

ИИ платформа «Виртуальный ассистент»:  
руководство по установке сервера

## 1. Системные требования

Требования к аппаратному и программному обеспечению для установки комплекта серверной части программного комплекса «Виртуальный Ассистент» в нераспределенную систему (обособленный компьютер либо виртуальный сервер в облачном сервисе):

- Центральный процессор: Intel Xeon E5 v4 (16 и более ядер)
- ОЗУ: 64 Гб и более
- ПЗУ: SSD либо HDD объемом 256 Гб и более
- Графический адаптер: дискретная видеокарта на основе чипсета GPU Nvidia поколения Pascal с архитектурой SM\_60 (и новее), поддерживающей набор инструкций CUDA 11 с объемом графического ОЗУ 6 Гб и более. Например: GeForce GTX 1060, RTX 2080, Tesla P4, Tesla T4 и т.д. Подробнее: см.  
<https://arnon.dk/matching-sm-architectures-arch-and-gencode-for-various-nvidia-cards/>
- Операционная система: Ubuntu Linux 18.04/20.04 LTS 64-bit

При установке в распределенную систему мощность каждого узла системы подбирается индивидуально, исходя из количества сервисов NLP (Natural Language Processing), STT (Speech-To-Text) и TTS (Text-To-Speech) необходимых для поддержки производительности (макс. количество одновременных запросов) системы заданной заказчиком

## 2. Предварительное конфигурирование операционной системы

При установке в распределенную систему произведите нижеуказанные действия для каждого узла кластера:

- Создайте пользователя с правами sudo в системе
- Установите сервис Docker: `$ sudo apt-get install docker.io`
- Установите сервис docker-compose: `$ sudo apt-get install docker-compose`
- Установите пакет 7zip: `$ sudo apt-get install p7zip-full`
- Установите комплект драйверов Nvidia, соответствующих CUDA 11.4 (или новее- см. соответствие версии пакета драйверов версии CUDA на <https://www.nvidia.com/download/index.aspx?lang=en-us>): `$ sudo apt-get install nvidia-driver-470` (только для узлов с графической картой)
- Перезагрузите машину после чего проверьте корректность установки драйверов командой `$nvidia-smi` (только для узлов с графической картой)
- Установите пакет nvidia-container-runtime из официального репозитория Nvidia (только для узлов с графической картой):
  - Установите текущее значение переменной окружения distribution:  
`$ sudo distribution=$(. /etc/os-release;echo $ID$VERSION_ID)`
  - Зарегистрируйте ключи официальных репозитория Nvidia в системе:  
`$ sudo curl -s -L https://nvidia.github.io/nvidia-docker/gpgkey | apt-key add`
  - Добавьте репозитории Nvidia в список локальных репозитория системы:  
`$ sudo curl -s -L https://nvidia.github.io/nvidia-docker/$distribution/nvidia-docker.list | sudo tee /etc/apt/sources.list.d/nvidia-docker.list`

- Обновите локальный список доступных версий пакетов:  
\$ sudo apt-get update
- Запустите установку пакета nvidia-container-runtime с зависимостями:  
\$ sudo apt-get install nvidia-container-runtime
- Перезагрузите машину

### 3. Создание сетевого хранилища общего доступа (NFS shared storage)

В случае установки серверного комплекта системы «Виртуальный Ассистент» в распределенную систему необходимо выбрать узел в системе используемый как сетевое хранилище стека и выполнить следующую процедуру установки сетевого хранилища:

- На узле, используемом как сетевое хранилище:

- Создайте папки сетевого хранилища:  
\$ sudo mkdir -p /data/nfs\_share/internal  
\$ sudo mkdir -p /data/nfs\_share/external  
\$ sudo mkdir -p /mount/internal  
\$ sudo mkdir -p /mount/external
- Запустите скрипт create\_stack\_mount\_dirs.sh из комплекта поставки системы Виртуальный Ассистент для создания системы папок, используемых сервисами стека
- Выполните маппинг папок сетевого хранилища на общие папки стека:  
Добавьте следующие строки в файл /etc/fstab:  
/data/nfs\_share/internal /mount/internal none defaults,bind 0 0  
/data/nfs\_share/external /mount/external none defaults,bind 0 0
- Установите пакет NFS server:  
\$ sudo apt-get install nfs-kernel-server
- Сконфигурируйте удаленный доступ к общим папкам хранилища стека:  
\$ sudo nano /etc/exports  
Добавьте следующие строки в вышеуказанный файл:  
/data/nfs\_share/internal <sub\_net> (rw,sync,no\_root\_squash,no\_subtree\_check)  
/data/nfs\_share/external <sub\_net> (rw,sync,no\_root\_squash,no\_subtree\_check)  
Здесь <sub\_net> - маска пространства IP-адресов узлов, входящих в кластер- напр. 192.168.0.0/16
- Примените вышеуказанные изменения в системе:  
\$ sudo exportfs -a  
\$ sudo systemctl restart nfs-kernel-server  
\$ sudo mount -a
- Запомните IP-адрес NFS сервера (NFS\_server\_IP)

- На остальных узлах кластера, используемых как клиенты для доступа к общему сетевому хранилищу стека:

- Создайте папки сетевого хранилища:  
\$ sudo mkdir -p /mount/internal  
\$ sudo mkdir -p /mount/external
- Установите пакет NFS Common:  
\$ sudo apt-get install nfs-common
- Добавьте следующие строки в файл /etc/fstab:  
<NFS\_server\_IP>:/data/nfs\_share/internal /mount/internal nfs rw,soft,intr,noatime,x-gvfs-show  
<NFS\_server\_IP>:/data/nfs\_share/external /mount/external nfs rw,soft,intr,noatime,x-gvfs-show
- Примените вышеуказанные изменения точек монтирования:

```
$ sudo mount -a
```

#### 4. Инициализация и настройка docker swarm

- Инициализируйте docker swarm на узле, выбранном как docker swarm manager: `$ docker swarm init`

- На узле, инициализированном как docker swarm manager получите команду для добавления дополнительных worker- узлов в swarm:

```
svc_va@ubuntu-va-02:~$ docker swarm join-token worker
```

To add a worker to this swarm, run the following command:

```
docker swarm join --token
SWMTKN-1-1x0evj19n4lcmw54xc70bwjz7b5ci3xxa5nu9pruw1ejcugd72-75qityad1yx3p45mz0umfwgfj
192.168.4.42:2377
```

- На узлах, которые должны быть использованы как worker выполните команду, полученную в вышеуказанном ответе на узле swarm manager- напр.:

```
svc_va@ubuntu-va-03: $ docker swarm join -token
SWMTKN-1-1x0evj19n4lcmw54xc70bwjz7b5ci3xxa5nu9pruw1ejcugd72-75qityad1yx3p45mz0umfwgfj
fj 192.168.4.42:2377
```

- В случае необходимости добавить дополнительные узлы как swarm manager получите ключ для добавления командой `$ docker swarm join-token manager` на одном из действующих менеджеров сворма

- Проверьте корректность сборки сворма командой `$ docker node ls`, запущенной с одного из менеджеров сворма, результат команды должен содержать в себе список узлов кластера и их текущий статус, напр.:

```
svc_va@ubuntu-va-02:~$ docker node ls
ID                HOSTNAME          STATUS    AVAILABILITY  MANAGER STATUS  ENGINE VERSION
ee0yhjr89skl17nziz3yv9orv  ubuntu-va-01    Ready    Active
xhmludugw5jdg1d405q26gp1m *  ubuntu-va-02    Ready    Active         Leader          20.10.7
6efdun4t000f0o3lwfgiqf8jx   ubuntu-va-03    Ready    Active          20.10.12
```

- Сконфигурируйте сервис докеров на каждом узле стека:

- Создайте файл `/etc/docker/daemon.json` на каждом узле стека в том случае если вышеуказанный файл отсутствует
- Добавьте следующие строки в файл `/etc/docker/daemon.json` для каждого узла с GPU:

```
{
  "dns": ["4.4.4.4", "8.8.8.8"],
  "max-concurrent-downloads": 10,
  "max-concurrent-uploads": 10,
  "log-driver": "json-file",
  "log-opts": {
    "max-size": "100m",
    "max-file": "5"
  },
  "runtimes": {
    "nvidia": {
      "path": "/usr/bin/nvidia-container-runtime",
      "default-runtime": "nvidia",
      "runtimeArgs": []
    }
  }
},
```

- ```

    "default-runtime": "nvidia"
  }

```
- Добавьте следующие строки в файл `/etc/docker/daemon.json` для каждого узла без GPU:

```

{
  "dns": ["4.4.4.4", "8.8.8.8"],
  "max-concurrent-downloads": 10,
  "max-concurrent-uploads": 10,
  "log-driver": "json-file",
  "log-opts": {
    "max-size": "100m",
    "max-file": "5"
  }
}

```
  - Перезапустите сервис докеров после внесения изменений в файл `/etc/docker/daemon.json`:

```

$ sudo systemctl restart docker

```

## 5. Созданий файлов сертификатов

Создайте приватный ключ и файл корневого сертификата `ca.key`, `ca.crt` либо воспользуйтесь существующими файлами корневого сертификата вашей организации.

Скопируйте файлы `ca.key` и `ca.crt` в папку `certs_unoto` из комплекта поставки.

Создайте файл `ca.pem`:

```
$ grep -A 9999 "BEGIN CERT" ca.crt > ca.pem
```

```
$ cat ca.key >> ca.pem
```

В папке `certs_unoto` создайте копии файлов `ca.key`, `ca.crt` и `ca.pem` под именами `server_auth.key`, `server_auth.crt`, `server_auth.pem` и `server_encrypt.key`, `server_encrypt.crt`, `server_encrypt.pem` соответственно.

Создайте клиентский сертификат (`.crt`) и сертификат панели администратора (`.p12`) для клиента (CN) "unoto" при помощи скрипта создания клиентских сертификатов `certs_unoto/cn.sh` из комплекта поставки.

## 6. Установка и запуск стека серверных сервисов Виртуальный Ассистент

- Скопируйте архив (напр. `va_install.7z`), полученный от поставщика системы и содержащий установочные файлы (файл архива образов `.dockers`, файл деплоймента стека `.yml`, скрипт создания точек монтирования докеров и др.) на все узлы кластера, созданного выше

- Извлеките из архива файлы, содержащие установочные модули системы на каждом узле кластера:

```
$ 7z x va_install.7z
```

- Добавьте образы компонент системы в локальный `docker` репозиторий каждого узла кластера:

```
$ docker load -i va_install.dockers
```

- Скопируйте папку `certs_unoto`, содержащую файлы сертификатов в сетевое хранилище общего доступа кластера (`/mount/internal/service/`)

- Разверните стек сервисов голосового ассистента, используя файл деплоя из комплекта поставки: `$ docker stack deploy -c docker-compose.yml va_server_demo va_server_demo`

- Подождите 15-45 минут (время старта стека сервера Голосового Ассистента может значительно отличаться в зависимости от производительности узлов кластера, стабильности и скорости сетевого соединения между узлами кластера, количества поддерживаемых языков и тонких настроек ключевых сервисов системы (NLP, STT, TTS))

- Проверьте успешность деплоя стека:

- Запустите следующую команду с любой Linux-машины (либо Bash- эмулятора- напр. Putty, Git Bash итд.), имеющей доступ к любому из узлов кластера, в котором развернут сервер:  

```
$ curl -s -w '%{http_code}' -o /dev/null -v --cert /certs_unoto/unoto.crt --key /certs_unoto/unoto.key -k -H Content-Type:audio/vnd.wave -X POST --data-binary @test_va.wav -H 'ACCESS-TOKEN: 1' 'https://<server_ip>:<port_id>/put_audio?device_id=1&with_debug=True&skip_tts=true'
```

/certs\_unoto - папка, содержащая файлы сертификатов из комплекта поставки  
test\_va.wav - тестовый WAV файл из комплекта поставки, может быть заменён любым WAV файлом  
<server\_ip>: IP-адрес любого из узлов кластера  
<port\_id>: внешний порт стека, данные с которого перенаправлены на внутренний порт 443 сервиса LoadBalancer, принимающий сообщения от клиентов работающих по протоколу HTTP (по умолчанию- порт 8033)
- Пример ожидаемого ответа при готовности сервера:

```
< HTTP/2 200
< server: nginx
< date: Wed, 15 Feb 2023 13:00:05 GMT
< content-type: audio/vnd.wave
< content-length: 0
< request_id: 15314
< error: '{"error_code": 0}'
< language: en
< question_text: 'Hi, good to see you. How about we do any intensity training today?'
< answer_text: 'Sorry, I did not understand. Could you please repeat?'
```
- В случае если код ответа сервера отличен от 200 (в большинстве случаев при неготовности система вернёт код ответа 502 Bad Gateway) воспользуйтесь средствами диагностики, позволяющими отслеживать статус docker- сервисов и просматривать логи docker- контейнеров (напр. Portainer – см. ниже)

## 7. Диагностика и поиск неисправностей

### 7.1 Базовая диагностика сервисов с помощью командной строки

- На любом docker manager- узле кластера запустите следующую команду `$ docker service ls |grep <stack_name>` (здесь <stack\_name>- имя docker- стека, в котором развернуто программное решение сервера Голосового Ассистента – напр., va\_server\_demo)

```

svc_va@ubuntu-va-02:~$ docker service ls |grep va_server_test
n6x1aeu5aywi va_server_test_admintool replicated 1/1 va/admintool:DE_improve_save_74
pc27u1m91jko va_server_test_controller replicated 1/1 va/controller:MLVA-443_13022023
d60k2i5quw52 va_server_test_dbbackup replicated 1/1 va/dbbackup:MLVA-436_08022023
fr6nz8tkgvmo va_server_test_etcd replicated 1/1 va/etcd:v3.3.25
do7o8bognuyb va_server_test_extsrv replicated 1/1 va/extsrv:r7
uk1qo51c3xaw va_server_test_loadbalancer replicated 1/1 va/loadbalancer:balancer_unoto_cert
>649/tcp, *:7896->446/tcp, *:8006->447/tcp, *:8016->643/tcp, *:8026->445/tcp, *:8036->443/tcp, *:8076->648/tcp, *:8086->645/tcp
urpn5t1a93ft va_server_test_myadmin replicated 1/1 va/phpmyadmin:va-bundle-4.2236.02
xt9sym5xc51l va_server_test_nlp0 replicated 1/1 va/nlp:va_23012023_meeting_report_stt_marks_unoto
uestale9g9tk va_server_test_percona-xtradb-cluster replicated 2/2 va/percona_cluster:va-bundle-4.2236.02
0014b7vviccm va_server_test_proxy replicated 1/1 va/percona_proxysql:percona_for_sip
>3306/tcp
37cgd03ps29q va_server_test_reporter replicated 1/1 va/reporter:olap_update-04
wwojzmcz2wx va_server_test_sip_client replicated 1/1 va/sip_client:MLVA-443_13022023
>22/tcp
izcyjckg8p9v va_server_test_stt0 replicated 1/1 va/sttv2:221222_unoto
nmgx2wxbffmj va_server_test_stt1 replicated 1/1 va/sttv2:va-bundle-4.2239.03_ftn
p71oy718hjqb va_server_test_stt2 replicated 1/1 va/stt:130123_unoto
s35ylyrd7p9b va_server_test_stt_aws replicated 1/1 va/stt_aws:va-bundle-4.2236.02
69tbtz1w2jx va_server_test_stt_skilltool replicated 1/1 va/stt_skilltool:latest_081122
oqk1me1n8gn4 va_server_test_sysmon global 3/3 va/sysmon:va-bundle-4.2236.02
cfobqbi40f2o va_server_test_tts0 replicated 1/1 va/tts:MLVA-413_01022023
5p5kumg6ynaz va_server_test_tts_engine replicated 1/1 va/tts_engine_sova:MLVA-413_06022023
qg5qv71lo1bb va_server_test_ws_manager replicated 1/1 va/ws_manager:customer_migration
svc_va@ubuntu-va-02:~$

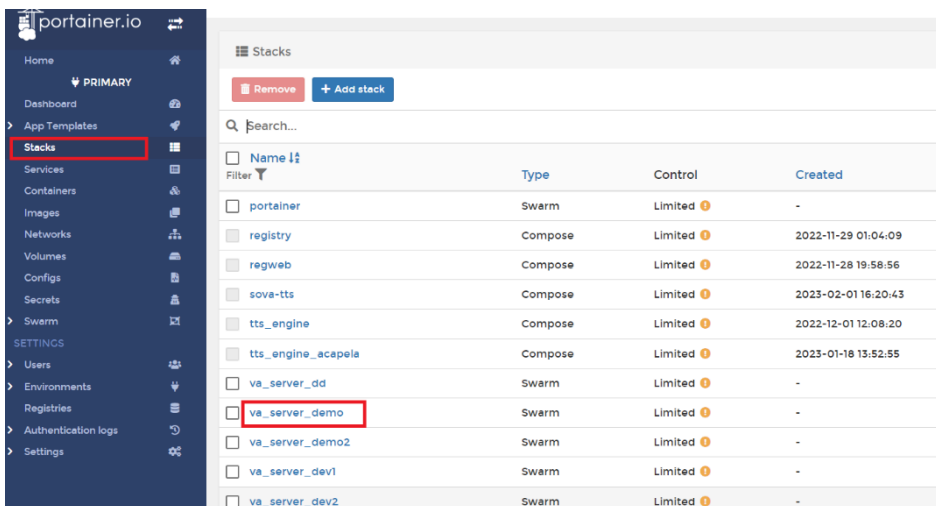
```

- Основные сервисы стека (ws\_manager, controller, etcd, percona-xtradb-cluster, proxy, loadbalancer, nlp, stt, tts, tts\_engine) должны быть активированы и запущены (1/1, 2/2 итд в столбце REPLICAS)

## 7.2 Диагностика сервисов с помощью программного комплекса Portainer

Portainer ([portainer.io](https://portainer.io)) – программный комплекс диагностики и управления контейнерами. Может быть установлен как отдельно из репозитория docker.io, так и включен в комплект поставки серверной платформы Виртуальный Ассистент.

- Загрузите основную страницу Portainer через браузер: [https://<server\\_ip>:9000](https://<server_ip>:9000) (с машины, имеющей доступ к кластеру через порт 9000 (может заменён на другой при разворачивании сервисов Portainer))
- Выберите необходимый стек из списка стеков:



- Выберите необходимый сервис для диагностики состояния контейнеров:

Stack details

va\_server\_demo

| Services                  |                                             |                  |         |
|---------------------------|---------------------------------------------|------------------|---------|
| Update Remove             |                                             |                  |         |
| Q Search...               |                                             |                  |         |
| Name                      | Image                                       | Scheduling Mode  | Publ    |
| va_server_demo_admintool  | va/admintool:DE_improve_save_74             | replicated 1 / 1 | Scale - |
| va_server_demo_controller | va/controller:with_nlp_answer_ids_in_stream | replicated 1 / 1 | Scale - |
| va_server_demo_dbbackup   | va/dbbackup:with_db_calls                   | replicated 1 / 1 | Scale - |

| Status   | Filter | Task                      | Actions   | Slot |
|----------|--------|---------------------------|-----------|------|
| running  |        | j87nf8iumzxhj56jy6n87ugeb | 📄 ⓘ ⚙️ >_ | 1    |
| shutdown |        | m7nn6u1na15o4crlc6bh0d89  | 📄 ⓘ       | 1    |
| shutdown |        | n91fdqhepuqlo6u1n4lw53kqh | 📄 ⓘ       | 1    |
| shutdown |        | qo1its19tfb2ast80x7dti6rc | 📄 ⓘ       | 1    |

- Проверьте статус выбранного сервиса- все контейнеры должен быть в состоянии "running" и лог не должен содержать критических ошибок типа exception

```
'STT': 'DIARIZATION_ALGORITHM' is overridden with 'speechbrain'  
'STT': 'SPEECH_RECOGNITION_ALGORITHM_BY_LANG' is overridden with '{\"en\": \"conformer\"}'  
'STT': 'DEFAULT_SPEECH_RECOGNITION_ALGORITHM' is overridden with 'conformer'  
'CONSUL': 'ADDRESS' is overridden with 'ubuntu-va-01'  
'GENERAL': 'CUSTOMER' is overridden with 'transcribe'  
'DATABASE': 'CHECK_DB_STRUCTURE' is overridden with 'false'  
[2023-02-15T14:48:35.615][ConformerHelper] [INFO] [None]: Started [STT RU_ConformerHelper:0] helper  
[2023-02-15T14:48:35.705][request_new_batch] [INFO] [None]: send GetFiles  
[2023-02-15T14:48:35.706][RU_TRANSCRIPT] [INFO] [None]: Ready to start loop  
[2023-02-15T14:48:35.708][LiveWatcher] [DEBUG] [None]: Removed 'ru prepare_request_loop' thread  
[2023-02-15T14:48:35.709][RU_LANG] [DEBUG] [None]: Started lang pipe loop  
[2023-02-15T14:48:35.709][LiveWatcher] [DEBUG] [None]: Removed 'ru lang_pipe_loop' thread  
[2023-02-15T14:48:35.709][RU_TRANSCRIPT] [INFO] [None]: Start loop  
[2023-02-15T14:48:35.716][SttConnector] [INFO] [None]: Initialized 'ru' language  
[2023-02-15T14:48:35.717][LiveWatcher] [DEBUG] [None]: Removed 'lang activator loop' thread  
[2023-02-15T14:48:35.897][SttServer] [INFO] [None]: ===== STT server started =====  
[2023-02-15T14:48:35.897][SttServer] [INFO] [None]: Supported Languages: ['en', 'ru']  
[2023-02-15T14:48:35.897][SttServer] [INFO] [None]: ===== Started Languages: ['en', 'ru'] =====  
[2023-02-15T14:48:35.897][SttServer] [INFO] [None]: ===== Supported audio types: ['OPUS', 'RAW', 'AAC', 'WEBM', 'CAF', 'OGG', 'WAV'] =====
```

### 7.3 Диагностика и решение проблем, характерных для первоначального деплоя стека

- Проблема: сервисы не стартуют по причине отсутствия точки монтирования

| va_server_dev2_loadbalancer         |        |                           |         |      |
|-------------------------------------|--------|---------------------------|---------|------|
| va/loadbalancer:balancer_unoto_cert |        |                           |         |      |
| replicated 0 / 1 Scale              |        |                           |         |      |
| Status                              | Filter | Task                      | Actions | Slot |
| rejected                            |        | i14mt1z4yp3jhcxjxbyfwj41z | 📄 ⓘ     | 1    |
| rejected                            |        | oumrwbvb4tyccd2kpotpf9u9o | 📄 ⓘ     | 1    |

#### Task details

Services > va\_server\_dev2\_loadbalancer > i14mt1z4yp3jhcxjxbyfwj41z

| Task status   |                                                                                                           |
|---------------|-----------------------------------------------------------------------------------------------------------|
| ID            | i14mt1z4yp3jhcxjxbyfwj41z                                                                                 |
| State         | rejected                                                                                                  |
| State Message | preparing                                                                                                 |
| Error message | invalid mount config for type "bind": bind source path does not exist: /mount/internal/service/cert_unoto |
| Image         | va/loadbalancer:balancer_unoto_cert                                                                       |
| Slot          | 1                                                                                                         |

Причина: отсутствует папка точки монтирования, содержащаяся в файле деплоя

Решение: проверить корректность описания точки монтирования в файле деплоймента, создать/скорректировать папку точки монтирования

| Type   | Source                                                 | Target                  | Read only                | Actions |
|--------|--------------------------------------------------------|-------------------------|--------------------------|---------|
| Bind   | /mount/internal/service/certs_unoto                    | /cert                   | <input type="checkbox"/> |         |
| Bind   | /mount/internal/service/certs_unoto                    | /cert_old               | <input type="checkbox"/> |         |
| Bind   | /mount/internal/service/cert_origin                    | /cert_origin            | <input type="checkbox"/> |         |
| Bind   | /mount/internal/service/logs/loadbalancer_logs/dev2    | /root/loadbalancer/logs | <input type="checkbox"/> |         |
| Bind   | /mount/internal/service/AdminTool/Private/dev2/Content | /srv/content            | <input type="checkbox"/> |         |
| Volume | va_server_dev2_dcs_locks - local                       | /var/dcs_locks          | <input type="checkbox"/> |         |

- Проблема: сервисы не подключаются к базе данных

Container logs  
Containers > va\_server\_dev2\_nlp01.rmh5jph5mj3t9om3mu26ua04 > Logs

Log viewer settings

Auto-refresh logs

Wrap lines

Display timestamps

Fetch: All logs

Search: Filter...

Lines: 100

Actions: Download logs, Copy, Copy selected lines, Unselect

```
r_nlp@'10.0.15.211' (using password: YES)
Retry 2 after 0.5 sec for MySQLHandler.get_connection
[2023-02-16T11:17:58.594][MySQLHandler][ERROR][None]: Cannot create connection pool [ProgrammingError]: 1045 (28000): ProxySQL Error: Access denied for user 'proxy'
r_nlp@'10.0.15.211' (using password: YES)
Retry 3 after 0.501 sec for MySQLHandler.get_connection
[2023-02-16T11:17:59.095][MySQLHandler][ERROR][None]: Cannot create connection pool [ProgrammingError]: 1045 (28000): ProxySQL Error: Access denied for user 'proxy'
r_nlp@'10.0.15.211' (using password: YES)
Retry 2 after 30.001 sec for MySQLHandler.check_availability
[2023-02-16T11:18:29.101][MySQLHandler][ERROR][None]: Cannot create connection pool [ProgrammingError]: 1045 (28000): ProxySQL Error: Access denied for user 'proxy'
r_nlp@'10.0.15.211' (using password: YES)
```

Причина: один или несколько контейнеров сервиса базы данных не запустились корректно (проверить логи сервиса persona\_cluster)

va\_server\_dev2\_persona-xtradb-cluster va/persona\_cluster:va-bundle-4.2236.02 replicated 1 / 2 Scale -

| Status   | Task                      | Actions | Slot | Node         |
|----------|---------------------------|---------|------|--------------|
| failed   | taj1iio7ybmaubd7xsypzeez2 |         | 2    | ubuntu-va-01 |
| failed   | xjo7bwwn89jm4ne3vhg8qcwdb |         | 1    | ubuntu-va-02 |
| ready    | jqkbh8s5wpjnj01avo7vo47wy |         | 2    | ubuntu-va-01 |
| running  | yruthe4qmnwfgwksi2o1v27yr |         | 1    | ubuntu-va-02 |
| shutdown | 00wa5tz4bzz9b96cthr5cuycp |         | 2    | ubuntu-va-01 |

## Task details

Services > va\_server\_dev2\_percona-xtradb-cluster > jqkhh85wpjnj0lavo7vo47wy > Logs

Log viewer settings

va\_server\_dev2\_percona-xtradb-cluster va/percona\_cluster:va-bundle-4.2236.02 replicated 1 / 2 Scale

Auto-refresh logs

Wrap lines

Display timestamps

Fetch running All logs

Search Filter...

Lines 100

Actions

```
2023-02-16T11:27:48.476704Z 0 [Note] [MY-000000] [Galera] Save the discovered primary-component to disk
2023-02-16T11:27:48.520189Z 0 [Note] [MY-000000] [Galera] New COMPONENT: primary = yes, bootstrap = no, my_idx = 0, memb_num = 2
2023-02-16T11:27:48.520327Z 0 [Note] [MY-000000] [Galera] STATE_EXCHANGE: sent state UUID: f0fab318-adee-11ed-ad7b-8a17fd019b96
2023-02-16T11:27:48.523208Z 0 [Note] [MY-000000] [Galera] STATE_EXCHANGE: sent state msg: f0fab318-adee-11ed-ad7b-8a17fd019b96
2023-02-16T11:27:48.523837Z 0 [Note] [MY-000000] [Galera] STATE_EXCHANGE: got state msg: f0fab318-adee-11ed-ad7b-8a17fd019b96 from 0 (8b5ba542d374)
2023-02-16T11:27:48.974647Z 0 [Note] [MY-000000] [Galera] STATE_EXCHANGE: got state msg: f0fab318-adee-11ed-ad7b-8a17fd019b96 from 1 (1a1f9fa5b7f8)
2023-02-16T11:27:48.974687Z 0 [Note] [MY-000000] [Galera] Quorum results:
version = 6
```

Решение: перезапустить сервис базы данных (percona\_cluster), дождаться корректного старта всех (обычно 2-3) контейнеров данного сервиса (сообщение “sshd is running” должно появляться в логе периодически) после чего перезапустить сервис DbBackup и дождаться корректного восстановления базы данных

## Stack details

va\_server\_dev2

### Services

Search...

| <input type="checkbox"/>            | Name ↓                                | Image                                                           | Scheduling Mode        |
|-------------------------------------|---------------------------------------|-----------------------------------------------------------------|------------------------|
| <input type="checkbox"/>            | va_server_dev2_admintool              | va/admintool:ofd_import292                                      | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_controller             | va/controller:MLVA-361_25012023                                 | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_dbbackup               | va/dbbackup:MLVA-436_08022023                                   | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_etcd                   | va/etcd:v3.3.25                                                 | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_extsrv                 | va/extsrv:r7                                                    | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_kafdrop                | obsidiandynamics/kafdrop:3.30.0                                 | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_kafka-broker-0         | va/kafka:1.1                                                    | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_kafka-broker-1         | va/kafka:1.1                                                    | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_kafka-broker-2         | va/kafka:1.1                                                    | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_kafka-ui               | provectuslabs/kafka-ui:7837622d5eaacae9bade7342f14732967e1b48d0 | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_loadbalancer           | va/loadbalancer:balancer_unoto_cert                             | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_myadmin                | va/phpmyadmin:va-bundle-4.2236.02                               | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_nlp0                   | va/nlp:MLVA-257_kafka_punctuator_09022023                       | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_nlp_punctuator_en      | va/nlp_punctuator:0.1_09022023_en                               | replicated 1 / 1 Scale |
| <input type="checkbox"/>            | va_server_dev2_nlp_punctuator_ru      | va/nlp_punctuator:0.1_09022023_ru                               | replicated 1 / 1 Scale |
| <input checked="" type="checkbox"/> | va_server_dev2_percona-xtradb-cluster | va/percona_cluster:va-bundle-4.2236.02                          | replicated 2 / 2 Scale |

## Container logs

Containers > va\_server\_dev2\_percona-xtradb-cluster.1.p6w6lr3errtpphckysbyes > Logs

Log viewer settings

Auto-refresh logs

Wrap lines

Display timestamps

Fetch: All logs

Search: Filter...

Lines: 100

Actions: [Download logs](#) [Copy](#) [Copy selected lines](#) [Unselect](#)

```
* sshd is running
mbind: Operation not permitted
mbind: Operation not permitted
* sshd is running
mbind: Operation not permitted
mbind: Operation not permitted
* sshd is running
mbind: Operation not permitted
mbind: Operation not permitted
* sshd is running
mbind: Operation not permitted
```

## Container logs

Containers > va\_server\_dev2\_dbbackup.1.915oywhaj7w02k44bgr5rmag3 > Logs

Log viewer settings

Auto-refresh logs

Wrap lines

Display timestamps

Fetch: All logs

Search: Filter...

Lines: 100

Actions: [Download logs](#) [Copy](#) [Copy selected lines](#) [Unselect](#)

```
[2023-02-16T11:39:58.554][MySQLMigration] [[INFO] [None]: Try to run installation command: REPLACE INTO `db_main`.`settings_client` (`client_id`, `key_name`, `value`, `value_type`) select id,
row `db_main`.`client` left join `db_main`.`settings_client` on `client`.`id`=`settings_client`.`client_id` and `settings_client`.`key_name`='output_audio_types' where version='unoto'
[2023-02-16T11:39:58.558][MySQLMigration] [[INFO] [None]: Command is completed
[2023-02-16T11:39:58.559][MySQLMigration] [[INFO] [None]: Try to run installation command: UPDATE `db_main`.`users` set `role`='admin,skill_tool,plugin_manager,developer,debugger' where id=1
[2023-02-16T11:39:58.563][MySQLMigration] [[INFO] [None]: Command is completed
[2023-02-16T11:39:58.563][MySQLMigration] [[INFO] [None]: Try to run installation command: REPLACE INTO `db_main`.`settings` VALUES ('CES_CUSTOMER', 'demo', 0)
[2023-02-16T11:39:58.567][MySQLMigration] [[INFO] [None]: Command is completed
[2023-02-16T11:39:58.568][MySQLMigration] [[INFO] [None]: Try to run installation command: REPLACE INTO `db_main`.`settings` VALUES ('CONTROLLER_SETTINGS_REGISTER_LIST', 'ADMINTOOL_URL, CLIENT,
[2023-02-16T11:39:58.572][MySQLMigration] [[INFO] [None]: Command is completed
[2023-02-16T11:39:58.572][MySQLMigration] [[INFO] [None]: Try to run installation command: COMMIT
[2023-02-16T11:39:58.575][MySQLMigration] [[INFO] [None]: Command is completed
[2023-02-16T11:39:58.575][DBBackup] [[INFO] [None]: All migrations are complete, add all users to ProxySQL
Param DATABASE_HOST is None, set to proxy
Warning: Permanently added 'proxy,10.0.15.23' (ECDSA) to the list of known hosts.
set env from file
All users were added to ProxySQL
-----Script to add user to the proxy was completed
[2023-02-16T11:40:01.758][DBBackup] [[INFO] [None]: Added all user to ProxySQL
[2023-02-16T11:40:01.760][DBBackup] [[INFO] [None]: Backup task timer for `db` is set by 18:50 every 1 days
[2023-02-16T11:40:01.762][DBBackup] [[INFO] [None]: Backup task timer for `db_nip` is set by 18:55 every 1 days
[2023-02-16T11:40:01.762][DBBackup] [[INFO] [None]: Backup task timer for `db_log` is set by 19:00 every 1 days
[2023-02-16T11:40:01.763][DBBackup] [[INFO] [None]: Backup task timer for `db_calls` is set by 19:05 every 1 days
[2023-02-16T11:40:01.763][DBBackup] [[INFO] [None]: Backup task timer xtrabackup is set by 18:30 every 1 days
[2023-02-16T11:40:01.764][DBBackup] [[INFO] [None]: Logrotate timer is set by 06:10 every 1 days
[2023-02-16T11:40:01.764][DBBackup] [[INFO] [None]: KPI timer is set by 00:30 every sunday
[2023-02-16T11:40:01.765][DBBackup] [[INFO] [None]: Languages download timer didn't set
```

- Проблема: ошибка CUDA (CUDA error) при старте сервисов, использующих GPU (STT, TTS Engine)

```

Lines 2000

Actions [Download logs] [Copy] [Copy selected lines] [Unselect]

[2023-02-16T11:56:30.971][STDERR] ][ERROR][None]: File "/root/tts_engine/connectors/tts_engine_connector.py", line 49, in __init__
[2023-02-16T11:56:30.971][STDERR] ][ERROR][None]: 'natasha': Synthesizer.from_config("config.yaml", name="natasha')
[2023-02-16T11:56:30.972][STDERR] ][ERROR][None]: File "/root/tts_engine/tts/synthesizer.py", line 278, in from_config
[2023-02-16T11:56:30.972][STDERR] ][ERROR][None]: params[mtype] = Synthesizer.module_from_config(modules_config, mtype, mname, device)
[2023-02-16T11:56:30.972][STDERR] ][ERROR][None]: File "/root/tts_engine/tts/synthesizer.py", line 293, in module_from_config
[2023-02-16T11:56:30.972][STDERR] ][ERROR][None]: return _modules_dict[mname](**module_config)
[2023-02-16T11:56:30.973][STDERR] ][ERROR][None]: File "/root/tts_engine/tts/backend_wrappers/waveglow.py", line 24, in __init__
[2023-02-16T11:56:30.973][STDERR] ][ERROR][None]: self.model.eval().to(device=self.device, dtype=self.dtype)
[2023-02-16T11:56:30.973][STDERR] ][ERROR][None]: File "/usr/local/lib/python3.8/dist-packages/torch/nn/modules/module.py", line 612, in to
[2023-02-16T11:56:30.973][STDERR] ][ERROR][None]: return self._apply(convert)
[2023-02-16T11:56:30.974][STDERR] ][ERROR][None]: File "/usr/local/lib/python3.8/dist-packages/torch/nn/modules/module.py", line 359, in _apply
[2023-02-16T11:56:30.974][STDERR] ][ERROR][None]: module._apply(fn)
[2023-02-16T11:56:30.974][STDERR] ][ERROR][None]: File "/usr/local/lib/python3.8/dist-packages/torch/nn/modules/module.py", line 381, in _apply
[2023-02-16T11:56:30.974][STDERR] ][ERROR][None]: param_applied = fn(param)
[2023-02-16T11:56:30.974][STDERR] ][ERROR][None]: File "/usr/local/lib/python3.8/dist-packages/torch/nn/modules/module.py", line 610, in convert
[2023-02-16T11:56:30.975][STDERR] ][ERROR][None]: return t.to(device, dtype if t.is_floating_point() else None, non_blocking)
[2023-02-16T11:56:30.975][STDERR] ][ERROR][None]: RuntimeError: CUDA out of memory. Tried to allocate 14.00 MiB (GPU 0; 7.79 GiB total capacity; 712.97 MiB already allocated; 9.75 MiB free; 714.00 MiB reserved in total by PyTorch)
[2023-02-16T11:56:31.062][TTSEngineWorker] ][DEBUG][None]: Destroy (TTSEngineWorker)

```

Причина: отсутствует GPU на узле, запускающем данный сервис; не установлены драйвера GPU; недостаточно оперативной памяти GPU

```

svc_va@ubuntu-va-02:~$ nvidia-smi
Thu Feb 16 15:02:34 2023

+-----+
| NVIDIA-SMI 470.74      | Driver Version: 470.74      | CUDA Version: 11.4      |
+-----+-----+-----+-----+-----+-----+
GPU  Name	Persistence-M	Bus-Id	Disp.A	Volatile Uncorr. ECC
Fan  Temp	Pwr:Usage/Cap	Memory-Usage	Memory	GPU-Util  Compute M.
+-----+-----+-----+-----+-----+-----+				
0   NVIDIA GeForce ...	Off	00000000:02:00:0	Off	0%         Default
30%  42C	P8	6300MiB / 7981MiB		N/A
+-----+-----+-----+-----+-----+-----+				
1   NVIDIA GeForce ...	Off	00000000:03:00:0	Off	0%         Default
27%  33C	P8	7285MiB / 7982MiB		N/A
+-----+-----+-----+-----+-----+-----+

Processes:
+-----+-----+-----+-----+-----+-----+
| GPU  GI  CI           | PID  Type  | Process name          | GPU Memory |
|   ID  ID  ID           |      |      |                       | Usage      |
+-----+-----+-----+-----+-----+-----+
0   N/A N/A	1604	G	/usr/lib/xorg/Xorg	8MiB
0   N/A N/A	1783	G	/usr/bin/gnome-shell	3MiB
0   N/A N/A	2650184	C	STT_RU_LANG	759MiB
0   N/A N/A	2651230	C	/usr/bin/python	951MiB
0   N/A N/A	2651608	C	STT_EN_SttSdkHelper:0	915MiB
0   N/A N/A	2651609	C	STT_EN_SttSdkHelper:0	915MiB
0   N/A N/A	2652299	C	STT_EN_SttSdkHelper:0	913MiB
0   N/A N/A	2652728	C	STT_RU_SttSdkHelper:0	915MiB
0   N/A N/A	2653742	C	STT_RU_SttSdkHelper:0	915MiB
1   N/A N/A	1604	G	/usr/lib/xorg/Xorg	4MiB
1   N/A N/A	2168567	C	STT_RU_LANG	759MiB
1   N/A N/A	2169002	C	/usr/bin/python	967MiB
1   N/A N/A	2171395	C	STT_RU_ConformerHelper:0	1891MiB
1   N/A N/A	2229854	C	STT_EN_SttSdkHelper:0	915MiB
1   N/A N/A	2231610	C	STT_EN_SttSdkHelper:0	915MiB
1   N/A N/A	2231650	C	STT_RU_SttSdkHelper:0	915MiB
1   N/A N/A	2239212	C	STT_RU_SttSdkHelper:0	915MiB
+-----+-----+-----+-----+-----+-----+

```

Решение: с помощью меток узлов кластера перенести контейнеры сервисы на узел с GPU/ достаточным количеством памяти GPU; установить драйвера GPU

The screenshot shows the Kubernetes dashboard interface. On the left sidebar, the 'Swarm' menu item is highlighted with a red box. The main panel displays the 'Nodes' section with a table of node details. The 'ubuntu-va-01' node is highlighted with a red box. The table shows the following information for each node:

| Name         | Role    | CPU | Memory  | Engine   | IP Address    | Status | Availability |
|--------------|---------|-----|---------|----------|---------------|--------|--------------|
| ubuntu-va-01 | worker  | 20  | 135 GB  | 20.10.9  | 192.168.4.24  | ready  | active       |
| ubuntu-va-02 | manager | 20  | 135 GB  | 20.10.7  | 192.168.4.42  | ready  | active       |
| ubuntu-va-03 | worker  | 12  | 50.4 GB | 20.10.12 | 192.168.4.189 | ready  | active       |

</> Node Details

Role: worker

Availability: Active

Status: ready

Node Labels

+ label

|      |        |       |      |  |
|------|--------|-------|------|--|
| name | DB     | value | true |  |
| name | STTSRV | value | 1    |  |

Apply changes

portainer.io

Home

PRIMARY

Dashboard

App Templates

Stacks

Services

Containers

Images

Networks

Volumes

Configs

Secrets

Swarm

Apply changes

Placement constraints

placement constraint

| Name               | Operator | Value |  |
|--------------------|----------|-------|--|
| node.labels.STTSRV | ==       | 1     |  |

Apply changes

Placement preferences

placement preference

There are no placement preferences for this service.