



cedrusdata

## Инструкция по установке CedrusData

ООО «Кверифай Лабс»

ОГРН 1217800163790

ИНН 7811766769

КПП 781101001

|  |          |
|--|----------|
| <b>Введение</b>                          | <b>3</b> |
| <b>1. Описание архива</b>                | <b>3</b> |
| <b>2. Предварительные действия</b>       | <b>4</b> |
| <b>3. Установка</b>                      | <b>4</b> |
| <b>4. Подключение через JDBC драйвер</b> | <b>5</b> |

# Введение

CedrusData это высокопроизводительная распределенная платформа для сквозного анализа всех данных предприятия в облаке и on-premise через единую точку доступа с SQL интерфейсом.

CedrusData основана на распределенном SQL-движке Trino (<https://trino.io/>) и включает дополнительный функционал управления и мониторинга (в том числе в облачных инфраструктурах), улучшения производительности, профессиональную документацию и поддержку.

CedrusData может быть установлена из архива или из Docker-образа. Краткая инструкция по установке доступна в документации:

- Установка из архива: <https://docs.cedrusdata.ru/latest/installation/archive.html>.
- Установка из Docker-образа: <https://docs.cedrusdata.ru/latest/installation/docker.html>.

Данный документ представляет собой руководство по установке CedrusData из архива.

CedrusData поддерживает операционные системы семейства Linux. Инструкции из данного руководства были проверены на операционных системах Astra Linux, ALT Linux, Debian, Ubuntu и Fedora.

## 1. Описание архива

Архив CedrusData представляет собой файл с расширением `tar.gz` и доступен по адресу <https://downloads.cedrusdata.ru/releases/cedrus-402-1.tar.gz>.

### 1.1. Структура директорий

Архив содержит в себе следующие файлы и директории:

- `etc/` - файлы конфигурации CedrusData.
- `etc/catalog/` - файлы конфигурации каталогов CedrusData.
- `bin/` - скрипты для запуска CedrusData и интерфейса командной строки `trino`.
- `jdbc/` - JDBC-драйвер Trino.
- `lib/` - бинарные файлы CedrusData.
- `plugin/` - бинарные файлы плагинов CedrusData.

### 1.2. Доступные каталоги

Архив CedrusData содержит конфигурацию по умолчанию, в которой присутствуют следующие каталоги:

- `tpch` - каталог TPC-H коннектора со сгенерированными TPC-H данными (см. <https://docs.cedrusdata.ru/latest/connector/tpch.html>).
- `tpcds` - каталог TPCDS коннектора со сгенерированными TPC-DS данными (см. <https://docs.cedrusdata.ru/latest/connector/tpcds.html>).
- `memory` - каталог Memory коннектора для хранения данных в памяти (см. <https://docs.cedrusdata.ru/latest/connector/memory.html>).

- `jmx` - каталог JMX коннектора для доступа к MBeans (см. <https://docs.cedrusdata.ru/latest/connector/jmx.html>).

## 2. Предварительные действия

Перед установкой CedrusData необходимо проделать следующие действия.

1. Убедитесь, что вы используете операционную систему семейства Linux.
2. Убедитесь, что у вас установлен Python версии 2.x или 3.x с помощью команды `python --version`. Если программа не установлена, воспользуйтесь инструкциями по установке с официального сайта <https://www.python.org/>.
3. Убедитесь, что у вас установлена JDK 17 с помощью команды `java --version`. Если программа не установлена, установите любой дистрибутив JDK 17. Например, Eclipse Temurin <https://adoptium.net/temurin/releases/>.
4. Убедитесь, что у вас установлены утилиты `wget` и `gzip`. Они не требуются для функционирования CedrusData, но будут использованы в данной инструкции.

## 3. Установка

### 3.1. Инструкции

1. Скачайте архив:

```
wget https://downloads.cedrusdata.ru/releases/cedrus-402-1.tar.gz
```

2. Распакуйте архив:

```
tar -xf cedrus-402-1.tar.gz
```

3. Перейдите в директорию распакованного архива:

```
cd cedrus-402-1
```

4. Запустите узел CedrusData:

```
bin/launcher start
```

На экране будет напечатан PID процесса CedrusData. Запуск узла занимает от нескольких секунд до минуты. В случае успешного запуска узла, вы увидите следующее сообщение в логе `data/cedrus/var/log/server.log`:

```
io.trino.server.Server ===== SERVER STARTED =====
```

5. Запустите SQL-запрос к данным CedrusData с помощью интерфейса командной строки:

```
bin/trino --execute "select count(*) from tpch.sf1.orders"
```

В случае успешного выполнения SQL-запроса, на экране будет напечатана следующая строка:

```
"1500000"
```

6. Остановите узел CedrusData:

```
bin/launcher stop
```

На экране будет напечатан PID остановленного процесса CedrusData.

### 3.1. Устранение неисправностей

При выполнении команды `bin/launcher start` вы можете увидеть в логе `data/cedrus/var/log/server.log` следующую ошибку:

```
java.io.UncheckedIOException: Failed to bind to /0.0.0.0:8080
```

Данная ошибка возникает, так как узел CedrusData по умолчанию требует доступ к порту `8080`, который оказался занят другим процессом. Для решения проблемы вы можете либо освободить порт `8080`, либо использовать другой порт, изменив его в двух местах в файле конфигурации `etc/config.properties`:

```
http-server.http.port=<укажите другой порт>  
discovery.uri=http://localhost:<укажите другой порт>
```

## 4. Подключение через JDBC драйвер

При промышленном использовании CedrusData запуск SQL-запросов наиболее часто происходит через JDBC драйвер. В данном разделе рассмотрено подключение к CedrusData через JDBC драйвер с использованием популярной утилиты `DBeaver`.

1. Запустите узел CedrusData:

```
bin/launcher start
```

2. Установите утилиту `DBeaver` с официального сайта <https://dbeaver.io/download/> и запустите ее.

3. Выберите `File -> New -> DBeaver -> Database Connection` и нажмите кнопку `Next`.

4. Выберите продукт `Trino` и нажмите кнопку `Next`.

5. Заполните следующие поля, после чего нажмите кнопку `Finish`:

- В поле `Host` укажите IP адрес или сетевое имя компьютера, на котором запущен узел. Если `DBeaver` запущен на том же компьютере, что и `CedrusData`, можно указать адрес `localhost`.
- В поле `Port` укажите значение `8080`. Если вы изменили порт `CedrusData` согласно инструкции из пункта 3.1, то укажите актуальное значение порта.
- В поле `Username` укажите любое непустое значение. Например, `admin`.

6. Нажмите на появившееся слева соединение правой кнопкой мыши, выберите `SQL Editor -> New SQL Script`. В открывшемся окне напишите и запустите следующий SQL-запрос:

```
select count(*) from tpch.sf1.orders
```

В случае успешного выполнения SQL-запроса, на экране появится таблица с одной колонкой и одной записью:

```
col0  
-----  
1,500,000
```

При нажатии на соединение правой кнопкой мыши, программа может запросить у вас установку JDBC драйвера Trino. В этом случае установите JDBC драйвер Trino, нажав кнопку [Download](#) в появившемся окне.