

# Tarantool Column Store

Руководство пользователя

Версия 1.0.0 от 23.08.23

## ● Термины и сокращения

В настоящем документе используются сокращения, перечисленные в таблице ниже.

Сокращение	Расшифровка
БД	База данных
ВМ	Виртуальная машина
ОС	Операционная система
ПО	Программное обеспечение
СУБД	Система управления базами данных
API	Application Programming Interface - Аппаратно-программный интерфейс. Описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой
JavaScript	Мультипарадигменный язык программирования. Поддерживает объектно-ориентированный, императивный и функциональный стили
JSON	JavaScript Object Notation – текстовый формат обмена данными, основанный на JavaScript
Доменная сущность	Объект, обладающий определённым смыслом (например, Лицо) и описываемый набором атрибутов
Атрибут	Именованная количественная или качественная характеристика (напр. “Адрес IP”)
Операция	Единичное действие, связанное с одной или несколькими Доменными Сущностями, описанное набором атрибутов
Матрица	Контейнер исторических данных об Операциях с колоночным видом хранения и представления данных
Индексная колонка	Колонка из матрицы, для которой строится индекс (например, ip-адрес или страна)
Счетчик	Рассчитываемое значение агрегата одной колонки в зависимости от условий
PreWhere	Условие фильтрации, которое использует индексную колонку
Batch	Список счетчиков, которые могут вычисляться параллельно при чтении с одним и тем же условием PreWhere. Счетчики в одном батче могут иметь разные условия срабатывания и глубину поиска
Глубина поиска	Количество записей, которые СУБД должна прочитать, чтобы сформировать расчет одного счетчика

Сокращение	Расшифровка
Материализованный счетчик	Рассчитанное историческое значение, сохраняемое в основном массиве данных с возможностью последующего рекуррентного использования для дальнейших расчётов
Узел, Инстанс, Нода	Единичный экземпляр (процесс) запущенного ПО

## ● 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 пользователей TCS в частной инсталляции.

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

### 1.2. Краткое описание возможностей

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.

### 1.3. Требования к квалификации пользователя

К пользователям TCS предъявляются следующие требования:

- Прохождение обучения от компании ООО «ВК ЦИФРОВЫЕ ТЕХНОЛОГИИ» по использованию компонентов, разработанных вендором;
- Навыки работы с ПК, в том числе способность самостоятельно осуществлять выполнение операций с файлами средствами ОС:
  - Создание;
  - Копирование;
  - Переименование;
  - Удаление;

- Перемещение и т.п.

## 1.4. Перечень эксплуатационной документации

Перечень эксплуатационной документации, с которой необходимо ознакомиться:

- Инструкция по установке экземпляра программного обеспечения TCS;
- Описание функциональных характеристик TCS;
- Руководство по установке TCS;
- Описание процессов поддержания жизненного цикла TCS;
- Документация к Tarantool на сайте [tarantool.io](https://tarantool.io).

## ● 2. Описание операций

### ○ 2.1. Описание всех выполняемых функций, задач, комплексов задач, процедур

Система выполняет следующие функции:

- Создание и управление блочно-колоночным хранилищем данных;
- Поддержка различных типов данных;
- Хранение исторических данных о выполненных операциях хранилища;
- Поддержка конфигурирования схемы данных:
  - Таблицы;
  - Атрибуты таблиц;
  - Типы данных;
  - Глубина хранения;
  - Индексы;
  - Агрегаты.
- Выполнение аналитических запросов к данным с группировками и расчетом итоговых значений (агрегатов).
- Обработка запросов на запись данных в хранилище;
- Сохранение единичной транзакции в хранилище;
- Выполнение вычислений агрегатов согласно условиям.
- Обработка аналитических запросов на чтение данных из хранилища;
- Обеспечение сжатия данных хранилища TCS.
- Поддержка импорта данных в хранилище TCS из следующих форматов:
  - CSV;
  - Parquet.

- Предоставление доступа к данным хранилища TCS по следующим протоколам:
  - HTTP API для запросов на чтение с SQL;
  - Arrow Flight RPC.

Так как система реализована на базе платформы Tarantool Enterprise, то она имеет те же ограничения:

- Однопоточность обработки данных;
- Доступность аппаратных ресурсов (RAM, CPU).

## 2.2. Описание операций технологического процесса обработки данных, необходимых для выполнения функций, комплексов задач (задач), процедур

Наименование операции	Предусловия	Подготовительные действия	Основные действия в требуемой последовательности	Результат
Создание и управление блочно-колоночным хранилищем			Данная операция выполняется администратором системы	
Сохранение единичной транзакции в хранилище	Создано хранилище	Установка соединения	Передать транзакцию одним из способов: HTTP API / Arrow Flight RPC	Получить подтверждение или ошибку
Выполнение импорта данных в хранилище	Создано хранилище	Установка соединения	Выполнить батчевую загрузку данных в хранилище из файлов формата parquet / csv	Получить подтверждение или ошибку
Выполнение вычислений агрегатов согласно условиям	Данные сохранены в хранилище		Вычислить агрегат по атрибутам данных хранилища	Обновить счетчики (агрегаты)
Выполнение сжатия данных	Данные сохранены в хранилище		Применить алгоритм сжатия для данных хранилища	Преобразовать данные по алгоритму

Выполнение индексирования данных	Данные сохранены в хранилище		Рассчитать значения для указанных индексных колонок в хранилище	Построить индексы для данных хранилища
Обработка аналитических запросов на чтение данных	Данные сохранены, сжаты, проиндексированы	Получен запрос на чтение данных	Интерпретировать полученный SQL-запрос. Выполнить поиск данных в хранилище.	Получить данные из хранилища
Настройка конфигурации хранилища	Создано хранилище		Данная операция выполняется администратором системы	
Управление пользователями и доступами	Создано хранилище		Данная операция выполняется администратором системы	

### 3. Аварийные ситуации

В случае возникновения аварийной ситуации пользователю необходимо обратиться к администратору системы и выполнить следующие действия:

- Передать информацию о том, что было сделано, прежде чем проблема появилась;
- Сформулировать и описать, в чем именно заключается проблема.