Les filtres et utilitaires

Un **filtre** (ou une commande filtre) est un programme sachant écrire et lire des données par les canaux standards d'entrée et de sortie. Il en modifie ou traite éventuellement le contenu. **wc** est un filtre.

Les utilitaires sans être obligatoirement des filtres permettent un certain nombre d'actions sur des fichiers ou leur contenu comme le formatage ou l'impression.

1. Extraction des noms et chemins

La commande **basename** permet d'extraire le nom du fichier dans un chemin.

```
$ basename /tmp/driss/liste
liste
```

La commande **dirname** effectue l'inverse, elle extrait le chemin.

```
$ dirname /tmp/driss/liste
/tmp/driss
```

2. Recherche de lignes

Il s'agit d'extraire des lignes d'un fichier selon divers critères. Pour cela vous disposez de trois commandes **grep**, **egrep** et **fgrep** qui lisent les données soit depuis un fichier d'entrée, soit depuis le canal d'entrée standard.

a. grep

La syntaxe de la commande **grep** est :

```
grep [Options] modèle [Fichier1...]
```

Le modèle se compose de critères de recherche ressemblant beaucoup aux critères déjà exposés pour les wildcard (expression régulière). Il ne faut pas oublier que ces critères doivent être interprétés par la commande **grep** et pas par le shell. Il faut donc verrouiller tous les caractères.

```
$ cat fic4
Cochon
Veau
Boeuf
rat
Rat
boeuf
$ grep "^[bB]" fic4
Boeuf
bœuf
```

La commande **grep** peut aussi prendre quelques options intéressantes.

- **-v** effectue la recherche inverse : toutes les lignes ne correspondant pas aux critères sont affichées.
- **-c** ne retourne que le nombre de lignes trouvées sans les afficher.
- **-i** ne différencie pas les majuscules et les minuscules.
- **-n** indique le numéro de ligne pour chaque ligne trouvée.
- **-l** dans le cas de fichiers multiples, indique dans quel fichier la ligne a été trouvée.

```
$ grep -i "^b" fic4
Boeuf
boeuf
```

b. egrep

La commande **egrep** étend les critères de recherche et peut accepter un fichier de critères en entrée. Elle est équivalente à un grep -E. Elle emploie comme critères des expressions régulières.

```
egrep -f fichier_critère fichier_recherche
```

Caractère spécial	Signification
|	Ou logique, l'expression située avant ou après doit apparaître.
(...)	Groupement de caractères.
[...]	Un caractère à cette position parmi ceux indiqués.
. (point)	Un caractère quelconque.
+	Répétition, le caractère placé avant doit apparaître au moins une fois.
*	Répétition, le caractère situé avant doit apparaître de zéro à n fois.

Caractère spécial	Signification
?	Le caractère situé avant doit apparaître une fois au plus.
{n}	Le caractère situé avant doit apparaître exactement n fois.
{n,}	Il apparaît n fois ou plus.
{n,m}	Il apparaît entre n et m fois.
^	En début de chaîne.
$	En fin de chaîne.

Seulement bonjour et bonsoir commençant par une majuscule ou une minuscule s'ils sont seuls sur une ligne :

```
^[bB]on(jour|soir)$
```

Vérification très sommaire de la validité d'une adresse IP :

```
echo $IP | egrep '([0-9]{1,3}\.){3}[0-9]{1,3}'
```

Voici comment cette ligne est décomposée :

- ([0-9]{1,3}\.){3} : www.xxx.yyy
  - [0-9] : un caractère entre 0 et 9
  - {1,3} : répété entre une et trois fois, donc x, xx ou xxx
  - \. : suivi d'un point
  - {3} : le tout trois fois
- Puis [0-9]{1,3} : zzz
  - [0-9] : un caractère entre 0 et 9
  - {1,3} : répété entre une et trois fois

c. fgrep

La commande **fgrep** est un grep simplifié et rapide (fast grep) et équivaut à un grep -F . Elle accepte aussi un fichier de critères de recherche mais il s'agit là de critères simples, sans caractères spéciaux. Vous saisissez dans le fichier de critères des lignes simples (du texte et des chiffres), une recherche par ligne. Fgrep va alors rechercher dans un fichier cible ou un flux en entrée les lignes

correspondant à chacun des critères.

d. sed

L’apprentissage de sed demanderait toute une année (bon ok ! un module big data à lui tout seul). Sed est un éditeur de flux (Stream Editor) permettant de filtrer et de transformer du texte. C’est un peu comme un éditeur permettant de modifier du texte via des commandes scripts, mais en une passe et sans édition interactive. Il utilise un jeu étendu de commandes issu de l’éditeur ed. Sa syntaxe de base est :

```
sed -e ’<cmd>’ fic
```

Pour utiliser sed, il faut apprendre et comprendre les expressions rationnelles. Le tableau de la commande **egrep** reprend la syntaxe de base des expressions. Tout ouvrage sur sed part de ces expressions, et réciproquement.

Sed est très souvent utilisé pour remplacer des valeurs par d’autres (substitution) ou supprimer des lignes particulières (bien que grep pourrait être utilisé dans ce cas). La syntaxe basique de substitution est la suivante :

```
s/<ancien>/nouveau/[g]
```

Le g final permet de faire un remplacement sur toute la ligne en cas de présence de plusieurs occurrences. Voici un exemple qui remplace __NOM__ par Toto :

```
$ echo "Je m’appelle __NOM__. Tu t’appelles __NOM__?" | sed -e 's/__
NOM__/Toto/'
Je m’appelle Toto. Tu t’appelles __NOM ?
$ echo "Je m’appelle __NOM__. Tu t’appelles __NOM__ ?" | sed -e 's/__
NOM__/Toto/g'
Je m’appelle Toto. Tu t’appelles Toto ?
```

Vous pouvez placer une valeur numérique dans le champ nouveau pour préciser, si la recherche comporte plusieurs éléments regroupés par des parenthèses, sur quel élément recherché travailler. Voici un simple exemple qui rajoute des étoiles autour du nom toto :

```
$ echo toto | sed -e "s/\(toto\)/**\1**/"
**toto**
```

Pour supprimer toutes les lignes vides ou ne contenant que des espaces :

```
$ sed -e ’/^ *$/d’ fichier
```

3. Colonnes et champs

La commande **cut** permet de sélectionner des colonnes et des champs dans un fichier.

a. Colonnes

La syntaxe est la suivante :

```
cut -cColonnes [fic1...]
```

Une colonne est la position d'un caractère dans la ligne. Le premier caractère est la colonne 1, le deuxième la colonne 2, et ainsi de suite. Une ligne de 80 caractères dispose de 80 colonnes. La numérotation commence à 1. C'est la méthode idéale pour des fichiers plats et à format fixe où chaque champ débute et finit à des positions données.

Le format de sélection de colonne est le suivant :

- une colonne seule (ex. -c2 pour la colonne 2).
- une plage (ex. -c2-4 pour les colonnes 2, 3 et 4).
- une liste de colonnes (ex. -c1,3,6 pour les colonnes 1, 3 et 6).
- les trois en même temps (ex. -c1-3,5,6,12-).

```
$ cat liste
Produit prix    quantites
souris  30      15
disque  100     30
ecran   300     20
clavier 45      30

$ cut -c1-5 liste
Produ
souri
disqu
ecran
clavi

$ cut -c1-3,10-12,15
Prorx   quantites
sou0    15
dis0    30
ecr0    20
cla530
```

b. Champs

La commande **cut** permet aussi de sélectionner des champs. Ces champs doivent être par défaut délimités par une tabulation, mais le paramètre -d permet de sélectionner un autre caractère (espace, ;). La sélection des champs est identique à celle des colonnes.

Le caractère séparateur doit être unique. Il n'est pas possible d'en mettre deux ou trois, ou une chaîne de séparateurs. Pour éliminer les caractères multiples,

utilisez tr. De même le séparateur par défaut est la tabulation. Or par défaut les tabulations sont souvent remplacées par des espaces au sein des éditeurs...

cut -dc -fChamps [fic1...]

Voici quelques exemples. Le fichier liste contient des champs séparés par des tabulations.

```
$ cat liste
Produit prix    quantites
souris  30      15
dur     100     30
disque  100     30
ecran   300     20
clavier 45      30
carte   45      30

$ cut -f1 liste
Produit
souris
dur
disque
ecran
clavier
carte

$ cut -f1,3 liste
Produit quantites
souris  15
dur     30
disque  30
ecran   20
clavier 30
carte   30
```

Notez que si vous inversez l'ordre des champs (-f3,1) vous n'obtenez pas l'effet escompté : les champs sortent toujours dans le sens 1,3.

S'il n'y a pas de délimiteur (tabulation ou autre) dans une ligne, cut affiche toute la ligne.

4. Décompte de lignes

La commande **wc** (word count) permet de compter les lignes, les mots et les caractères.

wc [-l] [-c] [-w] [-w] fic1

- -l : compte le nombre de lignes.
- -c : compte le nombre d'octets.
- -w : compte le nombre de mots.

- -m : compte le nombre de caractères.

```
$ wc liste
      12       48      234 liste
```

Le fichier liste contient 12 lignes, 48 mots et 234 caractères(octets).

5. Tri de lignes

La commande **sort** permet de trier des lignes. Par défaut le tri s'effectue sur tout le tableau et en ordre croissant. Le tri est possible sur un ou plusieurs champs. Le séparateur de champs par défaut est la tabulation ou au moins un espace. S'il y a plusieurs espaces, le premier est le séparateur, les autres des caractères du champ.

La syntaxe de sort a évolué depuis quelques années et Linux s'est mis en conformité. Aussi l'ancienne syntaxe basée sur +/- n'est plus utilisée. À la place, il faut utiliser le paramètre -k. La numérotation des champs commence à 1.

```
sort [options] [-k pos1[,pos2]] [fic1...]
```

```
$ cat liste
souris  optique 30     15
dur     30giga  100    30
dur     70giga  150    30
disque  zip     12     30
disque  souple  10     30
ecran   15      150    20
ecran   17      300    20
ecran   19      500    20
clavier 105     45     30
clavier 115     55     30
carte   son     45     30
carte    video   145    30
```

Voici comment trier par ordre alphabétique sur la première colonne :

```
$ sort -k 1 liste
carte   son     45     30
carte   video   145    30
clavier 105     45     30
clavier 115     55     30
disque  souple  10     30
disque  zip     12     30
dur     30giga  100    30
dur     70giga  150    30
ecran   15      150    20
ecran   17      300    20
ecran   19      500    20
souris  optique 30     15
```

Exemple, tri numérique sur le prix par produits en ordre décroissant :

```
$ sort -n -r -k 3 liste
ecran   19      500     20
ecran   17      300     20
ecran   15      150     20
dur     70giga  150     30
carte   video   145     30
dur     30giga  100     30
clavier 115     55      30
clavier 105     45      30
carte   son     45      30
souris  optique 30      15
disque  zip     12      30
disque  souple  10      30
```

Il est aussi possible de démarrer le tri à partir d'un certain caractère d'un champ. Pour cela vous devez spécifier le « .pos » : -k1.3 commencera le tri à partir du troisième caractère du champ 1.

Quelques paramètres

Option	Rôle
-d	Dictionnary sort (tri dictionnaire). Ne prend comme critère de tri que les lettres les chiffres et les espaces.
-n	Tri numérique, idéal pour les colonnes de chiffres.
-b	Ignore les espaces en début de champ.
-f	Pas de différences entre majuscules et minuscules (conversion en minuscules puis tri).
-r	Reverse, tri en ordre décroissant.
-tc	Nouveau délimiteur de champ c.

6. Suppression des doublons

La commande **uniq** permet de supprimer les doublons dans des flux en entrée ou des fichiers triés.

```
$ cat fic.txt
Monsieur François 1 rue de la poche 34000 Montpellier
Madame Louise 456 avenue de Maurin 34080 Montpellier
Madame Françoise 47 rue des pinsons 34080 Montpellier
Monsieur Fernand 642 avenue Louis Ravas 34090 Montpellier
Madame Louise 456 avenue de Maurin 34080 Montpellier

$ cat fic.txt | uniq

Monsieur François 1 rue de la poche 34000 Montpellier
Madame Louise 456 avenue de Maurin 34080 Montpellier
Madame Françoise 47 rue des pinsons 34080 Montpellier
Monsieur Fernand 642 avenue Louis Ravas 34090 Montpellier

$ cat fic.txt | uniq -d

Madame Louise 456 avenue de Maurin 34080 Montpellier
```

7. Jointure de deux fichiers

a. Ligne à ligne

La commande **paste** regroupe n fichiers en un. Pour cela elle concatène les lignes de chacun des fichiers en une seule ligne : ligne1 de fic1 avec ligne1 de fic2, ligne1 de fic 3, et ainsi de suite. C'est un peu l'inverse du cut. Le séparateur par défaut est la tabulation mais vous pouvez préciser un délimiteur avec -d.

```
$ cat fic1
liste_a
liste_b
liste_c

$ cat fic2
liste_a2
liste_b2
liste_c2

$ paste -d: fic1 fic2
liste_a:liste_a2
liste_b:liste_b2
liste_c:liste_c2
```

8. Découpage d’un fichier en morceaux

a. Découper

Voici une commande fort pratique, **split**, qui permet de découper un gros fichier en plusieurs morceaux d’une taille donnée. Les systèmes de fichiers ne sont pas tous égaux devant la taille maximale d’un fichier. Sous Linux le problème se pose peu, un système de fichiers de type ext4 supportant de fichiers de plusieurs dizaines de To. Mais les disques amovibles n’ont pas tous cette possibilité.

Une clé USB ou un disque externe sont généralement « formatés » avec un système de fichiers de type VFAT issu du monde Microsoft. Ce système de fichiers provenant de DOS puis Windows 9x garantit une compatibilité entre tous les systèmes (Unix, Windows, Mac OS), qui peut le plus peut le moins. VFAT (ou plutôt FAT16 ou FAT32) ne supporte que des fichiers d’une taille maximum de 4 Go. Une image ISO de DVD ou une archive de sauvegarde ou un fichier de données nosql de log ne peut y rentrer d’un seul bloc. Il faut donc découper le fichier en plusieurs morceaux.

split [-l n] [-b n[bkm] [fichier [préfixe]]

La commande peut fonctionner selon deux modes :

- découpage par lignes avec -l : les fichiers en sortie auront tous n lignes de texte (sauf éventuellement le dernier).
- découpage à taille fixe avec -b : les fichiers auront tous une taille fixe de n octets. Le suffixe b indique une taille de n blocs (512 octets), k indique n ko (1024 octets) et m indique n Mo (1024 ko).

Comme tout filtre **split** peut prendre un flux en entrée, ce qui est le cas si aucun fichier n’est précisé, ou si un tiret est présent. Un préfixe définit le nom des fichiers en sortie. Voici un fichier de 1 Go à découper en tranches de 150 Mo. Le préfixe est fic. Chaque fichier en sortie s’appelle ficaa, ficab, ficac, ficad, et ainsi de suite.

```
$ ls -l grosfichier
-rw-r--r-- 1 driss users 1073741824 mar 12 19:47 grosfichier
$ split -b 150m grosfichier fic
$ ls -l fic*
-rw-r--r-- 1 driss users 157286400 mar 12 20:15 ficaa
-rw-r--r-- 1 driss users 157286400 mar 12 20:15 ficab
-rw-r--r-- 1 driss users 157286400 mar 12 20:15 ficac
-rw-r--r-- 1 driss users 157286400 mar 12 20:16 ficad
-rw-r--r-- 1 driss users 157286400 mar 12 20:16 ficae
-rw-r--r-- 1 driss users 157286400 mar 12 20:16 ficaf
-rw-r--r-- 1 driss users 130023424 mar 12 20:16 ficag
```

b. Reconstruire

Une ligne suffit pour reconstruire un fichier splité à l'aide des redirections :

```
$ cat fic* > newfic
$ ls -l newfic
-rw-r--r-- 1 driss users 1073741824 mar 12 20:47 newfic
```

9. Remplacement de caractères

a. Liste de caractères

La commande **tr** permet de substituer des caractères à d'autres et n'accepte que des données provenant du canal d'entrée standard, pas les fichiers.

```
tr [options] original destination
```

L'original et la destination représentent un ou plusieurs caractères. Les caractères originaux sont remplacés par les caractères de destination dans l'ordre indiqué. Les crochets permettent de définir des plages.

Par exemple, remplacer le o par le e et le i par le a.

```
$ cat liste | tr "oi" "ea"
Preduat ebjet   prax    quantates
seuras  eptaque 30     15
dur     30gaga  100    30
dur     70gaga  150    30
dasque  zap     12     30
dasque  seuple  10     30
ecran   15      150    20
ecran   17      300    20
ecran   19      500    20
clavaer 105     45     30
clavaer 115     55     30
carte   sen     45     30
carte   vadee   145    30
```

Avec cette commande vous pouvez convertir une chaîne en majuscules ou en minuscules.

```
$ cat liste | tr "[a-z]" "[A-Z]"
PRODUIT OBJET   PRIX    QUANTITES
SOURIS  OPTIQUE 30     15
DUR     30GIGA  100    30
DUR     70GIGA  150    30
DISQUE  ZIP     12     30
DISQUE  SOUPLE  10     30
ECRAN   15      150    20
ECRAN   17      300    20
ECRAN   19      500    20
CLAVIER 105     45     30
CLAVIER 115     55     30
CARTE   SON     45     30
CARTE   VIDEO   145    30
```

Supprimer les répétitions

Surtout, tr admet deux paramètres, -s (squeeze) et -d (delete), qui permettent de supprimer des caractères en doublons ou non. C'est parfait dans le cas de séparateurs multiples. Voici un exemple pratique où l'on cherche à isoler l'adresse IP d'une machine.

```
$ /sbin/ifconfig eth0
eth0      Lien encap:Ethernet  HWaddr 00:13:D3:D7:A4:6C
          inet adr:10.9.238.170  Bcast:10.9.239.255  asque:255.255.252.0
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:15054381 errors:0 dropped:0 overruns:0 frame:0
          TX packets:4991811 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 lg file transmission:1000
          RX bytes:4157389034 (3964.7 Mb)  TX bytes:374974072 (357.6 Mb)
          Interruption:22 Adresse de base:0xcc00
```

Seule la deuxième ligne, contenant inet, nous intéresse :

```
$ /sbin/ifconfig eth0| grep "inet "
          inet adr:10.9.238.170  Bcast:10.9.239.255  Masque:255.255.252.0
```

Pour isoler l'adresse IP située après « inet adr: » le séparateur « : » semble intéressant mais dans ce cas un cut retournerait « 10.9.238.170 Bcast » ce qui ne convient pas. L'astuce consiste à remplacer tous les espaces par un seul « : ». Le paramètre -s remplace une chaîne de n caractères identiques par un seul. S'il n'est pas précisé c'est le même caractère, sinon un caractère de substitution donné.

```
$ /sbin/ifconfig eth0| grep "inet " | tr -s " " ":"
:inet:adr:10.9.238.170:Bcast:10.9.239.255:Masque:255.255.252.0
```

Il n'y a plus qu'à compter : l'adresse IP est en quatrième position (le premier champ avant le premier « : » est vide).

```
$ /sbin/ifconfig eth0| grep "inet " | tr -s " " ":" | cut -d: -f4
10.9.238.170
```

b. Tabulations et espaces

La plupart des éditeurs remplacent les tabulations par des espaces. Or certaines commandes s'attendent à obtenir des tabulations comme délimiteurs de champs (comme cut). S'il est possible de s'en sortir avec tr, deux commandes sont à votre disposition pour ce cas spécifique.

La commande **expand** convertit les tabulations en espaces. La commande **unexpand** convertit les espaces en tabulations. Soit le fichier liste selon le modèle précédent où les colonnes sont séparées par des espaces au lieu de tabulations. Dans le premier cas le résultat n'est pas du tout celui attendu. La commande **cut** tente de sortir le troisième champ d'un fichier tabulé. Comme il

n’y a pas de tabulations, il affiche toute la ligne.

```
$ cut -f1 liste
Produit     objet       prix    quantites
souris      optique     30      15
dur         30giga      100     30
dur         70giga      150     30
disque      zip         12      30
disque      souple      10      30
...
```

La commande **unexpand** avec le paramètre -a remplace toutes les séquences d’au moins deux espaces par le nombre nécessaire de tabulations. Cette fois le résultat est correct.

```
$ unexpand -a liste | cut -f1
Produit
souris
dur
dur
disque
disque
...
```

10. xargs

La commande **xargs** permet de lire des éléments en entrée standard (pipe, redirection), délimités par défaut par un espace ou un retour à la ligne, puis exécute une commande, par défaut **echo**, avec ces mêmes éléments, un par un ou reformatés.

Voici quelques exemples de manipulations possibles avec un fichier nommé input contenant une liste de lignes, de 1 à 20.

```
$ cat input
1
2
...
20
```

Toutes les lignes et mots regroupés :

```
$ cat input | xargs
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

En colonnes de deux :

```
$ cat input | xargs -n 2
1 2
3 4
5 6
7 8
9 10
11 12
...
```

Tous les utilisateurs de /etc/passwd dans une seule ligne, par ordre croissant :

```
$ cut -d: -f1 < /etc/passwd | sort | xargs echo
abrt adm avahi avahi-autoipd bin chrony colord daemon dbus dockerroot ftp
games gdm geoclue gnome-initial-setup hacluster halt haproxy libstoragemg
...
```

Supprimer tous les fichiers trouvés par une commande **find** :

```
$ find -name "core*" -size +1m -print | xargs rm -f
```

11. Visualisation de texte

a. En pleine page

Rien n'empêche de détourner un quelconque flux pour l'afficher sur l'écran ou l'imprimante. Voici quelques commandes.

- page par page : **pg**, **more**, **less**
- en bloc : **cat**
- à l'envers : **tac**
- en dump hexadécimal : **hexdump**
- en dump octal (par défaut) : **od**
- création d'une bannière : **banner**
- formatage pour impression : **pr**
- formatage de paragraphe : **fmt**
- numéroter les lignes : **cat -n** ou **nl**

Voici quelques exemples en utilisant toujours le même fichier input :

Sortie en hexadécimal :

```
$ hexdump input
0000000 0a31 0a32 0a33 0a34 0a35 0a36 0a37 0a38
0000010 0a39 3031 310a 0a31 3231 310a 0a33 3431
0000020 310a 0a35 3631 310a 0a37 3831 310a 0a39
0000030 3032 000a
0000033
$ od -Ax -x input
000000 0a31 0a32 0a33 0a34 0a35 0a36 0a37 0a38
000010 0a39 3031 310a 0a31 3231 310a 0a33 3431
000020 310a 0a35 3631 310a 0a37 3831 310a 0a39
000030 3032 000a
000033
```

Sortie en octal :

```
$ od input
0000000 005061 005062 005063 005064 005065 005066 005067 005070
0000020 005071 030061 030412 005061 031061 030412 005063 032061
0000040 030412 005065 033061 030412 005067 034061 030412 005071
0000060 030062 000012
0000063
```

Numérotation des lignes avec remplissage de zéros :

```
$ nl -nrz input
000001    1
000002    2
000003    3
000004    4
000005    5
000006    6
```

Soit le fichier suivant :

```
$ cat lorem.txt

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod
tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam,
quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo
consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse
cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat
non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.
```

La commande **fmt** va reformater le ou les paragraphes pour plus de lisibilité (ou pour l'impression). Par exemple, sur des colonnes de 60 caractères :

```
$ fmt -t -w 60  lorem.txt
Lorem ipsum dolor sit amet, consectetur adipiscing
   elit, sed do eiusmod tempor incididunt ut labore et
   dolore magna aliqua. Ut enim ad minim veniam, quis
   nostrud exercitation ullamco laboris nisi ut aliquip
   ex ea commodo consequat. Duis aute irure dolor in
   reprehenderit in voluptate velit esse cillum dolore eu
   fugiat nulla pariatur. Excepteur sint occaecat cupidatat
   non proident, sunt in culpa qui officia deserunt mollit
   anim id est laborum.
```

b. Début d'un fichier

Pour voir le début d'un fichier, utilisez la commande **head**.

```
head [-c nbcar] [-n nblignes] [fic1...]
```

Le paramètre -c permet de préciser un nombre d'octets d'en-tête à afficher. Par défaut dix lignes sont affichées. Le paramètre -n permet d'indiquer le nombre de lignes à afficher. Vous pouvez indiquer directement le nombre de lignes :

```
head [-nblignes] [fic1...]
```

```
$ head -3 liste
Produit objet   prix    quantites
```

```
souris  optique 30      15
dur     30giga  100     30
```

c. Fin et attente de fichier

Pour voir les dernières lignes d'un fichier, utilisez la commande **tail**.

```
tail [+/-valeur[b/c]] [-f] [fic1...]
```

Comme pour head, par défaut les dix dernières lignes sont affichées. La valeur -nblignes permet de modifier cet état. Précisez c pour indiquer un nombre de caractères. Un b indique un nombre de blocs (512 octets par bloc).

Enfin l'option -f laisse le fichier ouvert. Ainsi si le fichier continue d'être rempli (par exemple un fichier trace), son contenu s'affichera en continu sur l'écran jusqu'à interruption volontaire par l'utilisateur ([Ctrl] **C**).

```
$ tail -5 liste
ecran   19      500     20
clavier 105     45      30
clavier 115     55      30
carte   son     45      30
carte   video   145     30

$ tail -10c liste
eo      145     30
```

d. Formater une sortie

La commande **column** permet de formater une sortie sous forme de table. L'option -t détermine combien de colonnes contiennent la sortie et ajoute des espaces afin de les aligner. L'option -s permet d'indiquer quel est le séparateur.

```
driss@slyserver:~$ column -s: -t /etc/group
root            x  0
daemon          x  1
bin             x  2
sys             x  3
adm             x  4     driss,steph
tty             x  5
disk            x  6
```

12. Duplication du canal de sortie standard

Dans certains cas, comme par exemple la génération de fichiers traces (log), il peut être nécessaire de devoir à la fois placer dans un fichier le résultat d'une commande et de filtrer ce même résultat avec une autre commande. Utilisez pour cela la commande **tee** qui permet de dupliquer le flux de données. Elle lit le flux de données provenant d'une autre commande par le canal d'entrée, l'écrit dans un fichier et restitue ce flux à l'identique par le canal de sortie. Par défaut le fichier généré écrase l'ancien s'il existe.

```
tee [-a] nom_fic
```

Le paramètre -a signifie append. *Dans ce cas le fichier n'est pas écrasé mais complété à la fin*. Par exemple, vous voulez obtenir à la fois dans un fichier la liste des noms d'utilisateurs et afficher leur nombre sur écran.

```
$ cat /etc/passwd | cut -d: -f1 | tee users | wc -l
       65
$ cat users
root
nobody
nobodyV
daemon
bin
uucp
uucpa
auth
cron
lp
tcb
...
```

13. Comparaison de fichiers

Les deux commandes permettant de comparer le contenu de deux fichiers, ou d'un fichier et d'un flux sont les commandes **diff** et **cmp**.

a. diff

La commande **diff** indique les modifications à apporter aux deux fichiers en entrée pour que leur contenu soit identique.

```
diff [-b] [-e] fic1 fic2
```

L'option -b permet d'ignorer les espaces (blank), et l'option -e permet de générer un script ed (nous ne l'utiliserons pas). Cette commande renvoie trois types de messages :

- APPEND : ex 5 a 6,8 veut dire : à la ligne 5 de fic1 il faut raccrocher les lignes 6 à 8 de fic2 pour que leurs contenus soient identiques.
- DELETE : ex 7,9 d 6 veut dire : les lignes 7 à 9 de fic1 doivent être supprimées, elles n'existent pas derrière la ligne 6 de fic2.
- CHANGE : ex 8,12 c 9,13 veut dire : les lignes 8 à 12 de fic1 doivent être échangées contre les lignes 9 à 13 de fic2.

Dans tous les cas, le signe "<" indique les lignes de fic1 concernées, et le signe ">" les lignes de fic2 concernées.

```
$ cat liste
Produit objet   prix    quantites
```

```
souris  optique 30      15
dur     30giga  100     30
dur     70giga  150     30
disque  zip     12      30
disque  souple  10      30
ecran   15      150     20
ecran   17      300     20
ecran   19      500     20
clavier 105     45      30
clavier 115     55      30
carte   son     45      30
carte   video   145     30

$ cat liste2
Produit objet   prix    quantites
souris  boutons 30      15
dur     30giga  100     30
dur     70giga  150     30
disque  zip     12      30
disque  souple  10      30
ecran   15      150     20
ecran   17      300     20
ecran   19      500     20
ecran   21      500     20
clavier 105     45      30
clavier 115     55      30
```

Le fichier liste est l'original. Dans liste2, la deuxième ligne a été modifiée, une ligne écran a été ajoutée et les deux dernières lignes ont été supprimées.

```
$ diff liste liste2
2c2
< souris        optique 30      15
---
> souris        boutons 30      15
9a10
> ecran 21      500     20
12,13d12
< carte son     45      30
< carte video   145     30
```

- 2c2 : les lignes 2 de liste et liste2 doivent être échangées (elles doivent concorder soit en optique, soit en boutons).
- 9a10 : après la ligne 9 de liste (écran 19) il faut ajouter la ligne 10 (écran 21) de liste2.
- 12,13d12 : les lignes 12 et 13 de liste (carte son et vidéo) doivent être supprimés car elles n'existent pas après la ligne 12 de liste2.

b. cmp

La commande **cmp** compare les fichiers caractère par caractère. Par défaut la commande s'arrête dès la première différence rencontrée et indique la position de l'erreur.

cmp [-l] [-s] fic1 fic2

Le paramètre -l détaille toutes les différences en trois colonnes. La première colonne représente le numéro de caractère, la deuxième la valeur octale ASCII du caractère concerné de fic1 et la troisième la valeur octale ASCII du caractère concerné de fic2.

L'option -s retourne uniquement le code d'erreur (non visible), accessible par echo $?.

```
$ cmp liste liste2
liste liste2 differ: char 38, line 2
$ cmp -l liste liste2
   38 157 142
   39 160 157
   40 164 165
   41 151 164
   42 161 157
   43 165 156
   44 145 163
  182 143 145
  183 154 143
  ...
```

14. Délai d'attente

La commande **sleep** permet d'attendre le nombre de secondes indiqués. Le script est interrompu durant ce temps. Le nombre de secondes et un entier compris entre 0 et 4 milliards (136 ans).

```
$ sleep 10
```

15. Contrôler le flux

La commande **pv**, souvent méconnue, permet de répondre à une question fréquemment posée : que se passe-t-il dans le tube lorsque les données passent d'un processus à un autre : combien de données, combien de temps cela va-t-il mettre ? pv est un moniteur de flux. Il s'intercale généralement entre une commande et une autre (comme tee), et analyse le flux qu'il reçoit avant de le renvoyer à sa destination. De ce fait, il sait ce qui transite, et peut afficher une barre de progression, par exemple. Voici deux exemples :

Copier un fichier

```
[driss@localhost ~]$ pv test > test2
50MiO 0:00:02 [24,1MiB/s] [================>              ] 52% ETA 0:00:01
```

Durée de compression d'un fichier

```
pv /boot/vmlinuz-3.14.7-200.fc20.x86_64 | gzip > test.gz
5,26MiO 0:00:00 [6,05MiB/s] [==================================>] 100%
```

TRAVAUX PRATIQUES

1.	Le fichier /etc/passwd est un grand classique sous Unix. Il se compose de sept champs séparés par des « : » : login:passwd:UID:GID:Commentaire:homedir:shell. Récupérez la ligne de l'utilisateur root dans /etc/passwd : $ grep ^root: /etc/passwd
2.	De cette ligne, récupérez l'UID de root : $ grep ^root: /etc/passwd | cut -d: -f3
3.	Comptez le nombre d'utilisateurs contenus dans ce fichier à l'aide d'une redirection en entrée : $ wc -l < /etc/passwd
4.	Un peu plus compliqué : récupérez la liste des GID, triez-les par ordre croissant et supprimez les doublons : $ cut -d: -f4 /etc/passwd | sort -n | uniq
5.	De là, extrapolez le nombre de groupes différents utilisés : $ cut -d: -f4 /etc/passwd | sort -n | uniq | wc -l
6.	Convertissez tous les logins en majuscules : $ cut -d: -f1 /etc/passwd | tr "[a-z]" "[A-Z]"

7.	Isolez maintenant la huitième ligne de /etc/passwd. Il y a plusieurs solutions, en voici deux :
	
	$ head -8 /etc/passwd | tail -1
	
	Et :
	
	$ grep -n "" /etc/passwd | grep ^8: | cut -d: -f2-

EXERCICES

1 - Recherchez dans /etc/passwd la ligne correspondant à votre utilisateur, et seulement celui-ci.

2 - Recherchez dans /etc/passwd les utilisateurs n’ayant pas bash comme shell par défaut. Utilisez un tube (pipe).

3 - À l’aide d’une simple expression régulière, supprimez les lignes vides et les commentaires (commençant par #) de /etc/hosts.

4 - Affichez uniquement la liste des logins et UID associés (champ 3) de /etc/passwd.

5 - Quelle suite de commande permet de connaître le nombre de lignes réel (non vides et pas les commentaires) de /etc/hosts ? Utilisez deux fois les tubes et numérotez les lignes.

6 - Triez le fichier /etc/group par ordre de GID (troisième colonne) décroissant.

7 - Donnez la liste des différents shells utilisés par les utilisateurs déclarés sur votre machine. Il ne doit pas y avoir de doublons.

8 - Dans un fichier texte de votre choix (appelé ici fic) remplacez tous les é, è, ê, ë par un e (attention cette question ne fonctionne pas en unicode).

9 - Comment récupérer la 27e ligne d’un fichier fic (partant du principe que fic en contient plus de 27) ?

- A - cat -n fic | grep ^27 | cut -d" " -f2-
- B - grep -n ”” fic | grep ^27 | cut -d: -f2-
- C - tail -27 fic | head -1
- D - head -27 fic | tail -1

10 - Soit la commande

1>liste 2>&1

Pour lister l’intégralité des fichiers du système et placer tous les résultats quels qu’ils soient dans liste, vous redirigez le canal d’erreur dans le canal de sortie standard, puis la sortie standard dans liste, de droite à gauche, le 1 étant optionnel. Mais cette fois vous devez faire en sorte que toute la liste soit placée dans un fichier ET affichée.