

Руководство пользователя GPU SuperComputer

Самый быстрый путь к работе с искусственным интеллектом.

<https://cloud.mts.ru>

Отзывы и пожелания по продуктам #CloudMTS присылайте на feedback@cloud.mts.ru

Версия от 15 августа 2019 года.

Оглавление

Руководство пользователя GPU SuperComputer	1
GPU SuperComputer	3
Что такое GPU SuperComputer?	3
Аппаратные конфигурации виртуальных машин	3
Тарификация	4
Регионы доступности	4
Как получить доступ	4
Работа с виртуальной машиной в GPU SuperComputer	5
Подключение к серверу по SSH	5
Операционная система NVIDIA DGX OS	5
Проверка версии DGX OS	5
Установка обновлений DGX OS	6
Создание аккаунта NVIDIA GPU Cloud	6
Проверка доступных GPU и приложений, использующих GPU	6
Техническая поддержка и отзывы	7
Как обратиться в поддержку?	7
Как оставить отзыв или пожелание?	7

GPU SuperComputer

Что такое GPU SuperComputer?

GPU SuperComputer – облачная инфраструктура, предоставляющая возможность использования программно-аппаратных систем глубокого обучения класса NVIDIA DGX как сервис #CloudMTS.

Системы спроектированы для запуска программного стека NVIDIA DGX с поддержкой NVIDIA docker и оптимизированного каталога контейнеров NVIDIA GPU Cloud, архитектурно разработаны для глубокого обучения, высокопроизводительных вычислений, транскодинга и аналитики.

В настоящий момент доступны системы:

- на базе NVIDIA DGX-1 supercomputer в конфигурации “DGX-1 with V100-32” производительностью до 1 петаФЛОПС. Система содержит 8 видеокарт Nvidia Tesla V100 32Gb, соединенных по шине NVLink 300 Гбит/с.

В качестве виртуализации используется NVIDIA KVM с гарантированным выделением ресурсов.

Заказчику предоставляется доступ к виртуальной машине на GPU SuperComputer с необходимым количеством видеокарт по протоколу SSH, с правами администратора. В качестве ОС используется NVIDIA DGX OS (доработанная Ubuntu Linux Server без графического интерфейса, с интегрированным программным стеком NVIDIA, с поддержкой NVIDIA Docker).

Заказчик не ограничивается в выборе ПО каталогом NVIDIA GPU Cloud – может устанавливать и использовать любое ПО, которое сможет работать в серверной редакции Ubuntu Linux и использовать видеокарты. Для проверки совместимости ПО Заказчика с GPU SuperComputer может быть предоставлен тестовый период.

Аппаратные конфигурации виртуальных машин

На системе на базе NVIDIA DGX-1 supercomputer доступны следующие аппаратные конфигурации виртуальных машин:

- 1 x Tesla V100 32 Гб (5120 ядер CUDA, 640 ядер Tensor)
10 vCPU E5-2698 v2 @ 2.2 GHz, 61 GB Ram DDR4, 250 GB SSD.
- 2 x Tesla V100 32 Гб (10240 ядер CUDA, 1280 ядер Tensor)
20 vCPU E5-2698 v2 @ 2.2 GHz, 122 GB Ram DDR4, 500 GB SSD.
- 4 x Tesla V100 32 Гб (20480 ядер CUDA, 2560 ядер Tensor)
40 vCPU E5-2698 v2 @ 2.2 GHz, 215 GB Ram DDR4, 1000 GB SSD.
- 8 x Tesla V100 32 Гб (40960 ядер CUDA, 5120 ядер Tensor)
80 vCPU E5-2698 v2 @ 2.2 GHz, 490 GB Ram DDR4, 2000 GB SSD.

Системные диски виртуальных машин располагаются в массиве RAID 0 для максимальной производительности. К каждой виртуальной машине подключен диск данных **/raid** из массива RAID 1 для максимальной надежности при работе с данными.

Тарификация

Для услуги доступна модель потребления Allocation pool (выделенный пул ресурсов).

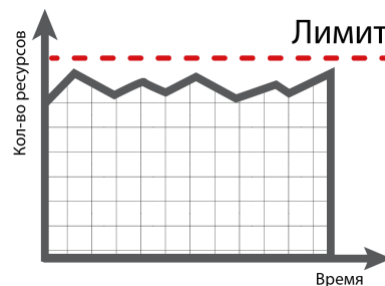
Данная модель подразумевает оплату фиксированных выделенных ресурсов с ежемесячной или ежедневной тарификацией (на выбор).

Ежемесячная тарификация подходит, если вам необходим постоянный доступ к ресурсам GPU SuperComputer по наиболее выгодной цене. Выбранная конфигурация заказывается на срок не менее месяца.

Ежедневная тарификация подходит, если вам необходим доступ к ресурсам GPU SuperComputer на небольшое время, например на несколько дней.

Услуга может быть заказана на фиксированный срок (тарификация на выбор) или на неопределенный срок (применяется ежемесячная тарификация).

При необходимости продления услуги, заказанной на фиксированный срок, необходимо до окончания срока отправить заявку на продление.



Регионы доступности

В настоящий момент GPU SuperComputer развернут в регионе: Москва.

Как получить доступ

Для получения консультации заполните заявку на сайте <https://cloud.mts.ru> или обратитесь к персональному менеджеру.

После консультации, при согласовании требуемой конфигурации с отделом продаж, вы получите письмо с доступами по протоколу SSH к виртуальной машине в GPU SuperComputer.

Работа с виртуальной машиной в GPU SuperComputer

Подключение к серверу по SSH

Для начала работы подключитесь к серверу по SSH с полученными данными для доступа.

```
ssh nvidia@ip-адрес
```

```
trukhinyuris-MacBook-Pro:Temp trukhinyuri$ ssh nvidia@10.97.36.165
Welcome to NVIDIA DGX KVM VM Server Version 4.1.0 (GNU/Linux 4.15.0-50-generic x86_64)

System information as of Mon Jun  3 06:05:51 PDT 2019

System load:  0.0          Processes:      208
Usage of /:   2.1% of 245.82GB  Users logged in:  1
Memory usage: 1%          IP address for enp1s0: 10.97.36.165
Swap usage:   0%          IP address for docker0: 172.17.0.1
Last login: Mon Jun  3 06:05:36 2019 from 194.54.148.72
nvidia@9-1g2:~$
```

В целях безопасности рекомендуем вам использовать доступ по ключу и отключить доступ по паролю к серверу.

Диск данных (data disk) находится в директории `/raid` и `/home/nvidia/raid` (символическая ссылка на `/raid`). Пожалуйста используйте диск данных для работы с вашими данными в целях максимальной надежности хранения.

Операционная система NVIDIA DGX OS

Виртуальная машина работает под управлением NVIDIA DGX OS, построенной на базе Ubuntu Server с интегрированными пакетами NVIDIA.

Основные интеграции:

- Обновляемый видеодрайвер NVIDIA
- Контейнерная виртуализация Docker CE
- NVIDIA Container Runtime for Docker
- NVIDIA System Management (NVSM). Обеспечиваем мониторинг здоровья и системные предупреждения для администраторов, простую возможность проверки здоровья из командной строки.
- Data Center GPU Management (DCGM). Обеспечивает управление GPU-кластерами и дата-центрами.

Вы можете запускать GPU-ускоряемые рабочие нагрузки так же, как на обычном linux-сервере с видеокартами NVIDIA.

Проверка версии DGX OS

```
grep VERSION /etc/dgx-release
```

```
nvidia@9-1g2:~$ grep VERSION /etc/dgx-release
DGX_SWBUILD_VERSION="4.1.0"
```

Установка обновлений DGX OS

```
sudo apt update && sudo apt full-upgrade
```

Создание аккаунта NVIDIA GPU Cloud

NVIDIA GPU Cloud (NGC) представляет простой доступ к GPU-оптимизированным инструментам для задач глубокого обучения и высокопроизводительных вычислений, полностью раскрывающих свои возможности на видеокартах NVIDIA. Аккаунт NGC предоставляет вам доступ к этим инструментам и возможность настройки вашего приватного репозитория для управления вашими кастомизированными инструментами.

[Каталог NVIDIA GPU Cloud.](#)

Проверка доступных GPU и приложений, использующих GPU

Для проверки используйте команду

```
nvidia-smi
```

```
nvidia@9-1g2:~$ nvidia-smi
Mon Jun  3 06:53:15 2019

+-----+
| NVIDIA-SMI 418.67                Driver Version: 418.67          CUDA Version: 10.1     |
+-----+-----+
| GPU   Name           Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|    0   Tesla V100-SXM2...    On   | 00000000:05:00.0 Off  |
| N/A   35C    P0      41W / 300W |  0MiB / 32480MiB |      0%      Default |
+-----+-----+

+-----+
| Processes:                         GPU Memory |
|  GPU       PID    Type    Process name                     Usage    |
+-----+-----+
| No running processes found         |
+-----+
```

Техническая поддержка и отзывы

Как обратиться в поддержку?

Нашим пользователям оказывается техническая поддержка, доступная 24/7.

Для обращения в техническую поддержку вы можете отправить запрос по электронной почте support@cloud.mts.ru или позвонить по телефону круглосуточной горячей линии **8-800-250-10-01** (звонок из России бесплатен).

Как оставить отзыв или пожелание?

Вы можете отправить отзыв о сервисе или ваши пожелания, используя адрес feedback@cloud.mts.ru. Письмо получит команда, занимающаяся развитием сервиса.