

Cloud computing

What is cloud computing?

(for scientists)

- You can “rent” access to computers and disk space from a commercial provider of same.
- This provides you with a way to scale your computation for “burst” periods, without investing in hardware.
- Or you can just use a bigger, faster computer.

Why “cloud”?!



...because the diagram that CS people use to represent abstract compute resources looks like a cloud.

PASSWORD ENTROPY IS RARELY RELEVANT. THE REAL MODERN DANGER IS PASSWORD REUSE.



SET UP A WEB SERVICE TO DO SOMETHING SIMPLE, LIKE IMAGE HOSTING OR TWEET SYNDICATION, SO A FEW MILLION PEOPLE SET UP FREE ACCOUNTS.



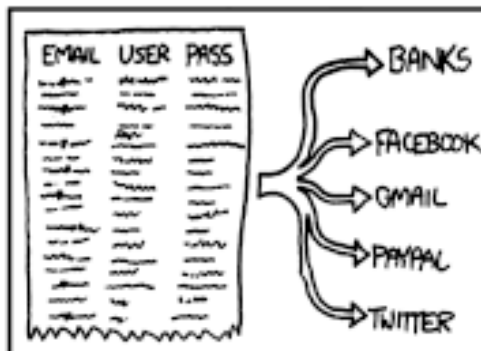
BAM, YOU'VE GOT A FEW MILLION EMAILS, DEFAULT USERNAMES, AND PASSWORDS.



TONS OF PEOPLE USE ONE PASSWORD, STRONG OR NOT, FOR MOST ACCOUNTS.



USE THE LIST AND SOME PROXIES TO TRY AUTOMATED LOGINS TO THE 20 OR 30 MOST POPULAR SITES, PLUS BANKS AND PAYPAL AND SUCH.



YOU'VE NOW GOT A FEW HUNDRED THOUSAND REAL IDENTITIES ON A FEW DOZEN SERVICES, AND NOBODY SUSPECTS A THING.



Amazon is a major cloud computing provider

- Did you know they rent computers!?
- Rumors are that it's more lucrative than their book selling division...

Terms

EC2 – Elastic Cloud Computing, computer rental from Amazon.

EBS – Elastic Block Storage, virtual hard drive rental from Amazon.

Some quick calculations:

1 small machine, / yr:

1.7gb of RAM, a ~ 1.0 GHz single-core CPU, 160gb of local disk.

\$.06 / hr

8760 hrs / year

=> $\sim \$525$ / year.

\sim a somewhat effective server replacement.

1 high-memory quadruple extra-large instance / yr:

68.4 gb of RAM, 8 core @ ~3.2 GHz, 1.7tb of local disk.

\$1.64 / hr

8760 hrs / year

=> \$14,400 / year

20 high-CPU extra large machines, for a day:

7gb of RAM, 8 x 2.5 GHz CPUs, 1.7tb of local disk.

\$0.58 / hr

24 hrs / day

20 machines

=> ~\$278/ day.

Why is EC2 so expensive??

- They cover *all* hardware, power, air conditioning and network costs.
- That's actually way more expensive than you think. (Talk to your sysadmin or HPC person...)
- They do not operate at 100% capacity,
- They want to make \$\$.

What are **we** using it for?

- Teaching workshops and classes.
- Running MG-RAST, RAST, and assembly rast with kBase backends.
- Scientific computing workhorses without the sysadmins
- Automated testing on clean machines with known software install.

Today's tutorials

1. Log in to a new (blank) machine from Amazon. (We have provided these for you).
2. Install NCBI BLAST
3. Download & format some databases
4. Run BLAST
5. Produce an excel spreadsheet of best hits
- ...
1. Run 2-way BLAST (ecoli x ecoli strains)
2. Calculate reciprocal best hits
3. Produce an excel spreadsheet of putative orthologs

Today's tutorials

1. Get some short-read data onto our instance
2. Download a reference sequence
3. Run short-read aligner BWA

This is a tool for when your data is bigger than your laptop, or for when your campus cluster is down.