

Unix

Tp 4

On s’amuse...

Florent DEVIN

7 octobre 2008

1 Introduction

En plus de vos portables, vous avez un compte de créé à l’école. Ce compte vous permet de consulter vos mails, de vous connecter sur AREL, et bien d’autres choses. Il peut aussi vous servir à sauvegarder vos données importantes.

Il est nécessaire de sauvegarder régulièrement vos données pour vous éviter de perdre votre travail dans le cas d’une casse matérielle, ou autre. Pour ce faire, il existe plusieurs façons de faire. Une façon pratique que vous pourriez mettre en place, c’est de sauvegarder périodiquement vos données sur votre compte.

Pour ce faire, il est nécessaire de connaître l’existence d’une machine accessible, et présent dans le réseau étudiant. Sur Pau¹, vous avez à “disposition” les machines de la salle 107 et 109. Inutile de vous précipiter sur ces machines. Nous allons apprendre comment se connecter à distance sur ces machines, et effectuer quelques manipulations standards.

Pour se connecter sur une machine, on a besoin de connaître soit son nom, soit son adresse IP, et utiliser la commande `ssh`. Voici comment la commande `ssh` s’utilise dans sa version simple : `ssh login@machine` Ce qui donne par exemple :

- `ssh fd@193.55.155.71`
- `ssh fd@pc10701.etude.pau.eisti.fr`

Ces deux commandes ont le même effet, elles permettent de se connecter sur la même machine, avec le même nom d’utilisateur (ici `fd`). La première commande utilise l’adresse IP de la machine (193.55.155.71), tandis que la seconde utilise son nom.

En fait, cette commande vous permet d’obtenir un shell sur une machine distante. Si il existe plusieurs autres façons de se connecter à distance sur une machine (`rlogin`, `telnet`, ...), `ssh` est la meilleure, chiffrement des données, échange de clés de chiffrement. Une fois connecté sur une machine distante vous pouvez travailler sur celle-ci exactement comme si

¹Pour les cergy-pontains, votre enseignant, ou l’équipe système, devrait pouvoir vous donner les informations adéquates. Pour les palois, dans la salle 107 il y a 40 machines, numérotées de `pc10701` à `pc10740`

c'était votre machine². Si vous passez l'option `-X` à la commande `ssh`, vous pourrez même lancer des applications graphiques.

Par ailleurs, vous pouvez remarquer que vous fournissez le login sur lequel vous souhaitez vous connecter. Il faut cependant connaître le mot de passe, puisque `ssh` vous demande le mot de passe associé au login.

Question 1 : Connectez vous sur une machine³ de la salle 107.

Réponse _____
`ssh login@pc107xx.etude.pau.eisti.fr`



Question 2 : Combien de personnes sont actuellement connectées sur cette machine ?

Réponse _____
`who`



Cependant il y a un inconvénient à la commande `ssh`, c'est qu'elle fait précisément ce pour quoi elle a été faite;-). Ce qui veut dire que lorsque vous utilisez la commande `ssh` vous êtes sur une autre machine, et donc vous n'avez plus accès aux données de votre machine initiale. C'est pourquoi la commande `scp` a été créé. De même il a été créé la commande `sftp`. Nous n'utiliserons pas cette dernière.

Question 3 : Comment fonctionne la commande `scp` ?

Réponse _____
`man scp`



Question 4 : Utiliser cette commande pour sauvegarder l'ensemble des Tps Unix que vous avez réalisé.

Réponse _____
`scp -r Unix/ login_eisti@193.55.155.71 : /`



Pour évitez toute surcharge sur les ordinateurs de l'école, il vaut mieux ne sauvegarder que les fichiers sources, c'est à dire votre production personnelle. Connectez vous sur une autre machine. Vous remarquez que vous avez toujours accès à vos fichiers. En fait, vos comptes EISTI sont stockés sur un serveur. C'est ce serveur qui gère vos données, et qui se charge de vous les fournir quand vous en avez besoin.

Question 5 : Nettoyez votre compte pour qu'il ne contienne que les fichiers que vous avez produit.

Copier le fichier `liste.txt` disponible sur AREL et enregistrez le sur votre disque dans un répertoire que vous créez.

²D'où un environnement fortement multi-utilisateurs

³Pour les cergy pontains, utiliser une salle de cergy

2 Exercices

2.1 Étudiants

Pour les questions suivantes, ne regardez pas sur internet comment faire, mais essayez de trouver avec l'aide installée sur vos machines. Il est important que vous soyez confronté avec cette démarche, car en examen, internet ne sera pas disponible.

Question 6 : Regardez le contenu du fichier avec `less` les colonnes indiquent :

- le nom de l'étudiant
- sa filière
- son sexe
- sa note dans la matière A
- sa note dans la matière B

Question 7 : Comptez le nombre de lignes, de mots

Réponse _____
`wc -l liste.txt ;wc -w liste.txt`



Question 8 : Triez le fichier

- (a) par ordre alphabétique des étudiants

Réponse _____
`sort -k1 liste.txt`



- (b) par note décroissante dans la matière A, puis B

Réponse _____
`sort -k4rn -k5rn liste.txt`



- (c) par matière

Réponse _____
`sort -k2 liste.txt`



Question 9 : Filtrez le fichier pour récupérer les étudiants d'informatique.

Réponse _____
`grep -w info liste.txt`



Question 10 : Faites de meme pour les étudiants de physique et de mathématique.

Réponse _____
`grep -v -w info liste.txt`



Question 11 : Filtrez le fichier pour récupérer la liste des étudiantes puis la liste des étudiants.

Réponse _____
`grep -w F liste.txt ; grep -w F liste.txt`



Question 12 : Filtre les étudiants par filière, et affichez UNIQUEMENT les prénoms.

Réponse _____
`sort -k2 liste.txt | cut -f 1`



Question 13 : Créez le fichier `math_hommes.txt` qui contient les étudiants classés par filières, par sexe, ainsi qu'en note décroissante.

Réponse _____
`grep -w M liste.txt | sort -k2,2 -k4rn -k5rn > math_hommes.txt`



2.2 Génome

Copiez le fichier `dmel.fasta`

Il contient des séquences d'ADN de drosophile (séquences intergéniques du chromosome 4), en format fasta.

On va utiliser ce fichier pour rechercher dans les séquences d'ADN certaines signatures pour des sites de fixation de facteurs de transcription⁴.

Les sites de fixation des facteurs de transcription sont des sous séquences (ou "motif") de quelques nucléotides (6-8), reconnues par les protéines qui se fixent à cet endroit. Souvent, un facteur de transcription reconnaît plusieurs motifs qui se ressemblent. Pour tenir compte de cette variabilité, on introduit d'autres lettres, en plus des A, C, G, T (code IUPAC) :

⁴Oui je sais ça semble complexe...

Lettre	Équivalence	Signification
A		Adenine
C		Cytosine
G		Guanine
T		Thymine
R	A ou G	puRines
Y	C ou T	pYrimidines
W	A ou T	Weak hydrogen bonding
S	G ou C	Strong hydrogen bonding
M	A ou C	aMino group at common position
K	G ou T	Keto group at common position
H	A, C ou T	not G
B	G, C ou T	not A
V	G, A ou C	not T
D	G, A ou T	not C
N	G, A, C ou T	aNy

Avec cette nomenclature, on peut désigner un motif par ARCCGKKY par exemple.

Question 14 : Traduisez en expression régulière le motif ARCCGKKN

Réponse _____
ARCCGKKN : A[AG]CCG[GT][GT].

Question 15 : Recherchez ce motif dans le fichier de séquences

Réponse _____
grep -E 'A[AG]CCG[GT][GT].' dmel.fasta

Voici quelques facteurs de transcription fameux chez la Drosophile, et leur sites de fixation consensus :

Hunchback (Hb)	SMANAAAAAA
Krueppel (Kr)	AMYGGGTTAN
Bicoid (Bcd)	SGGATTAN
Engrailed (en)	GTANTNN
suppressor of hairless	ANYGTGGGAAMCM

Question 16 : Transformez ces consensus en expressions régulières, et recherchez ces sites dans le fichier de séquences.

Réponse _____

- SMANAAAAAA : *grep -E '[GC][AC]A.A{6}' dmel.fasta*
- AMYGGGTTAN : *grep -E 'A[AC][CT]G{3}TTA.' dmel.fasta*
- SGGATTAN : *grep -E '[GC]GGATTA.' dmel.fasta*
- GTANTNN : *grep -E 'GTA.T..' dmel.fasta*

Question 17 : Cherchez aussi des combinaisons de sites de fixation (comme par exemple : Hb et Kr) ♣

Réponse _____

grep -E '([GC][AC]A.A{6}|A[AC][CT]G{3}TTA.){2}' dmel.fasta

Question 18 : Comment faire pour afficher les lignes qui commencent par ">xxxx" et qui indiquent dans quelle séquence intergénique on se trouve ? ♣

Réponse _____

grep -B 1 -E '[GC][AC]A.A{6}' dmel.fasta
