

Εισαγωγή στα Λειτουργικά Συστήματα



SET ΔΙΑΦΑΝΕΙΩΝ 16

SHELL SCRIPTS: ΦΙΛΤΡΑ

ΑΝΤΩΝΗΣ ΣΙΔΗΡΟΠΟΥΛΟΣ

Ορισμός

2

- Ο όρος φίλτρο χρησιμοποιείται στην ορολογία του Unix να αναφερθεί σε οποιοδήποτε πρόγραμμα που
 1. διαβάζει δεδομένα από την κανονική είσοδο,
 2. εκτελεί κάποια λειτουργία στα δεδομένα,
 3. και γράφει τα αποτελέσματα στην κανονική έξοδο.

Ορισμός

3

- Πιο συνοπτικά, ένα φίλτρο είναι οποιοδήποτε πρόγραμμα που μπορεί να χρησιμοποιηθεί ανάμεσα σε δύο άλλα προγράμματα σε μια διασωλήνωση. Έτσι, στο παράδειγμα
`ls -l | grep '^-' | wc`
 - η `grep` θεωρείται ένα φίλτρο
 - η `wc`, θεωρείται ένα φίλτρο.
 - η `ls` δεν είναι γιατί δεν διαβάζει την είσοδό της από την κανονική είσοδο.
- Άλλα παραδείγματα, η `cat` και η `sort` είναι τα φίλτρα, ενώ οι `who`, `date`, `cd`, `pwd`, `echo`, `rm`, `mv`, και `cp` δεν είναι.

Η φιλοσοφία

4

- Τα φίλτρα χρησιμοποιούνται για να «φιλτράρουν» δεδομένα... Ουσιαστικά ροές δεδομένων.
- Υπάρχουν πολλές εντολές σε ένα σύστημα UNIX που μπορούν να θεωρηθούν φίλτρα.
- Η ιδέα είναι ότι χρησιμοποιώντας συνδυασμούς από «απλά» φίλτρα μπορούμε να κάνουμε πολύπλοκες μετατροπές (\Leftrightarrow υπολογισμούς) σε δεδομένα.

Η φιλοσοφία

5

- Τα φίλτρα χρησιμοποιούνται συνήθως σε συνδυασμό με την διασωλήνωση (| pipelining).
- Παράδειγμα:
 - Να δώσετε έναν συνδυασμό εντολών με τις οποίες να βρίσκουμε ποιοι είναι οι υπο-φάκελοι ενός φακέλου, τα ονόματά των οποίων δεν καταλήγουν σε αριθμό και να εμφανίσετε όνομα φακέλου, μέγεθος και ημερομηνία τροποποίησης ταξινομημένα με βάση το πλήθος των i-nodes κατά φθίνουσα σειρά !!!!

Η φιλοσοφία

6

- Η λύση

```
bash-2.05a$ ls -l | grep '^d' | grep -v '[0-9]$\ ' | sort -  
k2nr | sed 's/^[^ ]* \{1,\}[^ ]* \{1,\}[^ ]* \{1,\}[^ ]*  
\{1,\}\/\ ' | sed 's/^\([0-9]*\) *\([^\ ]* *[0-9]* *[0-9:]*\)\  
\{1,\}\(. *\)$/\3 \1 \2/'  
public_html 4096 Nov 2 22:18  
perl 64 Dec 22 2002  
ttt 100 Dec 13 00:35  
usr 26 Aug 27 2003  
soft 42 Aug 27 2003  
courses 53 Jan 18 2003  
tests 57 Jun 6 2004  
mail 68 Dec 14 2004  
DBII 130 May 14 2004  
dbii bin 4096 Oct 7 13:43  
BAKS 27 Dec 31 2002  
valentinos vassiliou 35 Jan 19 2006  
test 4096 Dec 6 01:28  
bash-2.05a$
```

Ευκολότερο Παράδειγμα:

7

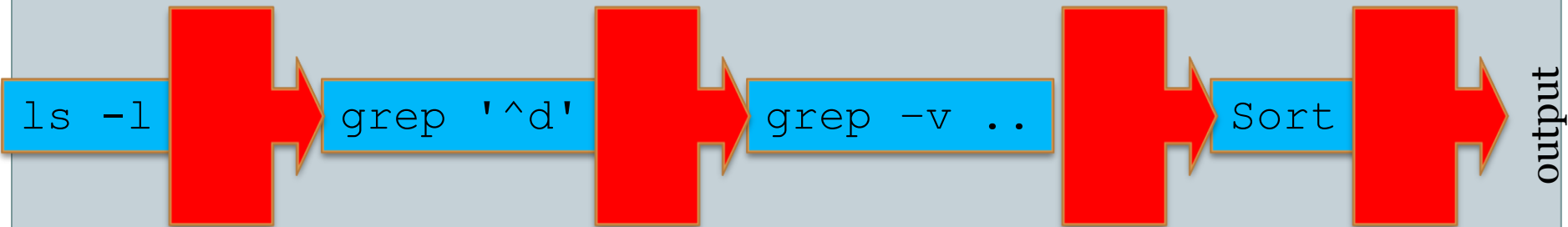
- Να δώσετε έναν συνδυασμό εντολών με τις οποίες να βρίσκουμε ποιοι είναι οι υπο-φάκελοι ενός φακέλου, τα ονόματά τους δεν καταλήγουν σε αριθμό και να τα εμφανίσετε ταξινομημένα με βάση το πλήθος των i-nodes κατά φθίνουσα σειρά:

```
bash-2.05a$ ls -l | grep '^d' | grep -v '[0-9]$\ ' | sort -k2nr
drwxr-x--x      7 asidirop it          4096 Nov  2 22:18 public_html
drwxr-xr-x      5 asidirop it           64 Dec 22  2002 perl
drwxr-xr-x      5 asidirop it          100 Dec 13 00:35 ttt
drwxr-xr-x      4 asidirop it           26 Aug 27  2003 usr
drwxr-xr-x      3 asidirop it           42 Aug 27  2003 soft
drwx-----    2 asidirop it           53 Jan 18  2003 courses
drwx-----    2 asidirop it           57 Jun  6  2004 tests
drwx-----    2 asidirop it           68 Dec 14  2004 mail
drwx-----    2 asidirop it          130 May 14  2004 DBII
drwx--x--x      2 asidirop it          4096 Oct  7 13:43 dbii_bin
drwxr-xr-x      2 asidirop it           27 Dec 31  2002 BAKS
drwxr-xr-x      2 asidirop it           35 Jan 19  2006 vassiliou
drwxr-xr-x      2 asidirop it          4096 Dec  6 01:28 test
```

Διασωλήνωση

8

```
ls -l | grep '^d' | grep -v '[0-9]$\ ' |  
sort -k2nr
```



Διασωλήνωση

9

```
ls -l | grep '^d' | grep -v '[0-9]$\ ' | sort -k2nr
```

ls -l

grep '^d'

grep -v ..

Sort

output

```
total 15
drwxr-xr-x  2 asidirop it      27 Dec 31  2002 BAKS
drwx----- 2 asidirop it    130 May 14  2004 DBII
drwxr-xr-x  2 asidirop it    115 May  4  2004 DBII_lab6
-rw-r--r--  1 asidirop it  1844 Oct  7 12:41 TEI_mails_all_2005-06e.txt
-rw-----  1 asidirop it    506 Dec  5  2004 Trash
-rw-r--r--  1 asidirop it  3253 Oct 10 12:54 all.txt
drwx----- 2 asidirop it     53 Jan 18  2003 courses
-rw-r--r--  1 asidirop it  2170 Oct 10 12:53 db.txt
drwx--x--x  2 asidirop it   4096 Oct  7 13:43 dbii_bin
-rw-r--r--  1 asidirop it      0 Nov 20 13:42 file1
-rw-r--r--  1 asidirop it 37166 Aug 30  2003 kernel-ntfs-2.4.20-20.9.i686.rpm
drwx----- 2 asidirop it     68 Dec 14  2004 mail
drwx--x--x 19 asidirop it   4096 Jun 17  2005 op2004-05e-T4-1
-rw-r--r--  1 asidirop it  1523 Oct 10 12:49 os.txt
drwxr-xr-x  5 asidirop it     64 Dec 22  2002 perl
drwxr-x--x  7 asidirop it   4096 Nov  2 22:18 public_html
drwxr-xr-x  3 asidirop it     42 Aug 27  2003 soft
drwxr-xr-x  2 asidirop it   4096 Dec  6 01:28 test
drwx----- 2 asidirop it     57 Jun  6  2004 tests
drwxr-xr-x  5 asidirop it    100 Dec 13 00:35 ttt
```

Διασωλήνωση

10

```
ls -l | grep '^d' | grep -v '[0-9]$\ ' | sort -k2nr
```

ls -l

grep '^d'

grep -v ..

Sort

output

```
total 15
drwxr-xr-x  2 asidirop it      27 Dec 31  2002 BAKS
drwx----- 2 asidirop it     130 May 14  2004 DBII
drwxr-xr-x  2 asidirop it     115 May  4  2004 DBII_lab6
-rw-r--r--  1 asidirop it    1844 Oct  7 12:41 TEI_mails_all_2005-06e.txt
-rw-----  1 asidirop it     506 Dec  5  2004 Trash
-rw-r--r--  1 asidirop it    3253 Oct 10 12:54 all.txt
drwx----- 2 asidirop it      53 Jan 18  2003 courses
-rw-r--r--  1 asidirop it    2170 Oct 10 12:53 db.txt
drwx--x--x  2 asidirop it    4096 Oct  7 13:43 dbii_bin
-rw-r--r--  1 asidirop it       0 Nov 20 13:42 file1
-rw-r--r--  1 asidirop it   37166 Aug 30  2003 kernel-ntfs-2.4.20-20.9.i686.rpm
drwx----- 2 asidirop it      68 Dec 14  2004 mail
drwx--x--x 19 asidirop it    4096 Jun 17  2005 op2004-05e-T4-1
-rw-r--r--  1 asidirop it    1523 Oct 10 12:49 os.txt
drwxr-xr-x  5 asidirop it      64 Dec 22  2002 perl
drwxr-x--x  7 asidirop it    4096 Nov  2 22:18 public_html
drwxr-xr-x  3 asidirop it      42 Aug 27  2003 soft
drwxr-xr-x  2 asidirop it    4096 Dec  6 01:28 test
drwx----- 2 asidirop it      57 Jun  6  2004 tests
drwxr-xr-x  5 asidirop it     100 Dec 13 00:35 ttt
```

Διασωλήνωση

11

```
ls -l | grep '^d'|grep -v '[0-9]$\'|sort -k2nr
```

ls -l

grep '^d'

grep -v ..

Sort

output

```
drwxr-xr-x    2 asidirop it          27 Dec 31  2002 BAKS
drwx-----    2 asidirop it        130 May 14  2004 DBII
drwxr-xr-x    2 asidirop it        115 May  4  2004 DBII_lab6
drwx-----    2 asidirop it         53 Jan 18  2003 courses
drwx--x--x    2 asidirop it       4096 Oct  7 13:43 dbii_bin
drwx-----    2 asidirop it         68 Dec 14  2004 mail
drwx--x--x   19 asidirop it       4096 Jun 17  2005 op2004-05e-T4-1
drwxr-xr-x    5 asidirop it         64 Dec 22  2002 perl
drwxr-x--x    7 asidirop it       4096 Nov  2 22:18 public_html
drwxr-xr-x    3 asidirop it         42 Aug 27  2003 soft
drwxr-xr-x    2 asidirop it       4096 Dec  6 01:28 test
drwx-----    2 asidirop it         57 Jun  6  2004 tests
drwxr-xr-x    5 asidirop it        100 Dec 13 00:35 ttt
drwxr-xr-x    4 asidirop it         26 Aug 27  2003 usr
drwxr-xr-x    2 asidirop it         35 Jan 19  2006 valentinos_vassiliou
```

Διασώληνωση

12

```
ls -l | grep '^d'|grep -v '[0-9]$\'|sort -k2nr
```



```
drwxr-xr-x    2 asidirop it          27 Dec 31   2002 BAKS
drwx-----    2 asidirop it        130 May 14   2004 DBII
drwxr-xr-x    2 asidirop it        115 May  4   2004 DBII_lab6
drwx-----    2 asidirop it         53 Jan 18   2003 courses
drwx--x--x    2 asidirop it       4096 Oct  7  13:43 dbii_bin
drwx-----    2 asidirop it         68 Dec 14   2004 mail
drwx--x--x    19 asidirop it      4096 Jun 17   2005 op2004-05e-T4-1
drwxr-xr-x    5 asidirop it         64 Dec 22   2002 perl
drwxr-x--x    7 asidirop it       4096 Nov  2  22:18 public_html
drwxr-xr-x    3 asidirop it         42 Aug 27   2003 soft
drwxr-xr-x    2 asidirop it       4096 Dec  6  01:28 test
drwx-----    2 asidirop it         57 Jun  6   2004 tests
drwxr-xr-x    5 asidirop it        100 Dec 13  00:35 ttt
drwxr-xr-x    4 asidirop it         26 Aug 27   2003 usr
drwxr-xr-x    2 asidirop it         35 Jan 19   2006 valentinos_vassiliou
```

Διασώληνωση

13

```
ls -l | grep '^d'|grep -v '[0-9]$\'|sort -k2nr
```



```
drwxr-xr-x    2 asidirop it          27 Dec 31   2002 BAKS
drwx-----    2 asidirop it        130 May 14   2004 DBII
drwx-----    2 asidirop it         53 Jan 18   2003 courses
drwx--x--x    2 asidirop it       4096 Oct  7  13:43 dbii_bin
drwx-----    2 asidirop it         68 Dec 14   2004 mail
drwxr-xr-x    5 asidirop it          64 Dec 22   2002 perl
drwxr-x--x    7 asidirop it       4096 Nov  2  22:18 public_html
drwxr-xr-x    3 asidirop it          42 Aug 27   2003 soft
drwxr-xr-x    2 asidirop it       4096 Dec  6  01:28 test
drwx-----    2 asidirop it         57 Jun  6   2004 tests
drwxr-xr-x    5 asidirop it        100 Dec 13  00:35 ttt
drwxr-xr-x    4 asidirop it          26 Aug 27   2003 usr
drwxr-xr-x    2 asidirop it          35 Jan 19   2006 valentinos_vassiliou
```

Διασωλήνωση

14

```
ls -l | grep '^d'|grep -v '[0-9]$\'|sort -k2nr
```

ls -l

grep '^d'

grep -v ..

Sort

output

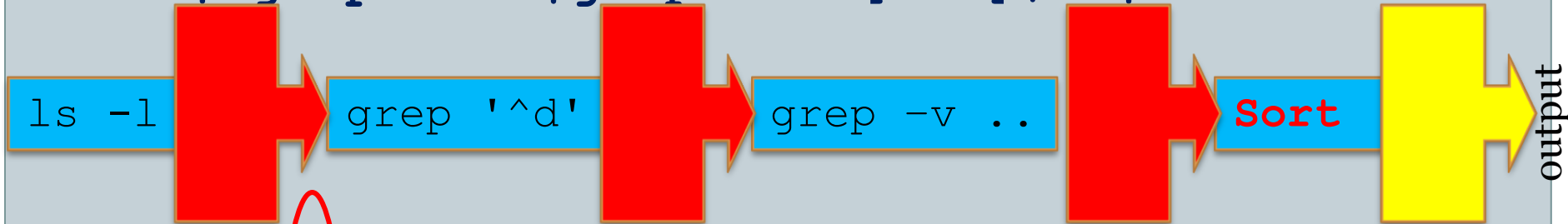
drwxr-xr-x	2	asidirop	it	27	Dec	31	2002	BAKS
drwx-----	2	asidirop	it	130	May	14	2004	DBII
drwx-----	2	asidirop	it	53	Jan	18	2003	courses
drwx--x--x	2	asidirop	it	4096	Oct	7	13:43	dbii_bin
drwx-----	2	asidirop	it	68	Dec	14	2004	mail
drwxr-xr-x	5	asidirop	it	64	Dec	22	2002	perl
drwxr-x--x	7	asidirop	it	4096	Nov	2	22:18	public_html
drwxr-xr-x	3	asidirop	it	42	Aug	27	2003	soft
drwxr-xr-x	2	asidirop	it	4096	Dec	6	01:28	test
drwx-----	2	asidirop	it	57	Jun	6	2004	tests
drwxr-xr-x	5	asidirop	it	100	Dec	13	00:35	ttt
drwxr-xr-x	4	asidirop	it	26	Aug	27	2003	usr
drwxr-xr-x	2	asidirop	it	35	Jan	19	2006	valentinos_vassiliou

- Η sort θα διαβάσει την είσοδό της και θα κάνει ταξινόμηση με βάση την 2^η στήλη, αριθμητικά και φθίνουσα σειρά.

Διασωλήνωση

15

```
ls -l | grep '^d'|grep -v '[0-9]$\'|sort -k2nr
```



drwxr-x--x	7	asidirop	it	4096	Nov	2	22:18	public_html
drwxr-xr-x	5	asidirop	it	100	Dec	13	00:35	ttt
drwxr-xr-x	5	asidirop	it	64	Dec	22	2002	perl
drwxr-xr-x	4	asidirop	it	26	Aug	27	2003	usr
drwxr-xr-x	3	asidirop	it	42	Aug	27	2003	soft
drwx-----	2	asidirop	it	130	May	14	2004	DBII
drwx-----	2	asidirop	it	53	Jan	18	2003	courses
drwx-----	2	asidirop	it	57	Jun	6	2004	tests
drwx-----	2	asidirop	it	68	Dec	14	2004	mail
drwxr-xr-x	2	asidirop	it	27	Dec	31	2002	BAKS
drwxr-xr-x	2	asidirop	it	35	Jan	19	2006	valentinos_vassiliou
drwxr-xr-x	2	asidirop	it	4096	Dec	6	01:28	test
drwx--x--x	2	asidirop	it	4096	Oct	7	13:43	dbii_bin

- Η sort θα τυπώσει το αποτέλεσμα στην κανονική έξοδο.

Διασωλήνωση

16

- Όταν χρησιμοποιούμε φίλτρα με διασωλήνωση, συνήθως δεν κάνουμε ανακατεύθυνση εισόδου/εξόδου.
- Μπορούμε μόνο να αλλάξουμε την κανονική είσοδο (stdin) της πρώτης εντολής και την κανονική έξοδο (stdout) της τελευταίας εντολής.
- **Αν αλλάξουμε την κανονική έξοδο μιας ενδιάμεσης εντολής, τότε χάνεται η «έννοια» της διασωλήνωσης.**

Παράδειγμα λάθους

17

- Στο προηγούμενο παράδειγμα αν είχαμε βάλει και μια ανακατεύθυνση στην 3^η εντολή:
- `ls -l | grep '^d' | grep -v '[0-9]$\ ' > file2 | sort -k2nr`
- μέχρι την 3^η εντολή θα εκτελούνταν κανονικά (όπως και πριν)
- Η έξοδος από την 3^η εντολή θα αποθηκευτεί στο αρχείο file2
- Η 3^η εντολή ΔΕΝ θα στείλει δεδομένα στην 4^η
- Η 4^η εντολή ΔΕΝ βρίσκει να διαβάσει κάτι (διαβάζει κενό αρχείο) και άρα δεν τυπώνει τίποτα.

Διασωλήνωση

18

- Αν αλλάξουμε την κανονική είσοδο μιας ενδιάμεσης εντολής, τότε χάνεται η «έννοια» της διασωλήνωσης.

Παράδειγμα λάθους

19

- Στο προηγούμενο παράδειγμα αν είχαμε βάλει και μια ανακατεύθυνση στην 3^η εντολή:
- `ls -l | grep '^d' | grep -v '[0-9]$\ ' < file2 | sort -k2nr`
- μέχρι την 2^η εντολή θα εκτελεστούν κανονικά (όπως και πριν)
- Η 3^η εντολή ΔΕΝ θα διαβάσει από την έξοδο της 2^{ης} , αλλά από το αρχείο file2.
- Η έξοδος από την 2^η εντολή δεν θα διαβαστεί ποτέ από κανέναν και θα χαθεί.
- Η 3^η εντολή ψάχνει τις γραμμές του αρχείου file2 που δεν τελειώνουν σε αριθμό και στις στέλνει στην επόμενη εντολή.
- Πρακτικά το προηγούμενο είναι ισοδύναμο με:
`grep -v '[0-9]$\ ' < file2 | sort -k2nr`
διότι ότι υπολογίζουμε νωρίτερα χάνεται.

Διασωλήνωση

20

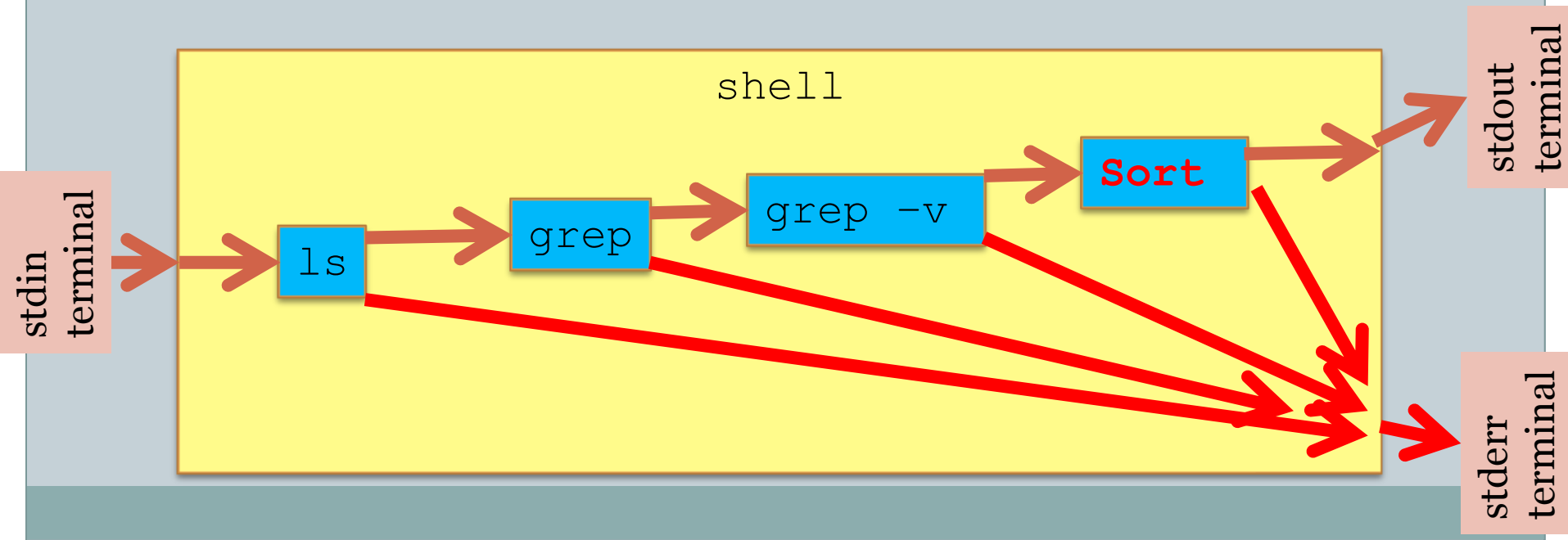
- Οι εντολές μιας διασωλήνωσης εκτελούνται «παράλληλα».
 - Ουσιαστικά εκτελούνται ψευδο-παράλληλα, εκτός και αν έχουμε υπολογιστή με πολλούς επεξεργαστές, οπότε σε αυτή την περίπτωση έχουμε πραγματική παράλληλη εκτέλεση.
- Συνήθως τα «φίλτρα» διαβάζουν γραμμή-γραμμή την είσοδό τους μέχρι να συναντήσουν τον χαρακτήρα EOF.

Διασωλήνωση & έξοδος λαθών

21

- Χρησιμοποιώντας την παρακάτω εντολή, δεν έχουμε αλλάξει σε καμία από τις εντολές την έξοδο λαθών. Άρα η έξοδος λαθών για την κάθε εντολή κληρονομείται από το shell, και πιθανόν θα είναι το τερματικό.

```
ls -l | grep '^d' | grep -v '[0-9]$\ ' | sort -k2nr
```

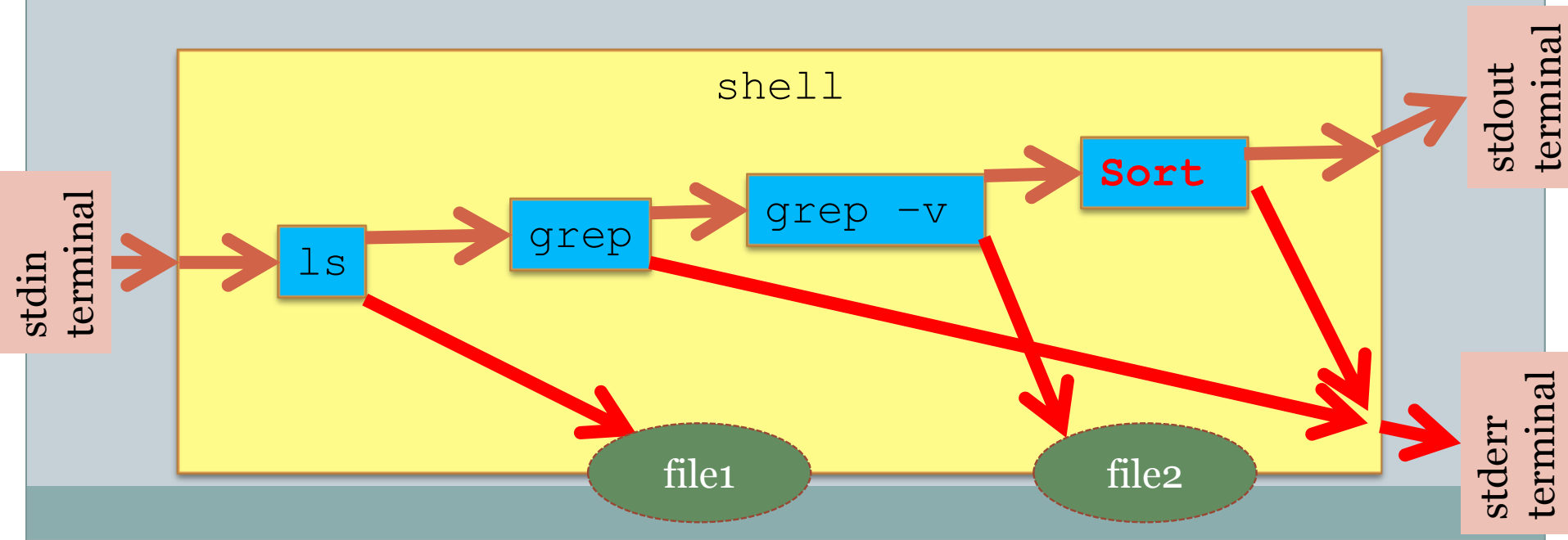


Διασωλήνωση & έξοδος λαθών

22

- Για κάθε εντολή (ή για κάποιες) θα μπορούσαμε να αλλάξουμε την έξοδο λαθών.

```
ls -l 2> file1 | grep '^d'|grep -v  
'[0-9]$\ ' 2> file2 |sort -k2nr
```



Οι εντολές μιας διασωλήνωσης εκτελούνται «παράλληλα».

23

- Για να πειραματιστούμε φτιάχνουμε το script `delay_ls`.
- Το δίπλα script εκτελεί την εντολή `ls -l` και δίνει τα αποτελέσματά της σε μια `while`.
- Η `while` διαβάζει γραμμή-γραμμή την είσοδό της και την τυπώνει. Μετά όμως από κάθε γραμμή κάνει μια χρονο-καθυστέρηση του ενός δευτερολέπτου.
- Εκτελέστε το `delay_ls`. Θα κάνει τόσα δευτερόλεπτα να ολοκληρωθεί όσα και τα αρχεία στον τρέχον φάκελο.

FILE: `delay_ls`

```
#!/bin/bash

ls -l | while read a; do
    echo "$a";
    sleep 1;
done
```

Οι εντολές μιας διασωλήνωσης εκτελούνται «παράλληλα».

24

- Αν εκτελέσετε την εντολή:
- `./delay_ls | grep '^-'`
θα διαπιστώσετε πως η εντολή `grep` δεν καθυστερεί να τυπώσει το αποτέλεσμα της (ξέρουμε ότι κάθε δευτερόλεπτο διαβάσει μια γραμμή)
- Μόλις διαβάσει μια γραμμή η `grep`, την ελέγχει και την τυπώνει εφόσον ταιριάζει στο πρότυπο.

```
asidirop@dellpc:/tmp$ ./delay_ls | grep '^-'
-rw-r--r-- 1 asidirop asidirop      0 2012-05-27 10:25 ACPIBatteryScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop      0 2012-05-27 10:25 ClearCalendarScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop      0 2012-05-27 10:25 ClearWeatherScreenlet.py.log
-rwxr-xr-x 1 asidirop asidirop    64 2012-05-27 11:36 delay_ls
-rw----- 1 root     root      1408 2012-05-27 10:25 filebMF0Cb
-rw-r--r-- 1 asidirop asidirop      0 2012-05-27 10:25 InfoPanelScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop      0 2012-05-27 10:25 MountScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop      0 2012-05-27 10:25 PlacesScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop      0 2012-05-27 10:25 screenlets-daemon.py.log
asidirop@dellpc:/tmp$
```


Οι εντολές μιας διασωλήνωσης εκτελούνται «παράλληλα».

25

- Με την διασωλήνωση δεν μας πειράζει ο ρυθμός εγγραφής/ανάγνωσης δεδομένων.
- Στο παρακάτω:
 - ls: τυπώνει "γρήγορα" τα δεδομένα
 - delay2: διαβάζει και τυπώνει "αργά" → Το 1^ο pipe συσσωρεύει δεδομένα μέχρι να τα διαβάσει η delay2
 - grep: διαβάζει γρήγορα, αλλά της έρχονται δεδομένα με αργό ρυθμό → μπαίνει συνεχώς σε αναμονή (I/O wait). → Το 2^ο pipe πάντα θα είναι σχεδόν άδειο.

FILE: delay2

```
#!/bin/bash

while read a; do
    echo "$a";
    sleep 1;
done
```

```
asidirop@dellpc:/tmp$ ls -l | ./delay2 | grep '^-'
-rw-r--r-- 1 asidirop asidirop 0 2012-05-27 10:25 ACPIBatteryScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop 0 2012-05-27 10:25 ClearCalendarScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop 0 2012-05-27 10:25 ClearWeatherScreenlet.py.log
-rwxr-xr-x 1 asidirop asidirop 64 2012-05-27 11:36 delay_ls
-rw----- 1 root root 1408 2012-05-27 10:25 filebMF0Cb
-rw-r--r-- 1 asidirop asidirop 0 2012-05-27 10:25 InfoPanelScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop 0 2012-05-27 10:25 MountScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop 0 2012-05-27 10:25 PlacesScreenlet.py.log
-rw-r--r-- 1 asidirop asidirop 0 2012-05-27 10:25 screenlets-daemon.py.log
asidirop@dellpc:/tmp$
```