OPTIMISER SON UTILISATION D'UNIX

BERNARD TATIN

BERNARD.TATIN@OUTLOOK.FR

RÉSUMÉ. Ce document vient des tréfonds de l'espace temps. Il a débuté il y a bien plus de trois ans de cela, repris de manière plus systématique et se trouve fortement complété aujourd'hui.

La première partie rappelle (rapidement) l'histoire et les concepts principaux des SHELLS. La deuxième partie est très orientée sur la recherche de qui~a~piraté~ma~machine mais peut être d'une grande utilité pour les débutants. La troisième partie, quant à elle, se focalise sur les scripts. Une quatrième partie donnera des notions des outils indispensables pour utiliser correctement son système \mathbf{Unix} .

Ce document et ses sources en LATEX sont disponibles sur GitHub.

Première partie 1. l'histoire et les concepts	6
1. Une histoire d' Unix	2
2. Les shells	5
2.1. Le fonctionnement	4
2.2. Quelques shells célèbres	ϵ
2.2.1. sh, le Bourne shell	(
2.2.2. csh , le C shell	(
2.2.3. tcsh ou le csh interactif	(
2.2.4. ksh, le Korn Shell	7
2.2.5. zsh , le Z shell	7
2.2.6. bash, Bourne Again shell	7
Deuxième partie 2. la configuration et la ligne de commande 3. La configuration	7
3.1. Le shell personnel	7
3.2. Configurer le prompt	8
4. la ligne de commande	12
4.1. les boucles	12
4.1.1. la boucle for	12
4.1.2. la boucle while	12
4.2. surprises avec stat, find et xargs	13
4.2.1. stat	13
4.2.2. réfléchissons un peu	14
4.2.3. tous les fichiers du monde	14

Date: Novembre 2013/Novembre 2016.

4.2.4. application pratique	15
4.3. les surprises de sudo	17
4.4. POSIX et GNU	19
Troisième partie 3. les scripts et les exemples	19
5. Les scripts shell	19
5.1. structure des scripts	19
5.2. choisir son Shell	20
5.3. les paramètres des scripts	20
5.4. tests et boucles	21
5.5. conditions, valeurs de retour des programmes	21
5.6. redirections	22
5.7. tubes ou <i>pipes</i>	22
6. exemples de manipulation de texte	22
6.1. des stats	22
6.2. peut-on faire mieux?	24
6.2.1. les options	24
6.2.2 avec hash	28

Table des matières

Première partie 1. l'histoire et les concepts

1. Une histoire d'**Unix**

Voici une (rapide) histoire d'**Unix**, choisie parmi d'autres, parmi celles qui évoluent avec le temps autant parce que des personnages hauts en couleur et ayant réussi à voler la vedette à de plus modestes collègues se font effacer eux-même par de plus brillants qu'eux, soit parce que, vieillissant ils se laissent aller à des confidences inattendues.

En nous basant sur Brève histoire d' \mathbf{Unix} , on rappelle que $AT\mathcal{E}T$ travaillait à la fin des années 60, sur un système d'exploitation $\mathbf{Multics}$ qui devait révolutionner l'histoire de l'informatique. Si révolution il y eut, ce fut dans les esprits : de nombreux concepts de ce système ont influencés ses successeurs, dont \mathbf{Unix} . Ken Thompson et Dennis Ritchie des fameux $Bell\ Labs$ et qui travaillaient (sans grande conviction, semble-t-il) sur $\mathbf{Multics}$, décidèrent de lancer leur propre projet d'OS :

baptisé initialement UNICS (UNiplexed Information and Computing Service) jeu de mot avec "eunuchs' (eunuque) pour "un Multics emasculé", par clin d'œil au projet Multics, qu'ils jugeaient beaucoup trop compliqué. Le nom fut ensuite modifié en Unix ¹.

L'essor d'Unix est très fortement lié à un langage de programmation, le C. À l'origine, le premier Unix était écrit en assembleur, puis Ken Thompson crée un nouveau langage, le B. En 1971, Dennis Ritchie écrit à son tour un nouveau

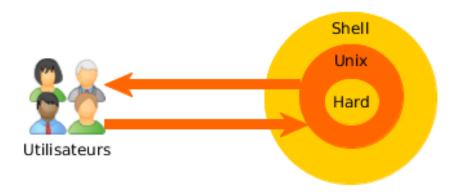
^{1.} cfl'article Multics de Wikipedia

langage, fondé sur le B, le C. Dès 1973, presque tout **Unix** est réécrit en C. Ceci fait probablement d'**Unix** le premier système au monde écrit dans un langage portable, c'est-à-dire autre chose que de l'assembleur².

Ce que j'ai surtout retenu de tout cela, c'est qu'**Unix** a banalisé autant l'utilisation des stations de travail connectées en réseau que le concept de shell, des système de fichiers hiérarchisés, des périphériques considérés comme de simples fichiers, concepts repris (et certainement améliorés) à **Multics** comme à d'autres. Pour moi, la plus grande invention d'**Unix**, c'est le langage C qui permet l'écriture des systèmes d'exploitations et des logiciels d'une manière très portable. N'oublions pas qu'aujourd'hui encore, C (mais pas C++) est un des langages les plus portable, même s'il commence à être concurrencé par Java par exemple.

2. Les shells

Un shell est une *coquille*, pour reprendre la traduction littérale, autour du système d'exploitation. Voici un magnifique diagramme (d'après ce que l'on trouve sur le WEB comme dans d'anciens ouvrages) donnant une idée du concept :



 $shell \ \mathbf{Unix}$ Source: le WEB, ouvrages divers

Entre mes débuts dans le monde de l'informatique et aujourd'hui, le concept de SHELL a quelque peu évolué. Certains qualifient l'explorateur de Windows comme un SHELL. Ontils raison? Certainement si l'on se réfère à l'image précédente : nos commandes (clique, clique et reclique) envoyée au shell graphique sont transmises au noyau qui nous renvoie, par l'intermédiaire du shell graphique, de belles images. Il faut avouer que l'explorateur Windows est le premier contact que l'utilisateur a avec sa machine. Et sur l'article interface système de Wikipedia, on trouve cette définition :

Une interface système (SHELL en anglais) est une couche logicielle qui fournit l'interface utilisateur d'un système d'exploitation. Il correspond à la couche la plus externe de ce dernier.

^{2.} cfBrève histoire d'Unix

Ce même article cite les:

shells graphiques fournissant une interface graphique pour l'utilisateur (GUI, pour Graphical User Interface)

Dans le monde \mathbf{Unix} , le concept de SHELL reste plus modeste, même si $\mathit{Midnight}$ $\mathit{Commander}$ (mc) est parfois considéré comme un SHELL :



mc dans une session Cygwin Source: mon PC

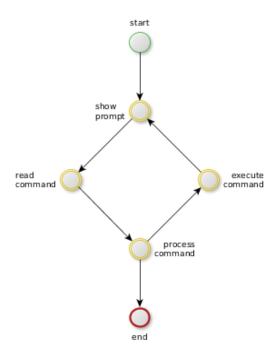
Pour nous et dans tout ce qui suit, nous considérons comme SHELL:

un interpréteur de commandes destiné aux systèmes d'exploitation **Unix** et de type **Unix** qui permet d'accéder aux fonctionnalités internes du système d'exploitation. Il se présente sous la forme d'une interface en ligne de commande accessible depuis la console ou un terminal. L'utilisateur lance des commandes sous forme d'une entrée texte exécutée ensuite par le SHELL. Dans les différents systèmes **Windows**, le programme analogue est command.com ou cmd.exe.

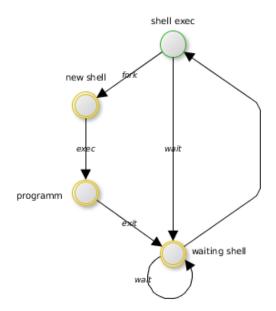
cfSHELL **Unix** sur Wikipedia

2.1. Le fonctionnement. Le fonctionnement général est assez simple, surtout si l'on ne tient pas compte de la gestion des erreurs comme dans le graphique suivant qui peut

être appliqué à tout bon interpréteur. Seuls les détails de process command et execute command vont réellement changer.



SHELL : $fonctionnement\ g\'{e}n\'{e}ral$ Source: créé avec yEd



SHELL : $exécution\ d$ 'un programmeSource: crée avec yEd

À noter que la commande exec se comporte différemment : elle correspond à l'appel système exec .

2.2. Quelques shells célèbres.

- 2.2.1. sh , $le\ Bourne$ SHELL. L'ancêtre, toujours vivant et avec lequel sont écrits une grande majorité des scripts actuels. Son intérêt essentiel est justement l'écriture de scripts. Pour l'interaction, il est absolument nul mais bien utile parfois pour dépanner.
- 2.2.2. csh , le C shell. Il se voulait le remplaçant glorieux de l'ancêtre sh avec une syntaxe considérée plus lisible car proche du C. Il est de plus en plus abandonné y compris par ses admirateurs les plus fervents, vieillissants dans la solitude la plus complète. Essayez d'écrire un script en csh d'un peu d'envergure sans faire de copié/collé! Il n'y a en effet pas de possibilité de créer des fonctions et, ce qui gêne peut-être encore plus les administrateurs système, il n'y a pas de gestion d'exception. Cependant, il fût certainement le premier à proposer l'historique des commandes.

A noter qu'il fût crée par Bill Joy, l'un des fondateurs historiques de la société Sun Microsystems.

2.2.3. tcsh ou le csh interactif. Le pendant interactif du précédent. Il lui reste des afficionados qui aiment bien sa gestion de l'historique et de la ligne de commande. Il est une extension de csh, i.e.tout ce qui peut-être fait par csh est fait par tcsh. Sur de nombreux systèmes (Mac OS X entre autre), ces deux SHELLs pointent sur le même exécutable (avec un lien symbolique).

En séquence *nostalgie*, je me souviens que c'est ce SHELL interactif que j'utilisais sur mon premier \mathbf{Unix} , en 87/88.

 $2.2.4.~{\tt ksh}$, $le~{\tt Korn}$ shell. Initialement écrit pour ${\tt Unix}~{\tt par}$ David Korn au début des années 80, ce shell a été repris par Microsoft pour Windows. Compatible avec ${\tt sh}$, il propose de nombreuses avancées comme beaucoup de fonctionnalités de ${\tt tcsh}$, des fonctions, des exceptions, des manipulations très évoluées de chaînes de caractères,

2.2.5. ${\tt zsh}$, ${\tt le}$ Z SHELL. C'est mon préféré pour l'interactivité, la complétion et bien d'autres choses encore dont il est capable depuis sa création ou presque. Comme ${\tt ksh}$, il est compilable en bytecode et propose des bibliothèques thématiques comme la couleur, les sockets, la gestion des dates...

2.2.6. bash , $Bourne\ Again\ {\tt SHELL}$. C'est le descendant le plus direct de ${\tt sh}$. C'est certainement le ${\tt SHELL}$ le plus répandu dans le monde Linux aujourd'hui.

Lors de ma découverte de Linux, je l'ai vite abandonné car il était très en retard pour la complétion en ligne de commande par rapport à d'autres, y compris tcsh qui commençait pourtant à vieillir un peu. Il a fallu beaucoup d'années (pratiquement 10) pour qu'il en vienne à peu près au niveau de zsh.

Aujourd'hui, c'est le SHELL par défaut de nombreuses distributions Linux et il commence à devenir très utilisé comme SHELL de script par défaut.

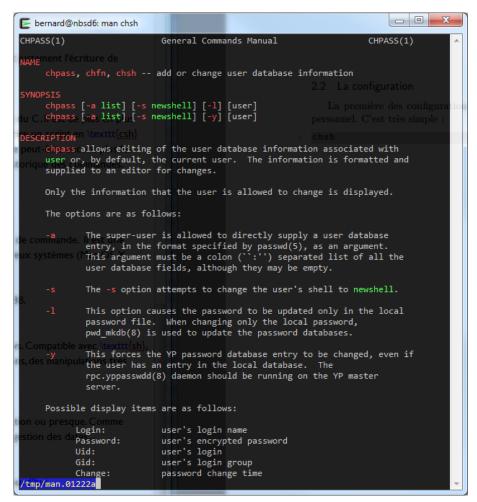
Deuxième partie 2. la configuration et la ligne de commande

3. La configuration

3.1. Le shell personnel. La première des configuration est le choix de son shell par défaut sur son compte personnel. C'est très simple :

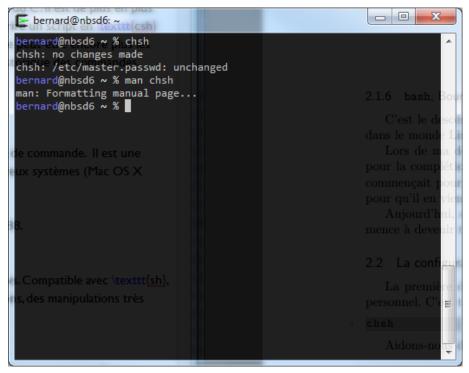
chsh

Aidons-nous du manuel (sous **NetBSD**):



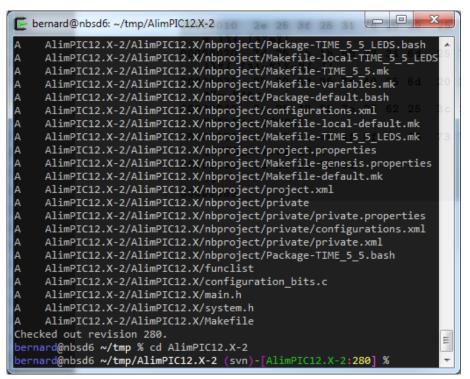
man chsh sous NetBSD Source: ma machine virtuelle

3.2. Configurer le prompt. Sur ma machine virtuelle NetBSD , j'obtiens quelque chose comme ceci :

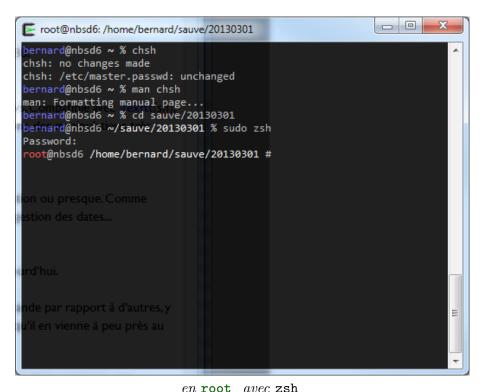


 $un\ prompt\ sous\ \mathbf{NetBSD}$, $avec\ \mathtt{zsh}$ Source: ma machine virtuelle

Le prompt, ce sont les caractères colorés que l'on voit en début de chaque lignes de commande. Ce prompt m'a aidé, voire sauvé plusieurs fois. Celui-ci m'affiche le nom de l'utilisateur courant en bleu, de la machine en blanc et du répertoire courant en blanc et gras. Lorsque j'ai des sessions sur plusieurs machines, je vois tout de suite où je me trouve avec son nom. Ensuite, lorsque je me déplace de répertoires en répertoires, je n'ai pas besoin de faire d'éternels <code>pwd</code> pour savoir où je me trouve. En plus, lorsque je trouve dans un dépôt SVN, j'ai un affichage me donnant les indications sur le répertoire de travail (on ne peut pas le faire sous <code>Cygwin</code>):



dans un répertoire de travail \mathbf{SVN} , avec \mathtt{zsh} Source: ma machine



en root avec zsn Source: ma machine

Tous les shells interactifs de ma connaissance ont au moins un fichier de configuration exécuté au lancement : avec zsh, c'est .zhrc, avec bash, c'est .bashrc et avec csh, c'est .cshrc. Aussi loin que mes souvenirs remontent, on personnalise le prompt avec la variable PS1 et ce, même pour le MS/DOS.

Voici un hexdump de mon PS1 :

```
00000000 25 7b 1b 5b 30 31 3b 33 31 6d 25 7d 25 28 3f 2e
      |%{.[01;31m%}%(?.|
  00000010
           2e 25 3f 25 31 76 20 29
                                      25 7b 1b 5b 33 37 6d 25
                                                                1.%?%1
     v )%{.[37m%|
  00000020 7d 25 7b 1b 5b 33 34 6d
                                      25 7d 25 6e 25 7b 1b 5b
      |}%{.[34m%}%n%{.[|
  00000030 30 30 6d 25 7d 40 25 6d
                                      20 25 34 30 3c 2e 2e 2e
                                                                100m%}
     @%m %40<...|
  00000040
           3c 25 42 25 7e 25 62 25
                                      3c 3c 20 24 7b 56 43 53
                                                                | <%B
5
     %~%b%<< ${VCS|
  00000050
           5f 49 4e 46 4f 5f 6d 65
                                      73 73 61 67 65 5f 30 5f
     _INFO_message_0_|
  00000060 7d 25 23 20 0a
                                                                |}%#
     . /
  00000065
```

4. LA LIGNE DE COMMANDE

Pour de multiples raisons déjà plus ou moins évoquées plus haut, j'ai choisi de travailler avec zsh comme shell par défaut. C'est ce que nous allons faire ici, autant sous **FreeBSD** que sous **Linux**, tout simplement pour prouver que l'utilisation du shell est assez indépendante du système sous-jacent. Mais comme je sais que certains systèmes viennent avec bash ou tcsh, sans possibilité de modification, je les évoqueraient donc, en particulier tcsh qui est utilisé très souvent, avec csh, pour l'administration. Ce n'est que fortuitement que j'examinerais ksh, autant par manque d'habitude que parce que je ne l'ai jamais rencontré.

4.1. les boucles. Il y a while et for .

 $4.1.1.\ la\ boucle$ for . On commence par celle-ci car elle en a dérouté plus d'un. Nous avons, avec ${\it zsh}\$ et ${\it bash}\$, deux syntaxes essentielles. La première ${\it parcourt}$ un ensemble de données :

```
for file_name in *.txt
do
cat $file_name
done
```

Il y a la boucle plus classique pour les spécialistes de Java :

```
1 for ((i=5; i < 8; i++))
2 do
3     echo $i
4 done</pre>
```

Avec tcsh, nous aurons:

```
foreach file_name (*.txt)
cat $file_name

end

foreach i ('seq 5 1 8')
echo $i
end
```

4.1.2. la boucle while . Elle permet de boucles infinies comme celle-ci avec zsh , bash et ksh :

```
while true; do date ''+%T''; sleep 1; done
```

Avec tcsh, nous écrirons en deux lignes :

```
while (1); date ''+%T''; sleep 1;
end
```

4.2. surprises avec stat, find et xargs.

4.2.1. **stat** . La commande **stat** permet de connaître bon nombre de détails à propos d'un fichier comme ici :

```
bernard@debian7 ~ % stat
   install:
   device
             70
3
             1837064
   inode
   mode
             16877
   nlink
             3
6
   uid
             1000
             1000
   gid
   rdev
             7337007
9
             512
   size
             1383862259
   atime
11
             1383085660
   mtime
12
   ctime
             1383085660
13
   blksize 16384
14
   blocks
15
   link
16
17
   userstart.tar.gz:
18
             70
   device
19
   inode
             1837156
20
             33188
   mode
21
   nlink
22
             1000
23
   uid
             1000
   gid
24
             7371584
   rdev
25
26
   size
             3498854
             1383855566
   atime
27
28
   mtime
             1383855552
             1383855552
   ctime
29
   blksize 16384
30
             6880
   blocks
31
   link
32
```

On obtient, sous zsh, un résultat totalement identique sous NetBSD et sous Linux. Si l'on fait un $which \ stat$, nous obtenons, sur les deux systèmes, le message stat: $shell \ built-in \ command$. C'est ce qui me plait sous zsh, les commandes non standard comme stat sont remplacées par des fonctions dont le résultat ne réserve pas de surprise. Si je veux une sortie plus agréable et n'afficher que la date de dernière modification $(cfThe \ zsh/stat \ module \ pour \ de \ plus \ amples \ explications)$:

```
3  userstart.tar.gz 2013-11-07 21:19:12
4  bernard@debian7 ~ %

1  bernard@NBSD-64bits ~ % stat -F "%Y-%m-%d_\%T" +mtime -n *
2  install 2013-11-14 09:52:56
3  userstart.tar.gz 2013-11-08 00:54:06
4  bernard@NBSD-64bits ~ %
```

4.2.2. réfléchissons un peu. Grâce à zsh, nous avons une méthode extrêmement portable entre Unix pour afficher des données détaillées des fichiers. Pour l'exemple, prenons le stat d'origine :

4.2.3. tous *les fichiers du monde*. Si je veux faire la même chose que précédemment, mais sur tous les fichiers de la machine, on peut tomber sur ce message d'erreur :

C'est là que xargs entre en jeu, mais avec un nouveau problème :

La commande xargs va chercher non pas la fonction de zsh mais le binaire qui se trouve sur le PATH de la machine. On doit donc faire :

Sous NetBSD:

4.2.4. application pratique. Sur le serveur, qui est sous **Linux**, faisons la même chose ou presque, on place la date en premier et c'est la surprise du jour :

En rajoutant l'option -print0 à find , l'option -0 à xargs , nous obtenons le bon résultat :

```
9 2013-11-13 17:16:04.377217037 +0100 /etc/apache2/sites-available

10 2013-11-13 17:16:04.377217037 +0100 /etc/phpmyadmin

11 2013-11-14 01:03:38.589252417 +0100 /etc/php5/conf.d/mysqli.ini

12 2013-11-14 01:05:14.997350491 +0100 /etc/php5/conf.d

13 2013-11-14 01:05:14.997350491 +0100 /etc/php5/conf.d/mcrypt.ini
```

En fait, les noms de fichier sous **Unix** peuvent contenir beaucoup de caractères étranges en dehors de / . xargs prend le caractère LF comme fin d'enregistrement de la part de son entrée standard. Si jamais un fichier contient ce caractère, plus rien ne va. Les nouvelles options permettent à find d'utiliser 0x00 comme séparateur d'enregistrement et à xargs de bien l'interpréter.

Il y a aussi une autre explication, depuis bien longtemps les outils \mathbf{GNU} fonctionnent comme ceci et ce n'est que très récemment que le couple \mathbf{find} /xargs fonctionne ainsi.

Après toutes ces considérations, on constate que le 14 Novembre 2013, un peu après 1 heure du matin, quelqu'un a modifié les fichiers /etc/php5/conf.d/mysqli.ini et /etc/php5/conf.d/mcrypt.ini, tout simplement pour remplacer les commentaires de type shell par des commentaires de type fichier ini.

Après une attaque du serveur, il est intéressant de faire le même exercice sur les répertoires vitaux comme /bin. Pour éviter des listings trop important, on limite la sortie à l'année 2013 et on fait une jolie boucle :

```
[bigserver] (694) ~ % for d in /bin /sbin /lib /lib32 /usr/bin /
    usr/sbin /usr/lib /usr/lib32; do

find $d -name "*" -print0 | xargs -0 stat --printf="%zu%n\n" |
    egrep "^2013"

done | sort
```

Nous obtenons un listing fort long, correspondant aux mises à jour faites le 8 et le 12 Novembre. Maintenant que nous savons que l'attaque a eu lieu avant le 8 Novembre, on sélectionne plus sévèrement :

```
[bigserver] (695) ~ % for d in /bin /sbin /lib /lib32 /usr/bin /
    usr/sbin /usr/lib /usr/lib32; do

find $d -name "*" -print0 | xargs -0 stat --printf="%zu%n\n" |
    egrep "^2013-11-0[1-7]"

done | sort
[bigserver] (696) ~ %
```

Cependant, rien ne prouve que nous n'avons pas eu de désordres un peu avant ou un peu pendant. Comme le gros des fichiers est dans /usr/lib, éliminons le de la liste :

```
[bigserver] (695) ~ % for d in /bin /sbin /lib /lib32 /usr/bin / usr/sbin; do

find $d -name "*" -print0 | xargs -0 stat --printf="%zu%n\n" | egrep "^2013"
```

```
3 done | sort
4 ...
5 [bigserver] (696) ~ %
```

Nous n'avons des modifications qu'entre le 8 et le 12 Novembre.

Plus fort encore, afficher les fichiers modifiés ce jour :

```
find / -name "*" -print0 | xargs -0 stat --printf="%z_{\perp}%n\n" | egrep "^$(date_\'+%Y-\m'\m')" | sort
```

On est débordé par l'affichage des fichiers système de Linux. Pour palier à cet inconvénient, on demande à find d'abandonner les répertoire /sys et /proc:

```
find / \( -path /proc -o -path /sys \) -prune -o -name "*" -
print0 | xargs -0 stat --printf="%z_%n\n" | egrep "^$(date_'+%Y-%m-%d')" | sort
```

4.3. **les surprises de sudo .** Reprenons l'exemple précédent en redirigeant la sortie standard vers /dev/null :

```
find / \( -path /proc -o -path /sys \) -prune -o -name "*" -
    print0 | xargs -0 stat --printf="%z_\%n\n" | egrep "^$(date_\'+\%Y-\%m-\%d')" > /dev/null
```

On aura une sortie comme celle-ci:

```
find: "/var/lib/postgresql/9.1/main": Permission non accordee

find: "/var/lib/sudo": Permission non accordee

find: "/var/cache/ldconfig": Permission non accordee

find: "/var/log/exim4": Permission non accordee

find: "/var/log/apache2": Permission non accordee

...
```

Pour éliminer les find: ... Permission non accordee, on utilise sudo :

```
sudo find / \( -path /proc -o -path /sys \) -prune -o -name "*" - print0 | xargs -0 stat --printf="%z_{\perp}%n\n" | egrep "^$(date_{\perp}'+% Y-%m-%d')" > /dev/null
```

C'est pire:

Nous avons demandé à **sudo** de traité **find** et avec le *pipe*, nous demandons à **xargs** de traiter les lignes de sorties avec **stat**. Ce dernier récupère un nom de fichier et le traite comme tel mais comme il n'est pas lancé avec **sudo**, nous avons ces erreurs. Essayons ceci :

```
sudo find / \( -path /proc -o -path /sys \) -prune -o -name "*" -
print0 | xargs -0 sudo stat --printf="%zu%n\n" | egrep "^$(
dateu'+%Y-%m-%d')" > /dev/null
```

C'est pas mieux, autant sous Linux que sous NetBSD :

```
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
```

Il faut bien l'avouer, je ne sais pas quoi dire de plus ici - sinon noter un TODO: comprendre ce qui ce passe. Ce qui est bien avec \mathbf{Unix} , c'est qu'il y a toujours un moyen de s'en sortir. On remarquera quelques différences entre les mondes \mathbf{Linux} et \mathbf{BSD} , en particulier avec \mathbf{man} \mathbf{sh} , où le premier nous renvoie sur \mathbf{bash} alors que le second traite bien directement de \mathbf{sh} . Dans tous les cas, \mathbf{sudo} \mathbf{sh} - \mathbf{c} ''...' est notre amie et nous obtenons avec \mathbf{NetBSD} :

Et sous **Linux**:

```
bernard@debian7 ~ % sudo sh -c "findu/u\(u-pathu/procu-ou-pathu/sysu\)u-pruneu-ou-nameuu'**'u-print0u|uxargsu-0uustatu--printf ='%zu%n\n'u|uegrepu'^$(dateu"+%Y-%m-%d")'"

2013-11-15 09:54:42.000000000 +0100 /var/lib/misc/statd.status
2013-11-15 09:54:41.000000000 +0100 /var/lib/urandom/random-seed
2013-11-15 09:55:09.000000000 +0100 /var/lib/dhcp/dhclient.em0.
leases
2013-11-15 09:54:49.000000000 +0100 /var/lib/exim4
```

```
6 2013-11-15 09:54:49.000000000 +0100 /var/lib/exim4/config.
autogenerated
7 2013-11-15 09:54:45.000000000 +0100 /var/lib/postgresql/9.1/main
8 2013-11-15 09:54:46.000000000 +0100 /var/lib/postgresql/9.1/main/global
```

4.4. **POSIX** et **GNU**. Profitons d'un moment de calme pour remarquer que de nombreuses commandes se comportent de manière très standard entre différents systèmes, y compris parfois, sous **MS/DOS**. Tout cela vient de **POSIX** ou de **GNU**.

Les guerres de religions qui opposent parfois violemment les mondes \mathbf{BSD} et \mathbf{Linux} , les supporters de Vi et Emacs, ... finissent par être absorbées avec le temps et seuls quelques irréductibles les raniment, souvent plus pour s'exposer aux yeux (blasés maintenant) du petit monde concerné. Seule reste l'opposition farouche entre tenants du libre et leurs opposants.

Ici, nous avons utilisé find de la même manière sous **Linux** et sous **NetBSD** ce qui n'a pas été toujours le cas, de même, sh se comporte de manière identique à quelques octets près sur les deux systèmes, ce qui n'était pas forcément vrai il y a quelques années. Pour revenir à find, nous avions un paquet de compatibilité **GNU** disponible sur plusieurs **BSD** qui reprenait le find que nous connaissons maintenant et l'on pouvait différencier gfind de bsdfind³.

Troisième partie 3. les scripts et les exemples

5. Les scripts shell

La magie des SHELLSest infinie, ils nous permettent en effet de créer des programmes complets, complexes... parfois aux limites du lisible. On les appelle scripts pour les opposer aux applications généralement créées à partir de langages compilés mais cela ne devrait rien changer au fait qu'ils doivent être conçus avec un soin égal à celui apporté aux autres langages comme C/C++, Java...

Dans tout ce qui suit, il ne faut pas perdre de vue que le SHELLest une coquille entourant le noyau d' \mathbf{Unix} . Certains aspects des SHELLSne font que recouvrir des appels systèmes.

5.1. **structure des scripts.** Ce qui est décrit ici est valable autant pour des langages interprétés comme l'horrible $Perl^4$, le sublime $Scheme^5$, le célèbre Python que pour n'importe quel shell.

^{3.} A vérifier dans les détails.

^{4.} je ne suis pas objectif, mais quand même...

^{5.} là, je me sens plus objectif... ou presque.

La première ligne d'un script est le *shebang*. Cette ligne est très importante car elle indique de manière sûre quel interpréteur il doit utiliser pour exécuter le corps du script. Voici quelques exemples :

sh: : #!/bin/sh bash: : #!/bin/sh

Perl: : #!/usr/bin/env perl

Python 2.7: : #!/usr/bin/env python2.7

Python: : #!/usr/bin/env python

awk: : #!/bin/awk -f

Les deux caractères #! sont considérés comme un nombre magique par le système d'exploitation qui comprend immédiatement qu'il doit utiliser le script dont le nom et les arguments suivent les deux caractères.

Dans un Shellinteractif, l'exécution d'un script suit l'algorithme suivant :

```
fork ();
   if (child) {
       open(script);
       switch(magic_number) {
           case Ox7f'ELF':
                exec_binaire();
                break;
           case '\#!':
                load_shell(first_line);
9
                exec_binaire(shellname, args);
10
11
12
13
  } else {
14
       wait(child);
15
```

- 5.2. **choisir son shell.** Par tradition autant que par prudence, on écrit ses scripts shellpour le shelld'origine, soit sh. Par prudence car on est certain qu'il sera présent sur la machine même si elle démarre en mode dégradé. Cependant, beaucoup de scripts sont applicatifs et ne pourront pas fonctionner en mode dégradé. Autant se servir d'un shellplus complet comme bash.
- 5.3. les paramètres des scripts. Les paramètres, leur nombre et leur taille n'ont de limites que de l'ordre de la dizaine de Ko. Il faut donc pouvoir y accéder. Le paramètre \$0 est le nom du script parfois avec le nom du répertoire. Les neufs suivants sont nommés \$1, ..., \$9. Pour accéder aux autres il faut ruser un peu avec l'instruction shift.

5.4. tests et boucles. Les tests se font avec if de cette manière :

Dans le même ordre d'idée, nous avons le while :

```
while condition
do
done
```

La construction des *condition* est tout un art, d'autant plus qu'en lieu et place du *if* nous pouvons écrire :

```
condition && condition_true && ...
```

ou bien:

```
condition || condition_false
```

Nous avons aussi une boucle for :

```
for index in ensemble
do
...
done
```

La détermination de **ensemble** est assez naturelle comme par exemple avec **\$(1s** *.java). Mais il faut être prudent : selon les SHELLSles résultats peuvent différer.

5.5. **conditions, valeurs de retour des programmes.** Tout les programmes sous **Unix** s'achèvent par un *return EXIT_CODE* ou bien un *exit(EXIT_CODE)* bien senti. La valeur *EXIT_CODE* est renvoyée au programme appelant, notre SHELL. On peut le récupérer depuis la variable *\$#*? puis étudier le cas :

```
myprogram arg1 arg2 ...
   case \ \ \ in
2
       0)
3
            its-okayyy
            ;;
5
       1)
6
            bad_parameterzzz
            ;;
       2|3|4)
            cant-open-filezzz
10
11
```

```
12 *)
13 unknow-error
14 ;;
15 esac
```

Unix considère que la valeur de retour 0 est signe que tout va bien et que tout autre valeur exprime une condition d'erreur. On peut donc utiliser cette propriété ainsi :

```
myprogram arg1 arg2 ... || onerror 'Error code \$\#?''
```

- 5.6. redirections.
- 5.7. tubes ou pipes.
 - 6. EXEMPLES DE MANIPULATION DE TEXTE
- 6.1. des stats. Voici un extrait d'un fichier /var/log/messages :

```
Nov 3 10:16:19 localhost org.gnome.zeitgeist.SimpleIndexer[2637]:
    ** \ldots

Nov 3 10:16:34 localhost org.freedesktop.FileManager1[2637]:
    Initializing \ldots

Nov 3 10:16:34 localhost nautilus: [N-A] Nautilus-Actions Menu
    Extender 3.2\ldots

Nov 3 10:16:34 localhost org.freedesktop.FileManager1[2637]:
    Initializing naut\ldots

Nov 3 10:16:34 localhost nautilus: [N-A] Nautilus-Actions Tracker
    3.2.3 initializing\ldots
```

Nous voulons déterminer les moments les plus actifs de ce fichier avec une granularité de une heure. La manipulation est simple :

```
afficher le fichier: cat file-name,
découper le ficher: cut -d ':' -f 1,
trier le fichier: sort,
compter les occurences: uniq -c,
trier en décroissant: sort -n.
```

Ce qui nous donne la commande :

```
1 cat $file-name |
2    cut -d ':' -f 1 |
3    sort |
4    uniq -c |
5    sort -n
```

On obtient rapidement un script (stat1.sh) à partir de cette ligne de commande :

```
#!/bin/sh
   scriptname="$(basename_\$0)"
3
4
5
   dohelp() {
       cat << DOHELP</pre>
6
   \{scriptname\} [-h|--help] : this text
   ${scriptname} file file \ldots : stats
   DOHELP
9
10
        exit 0
   }
11
12
   [ $# -eq 0 ] && dohelp
13
   case $1 in
14
       -h | --help)
15
            dohelp
16
            ;;
17
       *)
18
            cat "$0" | \
19
                 cut -d ':' -f 1 | \
20
                     sort | \
21
                          uniq -c | \
22
23
                               sort -n
24
25
```

On peut tester:

```
$ ./stat1.sh messages-1 /var/log/messages /var/log/messages.1
1
2
     152 Nov
               4 10
3
     155 Oct 27 11
4
     156 Oct 28 09
5
     164 Oct 28 17
6
     186 Oct 25 14
     213 Oct 28 11
8
               3 10
     216 Nov
     260 Oct 28 10
10
     636 Oct 26 15
11
     774 Oct 28 07
12
    1770 Nov
               3 09
13
    3844 Oct 28 14
14
   26201 Oct 28 15
```

- 6.2. **peut-on faire mieux?** Bien sûr! On peut avoir d'autres options que la simple aide, on peut aussi gérer correctement les erreurs, les *signaux* Unix . . .
- 6.2.1. les options. Depuis longtemps il existe une norme **POSIX** permettant de gérer les options de la ligne de commande. Malheureusement, il fut une époque où la norme avait beaucoup de variantes ce qui m'a poussé à faire ma propre gestion de ces paramètres.

Voici ma méthode, facile à mémoriser mais pas parfaite et un peu lourde :

l'aide: créer une fonction dohelp comme dans l'exemple précédent; le nom dohelp permet d'éviter un clash avec une éventuelle commande help.

s'assurer de l'existence de paramètres: il suffit de faire le test [\$# -eq 0] et exécuter le code nécessaire.

vider la liste des paramètres: une boucle while [\$# -ne 0] fait l'affaire.

Voici un exemple plus parlant (script stat2.sh):

```
#!/bin/sh
    scriptname="$(basename_$0)"
    dohelp() {
          cat << DOHELP
 6
    \{scriptname\} [-h|--help] : this text
    ${scriptname} [options] file file \ldots : stats
9
    options:
          -s|--size N : number of most important hours, default 5
10
          -b|--byhour : for each hours
11
    DOHELP
12
13
          exit 0
14
15
16
    byhour="cut_{\square}-d_{\square}':'_{\square}-f_{\square}1"
17
    is_byhour=0
    after=""
19
20
21
    set_size() {
          [ "$1" -lt 1 ] && onerror 3 "size_{\sqcup}must_{\sqcup}be_{\sqcup}>_{\sqcup}1"
22
          size=$1
23
24
   }
25
   set_byhour() {
          byhour="\{byhour\}_{\sqcup}|_{\sqcup}tr_{\sqcup}-s_{\sqcup}'_{\sqcup}'_{\sqcup}|_{\sqcup}cut_{\sqcup}-d_{\sqcup}'_{\sqcup}'_{\sqcup}-f_{\sqcup}3"
          after="_{\sqcup}|_{\sqcup}sort_{\sqcup}-k_{\sqcup}2"
27
    is_byhour=1
```

```
set_size 24
29
30
   doit() {
31
       end=1
32
       cmd="catu$@u|u${byhour}u|usortu|uuniqu-cu|usortu-nru|uheadu-nu
33
          {size}_{\perp} after {"}
       case ${is_byhour} in
34
       0)
               printf "%-7.7s_{\perp}%-6.6s_{\perp}%-2.2s\n" "occurences" "date" "
36
                   hour"
                printf "%-17.17s\n" "
38
       1)
39
         printf "%-7.7s_{\perp}%-2.2s\n" "occurences" "hour"
40
            printf "%-10.10s\n" "
42
     esac
43
   eval "${cmd}"
44
45
   }
46
47
   [ $# -eq 0 ] && dohelp
   end=0
48
49
   while [ $end -eq 0 ]
50
51
     case $1 in
52
        -h | --help)
53
          dohelp
54
55
          -s | --size)
56
               shift
57
               [ $# -eq 0 ] && onerror 2 "$1 needs a parameter"
                set_size "$1"
59
60
                shift
61
          -b | --byhour)
62
63
                set_byhour
64
65
        *)
66
               doit "$0"
67
68
       esac
```

70 done

Et maintenant avec le getopts ⁶:

```
#!/bin/sh
    scriptname=$(basename $0)
   dohelp() {
         cat << DOHELP</pre>
6
   ${scriptname} [-h] : this text
   ${scriptname} [options] file file \ldots : stats
9
    options:
         -s N : number of most important hours, default 5
10
         -b : for each hours
11
12 DOHELP
         exit 0
13
   }
14
    onerror() {
        local exit_code=$1
16
       shift
17
      local error_msg="$0"
19
    echo "ERROR:⊔$error_msg" 1>&2
         exit "$exit_code"
21
22
23
size=5
   byhour="cut_{\square}-d_{\square}':'_{\square}-f_{\square}1"
26 is_byhour=0
   after=""
27
29
    set_size() {
         [ "$1" -lt 1 ] && onerror 3 "size_{\square}must_{\square}be_{\square}>_{\square}1"
30
         size=$1
31
32
   }
   set_byhour() {
33
         byhour="\{byhour\}_{\sqcup}|_{\sqcup}tr_{\sqcup}-s_{\sqcup}'_{\sqcup}'_{\sqcup}|_{\sqcup}cut_{\sqcup}-d_{\sqcup}'_{\sqcup}'_{\sqcup}-f_{\sqcup}3"
34
         after="_{\sqcup}|_{\sqcup}sort_{\sqcup}-k_{\sqcup}2"
35
36
         is_byhour=1
         set_size 24
37
   }
38
   doit() {
39
         cmd="catu$@u|u${byhour}u|usortu|uuniqu-cu|usortu-nru|uheadu-nu
             {size}_{\perp} after "
```

^{6.} soyez prudents avec les (très) anciennes versions de Red Hat

```
case ${is_byhour} in
41
42
                printf "%-7.7s_{\square}%-6.6s_{\square}%-2.2s\n" "occurences" "date" "
                   hour"
                printf "%-17.17s\n" "
45
         1)
         printf "^{-7.7}s_{\perp}" -2.2s\n" "occurences" "hour"
47
                printf "%-10.10s\n" "
48
49
     esac
50
   eval "${cmd}"
51
52
53
   # [ $# -eq 0 ] && dohelp
54
55
   [ $# -eq 0 ] && echo "you need arguments" && dohelp
56
57
58
   while getopts "s:bh" opt
59
60
      case $opt in
       h)
61
                dohelp
62
                ;;
63
64
                set_size "$OPTARG"
65
66
        b)
67
                set_byhour
68
69
        :)
70
71
                onerror 2 "$OPTARG_needs_a_parameter"
72
73
        \?)
        onerror 7 "option」$OPTARG⊔is⊔unknown"
74
      ;;
75
76
77
   done
78
   shift $((OPTIND-1))
   doit "$@"
80
```

En fait, <code>getopts</code> ne sait traiter que les options courtes et classiques d'Unix . Les options longues à la mode Linux ne sont pas supportées. L'avantage de <code>getopts</code> est son mode de fonctionnement assez simple. Son inconvénient principal est d'être très <code>bash</code> même si <code>POSIX</code> le soutient, ce qui fait qu'il n'est pas forcément disponible partout.

Pour avoir les *options longues*, il faut utiliser l'outil **GNU** getopt (sans le s de fin) ⁷. Je reste donc sur ma méthode qui n'est finalement ni meilleure ni pire.

6.2.2. $avec \, bash$. On peut profiter des avantages de bash: boucles, tableaux, ... comme dans ce script (qui ne fonctionne pas avec sh):

```
#!/usr/bin/env bash
    scriptname="$(basename_$0)"
4
   dohelp() {
5
         cat << DOHELP
6
   \{scriptname\} [-h|--help] : this text
   ${scriptname} [options] file file \ldots : stats
    options:
9
10
         -s|--size N : number of most important hours, default 5
11
         -b|--byhour : for each hours
   DOHELP
12
13
14
15
   byhour="cut_{\square}-d_{\square}':'_{\square}-f_{\square}1"
17
   is_byhour=0
18
   after=""
19
   hours=
20
21
    set_size() {
22
         [ "$1" -lt 1 ] && onerror 3 "size, must, be, >, 1"
23
24
         size=$1
   }
25
26
   set_byhour() {
         local i
27
         byhour="\{byhour\}_{\sqcup}|_{\sqcup}tr_{\sqcup}-s_{\sqcup}'_{\sqcup}'_{\sqcup}|_{\sqcup}cut_{\sqcup}-d_{\sqcup}'_{\sqcup}'_{\sqcup}-f_{\sqcup}3"
28
         after="_||sort_-k_2"
29
         is_byhour=1
30
         set_size 24
31
         for ((i=0; i<24; i++)) {</pre>
32
              hours[i]=0
33
```

^{7.} voir cette discussion sur StackOverflow

```
35
   doit() {
36
37
       local ftmp=/tmp/tmp.file
       end=1
38
        cmd="catu$@u|u${byhour}u|usortu|uuniqu-cu|usortu-nru|uheadu-nu
39
           {size}_{\perp} after}"
        case ${is_byhour} in
40
        0)
41
                printf "%-7.7s_{\perp}%-6.6s_{\perp}%-2.2s\n" "occurences" "date" "
42
                    hour"
                printf "%-17.17s\n" "
44
45
         1)
          printf "\%-7.7s_{\perp}\%-2.2sn" "occurences" "hour"
46
             printf "%-10.10s\n" "
48
49
       esac
       eval "${cmd}" > ${ftmp}
50
       while IFS=' ' read count index; do
51
            local h=${index#0}
52
            hours [${h}] = ${count}
53
       done < $ftmp</pre>
       for ((i=0; i<24; i++)) {</pre>
            printf "%7d_{\perp}%02d_{n}" ${hours[i]} $i
56
57
58
59
   [ $# -eq 0 ] && dohelp
60
   end=0
61
62
   while [ $end -eq 0 ]
63
64
       case $1 in
65
66
           -h | --help)
           dohelp
67
68
            -s | --size)
69
              shift
70
               [ $# -eq 0 ] && onerror 2 "$1 needs a parameter"
71
                set_size "$1"
                shift
73
74
       -b | --byhour)
```

Conrairement aux précédents, si une tranche horaire n'est pas représentée dans les fichiers logs passés en paramètres, elle sera tout de même affichée avec la valeur 0.