**Программное обеспечение**

**«Neural Processing Service»**

**(ПО «NPS»)**

**Руководство по установке**

Листов

Москва 2022

**ОГЛАВЛЕНИЕ**

[ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИй 3](#_Toc114853282)

[ВВЕДЕНИЕ 4](#_Toc114853283)

[1 Требования к программному обеспечению 5](#_Toc114853284)

[2 Требования к аппаратному обеспечению 6](#_Toc114853285)

[3 Установка и настройка ПО «Neural Processing Service» 7](#_Toc114853286)

# ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИй

|  |  |
| --- | --- |
| Сокращение | Детальная расшифровка |
| CPU | Central Processing Unit |
| GPU | Graphics Processing Unit |
| ОС | Операционная система |
| БД | База данных |
| СУБД | Система управления базами данных |
| ПО | Программное обеспечение |

# ВВЕДЕНИЕ

В данном документе приведено руководство по установке программного обеспечения «Neural Processing Service» (далее Система).

Обслуживающий персонал должен иметь высшее техническое образование и квалификацию уровня инженер и выше.

## Требования к программному обеспечению

Минимальные программные требования для центрального сервера с установленными сервисами системы:

1. ОС семейства GNU/Linux (Ubuntu Linux 20.04 и новее, РЕД ОС и т.п.);
2. Установленные драйверы GPU nVidia последней версии;
3. СУБД PostgreSQL;
4. работоспособная платформа Docker на сервере;
5. доступ в локальную сеть или Интернет с выделенным IP-адресом.

## Требования к аппаратному обеспечению

Минимальные системные требования для центрального сервера с установленными сервисами системы:

1. CPU не ниже Intel Xeon Silver не ниже 1 поколения с не менее, чем 8 аппаратными ядрами с поддержкой HyperThreading;
2. GPU не ниже nVidia Quadro RTX 4000 для отработки нейросетевых алгоритмов;
3. не менее 64Гб оперативной памяти;
4. файловое хранилище для СУБД в системах RAID 5, 6 или 10;
5. не менее 400Гб свободного места на диске;
6. сетевое соединение не менее 1Гбит/сек.

В случае тестирования системы, ПО может быть установлено на более слабый вычислительный блок в составе:

1. центральный процессор Intel Core i7 6 поколения или выше с не менее чем 4 аппаратными ядрами с поддержкой HyperThreading (данное требование является минимальным для запуска фреймворка Intel OpenVino, начиная с версии 2021.4 LTS);
2. видеокарта не ниже nVidia GeForce GTX 1070 для отработки нейросетевых алгоритмов (данное требование по поколению является минимальным для запуска фреймворка Nvidia TensorRT, начиная с версии 7.x, также для некоторых нейросетевых алгоритмов требуется существенный объем памяти, поэтому минимальные рекомендации – от 6Гб и выше);
3. не менее 16Гб оперативной памяти.

## Установка и настройка ПО «Neural Processing Service»

Для запуска ПО системы требуется наличие и работоспособность платформы Docker на сервере.

* 1. В файле docker-compose существуют следующие разделы:
		1. ws – связующий сервис, получает команды от веб-клиента, подключается к сервису запуска алгоритмов машинного обучения (алгоритмический сервис) через localhost;
		2. analysis\_service – сервис анализа данных
		3. algorithm\_service – сервис запуска алгоритмов машинного обучения
	2. В файле docker-compose.yaml необходимо изменить следующие параметры:
		1. Удостовериться, что GPU nVidia «проброшено» в контейнер:



Рисунок 1 - настройки algorithm\_service

* + 1. В случае необходимости возможно изменить номер порта для подключения к algorithm\_service. Для этого необходимо изменить номер порта в подразделе ports соответствующего раздела конфигурации контейнера. Соответствующие изменения необходимо внести в файл config.ini сервиса:



Рисунок 2 - файл config.ini algorithm\_service

* 1. В файлах конфигурации сервисов analysis\_service (config.json) и ws (appsettings.prod.json) изменить настройки подключения к БД PostgreSQL в соответствии с необходимыми параметрами подключения:



Рисунок 3 - настройки подключения к БД analysis\_service



Рисунок 4 - настройки подключения к БД ws

* 1. Из каталога, содержащего файл «images.tar», нужно выполнить команду «docker load -i images.tar» и дождаться окончания выполнения процесса.
	2. Для запуска сервисов требуется из каталога, содержащего docker-compose файл, выполнить команду «docker-compose up –build» при первом запуске и «docker-compose up» при последующих.