

Функциональное описание продукта “Коннектор Б24”

Программа для ЭВМ «Коннектор Битрикс24 в Clickhouse (далее - Продукт) предназначена для выгрузки данных из Программы для ЭВМ “Битрикс24” (далее - Источник) в аналитические системы и представляет собой структурированный программный код, написанный на языке Python.

Продукт разбит на модули, каждый из которых описывает:

1. Получение данных из Источника путем обращения к программному интерфейсу Источника (далее - API)
2. Приведение к структуре и типам , предполагаемой для записи, полученных через API данных (далее - сериализация)
3. Создание структурированных таблиц в базах данных (далее - БД) системы управления базами данных (далее - Приёмник)
4. Запись сериализованных данных в таблицы Приёмника

Перечень обрабатываемых для выгрузки сущностей описан в [документации](#) по Продукту.

Продукт запрашивает данные get запросами из API Источника, сериализует и загружает данные в таблицу Приёмника. Таким образом Продукт относится к классу ETL (extract - transform - load).

Продукт поддерживает инкрементальную выгрузку основанную на данных (CDC data-based technology), что обеспечивает снижение нагрузки на Источник, Приёмник и Продукт.

Продукт поддерживает систему отслеживания создания и удаления объектов, что обеспечивает автогенерацию и наполнение данными таблиц для вновь созданных экземпляров объектов.



Продукт выгружает как типовые так и пользовательские поля. В том числе поля типа множественный список.

Продукт работает по расписанию благодаря планировщику заданий, входящему в поставку.

Периодичность выгрузки планируется для каждого объекта индивидуально.

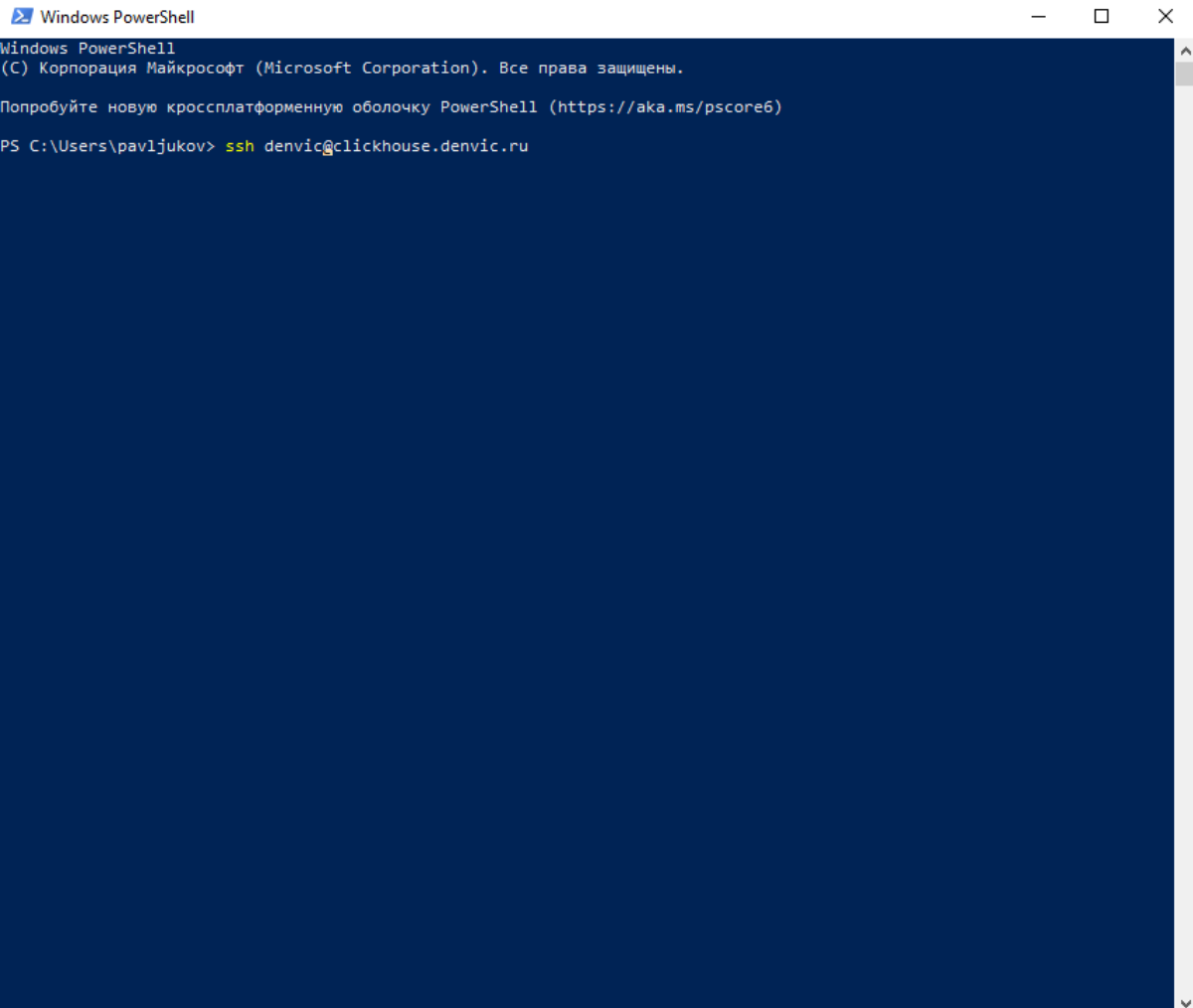
Например, если вы создали новый смарт процесс со своей структурой и типами данных полей, то для него будет создана и сопоставлена отдельная таблица в Приёмнике, которая будет наполняться данными по мере их добавления в Источник основываясь на дате последней выгрузки данных по этому Объекту

Пример таблицы:

- ▼  crm_deal
 - ▼  Колонки
 - ABC ID (String)
 - 123 OPPORTUNITY (Nullable(Decimal(18, 2)))
 - 🕒 dt_inserted (DateTime)
 - ABC TITLE (Nullable(String))
 - ABC TYPE_ID (Nullable(String))
 - ABC CATEGORY_ID (Nullable(String))
 - ABC STAGE_ID (Nullable(String))
 - ABC STAGE_SEMANTIC_ID (Nullable(String))
 - ABC IS_NEW (Nullable(String))
 - ABC IS_RECURRING (Nullable(String))
 - ABC IS_RETURN_CUSTOMER (Nullable(String))
 - ABC IS_REPEATED_APPROACH (Nullable(String))
 - 123 PROBABILITY (Nullable(Int64))
 - ABC CURRENCY_ID (Nullable(String))
 - ABC IS_MANUAL_OPPORTUNITY (Nullable(String))
 - 123 TAX_VALUE (Nullable(Float32))
 - ABC COMPANY_ID (Nullable(String))
 - ABC CONTACT_ID (Nullable(String))
 - 🔗 CONTACT_IDS (Array(Nullable(String)))
 - ABC QUOTE_ID (Nullable(String))
 - 🕒 BEGINDATE (Nullable(DateTime64(3)))
 - 🕒 CLOSEDATE (Nullable(DateTime64(3)))
 - ABC OPENED (Nullable(String))
 - ABC CLOSED (Nullable(String))
 - ABC COMMENTS (Nullable(String))
 - ABC ASSIGNED_BY_ID (Nullable(String))
 - ABC CREATED_BY_ID (Nullable(String))
 - ABC MODIFY_BY_ID (Nullable(String))
 - ABC MOVED_BY_ID (Nullable(String))
 - 🕒 DATE_CREATE (Nullable(DateTime64(3)))
 - 🕒 DATE_MODIFY (Nullable(DateTime64(3)))
 - 🕒 MOVED_TIME (Nullable(DateTime64(3)))
 - ABC SOURCE_ID (Nullable(String))
 - ABC SOURCE_DESCRIPTION (Nullable(String))
 - ABC LEAD_ID (Nullable(String))
 - ABC ADDITIONAL_INFO (Nullable(String))
 - ABC LOCATION_ID (Nullable(String))
 - ABC ORIGINATOR_ID (Nullable(String))
 - ABC ORIGIN_ID (Nullable(String))
 - ABC UTM_SOURCE (Nullable(String))
 - ABC UTM_MEDIUM (Nullable(String))
 - ABC UTM_CAMPAIGN (Nullable(String))
 - ABC UTM_CONTENT (Nullable(String))
 - ABC UTM_TERM (Nullable(String))
 - ABC PARENT_ID_142 (Nullable(String))
 - ABC PARENT_ID_149 (Nullable(String))
 - ABC PARENT_ID_185 (Nullable(String))

Установка

1. Получаем доступы к серверу, где необходимо развернуть скрипты выгрузки.
2. Подключаюсь к серверу клиента.
 - а. Проверяю есть ли необходимый ЯП и его версию (для работы необходим Python, версии не ниже 3.11.2)



```
Windows PowerShell
(C) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.
Попробуйте новую кроссплатформенную оболочку PowerShell (https://aka.ms/pscore6)
PS C:\Users\pavljukov> ssh denvic@clickhouse.denic.ru
```

3. Загружаю репозиторий с гита компании с файлами коннектора.

```
mc [denvic@clickhouse]:~/scripts/newConnector
Windows PowerShell
(C) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.

Попробуйте новую кроссплатформенную оболочку PowerShell (https://aka.ms/pscore6)

PS C:\Users\pavljukov> ssh denvic@clickhouse.denic.ru
PS C:\Users\pavljukov> ssh denvic@clickhouse.denic.ru
PS C:\Users\pavljukov> ssh denvic@10.10.1.115
denvic@10.10.1.115's password:
Linux clickhouse 4.19.0-20-amd64 #1 SMP Debian 4.19.235-1 (2022-03-17) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Tue Nov  5 11:20:20 2024 from 192.168.1.160
denvic@clickhouse:~$ mc

denvic@clickhouse:~/scripts/newConnector$ git clone https://git.denic.ru/denic/connector_b24_2_clickhouse.git
```

4. Переключаюсь на ветку "dev" репозитория

```
mc [denvic@clickhouse]:~/scripts/newConnector/connector_B24_2_clickhouse
Windows PowerShell
(C) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.

Попробуйте новую кроссплатформенную оболочку PowerShell (https://aka.ms/pscore6)

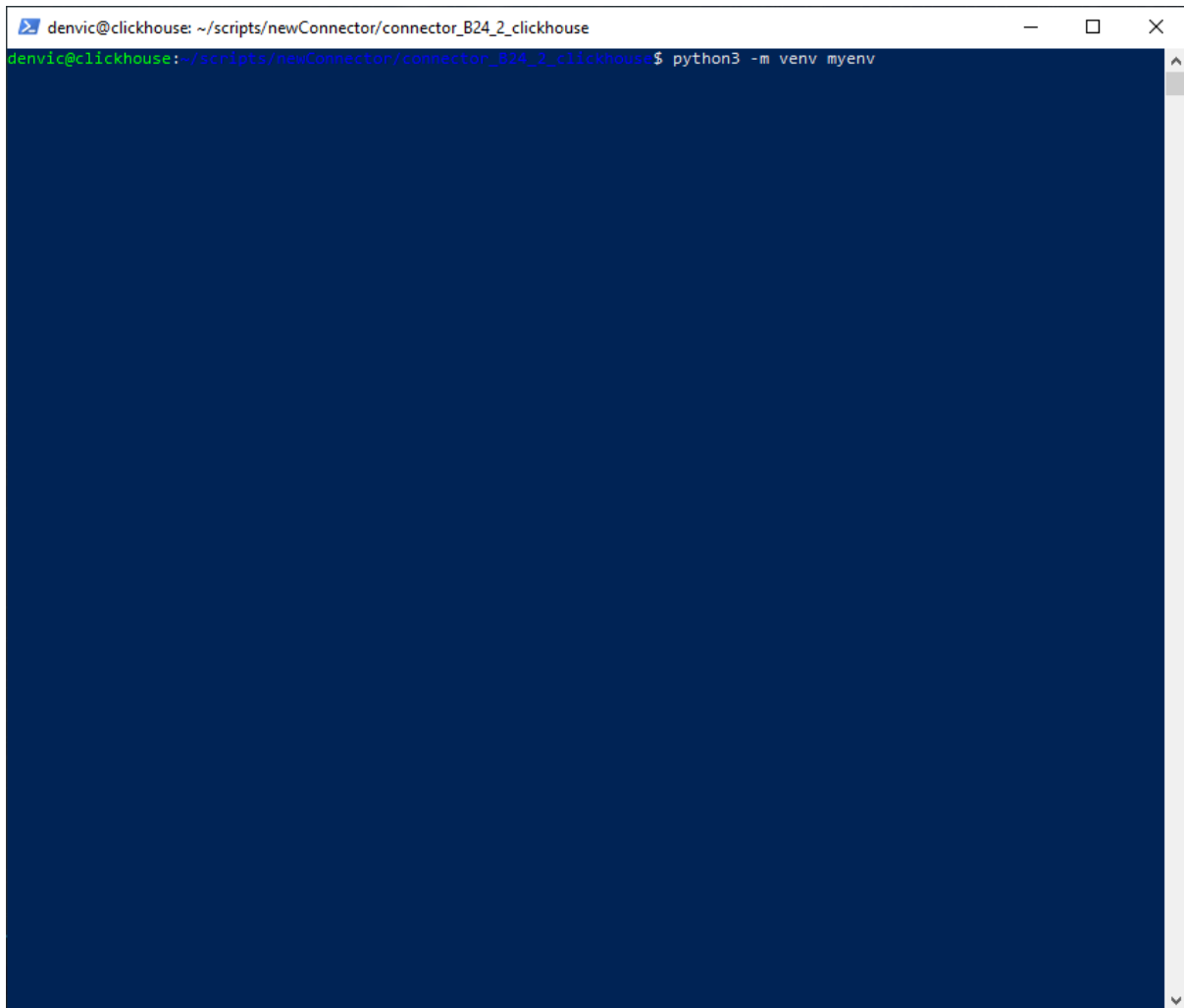
PS C:\Users\pavljukov> ssh denvic@clickhouse.denvic.ru
PS C:\Users\pavljukov> ssh denvic@clickhouse.denvic.ru
PS C:\Users\pavljukov> ssh denvic@10.10.1.115
denvic@10.10.1.115's password:
Linux clickhouse 4.19.0-20-amd64 #1 SMP Debian 4.19.235-1 (2022-03-17) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Tue Nov  5 11:20:20 2024 from 192.168.1.160
denvic@clickhouse:~$ mc

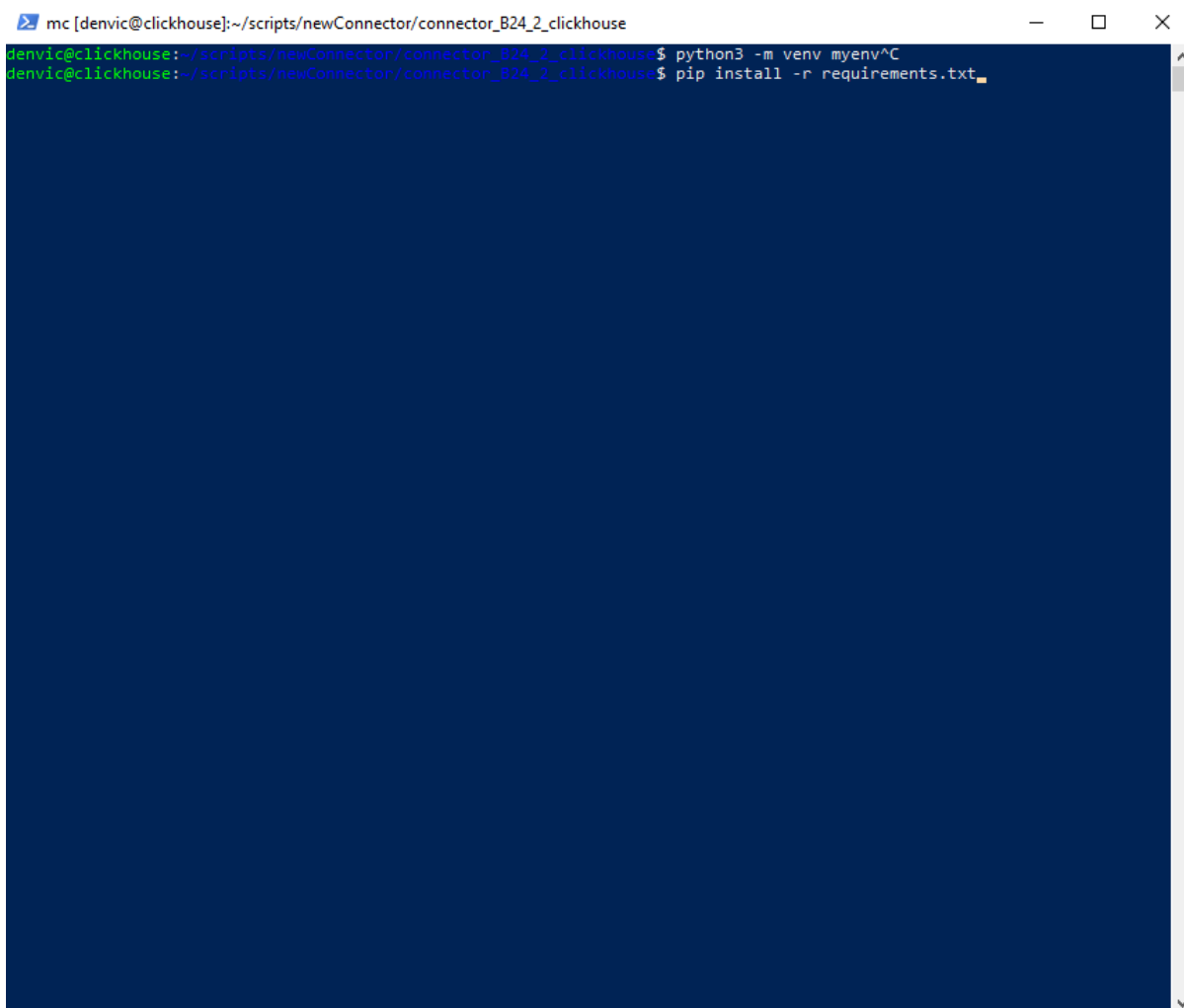
denvic@clickhouse:~/scripts/newConnector$ git clone https://git.denvic.ru/denvic/connector_B24_2_clickhouse.git^C
denvic@clickhouse:~/scripts/newConnector$
denvic@clickhouse:~/scripts/newConnector/connector_B24_2_clickhouse$ git checkout dev
```

5. Создаю Виртуальное окружение для Python

A terminal window with a dark blue background. The title bar shows the user 'denvic@clickhouse' and the current directory '~/scripts/newConnector/connector_B24_2_clickhouse'. The command 'python3 -m venv myenv' has been entered and is highlighted in green. The rest of the terminal area is empty.

```
denvic@clickhouse: ~/scripts/newConnector/connector_B24_2_clickhouse
denvic@clickhouse:~/scripts/newConnector/connector_B24_2_clickhouse$ python3 -m venv myenv
```

6. Загружаю необходимые для работы скрипта пакеты расширений. Необходимые пакеты загружаю с помощью менеджера пакетов "pip". Необходимые пакеты находятся в разделе проекта в файле requirements.txt



```
mc [denvic@clickhouse]:~/scripts/newConnector/connector_B24_2_clickhouse
denvic@clickhouse:~/scripts/newConnector/connector_B24_2_clickhouse$ python3 -m venv myenv^C
denvic@clickhouse:~/scripts/newConnector/connector_B24_2_clickhouse$ pip install -r requirements.txt
```

Настройка

1. Для настройки подключений, необходимо скопировать шаблон конфигурационного файла в главный раздел проекта (шаблон находится там же)

```

mc [denvic@clickhouse]:~/scripts/newConnector/connector_B24_2_clickhouse
Left File Command Options Right
<-- ..ripts/newConnector/connector_B24_2_clickhouse -.[^]> <-- ~ .[^]>
.n Name Size Modify time .n Name Size Modify time
/.. UP--DIR Sep 4 11:07 /.. UP--DIR Apr 1 2022
/.git 4096 Oct 16 14:33 /.cache 4096 Jun 20 15:03
/._pycache__ 4096 Sep 4 11:40 /.config 4096 Apr 4 2022
/etl 4096 Oct 16 14:33 /.dotnet 4096 May 31 2024
/myenv 4096 Dec 3 10:42 /.gnupg 4096 Apr 4 2022
/plugins 4096 Oct 2 10:09 /.local 4096 Mar 6 2024
/sql 4096 Oct 16 14:33 /.vscode-server 4096 Nov 5 15:49
/tests 4096 Oct 14 16:43 /scripts 4096 Sep 4 11:06
.gitignore 110 Sep 4 14:00 .bash_history 7833 Nov 15 07:53
README.md 2096 Sep 4 11:08 .bash_logout 220 Apr 1 2022
YandexInternalRootCA.crt 1862 Sep 4 11:21 .bashrc 3526 Apr 1 2022
arg_parser.py 1281 Sep 4 11:08 .gitconfig 59 Mar 6 2024
config.cfg 1557 Oct 2 11:34 .profile 807 Apr 1 2022
config_example.cfg 0 Apr 6 2022
config_parser.py 183 Nov 2 17:21
main.py
requirements.txt

Copy file "config_example.cfg" with source mask:
to:
r_B24_2_clickhouse/config.cfg
[ ] Follow links [ ] Dive into subdir if exists
[x] Preserve attributes [ ] Stable symlinks
[< OK >] [ Background ] [ Cancel ]

config_example.cfg 36G/197G (18%) UP--DIR 36G/197G (18%)
Hint: Please send any bug reports to mc-devel@gnome.org
denvic@clickhouse:~/scripts/newConnector/connector_B24_2_clickhouse$
1|help 2|Menu 3|view 4|Edit 5|Copy 6|RenMov 7|mkdir 8|Delete 9|PullDn 10|quit

```

2. Переименовываем скопированный шаблон конфигурации из “config_example.cfg” в “config.cfg”
3. Далее, указываем настройки подключения в новом файле конфигурации, который создали на предыдущем шаге.
 - a. раздел [Bitrix]
 - i. Переменная BitrixURL=*** - указываем веб-хук, с помощью которого, скрипт будет получать данные с портала Битрикс24
 - ii. Переменная operation_limit=*** - (только для облачных порталов) Служит для того, чтобы не заблокировали веб-хук по количеству обращений. Максимальный лимит, при котором происходит блокировка метода 480 секунд. Оптимальное значение для данной переменной 400 секунд (указывается просто 400)
 - b. раздел [Clickhouse]
 - i. Переменная url=*** - указываем хост, где располагается СУБД Clickhouse. Указывать необходимо с портом. Пример: `url=https://<хост>:<порт>`
 - ii. Переменная verify_path=*** - указываем путь до сертификата безопасности. Если СУБД расположена в кластере Яндекса, то указываем путь до сертификата Яндекса. Если СУБД располагается на собственном сервере и есть сертификат, то

указываем путь до собственного сертификата. Если сертификата нет, а подключение осуществляется не по защищенному протоколу (https), то комментируем строку

- iii. Переменная `database=***` - Указываем имя Базы данных, в которую будем загружать данные
 - iv. Переменная `clickhouse_host=***` - Указываем хост, где располагается СУБД Clickhouse. В данной переменной хост указывается без протокола и порта.
 - v. Переменная `clickhouse_user=***` - Указываем Имя пользователя от СУБД под которым данные будут поступать в Базу данных.
 - vi. Переменная `clickhouse_password=***` - Указываем пароль от пользователя под которым данные будут поступать в Базу данных.
 - vii. Переменная `clickhouse_port=***` - Указываем порт, по которому будет работать clickhouse client. Обычно 9440, если используется защищенное соединение. И 9000, если используется не защищенное соединение.
 - viii. Так же есть опциональная переменная `secure=***` - принимает 2 значения True/False. При значении True, происходит проверка сертификатов при подключении к Базе данных.
- c. раздел [ImportSettings]
- i. Переменная `batch_size=***` - данная переменная служит для определения размера пакета отправляемого в СУБД Clickhouse из Битрикс24.

```
mc [denvic@clickhouse]:~/scripts/newConnector/connector_B24_2_clickhouse
GNU nano 3.2 /home/denic/scripts/newConnector/connector_B24_2_clickhouse/config.cfg Modified
# config.cfg

[Bitrix]
BitrixURL=https://denvic-nfr.bitrix24.ru/rest/1/h9kp54jtyzm6j3a/
operating_limit=400

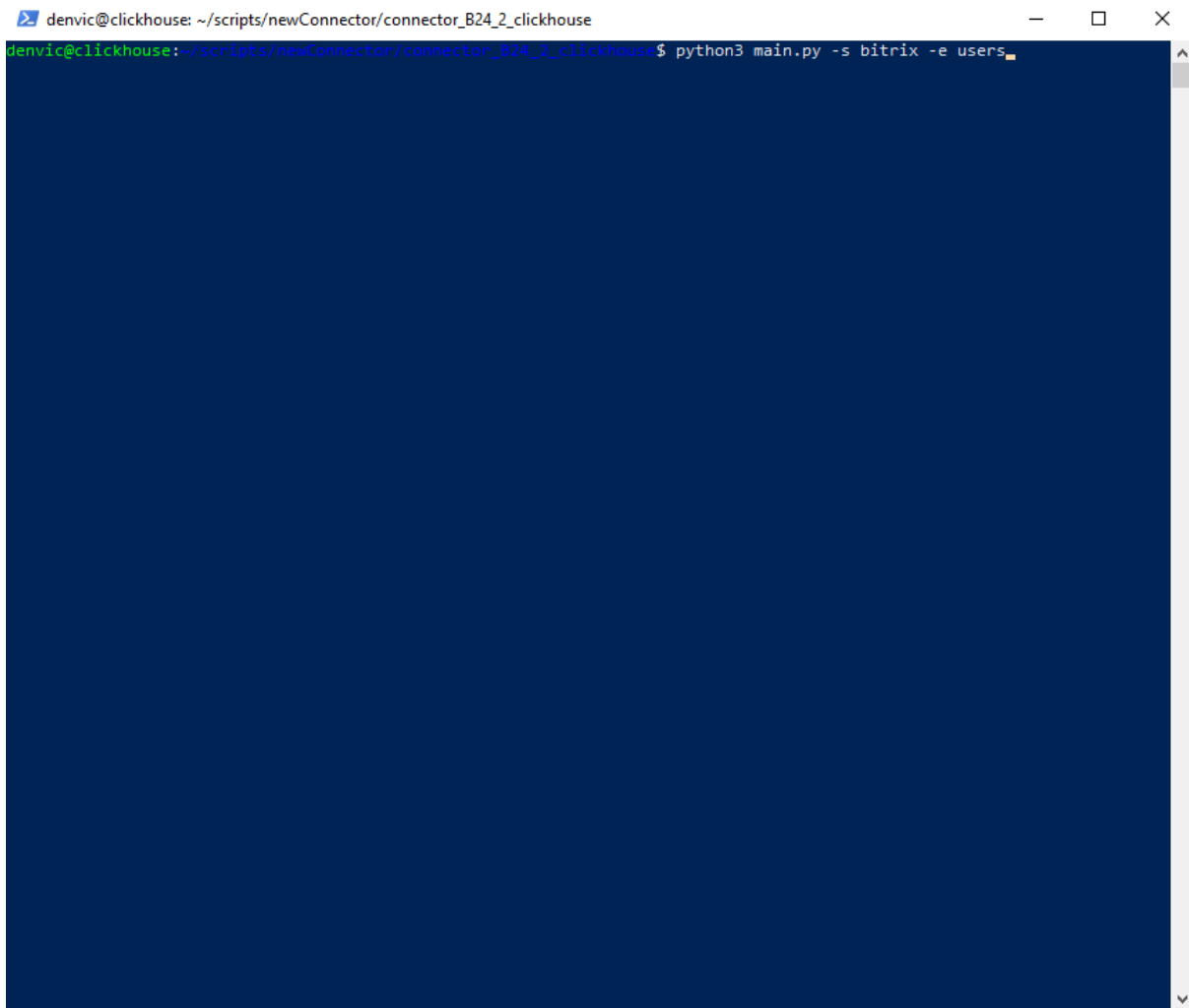
[Clickhouse]
url=https://rc1a-c9vohcmivp3cf57s.mdb.yandexcloud.net:8443
verify_path=/home/denic/scripts/newConnector/connector_B24_2_clickhouse/YandexInternalRootCA.crt
database=test_vlad
clickhouse_host=rc1a-c9vohcmivp3cf57s.mdb.yandexcloud.net
clickhouse_user=test_vlad
clickhouse_password=u9G-D2Z-LGJ-2CG
clickhouse_port=9440

[ImportSettings]
batch_size=2500

[BitrixApp]
BitrixURL=https://denvic-nfr.bitrix24.ru
client_id=local.66fcf30e6362b4.64298607
client_secret=v6jQ4W3zM8K63eFQmzpTGTdyGmmMxLqXTeSFgsbm67ngwpT70L
access_token=4403fd660071d7660066caca0000001000007c4abacad2467c2e94b5a800aab835547
refresh_token=348224670071d7660066caca0000001000007b7935d370d8df83b432d585bd7d10dd6

^G Get Help      ^O Write Out    ^M Where Is     ^K Cut Text     ^J Justify     ^C Cur Pos     ^-U Undo        ^-A Mark Text
^X Exit          ^R Read File   ^_ Replace     ^U Uncut Text  ^T To Spell   ^_ Go To Line  ^-E Redo       ^-G Copy Text
```

4. После указания всех настроек подключения, производим тестовую выгрузку. Обычно для тестовой выгрузки делаем выгрузку пользователей портала Битрикс24.

A terminal window with a dark blue background. The title bar at the top reads "denvic@clickhouse: ~/scripts/newConnector/connector_B24_2_clickhouse". The terminal prompt is "denvic@clickhouse: ~/scripts/newConnector/connector_B24_2_clickhouse\$". The command entered is "python3 main.py -s bitrix -e users". The rest of the terminal area is empty, indicating the command has been executed but its output is not visible.

```
denvic@clickhouse: ~/scripts/newConnector/connector_B24_2_clickhouse
denvic@clickhouse: ~/scripts/newConnector/connector_B24_2_clickhouse$ python3 main.py -s bitrix -e users
```

5. Ставим выгрузку на расписание с помощью стандартной linux-утилиты - crontab

denvic@clickhouse: ~

GNU nano 3.2

/tmp/crontab.YZnWox/crontab

Modified

```
# Edit this file to introduce tasks to be run by cron.
#
# Each task to run has to be defined through a single line
# indicating with different fields when the task will be run
# and what command to run for the task
#
# To define the time you can provide concrete values for
# minute (m), hour (h), day of month (dom), month (mon),
# and day of week (dow) or use '*' in these fields (for 'any').
#
# Notice that tasks will be started based on the cron's system
# daemon's notion of time and timezones.
#
# Output of the crontab jobs (including errors) is sent through
# email to the user the crontab file belongs to (unless redirected).
#
# For example, you can run a backup of all your user accounts
# at 5 a.m every week with:
# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h dom mon dow   command
*/5 * * * * /usr/bin/python3 /var/scripts/b24ToCH/connector_B24_2_clickhouse/main.py -s bitrix -e all_
```

^G Get Help ^O Write Out ^W Where Is ^X Cut Text ^J Justify ^C Cur Pos M-U Undo M-A Mark Text
^X Exit ^R Read File ^N Replace ^U Uncut Text ^T To Spell ^_ Go To Line M-E Redo M-C Copy Text