For the "toy" log file d100.log (or another log file), compute:

1. The total number of hits
Ans : 21

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$6 ~ "GET" && $7 ~ ".html$"' d100.log | wc -l
21
pranay@ubuntu:~/Desktop/Data Mining Week3$
```

2. Number of different IP addresses
Ans. 21

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ cat d100.log | wc
    99   1893   18677
pranay@ubuntu:~/Desktop/Data Mining Week3$ cat d100.log | wc -l
99
pranay@ubuntu:~/Desktop/Data Mining Week3$ cut -d" " -f1 d100.log | sort | uniq | wc -l
21
pranay@ubuntu:~/Desktop/Data Mining Week3$
```

3. What methods were used (GET/HEAD, etc), and how many hits used each method?
Ans: GET - 99
      HEAD - 1

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ cut -d" " -f6 d100.log | sort | uniq -c
     99 "GET
      1 "HEAD
```

4. Number of different files requested
Ans: 68

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ cut -d" " -f7 d100.log | sort | uniq | wc -l
68
```

5. The number of HTML pages requested.
Ans: 21

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$7 ~ ".html$"' d100.log | wc -l
21
```

6. Count files ending in .html, .htm, and /, (directories)
Ans: 28

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$7 ~ "(html|html/)$"' d100.log | wc
    21     353    3736
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$7 ~ "(html|html/)$"' d100.log | wc -l
21
```

Compute the same values for the large log file kdlog.zip.

1. The total number of hits
Ans. 7386

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$6 ~ "GET" && $7 ~ ".html$"' da-11-16.ipnt
ld.log | wc -l
7386
```

2. Number of different IP addresses
Ans. 2486

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ cat da-11-16.ipntld.log | wc -l
39974
pranay@ubuntu:~/Desktop/Data Mining Week3$ cut -d" " -f1 da-11-16.ipntld.log | sort | uniq
| wc -l
2486
```

3. What methods were used (GET/HEAD, etc), and how many hits used each method?
Ans:

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ cut -d" " -f6 da-11-16.ipntld.log | sort | uniq
-c
  39851 "GET
     71 "HEAD
     48 "OPTIONS
      4 "POST
```

4. Number of different files requested
Ans: 4566

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ cut -d" " -f7 da-11-16.ipntld.log | sort | uniq
| wc -l
4566
```

5. The number of HTML pages requested.
Ans: 7404

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$7 ~ ".html$"' da-11-16.ipntld.log | wc -l
7404
```

6. Count files ending in .html, .htm, and /, (directories)
Ans: 9540

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$7 ~ "(html|htm|/)$"' da-11-16.ipntld.log
| wc -l
9540
```