Lektion 12 – Laboration 2

5 st övningar med sed, grep och reguljära uttryck.

- Vi kommer fortsätta att gräva i Apache-loggfilerna från tidigare laboration. Använd samma, eller skapa nya. Du kan också använda apache-log-sample.gz från studentportalen.
- Använd https://regex101.com/ för att experimentera och få hjälp att tolka olika uttryck.
    - Välj ”Javascript” – dialekten för minst förvirring.
- Använd bara **zcat**, **sed**, **grep** och kanske **tail**, **bc** och **less**!

1: Använd ”grep” för att plocka ut rader ur loggen

- Lista alla rader där ip-adressen börjar på 200
- Lista alla rader där bytes – värdet (sista fältet) slutar på 3
- Lista alla rader där sökvägen börjar med ”en”
- Lista alla rader där sökvägen börjar med ”e” **eller** ”n”
- Lista alla rader där sekundvärdet i tiden slutar på noll

Överkurs: *Använd sed för att göra samma sak.*

2: Använd ”sed” för att ta bort tidszon från varje loggrad

- Använd en sök-och-ersätt – rad med ”sed” för att byta ut tidszonen mot ”inget”.
  Utmana dig själv genom att använda teckenklasser och kvantifierare. Prova flera lösningar.
- Syntax: “ sed -e 's/pattern/replacement/g' ”

Exempelrad efter bytet:

```
160.11.13.218 - - [20/Mar/2021:16:52:45] "PUT /posts/posts/explore HTTP/1.0" 200 5036
```

3: Använd ”sed” för att plocka ut ip-adressen från varje logg-rad

- Använd en sök-och-ersätt – rad med ”sed” för att ersätta hela raden med bara ip-adressen.

Tips:

- Du kan antingen byta ut det som inte är ip-adressen mot ”inget”, eller byta ut hela raden mot bara ip-adressen med hjälp av grupp-referenser. Prova helst båda varianterna.

4: Dölj känslig data

Vi har från vår infosäk-grupp fått reda på att anrop med metoden ”DELETE” kan innehålla känslig data som måste tas bort från våra loggar.

- Byt med hjälp av ”sed” ut sökvägen för alla rader med metoden ”DELETE” till ”**redacted**”.

Exempelvis borde raden:

```
129.86.40.126 - - [20/Mar/2021:19:36:48 +0000] "DELETE /wp-content HTTP/1.0" 200 4987
```

bytas ut mot:

```
129.86.40.126 - - [20/Mar/2021:19:36:48 +0000] "DELETE **redacted** HTTP/1.0" 200 4987
```

5: Trafiklogg

Vi har fått en förfrågan på en trafiklogg för att jämföra med en misstänkt attack mot vår webbserver.

- Trafikloggen ska innehålla tidsstämpel och datamängd för varje anrop till webbservern.
- Filen ska vara i ”csv” – format, alltså en rad per anrop, med komma-separerade fält.
- Vi är bara intresserade av rader med statuskod 500 (internal server error)
- Formatet på tidsstämpeln spelar ingen roll.

Exempel på resultat:

```
08/Apr/2021:14:46:18 +0000,  5001
08/Apr/2021:14:47:05 +0000,  4938
08/Apr/2021:14:50:27 +0000,  5033
08/Apr/2021:14:51:33 +0000,  5004
08/Apr/2021:15:00:44 +0000,  5077
08/Apr/2021:15:03:18 +0000,  5076
08/Apr/2021:15:07:27 +0000,  4850
08/Apr/2021:15:09:45 +0000,  5010
08/Apr/2021:15:11:18 +0000,  4988
08/Apr/2021:15:12:33 +0000,  4971
```

Överkurs:

- Använd bara **sed** för attt plocka ut rader med rätt status och formattera utskriften enligt ovan.

Tips:

- Använd ”grep” för att plocka ut rader med rätt status. Tänk på att avgränsa statuskoden så du inte matchar exempelvis ”5004” eller ”2500”.
- Använd ”sed” med en ”sök” - regex innehållande två grupper, en som matchar tiden och en som matchar datamängden.
    - Du använder parenteser för att skapa grupper
    - I ”ersätt” – delen av sed-kommandot kan du sen använda gruppreferenser, såhär: ”\1, \2” för att få ut grupperna kommaseparerade.