

INFOWATCH

InfoWatch Data Analysis Service. Руководство по установке, конфигурированию и администрированию

29/10/2026

© АО “ИнфоВотч”

Тел./Факс +7 (495) 229-00-22

http://www.infowatch.ru

СОДЕРЖАНИЕ

1 Аудитория 6

2 Комплект документов 7

3 Техническая поддержка пользователей 8

4 Обзор 9

4.1 Состав Data Analysis Service: 9

5 Типы установки и обновления, совместимость с другими продуктами, аппаратные и программные требования 10

5.1 Типы установки и совместимость с другими продуктами InfoWatch 10

5.2 Типы обновления 11

5.3 Аппаратные требования 11

5.4 Программные требования 13

6 Подготовка сервера 14

6.1 Шаги по подготовке сервера: 14

7 Настройка сетевых правил доступа 23

8 Установка Data Analysis Service 24

8.1 Варианты запуска установки Data Analysis Service 25

8.2 Установка Data Analysis Service в интерактивном режиме 25

8.3 Установка Data Analysis Service с указанием параметров в командной строке 31

9 Установка моделей 38

10 Устранение неполадок 41

11 Обновление Data Analysis Service 42

11.1 Варианты обновления Data Analysis Service 42

11.2 Особенности обновления кластера Центра расследований 43

11.3 Прямое обновление Data Analysis Service 44

11.4 Обновление Data Analysis Service с помощью сброса установки (reset) 47

12 Настройка Data Analysis Service 54

12.1 Настройка взаимодействия с продуктами Платформы ... 54
12.1.1 Включение технологий Data Analysis Service в продуктах Платформы ... 54
Activity Monitor ... 54
Data Discovery ... 55
Data Explorer ... 56
12.1.2 Задание нод обработки ... 56
12.2 Включение/выключение модулей Data Analysis Service ... 57
12.3 Настройка производительности ... 58
12.3.1 Оценка производительности модуля протокластеризации ... 58
12.3.2 Оценка производительности модулей Сервиса ... 58
12.3.3 Распределенная установка ... 59
12.4 Использование GPU ... 60
12.4.1 Аппаратные и программные требования к графическому процессору ... 60
12.4.2 Настройка доступа к GPU ... 61
12.5 Настройка модуля транскрибации ... 62
12.5.1 Включение модуля ... 62
12.5.2 Общее описание ... 62
12.5.3 Включение использования GPU ... 63
12.5.4 Выбор модели транскрибации ... 63
12.5.5 Проверка использования GPU ... 64
12.5.6 Настройки языков ... 64
Включение мультиязычной транскрибации ... 64
Задание языков транскрибации ... 65
12.6 Настройка модуля распознавания именованных сущностей ... 66
12.6.1 Включение модуля ... 66
12.6.2 Настройка параметров модуля ... 66
12.7 Настройка модуля детектирования объектов на изображениях ... 67
12.7.1 Включение модуля ... 67
12.7.2 Настройка параметров модуля ... 68
12.7.3 Включение использования GPU ... 69
12.8 Настройка модуля OCR ... 69
12.8.1 Включение модуля ... 69
12.8.2 Экстракторы ... 70
12.8.3 Выбор экстрактора ... 70
12.8.4 Настройка экстракторов ... 71
12.8.5 Включение использования GPU ... 72
12.9 Добавление языков для протокластеризации ... 73
12.10 Диагностика Data Analysis Service ... 74
12.10.1 Логирование ... 74
12.10.2 Скрипт диагностики ... 75
12.10.3 Диагностика технологий ... 78

13 Удаление Data Analysis Service 79

Настоящее руководство содержит сведения по установке, удалению и настройке компонентов сервиса InfoWatch Data Analysis Service (далее – Сервис или Data Analysis Service).

1 Аудитория

Информация, содержащаяся в Руководстве, предназначена для пользователей, работающих с Сервисом (выполняющих установку, удаление и т. п.).

Руководство рассчитано на пользователей, знакомых с основами работы в среде операционных систем Linux и в СУБД PostgreSQL.

2 Комплект документов

В документацию по InfoWatch Data Analysis Service входит:

- *InfoWatch Data Analysis Service. Руководство по установке, конфигурированию и администрированию*. Содержит описание установки, обновления, удаления и настройки Сервиса.

Сопутствующая документация по продукту InfoWatch Traffic Monitor включает в себя:

- *InfoWatch Traffic Monitor. Руководство по установке*. Содержит описание установки, обновления и удаления системы InfoWatch Traffic Monitor (далее Traffic Monitor);
- *InfoWatch Traffic Monitor. Руководство администратора*. Содержит информацию по администрированию Traffic Monitor (база данных, серверная часть);
- *InfoWatch Traffic Monitor. Руководство пользователя*. Содержит описание работы с Traffic Monitor (настройка конфигурации, экспорт/импорт данных, составление политик для обработки объектов);
- *InfoWatch Traffic Monitor. Справочник по конфигурационным файлам*. Содержит пояснения к часто используемым конфигурационным файлам.

Кроме основных руководств для пользователей и администраторов Сервиса предусмотрена База знаний lk.infowatch.ru, которая включает материалы по устранению неисправностей и ответы на вопросы. В случае отсутствия доступа к Базе знаний, просьба обращаться в службу технической поддержки компании InfoWatch по адресу support@infowatch.com.

3 Техническая поддержка пользователей

При возникновении проблем и вопросов, связанных с работой Сервиса, вы можете обратиться в службу технической поддержки:

- если вы приобрели продукт у партнера компании InfoWatch, то обратитесь в службу технической поддержки партнера;
- если продукт приобретен у компании InfoWatch напрямую, то обратитесь в службу технической поддержки компании InfoWatch по адресу support@infowatch.com.

Вы также можете посетить раздел технической поддержки на нашем сайте: https://www.infowatch.ru/services/support.

4 Обзор

Data Analysis Service предназначен для потоковой постобработки и анализа данных. Data Analysis Service обрабатывает данные, предоставляемые продуктами компании InfoWatch.

4.1 Состав Data Analysis Service:

- **Модуль протокластеризации текстовых данных**
  Анализирует текстовые документы и разбивает их на тематические группы. В ходе анализа для каждого документа и каждой группы определяется набор тегов – ключевых слов, по которым можно понять тематику и примерное содержание документа/группы документов.
  Модуль протокластеризации используется для работы следующий функций:
  - *Поиск аналогов* в InfoWatch Traffic Monitor. Подробнее см. *InfoWatch Traffic Monitor. Руководство пользователя. Решение задач при работе в консоли Traffic Monitor. Работа с объектами перехвата. Просмотр событий. Просмотр краткой формы события. Поиск аналогов*;
  - *Поиск похожих файлов* в InfoWatch Data Discovery. Подробнее см. *InfoWatch Data Discovery. Руководство пользователя. Раздел "Хранение файлов". Вкладка "Похожие файлы*".
  - Функция протокластеризации в InfoWatch Data Explorer.
- **Модуль транскрибации**
  Конвертирует аудиозаписи в текст. Модуль транскрибации используется при работе с аудиозаписями в InfoWatch Activity Monitor. Подробнее см. *InfoWatch Activity Monitor. Руководство пользователя. Раздел "Мониторинг". Вкладка "Аудиозаписи".*
- **Модуль детектирования объектов на изображениях**
  Распознает объекты на снимках экрана, в том числе непристойный контент. Детектирование объектов на изображениях используется в InfoWatch Activity Monitor. Подробнее см. *InfoWatch Activity Monitor. Руководство пользователя. Раздел "Мониторинг". Вкладка "Наблюдение". Таймлайн*.
- **Модуль распознавания именованных сущностей**
  Находит в тексте сущности заданных пользователем типов. Например, если указать тип "персона", то в тексте могут быть найдены такие сущности, как «Иванов Иван», «директор», «инженер 1й категории» и т.п.
  Данная технология реализована для InfoWatch Data Explorer и InfoWatch Activity Monitor.
- **Модуль оптического распознавания символов** (Далее также – **модуль OCR**)
  Распознает и извлекает текст из изображений. OCR используется для работы со снимками экрана в InfoWatch Activity Monitor. Подробнее см. *InfoWatch Activity Monitor. Руководство пользователя. Раздел "Мониторинг". Вкладка "Снимки экрана".*

5 Типы установки и обновления, совместимость с другими продуктами, аппаратные и программные требования

5.1 Типы установки и совместимость с другими продуктами InfoWatch

Data Analysis Service поддерживает интеграцию со следующими продуктами InfoWatch:

- **Traffic Monitor начиная с версии 7.14**
- **Activity Monitor 3.6**
- **Data Discovery 2.3**
- **Data Explorer 2.1**

> **Примечание:**
>
> Data Analysis Service может работать с одним или несколькими серверами Traffic Monitor.

Основой для установки Сервиса является Платформа. Инсталлятор Data Analysis Service содержит и Платформу, и компоненты Сервиса. Обе части обязательны к установке.
И компоненты Сервиса, и Платформа поддерживают установку в двух режимах:

- `central`;
- `office`.

Режимы установки Платформы и Сервиса вы указываете при запуске инсталлятора. Если оба режима не указаны пользователем, инсталлятор по умолчанию использует значение `central` для обоих компонентов.

Платформенные продукты поддерживают **распределенную установку**: центральная (`central`) и офисные (`office`) ноды объединены в кластер.
Вы можете установить Data Analysis Service как отдельно, так и в кластере с платформенными продуктами. При установке в кластере поддерживается установка как на центральную, так и на офисную ноду.

> **Важно!**
>
> При распределенной установке примите во внимание следующие ключевые аспекты:
>
> - Data Analysis Service и платформенные продукты в кластере должны быть установлены на одной версии Платформы – в номерах версий должны совпадать первые два числа. Версия центральной ноды Платформы должна быть самой новой в кластере.
> - Если вы устанавливаете в кластер Сервис или платформенный продукт на новой версии Платформы, то после этого обновите все остальные экземпляры Сервиса и продукты.
> - В одном кластере можно развернуть только одну центральную ноду Платформы и Data Analysis Service в режиме `central` только в единственном экземпляре.

- Для работы Data Analysis Service в режиме `office` необходимо, чтобы в кластере был экземпляр Сервиса в режиме `central`.
- Если в кластере уже установлен Data Analysis Service в режиме `central`, еще одна установка Сервиса в режиме `central` в том же кластере будет прервана ошибкой.

При установке Data Analysis Service 1.7 устанавливается Платформа версии 2.6.

Data Analysis Service может быть установлен на одном сервере со следующими продуктами InfoWatch:

- **Traffic Monitor начиная с версии 7.14**
- **Activity Monitor 3.6**
- **Data Discovery 2.3**
- **Data Explorer 2.1**
- **Vision 3.8**

Подробнее о процессе установки см. *Установка Data Analysis Service*.

5.2 Типы обновления

Data Analysis Service поддерживает:

- прямое обновление;
- обновление с помощью сброса и переустановки на старые данные.

Подробнее о процессе обновления и поддерживаемых вариантах см. *Обновление Data Analysis Service*.

5.3 Аппаратные требования

Минимальные системные требования, необходимые для установки:

Процессор	ОЗУ	Жесткий диск
Количество ядер – 4; Частота – 3.1 ГГц; Разрядность – 64 бит; Поддержка инструкций SSE 4.2.	16 ГБ	От 300 ГБ свободного места. От 15 ГБ свободного места на диске с директорией `/tmp`. Минимальное количество последовательных операций ввода-вывода – от 50 IOPS.

Требования к аппаратному обеспечению зависят от набора используемых модулей Data Analysis Service и необходимой скорости обработки данных. Далее приведены минимальные рекомендуемые аппаратные требования для работы модулей одного POD Сервиса.

	Процессор, исп	Модель, используемая при	Производительность модуля	Рекомендуемое аппаратное обеспечение

				При установке отдельно от других платформенных продуктов		При установке на один сервер с другим платформенным продуктом	
	используемый для обработки данных	транскрибации	Количество обрабатываемых текстовых контентов в час [1]	CPU, количество ядер/потоков	ОЗУ, ГБ	CPU, количество ядер/потоков	ОЗУ, ГБ
Модуль протокластеризации	CPU	–	30 000	4	16 (24 [2])	2	4 (12 [2])
–	–	–	Отношение времени обработки к длине аудиозаписи	–	–	–	–
Модуль транскрибации	CPU	малая	1:3	6	16	4	6
	GPU с 10 ГБ видеопамяти	малая	1:60	2	16	–	–
		большая	1:30	2	16	–	–

[1] Для текстовых контентов со средним размером 17,5 КБ.

[2] Если вы используете функциональность поиска похожих в Data Discovery при сканировании хранилищ большого размера (более 800 ГБ), то для модуля протокластеризации рекомендуется выделять количество ОЗУ, указанное в скобках.

Важно!

Аппаратные и программные требования к **графическому процессору** (**GPU**) для его использования при транскрибации приведены в статье *Использование GPU*.

При расчете аппаратных требований для совместной установки Data Analysis Service и Traffic Monitor рекомендуется суммировать их отдельные требования.

Примечание:

Чтобы точно рассчитать требования для конкретного проекта, обратитесь в службу технической поддержки компании InfoWatch.

5.4 Программные требования

Для установки и работы Data Analysis Service требуется использовать следующее программное обеспечение:

Тип ПО	Варианты
Операционная система	• Astra Linux Special Edition "Смоленск", версии: • 1.7.1.0 – 1.7.9.46 • 1.8.1.0 – 1.8.5.46 • Astra Linux Special Edition "Орел", версии: • 1.7.5.0 – 1.7.9.46 • Astra Linux Special Edition "Воронеж", версии: • 1.7.5.0 – 1.7.9.46 • Astra Linux Special Edition версии 1.8.5 сертифицированная • Red Hat Enterprise Linux, версии: • 7.9 • 8.4 – 8.10 • Oracle Linux, версии: • 7.9 • 8.4 – 8.10 • РЕД ОС, версии: • 7.3.3 – 7.3.6 • 8.0.0 • ОС Альт: • Сервер 10.1.0 – 10.2.2 • СП Сервер 10.1.0 – 10.2.2 **Примечание**: Вы можете установить Сервис на ОС Альт СП во внешний кластер kubernetes, развернутый до установки, или в кластер kubernetes, который создаст инсталлятор. • Мос.ОС Серверная Арбат 15.5 **Важно!** Модули **Netfilter** и **NAT** должны быть загружены в ядро ОС.

Примечание:

Data Analysis Service использует PostgreSQL 16.4.
СУБД поставляется вместе с продуктом и не требует отдельной установки.

6 Подготовка сервера

Перед установкой или обновлением Data Analysis Service подготовьте сервер.

6.1 Шаги по подготовке сервера:

1. Программа установки написана на языке Python 3. Установите интерпретатор этого языка версии 3.6 или более поздней.
   Пакет с интерпретатором **не** входит в состав дистрибутива Data Analysis Service.
2. На некоторых дистрибутивах ОС не установлена часть стандартных модулей Python 3, в том числе `importlib-metadata` и `pyyaml`. Установите их с помощью следующих команд:

   - **для Red Hat Enterprise Linux, Oracle Linux и РЕД ОС**

     ```
     yum install python3-importlib-metadata
     yum install python3-pyyaml
     ```

   - **для ОС Astra Linux**

     ```
     apt-get install python3-importlib-metadata
     apt-get install python3-distutils
     apt-get install python3-mako
     apt-get install python3-yaml
     ```

   - **для ОС Альт**

     ```
     apt-get install python-modules-json
     apt-get install python-modules-distutils
     apt-get install python3-modules-sqlite3
     apt-get install python3-module-importlib-metadata
     apt-get install python3-module-yaml
     ```

   - **для Мос.ОС**

     ```
     zypper install python3-importlib-metadata
     zypper install python3-PyYAML
     ```

     > **Примечание:**
     >
     > Пакеты не входят в состав дистрибутива Data Analysis Service.
     >
     > Пакеты могут отсутствовать в репозиториях некоторых операционных систем. В этом случае загрузите и установите требуемые пакеты вручную.

3. **На сервере под управлением Мос.ОС:**

Убедитесь, что доступны официальные репозитории ОС:

- `software`
- `update`
- `oss`

4. Настройте правила POD сети или отключите межсетевой экран (подробнее см. статьи Базы знаний *Конфликты при взаимодействии службы firewalld и Kubernetes*, *Полное отключение межсетевого экрана*). В противном случае установка Сервиса будет невозможна.
5. Для корректной работы Сервиса настройте сетевые правила доступа (подробнее см. *Настройка сетевых правил доступа*).
6. Data Analysis Service использует компонент Kubernetes. Для корректной установки Сервиса и работы Kubernetes на сервере должны быть установлены следующие пакеты:
   - `conntrack` или `conntrack-tools` (пакет необходимо установить, даже если `libnetfilter_conntrack` уже установлен).
     Убедитесь, что пакет установлен. Для этого введите в командной строке:

     ```
     conntrack
     ```

     В результате будут выведены сведения о работе с `conntrack` и его установленная версия.

```
Command line interface for the connection tracking system. Version 1.4.4
Usage: conntrack [commands] [options]

Commands:
  -L [table] [options]          List conntrack or expectation table
  -G [table] parameters         Get conntrack or expectation
  -D [table] parameters         Delete conntrack or expectation
  -I [table] parameters         Create a conntrack or expectation
  -U [table] parameters         Update a conntrack
  -E [table] [options]          Show events
  -F [table]                    Flush table
  -C [table]                    Show counter
  -S                            Show statistics

Tables: conntrack, expect, dying, unconfirmed

Conntrack parameters and options:
  -n, --src-nat ip                      source NAT ip
  -g, --dst-nat ip                      destination NAT ip
  -j, --any-nat ip                      source or destination NAT ip
  -m, --mark mark                       Set mark
  -c, --secmark secmark                 Set selinux secmark
  -e, --event-mask eventmask            Event mask, eg. NEW,DESTROY
  -z, --zero                            Zero counters while listing
  -o, --output type[,...]               Output format, eg. xml
  -l, --label label[,...]               conntrack labels

Expectation parameters and options:
  --tuple-src ip        Source address in expect tuple
  --tuple-dst ip        Destination address in expect tuple

Updating parameters and options:
  --label-add label     Add label
  --label-del label     Delete label

Common parameters and options:
  -s, --src, --orig-src ip              Source address from original direction
  -d, --dst, --orig-dst ip              Destination address from original direct
ion
  -r, --reply-src ip            Source addres from reply direction
  -q, --reply-dst ip            Destination address from reply direction
  -p, --protonum proto          Layer 4 Protocol, eg. 'tcp'
  -f, --family proto            Layer 3 Protocol, eg. 'ipv6'
  -t, --timeout timeout         Set timeout
  -u, --status status           Set status, eg. ASSURED
  -w, --zone value              Set conntrack zone
  --orig-zone value             Set zone for original direction
  --reply-zone value            Set zone for reply direction
  -b, --buffer-size             Netlink socket buffer size
  --mask-src ip                 Source mask address
  --mask-dst ip                 Destination mask address
```

Если пакет не установлен, установите его с помощью команды:

- **для Red Hat Enterprise Linux, Oracle Linux и РЕД ОС**

```
yum install conntrack
```

- **для ОС Astra Linux**

```
apt-get install conntrack
```

- **для ОС Альт**

```
apt-get install conntrack-tools
```

- **для Мос.ОС**

```
zypper install conntrack-tools
```

Если пакет недоступен в репозитории, скачайте его самостоятельно, например, из официального репозитория РЕД ОС. Затем перейдите в директорию с пакетом и установите его вручную с помощью команды вида:

- **для Red Hat Enterprise Linux, Oracle Linux, РЕД ОС и ОС Альт**

```
rpm -Uhv ./<имя_пакета>
```

- **для ОС Astra Linux**

```
dpkg -i ./<имя_пакета>
```

- **для Мос.ОС**

```
zypper install <имя_пакета>
```

- `socat` – аналогично пакету `conntrack`.

7. Если вы используете антивирус, настройте его для совместной работы с Data Analysis Service (см. статью Базы знаний *Настройка антивируса для продуктов Платформы*).
8. **На сервере под управлением Мос.ОС:**

   Установите пакеты `ethtool`, `app_armor` и `sysctl` с зависимостями:

```
zypper install ethtool
zypper install patterns-base-apparmor
zypper install procps
zypper install libprocps7
```

9. **На сервере под управлением Red Hat Enterprise Linux и РЕД ОС настройте SELinux.**

   Для корректной работы используемого `containerd` **обновите** политики SELinux или **отключите** систему контроля доступа:

   - **Чтобы обновить политики SELinux**, скачайте и установите пакет `container-selinux`.
     Вы можете сделать это, выполнив команду:

```
yum install container-selinux
```

     > **Примечание:**
     >
     > Пакет не входит в состав дистрибутива продукта.
     >
     > Пакет может отсутствовать в репозиториях. В этом случае загрузите и установите его вручную.

   - **Чтобы отключить SELinux**:
     a. Выполните команды от имени суперпользователя `root`:

```
setenforce 0
sed -i 's/SELINUX=enforcing/SELINUX=disabled/g' /etc/selinux/config
```

b. Перезагрузите сервер с помощью команды:

```
reboot
```

c. Проверьте статус системы SELinux с помощью команды:

```
sestatus
```

10. **Если вы устанавливаете Data Analysis Service на ОС Альт СП 10 во внешний кластер kubernetes, выполните следующие действия:**

a. В файле `/etc/apt/sources.list.d/altsp.list` раскомментируйте три последних строки, указывающие репозитории.

Пример содержимого altsp.list с раскомментированными строками

```
# update.altsp.su (IVK, Moscow)

# ALT Certified 10
#rpm [cert8] ftp://update.altsp.su/pub/distributions/ALTLinux c10f/branch/x86_64
classic gostcrypto
#rpm [cert8] ftp://update.altsp.su/pub/distributions/ALTLinux c10f/branch/x86_64-
i586 classic
#rpm [cert8] ftp://update.altsp.su/pub/distributions/ALTLinux c10f/branch/noarch
classic

rpm [cert8] http://update.altsp.su/pub/distributions/ALTLinux c10f/branch/x86_64
classic gostcrypto
rpm [cert8] http://update.altsp.su/pub/distributions/ALTLinux c10f/branch/x86_64-
i586 classic
rpm [cert8] http://update.altsp.su/pub/distributions/ALTLinux c10f/branch/noarch
classic
```

b. Обновите пакеты с помощью команды:

```
apt-get update
```

c. Установите требуемые пакеты с помощью команд:

```
apt-get install python3-module-yaml
apt-get install python3-modules-sqlite3
apt-get install conntrack-tools
```

d. Установите менеджер контейнеров Podman с помощью команды:

```
apt-get install podman
```

e. С помощью Podman загрузите образы:

```
podman pull registry.altlinux.org/k8s-c10f2/kube-apiserver:v1.31.1
podman pull registry.altlinux.org/k8s-c10f2/kube-controller-manager:v1.31.1
podman pull registry.altlinux.org/k8s-c10f2/kube-scheduler:v1.31.1
podman pull registry.altlinux.org/k8s-c10f2/kube-proxy:v1.31.1
podman pull registry.altlinux.org/k8s-c10f2/pause:3.10
podman pull registry.altlinux.org/k8s-c10f2/etcd:3.5.15-0
podman pull registry.altlinux.org/k8s-c10f2/coredns:v1.11.3
```

f. Для присвоения имени образу выполните команду:

```
podman tag registry.altlinux.org/k8s-c10f2/pause:3.10 registry.k8s.io/pause:3.10
```

g. Установите Kubernetes с помощью команд:

```
apt-get install kubernetes-kubeadm kubernetes-kubelet kubernetes-crio cri-tools
systemctl enable --now crio kubelet
swapoff -a
```

h. Для корректного создания кластера отключите использование прокси-соединения:

```
unset http_proxy
unset https_proxy
```

i. Создайте кластер:

```
kubeadm init --pod-network-cidr=10.244.0.0/16 --kubernetes-version=1.31.1 --
image-repository=registry.altlinux.org/k8s-c10f2
```

j. После создания кластера на экран будет выведена инструкция. Выполните предложенные действия.

> **Примечание:**
>
> **Рекомендуется** выполнить действия из инструкции.
> Если вы не будете выполнять все действия из инструкции, чтобы продолжить установку продукта, выполните команду для настройки переменной окружения:

```
export KUBECONFIG=/etc/kubernetes/admin.conf
```

Команда позволит использовать `kubectl` в рамках текущей сессии подключения. При следующем подключении или перезагрузке выполните команду повторно.

k. Установите для Kubernetes сетевой провайдер Flannel, для этого загрузите образы:

```
podman pull registry.altlinux.org/k8s-c10f2/flannel:v0.25.7
podman pull registry.altlinux.org/k8s-c10f2/flannel-cni-plugin:v1.4.0-flannel1
```

l. Загрузите манифест Flannel:

```
wget https://github.com/flannel-io/flannel/releases/latest/download/kube-
flannel.yml
```

m. Отредактируйте файл-манифест:
   i. Строку `image: docker.io/flannel/flannel-cni-plugin:v1.1.2` замените на:

Строка для замены:
`image: registry.altlinux.org/k8s-c10f2/flannel-cni-plugin:v1.4.0-flannel1`

   ii. Строки `image: docker.io/flannel/flannel:v0.22.0` замените на:

Строка для замены:
`image: registry.altlinux.org/k8s-c10f2/flannel:v0.25.7`

> **Примечание:**
>
> В конфигурационном файле две одинаковых строки вида image: `docker.io/flannel/flannel:v0.22.0` . Замените обе.
>
> В связи с периодическими обновлениями Flannel версии в заменяемых строках могут отличаться.

n. Откройте на редактирование конфигурацию CoreDNS:

```
kubectl edit cm -n kube-system coredns
```

o. Добавьте следующее содержимое:

```
     rewrite stop {
        name substring iwplatform cluster answer auto
     }
```

Вид конфигурационного файла после редактирования:

```
  Corefile: |
    .:53 {
        log
        errors
        health {
           lameduck 5s
        }
        rewrite stop {
           name substring iwplatform cluster answer auto
        }
        ready
        kubernetes cluster.local in-addr.arpa ip6.arpa {
           pods insecure
           fallthrough in-addr.arpa ip6.arpa
           ttl 30
        }
        prometheus :9153
        forward . /etc/resolv.conf {
           max_concurrent 1000
        }
        cache 30
        loop
        reload
        loadbalance
    }
```

p. Сохраните изменения и закройте редактор.
q. Примените изменения в манифесте и запустите Flannel с помощью команды:

```
kubectl apply -f kube-flannel.yml
```

r. Дождитесь запуска всех подов.
Вы можете проверить их статусы с помощью команды:

```
kubectl get pods -A
```

Все поды должны быть в статусе `Running`.
s. Смените директорию для временных файлов:

```
export TMPDIR=/var/tmp
```

11. Для установки и корректной работы Data Analysis Service в ОС Astra Linux Special Edition убедитесь, что настройка безопасности *astra-nochmodx-lock* отключена.
    Вы можете отключить эту настройку с помощью команды:

    ```
    astra-nochmodx-lock disable
    ```

12. Убедитесь, что на сервере установлены актуальные дата и время.

7 Настройка сетевых правил доступа

Для корректной работы должны быть разрешены соединения:

Соединение	Порт	Описание
Соединения с внешними системами		
Сервер Traffic Monitor → Сервер Data Analysis Service	TCP 30050	Порт сервера Data Analysis Service. Используется для обмена данными с Traffic Monitor по протоколу HTTP.
	TCP 30450	Порт сервера Data Analysis Service. Используется для обмена данными с Traffic Monitor по протоколу HTTPS.
Сервер Data Analysis Service → Сервер Traffic Monitor	TCP 443	Порт сервера Data Analysis Service. Используется для обмена данными с Traffic Monitor по протоколу HTTPS.
Работа в кластере		
Сервер Платформы в режиме `office` → Сервер Платформы в режиме `central`	TCP 6443 TCP 2379, 2380 TCP 10251, 10252, 10255	Порты сервера Платформы в режиме `central`. Используются для взаимодействия Платформы в режиме `office` с Платформой в режиме `central`
Сервер Платформы в режиме `central` → Сервер Платформы в режиме `office`	TCP 10251, 10255 TCP 30000–32767	Порты сервера Платформы в режиме `office`. Используются для взаимодействия Платформы в режиме `central` с Платформой в режиме `office`
Сервер Платформы в режиме `central` ← → Сервер Платформы в режиме `office`	UDP 8472 TCP 10250	Используются для туннелирования трафика между узлами кластера
Сервер Платформы в режиме `central` ← → Сервер Платформы в режиме `office`	TCP 10256	Порт должен быть доступен на всех нодах кластера. Используется для проверки состояния компонента `kube-proxy`

8 Установка Data Analysis Service

Для установки Сервиса используется архив `iw_das_setup_x.x.x.xx.tar.xz`, где x.x.x.xx – номер версии. Архив входит в состав дистрибутива. В нашем примере будет использована версия `1.7.0.98`, ей соответствует архив `iw_das_setup _1.7.0.98.tar.xz`. Архив содержит набор бинарных модулей образов контейнеров, необходимых для развертывания Сервиса без доступа к сети Интернет. Также архив содержит программу установки.

Для установки и работы Сервиса вам также понадобятся ресурсы и модели для тех технологий, которые вы собираетесь использовать. Они входят в состав дистрибутива. Их список вы можете найти в статье *Установка моделей*.

Установка Data Analysis Service производится вручную, с помощью командной строки.

Вы можете установить Data Analysis Service:

- на отдельном сервере;
- в кластере из серверов Data Analysis Service;
- на сервере совместно с продуктами InfoWatch, указанными в статье *Типы установки и обновления, совместимость с другими продуктами, аппаратные и программные требования*;
- в кластере с платформенными продуктами.

Важно!

Если вы устанавливаете Data Analysis Service в одном кластере с другими продуктами InfoWatch, то убедитесь, что версия Платформы совпадает у всех продуктов на всех нодах. В противном случае продукты, установленные на старой версии Платформы, необходимо обновить.

Чтобы узнать версию Платформы устанавливаемого Сервиса:

1. Распакуйте архив с дистрибутивом Data Analysis Service. Для этого перейдите в директорию с архивом и выполните команду:

   Пример для версии 1.7.0.98:

   ```
   tar -xvf iw_das_setup_1.7.0.98.tar.xz
   ```

2. Выполните команду:

   ```
   cat platform/setup.yaml | grep version
   ```

Важно!

Необходимо, чтобы все ноды кластера были доступны друг для друга в сети. Для этого должны быть открыты необходимые порты (см. *Настройка сетевых правил доступа*).

Также перед установкой убедитесь, что:

- настроено разрешение имен нод средствами DNS
  или

- в файле `/etc/hosts` каждой ноды определено соответствие между адресами и именами нод.

Для проверки доступности нод выполните команду между нодами кластера:

```
ping <имя ноды>
```

8.1 Варианты запуска установки Data Analysis Service

Вы можете запустить установку Data Analysis Service:

- **в интерактивном режиме** - инсталлятор последовательно запросит параметры установки (см. *Установка Data Analysis Service в интерактивном режиме*);
- **в тихом режиме** - указав параметры установки в команде запуска инсталлятора (см. *Установка Data Analysis Service с указанием параметров в командной строке*).

В обоих вариантах запуска инсталлятор поддерживает и совместную установку с платформенными продуктами на одном сервере, и распределенную установку в кластере.

8.2 Установка Data Analysis Service в интерактивном режиме

Примечание:

Если вы устанавливаете Data Analysis Service на сервер, на котором уже установлены какие-либо платформенные продукты, инсталлятор не запросит часть параметров, так как эти директории и настройки уже были указаны при первичной установке Платформы.

1. Проверьте, соответствует ли сервер аппаратным и программным требованиям (см. *Типы установки и обновления, совместимость с другими продуктами, аппаратные и программные требования*).
2. Выполните шаги по подготовке сервера к установке Data Analysis Service (см. *Подготовка сервера*).
3. Создайте новую директорию на сервере (например, `DAS`):

   ```
   mkdir DAS
   ```

4. Скопируйте архив **`iw_das_setup_1.7.x.xx.tar.xz`** в созданную директорию.
   В нашем примере будет использована версия `1.7.0.98`, соответственно, архив – `iw_das_setup _1.7.0.98.tar.xz`.
5. Перейдите в директорию, в которую был скопирован архив.
6. Распакуйте архив с дистрибутивом в эту директорию:

   ```
   tar -xvf iw_das_setup_1.7.0.98.tar.xz
   ```

7. **Для установки Data Analysis Service на ОС Альт СП 10 во внешний кластер kubernetes:**

Вы должны использовать Kubernetes из репозиториев ОС.
Для этого:

a. Выполните действия для ОС Альт СП 10 из статьи *Подготовка сервера*.
   Если вы уже выполнили подготовку, пропустите этот шаг.
b. Выполните команду:

```
kubectl taint nodes <hostname_ноды> node-role.kubernetes.io/control-plane:NoSched
ule-
```

c. На следующем шаге при запуске инсталлятора дополнительно укажите опцию `--externalK8S=True`

8. Запустите программу установки с помощью команды вида:

```
./setup.py install --nodemode=<режим_установки_Платформы> --
productmode=<режим_установки_продукта>
```

Примечание:

Чтобы установить:

- **на центральную ноду Платформы Data Analysis Service в режиме central:**

```
./setup.py install --nodemode=central --productmode=central
```

- **на центральную ноду Платформы Data Analysis Service в режиме office:**

```
./setup.py install --nodemode=central --productmode=office
```

- **на офисную ноду Платформы Data Analysis Service в режиме central:**

```
./setup.py install --nodemode=office --productmode=central
```

- **на офисную ноду Платформы Data Analysis Service в режиме office:**

```
./setup.py install --nodemode=office --productmode=office
```

Если вы не укажете режимы установки Платформы и Data Analysis Service, инсталлятор по умолчанию выберет режим `central`.

Если вы устанавливаете Data Analysis Service на сервер первым и укажете только параметр `--nodemode`, инсталлятор использует такое же значение и для параметра `--productmode`.

> **Важно!**
>
> Во время установки могут быть недоступны продукты Платформы, установленные на сервере.

9. Ознакомьтесь с условиями лицензионного соглашения. Лицензионное соглашение содержит несколько страниц. Для перехода на следующую страницу используйте клавишу **Enter**.
10. Чтобы принять лицензионное соглашение, введите "**y**" и нажмите **Enter**.

> **Примечание:**
>
> Если отклонить лицензионное соглашение, инсталлятор прервет процесс и завершит работу.

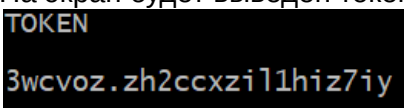
11. Если вы устанавливаете **центральную ноду Платформы** (`--nodemode=central`), введите IP-адрес сетевого интерфейса для взаимодействия с кластером в формате IPv4: "xxx.xxx.xxx.xxx" (по умолчанию: `0.0.0.0`) и нажмите **Enter**.
Если указать `0.0.0.0` , будут использованы все доступные сетевые интерфейсы.

> **Примечание:**
>
> Здесь и далее для использования значений, предложенных по умолчанию, нажмите **Enter** без ввода значений.

12. Если вы устанавливаете **офисную ноду Платформы** (`--nodemode=office`):
    a. Укажите адрес центральной ноды Платформы в формате IP-адреса или доменного имени (например: `192.0.2.0` или `host.example.com`).
    b. Укажите токен для подключения к центральной ноде Платформы.
    Чтобы получить токен для подключения к центральной ноде Платформы:
        i. Подключитесь по ssh к консоли центральной ноды Платформы;
        ii. Выполните в терминале команду для получения токена:

```
kubeadm token list
```

        На экран будет выведен токен:

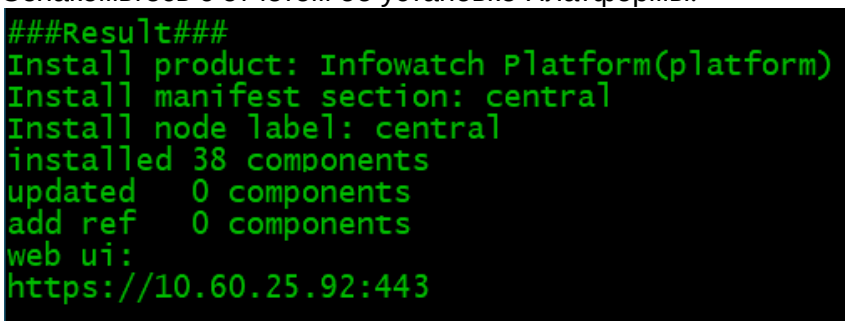


        Токен в нашем примере: `3wcvoz.zh2ccxzil1hiz7iy`
        iii. Укажите лейбл ноды (по умолчанию: `office`) и нажмите **Enter**.
        Лейбл может состоять из букв латинского алфавита и/или цифр. Длина лейбла не должна превышать 32 символа.
13. Выделите объем оперативной памяти для размещения данных Clickhouse (по умолчанию: `80%`) и нажмите **Enter**.
14. Укажите путь для размещения данных Clickhouse (по умолчанию: `/opt/chdata`) и нажмите **Enter**.

15. Укажите путь для размещения данных PostgreSQL (по умолчанию: `/opt/pgdata`) и нажмите **Enter**.
16. Укажите путь для размещения данных NATS (по умолчанию: `/opt/natsdata`) и нажмите **Enter**.
17. Укажите путь для размещения бинарных данных (по умолчанию: `/opt/dsdata`) и нажмите **Enter**.
18. Укажите путь для временного размещения событий Traffic Monitor перед их загрузкой в Платформу (по умолчанию: `/opt/factsldrtmdata`) и нажмите **Enter**.
19. Если вы устанавливаете **центральную ноду Платформы**, укажите порт подключения к веб-интерфейсу (по умолчанию: `443`) и нажмите **Enter**.

> **Важно!**
>
> При установке на один сервер с Traffic Monitor укажите порт отличный от 443, так как по умолчанию порт 443 используется в качестве порта подключения к веб-интерфейсу Traffic Monitor.

20. Дождитесь завершения установки Платформы.
21. Ознакомьтесь с отчетом об установке Платформы.



22. Введите через запятую лейблы нод для взаимодействия с платформенными продуктами (по умолчанию задается одна нода с лейблом `central`) и нажмите **Enter**. Data Analysis Service будет обрабатывать очередь сообщений каждой указанной ноды. Если такое взаимодействие не планируется, нажмите **Enter**.

> **Примечание:**
>
> InfoWatch Activity Monitor отправляет сообщения в очередь NATS на центральную ноду кластера.
>
> InfoWatch Data Discovery отправляет сообщения в очередь NATS на ту ноду кластера, на которой он установлен.

Чтобы узнать лейбл ноды:

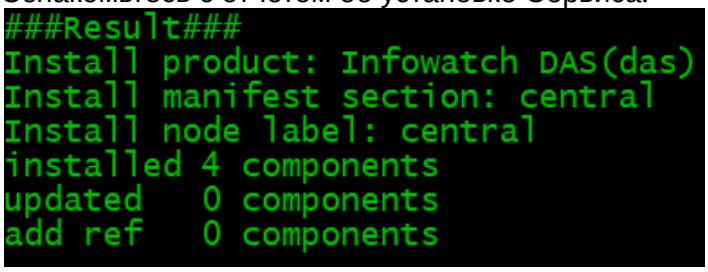
Выполните команду:

```
kubectl get node -n infowatch -o jsonpath="{range .items[*]}
{.metadata.annotations.flannel\.alpha\.coreos\.com/public\-ip}
{.metadata.labels.kubernetes\.io/hostname} {.metadata.labels.iwnode}{'\n'}{end}"
```

В результате будет выведен список нод кластера с указанием их лейблов.

> **Примечание:**
>
> Изменить список лейблов нод для взаимодействия можно после установки. Подробнее см. *Задание нод обработки*.

23. Укажите путь для размещения ресурсов Data Analysis Service (по умолчанию: `/opt/das_data`) и нажмите **Enter**.
24. Введите "**y**", чтобы включить модуль детектирования объектов на изображениях, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.
25. Введите "**y**", чтобы включить модуль распознавания именованных сущностей, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.
26. Введите "**y**", чтобы включить модуль OCR, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.
27. Введите "**y**", чтобы включить модуль протокластеризации, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.
28. Введите "**y**", чтобы включить модуль транскрибации, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.
29. Дождитесь окончания процесса установки Data Analysis Service.
30. Введите "**y**", чтобы выполнить миграцию значений переменных окружения, или "**n**", чтобы не выполнять, (по умолчанию: "**n**") и нажмите **Enter**.
    Если файл со значениями переменных не обнаружен в директории `/var/tmp/das_env_migration/raw/`, то этот шаг будет пропущен.
31. Ознакомьтесь с отчетом об установке Сервиса:



32. Чтобы вывести на экран информацию о Платформе и продуктах в кластере, выполните команду:

```
./setup.py showproducts
```

Вывод команды должен содержать строку `DbState: commited`. Например:

```
platform Title: Infowatch Platform Version: 2.6.0.333 Node: my-server(central) Mode:
central DbState: commited
```

33. Убедитесь, что все сервисы запущены, выполнив команду:

```
kubectl get pods -n infowatch
```

Команда выведет список сервисов. Каждый сервис должен иметь статус `Running`.

34. При установке Data Analysis Service на сервер, где уже установлен Traffic Monitor, убедитесь, что в конфигурационном файле `/opt/iw/tm5/etc/consul/consul.json` в поле "`bind_addr`" указан IP-адрес данного сервера.
35. Если до установки Data Analysis Service в кластере был установлен Device Monitor, обновите конфигурационный файл `n.yaml`:
    a. Подключитесь к центральной ноде Платформы.
    b. Если Device Monitor установлен на офисной ноде Платформы, скопируйте на центральную его дистрибутив и распакуйте.
    c. На центральной ноде Платформы перейдите в папку, где распакован дистрибутив Device Monitor.
    d. Запустите скрипт **setup.sh**, выполнив команду:

    ```
    ./setup.sh
    ```

    e. Выберите в меню пункт **Utilities**.
    f. Выберите **Fix n.yaml file**.
    g. Введите IP-адрес сервера Device Monitor и дождитесь окончания процесса.

    > **Примечание:**
    >
    > Если вы устанавливаете несколько продуктов в кластере, обновите **`n.yaml`** только после установки последнего продукта.

36. Установите модели для технологий, которые будете использовать. Подробнее см. статью *Установка моделей*.

    > **Важно!**
    >
    > Если модели для какой-либо технологии не будут установлены, то эта технология не будет работать.

Чтобы начать использование Data Analysis Service, настройте его взаимодействие с продуктом InfoWatch:

- с **InfoWatch Traffic Monitor** (подробнее см. "*InfoWatch Traffic Monitor. Руководство администратора. Настройка интеграции с Data Analysis Service*")
  и/или
- с продуктами Платформы, например, **InfoWatch Activity Monitor** или **InfoWatch Data Discovery** (подробнее см. в статье *Настройка взаимодействия с продуктами Платформы* и в документации на соответствующий продукт Платформы).

> **Важно!**
>
> При установке **Data Analysis Service 1.7.x, начиная с 1.7.1**, для взаимодействия с **InfoWatch Traffic Monitor версии 7.15.2 и ниже**, а также **7.16.0-7.16.1**, для каждого экземпляра Сервиса выполните команду:

```
kubectl set env deploy -n infowatch das-adapter-<node_label> AUTH_REQUIRED=False
```

где `<node_label>` – это лейбл ноды, на которую установлен экземпляр Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

8.3 Установка Data Analysis Service с указанием параметров в командной строке

Инсталлятор Data Analysis Service позволяет установить Сервис с минимальным участием пользователя. Для этого запустите программу установки, указав в командной строке все требуемые параметры.

Перед запуском инсталлятора Data Analysis Service:

1. Проверьте, соответствует ли сервер аппаратным и программным требованиям (см. *Типы установки и обновления, совместимость с другими продуктами, аппаратные и программные требования*).
2. Выполните шаги по подготовке сервера к установке Data Analysis Service (см. *Подготовка сервера*).
3. Создайте новую директорию на диске (например, `DAS`):

   ```
   mkdir DAS
   ```

4. Скопируйте архив с дистрибутивом Data Analysis Service **`iw_das_setup_1.7.x.xx.tar.xz`** в созданную директорию.
   В нашем примере будет использована версия `1.7.0.98`, соответственно, архив – `iw_das_setup_1.7.0.98.tar.xz`.
5. Перейдите в директорию, в которую был скопирован архив.
6. Распакуйте архив в эту директорию:

   ```
   tar -xvf iw_das_setup_1.7.0.98.tar.xz
   ```

Ознакомьтесь со списком ключей в таблице ниже, с их помощью вы можете указать инсталлятору параметры установки продукта.

Ключ	Описание
`--nodemode`	Режим установки Платформы **Доступные значения:** • `central` (по умолчанию) • `office`
`--productmode`	Режим установки продукта **Доступные значения:**

Ключ	Описание
	• `central` (по умолчанию) • `office`
`--licagree`	Принятие лицензионного соглашения **Доступные значения:** • `no` • `y es`
`--chmemsize`	Объем оперативной памяти, используемой для данных ClickHouse **Доступные значения:** • `1` - `100%` или числовое значение. Если не указать `%`, указанное значение будет принято за объем в ГБ
`--disable-check-free-disk`	Отключение проверки свободного места на диске: • 20 ГБ для размещения образов; • 100 ГБ для размещения данных в указанных директориях. **Примечание:** Если директории для хранения данных находятся на разных дисках, то проверятся наличие 100 ГБ свободного места на каждом из них.
`--disable-check-total-ram`	Отключение проверки объема ОЗУ на компьютере в 8 ГБ
`--chdbpath`	Путь для размещения данных ClickHouse
`--natspath`	Путь для размещения данных NATS
`--pgdbpath`	Путь для размещения файлов базы данных PostgreSQL
`--dsdatapath`	Путь для размещения бинарных данных
`--factsldrtmdatapath`	Путь для временного размещения событий Traffic Monitor перед их загрузкой в Платформу
`--advertise`	IP-адрес сетевого интерфейса, на котором работает сервис `kubelet`. Укажите значение при установке центральной ноды Платформы.
`--guiport`	Порт подключения к веб-интерфейсу. Укажите значение при установке центральной ноды Платформы. **Важно!**

Ключ	Описание
	При установке на один сервер с Traffic Monitor указывайте порт отличный от 443, так как по умолчанию порт 443 используется в качестве порта подключения к веб-интерфейсу Traffic Monitor.
`--cluster`	Адрес центральной ноды Платформы в кластере. Укажите значение при установке офисной ноды Платформы.
`--clustertoken`	Токен для подключения к центральной ноде Платформы. Чтобы получить токен, в терминале центральной ноды Платформы выполните команду: `kubeadm token list` Укажите значение при установке офисной ноды Платформы.
`--nodelabel`	Лейбл ноды Платформы; уникален в пределах кластера. Значение по умолчанию для офисной ноды – `office`. Если в кластере более одной офисной ноды, необходимо задать уникальные лейблы.
`--proto_clusterization_enable`	Включение модуля протокластеризации **Доступные значения:** • `n` • `y`
`--transcription_enable`	Включение модуля транскрибации **Доступные значения:** • `n` • `y`
`--optical_character_recognition_enable`	Включение модуля OCR **Доступные значения:** • `n` • `y`
`--image_content_detection_enable`	Включение модуля детектирования объектов на изображениях **Доступные значения:** • `n` • `y`
`--named_entities_r`	Включение модуля распознавания именованных сущностей **Доступные значения:**

Ключ	Описание
`ecognition_enable`	• `n` • `y`
`--subscribe-to`	Список лейблов нод для взаимодействия с платформенными продуктами. Если таких нод несколько, укажите их лейблы через запятую, взяв весь список в двойные кавычки, например, `--subscribe-to="central, office1, office2"`. Data Analysis Service будет обрабатывать очередь сообщений каждой указанной ноды. **Примечание:** InfoWatch Activity Monitor отправляет сообщения в очередь NATS на центральную ноду кластера. InfoWatch Data Discovery отправляет сообщения в очередь NATS на ту ноду кластера, на которой он установлен. **Чтобы узнать лейбл ноды:** На любой ноде кластера выполните следующую команду: `kubectl get node -n infowatch -o jsonpath="{range .items[*]} {.metadata.annotations.flannel\.alpha\.coreos\.com/public\-ip} {.metadata.labels.kubernetes\.io/hostname} {.metadata.labels.iwnode} {'\n'}{end}"` В результате будет выведен список нод кластера с указанием их лейблов. **Примечание:** Изменить список лейблов ноды для взаимодействия можно после установки. Подробнее см. *Задание нод обработки*.
`--das-data-path`	Путь для размещения ресурсов Data Analysis Service
`--migrate-env`	Выполнить миграцию значений переменных окружения. Файл с ними должен быть предварительно сформирован скриптом `migrate_env.sh` или инсталлятором и располагаться на сервере в директории `/var/tmp/das_env_migration/raw/`. **Доступные значения:** • `n` • `y`

Чтобы установить Data Analysis Service:

1. **Для установки Data Analysis Service на ОС Альт СП 10 во внешний кластер kubernetes:**

   Вы должны использовать Kubernetes из репозиториев ОС.
   Для этого:

   a. Выполните действия для ОС Альт СП 10 из статьи *Подготовка сервера*. Если вы уже выполнили подготовку, пропустите этот шаг.
   b. Выполните команду:

```
kubectl taint nodes <hostname_ноды> node-role.kubernetes.io/control-plane:NoSched
ule-
```

c. На следующем шаге при запуске инсталлятора дополнительно укажите опцию `--externalK8S=True`

2. Запустите программу установки Data Analysis Service с указанием параметров.
Пример команды установки центральной ноды Платформы и Data Analysis Service в режиме central:

```
./setup.py install \
--nodemode=central \
--productmode=central \
--licagree=yes \
--advertise=0.0.0.0 \
--chmemsize=60% \
--chdbpath=/opt/chdata \
--pgdbpath=/opt/pgdata \
--natspath=/opt/natsdata \
--dsdatapath=/opt/dsdata \
--factsldrtmdatapath=/opt/factsldrtmdata \
--guiport=443 \
--proto_clusterization_enable=y \
--transcription_enable=y \
--image_content_detection_enable=y \
--named_entities_recognition_enable=y \
--optical_character_recognition_enable=y \
--subscribe-to=central \
--das-data-path=/opt/das_data
```

Пример команды установки офисной ноды Платформы и Data Analysis Service в режиме office:

```
./setup.py install \
--nodemode=office \
--productmode=office \
--licagree=yes \
--cluster=10.60.25.92 \
--clustertoken=3wcvoz.zh2ccxzil1hiz7iy \
--nodelabel=office123
--chmemsize=60% \
--chdbpath=/opt/chdata \
--pgdbpath=/opt/pgdata \
--natspath=/opt/natsdata \
--dsdatapath=/opt/dsdata \
--factsldrtmdatapath=/opt/factsldrtmdata \
--proto_clusterization_enable=y \
--transcription_enable=n \
```

```
--image_content_detection_enable=n \
--named_entities_recognition_enable=n \
--optical_character_recognition_enable=n \
--subscribe-to=central \
--das-data-path=/opt/das_data
```

Пример команды установки Data Analysis Service в режиме central на сервер с установленной Платформой:

```
./setup.py install \
--productmode=central \
--licagree=yes \
--proto_clusterization_enable=y \
--transcription_enable=n \
--image_content_detection_enable=n \
--named_entities_recognition_enable=n \
--optical_character_recognition_enable=n \
--subscribe-to="central, office1, office2" \
--das-data-path=/opt/das_data
```

Если вы не указали часть параметров:

- для параметров `--nodemode`, `--productmode` будут использованы значения по умолчанию;
- для остальных обязательных параметров инсталлятор запустит интерактивный режим и попросит задать значения. Подробнее см. *Установка Data Analysis Service в интерактивном режиме*.

3. Дождитесь завершения установки Data Analysis Service.
4. Чтобы вывести на экран информацию о Платформе и продуктах в кластере, выполните команду:

```
./setup.py showproducts
```

5. Убедитесь что все сервисы запущены, выполнив команду:

```
kubectl get pods -n infowatch
```

Команда выведет список сервисов. Каждый сервис должен иметь статус `Running`.

6. При установке Data Analysis Service на сервер, где уже установлен Traffic Monitor, убедитесь, что в конфигурационном файле `/opt/iw/tm5/etc/consul/consul.json` в поле "`bind_addr`" указан IP-адрес данного сервера.
7. Если до установки Data Analysis Service в кластере был установлен Device Monitor, обновите конфигурационный файл `n.yaml`:
   a. Подключитесь к центральной ноде Платформы.
   b. Если Device Monitor установлен на офисной ноде Платформы, скопируйте на центральную его дистрибутив и распакуйте.
   c. На центральной ноде Платформы перейдите в папку, где распакован дистрибутив Device Monitor.
   d. Запустите скрипт **setup.sh**, выполнив команду:

```
./setup.sh
```

e. Выберите в меню пункт **Utilities**.
f. Выберите **Fix n.yaml file**.
g. Введите IP-адрес сервера Device Monitor и дождитесь окончания процесса.

> **Примечание:**
>
> Если вы устанавливаете несколько продуктов в кластере, обновите `n.yaml` только после установки последнего продукта.

8. Установите модели для технологий, которые будете использовать. Подробнее см. статью *Установка моделей*.

> **Важно!**
>
> Если модели для какой-либо технологии не будут установлены, то эта технология не будет работать.

Чтобы начать использование Data Analysis Service, настройте его взаимодействие с продуктом InfoWatch:

- с **InfoWatch Traffic Monitor** (подробнее см. InfoWatch Traffic Monitor. *Руководство администратора*. *Настройка интеграции с Data Analysis Service*)
  и/или
- с продуктами Платформы, например, **InfoWatch Activity Monitor** или **InfoWatch Data Discovery** (подробнее см. в статье *Настройка взаимодействия с продуктами Платформы* и в документации на соответствующий продукт Платформы).

> **Важно!**
>
> При установке **Data Analysis Service 1.7.x, начиная с 1.7.1**, для взаимодействия с **InfoWatch Traffic Monitor версии 7.15.2 и ниже**, а также **7.16.0-7.16.1**, для каждого экземпляра Сервиса выполните команду:
>
> ```
> kubectl set env deploy -n infowatch das-adapter-<node_label> AUTH_REQUIRED=False
> ```
>
> где `<node_label>` – это лейбл ноды, на которую установлен экземпляр Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

9 Установка моделей

Для работы технологий Data Analysis Service должны быть установлены модели:

- для протокластеризации – `proto_clusterization_resources-<version>.tar.gz`;
- для транскрибации – `transcription_resources-<version>.tar.gz`;
- для детектирования объектов на изображениях – `image_resources-<version>.tar.gz`;
- для распознавания именованных сущностей – `named_entities_recognition_resources-<version>.tar.gz`;
- для распознавания текста на изображениях (OCR) – `optical_character_recognition_resources-<version>.tar.gz`.

Чтобы установить модели:

1. Скопируйте из дистрибутива архивы с нужными вам моделями на сервер, например, в директорию `../DAS/tools/models_installation`.
2. Перейдите в директорию `../DAS/tools/models_installation`, где располагается скрипт распаковки архивов с моделями.
3. Убедитесь, что на диске с директорией `/tmp` есть 15 ГБ свободного места. Это необходимо для успешной распаковки архивов.
4. Запустите скрипт распаковки архивов `deploy_technology_resources.sh` с помощью команды:

   ```
   ./deploy_technology_resources.sh [-n <namespace>] [-i <input_directory>] [--node
   <node_label>] [--transcription] [--proto-clusterization] [--image-content-detection] [--
   named-entity-recognition] [--optical-character-recognition]
   ```

   со следующими аргументами:
   `[-n <namespace>]` – пространство имен Kubernetes. Если аргумент не указан, значение берется из текущего Kubernetes context. Рекомендуется указывать аргумент, значение аргумента должно быть равным `infowatch`.
   `[-i <input_directory>]` – директория, в которой находятся архивы моделей. Если аргумент не указан, то будут распакованы архивы из той директории, где находится скрипт;
   `[--node <node_label>]` – лейбл ноды, необходимый для определения имени POD `das-processor-<node_label>`. По умолчанию имеет значение `central`. Указывайте этот параметр при установке на офисной ноде;
   `[--transcription]` – распаковать модели для транскрибации;
   `[--proto-clusterization]` – распаковать модели для протокластеризации;
   `[--image-content-detection]` – распаковать модели для детектирования объектов на изображениях;
   `[--named-entity-recognition]` – распаковать модели для распознавания именованных сущностей;
   `[--optical-character-recognition]` – распаковать модели для OCR.
   Если вы не укажете параметрами, для каких технологий нужно распаковать модели, далее скрипт интерактивно спросит вас об этом.

   > **Примечание:**
   >
   > Справку по использованию скрипта вы можете получить, выполнив команду:

```
./deploy_technology_resources.sh --help
```

Примеры запуска скрипта

Пример 1

Data Analysis Service установлен на центральной ноде кластера. Предполагается использование только технологии протокластеризации. Модель для нее скопирована в директорию `/root/das/tools/models_installation`, где располагается скрипт распаковки архивов.
Чтобы распаковать модель, выполните команду:

```
./deploy_technology_resources.sh -n infowatch --proto-clusterization
```

Пример 2

Data Analysis Service установлен на центральной ноде. Предполагается использование технологий протокластеризации и детектирования объектов на изображениях. Модели для них скопированы в директорию, отличную от той, где располагается скрипт распаковки архивов, например, в директорию `/root/models`.
Чтобы распаковать модели, выполните команду:

```
./deploy_technology_resources.sh -n infowatch -i /root/models
```

Пример 3

Data Analysis Service установлен на офисной ноде с лейблом "office1". Модели скопированы в директорию `/root/das/tools/models_installation`, где располагается скрипт распаковки архивов. Предполагается использование технологии транскрибации.
Для распаковки выполните команду:

```
./deploy_technology_resources.sh -n infowatch --node office1 --transcription
```

5. Если вы не указали на предыдущем шаге, для каких технологий нужно распаковать модели, то скрипт интерактивно спросит вас об этом.
   Для распаковки моделей нужной вам технологии введите "**y**" и нажмите **Enter**. Для всех остальных нажмите **Enter**.
6. Если в директории назначения (по умолчанию: `/opt/das_data`) ранее уже были распакованы модели, то:
   - Чтобы заменить их на новые, подтвердите удаление старых файлов. Для этого введите "**y**" и нажмите **Enter**.
   - Чтобы загрузить модели, которые ранее не были установлены, а старые модели не удалять, введите "**n**" и нажмите **Enter**.
7. Дождитесь окончания процесса распаковки моделей.

Примечание:

Если для какой-либо выбранной технологии не найден архив с моделями, то выполнение скрипт прервется с ошибкой. Убедитесь, что в директории есть архивы для всех нужных технологий и вы корректно выбрали технологии.

8. Если вы собираетесь использовать большую модель для транскрибации, включите ее использование. Подробнее см. *Настройка модуля транскрибации*.
9. Если вы будете использовать технологию OCR, выберите экстрактор OCR. Подробнее см. *Настройка модуля OCR*.
10. Убедитесь, что необходимые модели установлены. Для этого проверьте логи с помощью команды:

```
kubectl logs $(kubectl get pods -n infowatch -o name|grep das-processor) -n infowatch| grep "Waiting for model"
```

При наличии в кластере нескольких экземпляров Data Analysis Service дополнительно укажите в команде лейбл ноды:

```
kubectl logs $(kubectl get pods -n infowatch -o name|grep das-processor-<node_label>) -n infowatch| grep "Waiting for model"
```

Если модели для включенных технологий установлены, то в результате не будет выведено ни одной строки. Если какая-либо технология включена, но модели для нее не будут найдены, то обработка данных этой технологией происходить не будет. При этом в логах появится сообщение об ошибке вида "`Waiting for model <model name>...`".

Важно!

Если в кластере установлено несколько экземпляров Data Analysis Service и в каком-либо из них включена транскрибация, но не установлены модели для нее, то данный экземпляр Сервиса будет забирать аудиофайлы из очереди, но не будет их обрабатывать. Таким образом, эти файлы будут недоступны для других экземпляров и не будут обработаны.

10 Устранение неполадок

Во время установки возможны сбои: прерывания работы программы установки, отключение электричества и т.д. В таком случае Сервис будет установлен некорректно. В случае возникновения подобной ситуации следует сбросить установку. После сброса все данные сохранятся на сервере, а установку необходимо произвести заново (см. *Установка Data Analysis Service*). В процессе установки на вопрос об удалении существующих данных выберите ответ **Нет**.

Чтобы сбросить установку:

1. Если вы используете технологию протокластеризации в Data Discovery, остановите все задачи сканирования в его веб-интерфейсе. Подробнее см. *InfoWatch Data Discovery. Руководство пользователя*.
2. Запомните или сохраните пути к текущим директориям хранения данных. Получить список директорий можно, выполнив команду:

   ```
   cat /var/lib/iwplatform/data/volumes.txt
   ```

3. Запомните или сохраните лейблы нод. Чтобы посмотреть список лейблов в кластере, выполните команду:

   ```
   kubectl get node -n infowatch -o jsonpath="{range .items[*]}
   {.metadata.annotations.flannel\.alpha\.coreos\.com/public\-ip}
   {.metadata.labels.kubernetes\.io/hostname} {.metadata.labels.iwnode}{'\n'}{end}"
   ```

4. Перейдите в директорию, в которую был распакован архив с дистрибутивом при установке.
5. Выполните команду:

   ```
   ./setup.py reset
   ```

6. После сброса при установке присвойте нодам кластера те же лейблы, которые были до начала процедуры сброса.

Важно!

При сбросе установки Сервиса автоматически производится сброс всех других продуктов Платформы, установленных на сервере.

Если произведен сброс установки на центральной ноде Платформы, станут недоступны компоненты всех продуктов в кластере. В этом случае выполните сброс на каждой офисной ноде.

После установки Data Analysis Service повторно установите и другие продукты в кластере.

11 Обновление Data Analysis Service

Data Analysis Service поддерживает обновление с версий **1.5.x**, **1.6.x** и **1.7.x**.

Обновление Data Analysis Service производится вручную, с помощью командной строки.

Для обновления Сервиса используется архив `iw_das_setup_x.x.x.xx.tar.xz`, где x.x.x.xx – номер версии. Архив входит в состав дистрибутива. Архив содержит набор бинарных модулей образов контейнеров, необходимых для развертывания Сервиса без доступа к сети Интернет. Также архив содержит программу обновления.

Для обновления и работы Сервиса вам также понадобятся ресурсы и модели для тех технологий, которые вы собираетесь использовать. Они входят в состав дистрибутива. Их список вы можете найти в статье *Установка моделей*.

Важно!

Для успешного обновления Data Analysis Service необходимо не менее 150 ГБ свободного места на жестком диске.

Примечание:

Все задачи, активные на момент начала обновления, не будут завершены. Контенты не будут обработаны.
Чтобы проверить, все ли задачи завершены, выполните команду:
`ps -aux|grep das_processor`
Если процесс **das _processor** не использует ресурсы сервера, то активных задач нет.

Если необходимы результаты обработки конкретных контентов, дождитесь , когда эти результаты будут отображены в продукте - источнике контента, например, результаты транскрибации – в InfoWatch Activity Monitor. И только после этого начинайте процедуру обновления.

11.1 Варианты обновления Data Analysis Service

Data Analysis Service 1.7 поддерживает:

- прямое стандартное обновление с версий 1.7.x;
- обновление с помощью сброса текущей установки и переустановки на старые данные.

Примечание:

Если вы обновляетесь с переходом на другой сервер или на новую ОС, используйте обновление со сбросом установки.

Если Data Analysis Service работает штатно, рекомендуется использовать стандартное обновление.

Чтобы **обновить распределенный кластер** Data Analysis Serviceи/или продуктов Платформы:

1. **Обновите центральную ноду** кластера и все установленные на ней продукты.

2. **Обновите офисные ноды** и все установленные на них продукты.

11.2 Особенности обновления кластера Центра расследований

Перед обновлением кластера рекомендуем ознакомиться с инструкциями всех установленных продуктов. Если в кластере установлено несколько продуктов, перед обновлением учтите особенности обновления каждого из них.

Особенности и рекомендации для обновления распределенного кластера Центра расследований:

- Перед обновлением рекомендуем выключить Систему и выполнить ее резервное копирование;
- Для планирования обновления ознакомьтесь с составом кластера. Чтобы отобразить информацию о Платформе и установленных продуктах, перейдите в директорию с инсталлятором и выполните команду:

```
./setup.py showproducts
```

- Если вы обновляете Data Analysis Service в одном кластере с другими продуктами InfoWatch, то убедитесь, что **версия Платформы совпадает у всех продуктов на всех нодах**. В противном случае продукты, установленные на старой версии Платформы, необходимо обновить;
  Чтобы узнать версию Платформы Сервиса:
  1. Распакуйте архив с дистрибутивом Data Analysis Service. Для этого перейдите в директорию с архивом и выполните команду:

     Пример для версии 1.7.0.98:

     ```
     tar -xvf iw_das_setup_1.7.0.98.tar.xz
     ```

  2. Выполните команду:

     ```
     cat platform/setup.yaml | grep version
     ```

- В кластере с **Prediction**, на время любого обновления **остановите расчет рейтинга** (подробнее см. *InfoWatch Prediction. Руководство по установке. Обновление Prediction*);
- Если вы обновляете продукты с помощью сброса установки (`reset`):
  - До сброса установки запомните или сохраните пути хранения данных. Получить список папок можно, выполнив команду:

    ```
    cat /var/lib/iwplatform/data/volumes.txt
    ```

  - В кластере с **Data Discovery**:
    - На время любого обновления остановите все задачи сканирования в интерфейсе Data Discovery;
    - Чтобы сохранить настройки зон и серверов в новой версии, перед сбросом установки подключитесь к ноде, на которой Data Discovery установлен в режиме `central` и **остановите сервис Discoperator** с помощью команды:

```
kubectl scale --replicas=0 deployment discoperator-$(kubectl -n infowatch get
pods | grep discoperator | awk -F '-' '//{print $2;}') -n infowatch
```

- В кластере с **Data Analysis Service** для переноса настроек, перед обновлением запустите скрипт миграции переменных. Подробнее см. *InfoWatch Data Analysis Service. Руководство по установке, конфигурированию и администрированию. Обновление Data Analysis Service*.

11.3 Прямое обновление Data Analysis Service

> **Важно!**
>
> Чтобы **обновить распределенный кластер** Data Analysis Service и/или продуктов Платформы:
>
> 1. **Обновите центральную ноду** кластера и все установленные на ней продукты.
> 2. **Обновите офисные ноды** и все установленные на них продукты.

Чтобы обновить Data Analysis Service:

1. Если вы используете технологию протокластеризации в Data Discovery, остановите все задачи сканирования в его веб-интерфейсе. Подробнее см. *InfoWatch Data Discovery. Руководство пользователя*.
2. Выполните шаги по подготовке сервера к обновлению (см. *Подготовка сервера*).
3. Создайте новую директорию на диске (например, `DAS`):

   ```
   mkdir DAS
   ```

4. Скопируйте архив **`iw_das_setup_1.7.x.xx.tar.xz`** в созданную директорию.
   В нашем примере будет использована версия `1.7.1.142`, ей соответствует архив `iw_das_setup_1.7.1.142.tar.xz`.
5. Перейдите в директорию, в которую был скопирован архив.
6. Распакуйте архив с дистрибутивом Data Analysis Service в эту директорию:

   ```
   tar -xvf iw_das_setup_1.7.1.142.tar.xz
   ```

7. Убедитесь, что в разделе с директорией `/var/lib/containerd` доступно не менее 150 ГБ дискового пространства.
   Чтобы проверить доступное дисковое пространство, используйте команду:

   ```
   df -h /var/lib/containerd
   ```

   > **Примечание:**

Чтобы освободить дисковое пространство, вы можете удалить образы неиспользуемых контейнеров предыдущих версий. Для этого выполните следующие шаги:

a. Добавьте в файл `/etc/crictl.yaml` (при необходимости создайте его) строки:

```
runtime-endpoint: unix:///run/containerd/containerd.sock
image-endpoint: unix:///run/containerd/containerd.sock
timeout: 5
pull-image-on-create: false
disable-pull-on-run: false
```

b. Выполните команды:

i. Для удаления контейнеров Платформы (где x.x.x.x – установленная версия Платформы):

```
crictl rmi $(crictl image | awk '($2 !~ /\<x.x.x.x\>|\<latest\>/ && $1
~ /docker.infowatch.ru\/platform-core-release/ && $1 !~ /flannel/)
{print $3}')
```

ii. Для удаления контейнеров Data Analysis Service (где x.x.x.x – установленная версия Data Analysis Service):

```
crictl rmi $(crictl image | awk '($2 !~ /\<x.x.x.x\>|\<latest\>/ && $1
~ /docker.infowatch.ru\/clust/) {print $3}')
```

Данные команды удаляют контейнеры всех версий кроме указанной. Не удаляйте контейнеры установленной версии Data Analysis Service и Платформы!

8. Запустите программу обновления Сервиса:

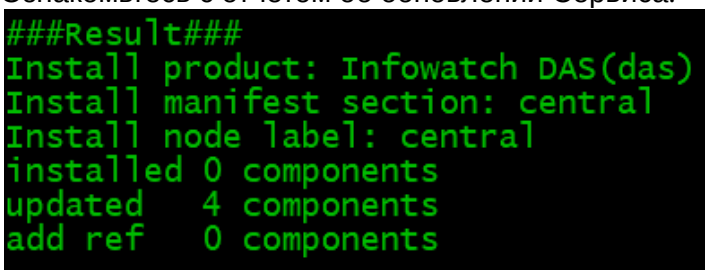
```
./setup.py update
```

Важно!

Во время обновления продукты Платформы, уже установленные на сервере, могут быть недоступны.

9. При необходимости инсталлятор обновит Платформу. В этом случае дождитесь окончания этого процесса.
10. Введите "**y**", чтобы выполнить миграцию значений переменных окружения, или "**n**", чтобы не выполнять, (по умолчанию: "**n**") и нажмите **Enter**.
Рекомендуется выполнить миграцию, чтобы перенести все настройки Сервиса.
11. Дождитесь окончания процесса обновления Data Analysis Service.

12. Ознакомьтесь с отчетом об обновлении Сервиса:



13. Чтобы вывести на экран информацию об установленных продуктах, выполните команду:

```
./setup.py showproducts
```

14. Чтобы проверить, что обновление прошло успешно, выполните команду:

```
kubectl get pods -n infowatch
```

Команда выведет список сервисов. Каждый сервис должен иметь статус `Running`.

15. Если на сервере установлены продукты Платформы, обновите их до версий, совместимых с Data Analysis Service (см. *Типы установки и обновления, совместимость с другими продуктами, аппаратные и программные требования*).
16. Если Data Analysis Service установлен в распределенном кластере, обновите другие экземпляры Сервиса и продукты Платформы на других серверах кластера до версий, совместимых с Data Analysis Service.
17. Если после обновления вам необходимо освободить место на диске, вы можете удалить неиспользуемые контейнеры аналогично п.7.

Важно!

При обновлении **Data Analysis Service на версию 1.7.x, начиная с 1.7.1**, для взаимодействия с **InfoWatch Traffic Monitor версии 7.15.2 и ниже**, а также **7.16.0-7.16.1**, для каждого экземпляра Сервиса выполните команду:

```
kubectl set env deploy -n infowatch das-adapter-<node_label> AUTH_REQUIRED=False
```

где `<node_label>` – это лейбл ноды, на которую установлен экземпляр Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

После обновления уровень логирования Сервиса будет сброшен и по умолчанию иметь значение INFO. Подробнее о логировании см. *Диагностика Data Analysis Service*.

11.4 Обновление Data Analysis Service с помощью сброса установки (reset)

Важно!

Чтобы обновить весь распределенный кластер Data Analysis Service и/или продуктов Центра расследований с помощью сброса (reset) установки, действуйте в следующей последовательности:

1. Выполните **сброс на офисных** нодах с продуктами в режиме `office`.
2. Выполните **сброс на офисных** нодах с продуктами в режиме `central`.
3. Выполните **сброс центральной** ноды.
4. **Установите центральную** ноду кластера.
5. **Установите офисные** ноды с продуктами в режиме `central`.
6. **Установите офисные** ноды с продуктами в режиме `office`.

В рамках обновления кластера выполняйте сброс установки только один раз для каждой ноды.

Важно!

При обновлении с помощью сброса установки происходит повторная массовая загрузка данных в OLAP-СУБД Clickhouse, которая является частью Платформы. Этот процесс может занимать длительное время. В Центре расследований могут временно не отображаться события.

Обновление состоит из следующих шагов:

1. Сброс установки **с помощью нового инсталлятора**. При сбросе данные Сервиса не будут удалены, но будут сброшены настройки. Во время обновления вы можете осуществить их миграцию.
2. Установка новых версий Платформы и Сервиса с указанием уже имеющихся папок с данными.

Примечание:

Если вы используете удаленное подключение по протоколу SSH, чтобы избежать проблем в случае разрыва соединения с сервером, рекомендуем использовать утилиту Screen. При отключении от утилиты запущенные в ней процессы не прервутся, что позволит безопасно продолжить обновление.

Чтобы обновить Data Analysis Service :

1. Выполните шаги по подготовке сервера к обновлению (см. *Подготовка сервера*).
2. Создайте новую директорию на диске (например, `DAS`):

```
mkdir DAS
```

3. Скопируйте архив **`iw_das_setup_1.7.x.xx.tar.xz`** в созданную директорию.
   В нашем примере будет использована версия `1.7.0.98`, ей соответствует архив `iw_das_setup_1.7.0.98.tar.xz`.
4. Перейдите в директорию, в которую был скопирован архив.
5. Распакуйте архив с дистрибутивом в эту директорию:

```
tar -xvf iw_das_setup_1.7.0.98.tar.xz
```

6. Чтобы после обновления не выполнять настройку Сервиса заново, вы можете сохранить значения переменных окружения для миграции. Вы можете это сделать двумя способами:
   - На каждой ноде кластера, где установлен Data Analysis Service, удалите Сервис с помощью команды:

     ```
     ./setup.py remove
     ```

     При удалении инсталлятор автоматически создаст файл в формате `.env` с настройками удаляемого экземпляра Data Analysis Service. Этот файл будет сохранен в директории `/var/tmp/das_env_migration/raw/`.
   - Или создайте файл с настройками, используя скрипт `migrate_env.sh`:
     a. На сервере Data Analysis Service перейдите в директорию `../DAS/tools/migration`.
     b. Запустите скрипт `migrate_env.sh` с помощью команды:

        ```
        ./migrate_env.sh prepare -n infowatch --source-version X.X.X
        ```

        где `X.X.X` – это версия ранее установленного Сервиса.

        Пример:

        ```
        ./migrate_env.sh prepare -n infowatch --source-version 1.6.2
        ```

        Скрипт создаст файл в формате `.env` с настройками данного экземпляра Data Analysis Service. Этот файл будет сохранен в директории `/var/tmp/das_env_migration/raw/`.
     c. Аналогичным образом запустите скрипт для **остальных экземпляров** Data Analysis Service в кластере.

        Примечание:

Справку по использованию скрипта вы можете получить, выполнив команду:

```
./migrate_env.sh --help
```

7. В дальнейшем, при установке новой версии присвойте нодам кластера те же лейблы, которые были до обновления.
   Чтобы посмотреть список лейблов в кластере, выполните команду:

```
kubectl get node -n infowatch -o jsonpath="{range .items[*]}
{.metadata.annotations.flannel\.alpha\.coreos\.com/public\-ip}
{.metadata.labels.kubernetes\.io/hostname} {.metadata.labels.iwnode}{'\n'}{end}"
```

8. До сброса установки запомните или сохраните пути к текущим директориям хранения данных. Получить список директорий можно, выполнив команду:

```
cat /var/lib/iwplatform/data/volumes.txt
```

9. Выполните сброс текущей установки **с помощью инсталлятора новой версии**:

```
./setup.py reset
```

> **Важно!**
>
> Если на сервере установлены продукты на базе Платформы, установки этих продуктов будут сброшены. При сбросе установки на центральной ноде продукты на офисных нодах перестанут функционировать.
>
> После установки новой версии Data Analysis Service повторно установите продукты Платформы, указав директории с имеющимися данными.
>
> Если на этом сервере вы **уже обновили продукт** Платформы до совместимой версии, **не выполняйте повторный сброс** при обновлении Data Analysis Service , так как требуемая версия Платформы уже установлена.

10. Дождитесь окончания процесса сброса установки Платформы и Сервиса.
11. Перезагрузите сервер с помощью команды:

```
reboot
```

12. Если вы обновляетесь с переходом на другой сервер или на новую ОС, скопируйте на другой раздел или на новый сервер директорию с данными PostgreSQL, которая была указана при установке.
    По умолчанию директории расположены по пути `/opt`. Обязательно копируйте их с сохранением прав и владельцев. Для копирования используйте стандартные средства, например утилиту `rsync`.

13. Убедитесь, что в разделе с директорией `/var/lib/containerd` доступно не менее 150 ГБ дискового пространства.
Чтобы проверить доступное дисковое пространство, используйте команду:

```
df -h /var/lib/containerd
```

> **Примечание:**
>
> Чтобы освободить дисковое пространство, вы можете удалить директорию с образами контейнеров прошлых версий `/var/lib/containerd`. Для этого выполните команду:
>
> ```
> rm -rf /var/lib/containerd
> ```
>
> **Данная команда удаляет все образы контейнеров. Не используйте ее, если на сервер уже установлены новые версии Data Analysis Service или других продуктов InfoWatch .**

14. Установите новую версию Data Analysis Service, в том числе новые модели (см. *Установка Data Analysis Service*).

> **Важно!**
>
> Когда инсталлятор потребует указать директории для размещения данных, указывайте директории с данными прошлой версии и выбирайте «**n**» при ответе на вопрос «Очищать ли папки перед установкой?».
>
> По умолчанию инсталлятор создает директории по пути `/opt`. Если до обновления вы использовали другой путь, то переместите данные в `/opt` или указывайте старые директории при установке новой версии. Для перемещения используйте стандартные средства, например, утилиту `rsync`.

15. Если до сброса на сервере были установлены и другие продукты, установите версии продуктов, совместимые с Data Analysis Service (см. *Типы установки и обновления, совместимость с другими продуктами, аппаратные и программные требования*).
16. Если Data Analysis Service установлен в распределенном кластере, обновите другие экземпляры Сервиса и продукты Платформы на других серверах кластера до версий, совместимых с Data Analysis Service .
17. Если после обновления вам необходимо освободить место на диске, вы можете удалить неиспользуемые контейнеры.
Для этого выполните следующие шаги:

    a. Добавьте в файл `/etc/crictl.yaml` (при необходимости создайте его) строки:

```
runtime-endpoint: unix:///run/containerd/containerd.sock
image-endpoint: unix:///run/containerd/containerd.sock
timeout: 5
```

```
pull-image-on-create: false
disable-pull-on-run: false
```

b. Выполните команды:
   i. Для удаления контейнеров Платформы (где x.x.x.x – установленная версия Платформы):

```
crictl rmi $(crictl image | awk '($2 !~ /\<x.x.x.x\>|\<latest\>/ && $1 ~ /
docker.infowatch.ru\/platform-core-release/ && $1 !~ /flannel/) {print
$3}')
```

   ii. Для удаления контейнеров Data Analysis Service (где x.x.x.x – установленная версия Data Analysis Service):

```
crictl rmi $(crictl image | awk '($2 !~ /\<x.x.x.x\>|\<latest\>/ && $1 ~ /
docker.infowatch.ru\/clust/) {print $3}')
```

Данные команды удаляют контейнеры всех версий кроме указанной. Не удаляйте контейнеры установленной версии Data Analysis Service и Платформы!

18. **Если при установке новой версии Сервиса не была выполнена миграция переменных окружения, но вы хотите сделать это:**

    a. На сервере Data Analysis Service перейдите в директорию `../DAS/tools/migration`.
    b. Запустите скрипт `migrate_env.sh`:

```
./migrate_env.sh -n infowatch finalize --target-version X.X.X --wait
```

    где `X.X.X` – это версия обновленного Сервиса.
    Скрипт перенесет значения переменных и перезапустит Сервис.

Пример команды запуска скрипта:

```
./migrate_env.sh -n infowatch finalize --target-version 1.7.0 --wait
```

    При распределенной установке необходимо отдельно запускать скрипт для каждого экземпляра Сервиса на той ноде, где он установлен.
    По умолчанию скрипт применяет настройки на текущей ноде. Но вы можете указать лейбл другой ноды с помощью параметра `--node <das_node_label>`.

19. **Если вы используете Data Explorer:**

    Для отображения карты кластеров Сервис должен рассчитать их координаты. После обновления Data Analysis Service это произойдет автоматически ближайшей ночью. Но если вы хотите, чтобы карта кластеров стала доступной сразу:

    a. На сервере Data Analysis Service перейдите в директорию `../DAS/tools/migration`. В ней располагается скрипт `create_coo.sh`.

b. Если Сервис установлен на офисной ноде кластера, скопируйте скрипт на центральную ноду.
c. На центральной ноде кластера запустите скрипт с помощью команды:

```
./create_coo.sh -n infowatch -l <das_node_label>
```

где `<das_node_label>` – лейбл ноды, на которую установлен центральный экземпляр Data Analysis Service (productmode=central). Если это центральная нода кластера, то параметр `-l` можно не указывать, либо используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.
d. Дождитесь окончания выполнения скрипта. Это может занять некоторое время.

20. Если вы обновляете Data Analysis Service с версии 1.6.0, ранее задавали языки для транскрибации и не выполнили миграцию переменных, задайте языки повторно, используя переменную `TRANSCRIPTION_LANGUAGES`. Подробнее см. *Настройки языков*.
21. Если до обновления Data Analysis Service в кластере был обновлен Device Monitor, обновите конфигурационный файл **`n.yaml`**:
    a. Подключитесь к центральной ноде Платформы.
    b. Если Device Monitor установлен на офисной ноде Платформы, скопируйте на центральную его дистрибутив и распакуйте.
    c. На центральной ноде Платформы перейдите в папку, где распакован дистрибутив Device Monitor.
    d. Запустите скрипт **`setup.sh`**, выполнив команду:

```
./setup.sh
```

    e. Выберите в меню пункт **Utilities**.
    f. Выберите **Fix n.yaml** file.
    g. Введите IP-адрес сервера Device Monitor и дождитесь окончания процесса.

> **Примечание:**
>
> Если вы обновляете несколько продуктов в кластере, обновите **`n.yaml`** только после обновления последнего продукта.

> **Важно!**
>
> При обновлении **Data Analysis Service на версию 1.7.x, начиная с 1.7.1**, для взаимодействия с **InfoWatch Traffic Monitor версии 7.15.2 и ниже**, а также **7.16.0-7.16.1**, для каждого экземпляра Сервиса выполните команду:
>
> ```
> kubectl set env deploy -n infowatch das-adapter-<node_label> AUTH_REQUIRED=False
> ```

где `<node_label>` – это лейбл ноды, на которую установлен экземпляр Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

После обновления уровень логирования Сервиса будет сброшен и по умолчанию иметь значение INFO. Подробнее о логировании см. *Диагностика Data Analysis Service*.

12 Настройка Data Analysis Service

В этом разделе:

- Настройка взаимодействия с продуктами Платформы
- Включение/выключение модулей Data Analysis Service
- Настройка производительности
- Использование GPU
- Настройка модуля транскрибации
- Настройка модуля распознавания именованных сущностей
- Настройка модуля детектирования объектов на изображениях
- Настройка модуля OCR
- Добавление языков для протокластеризации
- Диагностика Data Analysis Service

12.1 Настройка взаимодействия с продуктами Платформы

В этом разделе:

- Включение технологий Data Analysis Service в продуктах Платформы
- Задание нод обработки

12.1.1 Включение технологий Data Analysis Service в продуктах Платформы

Activity Monitor

Data Analysis Service позволяет:

- транскрибировать аудиозаписи;
- детектировать объекты на изображениях, в том числе непристойный контент;
- распознавать и извлекать текст из изображений.

Чтобы использовать эти функциональные возможности в Activity Monitor:

1. Убедитесь, что Data Analysis Service настроен на обработку центральной ноды кластера. Лейбл этой ноды указывается при установке Сервиса. Также его можно задать после установки. Подробнее см. *Задание нод обработки*.
2. Убедитесь, что модуль транскрибации/модуль детектирования объектов на изображениях/модуль OCR включены. Подробнее см. *Включение/выключение модулей Data Analysis Service*.
3. Включите **транскрибацию** данных в Activity Monitor, выполнив на центральной ноде кластера следующие команды:

   ```
   kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<am_central_node> | awk '//{print $1;}') -n infowatch -- psql -U postgres -d activitymonitor -c "update features set status='enabled' where name='stream_transcripts'"
   kubectl rollout restart deployment datacollector-<am_central_node> -n infowatch
   ```

   где `<am_central_node>` – лейбл ноды, где Activity Monitor установлен в режиме **`central`**.

> **Примечание:**
>
> Если Activity Monitor и Data Analysis Service установлены не в распределенном кластере, а на одном отдельном сервере, в качестве лейбла ноды укажите значение `central`.

Команды также приведены в документации Activity Monitor: *InfoWatch Activity Monitor. Руководство по установке. Установка Activity Monitor*.

4. Аналогичным образом включите **детектирование объектов**. Для этого на центральной ноде кластера выполните команды:

```
kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<am_central_node> | awk '//{print $1;}') -n infowatch -- psql -U postgres -d activitymonitor -c "update features set status='enabled' where name='stream_enrich'"
kubectl rollout restart deployment datacollector-<am_central_node> -n infowatch
```

   где `<am_central_node>` – лейбл ноды, где Activity Monitor установлен в режиме **`central`**.

5. Аналогичным образом включите **распознавание текста (OCR)**. Для этого на центральной ноде кластера выполните команды:

```
kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<am_central_node> | awk '//{print $1;}') -n infowatch -- psql -U postgres -d activitymonitor -c "update features set status='enabled' where name='screenshot_transcripts'"
kubectl rollout restart deployment datacollector-<am_central_node> -n infowatch
```

   где `<am_central_node>` – лейбл ноды, где Activity Monitor установлен в режиме **`central`**.

Data Discovery

Data Analysis Service позволяет осуществлять протокластеризацию текстовых документов для поиска похожих.

Чтобы использовать эту функциональную возможность в Data Discovery:

1. Установите Data Analysis Service в том же кластере.
2. Настройте Сервис на обработку ноды, на которой установлен Data Discovery в режиме **`central`**. Лейблы нод для обработки указываются при установке Сервиса. Также их можно задать после установки. Подробнее см. *Задание нод обработки*.
   Если вы будете использовать несколько серверов Data Analysis Service для взаимодействия с Data Discovery, настраивайте их все на ноду Data Discovery **`central`**.
3. На центральной ноде Data Discovery перезапустите сервис `discoperator` с помощью команды:

```
kubectl rollout restart deploy discoperator-<dd_central_node> -n infowatch
```

   где `<dd_central_node>` – лейбл центральной ноды Data Discovery.

Пример команды для ноды с лейблом ddcentral:

```
kubectl rollout restart deploy discoperator-ddcentral -n infowatch
```

После этого в Data Discovery вы сможете включить опцию "**Определять похожие файлы**" при создании или редактировании задачи сканирования.
Подробнее о задачах сканирования см. *InfoWatch Data Discovery. Руководство пользователя. Раздел "Хранение файлов". Вкладка "Задачи"*.
О поиске похожих файлов см. *InfoWatch Data Discovery. Руководство пользователя. Раздел "Хранение файлов". Вкладка "Похожие файлы"*.

Data Explorer

Data Analysis Service позволяет использовать в InfoWatch Data Explorer технологии протокластеризации и распознавания именованных сущностей. Для этого достаточно установить Сервис в одном кластере с Data Explorer.
Подробнее об использовании технологий см. *InfoWatch Data Explorer. Руководство пользователя*.

12.1.2 Задание нод обработки

При установке Data Analysis Service в кластере с продуктами Платформы, для Сервиса необходимо задать одну или несколько нод, очередь сообщений которых он будет обрабатывать.

Чтобы изменить данный параметр после установки, задайте необходимое значение переменной окружения `SUBSCRIBE_TO`. Для этого на сервере Data Analysis Service выполните следующую команду:

```
kubectl set env deploy -n infowatch das-processor-<das_node_label>
SUBSCRIBE_TO="<message_queue_node_labels>"
```

где `<das_node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.
`<message_queue_node_labels>` – список лейблов нод, на которые продукты Платформы отправляют сообщения для Data Analysis Service. Указывайте лейблы через запятые.

> **Примечание:**
>
> InfoWatch Activity Monitor отправляет сообщения в очередь NATS на центральную ноду кластера.
>
> InfoWatch Data Discovery отправляет сообщения в очередь NATS на ноду, где Data Discovery установлен в режиме `central`.

Чтобы узнать лейбл ноды, выполните команду:

```
kubectl get node -n infowatch -o jsonpath="{range .items[*]}
{.metadata.annotations.flannel\.alpha\.coreos\.com/public\-ip} {.metadata.labels.kubernetes\.io/hostname}
{.metadata.labels.iwnode}{'\n'}{end}"
```

В результате будет выведен список нод кластера с указанием их лейблов.

Чтобы узнать текущее значение переменной окружения `SUBSCRIBE_TO`, выполните команду:

```
kubectl get deployment -n infowatch das-processor-<das_node_label> -o
jsonpath='{range .spec.template.spec.containers[0].env[*]}{.name}={.value}{"\n"}{end}'|grep SUBSCRIBE_TO
```

12.2 Включение/выключение модулей Data Analysis Service

Вы можете включить модули во время установки Сервиса.

Чтобы включить или выключить какой-либо компонент Data Analysis Service после установки, присвойте соответствующее значение переменной окружения для необходимого компонента. Для этого на сервере Data Analysis Service выполните следующую команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> <name>=<value>
```

где `<node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`;
`<name>` – это имя переменной. Модулям Data Analysis Service соответствуют следующие переменные окружения:

Имя переменной окружения	Компонент Data Analysis Service
`PROTO_CLUSTERIZATION_ENABLE`	Модуль протокластеризации
`TRANSCRIPTION_ENABLE`	Модуль транскрибации
`IMAGE_CONTENT_DETECTION_ENABLE`	Модуль детектирования объектов на изображениях
`NAMED_ENTITY_RECOGNITION_ENABLE`	Модуль распознавания именованных сущностей
`OPTICAL_CHARACTER_RECOGNITION_ENABLE`	Модуль оптического распознавания символов (Модуль OCR)

`<value>` – это значение переменной. Оно может быть равно `TRUE` или `FALSE`.

Пример:

Для включения модуля транскрибации Сервиса, установленного на центральной ноде кластера, выполните команду:

```
kubectl set env deploy -n infowatch das-processor-central TRANSCRIPTION_ENABLE=TRUE
```

Чтобы вывести все переменные окружения Data Analysis Service и их значения, выполните команду:

```
kubectl get deployment -n infowatch das-processor-<node_label> -o
jsonpath='{range .spec.template.spec.containers[0].env[*]}{.name}={.value}{"\n"}{end}'
```

Примечание:

Рекомендуется выключать неиспользуемые функции в целях экономии ресурсов.

12.3 Настройка производительности

По умолчанию на сервере функционирует один POD Data Analysis Service, включающий в себя все модули Сервиса. Рекомендуемые аппаратные требования для модулей одного POD вы можете найти в статье *Типы установки, совместимость с другими продуктами, аппаратные и программные требования*.

Вы можете оценить производительность Data Analysis Service следующими способами:

- Оценка производительности модуля протокластеризации
- Оценка производительности модулей Сервиса

По итогам оценки вы можете увеличить производительность Data Analysis Service с помощью распределенной установки.

12.3.1 Оценка производительности модуля протокластеризации

Если выделенных для модуля протокластеризации ресурсов недостаточно для обработки текстовых данных, то в Консоли Управления Traffic Monitor кнопка **Найти аналоги** может не отображаться для новых событий. В таком случае оцените производительность модуля протокластеризации, проанализировав количество необработанных объектов в базе данных. Для этого на центральной ноде кластера в течение 2-3 часов каждые 30-60 минут выполняйте команду:

```
kubectl exec -ti -n infowatch $(kubectl get pods -n infowatch | grep postgresql-<das_central_node> | awk
'{print $1}') -- psql -U postgres -d das_adapter -c 'select count(*) from content_handling_progress;'
```

где `<das_central_node>` – это лейбл ноды, где установлен центральный экземпляр Data Analysis Service, т.е. при установке для параметра `--productmode` было задано значение `central`.

При росте числа необработанных объектов и нехватке производительности модуля протокластеризации рекомендовано ее увеличение.

12.3.2 Оценка производительности модулей Сервиса

Вы можете оценить производительность интересующего вас модуля, отследив изменение количества контентов в очереди службы NATS на входе модуля.

Чтобы посмотреть информацию об очередях NATS:

1. Выведите список подов, выполнив команду:

   ```
   kubectl get pods -n infowatch
   ```

2. Найдите среди них **das-processor** и скопируйте его полное название.

3. Получите значение переменной `NATSTOKEN` из окружения пода **das-processor**:

```
kubectl exec -it -n infowatch <das-processor_pod_name> -- env | grep -i NATSTOKEN
```

где `<das-processor_pod_name>` – полное название пода **das-processor**.

Пример команды

```
kubectl exec -it -n infowatch das-processor-central-84dc84cbfc-slqx8 -- env | grep -i
NATSTOKEN
```

4. Определите адрес сервиса **nats-central**:

```
kubectl get svc -A | grep nats-central
```

Запомните значение поля **`ClusterIP`**.

5. Выведите значение переменной окружения **`PATH`**:

```
echo $PATH
```

6. Установите утилиту **nats**. Для этого:
   a. Перейдите в директорию `../DAS/tools/natscli/`.
   b. Скопируйте бинарный файл **`nats`** в одну из директорий, заданных в переменной **`PATH`**, например, в `/usr/local/bin/`:

```
sudo cp nats /usr/local/bin/
```

   c. При необходимости сделайте файл исполняемым:

```
chmod +x /usr/local/bin/nats
```

7. Подключитесь к службе NATS:

```
nats -s nats://<NATSTOKEN>@<NATS_CENTRAL_IP>:8080 stream ls
```

где `<NATSTOKEN>` – это значение, полученное на 3 шаге;
`<NATS_CENTRAL_IP>` – значение, полученное на 4 шаге.

В результате будет выведен список очередей NATS, среди которых вы можете просмотреть те, которые относятся к нужной вам технологии Data Analysis Service.

12.3.3 Распределенная установка

Для увеличения производительности Data Analysis Service вы можете установить Сервис на одной или нескольких дополнительных нодах (подробнее см. *Установка Data Analysis Service*). Приходящие в Сервис контенты будут распределяться между всеми его экземплярами в кластере и обрабатываться параллельно.

Для параллельной обработки контентов от InfoWatch Traffic Monitor, настройте взаимодействие InfoWatch Traffic Monitor с любым экземпляром Сервиса в кластере (подробнее см. документ *InfoWatch Traffic Monitor. Руководство администратора. Настройка интеграции с Data Analysis Service*).

Для параллельной обработки контентов от продукта Платформы, настройте экземпляры Сервиса на ноду, куда этот продукт отправляет сообщения. Подробнее см. *Задание нод обработки*.

12.4 Использование GPU

12.4.1 Аппаратные и программные требования к графическому процессору

Для использования GPU вам необходима видеокарта NVIDIA с поддержкой Compute Unified Device Architecture (**CUDA**) версии **12.4**.
Поддерживаемые версии CUDA GPU **Compute Capability** (CC): **7.0-12.0** (подробнее см. на официальном сайте).
Поддерживаемые версии архитектур CUDA:

- Volta (2017)
- Turing (2018)
- Ampere (2020)
- Ada Lovelace (2022)
- Blackwell (2024)

На сервере должны быть установлены:

- Драйвер NVIDIA с версией не ниже 550.xx

  > **Примечание:**
  >
  > Рекомендуется устанавливать наиболее свежий драйвер.
  > Если вы установили драйвер недостаточной версии, то в логах процесса **das-processor** появится сообщение вида:
  > `[ERROR] (das_processor.tech.transcriber) CUDA failed with error CUDA driver version is insufficient for CUDA runtime version`

- NVIDIA Container Toolkit с версией не ниже 1.17.8. С инструкцией по установке можно ознакомиться на официальном сайте (страница доступна на английском языке)
- Kubernetes с версией не ниже 1.10
- NVIDIA Kubernetes Device Plugin версии 0.14.1. Его можно скачать с официального сайта (страница доступна на английском языке)

  > **Примечание:**
  >
  > В случае установки Data Analysis Service на офисной ноде плагин NVIDIA необходимо устанавливать на центральной ноде.

Ресурсы GPU должны быть доступны в кластере. Инструкция по настройке доступна на официальном сайте (страница доступна на английском языке).

12.4.2 Настройка доступа к GPU

Чтобы Data Analysis Service мог использовать GPU, настройте доступ Сервиса к нему. Для этого:

1. Убедитесь, что окружение удовлетворяет указанным выше требованиям:
   - Чтобы узнать версию установленного драйвера NVIDIA, используйте команду:

     ```
     cat /sys/module/nvidia/version
     ```

   - Чтобы узнать версию установленного NVIDIA Container Toolkit, используйте команду:

     ```
     nvidia-ctk --version
     ```

   - Чтобы узнать версию установленного Kubernetes, используйте команду:

     ```
     kubectl version --short
     ```

   - Чтобы проверить доступность GPU на сервере, используйте команду:

     ```
     gpu_count=$(kubectl describe -n infowatch node | grep nvidia.com | wc -l); if [[ ${gpu_count} > 0 ]]; then echo "Ok"; else echo "No GPU support"; fi
     ```

     Если в результате команды будет выведено сообщение "`No GPU support`", то:
     a. Выполните команды:

        ```
        sudo systemctl restart containerd
        kubectl rollout restart ds -n kube-system nvidia-device-plugin-daemonset
        ```

     b. Повторите проверку.
        Если снова выведено сообщение "`No GPU support`", то:
        i. Выполните команды:

           ```
           sudo nvidia-ctk runtime configure --runtime=containerd --set-as-default
           sudo systemctl restart containerd
           ```

        ii. Повторите проверку.
2. Добавьте GPU в ресурсы процесса **das-processor**. Для этого выполните команду:

   ```
   sudo kubectl -n infowatch set resources deploy das-processor-<node_label> --limits=nvidia.com/gpu=1
   ```

   где `<node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.
3. Перезагрузите сервер. Загрузка кластера может занять некоторое время.

4. Убедитесь, что все Pod запущены, выполнив команду:

```
sudo kubectl get pods -n infowatch
```

   Каждый Pod должен иметь статус Running.

Далее вы можете включить использование GPU в нужном вам модуле.

12.5 Настройка модуля транскрибации

В данной статье:

- Включение модуля
- Общее описание
- Включение использования GPU
- Выбор модели транскрибации
- Проверка использования GPU

12.5.1 Включение модуля

Чтобы включить модуль транскрибации после установки, на сервере Data Analysis Service выполните следующую команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> TRANSCRIPTION_ENABLE=TRUE
```

где `<node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

Подробнее см. *Включение/выключение модулей Data Analysis Service*.

12.5.2 Общее описание

Модуль транскрибации включает в себя две модели обработки: малую и большую. Использование большой модели существенно повышает качество транскрибации, но уменьшает ее скорость. По умолчанию используется малая модель.

> **Важно!**
>
> Для использования моделей, необходимо предварительно их установить. Подробнее см. *Установка моделей*.

Вы можете увеличить производительность модуля транскрибации, используя вместо центрального процессора (CPU) графический (GPU). Для этого включите использование GPU.

Чтобы ускорить обработку аудиоданных, используйте GPU.

Чтобы повысить качество транскрибации:

- Включите использование GPU;

- Используйте большую модель обработки.

Чтобы убедиться, что Сервис использует GPU, вы можете выполнить проверку.

Также вы можете задать языки транскрибации и использовать мультиязычную транскрибацию. Подробнее см. *Настройки языков*.

12.5.3 Включение использования GPU

Чтобы модуль транскрибации использовал GPU:

1. Настройте доступ Data Analysis Service к GPU. См. *Использование GPU*.
2. Укажите, что модуль транскрибации должен использовать GPU. Для этого выполните команду:

   ```
   kubectl -n infowatch set env deploy das-processor-<node_label> AGE_CLASSIFIER_USE_GPU=TRUE
   AUDIO_EMBEDDER_USE_GPU=TRUE EMOTION_DETECTOR_USE_GPU=TRUE
   FASTER_WHISPER_TRANSCRIBER_USE_GPU=TRUE WHISPER_DETECTOR_USE_GPU=TRUE
   T5_CORRECTOR_USE_GPU=TRUE
   ```

   где `<node_label>` – лейбл ноды, на которую установлен Data Analysis Service.

За использование GPU в модуле транскрибации отвечает набор переменных окружения. Если нужно переключиться обратно на использование CPU, выполните команду:

```
kubectl -n infowatch set env deploy das-processor-<node_label> AGE_CLASSIFIER_USE_GPU=FALSE
AUDIO_EMBEDDER_USE_GPU=FALSE EMOTION_DETECTOR_USE_GPU=FALSE FASTER_WHISPER_TRANSCRIBER_USE_GPU=FALSE
WHISPER_DETECTOR_USE_GPU=FALSE T5_CORRECTOR_USE_GPU=FALSE
```

12.5.4 Выбор модели транскрибации

Чтобы выбрать модель обработки, выполните команду:

- для использования малой модели:

  ```
  sudo kubectl set env deployment -n infowatch das-processor-<node_label>
  FASTER_WHISPER_TRANSCRIBER_MODEL_TYPE=small
  ```

- для использования большой модели:

  ```
  sudo kubectl set env deployment -n infowatch das-processor-<node_label>
  FASTER_WHISPER_TRANSCRIBER_MODEL_TYPE=large_v1
  ```

> **Примечание:**
>
> Большая модель может работать только на GPU.

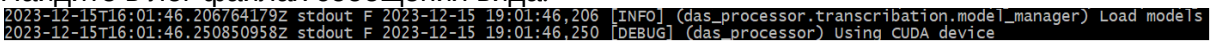
12.5.5 Проверка использования GPU

Чтобы убедиться, что модуль транскрибации использует GPU:

1. Если вы будете использовать большую модель обработки, выберите ее.
2. Установите для Сервиса режим логирования DEBUG:

```
sudo kubectl -n infowatch set env deploy das-processor-<node_label> SRV_LOG_LEVEL=DEBUG
```

3. Обеспечьте передачу в Data Analysis Service аудиоданных для обработки.
4. Перейдите в директорию `/var/log/pods/infowatch_das-processor-<node_label>-…/das-processor-<node_label>` , где находятся лог-файлы.
5. Найдите в лог-файлах сообщения вида:



   Такие сообщения появляются при загрузке моделей модуля транскрибации. В момент загрузки моделей в лог-файл записывается сообщение `Load models` . Сообщение `Using CUDA device` означает, что Сервис использует GPU.
   Модели загружаются для обработки аудиоданных и уничтожаются, если неактивны более часа. При уничтожении в лог-файл записывается сообщение `Drop models` .

Также вы можете найти сообщения в логах с помощью команды:

```
kubectl logs $(kubectl get pods -n infowatch -o name|grep das-processor-<node_label>) -n infowatch| grep "<текст_сообщения>"
```

12.5.6 Настройки языков

По умолчанию модуль транскрибации распознает речь целенаправленно только на одном языке – на том, который чаще всего встречается в конкретном аудиофайле. Модель определения языка рассчитывает, с какой вероятностью речь на том или ином языке присутствует в аудиозаписи. Модуль транскрибации выбирает самый вероятный.
Чтобы модуль распознавал речь на нескольких языках, а не на одном, вы можете включить **мультиязычную транскрибацию**. Тогда он будет разбивать каждую аудиозапись на отрезки, между которыми есть тишина, и в каждом из них отдельно определять язык.

Включение мультиязычной транскрибации

Чтобы включить или выключить мультиязычную транскрибацию, присвойте переменной окружения **`TRANSCRIPTION_MULTILINGUAL`** значение `TRUE` или `FALSE` соответственно:

```
kubectl set env deploy -n infowatch das-processor-<node_label> TRANSCRIPTION_MULTILINGUAL=<value>
```

где `<node_label>` – лейбл ноды, на которой установлен Data Analysis Service. Для центральной ноды используйте лейбл `central` . Для офисной ноды лейбл задается при установке, по умолчанию: `office` ;
`<value>` – это значение переменной.

По умолчанию мультиязычная транскрибация выключена.

Примечание:

В случае мультиязычной транскрибации модуль может работать медленнее, чем при обработке на одном языке.

Задание языков транскрибации

Если для модуля заданы языки обработки, то он будет распознавать в аудиофайлах только эти языки. Вы можете задать их, используя переменную окружения **`TRANSCRIPTION_LANGUAGES`**. Также при этом нужно удалить переменную **`TRANSCRIBER_LANGUAGE`**. Для этого выполните команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label>
TRANSCRIPTION_LANGUAGES=<language_code>,<language_code>,<language_code> TRANSCRIBER_LANGUAGE-
```

где `<node_label>` – лейбл ноды, на которой установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`;
`<language_code>` – это код языка.

Доступные языки обработки и их коды:

- "ru": русский,
- "en": английский,
- "de": немецкий,
- "zh": китайский,
- "tr": турецкий,
- "ar": арабский,
- "hi": хинди,
- "uk": украинский,
- "az": азербайджанский,
- "hy": армянский,
- "kk": казахский,
- "ka": грузинский,
- "be": белорусский,
- "tg": таджикский,
- "uz": узбекский,
- "tk": туркменский,
- "tt": татарский,
- "ba": башкирский
- и другие.

Полный список языков вы можете узнать у сотрудников InfoWatch.

Если **задать только один язык**, то модель определения языка не будет использоваться. Благодаря этому скорость обработки будет выше.
Также указание единственного языка может повысить точность транскрибации в случаях, когда качество аудиозаписей низкое и модель не всегда корректно определяет язык того или иного отрезка.

Пример:

Чтобы включить мультиязычную транскрибацию на русском и английском языках в Data Analysis Service, установленном на центральной ноде кластера, выполните команду:

```
kubectl set env deploy -n infowatch das-processor-central TRANSCRIPTION_MULTILINGUAL=TRUE
TRANSCRIPTION_LANGUAGES=ru,en TRANSCRIBER_LANGUAGE-
```

12.6 Настройка модуля распознавания именованных сущностей

12.6.1 Включение модуля

Чтобы включить модуль распознавания именованных сущностей после установки, на сервере Data Analysis Service выполните следующую команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> NAMED_ENTITY_RECOGNITION_ENABLE=TRUE
```

где `<node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

Подробнее см. *Включение/выключение модулей Data Analysis Service*.

12.6.2 Настройка параметров модуля

Для изменения параметров модуля используются переменные окружения.
Чтобы присвоить значение переменной окружения, используйте команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> <name>=<value>
```

где `<node_label>` – это лейбл ноды, на которую установлен Data Analysis Service;
`<name>` – это имя переменной.

Вы можете **ускорить работу технологии** распознавания именованных сущностей в Data Explorer, например, если при открытии большого количества документов замедлилась их обработка. Для этого увеличьте значения следующих переменных окружения:

Имя переменной	Значение по умолчанию	Описание
`NAMED_ENTITY_RECOGNITION_MAX_CONCURRENCY`	5	Количество файлов, которое модуль одновременно берет из очереди для параллельной обработки

Имя переменной	Значение по умолчанию	Описание
`NAMED_ENTITY_RECOGNIZER_REPLICAS`	5	Количество реплик модели

Значения переменных `NAMED_ENTITY_RECOGNITION_MAX_CONCURRENCY` и `NAMED_ENTITY_RECOGNIZER_REPLICAS` должны быть одинаковыми, чтобы контенты поступали на вход каждой реплики модели. Рекомендуется повышать значения переменных с шагом 2-3. Каждая реплика модели занимает примерно 1 ГБ ОЗУ.

Пример:

Чтобы присвоить переменным `NAMED_ENTITY_RECOGNITION_MAX_CONCURRENCY` и `NAMED_ENTITY_RECOGNIZER_REPLICAS` значение 7 на центральной ноде, выполните команду:

```
kubectl set env deploy -n infowatch das-processor-central
NAMED_ENTITY_RECOGNITION_MAX_CONCURRENCY=7 NAMED_ENTITY_RECOGNIZER_REPLICAS=7
```

В дальнейшем вы можете вывести значения переменных окружения Data Analysis Service с помощью команды:

```
kubectl get deployment -n infowatch das-processor-<node_label> -o
jsonpath='{range .spec.template.spec.containers[0].env[*]}{.name}={.value}{"\n"}{end}'
```

12.7 Настройка модуля детектирования объектов на изображениях

12.7.1 Включение модуля

Чтобы включить модуль детектирования объектов на изображениях после установки, на сервере Data Analysis Service выполните следующую команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> IMAGE_CONTENT_DETECTION_ENABLE=TRUE
```

где `<node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

Подробнее см. *Включение/выключение модулей Data Analysis Service*.

12.7.2 Настройка параметров модуля

Для изменения параметров модуля используются переменные окружения.
Чтобы присвоить значение переменной окружения, используйте команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> <name>=<value>
```

где `<node_label>` – лейбл ноды, на которую установлен Data Analysis Service;
`<name>` – это имя переменной;
`<value>` – значение переменной.

Модуль включает в себя 3 модели обработки:

- **Object Detector** – обнаружение распространенных объектов: оружие, животные, транспорт, электроника и др.;
- **NSFW Detector** – обнаружение взрослого, откровенного или неприемлемого материала;
- **Nude Detector** – обнаружение обнаженного тела; модель используется для подтверждения результатов работы NSFW Detector.

Каждый контент обрабатывается последовательно каждой моделью.

Вы можете **ускорить работу технологии** детектирования объектов на изображениях путем увеличения количества реплик моделей. Для этого увеличьте значения следующих переменных окружения:

Имя переменной	Значение по умолчанию	Описание
`IMAGE_CONTENT_DETECTION_MAX_CONCURRENCY`	1	Максимальное количество одновременно обрабатываемых задач. Т.е. то количество файлов, которое модуль одновременно берет из очереди для параллельной обработки
`OBJECT_DETECTOR_REPLICAS`	1	Количество реплик модели Object Detector
`NSFW_DETECTOR_REPLICAS`	1	Количество реплик модели NSFW Detector
`NUDE_DETECTOR_REPLICAS`	1	Количество реплик модели Nude Detector

Количество одновременно обрабатываемы файлов и количество реплик каждой моделей рекомендуется устанавливать одинаковыми, чтобы контенты поступали на вход каждой реплики. Рекомендуется повышать значения переменных с шагом 1-2. Каждая реплика модели занимает примерно 1 ГБ ОЗУ.

Пример:

Чтобы присвоить всем переменным значение 3 на центральной ноде, выполните команду:

```
kubectl set env deploy -n infowatch das-processor-central IMAGE_CONTENT_DETECTION_MAX_CONCURRENCY=3
 OBJECT_DETECTOR_REPLICAS=3 NSFW_DETECTOR_REPLICAS=3 NUDE_DETECTOR_REPLICAS=3
```

12.7.3 Включение использования GPU

Для ускорения работы технологии, вы также можете использовать GPU.

Чтобы модуль детектирования объектов на изображениях использовал GPU:

1. Настройте доступ Data Analysis Service к GPU. См. *Использование GPU*.
2. Укажите, что все модели модуля детектирования объектов на изображениях должны использовать GPU. Для этого выполните команду:

   ```
   kubectl -n infowatch set env deploy das-processor-<node_label> OBJECT_DETECTOR_USE_GPU=TRUE
   NSFW_DETECTOR_USE_GPU=TRUE NUDE_DETECTOR_USE_GPU=TRUE
   ```

   где `<node_label>` – лейбл ноды, на которую установлен Data Analysis Service.

Если нужно переключиться обратно на использование CPU, выполните команду:

```
kubectl -n infowatch set env deploy das-processor-<node_label> OBJECT_DETECTOR_USE_GPU=FALSE
NSFW_DETECTOR_USE_GPU=FALSE NUDE_DETECTOR_USE_GPU=FALSE
```

В дальнейшем вы можете вывести значения переменных окружения Data Analysis Service с помощью команды:

```
kubectl get deployment -n infowatch das-processor-<node_label> -o
jsonpath='{range .spec.template.spec.containers[0].env[*]}{.name}={.value}{"\n"}{end}'
```

12.8 Настройка модуля OCR

В данной статье:

- Включение модуля
- Экстракторы
- Выбор экстрактора
- Настройка экстракторов
- Включение использования GPU

12.8.1 Включение модуля

Чтобы включить модуль OCR после установки, на сервере Data Analysis Service выполните следующую команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> OPTICAL_CHARACTER_RECOGNITION_ENABLE=TRUE
```

где `<node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

Подробнее см. *Включение/выключение модулей Data Analysis Service*.

12.8.2 Экстракторы

Модуль OCR предоставляет на выбор 3 OCR-экстрактора:

- IWOCR
- Tesseract
- Tesseract_East

В зависимости от ваших возможностей и потребностей вы можете выбрать и использовать один из них. В любой момент вы можете переключиться на использование другого.

Для использования какого-либо экстрактора, кроме Tesseract, необходимо установить модели для OCR. Подробнее см. *Установка моделей*.

Экстрактор	Краткое описание	Используемый процессор
IWOCR	Разработан компанией InfoWatch. Обладает лучшими показателями качества среди имеющихся экстракторов. Рекомендуется использовать с GPU.	CPU или GPU
Tesseract	Open-source OCR экстрактор. Удовлетворительно распознает только отсканированные печатные документы. Не рекомендуется использовать для распознавания текста на снимках экрана.	CPU
Tesseract_East	Усовершенствованный Tesseract. Благодаря дополнительной предобработке, во время которой детектируются и выравниваются отдельные блоки текста на изображениях, экстрактор лучше распознает текст на снимках экрана.	CPU

По умолчанию каждый экстрактор извлекает с изображений тексты на русском и английском языках. При настройке экстракторов Tesseract и Tesseract_East можно указать другие языки, подробнее см. ниже.

12.8.3 Выбор экстрактора

Чтобы выбрать экстрактор OCR, присвойте переменной `OPTICAL_CHARACTER_RECOGNITION_ENGINE` одно из следующих значений:

- TESSERACT_EAST (по умолчанию)
- TESSERACT
- IWOCR

Пример команды выбора IWOCR на центральной ноде:

```
kubectl set env deploy -n infowatch das-processor-central OPTICAL_CHARACTER_RECOGNITION_ENGINE=IWOCR
```

12.8.4 Настройка экстракторов

Для настройки экстракторов используйте переменные окружения, приведенные в таблице ниже. Чтобы присвоить значение переменной окружения, используйте команду:

```
kubectl set env deploy -n infowatch das-processor-<node_label> <name>=<value>
```

где `<node_label>` – лейбл ноды, на которую установлен Data Analysis Service;
`<name>` – имя переменной;
`<value>` – значение переменной.

Значения по умолчанию в таблице приведены для внутренних параметров Сервиса. При этом изначально сами переменные окружения могут быть не заданы явно.

Экстрактор	Имя переменной	Значение по умолчанию	Описание
Любой	`OPTICAL_CHARACTER_RECOGNITION_ENABLE_ORIENTATION_CORRECTION`	None	Принудительное включение/выключение модели коррекции ориентации. Данная модель осуществляет предобработку изображений перед их отправкой в экстрактор. По умолчанию она управляется продуктом InfoWatch, который присылает в Data Analysis Service изображения для обработки. В большинстве случаев не требуется принудительно включать или выключать ее. Но если вы задали этой переменной значение, то оно будет иметь больший приоритет, чем значение в запросе продукта-источника.
	`ORIENTATION_CORRECTOR_REPLICAS`	1	Количество реплик модели коррекции ориентации изображений.
IWOCR	`IWOCR_REPLICAS`	1	Количество реплик экстрактора. Чем их больше реплик, тем быстрее он работает. При работе на GPU 1 реплика занимает примерно 5 ГБ видеопамяти; 3 реплики – 16 ГБ; 4 реплики – 24 ГБ.
	`IWOCR_USE_GPU`	FALSE	Флаг использования GPU. Возможные значения: `TRUE`, `FALSE`.

Экстрактор	Имя переменной	Значение по умолчанию	Описание
Tesseract	`TESSERACT_MODEL_TYPE`	standard	Тип модели экстрактора Tesseract по соотношению скорости и качества распознавания. Возможные значения: `standard` / `best` / `fast` .
	`TESSERACT_LANGUAGES`	"eng,rus"	Языки распознавания. Указываются через запятую. Tesseract поддерживает более 100 языков, полный список вы можете найти в официальных источниках. **Примеры поддерживаемых языков и их кодовых обозначений:** • Английский: eng • Русский: rus • Украинский: ukr • Белорусский: bel • Армянский: arm • Казахский: kaz • Киргизский: kir • Азербайджанский: aze • Таджикский: tgk • Узбекский: uzb • Турецкий: tur
	`TESSERACT_REPLICAS`	1	Количество реплик экстрактора.
Tesseract_East	`TESSERACT_EAST_MODEL_TYPE`	standard	Тип модели экстрактора Tesseract по соотношению скорости и качества распознавания. Возможные значения: `standard` / `best` / `fast` .
	`TESSERACT_EAST_LANGUAGES`	"eng,rus"	Языки распознавания. Указываются через запятую. Поддерживаются те же языки, что и у экстрактора Tesseract.
	`TESSERACT_EAST_REPLICAS`	1	Количество реплик экстрактора.

12.8.5 Включение использования GPU

Чтобы экстрактор IWOCR использовал GPU:

1. Настройте доступ Data Analysis Service к GPU. См. *Использование GPU*.
2. Выберите экстрактор IWOCR.
3. Укажите, что экстрактор IWOCR должен использовать GPU. Для этого используйте переменную `IWOCR_USE_GPU` из таблицы выше.

Например, для включения использования GPU экстрактором IWOCR на центральной ноде, выполните команду:

```
kubectl set env deploy -n infowatch das-processor-central IWOCR_USE_GPU=TRUE
```

Чтобы вывести значения переменных окружения Data Analysis Service, используйте команду:

```
kubectl get deployment -n infowatch das-processor-<node_label> -o
jsonpath='{range .spec.template.spec.containers[0].env[*]}{.name}={.value}{"\n"}{end}'
```

12.9 Добавление языков для протокластеризации

Data Analysis Service осуществляет протокластеризацию текстовых документов с учетом морфологии на тех языках, для которых добавлены словари. После установки Сервис работает с русским и английским языками. Чтобы обрабатывать с учетом морфологии документы на других языках, необходимо добавить в Data Analysis Service словари этих языков. Вы можете получить их у специалистов компании InfoWatch.

> **Примечание:**
>
> Теги для документов также создаются только для тех языков, для которых добавлены словари.

Словарь включает в себя:

1. *Морфологический словарь языка в формате JSON* – содержит различные словоформы и используется для определения того, что встречаемые в документах формы принадлежат определенному слову.
   Пример названия файла: `az.json`
2. *Файл с нормальными формами слов в формате JSON* – содержит проиндексированный список нормальных форм слов для данного языка.
   Пример названия файла: `az_tokenizer_vocabulary.json`
3. *Файл со стоп-словами в формате TXT* – содержит слова, исключаемые из анализа текста, так как те не несут смысловой нагрузки, например, предлоги.
   Пример названия файла: `stopwords_az.txt`

Чтобы добавить в Data Analysis Service словарь языка для протокластеризации:

1. Скопируйте словарь в директорию `<das_data_dir>/nlp/tokenization/dicts/<vocabulary_dir>`
   где `<das_data_dir>` – это директория для размещения ресурсов Data Analysis Service, которая задается при установке. По умолчанию: `/opt/das_data`.
   `<vocabulary_dir>` – конечная директория для файлов словаря. В качестве имени она должна иметь двухбуквенный код добавляемого языка в формате ISO 639-1.
   Например, для азербайджанского языка в Data Analysis Service 1.4, это будет директория `/opt/das_data/nlp/tokenization/dicts/az`.
   Если директория не существует, создайте ее.
2. Перезапустите процесс **das-processor** с помощью команды:

```
kubectl rollout restart deploy -n infowatch das-processor-<das_node_label>
```

где `<das_node_label>` – это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды используйте лейбл `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

Важно!

После добавления словаря не изменяйте входящие в него файлы. Если вы измените какой-либо файл, Сервис не сможет обрабатывать контенты на данном языке.

12.10 Диагностика Data Analysis Service

12.10.1 Логирование

Лог-файлы Сервиса вы можете найти в директориях:

- `/var/log/pods/infowatch_das-processor-<node_label>-…/`
- `/var/log/pods/infowatch_das-adapter-<node_label>-…/`
- `/var/log/pods/infowatch_das-proto-cl-prov-<node_label>-…/`

где `<node_label>` – здесь и далее это лейбл ноды, на которую установлен Data Analysis Service. Для центральной ноды лейбл имеет значение `central`. Для офисной ноды лейбл задается при установке, по умолчанию: `office`.

Чтобы вывести логи, выполните команду:

```
kubectl logs $(kubectl get pods -n infowatch -o name|grep das-processor) -n infowatch
```

При наличии в кластере нескольких экземпляров Data Analysis Service дополнительно укажите в команде лейбл ноды:

```
kubectl logs $(kubectl get pods -n infowatch -o name|grep das-processor-<node_label>) -n infowatch
```

Чтобы найти в логах конкретное сообщение, добавьте команду `grep` с его текстом:

```
kubectl logs $(kubectl get pods -n infowatch -o name|grep das-processor) -n infowatch| grep
"<текст_сообщения>"
kubectl logs $(kubectl get pods -n infowatch -o name|grep das-processor-<node_label>) -n infowatch| grep
"<текст_сообщения>"
```

Чтобы изменить уровень логирования, выполните команду:

```
sudo kubectl -n infowatch set env deploy das-processor-<node_label> das-adapter-<node_label> das-proto-cl-
prov-<node_label> SRV_LOG_LEVEL=<value>
```

где `<value>` – это уровень логирования, для которого вы можете установить следующие значения:

- INFO – краткая информация о работе сервиса и возникающих ошибках. Установлен по умолчанию;
- DETAIL – более детальная информация, чем при значении INFO;
- DEBUG – подробная информация о работе сервиса и возникающих ошибках;
- TRACE – более расширенная информация, чем при значении DEBUG. Используйте для углубленного анализа работы Сервиса, если недостаточно данных, выводимых при значении DEBUG. Логи могут занимать значительное место.

Например, чтобы установить уровень логирования DEBUG для Сервиса на центральной ноде, выполните команду:

```
sudo kubectl -n infowatch set env deploy das-processor-central das-adapter-central das-proto-cl-prov-central SRV_LOG_LEVEL=DEBUG
```

12.10.2 Скрипт диагностики

Вы можете получить отчет о состоянии Сервиса, запустив скрипт диагностики `das-info.sh`. Для этого:

1. Перейдите в директорию `<DAS>/tools/diagnostics`, где `<DAS>` – это директория, в которую был распакован архив с дистрибутивом при установке.
2. Запустите скрипт с помощью команды:

   ```
   ./das-info.sh -n infowatch
   ```

В результате будет выведена следующая информация:

Блок в отчете	Описание
DAS & Platform Versions on Nodes	Версия Data Analysis Service и Платформы
GPU Status	Доступность GPU. Возможные значения: • `No GPU support` – ресурсы графического процессора недоступны на текущей ноде; • `GPU plugin found` – плагин установлен в кластере и ресурсы GPU доступны на текущей ноде.
DAS Pods Status	Список и статус подов
Errors in DAS Pod Logs	Список ошибок для каждого пода

Блок в отчете	Описание
Technology Flags	Статус Enabled/Disabled для каждой технологии
Environment Variables	Значения переменных окружения Data Analysis Service
Installed Models	Список моделей, установленных на текущей ноде

Вся информация кроме блоков **GPU Status** и **Installed Models** будет выведена по всем экземплярам Сервиса в кластере.

```
[redacted]:~/[redacted]/tools/diagnostics$ sudo ./das-info.sh -n infowatch

==================== DAS & Platform Versions on Nodes ====================
POD                                          NODE            DAS_VERSION     PLATFORM_VERSION
--------------------------------------------------------------------------------------------
das-processor-lv-426-7f4669df5f-59jhd        qa-l6c9         [redacted]      [redacted]

==================== GPU Status ====================
No GPU support

==================== DAS Pods Status ====================
NAME                                         READY   STATUS    RESTARTS      AGE
das-adapter-central-77ddb4bff-lxgsn          1/1     Running   0             24h
das-processor-central-58ff966877-smxzn       1/1     Running   0             3h13m
das-proto-cl-prov-central-5b66fc5d5f-ngg78   1/1     Running   0             24h

==================== Errors in DAS Pod Logs ====================

>> Pod: das-adapter-central-77ddb4bff-lxgsn
   Current logs:
      - no errors found
   Previous logs (--previous):
      - previous logs unavailable (pod not restarted)

>> Pod: das-processor-central-58ff966877-smxzn
   Current logs:
      - no errors found
   Previous logs (--previous):
      - previous logs unavailable (pod not restarted)

>> Pod: das-proto-cl-prov-central-5b66fc5d5f-ngg78
   Current logs:
      - no errors found
   Previous logs (--previous):
      - previous logs unavailable (pod not restarted)

==================== Technology Flags ====================
>> das-processor-central-58ff966877-smxzn (node: ubuntu-niokr, iwnode: central)
   - PROTO_CLUSTERIZATION_ENABLE: Enabled
   - TRANSCRIPTION_ENABLE: Enabled
   - IMAGE_CONTENT_DETECTION_ENABLE: Enabled
   - CLUSTERIZATION_ENABLE: Disabled
   - NAMED_ENTITIES_RECOGNITION_ENABLE: Enabled
   - OPTICAL_CHARACTER_RECOGNITION_ENABLE: Enabled

==================== Environment Variables ====================

>> das-processor-central-58ff966877-smxzn (node: ubuntu-niokr, iwnode: central)
    - TRANSFORMERS_CACHE=/tmp/hf_cache
    - NUMBA_CACHE_DIR=/tmp/numba_cache
    - MPLCONFIGDIR=/tmp/mpl_config
    - PYANNOTE_CACHE=/tmp/pyannote_cache
    - HUGGINGFACE_HUB_CACHE=/tmp/hf_hub_cache
    - SUBSCRIBE_TO=central
    - DAS_DATA_PATH=/opt/das_data
    - PROMETHEUS_PORT=9001
    - PROTO_CLUSTERIZATION=True
    - TRANSCRIPTION=True
    - IMG_CONTENT_DETECTION=True
    - NAMED_ENTITY_RECOGNITION=True
    - CLUSTERIZATION=False
    - OPTICAL_CHARACTER_RECOGNITION=True
    - PGUSER=from secret:pg-admin key:username
    - PGPASSWORD=from secret:pg-admin key:password
    - PGPORT=8080
    - NATSTOKEN=from secret:nats-admin key:password
    - SRV_PASSWORD=from secret:das-processor-das-guard-secret key:service_guard_password
    - NATS_PASSWORD=from secret:nats-admin key:password
    - NATS_USER=from secret:nats-admin key:username
    - SRV_LOG_LEVEL=DEBUG
    - USE_GPU=TRUE
    - TRANSCRIBER=whisper_large_v1_transcriber

==================== Installed Models (from DAS_DATA_PATH) ====================
>> Pod: das-processor-central-58ff966877-smxzn (node: ubuntu-niokr, iwnode: central)
   DAS_DATA_PATH = /opt/das_data
   [audio]
     - age
     - emotion_detection
     - gender
     - overlap
     - speechbrain_spkrec_ecapa_voxceleb
     - whisper
     - whisper_detection
```

Чтобы проверить только статус GPU, запустите скрипт с помощью команды:

```
./das-info.sh -n infowatch gpu_status
```

12.10.3 Диагностика технологий

С помощью скрипта вы можете проверить работоспособность интересующей вас технологии. Для этого запустите скрипт `das-info.sh` и убедитесь в том, что:

1. В блоке **DAS Pods Status** все поды имеют статус `Running`.
2. В блоке **Technology Flags** технология имеет статус `Enabled` (включена).
3. Блок **Installed Models** содержит модели технологии:
    - для транскрибации – список `audio`;
    - для детектирования объектов на изображениях – список `image`;
    - для протокластеризации – модели `fasttext` и `nltk` в списке `nlp`;
    - для распознавания именованных сущностей – модель `gliner_multi_v2_1_onnx` в списке `nlp`;
    - для OCR – модели `ocr` и `orientation` в списке `cv`.
4. Если вы проверяете технологию транскрибации и используете GPU, то:
    - в блоке **GPU Status** выведено сообщение `GPU plugin found`;
    - в блоке **Environment Variables**:
        - следующие переменные имеют значение `TRUE`:
            - `AGE_CLASSIFIER_USE_GPU`
            - `AUDIO_EMBEDDER_USE_GPU`
            - `EMOTION_DETECTOR_USE_GPU`
            - `FASTER_WHISPER_TRANSCRIBER_USE_GPU`
            - `WHISPER_DETECTOR_USE_GPU`
            - `T5_CORRECTOR_USE_GPU`
        - в случае использования большой модели переменная `FASTER_WHISPER_ TRANSCRIBER_MODEL_TYPE` имеет значение `large_v1`.
5. В блоке **Errors in DAS Pod Logs** отсутствуют ошибки.
   При наличии ошибок проанализируйте лог-файлы.

13 Удаление Data Analysis Service

Data Analysis Service удаляется с помощью командной строки.

> **Примечание:**
>
> Все задачи, активные на момент начала удаления, не будут завершены. Контенты не будут обработаны.
> Чтобы проверить, все ли задачи завершены, выполните команду:
>
> ```
> ps -aux|grep das_processor
> ```
>
> Если процесс **das_processor** не использует ресурсы сервера, то активных задач нет.
>
> Если необходимы результаты обработки конкретных контентов, дождитесь, когда эти результаты будут отображены в продукте - источнике контента, например, результаты транскрибации – в InfoWatch Activity Monitor. И только после этого начинайте процедуру удаления.

Чтобы удалить Data Analysis Service:

1. Перейдите в директорию, в которую был распакован дистрибутив последней установленной версии Сервиса.
2. Выполните команду:

   ```
   ./setup.py remove
   ```

3. Дождитесь завершения процесса. Инсталлятор удалит компоненты Сервиса. Данные Data Analysis Service удалены не будут. При необходимости вы можете удалить данные вручную из директорий, указанных при установке.
   Инсталлятор создаст файл в формате `.env` с настройками данного экземпляра Data Analysis Service. Этот файл будет сохранен в директории `/var/tmp/das_env_migration/raw/` .
   Если в дальнейшем вы будете устанавливать Сервис заново, вы сможете применить сохраненные настройки.
4. Чтобы в веб-интерфейсе Activity Monitor не отображались настройки, связанные с технологиями Data Analysis Service, выключите их в Activity Monitor. Для этого на ноде, где Activity Monitor установлен в режиме `central`, выполните следующие команды:
   a. Для выключения транскрибации:

      ```
      kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<am_central_node>
      | awk '//{print $1;}') -n infowatch -- psql -U postgres -d activitymonitor -c
      "update features set status='disabled' where name='stream_transcripts'"
      kubectl rollout restart deployment datacollector-<am_central_node> -n infowatch
      ```

      где `<am_central_node>` – здесь и далее лейбл центральной ноды Activity Monitor.
   b. Для выключения детектирования объектов на изображениях:

```
kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<am_central_node>
| awk '//{print $1;}') -n infowatch -- psql -U postgres -d activitymonitor -c
"update features set status='disabled' where name='stream_enrich'"
kubectl rollout restart deployment datacollector-<am_central_node> -n infowatch
```

c. Для выключения распознавания текста (OCR):

```
kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<am_central_node>
| awk '//{print $1;}') -n infowatch -- psql -U postgres -d activitymonitor -c
"update features set status='disabled' where name='screenshot_transcripts'"
kubectl rollout restart deployment datacollector-<am_central_node> -n infowatch
```

Примечание:

В случае распределенной установки, если Data Analysis Service установлен на офисной ноде и на ней отсутствуют другие продукты, то после удаления Data Analysis Service также рекомендуется удалить компоненты платформы. Для этого после выполнения команды `./setup.py remove`, также выполните команду:

```
./setup.py reset
```

Если Data Analysis Service установлен на центральной ноде или если на офисной ноде вместе с ним также установлены другие продукты, выполнять команду `./setup.py reset` категорически не рекомендуется.

Чтобы узнать, какие продукты установлены в кластере, вы можете перейти в директорию с инсталлятором и выполнить команду:

```
./setup.py showproducts
```

Примечание:

Если вы удалили Data Analysis Service, установленный в режиме `central`, компоненты Data Analysis Service, установленные в режиме `office` на других нодах, не будут функционировать и не будут удалены автоматически. Удалите их вручную.