Tarantool/Silverbox
efficient in-memory storage

Юрий Востриков, Mail.Ru

26 октября 2010, HighLoad++

Высоконагруженные плоские таблицы

Примеры

- sessions
- passwd

Характерные особенности

- select only by (*primary*) key
- no joins
- no complex wheres, no aggregates
- sizeof(data) $<$ 256B
- высокая доступность

Скорость разных типов памяти

Tape is dead, disk is tape, RAM is disk.

Задержка *произвольного* доступа

- RAM — 25 ns
- HDD — 8’000’000 ns

Пропускная способность *последовательного* доступа

- RAM — 6’400 MB/s
- HDD — 170 MB/s

The Answer

Tarantool — framework

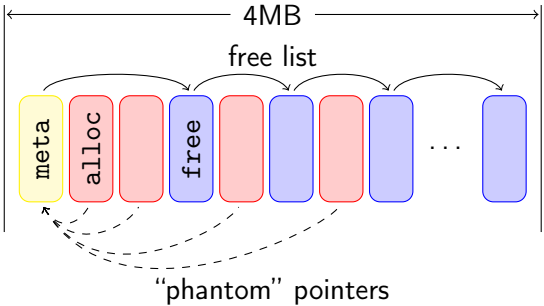
- In-memory storage
- Network IO
- Persistence facilities
- Hot Standby/Replication

SilverBox — Key-Value storage

- Tuple storage
- Native binary protocol
- Memcached protocol emulation

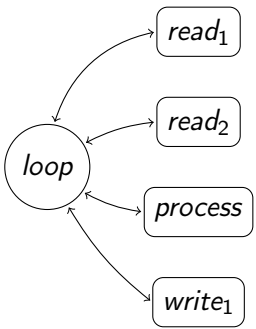
Устройство SLAB аллокатора

- Высокая скорость
- Рассчитан на хранение множества мелких объектов
- Низкие накладные расходы
- Отсутствие внешней фрагментации

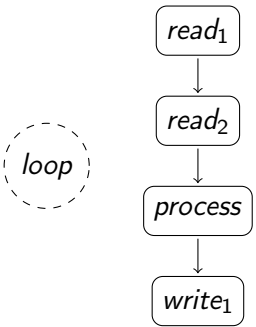


Массаракш: код наизнанку

FSM, libev only

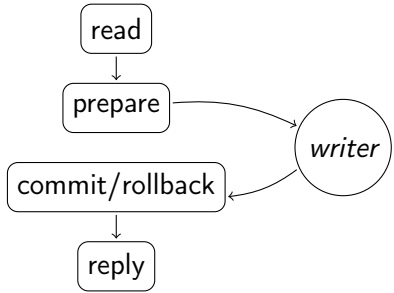


Fiber, libev+libcoro

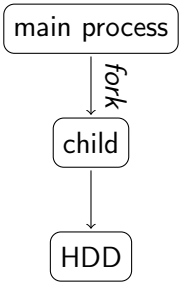


Persistance

WAL



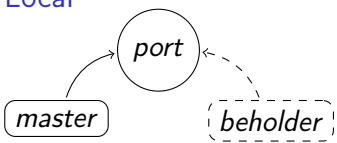
Snapshot



- Запись строго последовательная
- Copy on Write

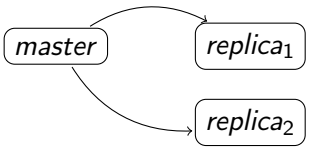
Hot Standby

Local



- Zero downtime апгрейд

Remote/Log Streaming



- Readonly реплики на удаленных серверах

IProto/Silverbox binary protocol

- async safe: каждый запрос маркирован sync числом
- tuple storage
  $< key_1,\ value_1,\ key_2,\ value_2,\ \ldots,\ value_n >$
- атомарные операции: $+/-$, *and*, *or*, *xor*
- namespaces
- индексы — u32 или string
- несколько индексов в одном namespace

Эмуляция memcached

- нет evictions
- expire — отложенный, с помощью background fiber

Производительность

- latency
- цена сискола
- пропускная способность дисков невелика
  $1\,\mathrm{KB} \times 100000\,\mathrm{rps} = 1\mathrm{MB} \times 100\mathrm{rps} = 100\,\mathrm{MB/sec}$
- u32 в два раза быстрее string key

Что дальше?

- новый формат WAL
- расширенная версия бинарного протокола
- Lua
- Эмуляция протокола redis

http://opensource.mail.ru/
http://github.com/mailru/
opensource@corp.mail.ru

Спасибо!

Вопросы?