

Tarantool Column Store

Эксплуатация экземпляра продукта,
предоставленного для проведения
экспертной проверки

Версия 1.0.1 от 23.10.23

1. Введение

1.1. Область применения

Tarantool Column Store (далее по тексту TCS) представляет собой распределенное блочно-колоночное хранилище данных для решений, подразумевающих комплексные аналитические запросы с группировками и расчетом итоговых значений (агрегатов), в сфере финтеха, real-time маркетинга, логистики, телекома и др., реализованное на базе in-memory платформы Tarantool Enterprise.

Настоящее руководство предназначено для администраторов TCS в частной инсталляции.

Область применения настоящего документа распространяется на все подсистемы, модули и разделы TCS в частной инсталляции.

1.2. Краткое описание возможностей

TCS представляет собой распределенное блочно-колоночное хранилище данных осуществляющее хранение данных не в виде строк, а в виде колонок. Такое хранение позволяет оптимизировать хранение данных для выполнения комплексных аналитических запросов с группировками и расчетами итоговых значений. Оптимизация осуществляется за счет снижения нагрузки на устройства хранения и обработки данных. Снижение же в свою очередь достигается за счет того, что при выполнении запроса на чтение фактически с диска происходит чтение только тех полей, которые фигурируют в запросе, в то время как в реляционных базах данных (записи хранятся построчно) запись считывается с диска и обрабатывается целиком.

Дополнительно можно отметить, что поколоночное хранение данных существенным образом влияет на возможность компрессии данных ввиду того, что данные в одной колонке как правило являются однотипными, чего нельзя сказать о данных, хранящихся в строке. Кроме того, компрессия может достигаться за счет применения ограничений, накладываемых предметной областью на хранимые данные.

Запись данных в TCS может осуществляться как построчно, так и блоком данных, состоящим из заданного количества строк (по умолчанию - 1024 записи). При этом происходит

автоматическое разбиение записываемой строки на колонки согласно определенной схеме данных.

Получение данных из TCS может осуществляться при помощи SQL запросов, отправляемых в виде POST запросов в storage инстанс Tarantool, на базе которого построен TCS.

Часто используемые SQL запросы могут быть сохранены в виде PREPARE STATEMENT с использованием параметров, указываемых при последующем выполнении данного SQL запроса. Сохранение и последующий вызов сохраненного SQL запроса осуществляется при помощи отправки соответствующего POST запроса в storage инстанс Tarantool, который осуществляет непосредственное хранение данных.

1.3. Требования к квалификации

Системные администраторы TCS отвечают за комплексную настройку инфраструктуры TCS, предоставляемой в качестве услуги конечным потребителям, устранение неисправностей, сбор диагностической информации и эскалацию неисправностей производителю аппаратной или программной составляющей хранилища данных.

Системные администраторы выполняют в рамках работы с TCS следующие функциональные обязанности:

- Настройка и диагностирование системы;
- Обслуживание системного и прикладного программного обеспечения системы;
- Управление хранилищем данных (создание, удаление, настройка конфигурации, управление доступом к данным);
- Управление пользователями (создание, удаление, выдача прав);
- Резервное копирование и восстановление данных;
- Производство регламентных работ и анализ их результатов.

К системным администраторам предъявляются следующие требования:

- Навыки системного администрирования Linux;
- Прохождение обучения от компании ООО «ВК ЦИФРОВЫЕ ТЕХНОЛОГИИ» по использованию компонентов, разработанных вендором;

- Навыки работы со службами TCS, в том числе способность самостоятельно осуществлять:

- Мониторинг программными средствами;
- Изменение настроек TCS.

Для обеспечения функционирования системы в круглосуточном режиме (24/7) необходимо наличие как минимум двух системных администраторов.

2. Установка и настройка экземпляра продукта, предоставленного для проведения экспертной проверки

Процесс установки и настройки экземпляра TCS, передаваемого для проведения экспертной оценки описан в документе:

- Tarantool Column Store_v.1.0.1_Инструкция по установке экземпляра программного обеспечения

TCS передается для проведения экспертной проверки в виде комплекта, содержащего следующие файлы:

- tarantool_column_store-0.9.4.x86_64.tar.gz – архив с TCS
- config.json – конфигурационный файл (содержит колоночную схему данных)

Минимальные системные требования, предъявляемые к рабочей станции, на которой осуществляется экспертная проверка:

- ОС - Linux (Ubuntu, Debian, CentOS, ALTLinux, AstraLinux, RedOS)
- 4 Core 2.6ГГц +
- 4Гб RAM

2.1 Подготовка к проведению экспертной оценки

В случае корректной установки и настройки экземпляра TCS все экземпляры Tarantool должны иметь статус RUNNING

- Перейти в директорию с разархивированным приложением

```
cd tarantool_column_store
```

- Проверить запуска экземпляров. Все экземпляры должны иметь статус RUNNING

```
./tt status
```

- В случае если экземпляры не запущены запустить их командой

```
./tt start
```

2.2 Добавление записей в хранилище

Данные в хранилище могут быть добавлены:

- По одной строке

- Подключиться к storage инстансу Tarantool

```
./tt connect tarantool_column_store:s1-1
```

- Проверить количество записей в строковом буфере до добавления

```
box.space.records:len()
```

- Добавить запись в TCS

```
tcs.storage.put_record{Attribute1 = 1}
```

- Проверить количество записей в строковом буфере после добавления.
Количество записей должно увеличиться

```
box.space.records:len()
```

- Выполнить команду

```
Ctrl + D
```

- Блоком строк (размер блока 1024 записи)

2.3 Получение записей из хранилища

Получение данных из TCS возможно с помощью SQL-запросов, отправляемых в виде POST запроса в storage инстанс Tarantool

```
curl localhost:7777/sql --data '
select "Attribute1", "Attribute2", "Attribute3", "Attribute4"
from attributes
limit 1000'
```

В TCS возможно сохранение SQL-запросов для последующего выполнения (PREPARE STATEMENT), в том числе возможно использование параметров, которые будут указаны при

последующем выполнении данного запроса. Сохранение SQL-запросов происходит при отправке соответствующего POST запроса в storage инстанс Tarantool

```
curl localhost:7777/sql --data '
PREPARE plan1(INT) AS
select "Attribute1", "Attribute2", "Attribute3", "Attribute4"
from attributes
WHERE "Attribute1" < $1
limit 1000'
```

После сохранения SQL-запроса он может быть вызван с указанием параметров, которые будут переданы в запрос

```
curl localhost:7777/exec/plan1?args=10
```

2.4 Завершение работы TCS

Для завершения работы TCS необходимо выполнить следующие операции

- Перейти в директорию с приложением

```
cd tarantool_column_store
```

- Завершить работы TCS

```
./tt stop
```

2.5 Поддержка

В случае возникновения проблем при эксплуатации TCS в рамках экспертной проверки вопросы могут быть направлены техническим специалистам:

- Алексей Андреев (al.andreev@corp.mail.ru)