

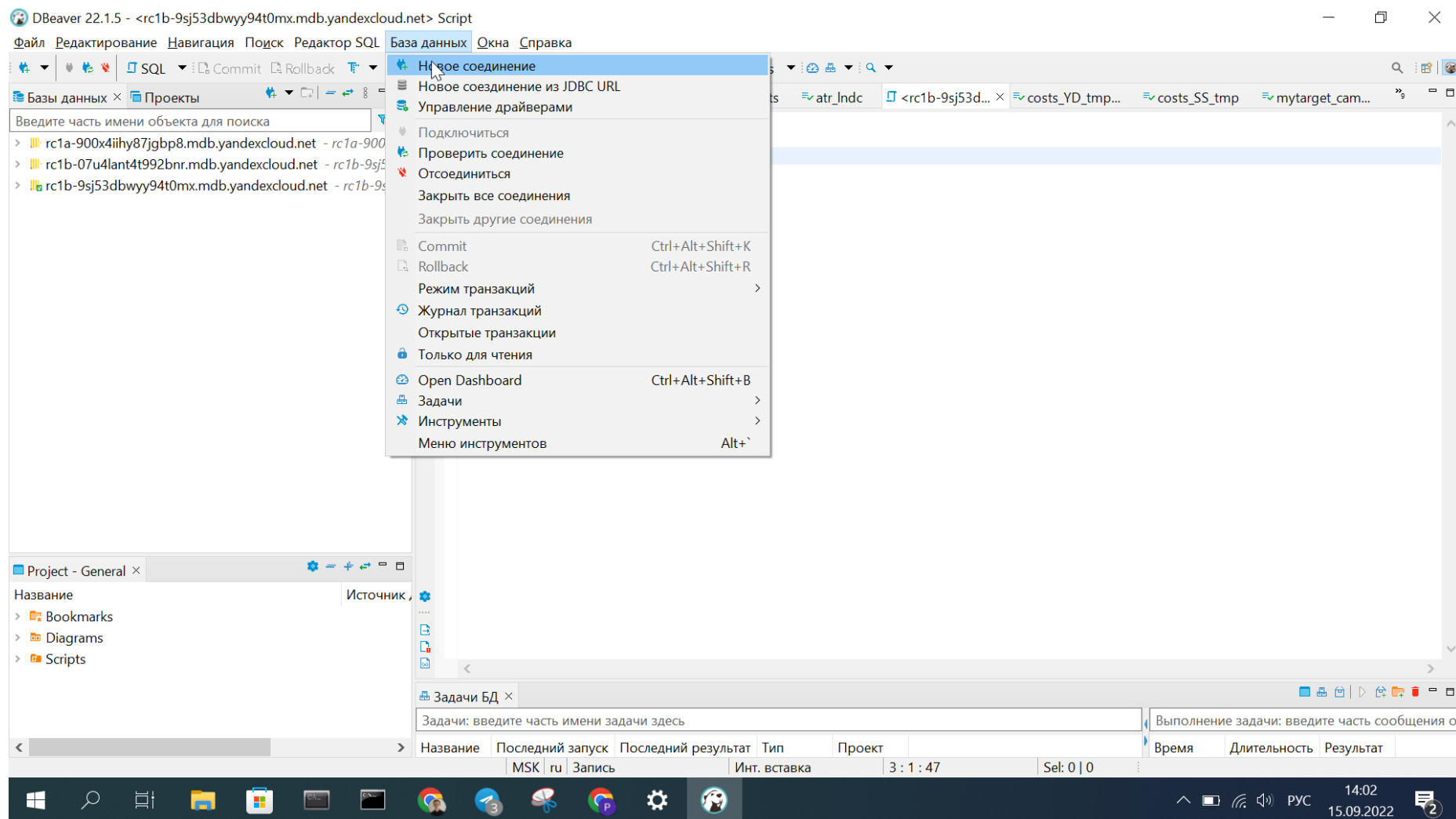
Инструкция по работе с данными в Clickhouse с помощью Python и DBeaver

Введение:

Настоящая инструкция содержит описание последовательности действий при работе с данными в СУБД ClickHouse (далее – CH) с использованием Python-скрипта (в среде Google Colab) и дальнейшей проверке корректности загрузки/изменения данных, содержащихся в CH, с помощью клиентского приложения DBeaver (так как у CH отсутствует собственный визуальный интерфейс, позволяющий осуществлять просмотр таблиц CH и написание SQL-запросов).

Последовательность действий:

- 1) Скачать DBeaver (<https://dbeaver.io/>) и открыть его.
- 2) В интерфейсе DBeaver создать новое соединение.



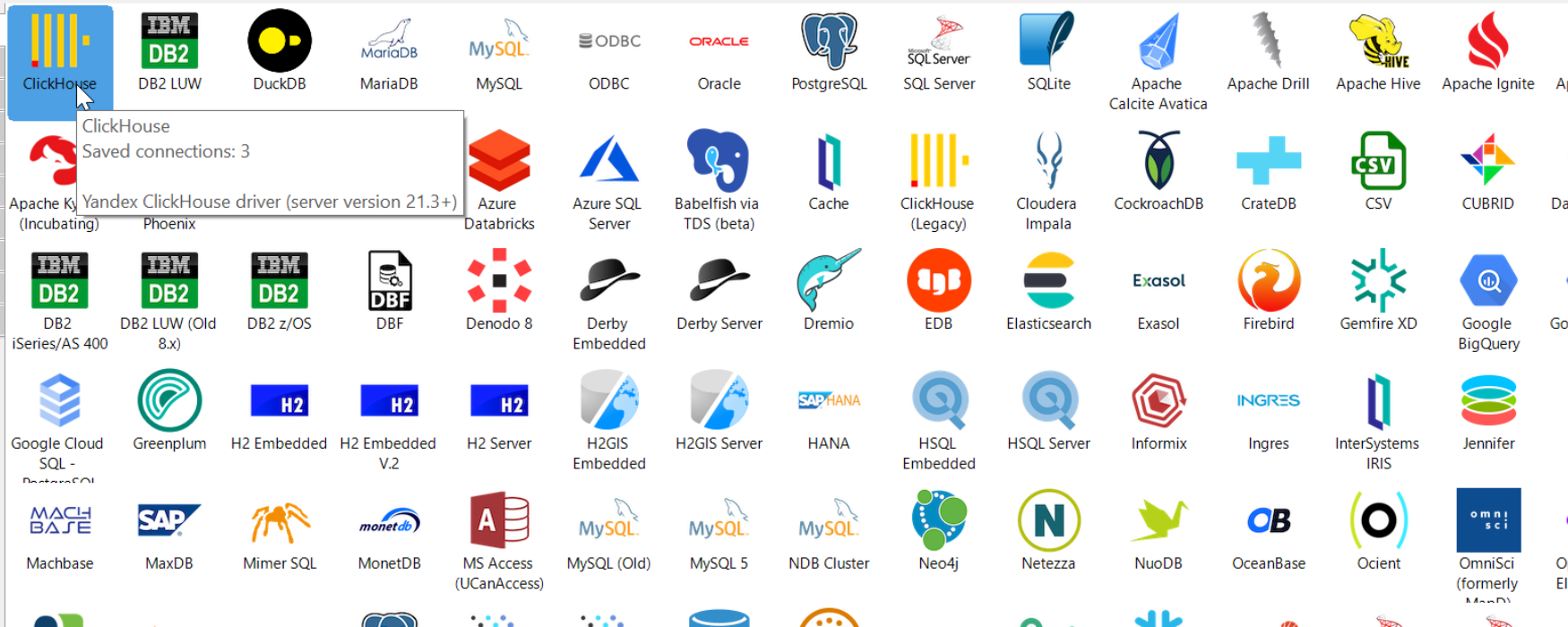
3) В предложенных вариантах найти и выбрать **ClickHouse**, дважды кликнув по нему.

Создать соединение

Выберите тип нового соединения

Создать соединение. Если вы не видите нужной базы данных в списке, то вы можете создать новый драйвер БД в менеджере драйверов.

- All
- Popular
- SQL
- NoSQL
- Analytical
- Timeseries
- Embedded
- Hadoop / BigData
- Full-text search
- Graph databases



Тест соединения ...

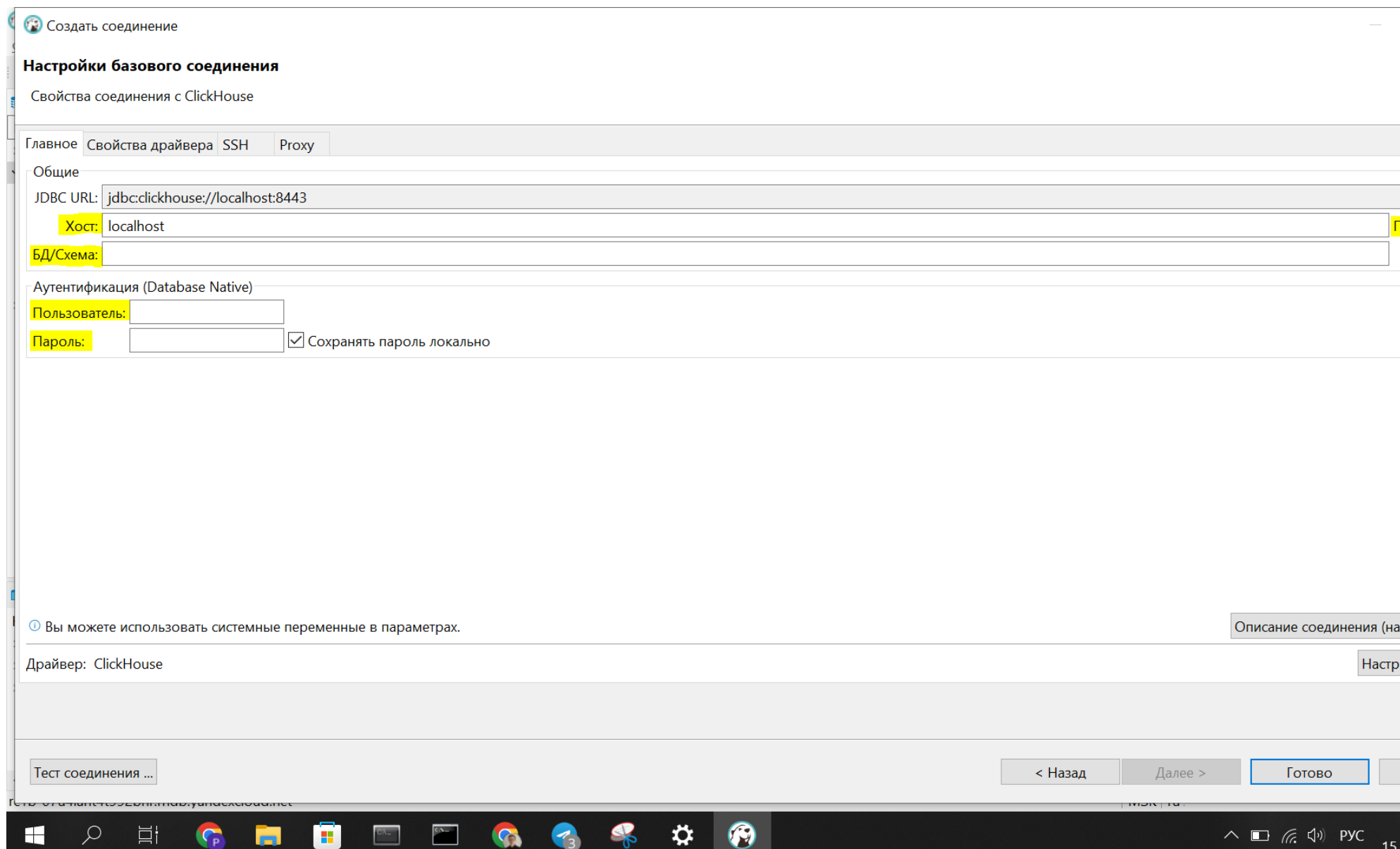
< Назад

Далее >

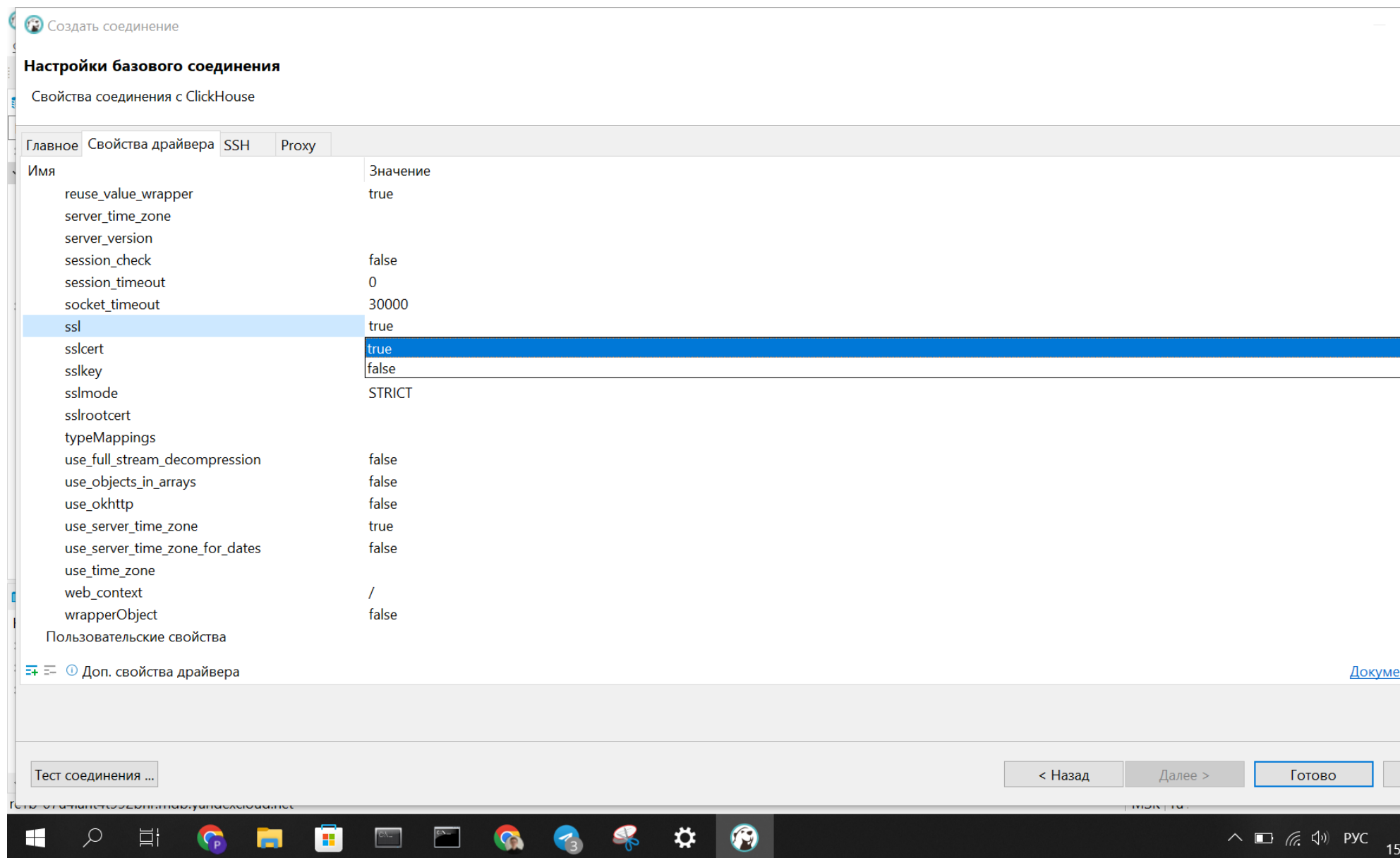
Готово

- 4) Во всплывающем окне заполнить в настройках поля “Хост”, “БД/Схема”, “Пользователь”, “Пароль”, в поле “Порт” указать **8443** (либо в некоторых случаях **9440**).

Важно: информацию “хост, БД/Схема, пользователь, пароль” нужно запросить у аккаунт-менеджера/руководителя.

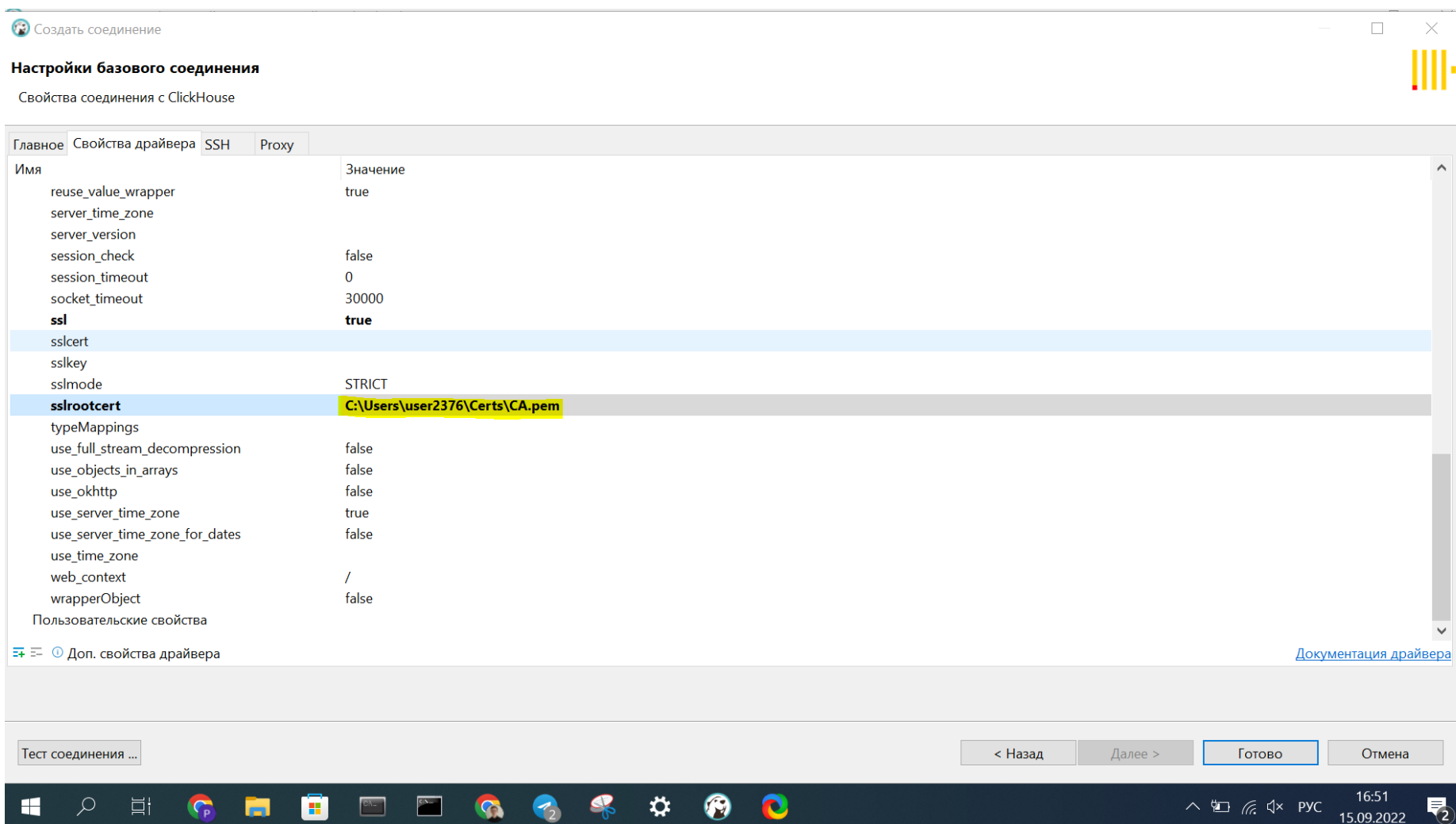


5) Перейти в раздел настроек “Свойства драйвера”, найти в списке поле **ssl**, установить значение **True**, как на скриншоте.



6) Указать в поле **sslrootcert** путь, где лежит SSL/TSL-сертификат, пример на скриншоте

Важно: SSL/TSL-сертификат нужно запросить у аккаунт-менеджера/руководителя (в некоторых случаях подойдет сертификат скачанный с официального сайта ClickHouse <https://storage.yandexcloud.net/cloud-certs/CA.pem>).



- 7) Нажать на кнопку “Тест соединения...”. В случае корректной настройки параметров соединения получим сообщение об успешном соединении с CH клиента RW.

Настройки соединения

Свойства соединения с ClickHouse



The screenshot shows the 'Настройки соединения' (Connection Settings) window for the ClickHouse JDBC Driver. The 'Свойства драйвера' (Driver Properties) tab is active, displaying a list of properties and their values. A 'Тест соединения' (Test Connection) dialog box is overlaid on the main window, indicating a successful connection to the ClickHouse server.

Имя	Значение
retry	true
reuse_value_wrapper	true
server_time_zone	
server_version	
session_check	false
session_timeout	0
socket_timeout	30000
ssl	true
sslcert	
sslkey	
sslmode	STRICT
sslrootcert	C:\Users\user2376\Certs...
typeMappings	
use_full_stream_decomp	false
use_objects_in_arrays	false
use_okhttp	false
use_server_time_zone	true
use_server_time_zone_fo	false
use_time_zone	
web_context	/
wrapperObject	false

Пользовательские свойства

Доп. свойства драйвера

[Документация драйвера](#)

Тест соединения ...

Тест соединения

Соединено (599 мс)

Сервер: ClickHouse 22.5.4.19

Драйвер: ClickHouse JDBC Driver clickhouse-jdbc 0.3.2 (revision: 1795272)

OK Сведения >>

OK Отмена

16:57 15.09.2022

- 8) В разделе интерфейса **“Базы данных”** должна появиться строчка о новом соединении, которую можно развернуть и увидеть существующие таблицы и представления внутри источника.

DBeaver 22.1.5 - <rc1b-9sj53dbwyy94t0mx.mdb.yandexcloud.net> Script

Файл Редактирование Навигация Поиск Редактор SQL База данных Окна Справка

SQL Commit Rollback Auto rc1b-9sj53d ... excloud.net rw_costs atr_Indc <rc1b-9sj53d... costs_YD_tmp costs_SS_tmp mytarget_cam...

Базы данных Проекты

Введите часть имени объекта для поиска

- > rc1a-900x4iihy87jg8p8.mdb.yandexcloud.net - rc1a-900x
- > rc1b-07u4lant4t992bnr.mdb.yandexcloud.net - rc1b-9sj53d
- > HIST
- > TMP
- > alfapartners
- > rw
- > rw_costs
 - > Таблицы
 - > costs_SS_tmp
 - > costs_SS_tmp_backup
 - > costs_YD_tmp
 - > costs_YD_tmp_backup
 - > costs_YD_tmp_test
 - > costs_YD_tmp_test_kr
 - > mytarget_campaign_stat_3463c4c710_agency_cli
 - > mytarget_campaign_stat_512486aa26_agency_cli
 - > mytarget_campaign_stat_5848659715_agency_cli
 - > mytarget_campaign_stat_6514213b56_agency_cli
 - > mytarget_campaign_stat_min_3463c4c710_agency_cli

Project - General

Название Источник

- > Bookmarks
- > Diagrams
- > Scripts

Задачи БД

Задачи: введите часть имени задачи здесь

Выполнение задачи: введите часть сообщения о

Название	Последний запуск	Последний результат	Тип	Проект	Время	Длительность	Результат
rc1b-07u4lant4t992bnr.mdb.yandexcloud.net							

MSK ru

17:07 15.09.2022

- 9) Используя скрипт в Google Collab по **ссылке** ниже произвести необходимые действия (создание\удаление, скачивание, преобразование, загрузку в CH и т.д.) с данными формата .xls , .xlsx, .csv:

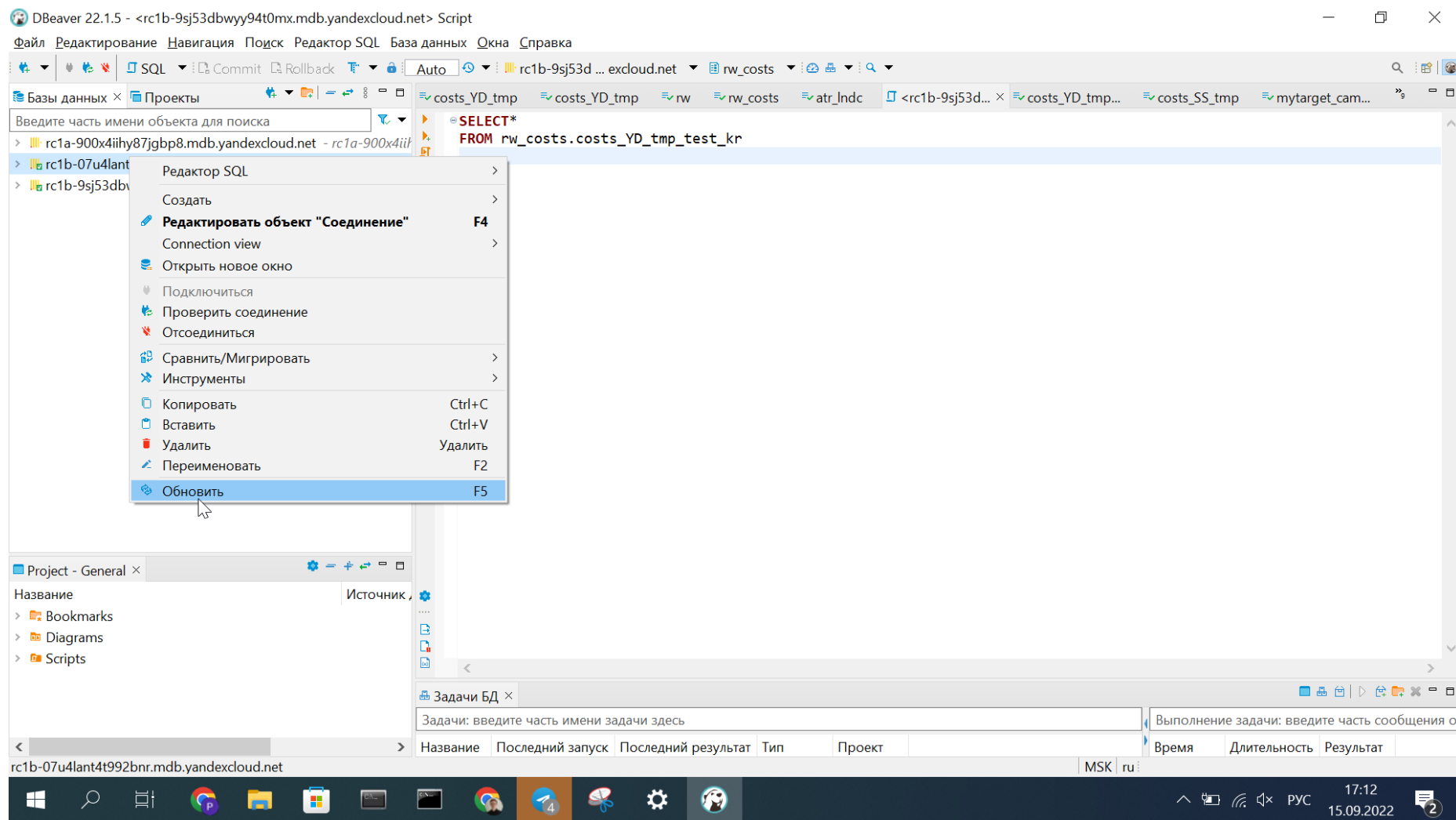
<https://colab.research.google.com/drive/11ONsK4gmAI2FU3v2NLRPYrFTWJ17s-Nc?usp=sharing>

На текущий момент скрипт позволяет осуществить:

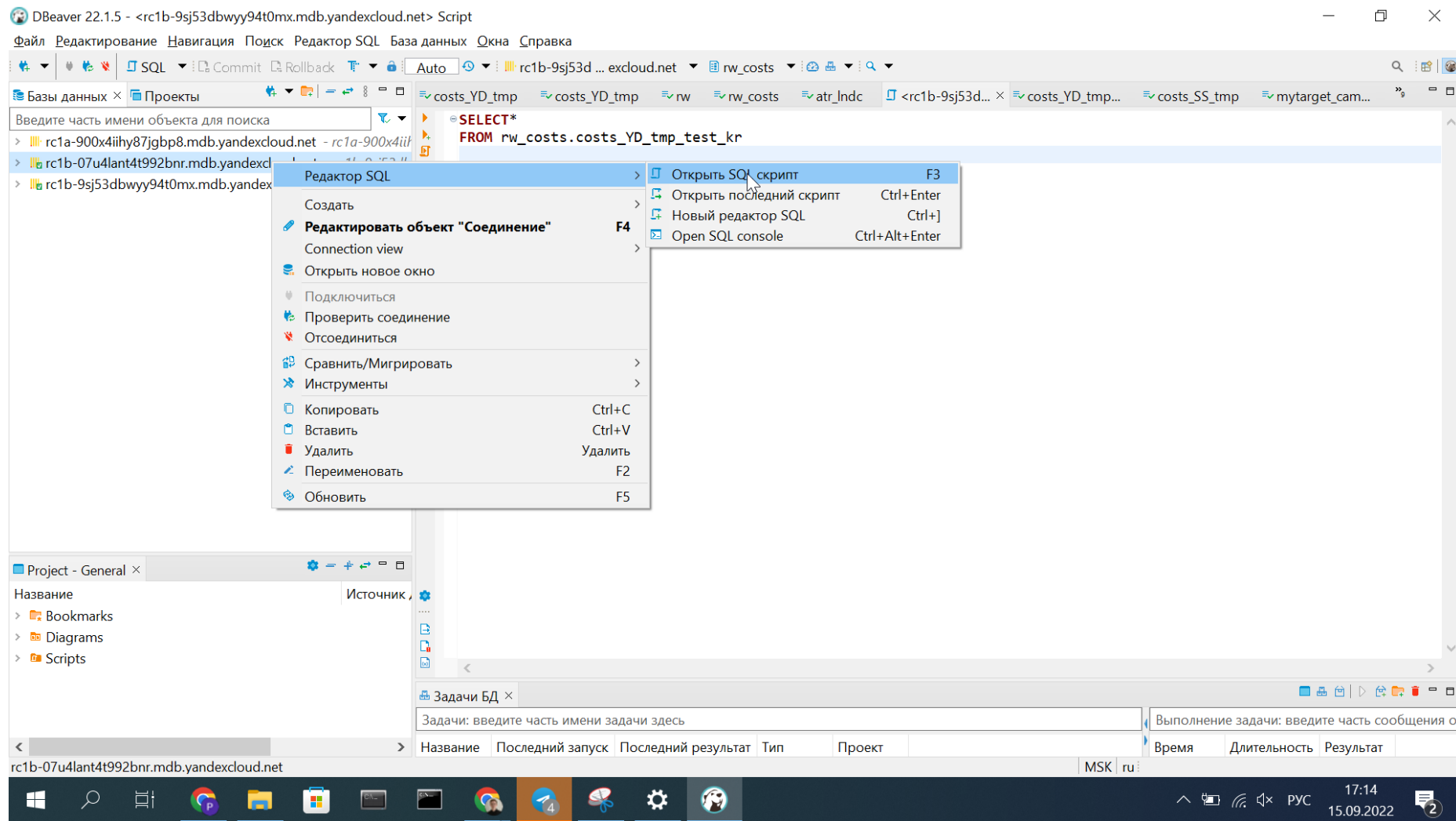
- 1) Создание и удаление таблиц в ClickHouse клиента RW.
- 2) Загрузку и преобразование в Python данных в формате Pandas Dataframe из:
 - GDrive, локального диска (.csv, xls, xlsx)
 - BigQuery
 - Clickhouse

с последующей загрузкой датафрейма в ClickHouse клиента RW.

- 10) Вернувшись в интерфейс DBeaver, обновить ресурс (кнопка **“Обновить”** выделена на скрине), чтобы подтянуть изменения, которые мы внесли в структуру и содержание источника скриптом из п.9



11) Используя редактор SQL-запросов, написать SQL-запрос к созданной/измененной таблице.



12) В поле под редактором SQL-запросов появятся данные в табличном виде, которые были загружены/изменены нами ранее скриптом из п.9. Последний шаг – проверка этих данных по количеству столбцов, строк, правильности формирования и заполнения кастомных столбцов (в примере для настоящей инструкции это столбец “product”).

DBEaver 22.1.5 - <rc1b-9sj53dbwyy94t0mx.mdb.yandexcloud.net> Script

Файл Редактирование Навигация Поиск Редактор SQL База данных Окна Справка

SQL Commit Rollback Auto rc1b-9sj53d ... excloud.net rw_costs costs_YD_tmp costs_SS_tmp mytarget_cam...

costs_YD_tmp costs_YD_tmp rw rw_costs atr_Indc <rc1b-9sj53d... costs_YD_tmp... costs_SS_tmp mytarget_cam... mytarget_cam...

Введите часть имени

- > rc1a-900x4iihy87jgpb8
- > rc1b-07u4lant4t992bni
- > rc1b-9sj53dbwyy94t0n

costs_YD_tmp_test_kr 1 x

SELECT* FROM rw_costs.costs_YD_tmp_test_kr

Введите SQL выражение чтобы отфильтровать результаты

	AdGroupId	Add	CriterionType	Criterion	CriterionId	views	clicks	cost	cost_nds	cost_with_nds_AK	product
1	h-se: 4728090949	M-11911690525	autotargeting	34573812395	557	3	638	766	766	RKO	
2	h-se: 4728090949	M-11911690526	autotargeting	34573812395	233	2	124	149	149	RKO	
3	h-se: 4728090949	M-11911690527	autotargeting	34573812395	45	0	0	0	0	RKO	
4	h-se: 4728090949	M-11911690528	autotargeting	34573812395	63	2	74	89	89	RKO	
5	h-se: 4728090949	M-11911690529	autotargeting	34573812395	12	0	0	0	0	RKO	
6	h-se: 4728090949	M-11911690530	autotargeting	34573812395	9	0	0	0	0	RKO	
7	h-se: 4728090949	M-11911690531	autotargeting	34573812395	8	0	0	0	0	RKO	
8	h-se: 4728090949	M-11911690532	autotargeting	34573812395	8	0	0	0	0	RKO	
9	h-se: 4728090949	M-11911690533	autotargeting	34573812395	2	1	16	20	20	RKO	
10	h-se: 4728090949	M-11911690534	autotargeting	34573812395	10	0	0	0	0	RKO	
11	h-se: 4728090949	M-11911690535	autotargeting	34573812395	8	0	0	0	0	RKO	
12	h-se: 4728090949	M-11911690536	autotargeting	34573812395	1	0	0	0	0	RKO	
13	h-se: 4728090949	M-11911690537	autotargeting	34573812395	8	0	0	0	0	RKO	
14	h-se: 4728090949	M-11911690538	autotargeting	34573812395	6	0	0	0	0	RKO	
15	h-se: 4728090949	M-11911690539	autotargeting	34573812395	26	0	0	0	0	RKO	
16	h-se: 4728090949	M-11911690540	autotargeting	34573812395	4	1	6	7	7	RKO	
17	h-se: 4728090949	M-11911690541	autotargeting	34573812395	12	0	0	0	0	RKO	

Save Cancel Script 16000 15 308 Rows: 1 15308 строк получено - 476ms (159ms получ.), 20%

Задачи БД

Задачи: введите часть имени задачи здесь

Выполнение задачи: введите часть сообщения об ошибке

Название	Последний запуск	Последний результат	Тип	Проект	Время	Длительность	Результат
MSK ru Запись	Инт. вставка	2 : 35 : 45	Sel: 0 0				

17:20 15.09.2022

13) Вся вышеуказанная последовательность действий по настройке DBeaver также доступна в обучающем видео по ссылке:

https://drive.google.com/file/d/1BLOU0WW3NDq-_v1aPA82xGgM14ITVP0I/view