INFOWATCH

InfoWatch HEML Adapter. A Guide for Installation, Configuration, and Running the Program.

31/01/2020

© JSC “InfoWatch”

Tel/Fax +7 (495) 229-00-22

http://www.infowatch.com

Table of Contents

1 Function of InfoWatch HEML Adapter 3

2 Hardware and Software Requirements for InfoWatch HEML Adapter 4

2.1 Software requirements: 4

2.2 Hardware requirements: 4

3 Installing InfoWatch HEML Adapter 5

3.1 InfoWatch HEML Adapter Installation Order 5

3.2 Installing Plugin for InfoWatch Traffic Monitor 5

3.3 Configuring InfoWatch HEML Adapter 6

4 Configuring Mail Server 11

5 Running InfoWatch HEML Adapter 12

1 Function of InfoWatch HEML Adapter

InfoWatch HEML Adapter is an additional module of InfoWatch Traffic Monitor. It allows you to:

- unload incoming and outgoing emails from .pst files or from MS Exchange mail server for a specific period of time
- send mail messages to Traffic Monitor for analysis as mail interception events (via SMTP)
- filter automatically generated emails (emails from robots), spam emails
- simultaneously extract and process emails from several PST archives or from mailboxes of different users/user groups

The module is implemented as a script and does not have a GUI. You can manage InfoWatch HEML Adapter with **bash** commands.

2 Hardware and Software Requirements for InfoWatch HEML Adapter

2.1 Software requirements:

- Docker environment
- Python version 3.6 or later
- Readpst console utility

2.2 Hardware requirements:

- Disk space when exporting emails: not less than 5 GB per archive
- RAM when filtering emails: 4 GB

3 Installing InfoWatch HEML Adapter

3.1 InfoWatch HEML Adapter Installation Order

Before installing the Module, please prepare the environment:

1. Install Docker.
2. Congigure Docker to work with proxy.
3. Add a new user to the Docker group.
4. Log into a local Docker registry under the new user account via browser or Docker login console.

To install and run the Module:

1. Create a new directory to install the Module.
2. Copy the **heml_adapter.1.0.0.x.tar.xz** image archive to this directory.
3. Unpack the archive:
   `tar -xf heml_adapter.1.0.0.tar.xz`
4. Start loading the image:
   `docker load -i heml_adapter.1.0.0`
5. Wait until the image loading is complete:

```
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
```

Make sure that the image is successfully loaded:
`docker images`

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

3.2 Installing Plugin for InfoWatch Traffic Monitor

To install a plugin for InfoWatch Traffic Monitor:

1. In the **Plugins** section, click + **Add plugin**.
2. In the window that opens, click **Load**.
3. Choose the plugin file (`.zip`). If you see the *Not licensed* status next to the plugin, then:
   a. Make sure that the plugin file (`.zip`) contains **manifest.json** and the license file (`.license`).
   b. Open **manifest.json** for editing and, in the `LICENSE` parameter, specify the path to the required license file. Specify the path relative to the **manifest.json** file. If the license file and the **manifest.json** file are located in the same directory, then:

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

c. Save changes and repeat steps 1-3.

4. Make sure that the *Licensed* status is set next to the plugin. It means that you have a correct license for intercepted event types.
5. To complete the installation, click **Install**.

3.3 Configuring InfoWatch HEML Adapter

Before running the Module, you should edit its configuration file and fill in required parameters with correct data. To edit the file, open it in any editor, make changes and then save them. When you run the Module, it checks if the values are correct and, if not, gives an error.

Below you can find a description of the Module configuration file (`.json`).

Code	Is required	Description
`{`		
`"python-env": "/usr/bin/python",`	Yes	The path to the Python interpreter
`"collection-type": "pst",`	Yes	The mail unloading type ('`pst`' - from PST files, '`imap`' - from a mail server by IMAP)
`"collection-step": true,`	Yes	Activating the stage of collecting emails
`"filter-step": true,`	Yes	Activating the stage of filtering emails: a spam filter and a filter for auto generated messages
`"sending-step": true,`	Yes	Activating the stage of sending emails to TM server via SMTP
`"load-name": "my_history_1",`	Yes	The *Unloading name* custom attribute
`"load-date": "2018-10-10T15:12:00+03:00",`		The *Unloading date* custom attribute. By default, the current date
`"load-host": "my_host",`		The *Host name* (where unloading took place) custom attribute. By default, an IP address of the computer where the script was running

`"tmp-dir": null,`		Specifying a temporary directory that will be automatically removed when all work is complete
`"verbose": true,`		Displaying additional information
`"log-dir": "logs",`	Yes	A directory for saving logs
`"log-level": "INFO",`		The logging level. The default value is `'INFO'`
`"IMAP_COLLECTOR": {`		Parameters of collecting emails via IMAP
`"out-dir": "tmp/imap_results/",`	Yes	A directory for saving results
`"host": "xxx.xxx.xxx.xxx",`	Yes	The hostname
`"port": 143,`		The port name. By default, `143`
`"ssl": false,`		Whether to connect to a mail server via SSL. By default, `'false'`
`"user": "user",`	Yes	The name of a mail server user who has the right to unload emails
`"aliases": "resourses/aliases.csv",`	Yes	A file with aliases unloaded from the mail server
`"start-date": "2000-01-01",`	Yes	Unloading start date (the Date field of the email is analyzed)
`"end-date": "2018-01-01",`	Yes	Unloading end date (the Date field of the email is analyzed)
`"jobs": -1,`		The number of used streams ('-1' - the number of streams = the number of streams available to the system; the default value is '1')
`"verbose": true,`		Displaying additional information
`"log-dir": "logs/imap_collector",`	Yes	A directory for saving logs
`"log-level": "INFO",`		The logging level. The default value is `'INFO'`

},		
"PST_COLLECTOR": {		Parameters of collecting emails from PST files
"proc-dir": "resourses/pst_my/",	Yes	A directory with PST files
"out-dir": "tmp/pst_results",	Yes	A directory for saving the result
"export-dir": "tmp/pst_export",		A directory for exporting contents of PST files
"aliases": "resourses/aliases.csv",	Yes	A file with aliases unloaded from the mail server
"start-date": "2000-01-01",	Yes	Unloading start date (the Date field of the email is analyzed)
"end-date": "2018-01-01",	Yes	Unloading end date (the Date field of the email is analyzed)
"jobs": -1,		The number of used streams ('-1' - the number of streams = the number of streams available to the system; the default value is '1')
"verbose": true,		Displaying additional information
"log-dir": "logs/pst_collector",	Yes	A directory for saving logs
"log-level": "INFO"		The logging level. The default value is `'INFO'`
},		
"EML_FILTER" : {		Parameters of filtering emails
"proc-dir": "tmp/pst_results",	Yes	A directory with emails (в формате .eml) to be filtered
"filter-dir": "tmp/filter",	Yes	A directory with filtered emails

"spam-stop-list": "resourses/spam_stop_list.txt",		A list of stop words for spam filter. By default: ['spam']
"step-list": ["spam", "autogen"],		Filtering stages: • spam - filtering spam messaging • autogen - filtering automatically generated messages By default: ["spam", "autogen"]
"verbose": true,		Displaying additional information
"log-dir": "logs/filter",	Yes	A directory for saving logs
"log-level": "INFO"		The logging level. The default value is 'INFO'
},		
"EML_SENDER": {		Parameters of sending emails to TM server (eml_sender)
"proc-dir": "tmp/pst_results",	Yes	A directory with emails to be sent
"host": "host@infowatch.ru",	Yes	The name of TM hosting server
"user": "user",		The name of a DB user on the server who can create tablespaces. **Important!** When running the script, you will be asked to type a password in the console.
"port": 25,		The port number. The default value is 25
"db-type": "postgres",	Yes	A DB type on the server. Possible values: 'postgres', 'oracle'
"clean": false,		Whether to clear the 'proc-dir' directory after sending. The default value is 'false'
"err-eml-dir": "tmp/sender_err_eml",	Yes	A directory for saving emails that cannot be sent
"excl-eml": null,		The path to the file with emails that should be excluded from sending
"start_date": null,		The initial sending date in 'YYYY-mm-dd' format

"min-parallel": null,		The minimum number of emails to start parallel processing. The default value is 10.000
"jobs": -1,		The number of used streams ('-1' - the number of streams = the number of streams available to the system; the default value is '1')
"verbose": true,		Displaying additional information
"log-dir": "logs/eml_sender",	Yes	A directory for saving logs
"log-level": "INFO"		The logging level. The default value is 'INFO'
}		
}		

4 Configuring Mail Server

To unload archives with emails from MS Exchange server, you need to get access to required user accounts. For this:

1. Open MS Exchange server console.
2. In the left area, choose **Recipient Configuration**.
3. From the drop-down list in the central area, right-click on the required user and choose the **Full Access Permission...** item of the **Manage** menu.
4. In the window that opens, click **+Add** and specify user(s) who will get full rights to this account.
5. Click **OK**, then click **Manage**, and then **Finish**.

5 Running InfoWatch HEML Adapter

To run the Module, you need to run a Python script which automatically runs the following internal scripts:

- a script for collecting data from mail PST archives
- a script for collecting messages from a mail server
- scripts for filtering messages
- a script for sending emails to Traffic Monitor.server

The Module operation settings are specified in a configuration file (see "Configuring InfoWatch HEML Adapter"). Please, fill in the required parameters. Depending on your purposes, you can use the configuration file to enable/disable the script stages (data collection, filtering, sending to TM).

Note:

Features:

1. You should specify paths to files relative to the `/mnt/tmp/` directory.
2. The mail server user account should provide full access to mailboxes of required users (login and password).
3. When collecting data, there should be no archived tablespaces for dates from the range that was specified during configuration. If such tablespaces exist, they should be restored before running the script.

To run the Module:

1. Create a directory (you may create it in the directory where the Module is installed), which will be mounted in docker:
   `mkdir tmp`
2. Run the container. This initiates mounting the `temp_dir` directory, which will store results of the Module operation. You can copy files used by the Module (aliases, pst) to this directory.
   `docker run -it --network=host -v /home/tmp:/mnt/tmp/ heml_adapter -c /mnt/tmp/<config_path>`

After this, you can view the script results in subfolders: `/pst, /pst-export, /pst-results.` Log files are located in the `/logs` folder.

Emails received from the Module are available to Traffic Monitor users and mail events. Such events have the following additional attributes:

- the archive unloading name;
- the archive extraction date;
- the workstation name.