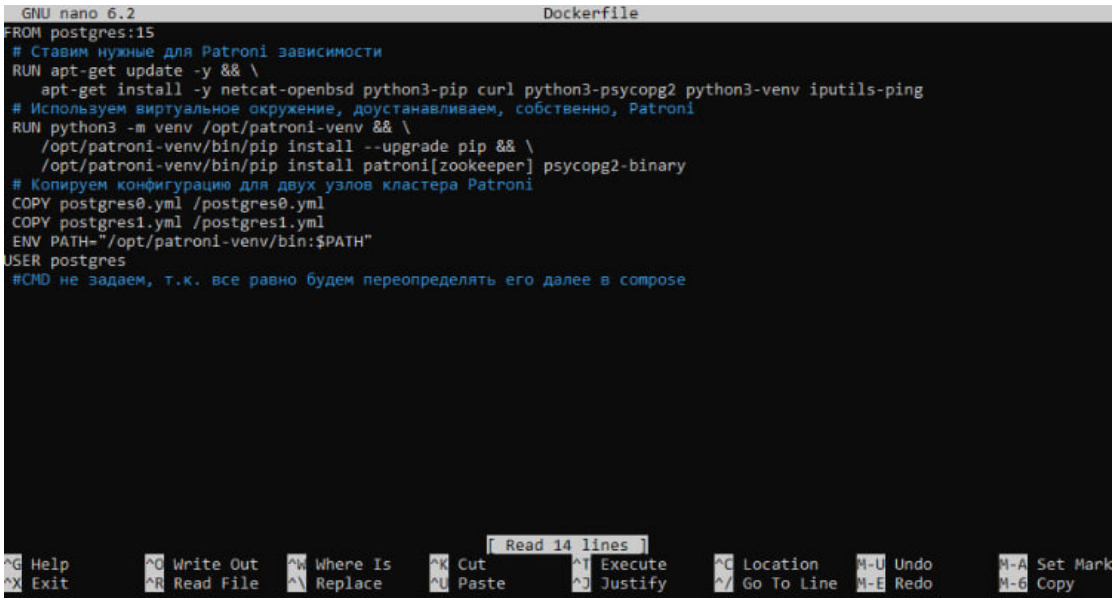
Лабораторная №3

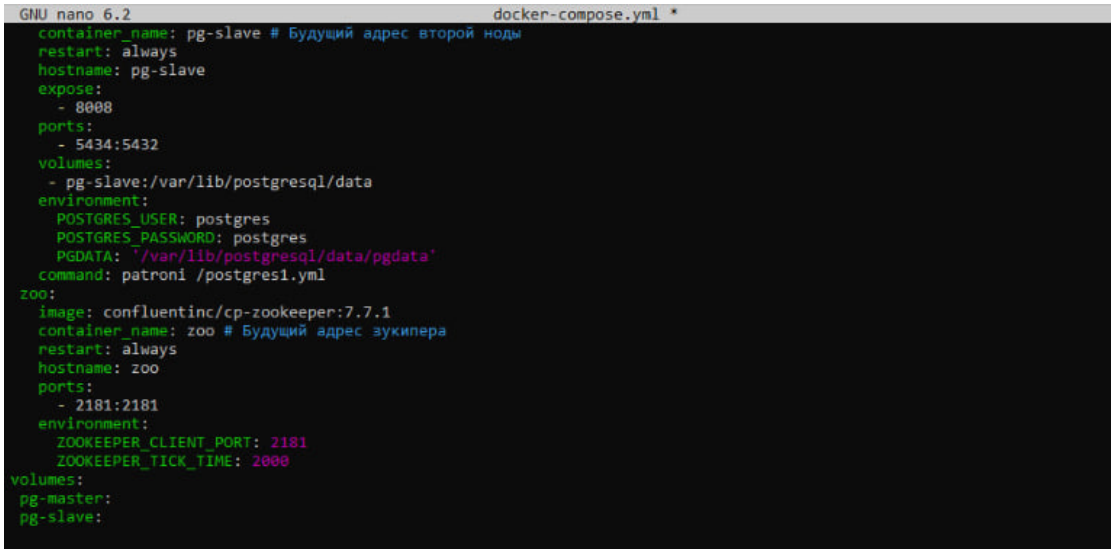
HA Postgres Cluster

Часть1 - Поднимаем Postgres

Подготавливаем Dockerfile для нашего постгреса. Кластеризацию будем делать с помощью Patroni, а ему необходим доступ к бинарникам самого постгреса. Поэтому будем билдить образ, который сразу содержит в себе Postgres + Patroni



Подготавливаем compose файл, в котором описываем наш деплой постгреса. Так же добавляем в него Zookepeer, который нужен для непосредственного управления репликацией и определением “лидера” кластера



Создаем упомянутые выше postgres0.yml и затем на основе него — postgres1.yml (надо будешь лишь поменять имя, адреса и место хранения данных ноды с первой на вторую)

```
 GNU nano 6.2                                  postgres0.yml *
scope: my_cluster # Имя нашего кластера
name: postgresql0 # Имя первой ноды
restapi: # Адреса первой ноды
 listen: pg-master:8008
 connect_address: pg-master:8008
zookeeper:
   hosts:
     - zoo:2181 # Адрес Zookeeper
bootstrap:
 dcs:
   ttl: 30
   loop_wait: 10
   retry_timeout: 10
   maximum_lag_on_failover: 10485760
   master_start_timeout: 300
   synchronous_mode: true
   postgresql:
     use_pg_rewind: true
     use_slots: true
     parameters:
       wal_level: replica
       hot_standby: "on"
       wal_keep_segments: 8
       max_wal_senders: 10
       max_replication_slots: 10
       wal_log_hints: "on"
       archive_mode: "always"
       archive_timeout: 1800s
       archive_command: mkdir -p /tmp/wal_archive && test ! -f /tmp/wal_archive/%f && cp %p /tmp/wal_archive/%f
 pg_hba:
 - host replication replicator 0.0.0.0/0 md5
 - host all all 0.0.0.0/0 md5
postgresql:
 listen: 0.0.0.0:5432
 connect_address: pg-master:5432 # Адрес первой ноды
 data_dir: /var/lib/postgresql/data/postgresql0 # Место хранения данных первой ноды
 bin_dir: /usr/lib/postgresql/15/bin
 pgpass: /tmp/pgpass0
 authentication:
   replication: # логопасс для репликаци, при желании можно поменять
     username: replicator
     password: rep-pass
   superuser: # админский логопасс, при желании можно поменять (в том числе в файле compose)
     username: postgres
     password: postgres
 parameters:
   unix_socket_directories: '.'
watchdog:
 mode: off
tags:
   nofailover: false
   noloadbalance: false
   clonefrom: false
   nosync: false
```

Деплоим. Проверяем в логах, что зукипер запустился, и что одна нода постгреса из двух стала лидером/овнером/мастером (есть вероятность, что вопреки названию это будет НЕ pg-master, это нормально!)

```
xt time you run initdb.
syncing data to disk ... ok


Success. You can now start the database server using:

    /usr/lib/postgresql/15/bin/pg_ctl -D /var/lib/postgresql/data/postgresql1 -l logfile start

2024-12-09 16:59:50,837 INFO: postmaster pid=31
2024-12-09 16:59:50.845 UTC [31] LOG:  starting PostgreSQL 15.10 (Debian 15.10-1.pgdg120+1) on x86_64-pc-linux-gnu, comp
iled by gcc (Debian 12.2.0-14) 12.2.0, 64-bit
2024-12-09 16:59:50.846 UTC [31] LOG:  listening on IPv4 address "0.0.0.0", port 5432
2024-12-09 16:59:50.851 UTC [31] LOG:  listening on Unix socket "./.s.PGSQL.5432"
2024-12-09 16:59:50.862 UTC [35] LOG:  database system was shut down at 2024-12-09 16:59:47 UTC
2024-12-09 16:59:50.901 UTC [36] FATAL:  the database system is starting up
localhost:5432 - rejecting connections
2024-12-09 16:59:50.909 UTC [31] LOG:  database system is ready to accept connections
localhost:5432 - accepting connections
2024-12-09 16:59:51,015 INFO: establishing a new patroni heartbeat connection to postgres
2024-12-09 16:59:51,050 INFO: running post_bootstrap
2024-12-09 16:59:51,097 INFO: Reaped pid=45, exit status=0
2024-12-09 16:59:51,122 INFO: initialized a new cluster
2024-12-09 17:00:01,136 INFO: Lock owner: postgresql1; I am postgresql1
2024-12-09 17:00:01,163 INFO: Enabled synchronous replication
2024-12-09 17:00:01,177 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:00:11,105 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:00:21,102 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:00:31,095 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:00:41,103 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:00:51,096 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:01:01,110 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:01:11,096 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-09 17:01:21,100 INFO: no action. I am (postgresql1), the leader with the lock
root@hakuna-matata:~/post#
```

Часть 2 - Проверяем репликацию

Запустим оба нода

Создаём таблицу на pg-master

Query Query History

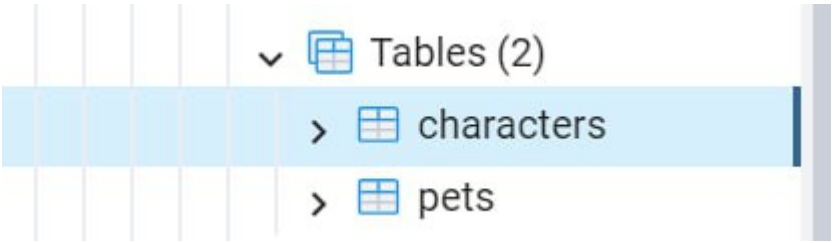
```
1  CREATE TABLE Pets (
2      id int,
3      name varchar,
4      type varchar
5  )
```

Data Output Messages Notifications

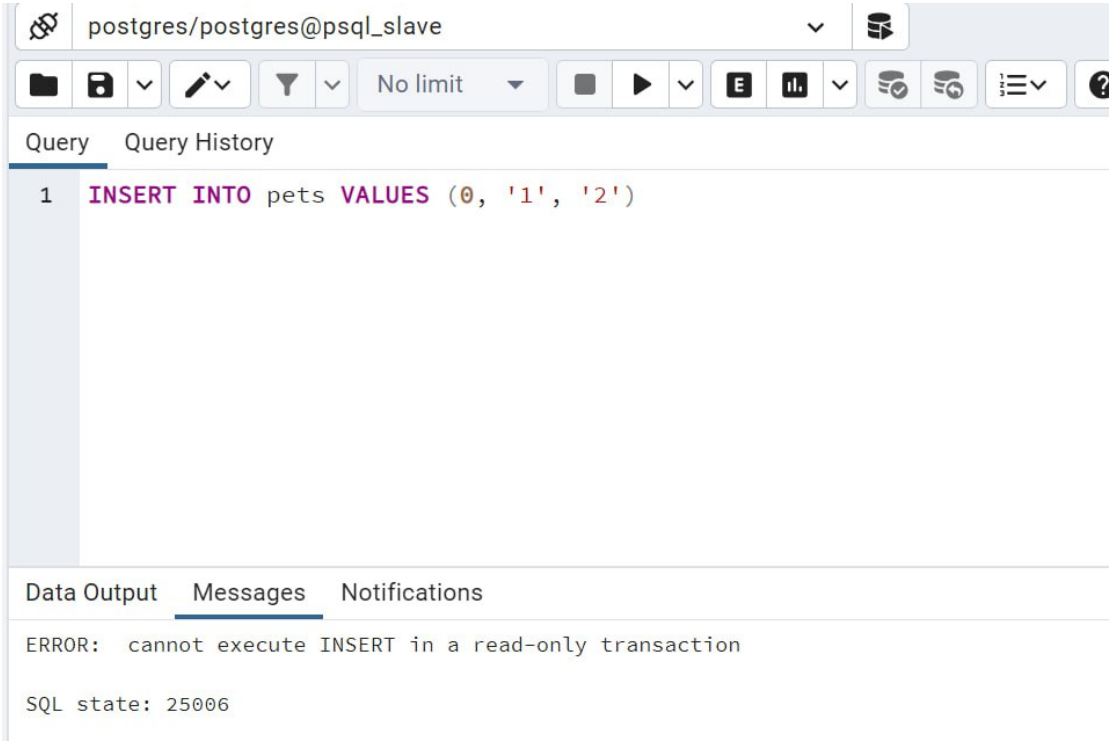
```
CREATE TABLE

Query returned successfully in 127 msec.
```

И видим, что она также появилась и на pg-slave

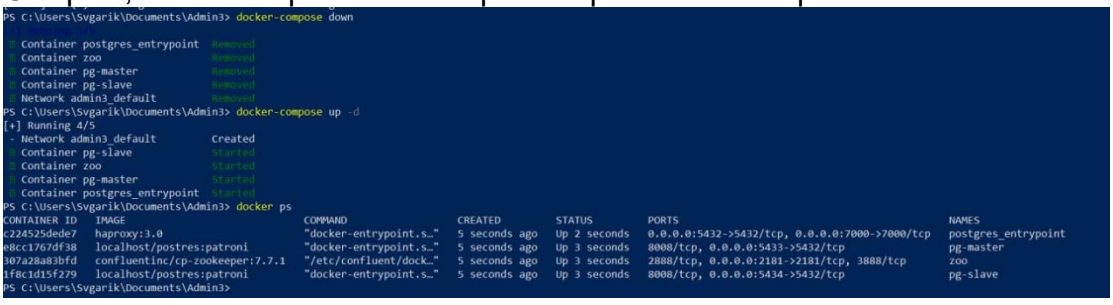


В slave мы не можем записать данные

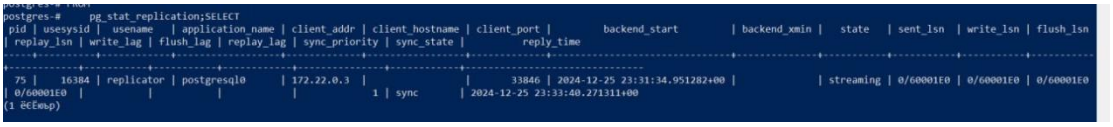


Часть 3 - ***Делаем ~~среднего роста~~ высокую доступность***

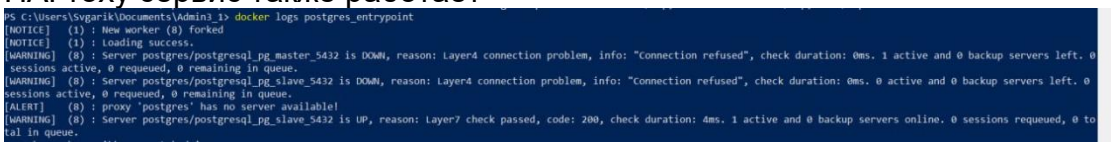
Дополним docker-compose.yml а также создадим haproxy.cfg
Теперь проверим работоспособность после перезапуска
Сморим, что мастер - это мастер и его реплика - это раб



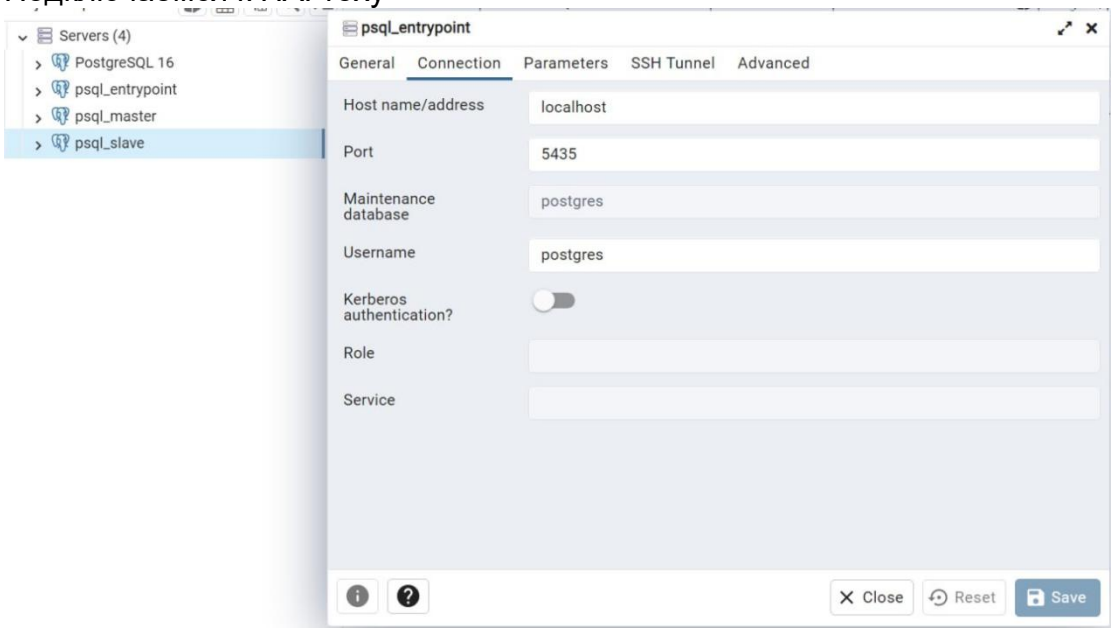
В текущем запуске оказалось наоборот



HAProxy сервис также работает



Подключаемся к HAProxy



Бонусом было проверено, что таблица созданная в psql_entrypoint создалась также и в двух остальных нодах

Задание

Сначала остановим текущего ведущего

```
PS C:\Users\Svgarik\Documents\Admin3_1> docker stop pg-slave
pg-slave
```

Также понятно, что мы можем изменить данные через entrypoint

```
PS C:\Users\Svgarik\Documents\Admin3_1> psql -h localhost -p 5435 -U postgres
Пароль пользователя postgres:
psql (16.0, сервер 15.10 (Debian 15.10-1.pgdg120+1))
ПРЕДУПРЕЖДЕНИЕ: Кодовая страница консоли (866) отличается от основной
                страницы Windows (1251).
                8-битовые (русские) символы могут отображаться некорректно.
                Подробнее об этом смотрите документацию psql, раздел
                "Notes for Windows users".
Введите "help", чтобы получить справку.

postgres=# INSERT INTO Characters VALUES (2, 'Inko', 'Roga')
postgres-# ;
INSERT 0 1
postgres=#
```

pg-master продолжил считать себя репликой, но ответа от pg-slave не было

```
2024-12-26 03:03:14 2024-12-26 00:03:14,729 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-26 03:03:23 2024-12-26 00:03:23.427 UTC [73] FATAL:  could not receive data from WAL stream: server closed the connection u
nexpectedly
2024-12-26 03:03:23            This probably means the server terminated abnormally
2024-12-26 03:03:23            before or while processing the request.
2024-12-26 03:03:23 2024-12-26 00:03:23.433 UTC [28] LOG:  invalid record length at 0/7000110: wanted 24, got 0
2024-12-26 03:03:23 2024-12-26 00:03:23.557 UTC [717] FATAL:  could not connect to the primary server: connection to server at "pg-
slave" (172.22.0.4), port 5432 failed: Connection refused
2024-12-26 03:03:23            Is the server running on that host and accepting TCP/IP connections?
2024-12-26 03:03:23 2024-12-26 00:03:23.558 UTC [28] LOG:  waiting for WAL to become available at 0/7000128
2024-12-26 03:03:24 2024-12-26 00:03:24,777 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-26 03:03:34 2024-12-26 00:03:34,732 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-26 03:03:36 2024-12-26 00:03:36.451 UTC [719] FATAL:  could not connect to the primary server: could not translate host nam
e "pg-slave" to address: Temporary failure in name resolution
```

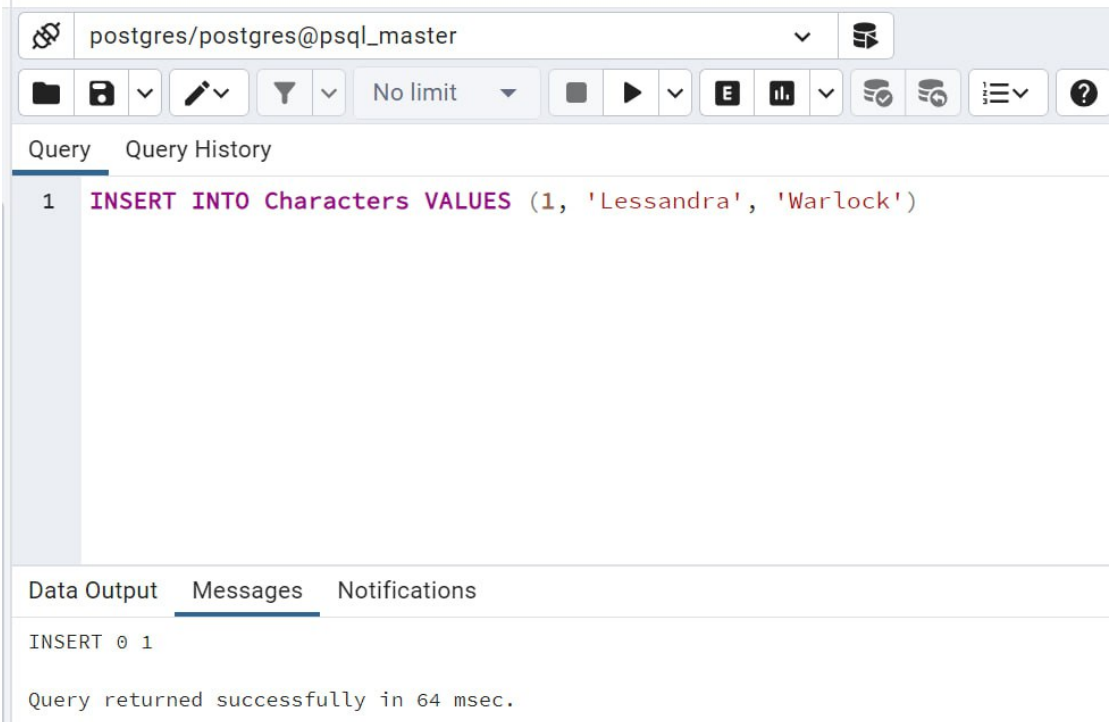
Через некоторое время pg-master стал лидером

```
2024-12-26 03:04:08 2024-12-26 00:04:08.530 UTC [741] FATAL:  could not connect to the primary server: could not translate host nam
e "pg-slave" to address: Temporary failure in name resolution
2024-12-26 03:04:08 2024-12-26 00:04:08.537 UTC [28] LOG:  redo done at 0/70000D8 system usage: CPU: user: 0.06 s, system: 0.07 s,
elapsed: 1964.49 s
2024-12-26 03:04:08 2024-12-26 00:04:08.537 UTC [28] LOG:  last completed transaction was at log time 2024-12-25 23:59:58.774314+00
2024-12-26 03:04:08 2024-12-26 00:04:08.548 UTC [28] LOG:  selected new timeline ID: 4
2024-12-26 03:04:08 2024-12-26 00:04:08.920 UTC [28] LOG:  archive recovery complete
2024-12-26 03:04:08 2024-12-26 00:04:08.952 UTC [26] LOG:  checkpoint starting: force
2024-12-26 03:04:08 2024-12-26 00:04:08.963 UTC [24] LOG:  database system is ready to accept connections
2024-12-26 03:04:09 2024-12-26 00:04:09.064 UTC [26] LOG:  checkpoint complete: wrote 2 buffers (0.0%); 0 WAL file(s) added, 0 remo
ved, 0 recycled; write=0.026 s, sync=0.008 s, total=0.112 s; sync files=2, longest=0.005 s, average=0.004 s; distance=16226 kB, est
imate=44633 kB
2024-12-26 03:04:09 2024-12-26 00:04:09,797 INFO: Lock owner: postgresql0; I am postgresql0
2024-12-26 03:04:09 2024-12-26 00:04:09,827 INFO: Reaped pid=766, exit status=0
2024-12-26 03:04:09 2024-12-26 00:04:09,850 INFO: no action. I am (postgresql0), the leader with the lock
2024-12-26 03:04:19 2024-12-26 00:04:19,789 INFO: Lock owner: postgresql0; I am postgresql0
2024-12-26 03:04:19 2024-12-26 00:04:19,808 INFO: Dropped unknown replication slot 'postgresql1'
2024-12-26 03:04:19 2024-12-26 00:04:19,817 INFO: no action. I am (postgresql0), the leader with the lock
2024-12-26 03:04:29 2024-12-26 00:04:29,830 INFO: no action. I am (postgresql0), the leader with the lock
2024-12-26 03:04:39 2024-12-26 00:04:39,791 INFO: no action. I am (postgresql0), the leader with the lock
```

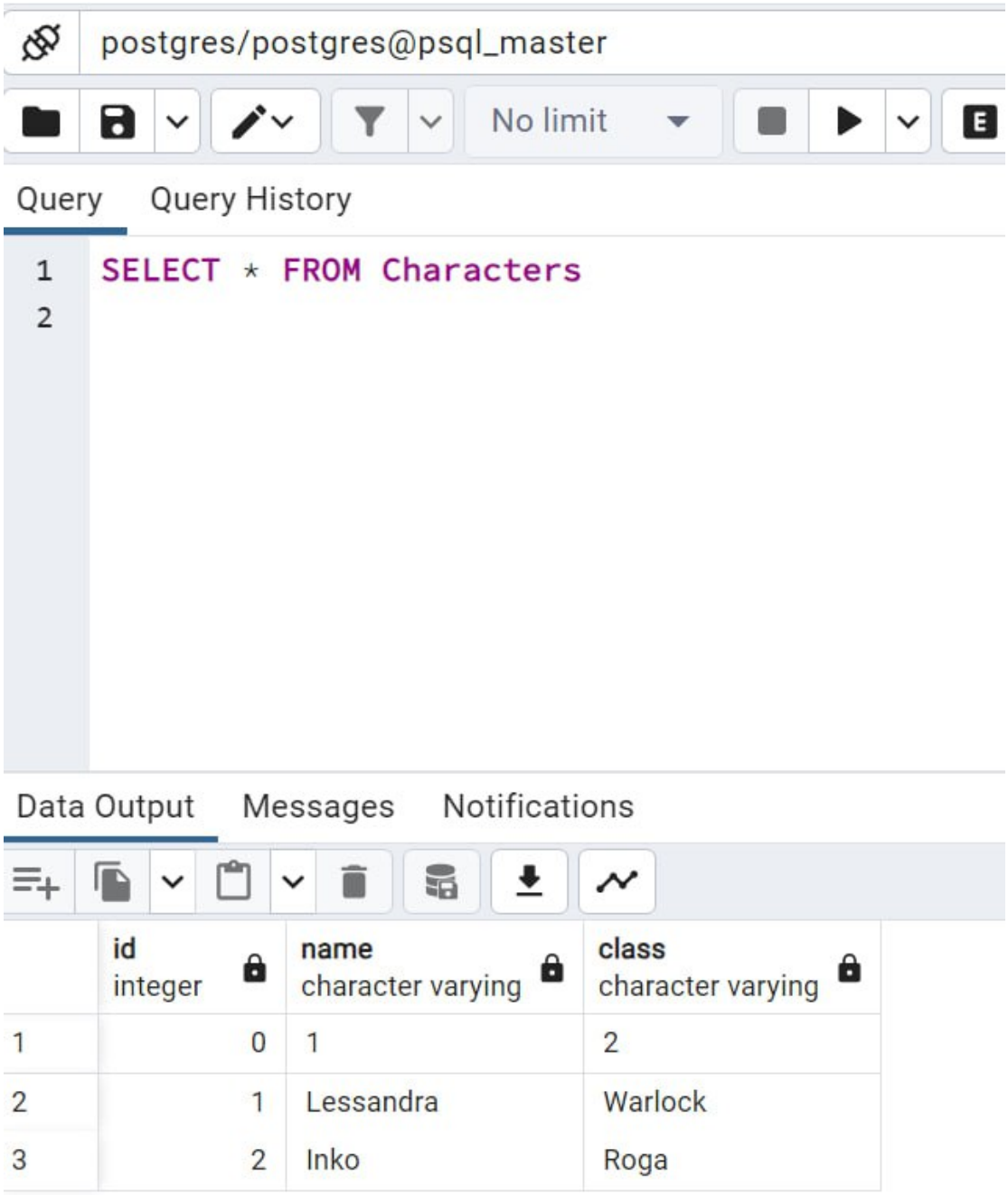
Также видим, что и в логах HAProxy зафиксировалось падение pg-slave и подъём pg-master

```
duration: 4ms. 1 active and 0 backup servers online. 0 sessions requeued, 0 total in queue.
2024-12-26 03:03:39 [WARNING]  (8) : Server postgres/postgresql_pg_slave_5432 is DOWN, reason: Layer4 timeout, check duration: 3004
ms. 0 active and 0 backup servers left. 0 sessions active, 0 requeued, 0 remaining in queue.
2024-12-26 03:03:39 [ALERT]    (8) : proxy 'postgres' has no server available!
2024-12-26 03:04:05 [WARNING]  (8) : Server postgres/postgresql_pg_master_5432 is UP, reason: Layer7 check passed, code: 200, check
 duration: 3ms. 1 active and 0 backup servers online. 0 sessions requeued, 0 total in queue.
```

В pg-master можем обновить таблицу



И изменения зафиксировались в pg-master



Ответы на вопросы:

- Порты 8008 и 5432 вынесены в разные директивы, expose и ports. По сути, если записать 8008 в ports, то он тоже станет exposed. В чем разница?

- expose позволяет контейнерам, в нашей лабе pg-master и pg-slave, взаимодействовать друг с другом внутри Docker-сети без необходимости публикации портов наружу, а ports открывает доступ извне, то есть в нашей лабе для подключения к PostgreSQL

- При обычном перезапуске композ-проекта, будет ли сбилден заново образ? А если предварительно отредактировать файлы postgresX.yml? А если содержимое самого Dockerfile? Почему?

При обычном перезапуске композ-проекта образ не будет сбилден заново, потому что композ использует уже существующие образы, которые были ранее собраны.
Если же отредактировать файлы postgresX.yml, образ также не будет сбилден заново, потому что такие файлы обычно монтируются в контейнер как внешние тома и не вызывают пересборку образа, а изменения будут применены при следующем запуске контейнера.
Если же отредактировать содержимое самого Dockerfile, образ будет сбилден заново только при использовании флага build (docker-compose up --build), так как Docker не отслеживает такие изменения автоматически.