LESSON 11

report.awk

```
awk '{ ip[$1]++ }
END { for (i in ip)
  print ip[i], i }'
  access.log
```

AWK

Real-World AWK

Learn it all @MyWebUniversity



A FREE MYWEBUNIVERSITY COURSE

AWK SCRIPTING

LESSON 11

Real-World AWK

Web server logs, CSV reports, system monitoring and data cleaning

Analyzing a Web Server Log

terminal

```
$ awk '{ ip[$1]++ } END { for (i in ip) print ip[i], i }' access.log | sort -rn | head -5
1520 203.0.113.7
 988 198.51.100.23

$ awk '{ code[$9]++ } END { for (c in code) print c, code[c] }' access.log
200 18233
404 412
500 9
```

- An access log has one request per line: IP address, date, request, status code and size. …
- In one line we find the top visitors, and in another we count each status code, so errors …

01

CSV Sales Report

sales.awk

```
BEGIN { FS = "," }
NR == 1 { next }
{ sales[$2] += $4; total += $4 }
END {
    for (r in sales) printf "%-10s %10.2f %5.1f%%\n", r, sales[r], 100 * sales[r] / total
    printf "%-10s %10.2f\n", "TOTAL", total
}
```

- A sales CSV has a header, then date, region, product and amount. We skip the header, sum by …
- Real CSV can contain quoted commas. gawk has FPAT for that, which describes what a field …

02

System Monitoring

terminal

```
$ df -P | awk 'NR > 1 { u = $5 + 0; if (u > 90) print "WARNING:", $6, u "%" }'
WARNING: /var 93%

$ free | awk '/Mem:/ { printf "memory used: %.1f%%\n", $3 / $2 * 100 }'
memory used: 61.4%
```

- AWK reads the files in slash proc and the output of commands like df and free, so it is …
- Here we warn when a disk is over 90 percent full and print memory use as a percentage.

03

Cleaning Data

clean.awk

```
BEGIN { FS = OFS = "," }
NF == 0 { next }                          # empty line
NF != 4 { print > "rejected.csv"; next }      # wrong shape
{
    for (i = 1; i <= NF; i++) gsub(/^ +| +$/, "", $i)
    $2 = toupper($2); print
}
```

- Before analysis, data must be clean. AWK can drop empty lines, fix spaces, standardize …
- Send bad rows to a separate file so nothing is lost and you can check them later.

04

Speed and Safety Tips



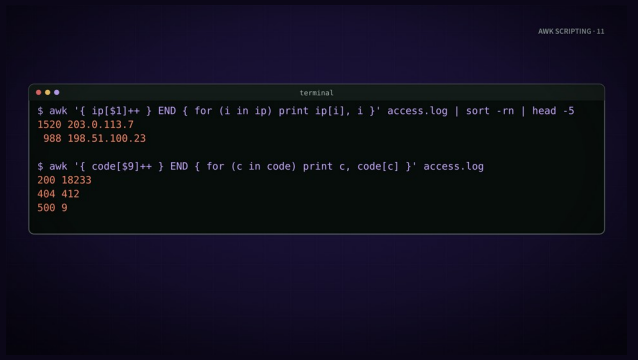
- LC_ALL=C — faster on big files
- filter early — pattern before action
- 'single quotes' — shell-safe program

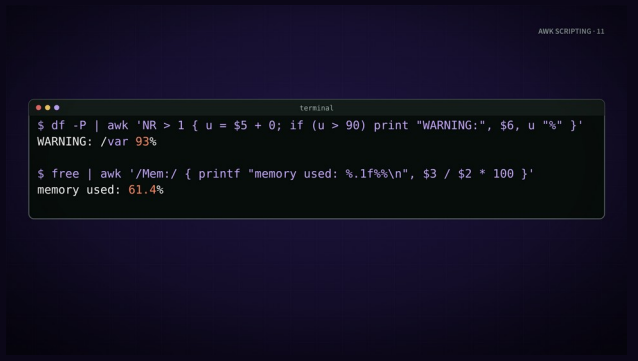
05

Lesson Map


1
Analyzing a Web Server Log
2
CSV Sales Report
3
System Monitoring
4
Cleaning Data
5
Speed and Safety Tips

Sources and Further Reading





Thank you for watching

MyWebUniversity.com

Free technical education since 2000 — English and Dari

@MyWebUniversity · @MyWebUniversity-Dari