INFOWATCH

Руководство по установке, конфигурированию и администрированию

17/09/2024

© АО “ИнфоВотч”

Тел./Факс +7 (495) 229-00-22

http://www.infowatch.ru

СОДЕРЖАНИЕ

1 Аудитория 5

2 Комплект документов 6

2.1 Техническая поддержка пользователей 6

3 Обзор 7

3.1 Состав Data Analysis Service: 7

4 Типы установки, совместимость с другими продуктами, аппаратные и программные требования 8

4.1 Типы установки и совместимость с другими продуктами InfoWatch 8

4.2 Аппаратные требования 9

4.3 Программные требования 10

5 Подготовка сервера к установке/обновлению Data Analysis Service 12

6 Установка Data Analysis Service 14

6.1 Варианты запуска установки Data Analysis Service 14

6.2 Установка Data Analysis Service в интерактивном режиме 14

6.3 Установка Data Analysis Service в тихом режиме 19

7 Обновление Data Analysis Service 27

7.1 Варианты обновления Data Analysis Service 27

7.2 Прямое обновление Data Analysis Service 28

7.3 Обновление Data Analysis Service с помощью сброса установки (reset) 32

8 Настройка Data Analysis Service 34

8.1 Настройка взаимодействия с продуктами Платформы 34

8.1.1 Задание ноды обработки 34

8.1.2 Включение функциональностей Data Analysis Service в продуктах Платформы 35

Activity Monitor 35

Data Discovery 36

8.2 Включение/выключение модулей Data Analysis Service 36

8.3 Настройка производительности 37

8.3.1 Создание реплик POD 37

8.3.2 Многонодовая установка 38

8.3.3 Оценка производительности модуля протокластеризации 38
8.4 Настройка использования ресурсов сервера 38
8.5 Настройка модуля транскрибации 39
9 Удаление Data Analysis Service 43

Настоящее руководство содержит сведения по установке, удалению и настройке компонентов сервиса InfoWatch Data Analysis Service (далее также 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 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, настройки 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.

2.1 Техническая поддержка пользователей

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

- если вы приобрели Data Analysis Service 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.

Часы работы службы технической поддержки – с 7:00 до 21:00 по московскому времени с понедельника по пятницу, исключая официальные выходные и праздничные дни РФ. Вы также можете посетить раздел технической поддержки на нашем сайте: https://www.infowatch.ru/services/support.

3 Обзор

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

3.1 Состав Data Analysis Service:

- **Модуль протокластеризации текстовых данных**
  Данный модуль позволяет анализировать текстовые документы и разбивать 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", глава "Решение задач при работе в консоли Traffic Monitor", раздел "Работа с объектами перехвата" -> "Просмотр событий" -> "Просмотр краткой формы события", статья "Поиск аналогов";
  - *Поиск похожих файлов* в InfoWatch Data Discovery. Подробнее см. "InfoWatch Data Discovery. Руководство пользователя", "Раздел "Хранение файлов", "Вкладка "Похожие файлы".
- **Модуль транскрибации**
  Данный модуль позволяет конвертировать аудиозапись в текст. Модуль транскрибации используется при работе с аудиозаписями в InfoWatch Activity Monitor. Подробнее см. "InfoWatch Activity Monitor. Руководство пользователя", глава "Интерфейс Центра расследований", "Раздел "Мониторинг", "Вкладка "Аудиозаписи", статья "Работа с аудиозаписями".
- **Модуль детектирования объектов на изображениях**
  Данный модуль позволяет распознавать объекты на снимках экрана, в том числе непристойный контент. Детектирование объектов на изображениях реализовано для InfoWatch Activity Monitor.

4 Типы установки, совместимость с другими продуктами, аппаратные и программные требования

4.1 Типы установки и совместимость с другими продуктами InfoWatch

Data Analysis Service поддерживает интеграцию со следующими продуктами InfoWatch:

- **Traffic Monitor начиная с версии 7.8**
- **Activity Monitor начиная с версии 3.0**
- **Data Discovery 1.10**

> **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. Если вы установите в кластер Сервис или продукт на новой версии Платформы, обновите все экземпляры Сервиса и все продукты.
> - В одном кластере можно развернуть только одну центральную ноду Платформы и Data Analysis Service в режиме `central` только в единственном экземпляре.
> - 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.
> - Каждый экземпляр Data Analysis Service может обрабатывать очередь сообщений только одной ноды. Если платформенные продукты отправляют сообщения для Сервиса на разные ноды, то для интеграции с каждым из них понадобится

отдельный экземпляр Сервиса. Подробнее см. "Настройка взаимодействия с продуктами Платформы".

При установке Data Analysis Service на сервер, где Платформа ещё не установлена, устанавливается следующая ее версия:

- 2.0.1 – для Data Analysis Service 1.2.0
- 2.0.2 – для Data Analysis Service 1.2.1

Data Analysis Service 1.2.1 совместим с версиями Платформы от 2.0.0 до 2.1.0.

Сервис может быть установлен на одном сервере со следующими продуктами InfoWatch:

- **Traffic Monitor начиная с версии 7.8**
- **Activity Monitor начиная с версии 3.0**
- **Data Discovery начиная с версии 1.9**
- **Vision начиная с версии 3.2**

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

4.2 Аппаратные требования

Минимальные системные требования, необходимые для установки:

Процессор	ОЗУ	Жесткий диск
Количество ядер – 4; Частота – 3.1 ГГц; Разрядность – 64 bit; Поддержка инструкций SSE 4.2.	16 Гбайт	от 150 Гбайт 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, используемый для обработки данных	Модель, используемая при транскрибации	Производительность модуля	Рекомендуемое аппаратное обеспечение			
				При установке отдельно от других платформенных продуктов		При установке на один сервер с другим платформенным продуктом	
			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	2	4
–	–	–	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

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.

4.3 Программные требования

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

Тип ПО	Варианты
Операционная система	• Astra Linux Special Edition 1.6 "Смоленск": • Update 10; • Astra Linux Special Edition 1.7 "Смоленск": • версии 1.7.x (начиная с 1.7.0-11.06.2021_12.40); • Red Hat Enterprise Linux: • версии 7.x (начиная с 7.7); • версии 8.x (начиная с 8.0); • Oracle Linux: • версии 7.x (начиная с 7.9); • версии 8.x (начиная с 8.4); • РЕД ОС: • версии 7.x.x (начиная с версии 7.3.1).

Примечание:

Data Analysis Service использует PostgreSQL 13.8. СУБД поставляется вместе с продуктом и не требует отдельной установки.

Примечание:

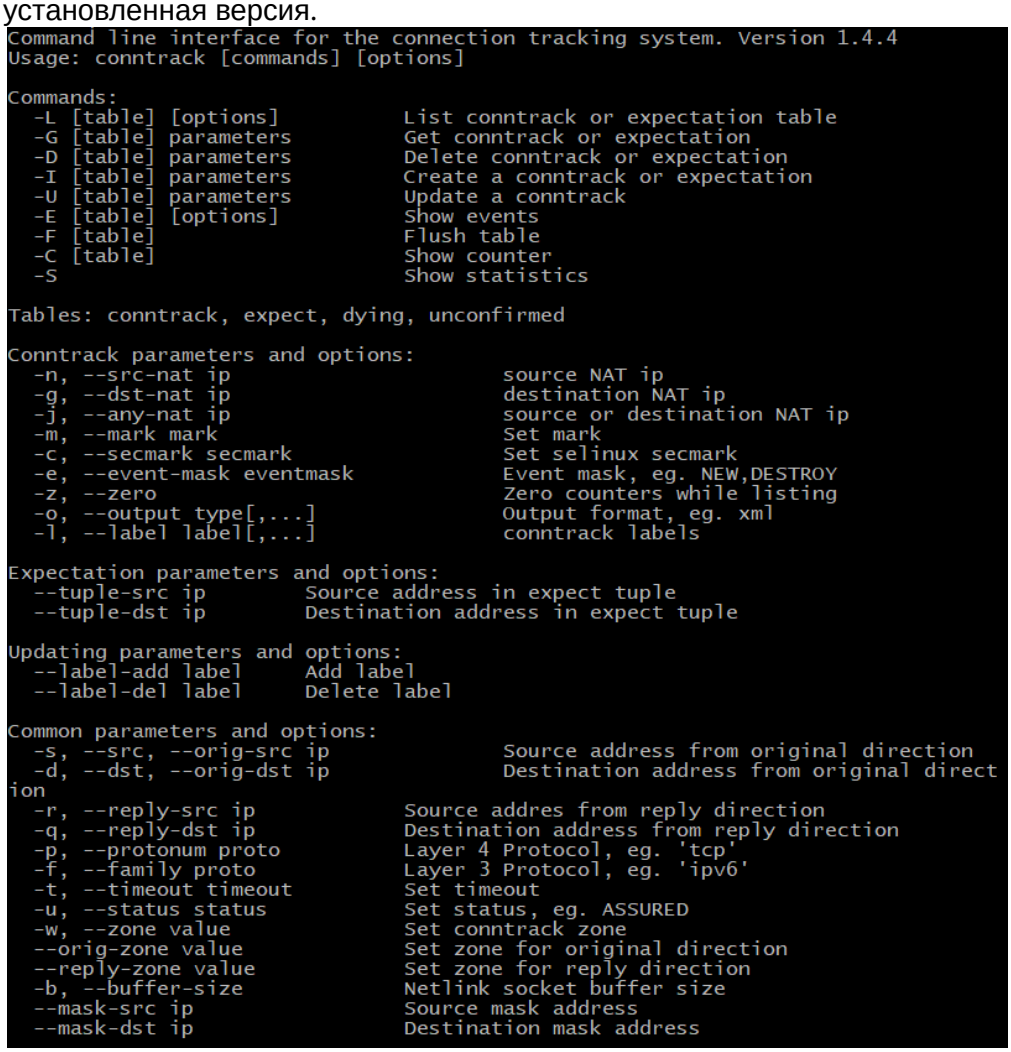
Требования к графическому процессору (GPU) для его использования при транскрибации приведены в статье Настройка модуля транскрибации.

5 Подготовка сервера к установке/обновлению Data Analysis Service

Перед установкой или обновлением Data Analysis Service обязательно подготовьте сервер:

1. Программа установки написана на языке Python версии 2.7, поэтому для установки Сервиса установите интерпретатор этого языка. Пакет с интерпретатором не входит в состав дистрибутива Сервиса.
2. В окружении Red Hat Enterprise Linux и РЕД ОС обновите политики SELinux. Для этого скачайте и установите пакет `container-selinux`.
   Вы можете установить пакет, выполнив команду:
   `yum install container-selinux`
3. На сервере под управлением ОС Red Hat Enterprise Linux и РЕД ОС отключите систему контроля доступа SELinux. Для этого:
   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`
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 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`
6. Если вы используете антивирус, настройте его для совместной работы с Data Analysis Service (см. статью Базы знаний "Настройка антивируса для продуктов Платформы").
7. Data Analysis Service использует компонент Kubernetes. Для корректной установки Сервиса и работы Kubernetes на сервере должен быть установлен пакет *conntrack*. Убедитесь, что пакет установлен. Для этого введите в командной строке:
   `conntrack`
   В результате будут выведены сведения о работе с модулем *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 модуль не установлен, установите его с помощью команды:

- `yum install conntrack` – для ОС Red Hat Enterpise Linux, Oracle Linux и РЕД ОС;
- `apt-get install conntrack` – для ОС Astra Linux.

Если пакет недоступен в репозитории, скачайте его, перейдите в директорию с пакетом и установите его вручную с помощью команды вида:

- `rpm -Uhv ./<имя_пакета>` – для ОС Red Hat Enterpise Linux, Oracle Linux и РЕД ОС;
- `dpkg -i ./<имя_пакета>` – для ОС Astra Linux.

6 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 `1.2.0.38`, ей соответствует архив `iw_das_setup_1.2.0.38.tar.xz`. Архив содержит полный набор бинарных модулей образов контейнеров, необходимых для развертывания Сервиса без доступа к сети Интернет. Также архив содержит программу установки.

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, совместимость с другими продуктами, аппаратные и программные требования";
- в кластере с платформенными продуктами.

6.1 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 в тихом режиме").

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

6.2 Установка Data Analysis Service в интерактивном режиме

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

1. Выполните шаги по подготовке сервера к установке Data Analysis Service (см. "Подготовка сервера к установке/обновлению Data Analysis Service").
2. Если вы устанавливаете Сервис на отдельный сервер в кластере с платформенными продуктами, и на других нодах кластера версия Платформы выше, чем в дистрибутиве Сервиса, например, 2.1.0 (подробнее о совместимости Сервиса с продуктами и Платформой см. "Типы установки, совместимость с другими продуктами, аппаратные и программные требования"), то установите на сервере Data Analysis Service Платформу более высокой версии.
   Для этого:
   a. На сервере Data Analysis Service создайте новую директорию (например, `Cluster_platform`):

```
mkdir Cluster_platform
```

b. Скопируйте в созданную директорию архив с инсталлятором платформенного продукта, который имеет необходимую версию Платформы. Рекомендуется использовать инсталлятор продукта, который уже установлен в кластере. В нашем примере это будет InfoWatch Data Discovery 1.10. Для него будет использоваться архив `iw_discovery_setup_1.10.0.193.tar.xz` .

> **Примечание:**
>
> Если архив с инсталлятором платформенного продукта уже распакован на другом сервере, вместо копирования всего архива вы можете:
>
> i. Перейти в директорию, куда распакован архив.
> ii. Любым удобным способом, например, с помощью внешнего носителя, скопировать папку `platform` на сервер Data Analysis Service.
> iii. Перейти в директорию, куда была скопирована папка `platform` .
> iv. Пропустить шаги **c** и **d** и перейти к шагу **e**.

c. Перейдите в директорию, в которую был скопирован архив.
d. Распакуйте архив в эту директорию:

```
tar -xvf iw_discovery_setup_1.10.0.193.tar.xz
```

e. Перейдите в директорию `platform` .
f. Запустите программу установки офисной ноды Платформы:

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

g. Выполните шаги с 11 по 19 данной инструкции.
h. Выполните шаги 25 и 26 данной инструкции.
Убедитесь, что на сервере Data Analysis Service установлена Платформа той же версии, что и на других нодах кластера.

3. Создайте новую директорию на сервере (например, `DAS`):

```
mkdir DAS
```

4. Скопируйте архив **`iw_das_setup_1.2.x.xx.tar.xz`** в созданную директорию.
В нашем примере будет использована версия `1.2.0.38` , соответственно, архив - `iw_das_setup_1.2.0.38.tar.xz` .
5. Перейдите в директорию, в которую был скопирован архив.
6. Распакуйте архив с дистрибутивом в эту директорию:

```
tar -xvf iw_das_setup_1.2.0.38.tar.xz
```

7. 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`.

Если на шаге 2 вы уже установили Платформу, то параметр `--nodemode` не нужно указывать.

Важно!

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

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

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

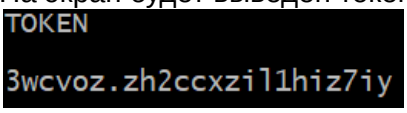
10. Если вы устанавливаете **центральную ноду Платформы** (`--nodemode=central`), введите IP-адрес сетевого интерфейса для взаимодействия с кластером в формате IPv4: "xxx.xxx.xxx.xxx" (по умолчанию: `0.0.0.0`) и нажмите **Enter**.
Если указать `0.0.0.0` , будут использованы все доступные сетевые интерфейсы.

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

11. 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.
12. Выделите объем оперативной памяти для размещения данных Clickhouse (по умолчанию: `80%`) и нажмите **Enter**.
13. Укажите путь для размещения данных Clickhouse (по умолчанию: `/mnt/chdata`) и нажмите **Enter**.
14. Укажите путь для размещения данных PostgreSQL (по умолчанию: `/mnt/pgdata`) и нажмите **Enter**.
15. Укажите путь для размещения данных NATS (по умолчанию: `/mnt/natsdata`) и нажмите **Enter**.
16. Укажите путь для размещения бинарных данных (по умолчанию: `/mnt/dsdata`) и нажмите **Enter**.
17. 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.

18. Дождитесь завершения установки Платформы.
19. Ознакомьтесь с отчетом об установке Платформы.



20. Введите лейбл ноды для взаимодействия с платформенными продуктами (по умолчанию: `central`) и нажмите **Enter**. Data Analysis Service будет обрабатывать очередь сообщений 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:

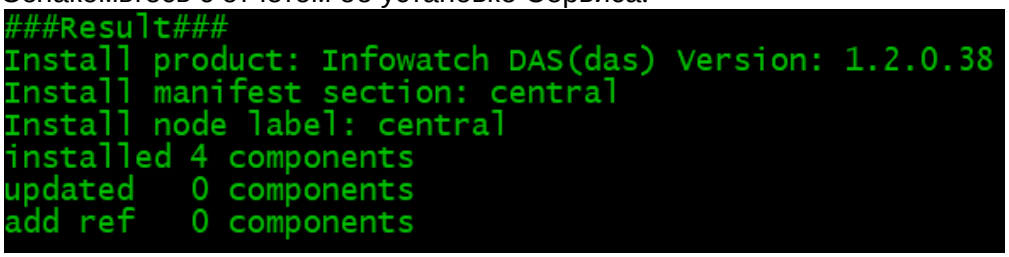
```
kubectl get node -n infowatch -o jsonpath="{range .items[*]}
{.metadata.annotations.flannel\.alpha\.coreos\.com/public\-ip}
{.metadata.labels.kubernetes\.io/hostname} {.metadata.labels.iwnode}{'\n'}{end}"
```

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

> **Примечание:**
>
> Изменить лейбл ноды для взаимодействия можно после установки. Подробнее см. "Настройка взаимодействия с продуктами Платформы".

21. Введите "**y**", чтобы включить модуль транскрибации, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.

22. Введите "**y**", чтобы включить модуль детектирования объектов на изображениях, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.
23. Дождитесь окончания процесса установки Data Analysis Service.
24. Ознакомьтесь с отчетом об установке Сервиса:



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

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

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

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

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

27. 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.

Чтобы начать использование Data Analysis Service, настройте его взаимодействие с InfoWatch Traffic Monitor (подробнее см. документ "*InfoWatch Traffic Monitor. Руководство администратора*", статья "Настройка интеграции с Data Analysis Service") или продуктом Платформы (подробнее см. в статье "Настройка взаимодействия с продуктами Платформы" и в документации на соответствующий продукт Платформы).

6.3 Установка Data Analysis Service в тихом режиме

Для минимизации участия в процессе установки Data Analysis Service, в инсталляторе предусмотрен тихий режим установки.

Перед запуском инсталлятора Data Analysis Service:

1. Выполните шаги по подготовке сервера к установке Data Analysis Service (см. "Подготовка сервера к установке/обновлению Data Analysis Service").
2. Если вы устанавливаете Сервис на отдельный сервер в кластере с платформенными продуктами, и на других нодах кластера версия Платформы выше, чем в дистрибутиве Сервиса, например, 2.1.0 (подробнее о совместимости Сервиса с продуктами и Платформой см. "Типы установки, совместимость с другими продуктами, аппаратные и программные требования"), то установите на сервере Data Analysis Service Платформу более высокой версии.
   Для этого:
   a. На сервере Data Analysis Service создайте новую директорию (например, `Cluster_platform`):

```
mkdir Cluster_platform
```

b. Скопируйте в созданную директорию архив с инсталлятором платформенного продукта, который имеет необходимую версию Платформы. Рекомендуется использовать инсталлятор продукта, который уже установлен в кластере. В нашем примере это будет InfoWatch Data Discovery 1.10. Для него будет использоваться архив `iw_discovery_setup_1.10.0.193.tar.xz` .

> **Примечание:**
>
> Если архив с инсталлятором платформенного продукта уже распакован на другом сервере, вместо копирования всего архива вы можете:
>
> i. Перейти в директорию, куда распакован архив.
> ii. Любым удобным способом, например, с помощью внешнего носителя, скопировать папку `platform` на сервер Data Analysis Service.
> iii. Перейти в директорию, куда была скопирована папка `platform` .
> iv. Пропустить шаги **c** и **d** и перейти к шагу **e**.

c. Перейдите в директорию, в которую был скопирован архив.
d. Распакуйте архив в эту директорию:

```
tar -xvf iw_discovery_setup_1.10.0.193.tar.xz
```

e. Перейдите в директорию `platform` .
f. Запустите программу установки офисной ноды Платформы с помощью команды со следующими параметрами:

```
./setup.py install \
--nodemode=office \
--cluster=10.60.25.92 \
--clustertoken=3wcvoz.zh2ccxzil1hiz7iy \
--nodelabel=office123
--chmemsize=60% \
--chdbpath=/mnt/chdata \
--pgdbpath=/mnt/pgdata \
--natspath=/mnt/natsdata \
--dsdatapath=/mnt/dsdata
```

Параметры, указанные в команде, описаны в таблице ниже и могут иметь значения отличные от приведенных.

g. Убедитесь что все сервисы запущены, выполнив команду:

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

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

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

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

Убедитесь, что на сервере Data Analysis Service установлена Платформа той же версии, что и на других нодах кластера.

3. Создайте новую директорию на диске (например, `DAS`):

```
mkdir DAS
```

4. Скопируйте архив с дистрибутивом Data Analysis Service **`iw_das_setup_1.2.x.xx.tar.xz`** в созданную директорию.
   В нашем примере будет использована версия `1.2.0.38`, соответственно, архив – `iw_das_setup_1.2.0.38.tar.xz`.
5. Перейдите в директорию, в которую был скопирован архив.
6. Распакуйте архив в эту директорию:

```
tar -xvf iw_das_setup_1.2.0.38.tar.xz
```

Ознакомьтесь со списком ключей в таблице ниже, с их помощью вы можете указать инсталлятору параметры установки продукта. Указанные значения по умолчанию - значения, предлагаемые инсталлятором в интерактивном режиме.

> **Примечание:**
>
> Если вы не укажете режимы установки Платформы и продукта, инсталлятор использует значение `central`.

Ключ	Значение по умолчанию	Доступные значения	Описание
`--nodemode`	`central`	`central, office`	Режим установки Платформы
`--productmode`	`central`	`central, office`	Режим установки Data Analysis Service
`--licagree`	`n`	`y, yes, n, no`	Принятие лицензионного соглашения
`--advertise`	`0.0.0.0`		IP-адрес сетевого интерфейса для взаимодействия с кластером.

Ключ	Значение по умолчанию	Доступные значения	Описание
			Указывайте при установке центральной ноды Платформы
`--chmemsize`	`80%`	1-100% или числовое значение Если не указать %, указанное значение будет принято за объем в GB	Объем оперативной памяти, используемой для данных Clickhouse
`--chdbpath`	`/mnt/chdata`		Путь для размещения данных ClickHouse
`--pgdbpath`	`/mnt/pgdata`		Путь для размещения файлов базы данных PostgreSQL
`--natspath`	`/mnt/natsdata`		Путь для размещения данных NATS
`--dsdatapath`	`/mnt/dsdata`		Путь для размещения бинарных данных.
`--guiport`	`443`		Порт подключения к веб-интерфейсу. Указывайте при установке центральной ноды Платформы. **Важно!** 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.

Ключ	Значение по умолчанию	Доступные значения	Описание
`--cluster`			Адрес центральной ноды Платформы в кластере. Указывайте 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` Указывайте при установке офисной ноды Платформы
`--nodelabel`	`office`		Лейбл ноды. Указывайте при установке офисной ноды Платформы
`--pclust-enable`	`y`	`y, n`	Включение модуля протокластеризации
`--transcribation-enable`	`n`	`y, n`	Включение модуля транскрибации
`--imcdet-enable`	`n`	`y, n`	Включение модуля детектирования объектов на изображениях
`--subscribe-to`	`central`		Лейбл ноды для взаимодействия с платформенными продуктами. Data Analysis Service будет обрабатывать очередь сообщений указанной ноды. **Примечание:** InfoWatch Activity Monitor отправляет сообщения в очередь NATS на центральную ноду кластера.

Ключ	Значение по умолчанию	Доступные значения	Описание
			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 лейбл ноды для взаимодействия можно после установки. Подробнее см. "Настройка взаимодействия с продуктами Платформы".

Чтобы установить Data Analysis Service:

1. Запустите программу установки Data Analysis Service с указанием параметров. Если на сервере уже установлена Платформа, то не указывайте относящиеся к ней параметры.
   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=/mnt/chdata \
--pgdbpath=/mnt/pgdata \
--natspath=/mnt/natsdata \
--dsdatapath=/mnt/dsdata \
--guiport=443 \
--pclust-enable=y \
--transcribation-enable=y \
--imcdet-enable=y
```

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

```
./setup.py install \
--nodemode=office \
--productmode=central \
--licagree=yes \
--cluster=10.60.25.92 \
--clustertoken=3wcvoz.zh2ccxzil1hiz7iy \
--nodelabel=office123
--chmemsize=60% \
--chdbpath=/mnt/chdata \
--pgdbpath=/mnt/pgdata \
--natspath=/mnt/natsdata \
--dsdatapath=/mnt/dsdata \
--pclust-enable=y \
--transcribation-enable=n \
--imcdet-enable=n \
--subscribe-to=central
```

Пример команды установки Data Analysis Service в режиме central на ранее установленную Платформу:

```
./setup.py install \
--productmode=central \
--licagree=yes \
--pclust-enable=y \
--transcribation-enable=n \
--imcdet-enable=n \
--subscribe-to=central
```

2. Если вы указали все обязательные параметры, инсталлятор запустит установку.
   Если вы не указали часть параметров, инсталлятор запустит интерактивный режим и попросит задать значения.
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.

Чтобы начать использование Data Analysis Service, настройте его взаимодействие с InfoWatch Traffic Monitor (подробнее см. документ "*InfoWatch Traffic Monitor. Руководство администратора*", статья "Настройка интеграции с Data Analysis Service") или продуктом Платформы (подробнее см. в статье "Настройка взаимодействия с продуктами Платформы" и в документации на соответствующий продукт Платформы).

7 Обновление Data Analysis Service

Важно!

Поддерживается обновление с версии Data Analysis Service 1.1.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 полный набор бинарных модулей образов контейнеров, необходимых для развертывания Сервиса без доступа к сети Интернет. Также архив содержит программу обновления.

Важно!

Для успешного обновления Data Analysis Service необходимо не менее 150 Гбайт 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.

7.1 Варианты обновления Data Analysis Service

Data Analysis Service поддерживает:

- прямое стандартное обновление;
- 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 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.

> **Примечание:**
>
> Продукты в кластере должны быть установлены на одной версии Платформы. Если вы установите в кластер продукт на новой версии Платформы, для корректной работы вам будет необходимо обновить все продукты.

7.2 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. Выполните шаги по подготовке сервера к обновлению (см. "Подготовка сервера к установке/обновлению Data Analysis Service").
2. Если на сервере установлены продукты Платформы, обновите их до версий, совместимых с Data Analysis Service (см. "Типы установки, совместимость с другими продуктами, аппаратные и программные требования").
3. Если вы обновляете Сервис, установленный на отдельном сервере в распределенном кластере с платформенными продуктами, и после обновления кластера версия Платформы на других нодах будет выше, чем в дистрибутиве Сервиса, например, 2.1.0 (подробнее о совместимости Сервиса с продуктами и Платформой см. "Типы установки, совместимость с другими продуктами, аппаратные и программные требования"), то на сервере Data Analysis Service обновите Платформу до более высокой версии.
   Для этого:
   a. На сервере Data Analysis Service создайте новую директорию (например, `Cluster_platform`):

   ```
   mkdir Cluster_platform
   ```

   b. Скопируйте в созданную директорию архив с инсталлятором платформенного продукта, который имеет необходимую версию Платформы. Рекомендуется использовать инсталлятор продукта, который уже обновлен в кластере. В нашем примере это будет InfoWatch Data Discovery 1.10. Для него будет использоваться архив `iw_discovery_setup_1.10.0.193.tar.xz` .

   > **Примечание:**

Если архив с инсталлятором платформенного продукта уже распакован на другом сервере, вместо копирования всего архива вы можете:

i. Перейти в директорию, куда распакован архив.
ii. Любым удобным способом, например, с помощью внешнего носителя, скопировать папку `platform` на сервер Data Analysis Service.
iii. Перейти в директорию, куда была скопирована папка `platform`.
iv. Пропустить шаги **c** и **d** и перейти к шагу **e**.

c. Перейдите в директорию, в которую был скопирован архив.
d. Распакуйте архив в эту директорию:

```
tar -xvf iw_discovery_setup_1.10.0.193.tar.xz
```

e. Перейдите в директорию `platform`.
f. Запустите программу обновления Платформы:

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

g. Выполните шаги 15 и 16 данной инструкции.
Убедитесь, что на сервере Data Analysis Service Платформа обновлена до той же версии, что и на других обновленных нодах кластера.

4. Создайте новую директорию на диске (например, `DAS`):

```
mkdir DAS
```

5. Скопируйте архив **`iw_das_setup_1.2.x.xx.tar.xz`** в созданную директорию.
В нашем примере будет использована версия `1.2.0.38`, ей соответствует архив `iw_das_setup_1.2.0.38.tar.xz`.
6. Перейдите в директорию, в которую был скопирован архив.
7. Распакуйте архив с дистрибутивом Data Analysis Service в эту директорию:

```
tar -xvf iw_das_setup_1.2.0.38.tar.xz
```

8. 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!

9. Запустите программу обновления Сервиса:

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

Важно!

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

10. При необходимости инсталлятор обновит Платформу. В этом случае дождитесь окончания этого процесса.
11. Введите лейбл ноды для взаимодействия с платформенными продуктами (по умолчанию: `central`) и нажмите **Enter**. Data Analysis Service будет обрабатывать очередь сообщений 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:

a. Подключитесь по ssh к консоли той ноды Платформы, с которой необходимо настроить взаимодействие;
b. Перейдите в директорию, где был распакован инсталлятор продукта Платформы;
c. Выполните команду:

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

В результате будет выведен список установленных продуктов в кластере. Для каждого в графе `Node` будут указаны имя сервера и лейбл ноды в скобках.

> **Примечание:**
>
> Изменить лейбл ноды для взаимодействия можно после установки. Подробнее см. "Настройка взаимодействия с продуктами Платформы".

12. Введите "**y**", чтобы включить модуль детектирования объектов на изображениях, или "**n**", чтобы оставить его выключенным, (по умолчанию: "**n**") и нажмите **Enter**.
13. Дождитесь окончания процесса обновления Data Analysis Service.
14. Ознакомьтесь с отчетом об установке Сервиса:



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

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

Команда выведет список сервисов. Каждый сервис должен иметь статус `Running`.
16. Чтобы вывести на экран информацию об установленных продуктах, выполните команду:

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

17. Если Data Analysis Service установлен в распределенном кластере, обновите Платформу, другие экземпляры Сервиса и продукты Платформы на других серверах кластера до версий, совместимых с Data Analysis Service.

18. Если после обновления вам необходимо освободить место на диске, вы можете удалить неиспользуемые контейнеры аналогично п.8.

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

> **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.

Обновление состоит из следующих шагов:

1. Сброс установки. При сбросе данные Сервиса не будут удалены.
2. Установка новых версий Платформы и Сервиса с указанием уже имеющихся папок с данными.

Чтобы обновить Data Analysis Service:

1. Выполните шаги по подготовке сервера к обновлению (см. "Подготовка сервера к установке/обновлению Data Analysis Service").
2. Создайте новую директорию на диске (например, `DAS`):

   ```
   mkdir DAS
   ```

3. Скопируйте архив **`iw_das_setup_1.2.x.xx.tar.xz`** в созданную директорию.
   В нашем примере будет использована версия `1.2.0.38`, ей соответствует архив `iw_das_setup_1.2.0.38.tar.xz`.
4. Перейдите в директорию, в которую был скопирован архив.
5. Распакуйте архив с дистрибутивом в эту директорию:

   ```
   tar -xvf iw_das_setup_1.2.0.38.tar.xz
   ```

6. При обновлении вы можете присвоить 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}"
   ```

7. Выполните сброс текущей установки:

   ```
   ./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.

8. Дождитесь окончания процесса сброса установки Платформы и Сервиса.
9. Перезагрузите сервер с помощью команды:

```
reboot
```

10. Если вы обновляетесь с переходом на другой сервер или на новую ОС:
    a. Скопируйте на другой раздел или на новый сервер директорию с данными PostgreSQL, которая была указана при установке.
    По умолчанию директория расположена по пути `/mnt/pgdata`. Обязательно копируйте директорию с сохранением прав и владельцев. Для копирования используйте стандартные средства, например утилиту **`rsync`**.
    b. После сброса текущей установки скопируйте со старого сервера на новый файл с паролями базы данных `/var/lib/iwplatform/` **`dbadmin.yaml`**.
11. 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
```

12. Установите новую версию Data Analysis Service (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?»

13. Если до сброса на сервере были установлены и другие продукты, установите версии продуктов, совместимые с Data Analysis Service (см. "Типы установки, совместимость с другими продуктами, аппаратные и программные требования").
14. Если Data Analysis Service установлен в распределенном кластере, обновите Платформу и продукты на других серверах кластера до версий, совместимых с Data Analysis Service.
15. Если после обновления вам необходимо освободить место на диске, вы можете удалить неиспользуемые контейнеры. Подробнее см. п.8 статьи "Прямое обновление Data Analysis Service".

8 Настройка Data Analysis Service

В этом разделе:

- Настройка взаимодействия с продуктами Платформы
- Включение/выключение модулей Data Analysis Service
- Настройка производительности
- Настройка использования ресурсов сервера
- Настройка модуля транскрибации

8.1 Настройка взаимодействия с продуктами Платформы

- Задание ноды обработки
- Включение функциональностей Data Analysis Service в продуктах Платформы
  - Activity Monitor
  - Data Discovery

8.1.1 Задание ноды обработки

При установке Data Analysis Service в кластере с продуктами Платформы, для Сервиса необходимо задать ноду, очередь сообщений которой он будет обрабатывать. Для каждого экземпляра Data Analysis Service может быть задана только одна такая нода.

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_label>
```

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_label>` – лейбл ноды, на которую продукты Платформы отправляют сообщения для Data Analysis Service.

> **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.

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
```

8.1.2 Включение функциональностей Data Analysis Service в продуктах Платформы

Activity Monitor

Data Analysis Service позволяет использовать в Activity Monitor траснкрибацию аудиозаписей и детектирование объектов, в том числе непристойного контента, на изображениях.

Чтобы включить эти функциональности 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. выше.
2. Убедитесь, что модуль транскрибации и/или модуль детектирования объектов на изображениях включены. Подробнее см. "Включение/выключение модулей Data Analysis Service".
3. Включите транскрибацию данных, выполнив на центральной ноде кластера следующие команды:

   ```
   kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<AM_NODE_LABEL> | 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_NODE_LABEL> -n infowatch
   ```

   где `<AM_NODE_LABEL>` – это лейбл ноды, на которой установлен Activity Monitor.

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

4. Включите детектирование данных, выполнив на центральной ноде кластера следующие команды:

   ```
   kubectl exec -it $(kubectl get pods -n infowatch | grep postgresql-<AM_NODE_LABEL> | 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_NODE_LABEL> -n infowatch
   ```

   где `<AM_NODE_LABEL>` – это лейбл ноды, на которой установлен Activity Monitor.

   > **Примечание:**

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

Data Discovery

Data Analysis Service позволяет использовать в Data Discovery протокластеризацию текстовых документов для поиска похожих.

Чтобы эта функциональность работала в Data Discovery, достаточно установить Data Analysis Service в том же кластере и настроить Сервис на обработку ноды, на которой установлен Data Discovery. Лейбл этой ноды указывается при установке Сервиса. Также его можно задать после установки. Подробнее см. выше.

После этого в Data Discovery вы сможете включить опцию "**Определять похожие файлы**" при создании или редактировании задачи сканирования. Подробнее о задачах сканирования см. документ "InfoWatch Data Discovery. Руководство пользователя", Раздел "Хранение файлов" > Вкладка "Задачи" > подразделы "Задача сканирования", "Создание, настройка и запуск задачи".

8.2 Включение/выключение модулей Data Analysis Service

Модуль протокластеризации включен по умолчанию. Модуль транскрибации и модуль детектирования объектов на изображениях можно включить 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`	Модуль протокластеризации
`TRANSCRIBATION`	Модуль транскрибации
`IMG_CONTENT_DETECTION`	Модуль детектирования объектов на изображениях

`<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 TRANSCRIBATION="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:

Чтобы включить или выключить модуль транскрибации или модуль детектирования объектов на изображениях, вы также можете переустановить Data Analysis Service и при установке выбрать необходимую опцию.

Примечание:

Рекомендуется выключать неиспользуемые функции в целях экономии ресурсов.

8.3 Настройка производительности

Существуют следующие способы увеличения производительности Data Analysis Service:

- Создание реплик POD
- Многонодовая установка

8.3.1 Создание реплик POD

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".

Чтобы увеличить производительность Сервиса, вы можете создать дополнительные реплики POD. В этом случае будет увеличена производительность всех включенных модулей.

Если вы хотите увеличить производительность модуля протокластеризации, вы можете сначала оценить ее. Оценка производительности модуля протокластеризации описана ниже.

Для каждого нового POD необходимо выделить ресурсы на сервере.

Пример

Если у вас работает один POD Data Analysis Service, использующий 6 CPU и 16 GB оперативной памяти, то при создании двух реплик, вам необходимо иметь на сервере дополнительно 12 CPU и 32 GB оперативной памяти. Таким образом, всего для трех POD Сервиса потребуется 18 CPU и 48 GB оперативной памяти.

Создать реплики POD можно с помощью команды:

```
kubectl scale deploy --replicas=<количество реплик> -n infowatch das-processor-<node_label>
```

где `<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`.

При необходимости вы можете ограничить использование ресурсов каждым POD Сервиса. Подробнее см. "Настройка использования ресурсов сервера".

8.3.2 Многонодовая 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*", статья "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. "Настройка взаимодействия с продуктами Платформы".

8.3.3 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. Это можно сделать способами, описанными выше.

8.4 Настройка использования ресурсов сервера

В Data Analysis Service вы можете ограничить использование Сервисом ресурсов сервера, например, если на одном сервере с Data Analysis Service установлены другие продукты Платформы. Данные ограничения распространяются на каждый POD Сервиса.

> **Примечание:**
>
> Получить информацию об использовании ресурсов сервера можно с помощью утилиты **htop**.

По умолчанию Data Analysis Service использует все доступные ядра центрального процессора и всю доступную оперативную память. Ограничить использование ресурсов можно с помощью скрипта `resources_setter.sh`. Для этого:

1. Перейдите в директорию `<SETUP_FOLDER>/tools/resource_management`, где `<SETUP_FOLDER>` – это директория, куда был распакован Data Analysis Service.
2. Запустите скрипт с помощью команды:

```
./resources_setter.sh -c <CPU_LIMIT> -m <MEMORY_LIMIT>
```

где `CPU_LIMIT` – количество ядер центрального процессора. Может принимать одно из следующих значений: **0.5**, **1**, **2**, **3** и т.д.
`MEMORY_LIMIT` – объем оперативной памяти. Указывается в байтах в формате `<value>` или в гигабайтах в формате `<value>G`.
Параметры `-c` и `-m` можно указать, как оба сразу, так и один из них.

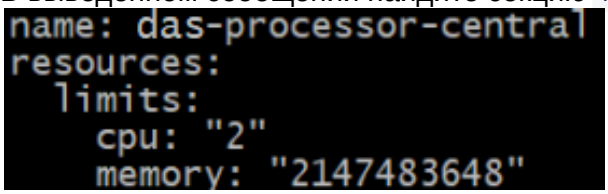
> **Пример команды запуска скрипта:**
>
> Чтобы Data Analysis Service использовал не более 4 ядер центрального процессора и не более 10 гигабайтов оперативной памяти, выполните команду:
>
> ```
> ./resources_setter.sh -c 4 -m 10G
> ```

Чтобы проверить, какие ограничения использования ресурсов установлены в текущий момент:

1. Выполните команду:

```
kubectl get deploy -o yaml -n infowatch das-processor-<node_label>
```

где `<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`.
2. В выведенном сообщении найдите секцию `resources`:



Объем оперативной памяти в сообщении указан в байтах.

8.5 Настройка модуля транскрибации

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.

Вы можете увеличить производительность модуля транскрибации, используя вместо центрального процессора (CPU) графический (GPU):

- Чтобы ускорить обработку аудио-данных, включите использование GPU.
- Чтобы повысить качество транскрибации:
  - Включите использование GPU;

- Выберите большую модель обработки.

Аппаратные и программные требования

Для использования GPU вам необходима видеокарта NVIDIA с поддержкой Compute Unified Device Architecture (CUDA).

На сервере должны быть установлены:

- Драйвер NVIDIA с версией не ниже 384.81
- NVIDIA Container Toolkit с версией не ниже 1.7.0 (для использовании встроенного GPU в системе на базе NVIDIA Tegra – не ниже 1.11.0). 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).

Чтобы включить использование GPU:

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.

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.

5. Укажите, что при вычислениях процесс **das-processor** должен использовать GPU. Для этого выполните команду:

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

Чтобы убедиться, что Data Analysis Service использует GPU:

1. Установите для Сервиса режим логирования DEBUG:

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

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

```
2023-12-15T16:01:46.206764179Z stdout F 2023-12-15 19:01:46,206 [INFO] (das_processor.transcribation.model_manager) Load models
2023-12-15T16:01:46.250850958Z stdout F 2023-12-15 19:01:46,250 [DEBUG] (das_processor) Using CUDA device
```

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 аудио-данных и уничтожаются, если неактивны более часа. При уничтожении в лог-файл записывается сообщение `Drop models`.

Примечание:

За использование GPU в Сервисе отвечает переменная окружения `USE_GPU`. Чтобы переключиться обратно на использование CPU, выполните команду:

```
sudo kubectl -n infowatch set env deploy das-processor-<node_label> USE_GPU=FALSE
```

Чтобы выбрать модель обработки, выполните команду:

- для использования малой модели:

```
sudo kubectl set env deployment -n infowatch das-processor-<node_label>
TRANSCRIBER=whisper_small_transcriber
```

- для использования большой модели:

```
sudo kubectl set env deployment -n infowatch das-processor-<node_label>
TRANSCRIBER=whisper_large_v1_transcriber
```

Примечание:

Большая модель может работать только на GPU.

9 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.

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