

Что есть

Серверная часть нашего проекта собирает, сохраняет и показывает в реальном времени различную телеметрию с 20+ тысяч устройств.

Каждое устройство присылает несколько пакетов с телеметрией минимум раз в минуту. Есть примерно десяток разных видов данных: GPS-координаты, температуры, события, отчёты о работе.

Данные в основном просто сохраняются в базу данных чтобы потом отобразиться в виде графиков, списков и GPS-треков в личном кабинете. Никакой аналитики на лету не считается. Данные для просмотра в личном кабинете загружаются всегда в ограниченном диапазоне времени ([mintime, maxtime]), без сложных условий выборки.

Технические детали

Используется база данных MongoDB 3.4. Replica Set из двух рабочих серверов, одного арбитра и одного Delayed Secondary, пока без шардинга.

Статистика базы данных:

```
rs0:PRIMARY> db.stats(1024*1024)
{
  "db" : "autogps",
  "collections" : 71,
  "views" : 0,
  "objects" : 132356416,
  "avgObjSize" : 15552.082602070457,
  "dataSize" : 1963060.2975330353,
  "storageSize" : 445624.34375,
  "numExtents" : 0,
  "indexes" : 117,
  "indexSize" : 7353.0234375,
  "ok" : 1
}
```

Вся серверная часть работает на тех же двух серверах, которые являются рабочими нодами MongoDB. Каждый из этих серверов имеет следующую конфигурацию: i7-6700 / 32Gb / 3×500Gb SSD. До трети RAM может быть занято процессами, обслуживающими бизнес-логику.

Задача

Провести аудит эффективности использования базы данных:

- Эффективно ли мы используем базу данных с точки зрения производительности работы и объёма, занимаемого на диске
- Какие меры могли бы повысить эффективность в ближайшей перспективе?
- Сформулировать стратегию роста: что делать если завтра количество собираемых данных возрастёт в 2, 5, 10 раз?
- Оценить целесообразность перехода на колоночную базу данных, например, ClickHouse

