

Tarantool Clusters Federation

Описание технической
архитектуры программного
обеспечения

Оглавление

Термины и сокращения	3
1. Введение	4
1.1. Назначение системы	4
1.2. Цели создания системы	4
2. Функциональные характеристики системы	5
2.1. Перечень составных частей решения	5
2.1.1. Подсистема Replicator	6
2.1.2. Подсистема TCF Role	7
2.2. Способы и средства связи для информационного обмена между компонентами подсистем	7
2.3. Функции, выполняемые системой	7

Термины и сокращения

В настоящем документе используются сокращения, перечисленные в таблице ниже.

Сокращение	Расшифровка
БД	База данных
ОС	Операционная система
ПИКД	Подсистема идентификации и контроля доступа
ПО	Программное обеспечение
СУБД	Система управления базами данных
Lua	Процедурный динамически типизированный модульный язык с автоматическим управлением памятью

1. Введение

1.1. Назначение системы

Tarantool Clusters Federation (TCF) — продукт, позволяющий организовать катастрофоустойчивую конфигурацию из двух и более кластеров Tarantool Enterprise Edition, находящихся в разных дата-центрах. TCF обеспечивает скорость работы, приближенную к реальному времени, и отказоустойчивость до 99,99%.

1.2. Цели создания системы

TCF является механизмом переключения кластеров и синхронизации данных и используется для построения Mission Critical систем хранения данных, для которых необходимо выполнение следующих требований:

- гарантия постоянной доступности данных
- обеспечение отсутствия потерь данных
- минимизация времени отклика для Клиента

TCF используется для минимизации негативных последствий от возникновения нештатных ситуаций в работе распределенных кластеров и для обеспечения бесперебойного доступа к данным.

2. Функциональные характеристики системы

2.1. Перечень составных частей решения

В состав решения Tarantool Clusters Federation входят:

- Application – приложение клиента, которое работает с данными, хранящимися в кластере
- Tarantool Cartridge с ролью “TCF Role” – обеспечивает управление переключением статусов кластеров.
- Etcd (представлен на схеме логотипом) – узел etcd используемый для консистентного хранения состояния кластеров (Active/Passive)
- Balancer – программное или аппаратное решение, используемое для распределения (балансировки) всех запросов приложения по кластеру с целью оптимизации использования вычислительных ресурсов
- Router – узел Tarantool Enterprise Edition с установленной ролью Cartridge-router
- Master/Replica - узел Tarantool Enterprise Edition с установленной ролью Cartridge-storage используется для непосредственного хранения данных.
 - Master – обеспечивает операции чтения и записи данных.
 - Replica – обеспечивает только операции чтения данных
- Replicator – подсистема TCF обеспечивающая отправку данных в соседний кластер по протоколу gRPC

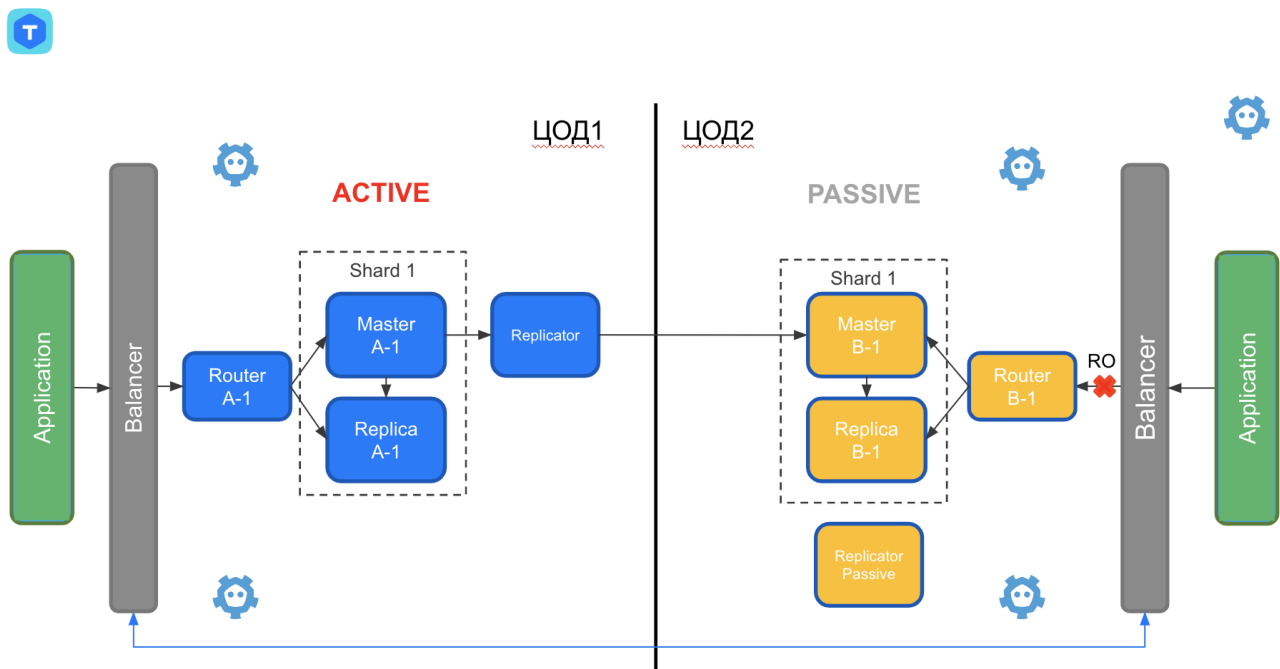


Рисунок 1 – Логическая схема

2.1.1. Подсистема Replicator

Replicator – подсистема, которая выполняет передачу данных между 2 кластерами, чтобы синхронизировать хранимые данные.

Состоит из 2-х частей:

- Gateway
- Applier

Gateway – подключается к шарду в кластере по протоколу репликации, в режиме анонимной реплики, для получения списка транзакций от кластера. После чего сериализует поток транзакций в gRPC.

Applier – подключается к Gateway по gRPC и запрашивает у него поток транзакций, после чего выполняет десериализацию этих транзакций и запись их в кластер по протоколу IProto.

2.1.2. Подсистема TCF Role

TCF Role – роль для Tarantool Cartridge, которая хранит состояние кластеров в etcd и позволяет управлять статусами кластеров.

Для управления используется HTTP rote.

2.2. Способы и средства связи для информационного обмена между компонентами подсистем

Информационный обмен между компонентами Системы производится в сетевом режиме передачи информации.

Взаимодействие между внутренними подсистемами осуществляется по протоколу Tarantool's binary protocol. Tarantool's binary protocol – бинарный протокол передачи данных, разработанный с учетом потребностей асинхронного ввода-вывода для облегчения интеграции с прокси-серверами. Каждый клиентский запрос начинается с бинарного заголовка переменной длины. В заголовке указывается идентификатор и тип запроса, идентификатор экземпляра, номер записи в журнале и т.д.

Также в заголовке обязательно указывается длина запроса, что облегчает обмен данными с клиентом или прокси. Ответ на запрос посылается по мере готовности. В заголовке ответа указывается тот же идентификатор и тип запроса, что и в изначальном запросе. По идентификатору легко соотнести запрос с ответом, даже если ответы приходят в произвольном порядке.

Взаимодействия с прочими компонентами осуществляется по протоколу HTTP.

2.3. Функции, выполняемые системой

- Передача данных из одного кластера в другой, с использованием механизма Change Data Capture
- Ручное переключение активного кластера
- Автоматическое переключение кластера