

## Re: Sorting out of sequence log Apache log file

---

*Source:* <http://linux.derkeiler.com/Mailing-Lists/RedHat/2006-04/msg00450.html>

---

- *From:* Ray Van Dolson <[rayvd@xxxxxxxxxxxxxxxxxx](mailto:rayvd@xxxxxxxxxxxxxxxxxx)>
  - *Date:* Wed, 26 Apr 2006 12:43:20 -0700
- 

On Wed, Apr 26, 2006 at 12:13:51PM -0700, Chris W. Parker wrote:

Hello,

I had a hiccup with syslog/apache/logrotate recently and as a result some of the Apache log files are out of sequence. This is bad because Webalizer no longer recognizes the out of sequence lines and my reporting results are skewed.

Is there a command line util that will sort the records correctly? I've been looking around through Google without any luck so far.

I assume by out of sequence you mean the time stamps are all off?

The following quickie Python hack works for me. Basically call it as follows:

```
% cat access_log | /path/to/sort_apache.py > sorted_log.log
```

If it's a huge logfile the script may give you some problems. Basically it reads in all the lines in the file and sorts by the date and time and then spits it out in the right order.

```
#!/usr/bin/python
#
# Simple script to sort an Apache log based on it's time/date field.
#
# Ray Van Dolson <rayvd@xxxxxxxxxxxxxxxxxx>
#

import re
import sys
from time import strptime, mktime

def main():

line_dict = {}

while 1:
```

Re: Sorting out of sequence log Apache log file

```
buf = sys.stdin.readline()
if buf:
# We have data to process.
t = re.match(".*\\[(\\d\\d\\d\\d\\d\\d[A-Za-z]{3}\\d{4}:\\d{2}:\\d{2}) .+?\\].*", buf)
if t:
ts = mktime(strptime(t.group(1), "%d/%b/%Y:%H:%M:%S"))
if not line_dict.has_key(ts):
line_dict[ts] = []
line_dict[ts].append(t.group(0))

else:
break

keys = line_dict.keys()
keys.sort()

for entry in keys:
for line in line_dict[entry]:
print line

if __name__ == '__main__':
main()
```

--

redhat-list mailing list

unsubscribe <mailto:redhat-list-request@xxxxxxxxxx?subject=unsubscribe>

<https://www.redhat.com/mailman/listinfo/redhat-list>