

ЛР 3. HA Postgres Cluster

Часть 1. Поднимаем Postgres

Установим Docker и Docker Compose

```
root@daniil:~# docker-compose --version
Docker Compose version v2.20.3
root@daniil:~# docker --version
Docker version 24.0.7, build 24.0.7-0ubuntu2~22.04.1
```

Подготавливаем Dockerfile, postgres0.yml, postgres1.yml и запускаем docker-compose.yml

```
root@daniil:~/postgres_cluster# nano Dockerfile
root@daniil:~/postgres_cluster# nano postgres0.yml
root@daniil:~/postgres_cluster# nano postgres1.yml
root@daniil:~/postgres_cluster# ls
Dockerfile  postgres0.yml  postgres1.yml
root@daniil:~/postgres_cluster# nano docker-compose.yml
root@daniil:~/postgres_cluster# docker-compose up -d
Parsing /root/.docker/cli-plugins/docker-compose.yml: yaml: line 45: did not find expected key
root@daniil:~/postgres_cluster# nano docker-compose.yml
root@daniil:~/postgres_cluster# docker-compose up -d
[+] Running 16/16
! pg-sql Warning                                0.0s
! zoo 13 layers [#####]                        00/00    Pulled    20.4s
✓ 56f27190e824 Pull complete                    0.0s
✓ 8e70b9b0b078 Pull complete                   10.1s
✓ 732c9ebb730c Pull complete                    0.5s
✓ ed746366f1b8 Pull complete                    1.8s
✓ 10894799cc09 Pull complete                    1.3s
✓ 8d377259558c Pull complete                    4.4s
✓ e7688095d1e6 Pull complete                    4.1s
✓ 8eab815b3593 Pull complete                    4.8s
✓ 00de6dd259e Pull complete                    4.8s
✓ 290f622c8150 Pull complete                    5.2s
✓ 4ee3050c4ff6 Pull complete                    5.5s
✓ 98acab318002 Pull complete                    9.4s
✓ 878348106a95 Pull complete                    5.8s
! pg-master Warning                             0.0s
[+] Building 86.5s (14/16)                        docker:default

[+] Running 6/6
✓ Network postgres_cluster_default                Created    0.2s
✓ Volume "postgres_cluster_pg-master"            Created    0.0s
✓ Volume "postgres_cluster_pg-slave"             Created    0.0s
✓ Container pg-slave                             Started    0.0s
✓ Container zoo                                  Started    0.0s
✓ Container pg-master                            Started    0.0s
root@daniil:~/postgres_cluster# docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS                                NAMES
c70ed1479310   localhost/postgres:patroni         "docker-entrypoint.s..." About a minute ago Up About a minute 8080/tcp, 0.0.0.0:5434->5432/tcp, :::5434->5432/tcp pg-slave
b2d1d2caa772   localhost/postgres:patroni         "docker-entrypoint.s..." About a minute ago Up About a minute 8080/tcp, 0.0.0.0:5433->5432/tcp, :::5433->5432/tcp pg-master
b104adb1d996   confluentinc/cp-zookeeper:7.7.1   "/etc/confluent/dock..." About a minute ago Up About a minute 2888/tcp, 0.0.0.0:2181->2181/tcp, :::2181->2181/tcp, 3888/tcp zoo
```

Часть 2. Проверяем репликацию

Видим, что репликация работает, и в pg-slave нельзя вставить данные

```

root@daniil:~/postgres_cluster# psql -h 127.0.0.1 -p 5433 -U postgres -d postgres
Password for user postgres:
psql (14.15 (Ubuntu 14.15-0ubuntu0.22.04.1), server 15.10 (Debian 15.10-1.pgdg120+1))
WARNING: psql major version 14, server major version 15.
        Some psql features might not work.
Type "help" for help.

postgres=# CREATE TABLE test_table (id SERIAL PRIMARY KEY, data TEXT);
ERROR:  cannot execute CREATE TABLE in a read-only transaction
postgres=# SELECT * FROM test_table;
 id |      data
----+-----
  1 | test data 1
  2 | test data 2
(2 rows)

root@daniil:~/postgres_cluster# psql -h 127.0.0.1 -p 5434 -U postgres -d postgres
Password for user postgres:
psql (14.15 (Ubuntu 14.15-0ubuntu0.22.04.1), server 15.10 (Debian 15.10-1.pgdg120+1))
WARNING: psql major version 14, server major version 15.
        Some psql features might not work.
Type "help" for help.

postgres=# CREATE TABLE test_table (id SERIAL PRIMARY KEY, data TEXT);
CREATE TABLE
postgres=# INSERT INTO test_table (data) VALUES ('test data 1'), ('test data 2');
INSERT 0 2

```

Часть 3. Делаем высокую доступность

Добавляем haproxy.cfg

```

root@daniil:~/postgres_cluster# nano haproxy.cfg

```

Перезапускаем проект, проверяем что:

Обе ноды корректно поднялись и распределили между собой роли мастера и слейва

Postgres Cluster Зукипер подцепился к кластеру без ошибок

Хапрокси поднялась без ошибок

```

root@daniil:~/postgres_cluster# docker-compose down
[+] Removing 4/4
✔ Container pg-master      Removed    11.2s
✔ Container zoo            Removed    0.5s
✔ Container pg-slave       Removed    11.1s
✔ Network postgres_cluster_default Removed    0.3s

```

```

root@danili:~/postgres_cluster# docker-compose up -d
[+] Running 7/7
  ✓ haproxy 6 layers [████████] 00/00 Pulled
  ✓ bc0965b23a04 Already exists
  ✓ 39e67f4ddc51 Pull complete
  ✓ 42ad7db09397 Pull complete
  ✓ 0e85c8005b0e Pull complete
  ✓ 21ad4bb2a562 Pull complete
  ✓ 4f4fb700ef54 Pull complete
[+] Building 2.1s (14/16)
  => [pg-master internal] load build definition from Dockerfile
  => => transferring dockerfile: 880B
  => [pg-slave internal] load build definition from Dockerfile
  => => transferring dockerfile: 880B
  => [pg-slave internal] load metadata for docker.io/library/postgres:15
  => [pg-slave internal] load .dockerignore
  => => transferring context: 2B
  => [pg-master internal] load .dockerignore
  => => transferring context: 2B
  => [pg-master internal] load build context
  => => transferring context: 60B
  => [pg-slave 3/5] FROM docker.io/library/postgres:15@sha256:6cbb0d16e7ea1fa08873fd05203fd1410a8311bf4e4d62575f043f88fe
  => [pg-slave internal] load build context
  => => transferring context: 60B
  => CACHED [pg-master 5/5] COPY postgres1.yml /postgres1.yml
  => CACHED [pg-master 2/5] RUN apt-get update -y && apt-get install -y netcat-openbsd python3-pip curl python3-psycopg2 python3-venv iputils-ping
  => CACHED [pg-master 3/5] RUN python3 -m venv /opt/patroni-venv && /opt/patroni-venv/bin/pip install --upgrade pip && /opt/patroni-venv/bin/pip install patroni[zookeeper] psycopg2-binary
  => CACHED [pg-master 4/5] COPY postgres0.yml /postgres0.yml
  => [pg-master] exporting to image
  => => exporting layers
  => => writing image sha256:25ee7bd9d3c875e090a806f7ac262ed5e8da96c1915a0f1b95a064c4bd300
  => => naming to docker.io/library/postgres_cluster-pg-master
  => [pg-slave] exporting to image
  => => exporting layers
  => => writing image sha256:9fdacbb18d0db58da161eae2009aa1543aaf55a0abdab0c4d6cc0b52afcd73ef
  => => naming to docker.io/library/postgres_cluster-pg-slave
[+] Running 5/5
  ✓ Network postgres_cluster_default Created
  ✓ Container zoo Started
  ✓ Container pg-master Started
  ✓ Container pg-slave Started
  ✓ Container postgres_entrypoint Started

root@danili:~/postgres_cluster# docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS                                                                 NAMES
f413a6cf6768   haproxy:3.0                        "docker-entrypoint.s..." 52 seconds ago Up 50 seconds 0.0.0.0:5432->5432/tcp, :::5432->5432/tcp, 0.0.0.0:7000->7000/tcp, :::7000->7000/tcp postgres_entrypoint

5090d0c13a     confluentinc/cp-zookeeper:7.7.1   "/etc/confluent/docker..." 53 seconds ago Up 51 seconds 2888/tcp, 0.0.0.0:2181->2181/tcp, :::2181->2181/tcp, 3888/tcp zoo
9b157dbee38     postgres_cluster-pg-slave          "docker-entrypoint.s..." 53 seconds ago Up 51 seconds 8080/tcp, 0.0.0.0:5434->5432/tcp, :::5434->5432/tcp pg-slave
178d36ce5c11     postgres_cluster-pg-master        "docker-entrypoint.s..." 53 seconds ago Up 51 seconds 8080/tcp, 0.0.0.0:5433->5432/tcp, :::5433->5432/tcp pg-master

```

```

server signaled
2024-12-08 09:35:16.500 UTC [55] LOG:  fetching timeline history file for timeline 2 from primary server
2024-12-08 09:35:16.500 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-08 09:35:16.624 UTC [55] LOG:  started streaming WAL from primary at 0/5000000 on timeline 1
2024-12-08 09:35:16.625 UTC [55] LOG:  replication terminated by primary server
2024-12-08 09:35:16.625 UTC [55] DETAIL:  End of WAL reached on timeline 1 at 0/50000A0.
2024-12-08 09:35:16.626 UTC [55] FATAL:  terminating walreceiver process due to administrator command
2024-12-08 09:35:16.626 UTC [28] LOG:  new target timeline is 2
2024-12-08 09:35:16.627 UTC [28] LOG:  invalid record length at 0/50000A0: wanted 24, got 0
2024-12-08 09:35:16.650 UTC [61] LOG:  started streaming WAL from primary at 0/5000000 on timeline 2
2024-12-08 09:35:26.386 INFO: Lock owner: postgresql1; I am postgresql0
2024-12-08 09:35:26.391 INFO: Local timeline=2 lsn=0/50001E0
2024-12-08 09:35:26.419 INFO: primary timeline=2
2024-12-08 09:35:26.430 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-08 09:35:36.387 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-08 09:35:46.386 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-08 09:35:56.387 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-08 09:36:06.386 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-08 09:36:16.387 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
2024-12-08 09:36:26.385 INFO: no action. I am (postgresql0), a secondary, and following a leader (postgresql1)
[2024-12-08 09:35:03,023] INFO Created server with tickTime 2000 ms minSessionTimeout 4000 ms maxSessionTimeout 40000 ms ClientCnxnSocketImplFactory org.eclipse.jetty.io.ssl.JettySSLContextFactory
[2024-12-08 09:35:03,023] INFO org.apache.zookeeper.server.ZooKeeperServer
[2024-12-08 09:35:03,654] INFO Logging initialized @674ms to org.eclipse.jetty.util.log.Slf4jLog (org.eclipse.jetty.util.log)
[2024-12-08 09:35:03,739] WARN o.e.j.s.ServletContextHandler@229f66ed{/null,STOPPED} contextPath ends with /* (org.eclipse.jetty.server.handler.ContextHandler)
[2024-12-08 09:35:03,739] WARN Empty contextPath (org.eclipse.jetty.server.handler.ContextHandler)
[2024-12-08 09:35:03,764] INFO jetty-9.4.54.v20240208; built: 2024-02-08T19:42:39.027Z; git: cef3fbd6d736a21e7d541a5db490381d95a2047d; jvm 17.0.12+7-LTS (org.eclipse.jetty.server)
[2024-12-08 09:35:03,822] INFO DefaultSessionIdManager workerName=node0 (org.eclipse.jetty.server.session)
[2024-12-08 09:35:03,822] INFO No SessionScavenger set, using defaults (org.eclipse.jetty.server.session)
[2024-12-08 09:35:03,824] INFO node0 Scavenging every 600000ms (org.eclipse.jetty.server.session)
[2024-12-08 09:35:03,829] WARN ServletContext@o.e.j.s.ServletContextHandler@229f66ed{/null,STARTING} has uncovered http methods for path: /* (org.eclipse.jetty.server)
[2024-12-08 09:35:03,842] INFO Started o.e.j.s.ServletContextHandler@229f66ed{/null,AVAILABLE} (org.eclipse.jetty.server.handler.ContextHandler)
[2024-12-08 09:35:03,856] INFO Started ServerConnector@564fabcb(HTTP/1.1, (http/1.1)){0.0.0.0:8080} (org.eclipse.jetty.server.AbstractConnector)
[2024-12-08 09:35:03,857] INFO Started @882ms (org.eclipse.jetty.server.Server)
[2024-12-08 09:35:03,857] INFO Started AdminServer on address 0.0.0.0, port 8080 and command URL /commands (org.apache.zookeeper.server.admin.JettyAdminServer)
[2024-12-08 09:35:03,862] INFO Using org.apache.zookeeper.server.NIOServerCnxnFactory as server connection factory (org.apache.zookeeper.server.ServerCnxnFactory)
[2024-12-08 09:35:03,864] WARN maxCnxns is not configured, using default value 0. (org.apache.zookeeper.server.ServerCnxnFactory)
[2024-12-08 09:35:03,865] INFO Configuring NIO connection handler with 10s sessionless connection timeout, 1 selector thread(s), 4 worker threads, and 64 kbytes (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2024-12-08 09:35:03,867] INFO binding to port 0.0.0.0/0.0.0.0:2181 (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2024-12-08 09:35:03,886] INFO Using org.apache.zookeeper.server.watch.WatchManager as watch manager (org.apache.zookeeper.server.watch.WatchManagerFactory)
[2024-12-08 09:35:03,887] INFO Using org.apache.zookeeper.server.watch.WatchManager as watch manager (org.apache.zookeeper.server.watch.WatchManagerFactory)
[2024-12-08 09:35:03,887] INFO zookeeper.snapshotSizeFactor = 0.33 (org.apache.zookeeper.server.ZKDatabase)
[2024-12-08 09:35:03,887] INFO zookeeper.commitLogCount=500 (org.apache.zookeeper.server.ZKDatabase)
[2024-12-08 09:35:03,894] INFO zookeeper.snapshot.compression.method = CHECKED (org.apache.zookeeper.server.persistence.SnapStream)
[2024-12-08 09:35:03,894] INFO Snapshotting: 0x0 to /var/lib/zookeeper/data/version-2/snapshot.0 (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-12-08 09:35:03,898] INFO Snapshot loaded in 11 ms, highest zxid is 0x0, digest is 1371985504 (org.apache.zookeeper.server.ZKDatabase)
[2024-12-08 09:35:03,899] INFO Snapshotting: 0x0 to /var/lib/zookeeper/data/version-2/snapshot.0 (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-12-08 09:35:03,900] INFO Snapshot taken in 1 ms (org.apache.zookeeper.server.ZooKeeperServer)
[2024-12-08 09:35:03,915] INFO zookeeper.request_throttler.shutdownTimeout = 10000 ms (org.apache.zookeeper.server.RequestThrottler)
[2024-12-08 09:35:03,916] INFO PrepRequestProcessor (sid:0) started, reconfigEnabled=false (org.apache.zookeeper.server.PrepareRequestProcessor)
[2024-12-08 09:35:03,933] INFO Using checkIntervalMs=60000 maxPerMinute=10000 maxNeverUsedIntervalMs=0 (org.apache.zookeeper.server.ContainerManager)
[2024-12-08 09:35:03,934] INFO ZooKeeper audit is disabled. (org.apache.zookeeper.audit.ZKAuditProvider)

```

```

root@daniil:~/postgres_cluster# psql -h localhost -p 5432 -U postgres -d postgres
Password for user postgres:
psql (14.15 (Ubuntu 14.15-0ubuntu0.22.04.1), server 15.10 (Debian 15.10-1.pgdg120+1))
WARNING: psql major version 14, server major version 15.
        Some psql features might not work.
Type "help" for help.

postgres=# SELECT pg_is_in_recovery();
 pg_is_in_recovery 
-----
 f
(1 row)

```

Задание

Любым способом выключаем доступ до ноды, которая сейчас является мастером (например, через `docker stop`). Некоторое время ждем, после этого анализируем логи и так же пытаемся считать/записать что-то в БД через `entrypoint` подключение. Затем необходимо расписать, получилось или нет, а так же объяснить, что в итоге произошло после принудительного выключения мастера (со скриншотами)

Останавливаем мастер-ноду

```

root@daniil:~/postgres_cluster# docker stop pg-master
pg-master

```

```

2024-12-08 09:35:05,199 INFO: establishing a new patroni heartbeat connection to postgres
2024-12-08 09:35:05,292 INFO: establishing a new patroni restapi connection to postgres
2024-12-08 09:35:06,866 INFO: Got response from postgresql0 http://pg-master:8008/patroni: {"state": "running", "postmaster_start_time": "2024-12-08 09:35:05.319735+00:00", "role": "replica", "server_version": 150010, "xlog": {"received_location": 83886080, "replayed_location": 83886080, "replayed_timestamp": null, "paused": false}, "timeline": 1, "cluster_unlocked": true, "dcs_last_seen": 1733658506, "database_identifier": "7445960677449695256", "patroni": {"version": "4.0.4", "scope": "my_cluster", "name": "postgresql0"}}
2024-12-08 09:35:06,901 INFO: promoted self to leader by acquiring session lock
2024-12-08 09:35:06,903 UTC [28] LOG: received promote request
2024-12-08 09:35:06,903 UTC [28] LOG: redo is not required
server promoting
2024-12-08 09:35:06,911 UTC [28] LOG: selected new timeline ID: 2
2024-12-08 09:35:06,986 UTC [28] LOG: archive recovery complete
2024-12-08 09:35:07,002 UTC [26] LOG: checkpoint starting: force
2024-12-08 09:35:07,005 UTC [24] LOG: database system is ready to accept connections
2024-12-08 09:35:07,062 UTC [26] LOG: checkpoint complete: wrote 3 buffers (0.0%); 0 WAL file(s) added, 0 removed, 0 recycled; write=0.026 s, sync=0.006 s, total=0.061 s; sync files=2, longest=0.004 s, average=0.003 s; distance=0 kB, estimate=0 kB
2024-12-08 09:35:07,069 INFO: Reaped pid=57, exit status=0
2024-12-08 09:35:07,984 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-08 09:35:17,920 INFO: Lock owner: postgresql1; I am postgresql1
2024-12-08 09:35:17,926 INFO: Assigning synchronous standby status to ['postgresql0']
2024-12-08 09:35:17,928 UTC [24] LOG: received SIGHUP, reloading configuration files
2024-12-08 09:35:17,929 UTC [24] LOG: parameter "synchronous_standby_names" changed to "postgresql0"
server signaled
2024-12-08 09:35:18,040 UTC [64] LOG: standby "postgresql0" is now a synchronous standby with priority 1
2024-12-08 09:35:18,040 UTC [64] STATEMENT: START REPLICATION SLOT "postgresql0" 0/50000000 TIMELINE 2
2024-12-08 09:35:20,048 INFO: Synchronous standby status assigned to ['postgresql0']
2024-12-08 09:35:20,055 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-08 09:35:27,929 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-08 09:35:27,923 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-08 09:35:47,920 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-08 09:35:57,925 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-08 09:36:07,919 INFO: no action. I am (postgresql1), the leader with the lock
2024-12-08 09:36:17,919 INFO: no action. I am (postgresql1), the leader with the lock

```

Видим, что мы не можем связаться с pg-master

```

root@daniil:~/postgres_cluster# docker logs postgres_entrypoint
[NOTICE] (1) : New worker (0) forked
[NOTICE] (1) : Loading success.
[WARNING] (8) : Server postgres/postgresql_pg_master_5432 is DOWN, reason: Layer4 connection problem, info: "Connection refused", check duration: 0ms. 1 active and 0 backup servers left. 0 sessions active, 0 requested, 0 remaining in queue.
[WARNING] (8) : Server postgres/postgresql_pg_slave_5432 is DOWN, reason: Layer4 connection problem, info: "Connection refused", check duration: 0ms. 0 active and 0 backup servers left. 0 sessions active, 0 requested, 0 remaining in queue.
[ALERT] (0) : proxy 'postgres' has no server available!
[WARNING] (8) : Server postgres/postgresql_pg_slave_5432 is UP, reason: Layer7 check passed, code: 200, check duration: 3ms. 1 active and 0 backup servers online. 0 sessions requested, 0 total in queue.

```

HAПрoxy переключился на новую ноду

```

[WARNING] (8) : Server postgres/postgresql_pg_slave_5432 is UP, reason: Layer7 check passed, code: 200,

```


Подключились к новому мастеру и сделали запись:

```
root@daniil:~/postgres_cluster# psql -h localhost -p 5432 -U postgres -d postgres
postgres=# SELECT pg_is_in_recovery();
 pg_is_in_recovery
-----
 f
(1 row)

postgres=# CREATE TABLE test_table (id SERIAL PRIMARY KEY, data TEXT);
ERROR:  relation "test_table" already exists
postgres=# ERROR:  relation "test_table" already existsERROR:  relation "test_table" already exists
postgres=# CREATE TABLE test_table2 (id SERIAL PRIMARY KEY, data TEXT);
ERROR:  syntax error at or near "ERROR"
LINE 1: ERROR:  relation "test_table" already existsERROR:  relation...
          ^
postgres=#
postgres=# CREATE TABLE test_tab (id SERIAL PRIMARY KEY, data TEXT);
CREATE TABLE
postgres=# INSERT INTO test_tab (data) VALUES ('test data 1'), ('test data 2');
INSERT 0 2
```

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?

ports — порт на контейнере открывается для внешнего мира через хост. Необходимо для приложений и сервисов, которые требуют внешнего доступа.

expose — порт на контейнере доступен исключительно для других контейнеров в рамках сети Docker. Не предназначен для подключения извне.

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

При стандартном запуске образы не будут пересобраны. Если внести изменения в файл postgresX.yml, контейнеры перезапустятся с обновленными настройками, но образы сохранят свою прежнюю версию. Если изменить Dockerfile, чтобы применить изменения, нужно использовать флаг --build, иначе будут использованы старые образы.