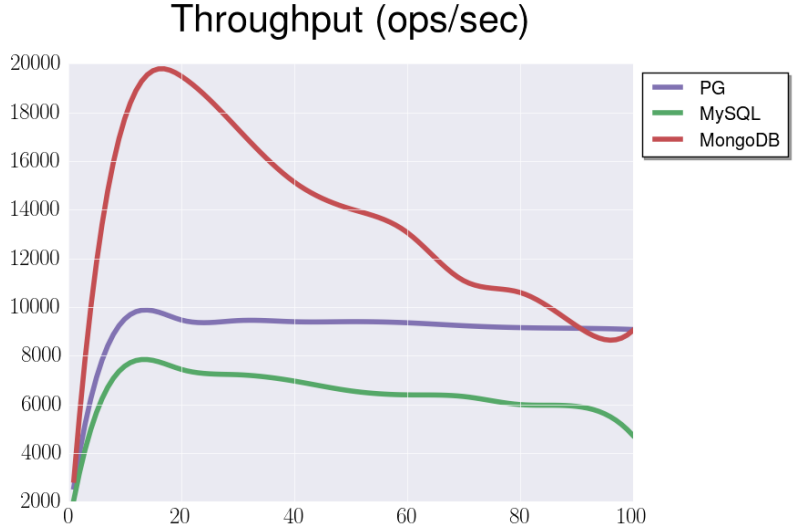
zalando

Срез по документу

”Большой документ”

Из документа выбирается 10 полей

zalando


Throughput (ops/sec)
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando


SET STORAGE EXTERNAL

Масштабируемость

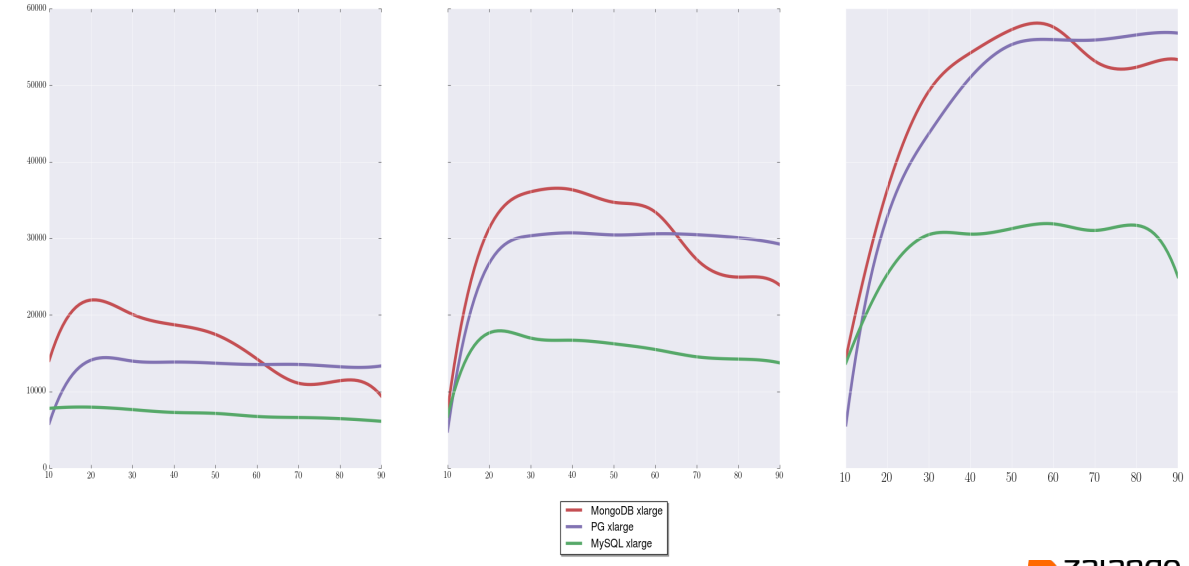
”Простой документ”

m4.large

m4.xlarge

m4.2xlarge

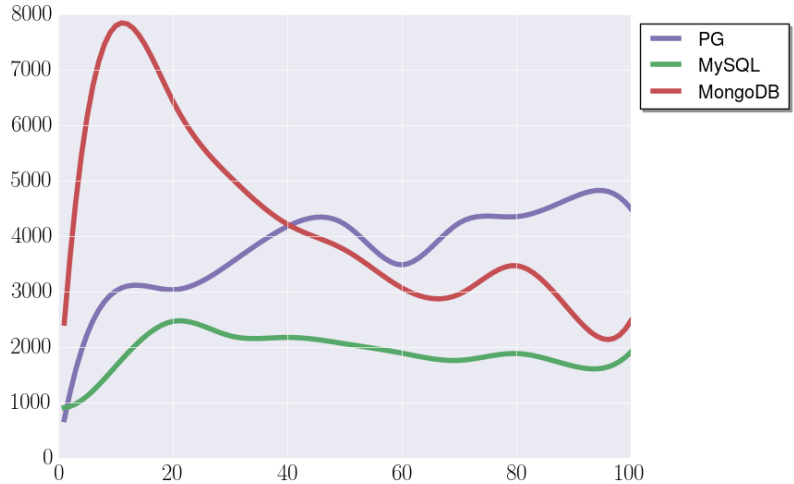
Throughput (ops/sec)



zalando

Вставка документов

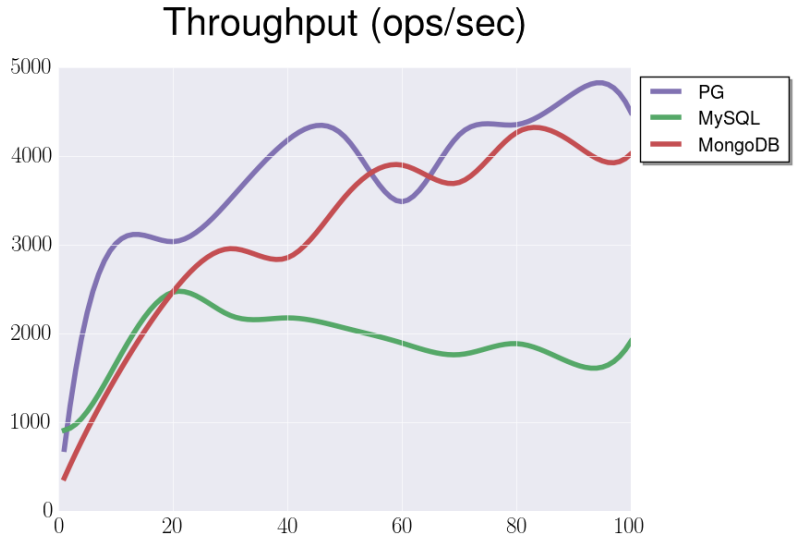
”Простой документ”
default write concern


Throughput (ops/sec)
8000
7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando

Вставка документов

”Простой документ”
journaled


Throughput (ops/sec)
5000
4000
3000
2000
1000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

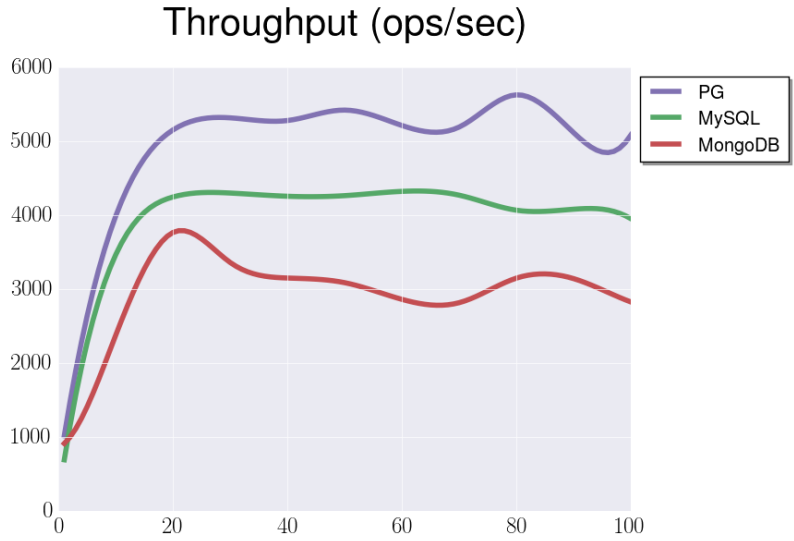
zalando


Average latency (μs)
700000
600000
500000
400000
300000
200000
100000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

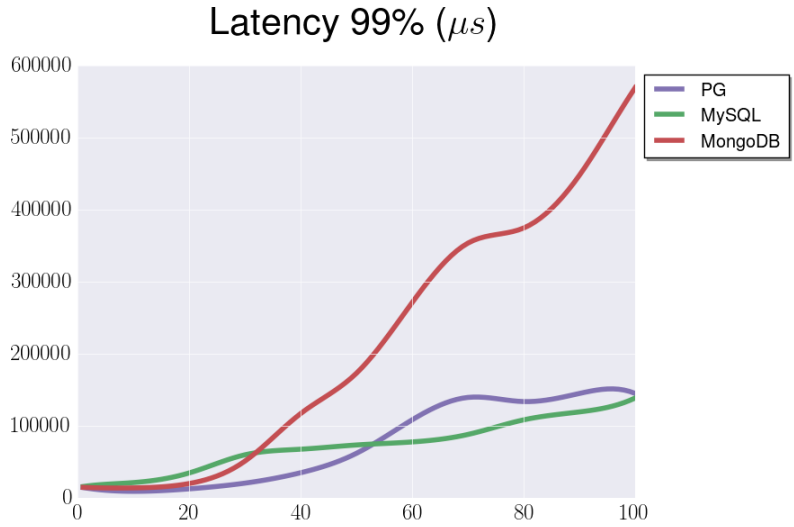
zalando

Выборка 50%, обновление 50%

”Простой документ”
обновление одного поля
transaction_sync


Throughput (ops/sec)
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

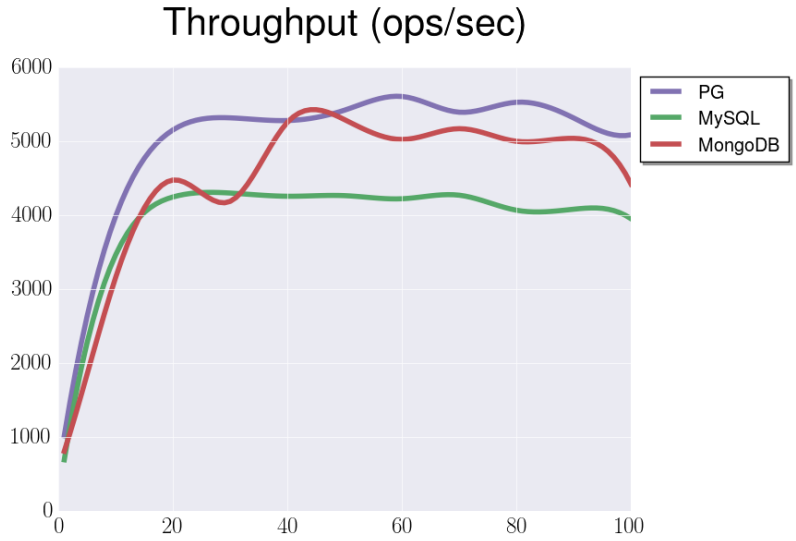
zalando


Latency 99% (μs)
600000
500000
400000
300000
200000
100000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando

Выборка 50%, обновление 50%

”Простой документ”
обновление одного поля
journaled

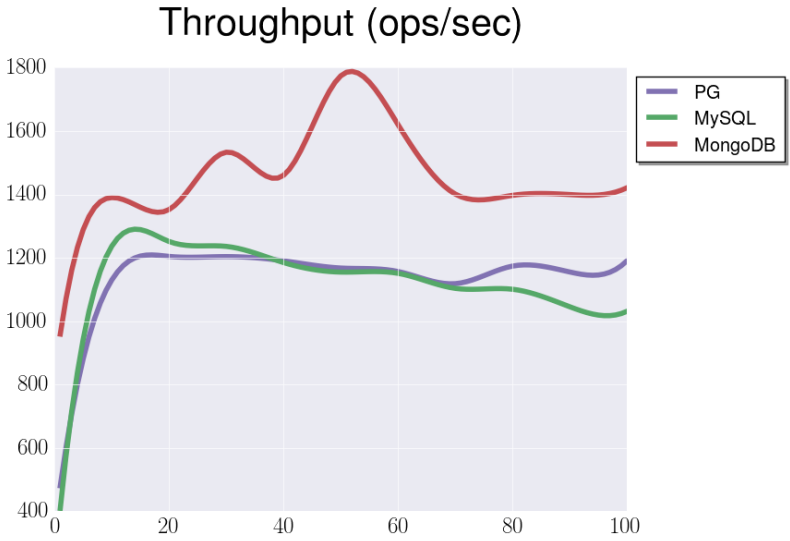

Throughput (ops/sec)
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando

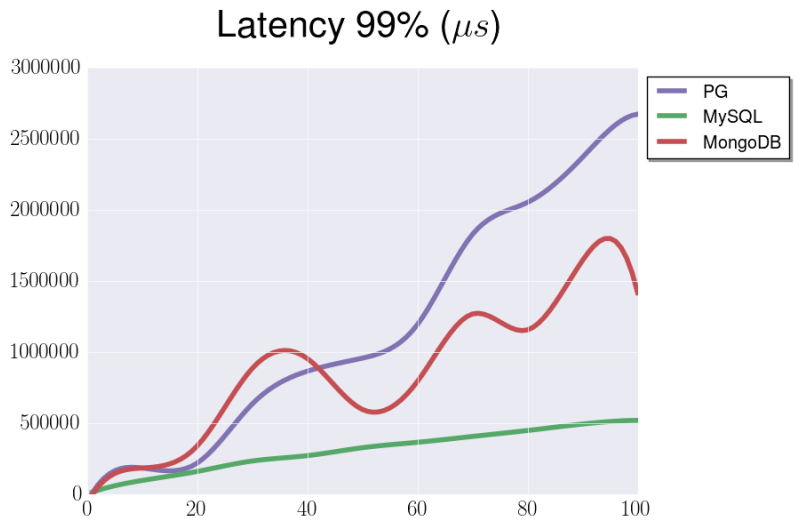
Выборка 50%, обновление 50%

”Большой документ”

обновление одного поля


Throughput (ops/sec)
1800
1600
1400
1200
1000
800
600
400
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando


Latency 99% (μs)
3000000
2500000
2000000
1500000
1000000
500000
0
0
20
40
60
80
100
PG
MySQL
MongoDB

zalando

➔ Документов в RDBMS можно не бояться

➔ Документов в RDBMS можно не бояться

➔ Приведенные бенчмарки - ”подсказки”

- ➔ Документов в RDBMS можно не бояться
- ➔ Приведенные бенчмарки - ”подсказки”
- ➔ Необходимы свои тесты для нагрузки

- ➔ Документов в RDBMS можно не бояться
- ➔ Приведенные бенчмарки - ”подсказки”
- ➔ Необходимы свои тесты для нагрузки
- ➔ Где кластер?

ВОПРОСЫ?

 github.com/erthalion
 @erthalion
 9erthalion6 at gmail dot com

zalando