1.2.2 (Unix)

In [71]:

```
# Extract username and store them into "tweets_username.txt"
!grep -P -o "#[^ \t]+#[^ \t]+" tweets.txt > tweets_twohashtags.txt
```

In [88]:

```
# First 10 lines in "tweets_twohashtags.txt"
!head -10 tweets_twohashtags.txt
```

```
#trueshit...#scaryshit...but
#!%#$^!@%&
#honey.......#imjussayn
#pause.....#megapause
#9:#Virtualization
#8:#Offshore
#thatisall--#agree
#FF--#FF
#LENT....#SMH
#UCanTakeTheKidOutHoodBut#UCanTakeTheKidOutHoodBut#UCanTakeTheKidOu
tHoodBut#UCanTakeTheKidOutHoodBut#UCanTakeTheKidOutHoodBut
```

In [89]:

```
# Calculate frequence of hashtags and store the result into "result_twohastags_7
50000.txt"
!sort tweets_twohashtags.txt | uniq --count | sort -nr > result_twohashtags_7500
00.txt
```

In [90]:

```
# Result of top 10 tweets that have at least two hashtags
! head -10 result_twohashtags_750000.txt
```

```
      8 #affiliate#marketing
      5 #zewdy#zewdy
      5 #BGC#BGC
      5 ####
      4 #??PFoundersday#??PFoundersday
      4 #Celebrity,#Philanthropy
      3 #AKA#AKA
      3 #39;What&#39;s
      3 #39;streaming&#39;
      3 #39;Green&#39;
```

In [104]:

```
# Shell script for 1.2 (2)
!cat 1_3.sh
```

```
#!/bin/sh
grep -P -o "#[^ \t]+#[^ \t]+" tweets.txt > tweets_twohashtags.txt
sort tweets_twohashtags.txt | uniq --count | sort -nr > result_twoh
ashtags_750000.txt
head -10 result_twohashtags_750000.txt
```

In [105]:

```
# Runtime of 1.2 (2) using Unix command
!time bash 1_3.sh
```

```
      8 #affiliate#marketing
      5 #zewdy#zewdy
      5 #BGC#BGC
      5 ####
      4 #??PFoundersday#??PFoundersday
      4 #Celebrity,#Philanthropy
      3 #AKA#AKA
      3 &#39;What&#39;s
      3 &#39;streaming&#39;
      3 &#39;Green&#39;
0.17user 0.01system 0:00.18elapsed 100%CPU (0avgtext+0avgdata 3108m
axresident)k
0inputs+48outputs (0major+828minor)pagefaults 0swaps
```