

INFOWATCH

InfoWatch HEML Adapter. Руководство по установке, конфигурированию и запуску

12/11/2019

© АО “ИнфоВотч”

Тел./Факс +7 (495) 229-00-22

http://www.infowatch.ru

СОДЕРЖАНИЕ

1 Функции InfoWatch HEML Adapter 3

2 Аппаратное и программное обеспечение для InfoWatch HEML Adapter . 4

2.1 Программное обеспечение: 4

2.2 Аппаратное обеспечение: 4

3 Установка InfoWatch HEML Adapter 5

3.1 Порядок установки InfoWatch HEML Adapter 5

3.2 Установка плагина для InfoWatch Traffic Monitor 5

3.3 Конфигурирование InfoWatch HEML Adapter 6

4 Настройка почтового сервера 11

5 Запуск InfoWatch HEML Adapter 12

1 Функции InfoWatch HEML Adapter

InfoWatch HEML Adapter представляет собой дополнительный модуль InfoWatch Traffic Monitor, позволяющий:

- выгружать входящие и исходящие письма из pst-файлов или с почтового сервера MS Exchange за конкретный период времени;
- передавать почтовые сообщения на анализ в Traffic Monitor как события перехвата почты (протокол SMTP);
- отфильтровать автоматически сгенерированные письма (письма от роботов) и SPAM-сообщения;
- параллельно извлекать и обрабатывать письма из нескольких PST-архивов либо из почтовых ящиков различных пользователей или групп пользователей;

Модуль реализован в виде скрипта и не имеет графического интерфейса. Управление InfoWatch HEML Adapter осуществляется с помощью команд в **bash.**

2 Аппаратное и программное обеспечение для InfoWatch HEML Adapter

2.1 Программное обеспечение:

- окружение docker;
- Python версии 3.6 и более поздние;
- консольная утилита Readpst.

2.2 Аппаратное обеспечение:

- объем свободного места на диске при экспорте писем: не менее 5 ГБ на архив;
- объем оперативной памяти при фильтрации писем: 4 ГБ.

3 Установка InfoWatch HEML Adapter

3.1 Порядок установки InfoWatch HEML Adapter

Перед установкой Модуля необходимо подготовить окружение:

1. Установить docker;
2. Настроить docker на работу с proxy;
3. Добавить нового пользователя в группу docker;
4. Войти под учетной записью нового пользователя в локальный docker registry через браузер или консоль docker login.

Чтобы установить и запустить Модуль:

1. Создайте директорию для установки Модуля.
2. Скопируйте архив с образом **heml_adapter.1.0.0.x.tar.xz** в эту директорию.
3. Разархивируйте архив:
   `tar -xf heml_adapter.1.0.0.tar.xz`
4. Запустите загрузку образа:
   `docker load -i heml_adapter.1.0.0`
5. Дождитесь окончания загрузки образа:

```
90d1009ce6fe: Loading layer [==================================================>]  105.5MB/105.5MB
c23711a84ad4: Loading layer [==================================================>]  23.99MB/23.99MB
8f7ee6d76fd9: Loading layer [==================================================>]  8.005MB/8.005MB
f75e64f96dbc: Loading layer [==================================================>]  146.4MB/146.4MB
e02b32b1ff99: Loading layer [==================================================>]  570.6MB/570.6MB
34929ec591c4: Loading layer [==================================================>]  17.35MB/17.35MB
5129f19da2c9: Loading layer [==================================================>]   66.1MB/66.1MB
f89067d6e30e: Loading layer [==================================================>]  4.608kB/4.608kB
27951393f8e7: Loading layer [==================================================>]  6.353MB/6.353MB
fa04c3baf4ca: Loading layer [==================================================>]   2.56kB/2.56kB
84fbc275b6f9: Loading layer [==================================================>]  3.072kB/3.072kB
aabe4fee45d6: Loading layer [==================================================>]  288.2MB/288.2MB
f6e1dbead27b: Loading layer [==================================================>]  3.072kB/3.072kB
e2d14398c3f5: Loading layer [==================================================>]  2.934MB/2.934MB
51497a0fb86f: Loading layer [==================================================>]  3.584kB/3.584kB
74dd013a7882: Loading layer [==================================================>]   2.56kB/2.56kB
bac503b6c251: Loading layer [==================================================>]   2.56kB/2.56kB
63a8c77bc3d6: Loading layer [==================================================>]  26.91MB/26.91MB
43ee602e7490: Loading layer [==================================================>]   98.3kB/98.3kB
f474a75c5684: Loading layer [==================================================>]  3.072kB/3.072kB
Loaded image: heml_adapter:latest
```

   Убедитесь, что образ был успешно загружен:
   `docker images`

```
heml_adapter        latest              6e538af90c32        8 min ago           1.23GB
```

3.2 Установка плагина для InfoWatch Traffic Monitor

Чтобы установить плагин для InfoWatch Traffic Monitor:

1. В разделе **Плагины** нажмите  **Добавить плагин.**
2. В появившемся окне нажмите **Загрузить.**
3. Выберите файл плагина (`.zip`). Если напротив плагина установлен статус *Не лицензирован*, то:
   a. убедитесь, что в файле плагина (`.zip`) присутствуют **manifest.json** и файл лицензии (`.license`).

b. откройте **manifest.json** на редактирование и пропишите в параметре `LICENSE` путь до нужного файла лицензии. Путь укажите относительно файла **manifest.json**. Если файл лицензии и **manifest.json** находятся в одной директории,то:

```
"LICENSE": [
       { "PATH" : "tm_license.license" }
   ]
```

c. сохраните изменения и повторите действия пунктов 1-3.

4. Убедитесь, что напротив плагина установлен статус *Лицензирован*. Установлена корректная лицензия перехватываемых типов событий.
5. Для завершения установки нажмите **Установить**.

3.3 Конфигурирование InfoWatch HEML Adapter

Перед запуском Модуля необходимо отредактировать его конфигурационный файл и заполнить обязательные параметры корректными данными. Для редактирования файла откройте его в любом редакторе, внесите и сохраните изменения. При запуске Модуль проверяет правильность значений и, в случае их некорректности, выдает ошибку.

Описание конфигурационного файла Модуля (`.json`) приведено ниже.

Код	Обязательный параметр	Описание
`{`		
`"python-env": "/usr/bin/python",`	Да	Путь до python-интерпретатора
`"collection-type": "pst",`	Да	Тип выгрузки почты (`'pst'` - из pst-файлов, `'imap'` - из почтового сервера по IMAP)
`"collection-step": true,`	Да	Активация стадии сбора почты
`"filter-step": true,`	Да	Активация стадии фильтрации почты: спам-фильтр и фильтр автосгенерированных сообщений
`"sending-step": true,`	Да	Активация стадии посылки почты на сервер TM по протоколу SMTP
`"load-name": "my_history_1",`	Да	Пользовательский атрибут *Название выгрузки*

`"load-date": "2018-10-10T15:12:00 +03:00",`		Пользовательский атрибут *Дата выгрузки*. По умолчанию: текущая дата
`"load-host": "my_host",`		Пользовательский атрибут *Имя хоста*, с которого производилась выгрузка. По умолчанию: ip-адрес компьютера, где запускался скрипт
`"tmp-dir": null,`		Указание временной директории, которая автоматически удалится по завершении всех работ
`"verbose": true,`		Вывод дополнительной информации
`"log-dir": "logs",`	Да	Директория для сохранения логов
`"log-level": "INFO",`		Уровень логирования. По умолчанию: `'INFO'`
`"IMAP_COLLECTOR": {`		Параметры сбора писем по IMAP
`"out-dir": "tmp/imap_results/",`	Да	Директория для сохранения результата
`"host": "xxx.xxx.xxx.xxx",`	Да	Имя хоста
`"port": 143,`		Номер порта. По умолчанию: `143`
`"ssl": false,`		Выполнять подключение к почтовому серверу по SSL. По умолчанию: `'false'`
`"user": "user",`	Да	Имя пользователя почтового сервера, имеющего права на выгрузку почты
`"aliases": "resourses/aliases.csv",`	Да	Файл с алиасами ,выгруженный из почтового сервера
`"start-date": "2000-01-01",`	Да	Начальная дата выгрузки (анализируется поле Date письма)
`"end-date": "2018-01-01",`	Да	Конечная дата выгрузки (анализируется поле Date письма)

"jobs": -1,		Количество используемых потоков ('-1' - число потоков = числу потоков доступных системе, по умолчанию: '1')
"verbose": true,		Вывод дополнительной информации
"log-dir": "logs/imap_collector",	Да	Директория для сохранения логов
"log-level": "INFO",		Уровень логирования. По умолчанию: 'INFO'
},		
"PST_COLLECTOR": {		Параметры сбора писем из PST-файлов
"proc-dir": "resourses/pst_my/",	Да	Директория с PST-файлами
"out-dir": "tmp/pst_results",	Да	Директория для сохранения результата
"export-dir": "tmp/pst_export",		Директория для экспорта содержимого pst-файлов
"aliases": "resourses/aliases.csv",	Да	Файл с алиасами, выгруженный из почтового сервера
"start-date": "2000-01-01",	Да	Начальная дата выгрузки (анализируется поле Date письма)
"end-date": "2018-01-01",	Да	Конечная дата выгрузки (анализируется поле Date письма)
"jobs": -1,		Количество используемых потоков ('-1' - число потоков = числу потоков доступных системе, по умолчанию: '1')
"verbose": true,		Вывод дополнительнойц информации
"log-dir": "logs/pst_collector",	Да	Директория для сохранения логов
"log-level": "INFO"		Уровень логирования. По умолчанию: 'INFO'

},		
"EML_FILTER" : {		Параметры фильтра почтовых сообщений
"proc-dir": "tmp/pst_results",	Да	Директория с письмами (в формате .eml) для фильтрации
"filter-dir": "tmp/filter",	Да	Директория с письмами, отсеенными фильтрами
"spam-stop-list": "resourses/spam_stop_list.txt",		Список стоп-слов для спам-фильтра. По умолчанию: `['spam']`
"step-list": ["spam", "autogen"],		Стадии фильтрации: • `spam` - фильтрация спам-сообщений, • `autogen` - фильтрация автоматически сгенерированных сообщений) По умолчанию: `["spam", "autogen"]`
"verbose": true,		Вывод доп. информации
"log-dir": "logs/filter",	Да	Директория для сохранения логов
"log-level": "INFO"		Уровень логирования. По умолчанию: `'INFO'`
},		
"EML_SENDER": {		Параметры отправки почтовых сообщений на сервер TM (eml_sender)
"proc-dir": "tmp/pst_results",	Да	Директория с письмами для отправки
"host": "host@infowatch.ru",	Да	Имя хост-сервера TM

`"user": "user",`		Имя пользователя БД на сервере, способного создать табличные пространства **Важно!** Пароль учетной записи будет запрошен в консоли при запуске скрипта
`"port": 25,`		Номер порта. По умолчанию: `25`
`"db-type": "postgres",`	Да	Тип БД на сервере. Возможные значения: '`postgres`', '`oracle`'
`"clean": false,`		Очищать директорию 'proc-dir' после отправки. По умолчанию: '`false`'
`"err-eml-dir": "tmp/sender_err_eml",`	Да	Директория для сохранения писем, которые не удалось отправить
`"excl-eml": null,`		Путь до файла с письмами, которые необходимо исключить из процесса отправки
`"start_date": null,`		Начальная дата отправки в формате '`YYYY-mm-dd`'
`"min-parallel": null,`		Минимальное количество писем для начала параллельной обработки. По умолчанию: `10.000`
`"jobs": -1,`		Количество используемых потоков ('-1' - число потоков = числу потоков доступных системе, по умолчанию: '1')
`"verbose": true,`		Вывод дополнительной информации
`"log-dir": "logs/eml_sender",`	Да	Директория для сохранения логов
`"log-level": "INFO"`		Уровень логирования. По умолчанию: '`INFO`'
`}`		
`}`		

4 Настройка почтового сервера

Чтобы выгружать архивы с почтовыми сообщениями с сервера MS Exchange необходимо получить доступ к целевым учетным записям пользователей. Для этого:

1. Зайдите в Консоль управления MS Exchange
2. В левой части выберите раздел **Recipient Configuration**
3. В раскрывшемся списке на центральной области выберите нужного пользователя, щелкните правой кнопкой мыши и выберите пункт меню Manage **Full Access Permission...**
4. В открывшемся окне нажмите **+Add** и укажите пользователя(ей), которые получат полные права к данной учетной записи.
5. Нажмите **Ок**, затем **Manage**, затем **Finish**.

5 Запуск InfoWatch HEML Adapter

Пуск Модуля осуществляется запуском python-скрипта, который организует автоматический запуск внутренних скриптов:

- скрипт сбора данных из почтовых PST-архивов;
- скрипт сбора сообщений с почтового сервера;
- скрипты для фильтрации сообщений;
- скрипт для отправки почтовых сообщений на сервер Traffic Monitor.

Настройки работы Модуля задаются в конфигурационном файле (см. "Конфигурирование InfoWatch HEML Adapter"). Заполните значения обязательных параметров. С помощью конфигурационного файла можно включать или отключать стадии работы скрипта (сбор данных, фильтрация, отправка в TM), исходя из поставленной задачи.

> ⚠ **Примечание:**
>
> Особенности:
>
> 1. При указании путей до файлов следует указывать пути относительно директории `/mnt/tmp/`.
> 2. Учетная запись пользователя почтового сервера должна предоставлять полный доступ к почтовым ящикам интересующих пользователей (логин и пароль).
> 3. При сборе данных не должно быть заархивированных табличных пространств для дат из диапазона, который был указан при конфигурировании. Если такие пространства есть, то их необходимо восстановить перед запуском скрипта.

Чтобы запустить Модуль:

1. Создайте директорию (можно в ранее созданной директории для установки Модуля), которая будет монтироваться в докер:
   `mkdir tmp`
2. Запустите контейнер. При запуске монтируется директория `temp_dir`, в которую будут сохраняться результаты работы Модуля. В нее можно скопировать файлы, используемые Модулем (aliases, pst).
   `docker run -it --network=host -v /home/tmp:/mnt/tmp/ heml_adapter -c /mnt/tmp/<config_path>`

После запуска Результаты работы скрипта можно посмотреть во вложенных папках: `/pst, /pst-export, /pst-results.` Лог-файлы находятся в папке `/logs.`

Полученные от Модуля письма становятся доступны пользователю Traffic Monitor как события почты. Такие события имеют несколько дополнительных специфичных атрибутов:

- Название выгрузки архива;
- Дата извлечения архива;
- Имя рабочей станции.