taran tool

План

- Транзакционность
- Масштабирование

Транзакционность

- ACID
- Serializable (No-yield)
- Нет интерактивных (только 2.6+)

Транзакция

```
box.atomic(function()
  box.space.theanswers:insert(
    {"В чем ответ на вопрос о мироздании?", "42"})
  box.space.counters:update('questions', {{'+', 'value', 1}})
end)
```

- Транзакции работают для произвольного количества документов в рамках одного узла

Модульность

Приложение на LuaJIT

- Lua: простой скриптовый язык для инженеров
- Высокоэффективная JIT-компиляция
- Работа рядом с данными

Приложение на LuaJIT

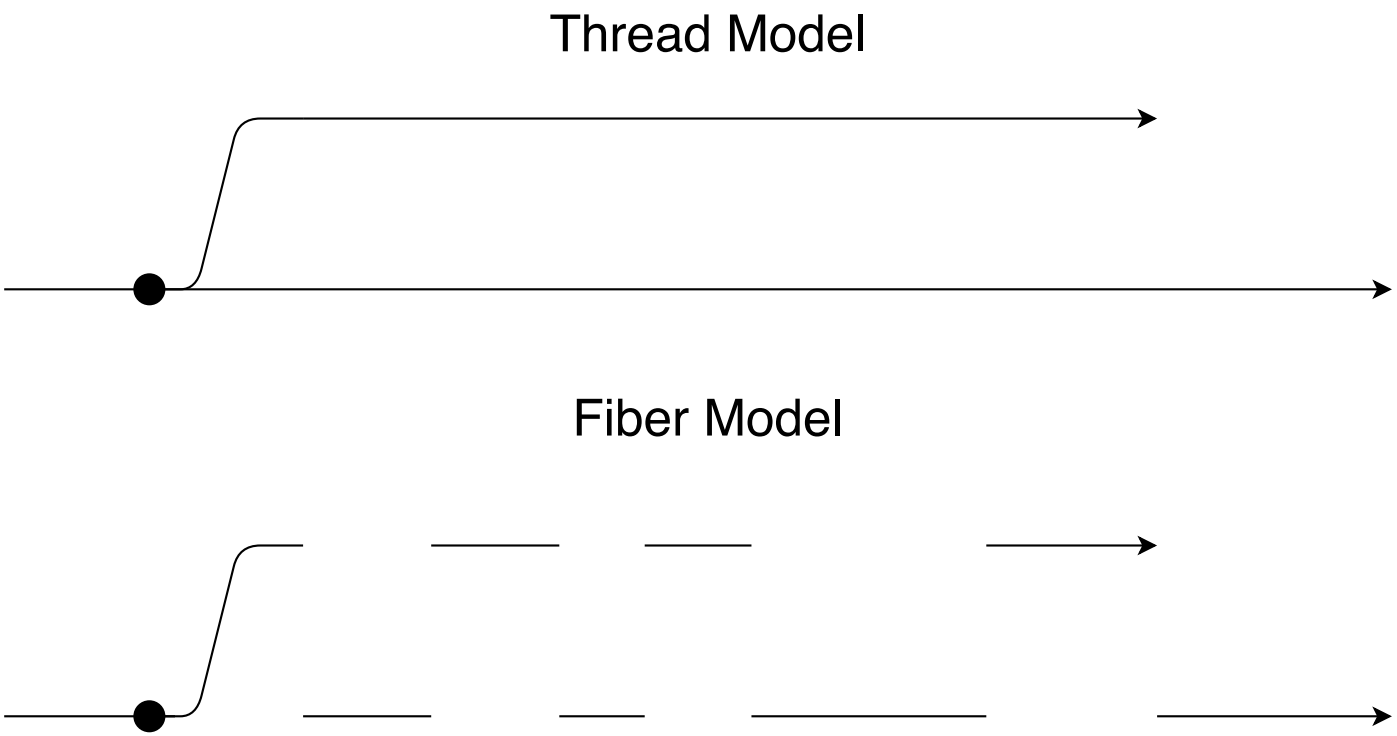
- Lua: простой скриптовый язык для инженеров
- Высокоэффективная JIT-компиляция
- Работа рядом с данными
- Не хранимые процедуры, а кооперативный рантайм!

Файберы и кооперативная многозадачность

- **Файбер (`fiber`)**
  Легковесная нить исполнения, реализующая кооперативную многозадачность

Файберы и кооперативная многозадачность

- **Файбер (`fiber`)**
  Легковесная нить исполнения, реализующая кооперативную многозадачность
- **Кооперативная многозадачность**
  Следующая задача выполняется после того, как текущая объявит о передаче управления


Thread Model
Fiber Model

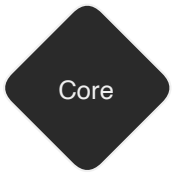
Сервер приложений

- Событийный цикл с файберами
- Неблокирующая работа с сокетами
- Коллекция библиотек для работы с сетью и данными
- Функции для работы с БД

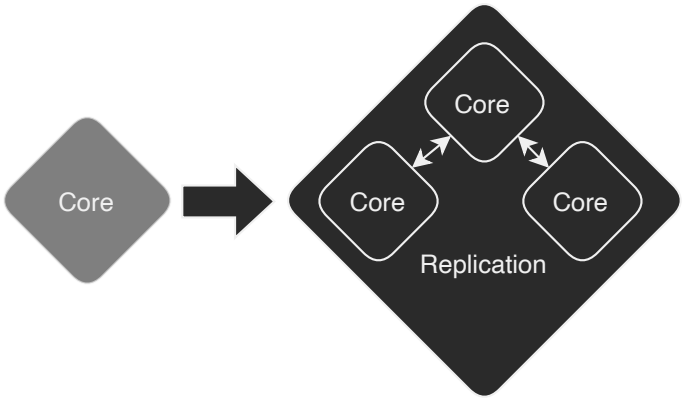
Встроенные модули

- clock
- http
- uri
- csv
- fun
- buffer
- digest
- crypto
- socket
- fio
- json
- yaml
- msgpack
- iconv
- utf8
- uuid

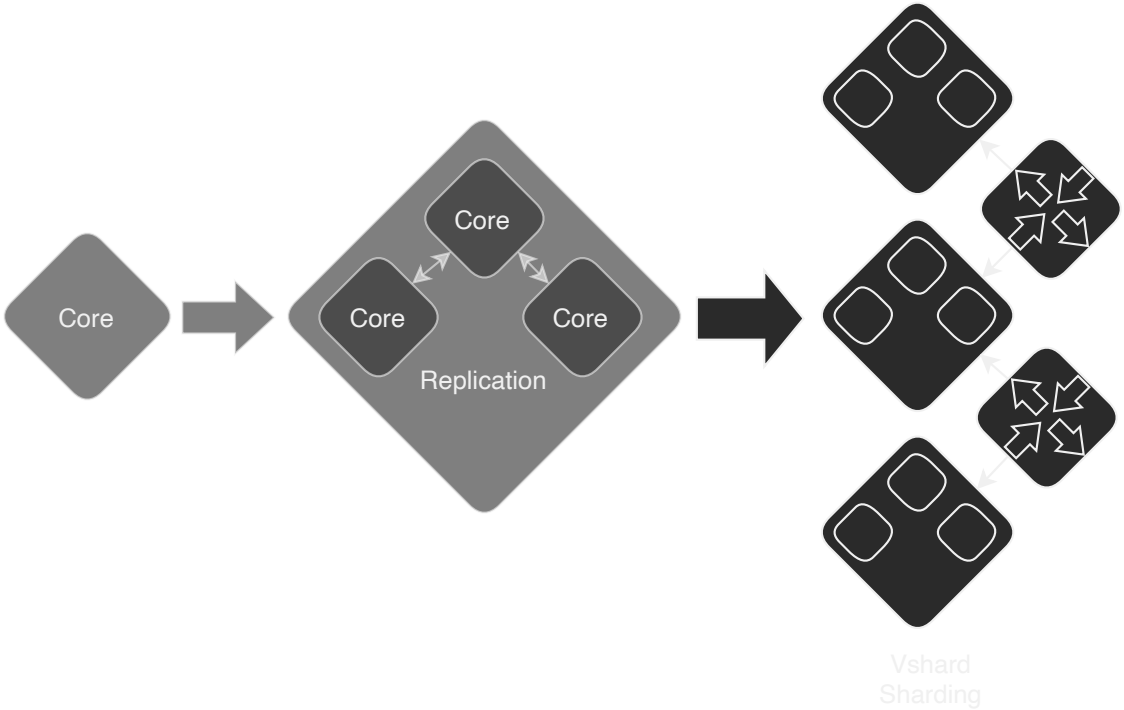
Масштабирование



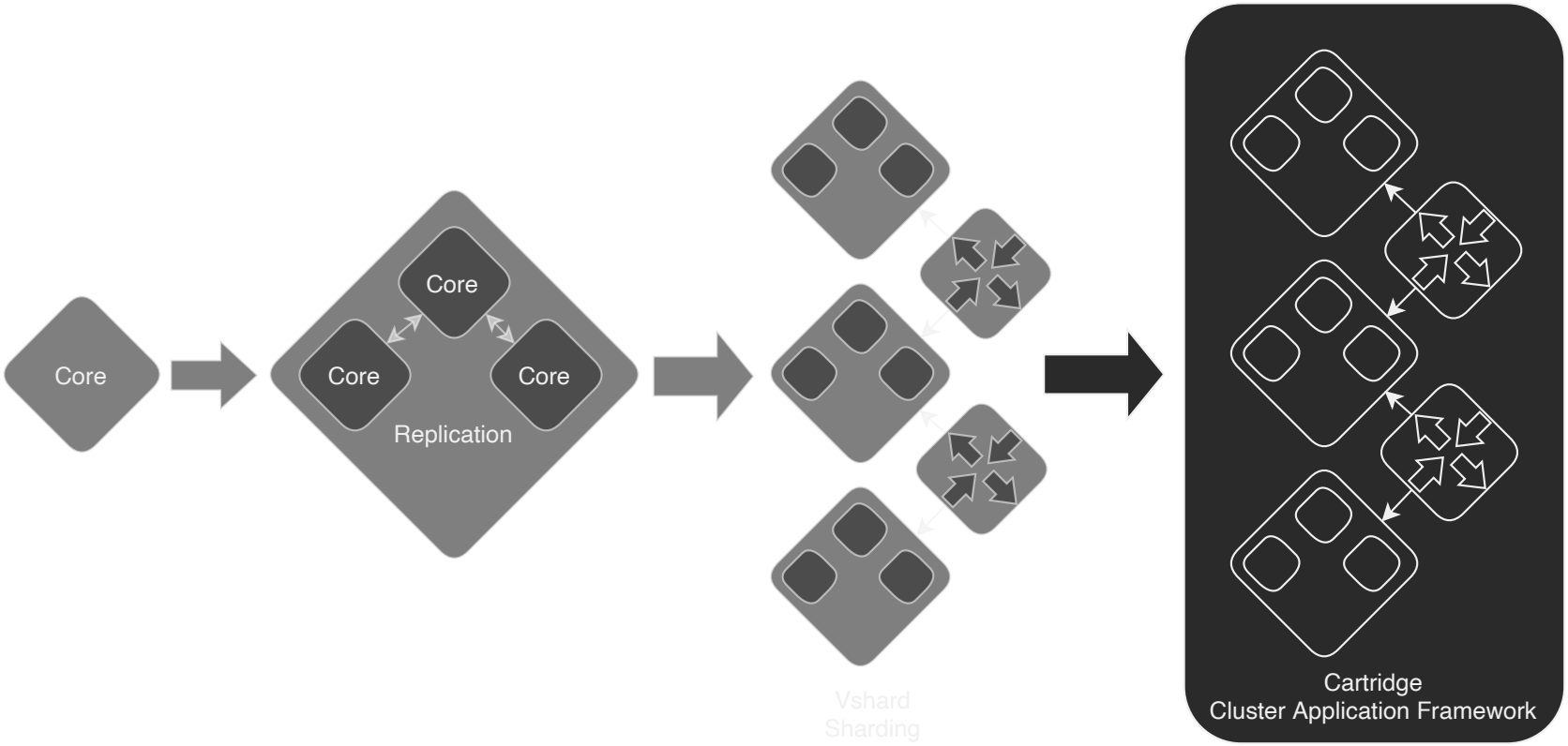
Масштабирование



Масштабирование



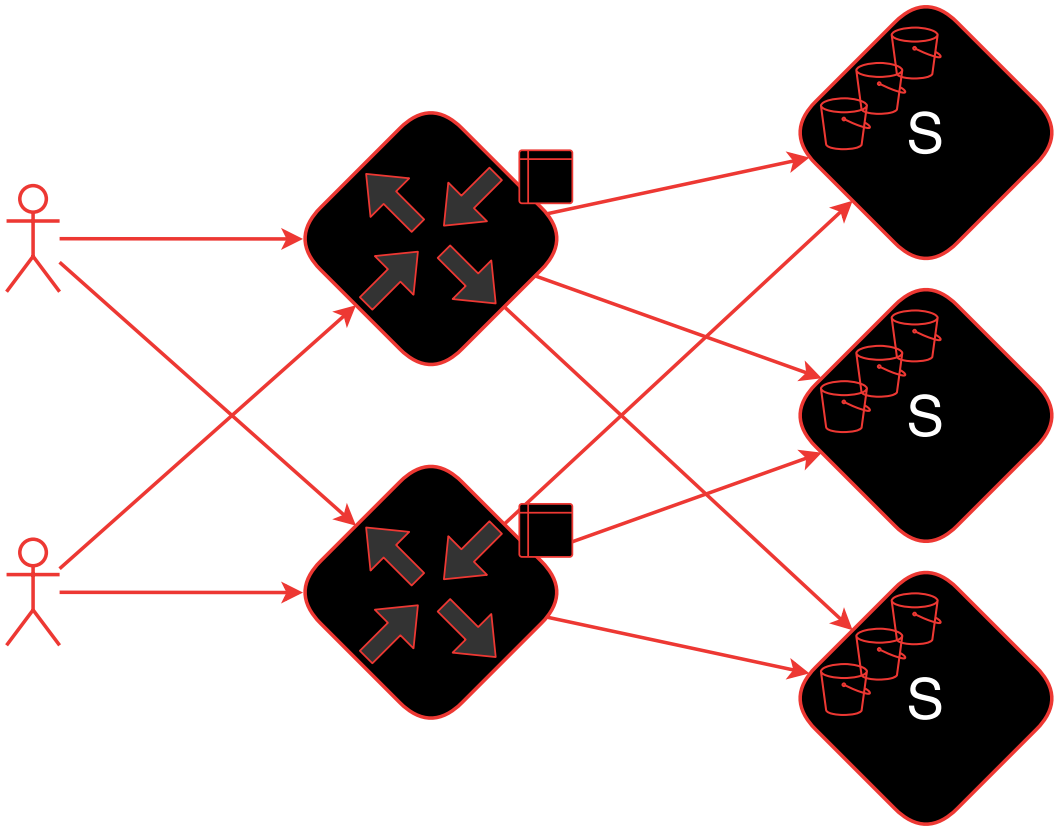
Масштабирование



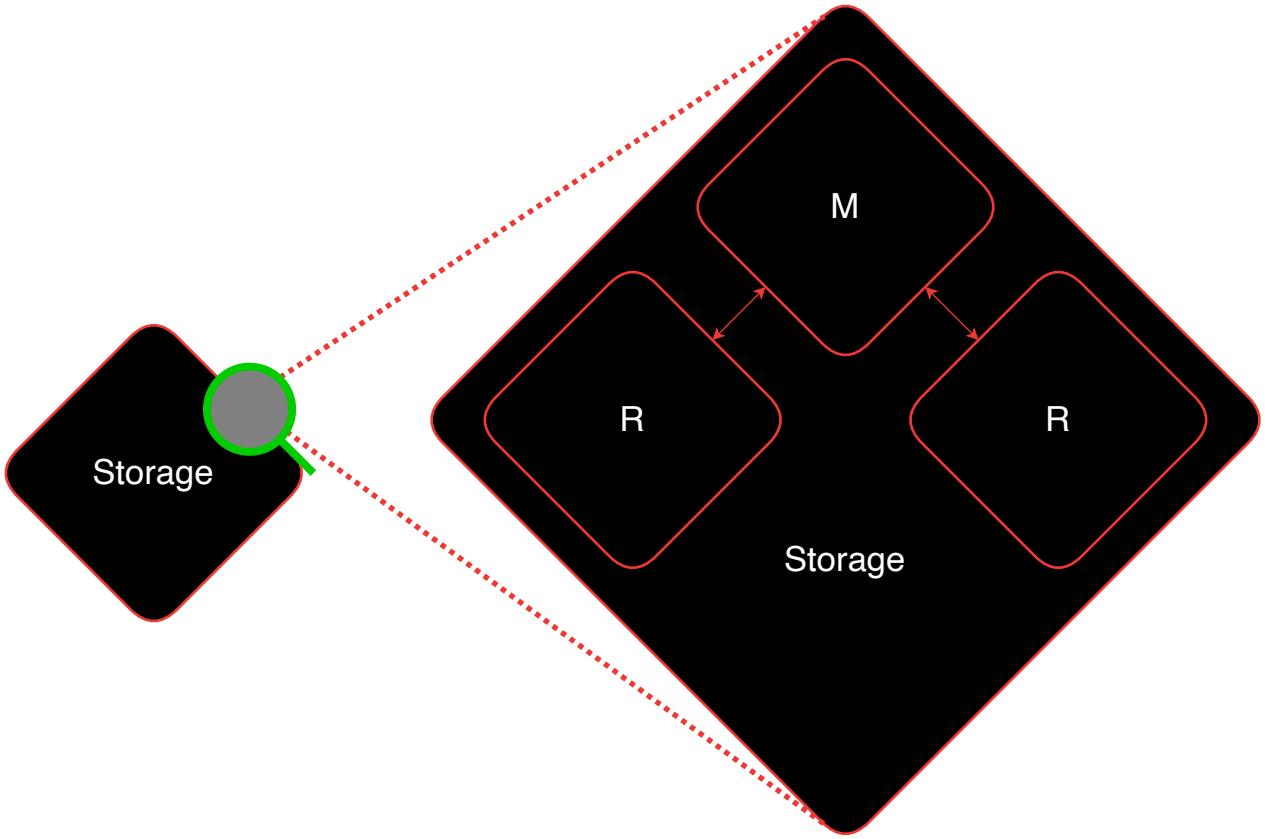
VShard

- Фреймворк для шардирования данных
- Реализует виртуальное шардирование бакетами
- Для каждого тапла задётся bucket_id
- Роутер значет на каких серверах, какие bucket_id хранятся

Схема шардированного кластера



VShard: Storage



VShard: Router

- Группа серверов: stateless*
- Знают карту `bucket → storage`
- Постоянно следят за изменениями карты
- Маршрутизируют запросы
- Линейно горизонтально масштабируются
- Выполняют bootstrap (первоначальную конфигурацию)

VShard: Storage

- Группа серверов: replicaset: физический шард
- Хранит подмножество бакетов
- Хранит пользовательские данные
- Не зависит от других шардов
- Избыточность при помощи репликации
- Доступ через роутер

```
box.space._buckets {
    { bucket_id_i, status = active },
    ...
    { bucket_id_j, status = active },
}
```

Cartridge

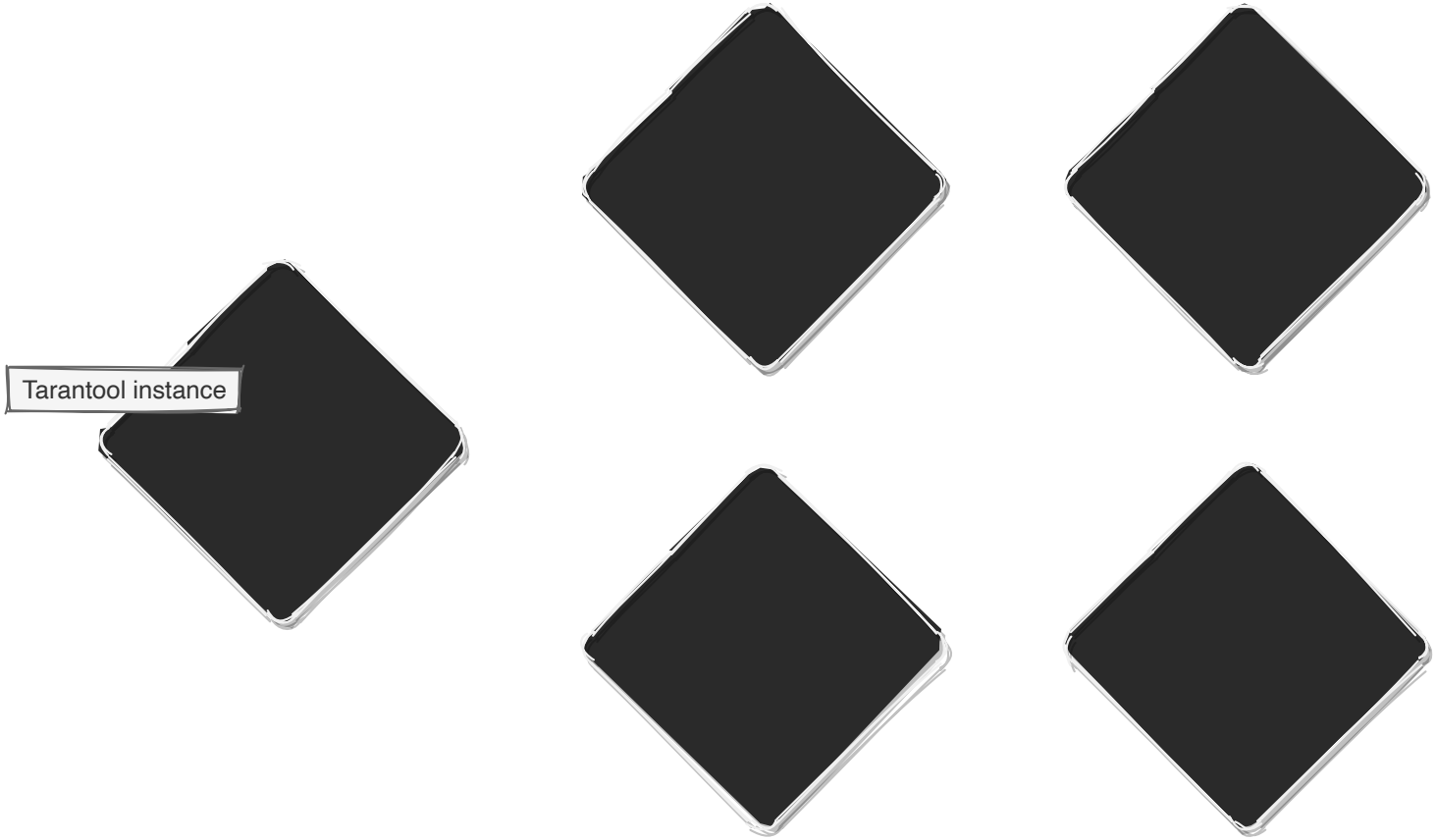
Фреймворк для построения кластерного приложения на Tarantool

Cartridge роли

Роль — Lua модуль реализовывающий бизнес API
Для повседневных задач уже реализованы:

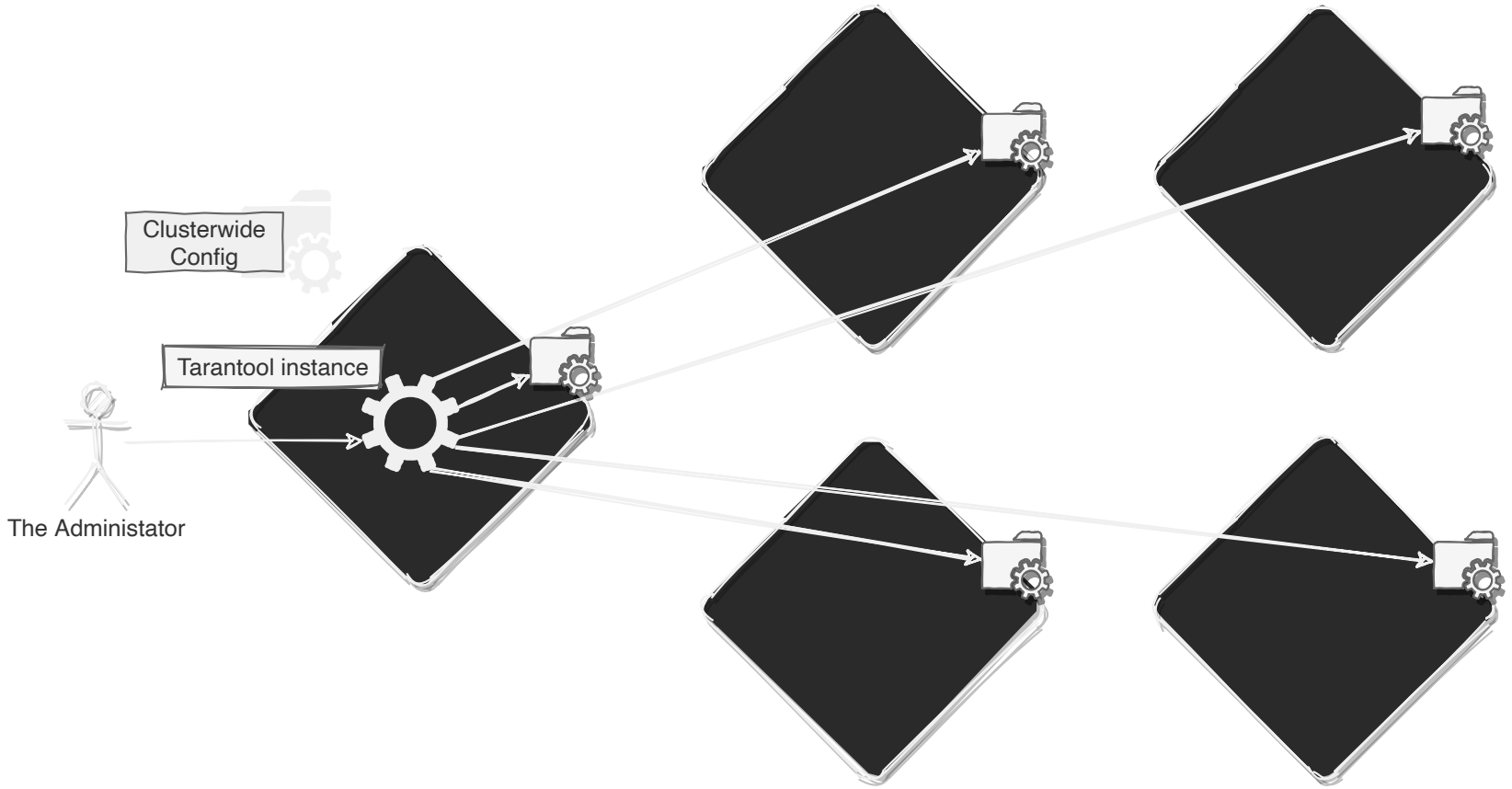
- crud — роли для выполнения кластерных запросов
- metrics — роль для сбора метрик работы приложения

Архитектура Cartridge — Запуск

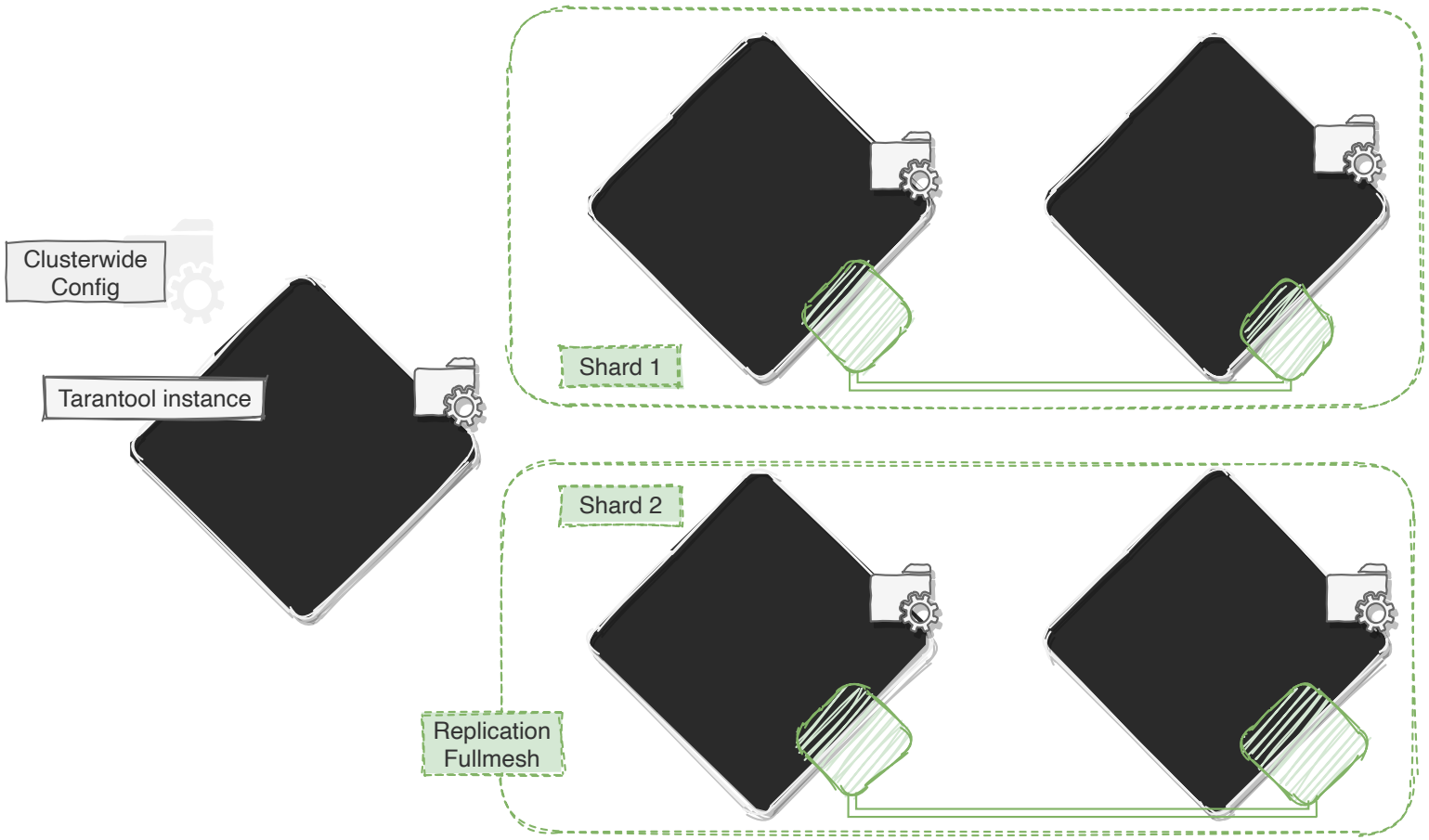


Обнаружение друг друга

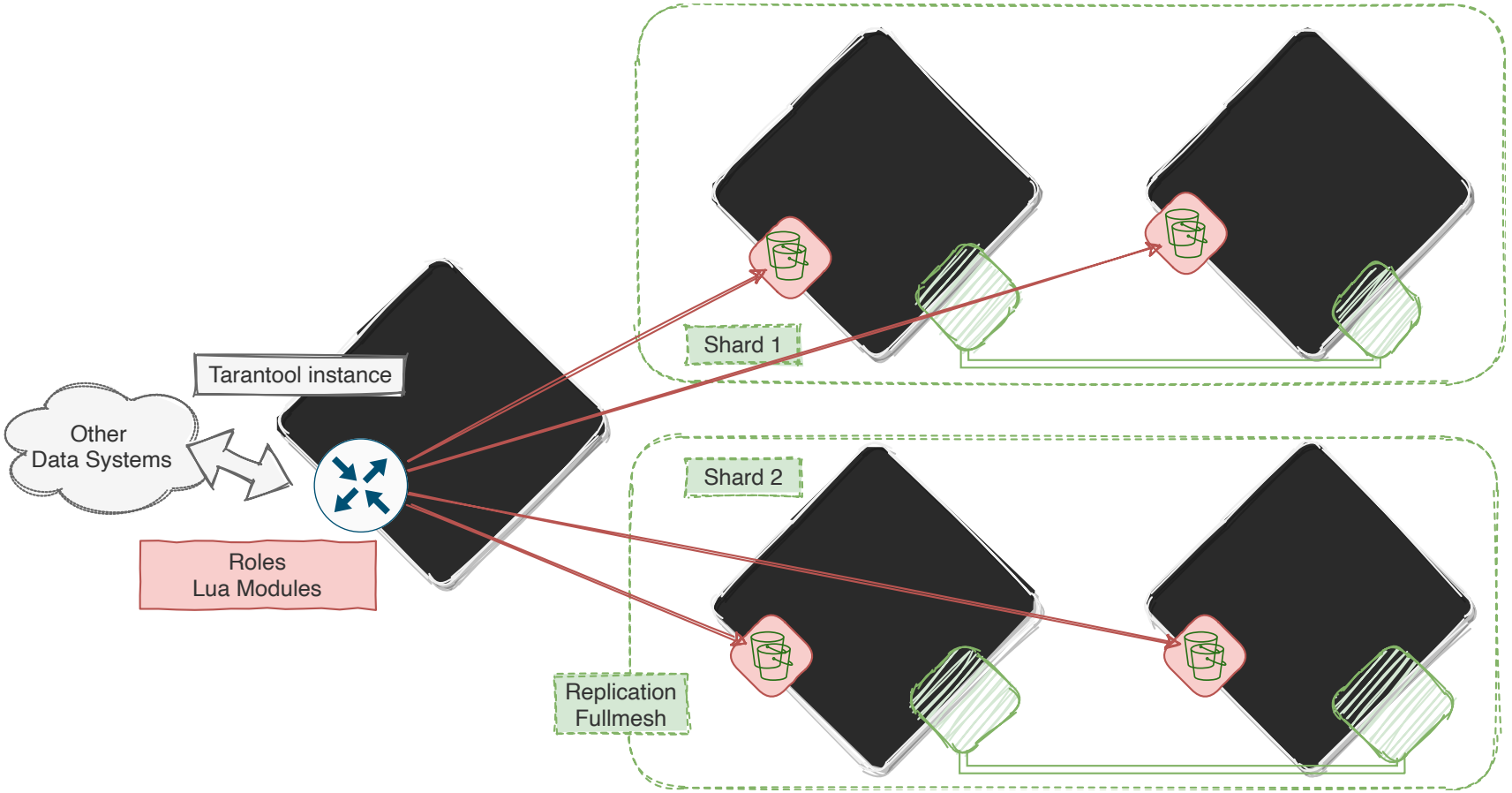
Конфигурация кластера



Топология


Clusterwide
Config
Tarantool instance
Shard 1
Shard 2
Replication
Fullmesh

Роли узлов


Tarantool instance
Other
Data Systems
Roles
Lua Modules
Shard 1
Shard 2
Replication
Fullmesh

Crud

- Модуль для cartridge для кластерных запросов

- **Кластерные** Create-Replace-Update-Delete операции
- Транзакции только в рамках одного узла!
- insert/insert_object
- get
- update
- delete
- replace/replace_object

Crud — insert_object

```
crud.insert_object("theanswers", {
    question="В чем ответ на вопрос о мироздании?",
    answer=42},
  {bucket_id=1})
```

Crud — delete

```
crud.insert_object("theanswers",
    {"В чем ответ на вопрос о мироздании?"},
  {bucket_id=1})
```

Crud — select

- Выборка данных из кластера
- Фильтрация по нескольким условиям (AND)
  - Range Scan
- Пагинация
- Балансировка

Crud — select

```
crud.select('theanswers', {
    {'>=', 'answer', 41},
    {'<', 'answer', 43}})
```

В заключение

Tarantool — инструмент для построения кластерных in-memory NoSQL приложений

- Обработка рядом с данными (LuaJIT)
- Однопоточный lock-free поток транзакций — файберы
- Индексы и итераторы
- Асинхронная репликация
- Шардирование бакетами
- UI управления кластером
- Ansible роль управления кластером

Ссылки

- Документация
- Design principles of Tarantool
- Engineering principles of Tarantool
- How Tarantool works with memory
- Tarantool data structures (internals)
- VShard — горизонтальное масштабирование в Tarantool
- Raft в Tarantool. Как это работает и как этим пользоваться
- Синхронная репликация в Tarantool
- db-engines.com — Tarantool System Properties