

# Описание

Модуль работы с ГАР

[Основная документация ФИАС](#)

## Импорт ГАР ФИАС

### Предварительная подготовка

- Проверить наличие zip-распаковщика. При его отсутствии установить следующей командой:

```
sudo apt install unzip
```

- Проверить наличие утилиты xml2. При ее отсутствии установить следующей командой:

```
sudo apt install xml2
```

- Скопировать папку fullsh;
- Исправить файлы, убрав переносы командой:

```
find . -type f -name '*.sh' -exec sed -i 's/\r$//' {} +
```

### Запуск

Главный файл, [autoImport.sh](#), поддерживает некоторые флаги и параметры. Итак, кратко о них:

1) -p = "..." - путь, куда отправлять уведомления о шагах. Не работает если также не указан идентификатор пользователя; 2) -u = "123" - идентификатор пользователя, для отправки уведомления о шагах. Не работает, если также не указан путь; 3) -d = "fias" - имя базы данных, в которую бет загружен ФИАС. По умолчанию: fias; 4) tmpFolder = "..." - путь, по которому будет скачен и распакован архив. Также там будут проводиться все манипуляции с файлами. По умолчанию: ./xml; 5) -r - отключает удаление, скачивание, проверку статуса, проверку версий и запись статуса в БД. Полезно, когда надо вручную запустить обновление и при этом не хочется заново качать архив; 6) -v - отключает проверку статуса и проверку версий, а также запись статуса в БД; 7) -g включает разархивацию архива и парсинг xml в csv. Актуально только при включенном флаге -r.

Скрипт необходимо запускать от пользователя postgres, сделать это можно так:

```
sudo -u postgres bash autoImport.sh
```

### Описание работы

- создание папки для хранения файлов ( tmpFolder );
- загрузка gar из интернета;
- распаковка данных (данные распакуются по регионам);
- удаление архива .zip;
- конвертация полученных .xml в .csv (см. [конвертация](#));
- удаление всех .xml;
- удаление старых временных таблиц;
- создание новых временных таблиц (см. [создание временных таблиц](#));
- создание схем по каждому региону;
- создание временных таблиц для каждого региона (см. [создание регионов](#));
- вставка только актуальных данных для каждого региона (см [вставка данных](#));
- удаление папки с данными;
- выполнение основного скрипта (см. [основной скрипт обновления](#));

- данные из каждого региона обновляются и переливаются в основную временную таблицу;
- после каждой операции удаляется схема отработанного региона со всеми данными;
- выполняется скрипт подмены временных таблиц на настоящие (с удалением настоящих) (см. [замена](#)).

Среднее время выполнения скрипта зависит от скорости интернета (для скачивания архива) и от частоты процессора (конвертация данных). Последний запуск на временном контуре составил около трёх часов.

## Отличия от fias

- теперь aoguids хранится, как uuid;
- уменьшено количество данных, за счёт отказа от неиспользуемых данных (`actstatus != 1 AND currstatus != 0`).