cloud

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Friday 6th June, 2014 11:43

outline

outline

no more installing software promise

- in the first class when we installed GIT and npp
  i promised that we won't be installing anything else
- today we run Python and there is no Python at the lab...

software

- you have noticed that there are lots of problems when everybody manages data from her own PC
  - lots of software installing
  - software works differently on different machines
  - and code works differently, too
  - it's difficult to make it PC independent: packages, paths, etc
- much of the time you and I spend on this class is resolving inconsistencies accross different PCs
- this is a great practice for data management (no kidding)...

server/research cloud

- ... but the future is “cloud computing”
  - that is, everybody connects and works on the same server
- beyond learning portability and hacking with dropbox you need to learn remote/cloud computing
  - draw a cloud picture with PC’s connecting to it...
- every Uinversity already has it
  - nobody uses it because it is little clunky/slow/outdated
  - still, it is great, undervalued, and it is the future !
- UTD has a server running Python and R (little outdated)
- I have a server (unsecure)

Linux/Unix

- and cloud/server is usually Linux/Unix
- then it has lots of great utilities that are amazing for data managemt
- we will be covering those later; e.g.:
  - wget
  - grep
  - curl
- but Python can do those things, too

like your PC ?

- if you like your PC or a lab PC, fine
- you can do everything the old way (run on a local machine)
- it's entirely up to you

my server: disclaimer

- no data security
  - do your backups
  - don't put there anyyhing private/sensitive
- it may slow down if many people do complicated things
- save RAM – only 4g
- it may break...

outline

UTD's server

- http://www.utdallas.edu/ir/labs/unixintro.htm
- Linux, Mac: shell
- Win: Putty
- ssh <your netid>@apache.utdallas.edu
- and run "R" or "python"
- I/O: you can move stuff using winscp – let's check it out
  - (Mac people can use http://cyberduck.ch/)
- you can also use scp

```
scp -p /home/aok/Desktop/app_eco/c/c4_matrix.do
ajo021000@apache.utdallas.edu:~/public_html/class/app_reg/
```

- it is secure, but little outdated and missing some packages
  - basic stuff will work but more fancy things may not work

my server: webshell

- we will use something called webshell
  - a bash shell on a website
- `https://aok.us.to:81/` (don't forget "s" on https)
- just hit enter for defaults for
  - Host/IP
  - Port
- user name is your (GIT) name
- and passwd is "data"
- and you are on the server

webshell

- to run R, say R
- to run Python, say python
- and there is git, too
- and of course you can run any Linux utilities, like wget
- sorry, no Stata – I just have one-person license
- please do not use too much RAM – i just have 4g
- don't worry about HD: 500g
- and good CPU: 3.6 Ghz Intel i-7

howto run stuff on webshell

- just open code in npp
- copy it
- and paste into webshell using
- shift INSERT

I/O

- to get data off the server:
- `http://aok.us.to:82/<yourname>`
- to get it on the server:
  use dropbox or something like that
  and download it using R or Python