Lumu Technologies
# Backend Integrations Test

## DNS QUERIES COLLECTOR

Lumu customers often require to capture DNS information from their internal DNSs servers and send it using our Collectors API.

You have been tasked with building a Python script to parse a BIND Server log and send it to Lumu using the *Send DNS Queries* method of the *Custom Collector API* (see: https://bit.ly/3oqQzul)

Your script must receive the file to be parsed as an argument, parse it and send the data in chunks of up to 500 records and finally, print a statistic of the parsed data as follows:

```
Total records 6000

Client IPs Rank
--------------  ---  -----
45.238.196.2    426  7.10%
179.0.4.14      324  5.40%
76.237.82.109   305  5.08%
87.190.45.164   261  4.35%
186.117.159.151 183  3.05%
--------------  ---  -----


Host Rank
----------------------------------------------------  ---  -----
TTRTX.local                                           305  5.08%
outlook.office.com                                     29  0.48%
global.asimov.events.data.trafficmanager.net           26  0.43%
MNZ-efz.ms-acdc.office.com                             24  0.40%
skypedataprdcolcus05.cloudapp.net                      21  0.35%
----------------------------------------------------  ---  -----
```

**Requirements**

- Use d39a0f19-7278-4a64-a255-b7646d1ace80 as *lumu-client-key* and 5ab55d08-ae72-4017-a41c-d9d735360288 as *collector-id*
- Be sure you send all the parsed data successfully. You can verify your collector is receiving the data by checking the collector stats at

[https://portal.lumu.io/collectors/custom-collectors/show/5ab55d08-ae72-4017-a41c-d9d735360288](https://portal.lumu.io/collectors/custom-collectors/show/5ab55d08-ae72-4017-a41c-d9d735360288) (user: [lumutest@spam4.me](mailto:lumutest@spam4.me), password: uCNgrBby28RYEKv)

- The stats should be printed as specified above, highlighting the number of records processed, the rank of clients (IPs) performing the most queries, as well as the most queried hosts. The rank should include the total hits and the percentage those hits represent from the total of records analyzed.
- Explain what is the computational complexity of your ranking algorithm.
- Provide instructions on how to run your script.
- Upload your project to GitHub and share it with us.
- If you feel that sending the data to the Lumu API is too much for a test, just parse the file and print the statistics as specified.

The Lumu technical tests are not in any sense intended to perform a detailed assessment of technical skills. They are mostly intended to understand the reasoning process and problem solving abilities of the candidates. Therefore, it is important to track your progress and elaborate on your analytical steps while achieving the solution. You could do it in the comments of your commits.

**Notes:**

You can find a sample file here: [https://bit.ly/33ZEnHy](https://bit.ly/33ZEnHy)

In this file you will find lines as follows:

```
18-May-2021 16:34:13.003 queries: info: client @0x55adcc672cc0 45.231.61.2#80
(pizzaseo.com): query: pizzaseo.com IN A +E(0)D (172.20.101.44)
```

Date: *18-May-2021*, time: *16:34:13.003*, an hexadecimal representation of the client which made the query: *0x55adcc672cc0*, the ip from where the query was made: *45.231.61.2*, and the host queried: *pizzaseo.com*, class of query: *IN*, type of query: *A*.

[https://superuser.com/questions/1264211/understanding-bind9-query-log](https://superuser.com/questions/1264211/understanding-bind9-query-log)