Piwik Server Log Analytics

Piwik (http://piwik.org[1]) allows to also parse log files from your server.

See: http://piwik.org/log-analytics/[2]

Example of command run on seeds4c to parse log files from antoher server (r.t.o) which didn't have piwik installed one year ago, but at least it had apache2 log files saved on disk server side. I've made a copy of those log files to seeds4c (where piwik is installed and recently updated to 2.11.1), and there I can run this type of command:

```
python /var/www/clients/client1/web20/web/misc/log-analytics/import_logs.py --
url=http://piwik.seeds4c.org --dry-run --show-progress --idsite=9
/home/xavi/r.t.o/access.log
```

which produces something like (note it's a "dry run"; i.e.: demo run):

```
root@seeds4c:~# python /var/www/clients/client1/web20/web/misc/log-
analytics/import_logs.py --url=http://piwik.seeds4c.org --dry-run --show-progress -
-idsite=9  /home/xavi/r.t.o/access.log
0 lines parsed, 0 lines recorded, 0 records/sec (avg), 0 records/sec (current)
Parsing log /home/xavi/r.t.o/access.log...

Logs import summary
-------------------

    440 requests imported successfully
    11 requests were downloads
    2639 requests ignored:
        43 HTTP errors
        1666 HTTP redirects
        0 invalid log lines
        0 requests did not match any known site
        0 requests did not match any --hostname
        835 requests done by bots, search engines...
        95 requests to static resources (css, js, images, ico, ttf...)
        0 requests to file downloads did not match any --download-extensions

Website import summary
----------------------

    440 requests imported to 1 sites
```

```
         1 sites already existed
         0 sites were created:

    0 distinct hostnames did not match any existing site:



Performance summary
-------------------

    Total time: 0 seconds
    Requests imported per second: 1477.97 requests per second

root@seeds4c:~#
```

But I have many log files there:

```
root@seeds4c:~# ls /home/xavi/r.t.o/acces*
/home/xavi/r.t.o/access.log        /home/xavi/r.t.o/access.log.22.gz
/home/xavi/r.t.o/access.log.35.gz  /home/xavi/r.t.o/access.log.48.gz
/home/xavi/r.t.o/access.log.1      /home/xavi/r.t.o/access.log.23.gz
/home/xavi/r.t.o/access.log.36.gz  /home/xavi/r.t.o/access.log.49.gz
/home/xavi/r.t.o/access.log.10.gz  /home/xavi/r.t.o/access.log.24.gz
/home/xavi/r.t.o/access.log.37.gz  /home/xavi/r.t.o/access.log.4.gz
/home/xavi/r.t.o/access.log.11.gz  /home/xavi/r.t.o/access.log.25.gz
/home/xavi/r.t.o/access.log.38.gz  /home/xavi/r.t.o/access.log.50.gz
/home/xavi/r.t.o/access.log.12.gz  /home/xavi/r.t.o/access.log.26.gz
/home/xavi/r.t.o/access.log.39.gz  /home/xavi/r.t.o/access.log.51.gz
/home/xavi/r.t.o/access.log.13.gz  /home/xavi/r.t.o/access.log.27.gz
/home/xavi/r.t.o/access.log.3.gz   /home/xavi/r.t.o/access.log.52.gz
/home/xavi/r.t.o/access.log.14.gz  /home/xavi/r.t.o/access.log.28.gz
/home/xavi/r.t.o/access.log.40.gz  /home/xavi/r.t.o/access.log.5.gz
/home/xavi/r.t.o/access.log.15.gz  /home/xavi/r.t.o/access.log.29.gz
/home/xavi/r.t.o/access.log.41.gz  /home/xavi/r.t.o/access.log.6.gz
/home/xavi/r.t.o/access.log.16.gz  /home/xavi/r.t.o/access.log.2.gz
/home/xavi/r.t.o/access.log.42.gz  /home/xavi/r.t.o/access.log.7.gz
/home/xavi/r.t.o/access.log.17.gz  /home/xavi/r.t.o/access.log.30.gz
/home/xavi/r.t.o/access.log.43.gz  /home/xavi/r.t.o/access.log.8.gz
/home/xavi/r.t.o/access.log.18.gz  /home/xavi/r.t.o/access.log.31.gz
/home/xavi/r.t.o/access.log.44.gz  /home/xavi/r.t.o/access.log.9.gz
/home/xavi/r.t.o/access.log.19.gz  /home/xavi/r.t.o/access.log.32.gz
/home/xavi/r.t.o/access.log.45.gz
/home/xavi/r.t.o/access.log.20.gz  /home/xavi/r.t.o/access.log.33.gz
/home/xavi/r.t.o/access.log.46.gz
/home/xavi/r.t.o/access.log.21.gz  /home/xavi/r.t.o/access.log.34.gz
/home/xavi/r.t.o/access.log.47.gz
```

```
root@seeds4c:~#
```

So I'd better write a script to parse them all at once for me:

```
#!/bin/bash
 # see http://tldp.org/HOWTO/Bash-Prog-Intro-HOWTO-7.html

date # show a timestamp

# Parse them all at once ;-)
        for i in $( ls /home/xavi/r.t.o/acces* ); do echo ********** item: $i
**********; python /var/www/clients/client1/web20/web/misc/log-
analytics/import_logs.py --url=http://piwik.seeds4c.org --dry-run --show-progress -
-idsite=9  $i; done

date # show a timestamp
```

---

[1] http://piwik.org

[2] http://piwik.org/log-analytics/