

1) Total number of hits

Ans: 21

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

2) Number of successful (code 200 or 304) GETs

Ans: 97

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$6 ~ "GET" && $9 == 200 || $9 == 304' d100.log | wc -l
97
```

3) Number of requests with 404 (not found) status code

Ans: 1

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ gawk '$6 ~ "GET" && $9 == 404' d100.log | wc -l
1
```

4) Visit start (as HHMMSS)

Ans: Sorting the timestamps and starting time is the time when the visit started its 16th Nov 2015 at 00:00:43

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ sort -t" " -k 3 d100.log | cut -d" " -f4 | head -1
[16/Nov/2005:00:00:43
```

5) Visit length (in seconds)

```
pranay@ubuntu:~/Desktop/Data Mining Week3$ sort -t" " -k 3 d100.log | cut -d" " -f4 | head -1
[16/Nov/2005:00:00:43
pranay@ubuntu:~/Desktop/Data Mining Week3$ sort -t" " -k 3 d100.log | cut -d" " -f4 | tail -1
[16/Nov/2005:00:08:52
pranay@ubuntu:~/Desktop/Data Mining Week3$
```

6) Visit agent (assume that the user agent is the same and take it from the first request).