Shell scripting and system variables

HORT 530

Lab 5

Instructor: Kranthi Varala

Bash comparisons

	Operator	Example	Result
String comparison	>	[[a > b]]	FALSE
	<	[[a < b]]	TRUE
	==	[[aa == bb]]	FALSE
	!=	[[aa != bb]]	TRUE
Integer comparison	-gt	[[25 -gt 5]]	TRUE
	-lt	[[25 -lt 5]]	FALSE
	-ge	[[10 -ge 9]]	TRUE
	-le	[[10 -le 5000]]	TRUE
	-eq	[[25 -eq 5]]	FALSE
	-ne	[[25 -ne 5]]	FALSE
Pattern	=	[[Tree = Tr*]]	TRUE
RegEx	=~	[[Tree =~ T[A-Z][0-9]+]]	FALSE

Table source: http://mywiki.wooledge.org/BashFAQ/031

Bash arithmetic

- Basic operators for multiplication, division, addition and subtraction are supported.

- Order of precedence is : *,/,+,-

- `$(( 4*5/2+3-1 ))` = 12 but, `$(( 4+5*2/2-1))` = 8

- Additional operators:
  - x++ : Add 1 to x
  - x-- : Subtract 1 from x
  - x % y : Divide x by y and return the remainder
  - x ** y : Raise x by y

- Order of precedence is : `**,*,/,%,+,-`

Exercise 1

- Write a shell script to split the fastq file /home/kvarala/Files/SRR6473849.fastq into 10 equal pieces.

- Notes:
    - You don't need to copy the file to your directory since this file is readable to you.
    - First figure out how many sequences are in this file, then divide that by 10 (let's call this y).
    - Use a loop where x goes from 1 to 10:
        - head -n (x*y) <fastq> | tail -n y > Part$x

Exercise 2

- Write a shell script to download the fastq files corresponding to the samples listed in the file /scratch/scholar/kvarala/ICB/Week5/SraRunTable.txt
  - The SraRunTable.txt was downloaded from NCBI's GEO database and lists all the sequencing runs associated with one experiment (GSE109388).
  - One of the columns in this file contains the Run ID for each sample. The run IDs always have the format SRRXXXXXXX where 'XXXXXXX' is a number.
  - Once you have extracted the list of SRR IDs, you can download the associated fastq file using the /scratch/scholar/kvarala/ICB/Week5/dumpFastq.sh script
  - Use a loop to download all the 12 fastq files related to this experiment.
  - After each file is downloaded check to confirm that the number of lines in the file is divisible by 4.

Exercise 3

- Write a shell script to process `/scratch/scholar/kvarala/ICB/Week5/Matrix.txt` and count genes whose mean expression in Wild-type is greater than X. Calculate the number of genes for each value of X in the range 4-16.

- Note:
  - Mean expression in Wild-type is the average value of columns 2,3,4,8,9 and 10.
  - To pass shell variables to awk use the –v switch e.g.,

```
GC=`awk -v t=$X -F "\t" '{WT=(\$2+\$3+\$4+\$8+\$9+\$10)/6;if(WT > t){print \$0}}' Matrix.txt |wc -l`
```

X is a shell variable