InfoWatch Data Explorer. Руководство по установке, конфигурированию и администрированию

26/01/2026

© АО “ИнфоВотч”

Тел./Факс +7 (495) 229-00-22

http://www.infowatch.ru

СОДЕРЖАНИЕ

1 Аудитория 4

2 Комплект документов 5

3 Аппаратные и программные требования Data Explorer 6

3.1 Аппаратные требования 6

3.2 Программные требования 6

4 Подготовка сервера к установке/обновлению Data Explorer 8

5 Установка Data Explorer 9

5.1 Чтобы установить Систему: 9

6 Обновление Data Explorer 11

6.1 Варианты обновления Data Explorer 11

6.2 Обновление Data Explorer с помощью сброса установки (reset) 11

6.2.1 Чтобы обновить Data Explorer с помощью сброса установки: 11

7 Удаление Data Explorer 13

8 Добавление плагина Traffic Monitor 14

9 Настройка интеграции с сервером Traffic Monitor 15

10 Компонент Автолингвист 17

11 Добавление языков в Data Explorer 18

12 Очистка дискового пространства 21

В настоящем руководстве вы сможете найти сведения по установке, обновлению, удалению и настройке компонентов системы InfoWatch Data Explorer.
InfoWatch Data Explorer (далее Система или Data Explorer) – программный продукт, позволяющий автоматизировать формирование политик DLP-системы с помощью выделения тематических кластеров документов в информационных потоках компании.

1 Аудитория

Информация, содержащаяся в Руководстве, предназначена для пользователей, работающих с Системой (выполняющих настройку конфигурации, анализ информационных объектов 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 Explorer входят:

- «InfoWatch Data Explorer. Руководство по установке, конфигурированию и администрированию». Содержит описание установки, обновления, удаления и настройки продукта;
- «InfoWatch Data Explorer. Руководство пользователя». Содержит описание работы с продуктом.

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 Системы (база данных, серверная часть);
- «InfoWatch Traffic Monitor. Руководство пользователя». Содержит описание работы с InfoWatch 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.

3 Аппаратные и программные требования Data Explorer

Для того чтобы Data Explorer мог получать данные, необходима интеграция с InfoWatch Traffic Monitor 7.9 и выше.

3.1 Аппаратные требования

Системные требования для установки:

Процессор	ОЗУ	Жесткий диск
Количество ядер – 4; Частота – 3.1 ГГц; Разрядность – 64 bit; Поддержка инструкций SSE 4.2.	16 Гбайт	от 100 Гбайт

Важно!

Минимальное количество последовательных операций ввода-вывода в секунду для диска 7200 оборотов в минуту равно 50 IOPS. В момент первоначальной загрузки данных в Систему нагрузка возрастает и может потребоваться скорость 500 IOPS. При дальнейшей периодической синхронизации нагрузка будет снижена.

Важно!

Data Explorer должен быть установлен на отдельный выделенный сервер. Совместная установка в кластере с другими продуктами не поддерживается.

Примечание:

Для точного расчета требований под конкретный проект обратитесь к специалистам компании InfoWatch.

3.2 Программные требования

Для установки и работы Data Explorer может быть использовано 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);

Тип ПО	Варианты
	• 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).
Браузер	• Google Chrome; • Яндекс.Браузер.

Примечание:

Data Explorer использует PostgreSQL 12.5. СУБД поставляется вместе с продуктом и не требует отдельной установки.

Важно!

Data Explorer должен быть установлен на отдельный сервер. Не рекомендуется совместная установка с другими продуктами.

4 Подготовка сервера к установке/обновлению Data Explorer

Перед установкой или обновлением Data Explorer обязательно подготовьте сервер:

1. Программа установки написана на языке Python версии 2.7, поэтому для установки продукта в целевой ОС установите интерпретатор этого языка. Пакет с интерпретатором не входит в состав дистрибутива продукта.
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 командой:
   `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. Для отключения выполните команды от имени суперпользователя `root`:
      ```
      setenforce 0
      sed -i 's/SELINUX=enforcing/SELINUX=disabled/g' /etc/selinux/config /etc/selinux/config
      ```
   b. Перезагрузите сервер командой:
      `reboot`
   c. Проверьте статус системы SELinux командой:
      `sestatus`
4. Настройте правила POD сети или отключите межсетевой экран (подробнее см. в базе знаний статьи "Конфликты при взаимодействии службы firewalld и Kubernetes", "Полное отключение межсетевого экрана"). В противном случае установка продукта будет невозможна.
5. Для установки и корректной работы Data Explorer 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 Explorer (см. статью Базы знаний "Настройка антивируса для продуктов Платформы").
7. Data Explorer использует компонент Kubernetes. Для корректной установки продукта и работы Kubernetes на сервере должен быть установлен пакет `conntrack`.
   Убедитесь, что пакет установлен, для этого введите в командной строке:
   `conntrack`
   В ответ будут выведены сведения о работе с данной системой и ее установленная версия.
   Если conntrack не установлен, обязательно установите его с помощью команды:
   - `yum install conntrack` - для ОС Red Hat Enterpise Linux, Oracle Linux и для РЕД ОС;
   - `apt-get install conntrack` - для ОС Astra Linux.

   Если пакет недоступен в резпозитории, скачайте его, перейдите в директорию с пакетом и установите 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.

5 Установка Data Explorer

> **Важно!**
>
> Data Explorer должен быть установлен на отдельный выделенный сервер. Совместная установка в кластере с другими продуктами не поддерживается.

Для установки продукта используется поставляемый архив `iw_de_setup_x.x.x.xx.tar.xz` из состава дистрибутива, где x.x.x.xx – номер версии. В нашем примере будет использована версия `1.5.0.25`, ей соответствует архив `iw_de_setup_1.5.0.25.tar.xz`. Архив содержит полный набор бинарных модулей образов контейнеров, необходимых для развертывания продукта без доступа к сети Интернет (offline-установка). Также архив содержит программу установки.

5.1 Чтобы установить Систему:

1. Выполните шаги по подготовке сервера к установке Data Explorer (см. "Подготовка сервера к установке/обновлению Data Explorer").
2. Создайте новую директорию на диске (например, `de`):
   `mkdir de`
3. Скопируйте архив **`iw_de_setup_ 1.5.0.25 .tar.xz`** в созданную директорию.
4. Перейдите в директорию, в которую был скопирован архив.
5. Распакуйте архив с дистрибутивом продукта в эту директорию:
   `tar -xvf iw_de_setup_ 1.5.0.25 .tar.xz`
6. Запустите программу установки продукта:
   `./setup.py install`
7. Ознакомьтесь с условиями лицензионного соглашения. Лицензионное соглашение содержит несколько страниц. Для перехода на следующую страницу используйте клавишу **Enter**.
8. Введите "**y**", чтобы принять лицензионное соглашение, и нажмите **Enter**.
9. 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**.

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

10. Выделите объем оперативной памяти для размещения данных Clickhouse (по умолчанию: `80%`) и нажмите **Enter**.
11. Укажите путь для размещения данных Clickhouse (по умолчанию: `/mnt/chdata`) и нажмите **Enter**.
12. Укажите путь для размещения данных Tarantool (по умолчанию: `/mnt/trdata`) и нажмите **Enter**.
13. Укажите путь для размещения данных PostgreSQL (по умолчанию: `/mnt/pgdata`) и нажмите **Enter**.
14. Укажите путь для размещения бинарных данных (по умолчанию: `/mnt/dsdata`) и нажмите **Enter**.
15. Укажите порт подключения к веб-интерфейсу (по умолчанию: `443`) и нажмите **Enter**.

16. Дождитесь окончания процесса установки продуктовой платформы.
17. Ознакомьтесь с отчетом об установке платформы (в нашем примере была установлена платформа версии 1.7.0.327). В графе **`web ui`** указаны адрес и порт для подключения к веб-интерфейсу (в нашем примере – `https://10.60.22.24:443`):

```
###Result###
Install product: Infowatch Platform(platform) Version: 1.7.0.327
Install node mode: central
Install node label: central
installed 36 components
updated   0 components
add ref   0 components
web ui:
https://10.60.22.24:443
```

18. Укажите путь для размещения данных RabbitMQ (по умолчанию `/mnt/rmqdata`) и нажмите **Enter**.
19. Укажите путь для размещения данных Redis (по умолчанию `/mnt/redisdata`) и нажмите **Enter**.
20. Укажите путь для размещения данных MongoDB (по умолчанию `/mnt/mongodata`) и нажмите **Enter**.
21. Дождитесь окончания процесса установки продукта Data Explorer.
22. Ознакомьтесь с отчетом об установке продукта.
23. Убедитесь, что все сервисы запущены, выполнив команду:
    `kubectl get pods -n infowatch`
    Команда выведет список сервисов. Каждый сервис должен иметь статус `Running`.
24. Чтобы начать использование Data Explorer, введите в браузере адрес и порт для подключения к веб-интерфейсу (в нашем примере – `https://10.60.22.24:443`).

После установки настройте интеграцию InfoWatch Data Explorer с сервером InfoWatch Traffic Monitor.

6 Обновление Data Explorer

Обновление Data Explorer производится вручную, с помощью командной строки.

Для обновления Системы используется архив `iw_de_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 `1.5.0.25`, ей соответствует архив `iw_de_setup_1.5.0.25.tar.xz`. Архив содержит полный набор бинарных модулей образов контейнеров, необходимых для развертывания Системы без доступа к сети Интернет.

> **Важно!**
>
> Для успешного обновления Data Explorer необходимо не менее 100 Гбайт свободного места на жестком диске.

6.1 Варианты обновления Data Explorer

Data Explorer 1.5 поддерживает только обновление с помощью сброса установки.

> **Важно!**
>
> Переустановка на данные предыдущей версии не поддерживается.

6.2 Обновление Data Explorer с помощью сброса установки (reset)

Обновление состоит из следующих шагов:

1. Сброс установки.
2. Установка новой версии Системы.

6.2.1 Чтобы обновить Data Explorer с помощью сброса установки:

1. Выполните шаги по подготовке сервера к обновлению (см. "Подготовка сервера к установке/обновлению Data Explorer").
2. Создайте новую директорию на диске (например, `DE`):
   `mkdir DE`
3. Скопируйте архив **`iw_de_setup_x.x.x.xx.tar.xz`** в созданную директорию.
   В нашем примере будет использована версия `1.5.0.25`, ей соответствует архив `iw_de_setup_1.5.0.25.tar.xz`.
4. Перейдите в директорию, в которую был скопирован архив.
5. Распакуйте архив с дистрибутивом Data Explorer в эту директорию:
   `tar -xvf iw_de_setup_1.5.0.25.tar.xz`
6. Выполните сброс текущей установки:
   `./setup.py reset`
7. Удалите данные прошлой версии из директорий, указанных ранее при установке.

> **Важно!**
>
> Использование данных прошлой версии не поддерживается.

8. Установите новую версию Data Explorer.
9. Очистите кеш браузера, в котором работали до обновления продукта.

7 Удаление Data Explorer

Data Explorer удаляется с помощью командной строки.

Чтобы удалить Data Explorer:

1. Перейдите в папку, в которой был распакован инсталлятор с последней версией продукта.
2. Выполните команду:
   `./setup.py remove`
3. Дождитесь завершения процесса.

> **Примечание:**
>
> Инсталлятор удалит сервисы и компоненты Системы. После удаления на сервере останутся результаты работы продукта и база данных. Данные останутся в директориях, которые были указаны при установке Data Explorer.

8 Добавление плагина Traffic Monitor

Чтобы Data Explorer мог подключаться к Traffic Monitor и отправлять данные, добавьте плагин в Консоли управления Traffic Monitor.

Плагин и файлы находятся в составе дистрибутива и представляют собой ZIP-архивы. Архив плагина имеет вид: `tmplugin/de.zip .`

Чтобы добавить плагин:

1. Перейдите в раздел **Управление - Плагины**.
2. В области **Плагины** нажмите .
3. Нажмите **Добавить**.
4. Просмотрите описание плагина и нажмите **Установить**.

Информация о плагине будет доступна в разделе **Управление** -> **Плагины**. Для последующей отправки данных в Traffic Monitor необходим токен, который создается автоматически и доступен в разделе **Управление** -> **Плагины** -> **Токены**. Скопируйте выданный вам токен и укажите его в поле **Токен** в Консоли управления Data Explorer при последующей настройке интеграции с InfoWatch Traffic Monitor.

9 Настройка интеграции с сервером Traffic Monitor

Data Explorer использует данные, извлекаемые из событий Traffic Monitor.
Для доступа компонента Data Explorer к событиям обязательно настройте интеграцию с сервером Traffic Monitor.

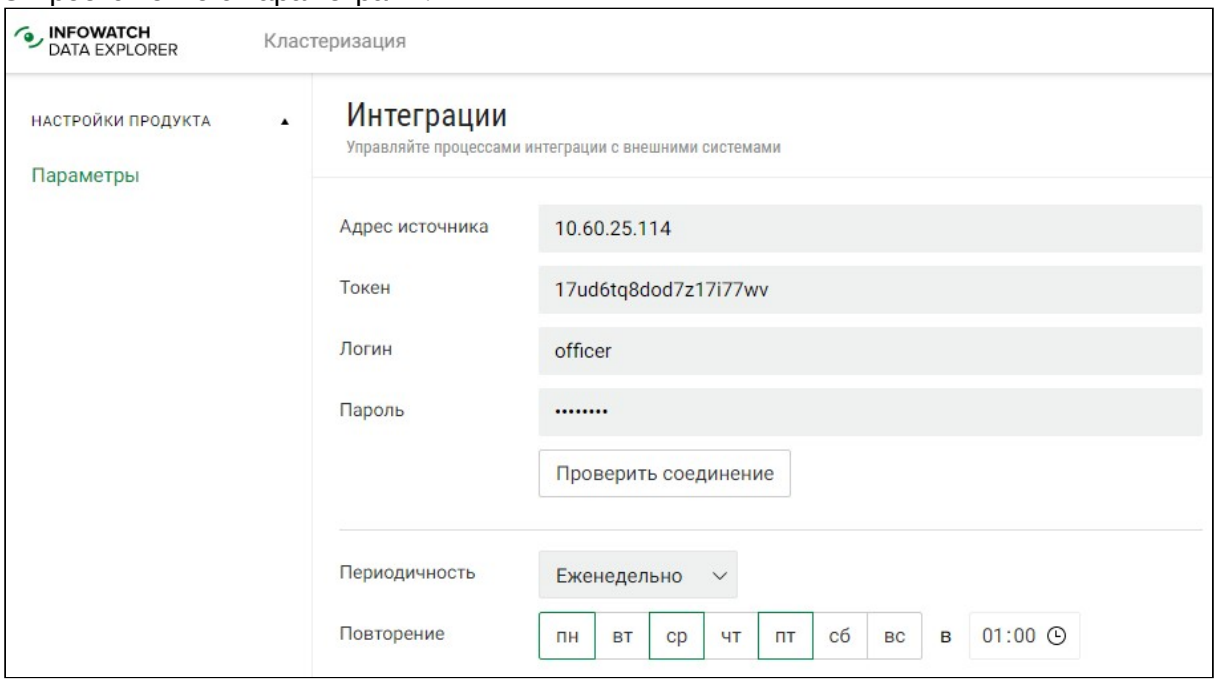
> **Примечание:**
>
> Одновременно поддерживается интеграция только с одним сервером Traffic Monitor.

Настройка интеграции выполняется в Консоли управления Data Explorer. Для настройки:

1. Откройте Консоль управления Data Explorer и авторизуйтесь.

   > **Примечание:**
   >
   > Если пароль пользователя Консоли не менялся, используйте для авторизации стандартные учетные данные:
   >
   > - логин – `officer`
   > - пароль – `xxXX1234`

2. Для перехода в раздел настроек в правом верхнем углу нажмите кнопку .
   Откроется окно с параметрами:



3. Заполните все указанные поля:
   - **Адрес источника** – IP-адрес сервера Traffic Monitor;
     В случае распределенной установки Traffic Monitor укажите адрес сервера, на котором установлены компоненты консоли управления Traffic Monitor (шаблон установки Веб-консоль).

- **Токен** – токен для подключения к серверу Traffic Monitor по программному интерфейсу приложения (API), полученный после добавления плагина;
- **Логин** – логин пользователя, от имени которого осуществляется соединение;
- **Пароль** – пароль пользователя, от имени которого осуществляется соединение.

4. После заполнения всех полей вы можете проверить корректность введенных данных и доступность сервера Traffic Monitor с помощью кнопки **Проверить соединение**.
5. Укажите периодичность, с которой InfoWatch Data Explorer будет выгружать данные из Traffic Monitor для последующей кластеризации.
   Вы можете выбрать один из вариантов:
   - **Ежечасно** – укажите интервал в часах, чтобы выгружать данные через каждые N часов;
   - **Ежедневно** – укажите время суток;
   - **Еженедельно** – укажите дни недели и время суток.

При настройках, указанных на изображении выше, Data Explorer будет подключаться к серверу Traffic Monitor по адресу *10.60.25.114* от имени пользователя *officer* и выгружать данные в *01:00* по *понедельникам*, *средам* и *пятницам*.

10 Компонент Автолингвист

Компонент Автолингвист реализован в виде технологии анализа в системе Traffic Monitor.

Установка, обновление и удаление Автолингвиста выполняются только в составе Traffic Monitor. Данные процессы описаны в документе «*InfoWatch Traffic Monitor. Руководство по установке*».

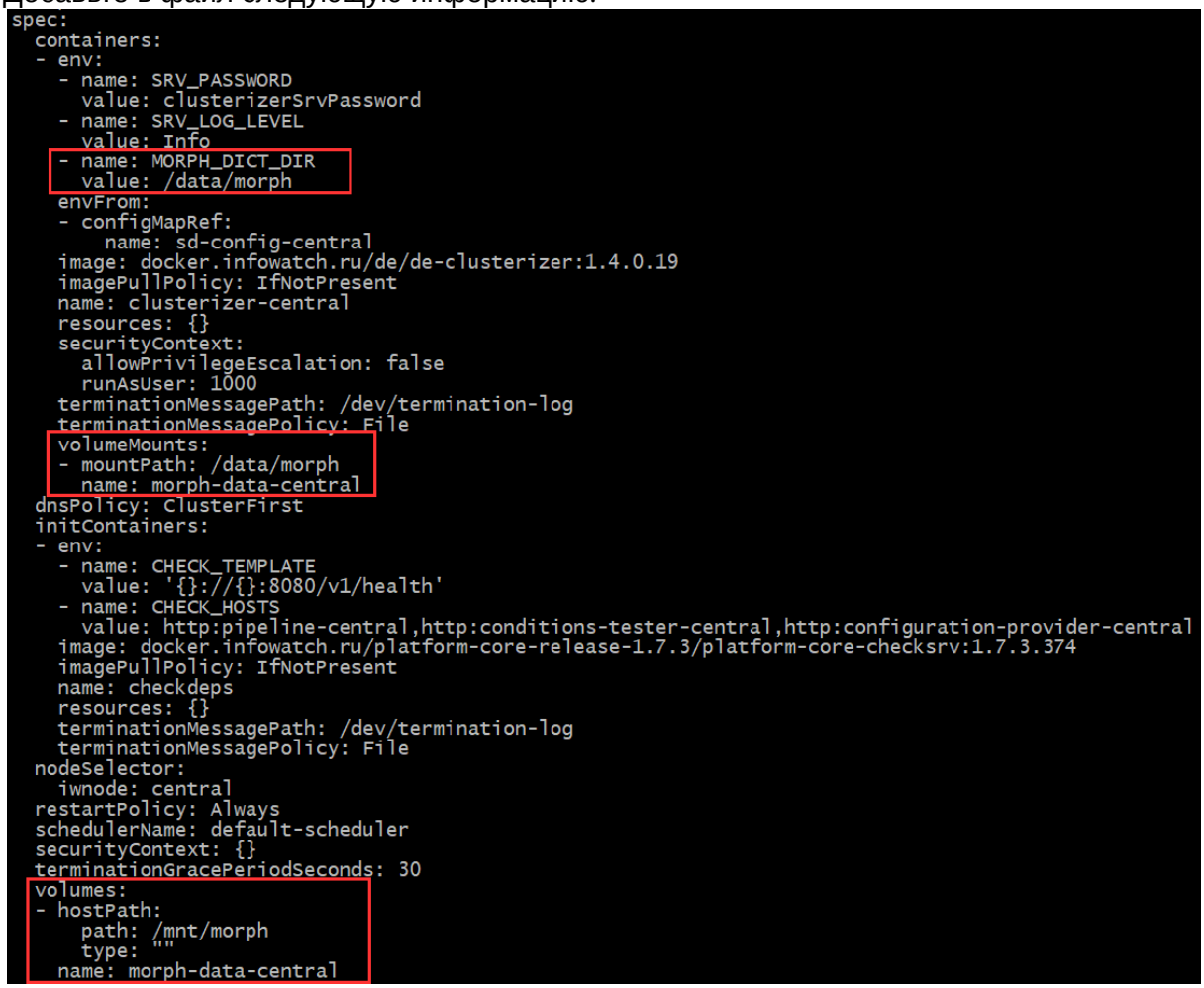
11 Добавление языков в Data Explorer

Data Explorer может осуществлять кластеризацию текстовых документов на языках, поддерживаемых InfoWatch Traffic Monitor. По умолчанию Data Explorer работает с русским и английским языками. Словари других языков вы можете получить у специалистов компании InfoWatch.

Чтобы добавить в Data Explorer новый язык для кластеризации:

1. Если у вас уже установлен Data Explorer, то удалите его с помощью командной строки:
   a. Перейдите в папку, в которую был распакован инсталлятор с последней установленной версией продукта.
   b. Выполните команду:
   `./setup.py reset`
   c. Дождитесь завершения процесса.
2. Удалите данные, оставшиеся от прошлой версии. Для этого выполните команду:
   `rm -rf /mnt/*`
3. Установите новую версию Data Explorer. Подробнее см. "Установка Data Explorer".
4. Создайте директорию, где будет храниться словарь нового языка:
   `mkdir /mnt/morph`
5. Скопируйте словарь в созданную директорию. Словарь представляет собой JSON-файл. В нашем примере это будет словарь казахского языка
   `sklearn_tokenizer_morph_vocab_kz.json` .
6. Откройте для редактирования конфигурационный файл `clusterizer-central.yaml` , выполнив команду:
   `kubectl edit deployment clusterizer-central -n infowatch`

7. Добавьте в файл следующую информацию:



```
spec:
  containers:
  - env:
    - name: SRV_PASSWORD
      value: clusterizerSrvPassword
    - name: SRV_LOG_LEVEL
      value: Info
    - name: MORPH_DICT_DIR
      value: /data/morph
    envFrom:
    - configMapRef:
        name: sd-config-central
    image: docker.infowatch.ru/de/de-clusterizer:1.4.0.19
    imagePullPolicy: IfNotPresent
    name: clusterizer-central
    resources: {}
    securityContext:
      allowPrivilegeEscalation: false
      runAsUser: 1000
    terminationMessagePath: /dev/termination-log
    terminationMessagePolicy: File
    volumeMounts:
    - mountPath: /data/morph
      name: morph-data-central
  dnsPolicy: ClusterFirst
  initContainers:
  - env:
    - name: CHECK_TEMPLATE
      value: '{}://{}:8080/v1/health'
    - name: CHECK_HOSTS
      value: http:pipeline-central,http:conditions-tester-central,http:configuration-provider-central
    image: docker.infowatch.ru/platform-core-release-1.7.3/platform-core-checksrv:1.7.3.374
    imagePullPolicy: IfNotPresent
    name: checkdeps
    resources: {}
    terminationMessagePath: /dev/termination-log
    terminationMessagePolicy: File
  nodeSelector:
    iwnode: central
  restartPolicy: Always
  schedulerName: default-scheduler
  securityContext: {}
  terminationGracePeriodSeconds: 30
  volumes:
  - hostPath:
      path: /mnt/morph
      type: ""
    name: morph-data-central
```

- описание переменной окружения `MORPH_DICT_DIR` в секцию "`spec`" -> "`containers`" -> "`env`":

```
        - name: MORPH_DICT_DIR
          value: /data/morph
```

- описание точки монтирования словаря в секцию "`spec`" -> "`containers`":

```
        volumeMounts:
        - mountPath: /data/morph
          name: morph-data-central
```

- описание тома монтирования в секцию "`spec`":

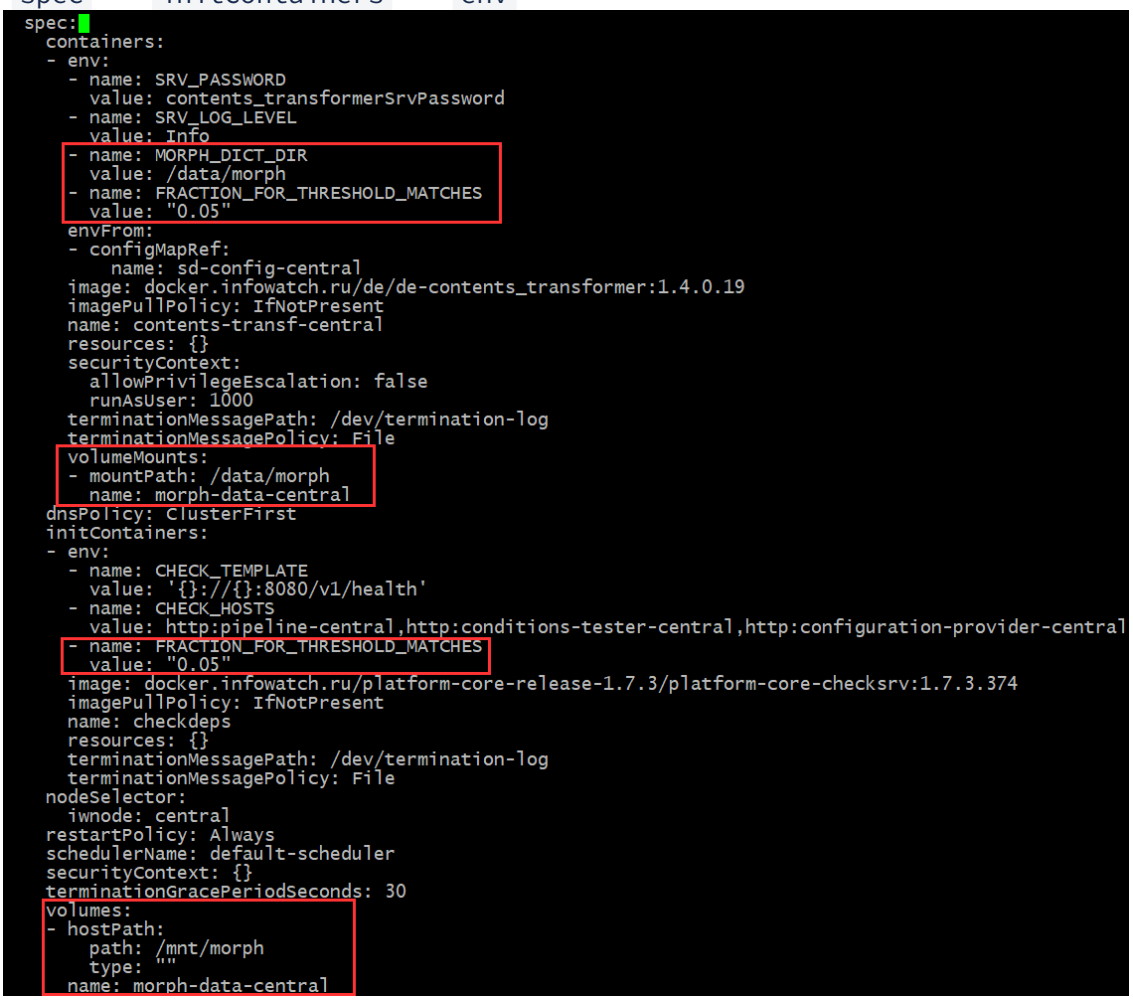
```
      volumes:
      - hostPath:
          path: /mnt/morph
          type: ""
        name: morph-data-central
```

8. Откройте для редактирования конфигурационный файл `contents-transf-central.yaml` , выполнив команду:
   `kubectl edit deployment contents-transf-central -n infowatch`
9. Добавьте в файл информацию, указанную в п.7.
10. Дополнительно добавьте в файл описание переменной окружения `FRACTION_FOR_THRESHOLD_MATCHES`

```
        - name: FRACTION_FOR_THRESHOLD_MATCHES
          value: "0.05"
```

в секции

- " `spec` " -> " `containers` " -> " `env` "
- " `spec` " -> " `initContainers` " -> " `env` "



Если новый язык успешно добавлен, то в логах процессов `clusterizer-central` и `contents-transf-central` появятся сообщения вида:

```
{"log":"2024-04-03 18:08:29,954 [INFO] (clusterizer) Directory '/data/morph' contains 1 morph dicts\n","str
eam":"stdout","time":"2024-04-03T15:08:29.95513779Z"}
{"log":"2024-04-03 18:08:30,262 [INFO] (clusterizer) Uploaded 1/1 files:
sklearn_tokenizer_morph_vocab_kz.json\n","stream":"stdout","time":"2024-04-03T15:08:30.262468034Z"}
```

12 Очистка дискового пространства

Чтобы процесс кластеризации проходил быстрее, Data Explorer каждую ночь загружает контенты из Traffic Monitor в свою базу данных. Из-за этого постепенно уменьшается свободное пространство на жестком диске. В связи с этим в Data Explorer реализован следующий механизм автоматической очистки дискового пространства:

1. Каждую ночь Система проверяет, сколько свободного места осталось на диске.
2. Если процент занятого места на диске больше верхнего порогового значения, то Система начинает удалять из БД старые обработанные контенты.
   Чтобы задать порог начала удаления, выполните команду:

   ```
   kubectl set env deployment task-scheduler-central
   SR_MAX_THRESHOLD_PERCENT=<верхний_порог_в_процентах> -n infowatch
   ```

   > **Примечание:**
   >
   > Контенты, которые участвуют хотя бы в одной задаче кластеризации, не будут удалены.

3. Если на диске появилось свободное пространство и процент занятого места стал равным нижнему пороговому значению, то Система перестает удалять контенты. Этот процесс также прекратится, если доступные для удаления контенты закончились.
   Чтобы задать порог окончания удаления, выполните команду:

   ```
   kubectl set env deployment task-scheduler-central
   SR_MIN_THRESHOLD_PERCENT=<нижний_порог_в_процентах> -n infowatch
   ```

По умолчанию верхнее и нижнее пороговые значения равны:

- `SR_MAX_THRESHOLD_PERCENT=80` - процент занятого места на жестком диске, при достижении которого начинается процесс удаления контентов;
- `SR_MIN_THRESHOLD_PERCENT=60` - процент занятого места на жестком диске, при достижении которого процесс останавливается.