

D4Science Infrastructure - Task #1854

Tool for usage statistics on DataMiner

Dec 18, 2015 05:41 PM - Gianpaolo Coro

Status:	Closed	Start date:	Dec 18, 2015
Priority:	Normal	Due date:	
Assignee:	Andrea Dell'Amico	% Done:	100%
Category:	High-Throughput-Computing	Estimated time:	0.00 hour
Target version:	Data Processing		
Infrastructure:	Production		

Description

A tool is required to deeply analyse the requests to DataMiner and their provenance.

This operation is currently done using Linux commands on the wps.statistical server but it could be done more flexibly on DataMiner.

In particular, Andrea Dell'Amico indicated the following tools:

<http://piwik.org/>

<http://www.awstats.org/>

Note that this is different from users' accounting because produced information will be for statistical purposes only, to be reported in the course of projects reviews. Furthermore, this system should report all the requests from non-portal users too (e.g. FishBase).

History

#2 - Jan 20, 2016 04:04 PM - Andrea Dell'Amico

I see that piwik can also ingest log files, so it's the best candidate: it's more modern and more powerful.

#3 - Feb 10, 2016 12:25 PM - Andrea Dell'Amico

- Related to Task #2221: Experiment with a new monitoring/metrics tool: influxdata (influxdb) added

#4 - Feb 10, 2016 12:26 PM - Andrea Dell'Amico

And influxdata could be even better, see <https://support.d4science.org/issues/2221>

#5 - Mar 16, 2016 03:36 PM - Andrea Dell'Amico

- Related to deleted (Task #2221: Experiment with a new monitoring/metrics tool: influxdata (influxdb))

#6 - Mar 16, 2016 03:37 PM - Andrea Dell'Amico

- Status changed from New to In Progress

#7 - Mar 16, 2016 03:46 PM - Andrea Dell'Amico

Hostname is going to be: analytics.d4science.org, IP 146.48.122.20

#8 - Mar 17, 2016 09:34 PM - Andrea Dell'Amico

- % Done changed from 0 to 40

The basic configuration is ready. The server answers at <http://analytics.d4science.org>

It authenticates against ldap but without support to ldap groups (it can use one group only), so the users have to be manually authorized after their first login.

Now I need a procedure to automatically process the dataminer log files.

#9 - Mar 18, 2016 01:57 PM - Andrea Dell'Amico

I'm going to change the dataminer{1,2}-p-d4s.d4science.org nginx configuration to let the access logs be available for download from the analytics server. The new configuration entry will be something like

```
location /logs {
    alias /var/log/nginx;
```

```
allow 146.48.122.20; # analytics.d4science.org
deny all;
}
```

#10 - Mar 18, 2016 06:41 PM - Andrea Dell'Amico

- Status changed from *In Progress* to *Feedback*
- % Done changed from 40 to 100

I've imported the logs of the last two days. The procedure will run every hour from now on.

The imported information is very partial because they are not real visits, but the URLs are complete and the calling IP address is there.

#11 - May 05, 2016 03:37 PM - Gianpaolo Coro

- Assignee changed from *Andrea Dell'Amico* to *Gianpaolo Coro*

#13 - May 06, 2016 05:30 PM - Massimiliano Assante

- Parent task set to #688

#14 - May 24, 2016 04:23 PM - Gianpaolo Coro

- Status changed from *Feedback* to *In Progress*
- Assignee changed from *Gianpaolo Coro* to *Andrea Dell'Amico*

Is it possible to give me access to <http://analytics.d4science.org>?

I would like to evaluate this solution before we go investigating other ones (e.g. elastic search + grafana)

#15 - May 26, 2016 07:43 PM - Andrea Dell'Amico

You should already have access. The credentials are the same of the production portals and redmine.

#16 - May 26, 2016 07:46 PM - Andrea Dell'Amico

All the data logs are frozen at march 26th, but the import script seems working. I need to investigate the behaviour.

#17 - May 30, 2016 10:12 AM - Gianpaolo Coro

It is true, the analytics arrive up to March. Further, defining a new Goal does not filter data, it is like the Goals processing tools is not communicating with the data repository.

#18 - Apr 11, 2017 07:31 PM - Andrea Dell'Amico

- Status changed from *In Progress* to *Closed*