Cold Spring Harbor Laboratory

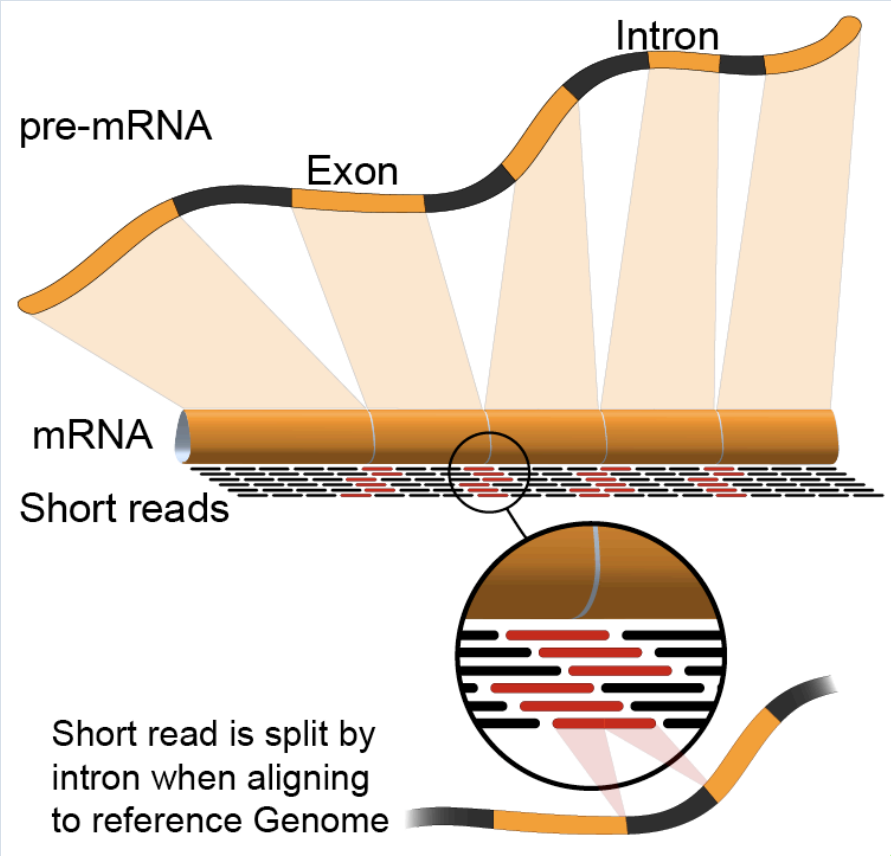
Advanced Sequencing Technologies & Applications

http://meetings.cshl.edu/courses.html

Cold Spring Harbor Laboratory

Module 0
Introduction to cloud computing
(slides modified with permission from Francis Ouellette)

Malachi Griffith, Obi Griffith, Jason Walker, Ben Ainscough
Advanced Sequencing Technologies & Applications
November 11-23, 2014

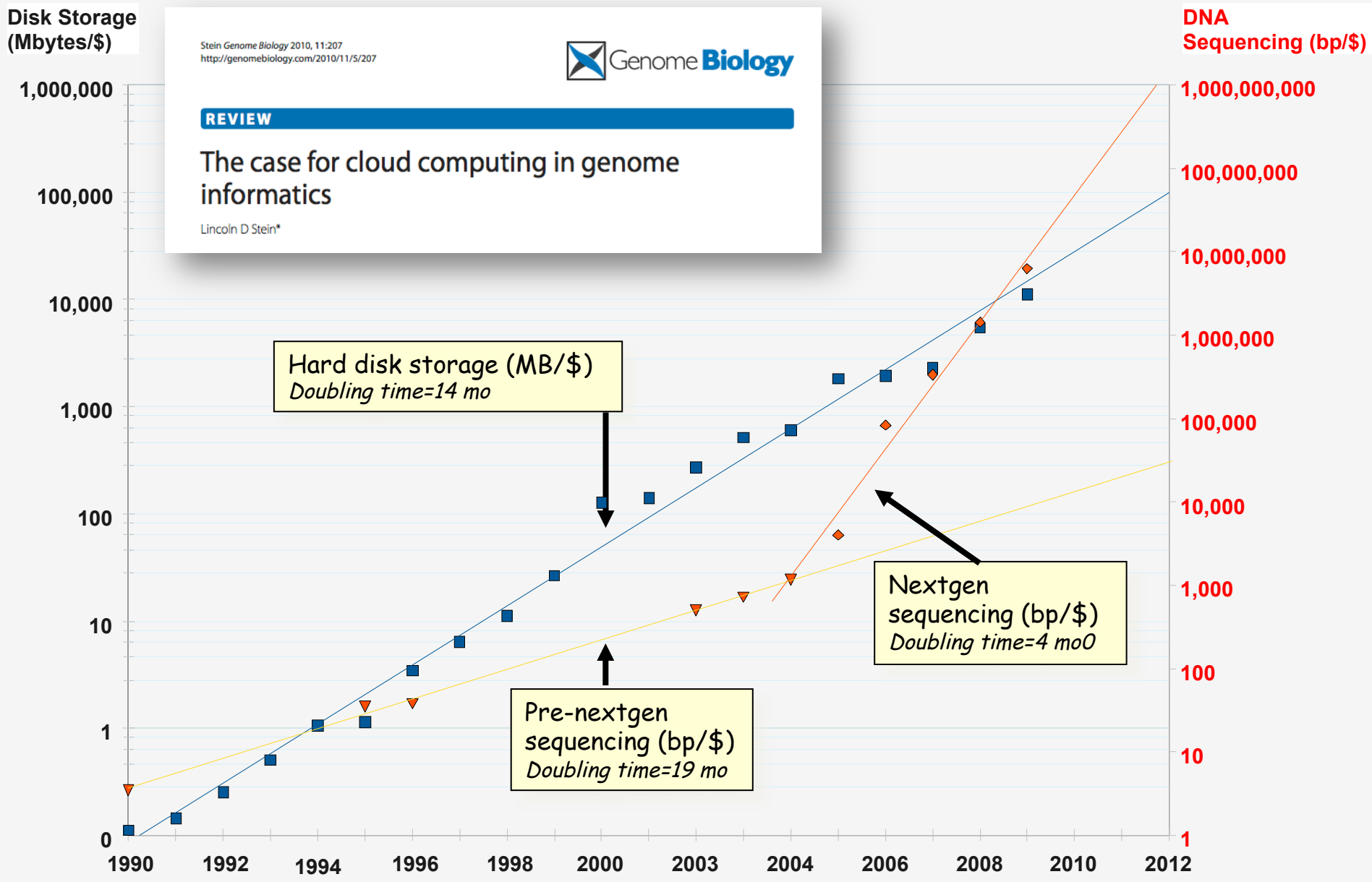


THE GENOME INSTITUTE at Washington University

Learning Objectives of Module

- Introduction to cloud computing
- Use of the wiki in this workshop
- How to log into the cloud

Disk Capacity vs Sequencing Capacity, 1990-2012



Stein *Genome Biology* 2010, 11:207
http://genomebiology.com/2010/11/5/207

Genome Biology

REVIEW

The case for cloud computing in genome informatics

Lincoln D Stein*

About DNA and computers

- We'll hit the $1000 genome during 2013-?, then need to think about the $100 genome.
- The doubling time of sequencing has been ~5-6 months.
- The doubling time of storage and network bandwidth is ~12 months.
- The doubling time of CPU speed is ~18 months.
- The cost of sequencing a base pair will eventually equal the cost of storing a base pair

What is the general biomedical scientist to do?

- Lots of data
- Poor IT infrastructure in many labs
- Where do they go?
- Write more grants?
- Get bigger hardware?

Amazon Web Services (AWS)

- Infinite storage (scalable): S3 (simple storage service)
- Compute per hour: EC2 (elastic cloud computing)
- Ready when you are High Performance Computing
- Multiple football fields of HPC throughout the world
- HPC are expanded at one container at a time:

amazon web services™

http://goo.gl/7PVAl

Some of the challenges of cloud computing:

- Not cheap!
- Getting files to and from there
- Not the best solution for everybody
- Standardization
- PHI: personal health information & security concerns
- In the USA: HIPAA act, PSQIA act, HITECH act, Patriot act, CLIA and CAP programs, etc.
  - http://www.biostars.org/p/70204/

Some of the advantages of cloud computing:

- At the CBW: we received a grant from Amazon, so supported by 'AWS in Education grant award.
- There are better ways of transferring large files, and now AWS makes it free to upload files.
- A number of datasets exist on AWS (e.g. 1000 genome data).
- Many useful bioinformatics AMI's (Amazon Machine Images) exist on AWS: e.g. cloudbiolinux & CloudMan (Galaxy)
- Many flavors of cloud available, not just AWS

amazon web services™

In this workshop:

- Some tools (data) are
  - on your computer
  - on the web
  - on the cloud.
- You will become efficient at traversing these various spaces, and finding resources you need, and using what is best for you.
- There are different ways of using the cloud:
  1. Command line (like your own very powerful Unix box)
  2. With a web-browser (e.g. Galaxy): not in this workshop

Things we have set up:

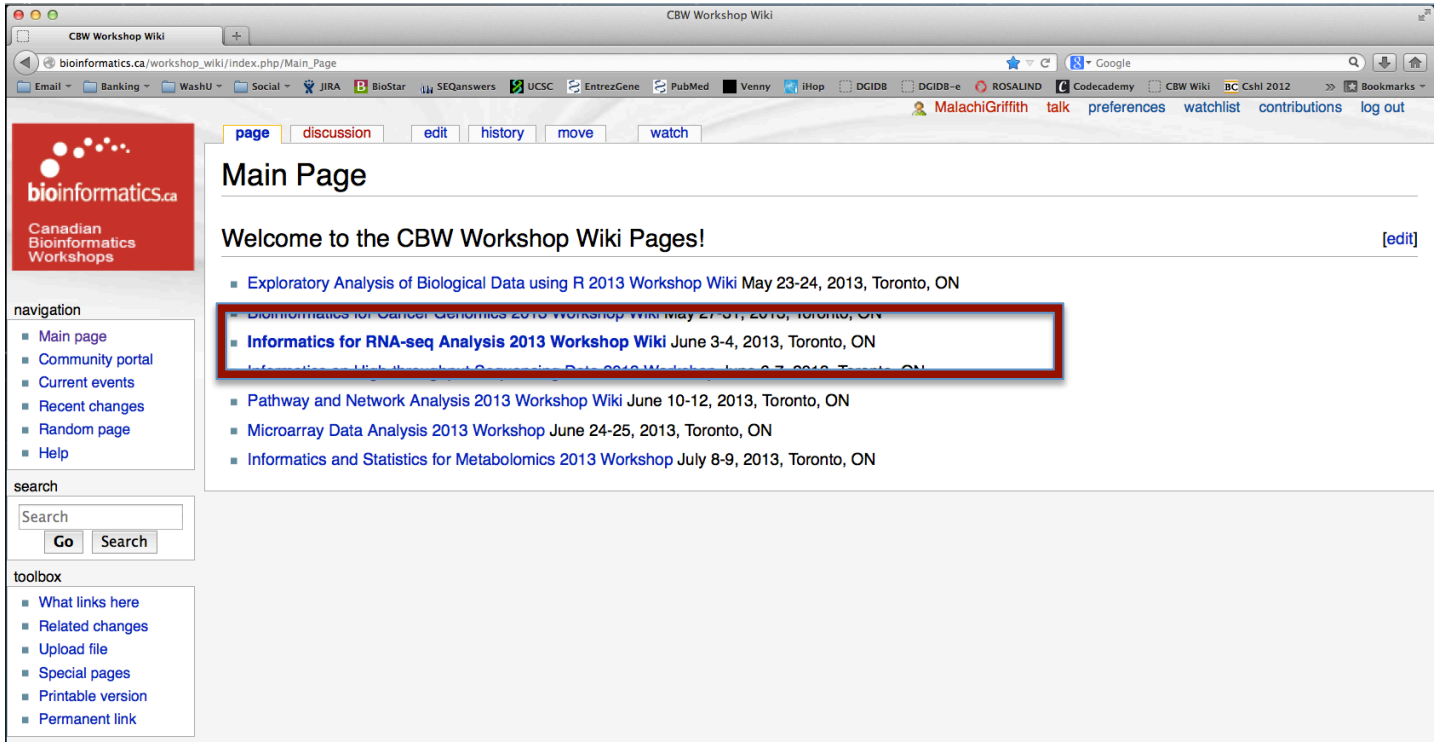
- Loaded data files to an S3 bucket
- We brought up an Ubuntu (Linux) instance, and loaded a whole bunch of software for NGS analysis.
- We then cloned this, and made separate instances for everybody in the class.
- We've simplified the security: you basically all have the same login and file access, and opened ports. In your own world you would be more secure.

For this workshop: all on Wiki!

http://bioinformatics.ca/workshop_wiki/

Login: FirstnameLastname

Password: guest



bioinformatics.ca
Canadian Bioinformatics Workshops

page | discussion | edit | history | delete | move | protect | watch

Informatics for RNA-seq Analysis 2014 Workshop Wiki

navigation
- Main page
- Community portal
- Current events
- Recent changes
- Random page
- Help

search
Search
Go | Search

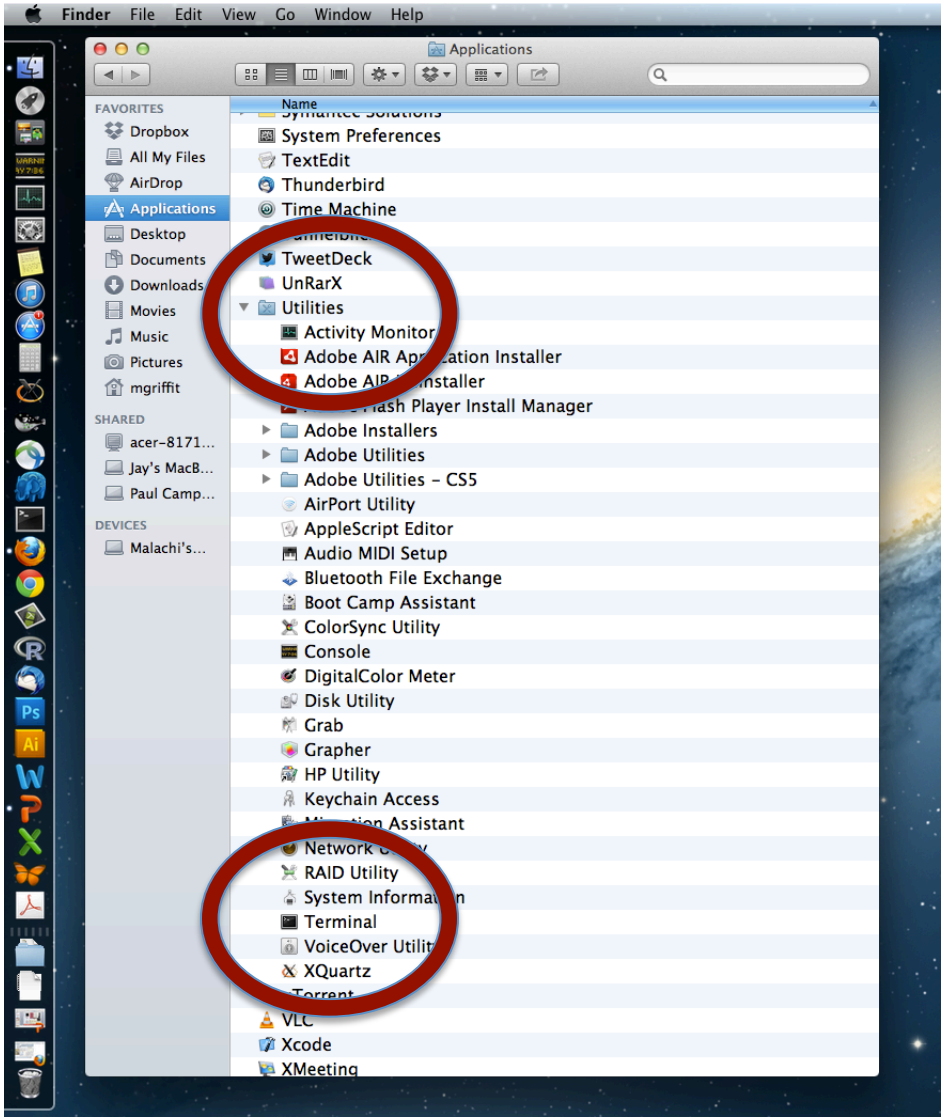
toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

Contents [hide]

1 Parking Lot Postings
2 Course Schedule
3 Laptop Setup Instructions
4 Pre-Workshop Tutorials
6 Logging into the Amazon cloud
  6.1 Logging in with ssh (Mac/Linux)
  6.2 Logging in with Putty (Windows)
  6.3 File system layout
  7.1 Welcome
  7.2 Module 0: Introduction to cloud computing
  7.3 Module 1: Introduction to RNA sequencing and analysis
  7.4 Module 2: RNA-seq alignment and visualization
  7.5 Integrated Assignment
8 Day 2
  8.1 Module 3: Expression and Differential Expression
  8.2 Module 4: Isoform discovery and alternative expression
9 Keeping Up-to-date with RNA-seq Analysis Developments

Macintosh users

Opening a 'terminal session'



```
[mgriffit@Malachis-MacBook-Pro-2:~]$ pwd
/Users/mgriffit
[mgriffit@Malachis-MacBook-Pro-2:~]$ ls -l
total 16
drwxr-xr-x   2 mgriffit  staff    68 Jan  8 23:27 Applications
drwx------+  8 mgriffit  staff   272 Jun  1 15:58 Desktop
drwx------+  7 mgriffit  staff   238 Apr 17 22:00 Documents
drwx------+ 12 mgriffit  staff   408 Jun  1 15:38 Downloads
drwx------@ 12 mgriffit  staff   408 Jun  1 15:13 Dropbox
drwx------@ 51 mgriffit  staff  1734 Feb 27 10:00 Library
drwx------+  4 mgriffit  staff   136 May 30  2012 Movies
drwx------+  4 mgriffit  staff   136 Mar  9  2012 Music
drwx------+  8 mgriffit  staff   272 Dec  9 21:39 Pictures
drwxr-xr-x+  5 mgriffit  staff   170 Mar  5  2012 Public
-rwxr-xr-x   1 mgriffit  staff   594 Mar  6  2012 addSShkey.sh
drwxr-xr-x  16 mgriffit  staff   544 Jan 17 21:42 backup
drwxr-xr-x   5 mgriffit  staff   170 May  7 18:32 git
drwxr-xr-x   6 mgriffit  staff   204 Mar  6  2012 igv
-rwxr-xr-x   1 mgriffit  staff   552 Dec  3 21:53 mac2unix
drwxr-xr-x   3 mgriffit  staff   102 Jun  1 15:55 notes
drwxr-xr-x  95 mgriffit  staff  3230 Dec 29 23:42 temp
drwxr-xr-x   3 mgriffit  staff   102 Oct 27  2012 workspace
[mgriffit@Malachis-MacBook-Pro-2:~]$ mkdir cbw
[mgriffit@Malachis-MacBook-Pro-2:~]$ cd cbw
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ ls -la
total 0
drwxr-xr-x   2 mgriffit  staff    68 Jun  1 16:03 .
drwxr-xr-x+ 64 mgriffit  staff  2176 Jun  1 16:03 ..
[mgriffit@Malachis-MacBook-Pro-2:cbw]$
```

Creating a working directory on your laptop

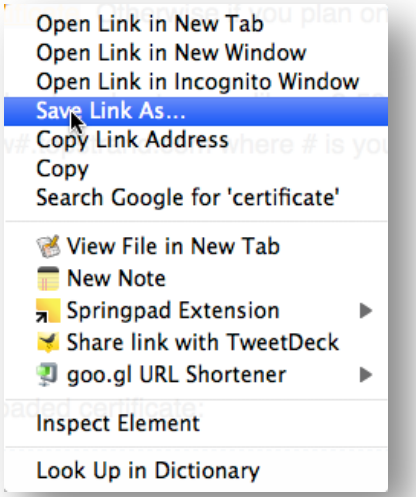
```
[mgriffit@Malachis-MacBook-Pro-2:~]$ pwd
/Users/mgriffit
[mgriffit@Malachis-MacBook-Pro-2:~]$ ls -l
total 16
drwxr-xr-x   2 mgriffit  staff    68 Jan  8 23:27 Applications
drwx------+  8 mgriffit  staff   272 Jun  1 15:58 Desktop
drwx------+  7 mgriffit  staff   238 Apr 17 22:00 Documents
drwx------+ 12 mgriffit  staff   408 Jun  1 15:38 Downloads
drwx------@ 12 mgriffit  staff   408 Jun  1 15:13 Dropbox
drwx------@ 51 mgriffit  staff  1734 Feb 27 10:00 Library
drwx------+  4 mgriffit  staff   136 May 30  2012 Movies
drwx------+  4 mgriffit  staff   136 Mar  9  2012 Music
drwx------+  8 mgriffit  staff   272 Dec  9 21:39 Pictures
drwxr-xr-x+  5 mgriffit  staff   170 Mar  5  2012 Public
-rwxr-xr-x   1 mgriffit  staff   594 Mar  6  2012 addSShkey.sh
drwxr-xr-x  16 mgriffit  staff   544 Jan 17 21:42 backup
drwxr-xr-x   5 mgriffit  staff   170 May  7 18:32 git
drwxr-xr-x   6 mgriffit  staff   204 Mar  6  2012 igv
-rwxr-xr-x   1 mgriffit  staff   552 Dec  3 21:53 mac2unix
drwxr-xr-x   3 mgriffit  staff   102 Jun  1 15:55 notes
drwxr-xr-x  95 mgriffit  staff  3230 Dec 29 23:42 temp
drwxr-xr-x   3 mgriffit  staff   102 Oct 27  2012 workspace
[mgriffit@Malachis-MacBook-Pro-2:~]$ mkdir cbw
[mgriffit@Malachis-MacBook-Pro-2:~]$ cd cbw
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ ls -la
total 0
drwxr-xr-x   2 mgriffit  staff    68 Jun  1 16:03 .
drwxr-xr-x+ 64 mgriffit  staff  2176 Jun  1 16:03 ..
[mgriffit@Malachis-MacBook-Pro-2:cbw]$
```

Obtaining your AWS 'key' file from the wiki

Logging into the Amazon cloud [edit]

- We have set up 40 instances on the Amazon cloud - one for each student. In order to log in to your instance, you will need a security certificate.
  - If you plan on using **Linux** or **Mac OS X**, please download this certificate.
  - If you plan on using **Windows** (with Putty and Winscp), please download this certificate.

On Mac:
Control+




```
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ ls -la
total 0
drwxr-xr-x   2 mgriffit  staff    68 Jun  1 16:10 .
drwxr-xr-x+ 65 mgriffit  staff  2210 Jun  1 16:10 ..
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ ls -la
total 8
drwxr-xr-x   3 mgriffit  staff   102 Jun  1 16:10 .
drwxr-xr-x+ 64 mgriffit  staff  2176 Jun  1 16:10 ..
-rw-r--r--@  1 mgriffit  staff  1692 Jun  1 16:09 CBWRNA.pem
[mgriffit@Malachis-MacBook-Pro-2:cbw]$
```

Viewing the 'key' file once downloaded

```
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ ls -la
total 0
drwxr-xr-x   2 mgriffit  staff    68 Jun  1 16:10 .
drwxr-xr-x+ 65 mgriffit  staff  2210 Jun  1 16:10 ..
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ ls -la
total 8
drwxr-xr-x   3 mgriffit  staff   102 Jun  1 16:10 .
drwxr-xr-x+ 64 mgriffit  staff  2176 Jun  1 16:10 ..
-rw-r--r--@  1 mgriffit  staff  1692 Jun  1 16:09 CBWRNA.pem
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ cat CBWRNA.pem
-----BEGIN RSA PRIVATE KEY-----
MIIEogIBAAKCAQEAhxeHhq0kXb16rZ7GbZDBRi2OyJ1j81jWmOfltxcCJEfjzDCc8ZEtyZthHDVU
gdS0iVb5exBmlqxE+nL4POd2jCBrT4DsyiVAzMMmmrsM0PX4wnoO8hURurgY9dz8j89BPjGhJ3wP
5Dhc+K0bM1IADR/b4ULdLzWaoPwRxouiRAN7PEiMaTXnGlbqiP9tUbF+Ep8F5NVPlKUWC2Ujf5ww
z0edpreHSiQ3juuCRXhVM7+0yY92s5KTLRIfwmzDx3dB0U0/5gZcY5I4dwaXHugHZc5vWctFZDab
pDxcaQatIa5pA5XDAFmwUkFrFPYdNaNvRbav1QGO3LvIfRr9fWjJiQIDAQABAoIBAAZZTL0/I5Jk
PUe7loe/rdWmfndfLjtZiXxWKa2ZHzilSEdoYXfY4tX+v2mb+ATCAloSBFeMRcX/Ap7imFgEaMXd
IC30WSyG46y73yyWbdvEM5SwN9OYueSUkle7swLvMpnBq1k3s2d+XqAxQn3+9JGpVaQYiFyXoFTA
g/VYyp93mNqUbyNl6204h+S6fyVcmgGJhSuW240sfG8lfANnpWu7FW0Bh8ErhdUFXpnRdw5B97fI
QLxbo0wlDZ3BY17JIsudbOqma8Sk+r1xxFxIPx4P+Jod+YEnoQUWwFMkNBCdaNqPI1Xhj+rX5T2z
tD2bEG7rOKy6VabFskk9CAlP0iECgYEAzTJDH+zFjSybh0y0sx6WVvRt2X7LyM+pMyRveNiFGs7W
7Ehrvp/4UZFc+9cKV6n8vSSE4U3w4VnL0yJmhzVhscVOc38aTcs8cfEiqfCOvfoEb3nCBEk+ow4k
Se71MOkmQH+9Qpt44Efl2m/KybMLkwvhMyFGjBu6arp6oWZTegsCgYEAqInqBMUpTQiqVtRdBgOV
jawx8h/iXd+UBdeBtI2Eu0o0wKeGkUAspXJRtwoE26ZEzEdJtZh31RW9LSPDpkfm4sTUMsu2/mG/
jSMa91E1GQV2/l8oNIs/mnA2NY+igc4FC0KuyR6nJXNu11P39vmsouBtyoDyAKrIwkEFUus1mzsC
gYBbzQjK15rqNfK+7UpKeHy6GlCq5Vc2V/FSuBJG9ldWTwNIYn4Rom/tziLTaLlvztULqbLpOX1y
1EGkC9fi1bcHWUZn1Y1nEG3SbBceAElcJOanGMn8MOic+k6x5P49C5HZeUyq5/OEI6Zpu2n2Iuwi
tUXjIhf1xtH3vcXvP2BJUQKBgBF5cy0ZcCXPjzG/hQGZn4j34JnVGnhXNRbB+G3K+occs4mCFR68
Q6THzJcoz8YiYCG6BndwnoQWY+3pNmDoNZUviTzXCFspsV2l1SbtYY79lQ/ct+NyyJ5vof89rEC/
JqC7QjoSFox4gnaK+zu9OVwN3vRPGEhfbL9PVIj7YCx/AoGAEY06Coij0Jp19PShWfrTP72Wamk+
W9RHkrBS3FfNY6RZsppXY7478kAeCTtQMI5CdOcPjNQ2Tbwq9M+NMSltzA+Gi6B1YM1JPPt9XzDK
7cIg3+TqCgkwVtRxNmc+Kzg4k2H3W5IV7qvZlxbJzXFb7Xkz64HWOmQXtn3epANQ5Vw=
[mgriffit@Malachis-MacBook-Pro-2:cbw]$
```

Changing file permissions of your 'key' file

- ls -l (long listing)

```
drwx------+ 67 francis  staff  2278 22 May 21:25 ../
-rw-r--r--@  1 francis  staff  1696 22 May 21:31 CBWRNA.pem
 rwx : owner
    rwx : group
       rwx: world
 r read    (4)
 w write   (2)
 x execute (1)
```

Which ever way you add these 3 numbers, you know which integers were used (6 is always 4+2, 5 is 4+1, 4 is by itself, 0 is none of them etc …)

So, when you have:

`chmod 600 <file name>`

It is "rw" for the the file owner **only**

Logging in to AWS

Logging in

- Make sure the permissions on your certificate are secure. Use chmod on your downloaded certificate:

```
chmod 600 CBWRNA.pem
```

- To log in to the node, use the -i command line argument to specify your certificate:

Use your assigned student #

```
ssh -i CBWRNA.pem ubuntu@cbw##.ssh01.com
```

(where # is your assigned student number. Your student number is the number on the participant list. If your number less than 10, please add 0 in front of it.)

Copying files from AWS to your computer

Copying files to your computer

- To copy files from a node, use scp in a similar fashion (in this case to copy a file called nice_alignments.bam):

```
scp -i CBWRNA.pem ubuntu@cbw##.ssh01.com:workspace/nice_alignments.bam .
```

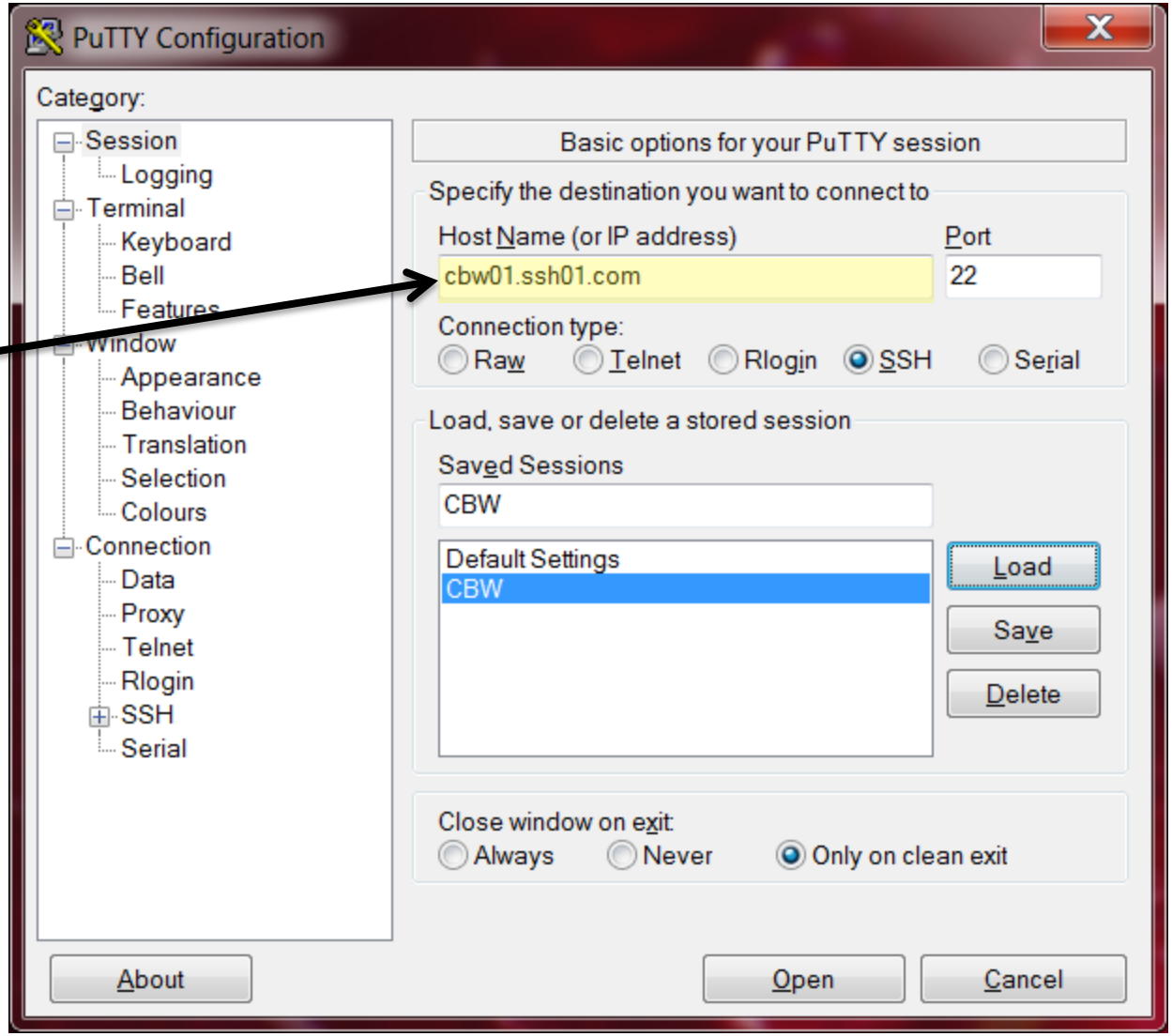
- Everything created in your workspace on the cloud is also available by a web server on your cloud instance. Simply go to the following in your browser:

```
http://cbw##.ssh01.com/
```

Windows users

PuTTY configuration for windows users

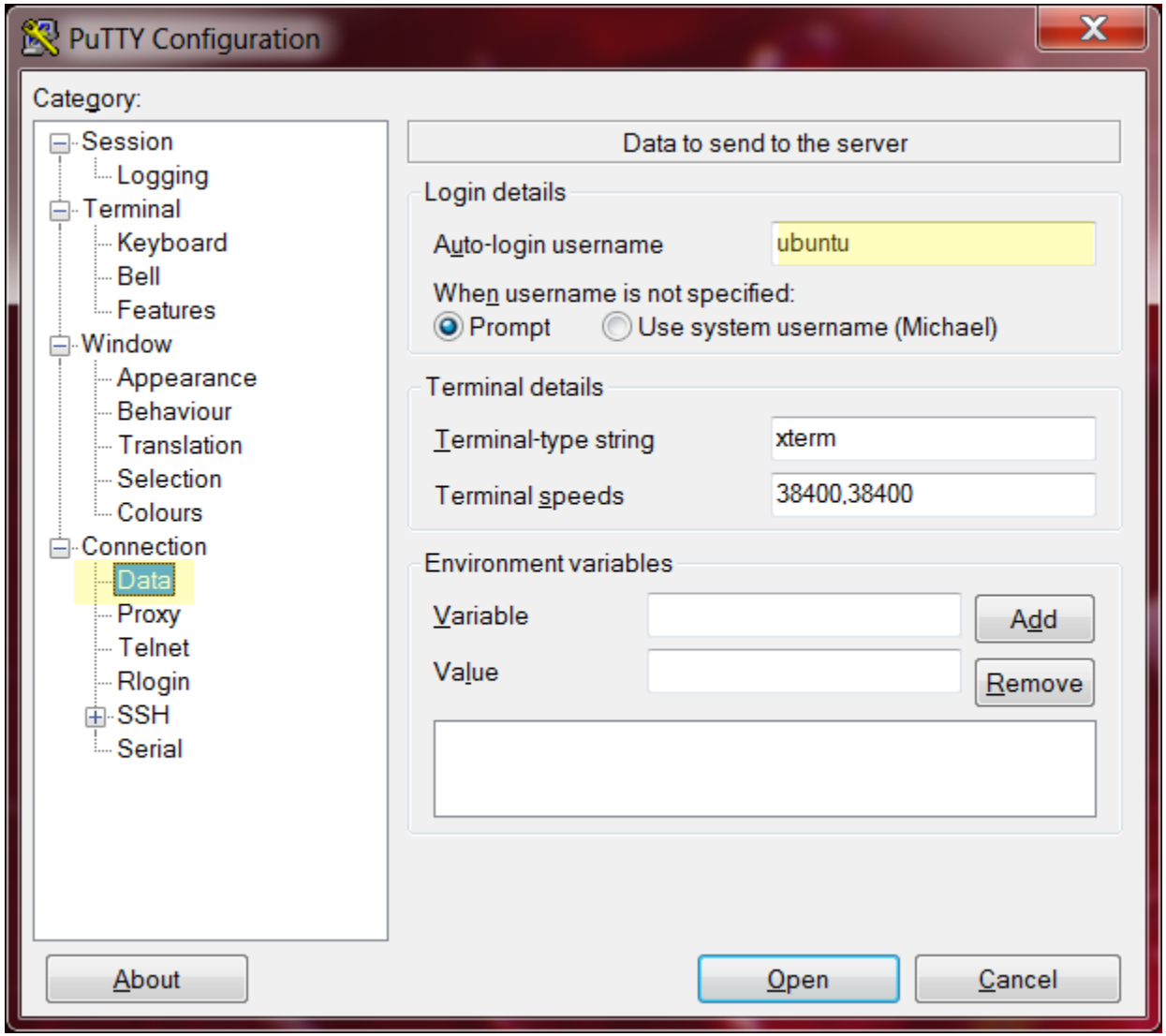
Use your assigned student #



1

PuTTY configuration for windows users

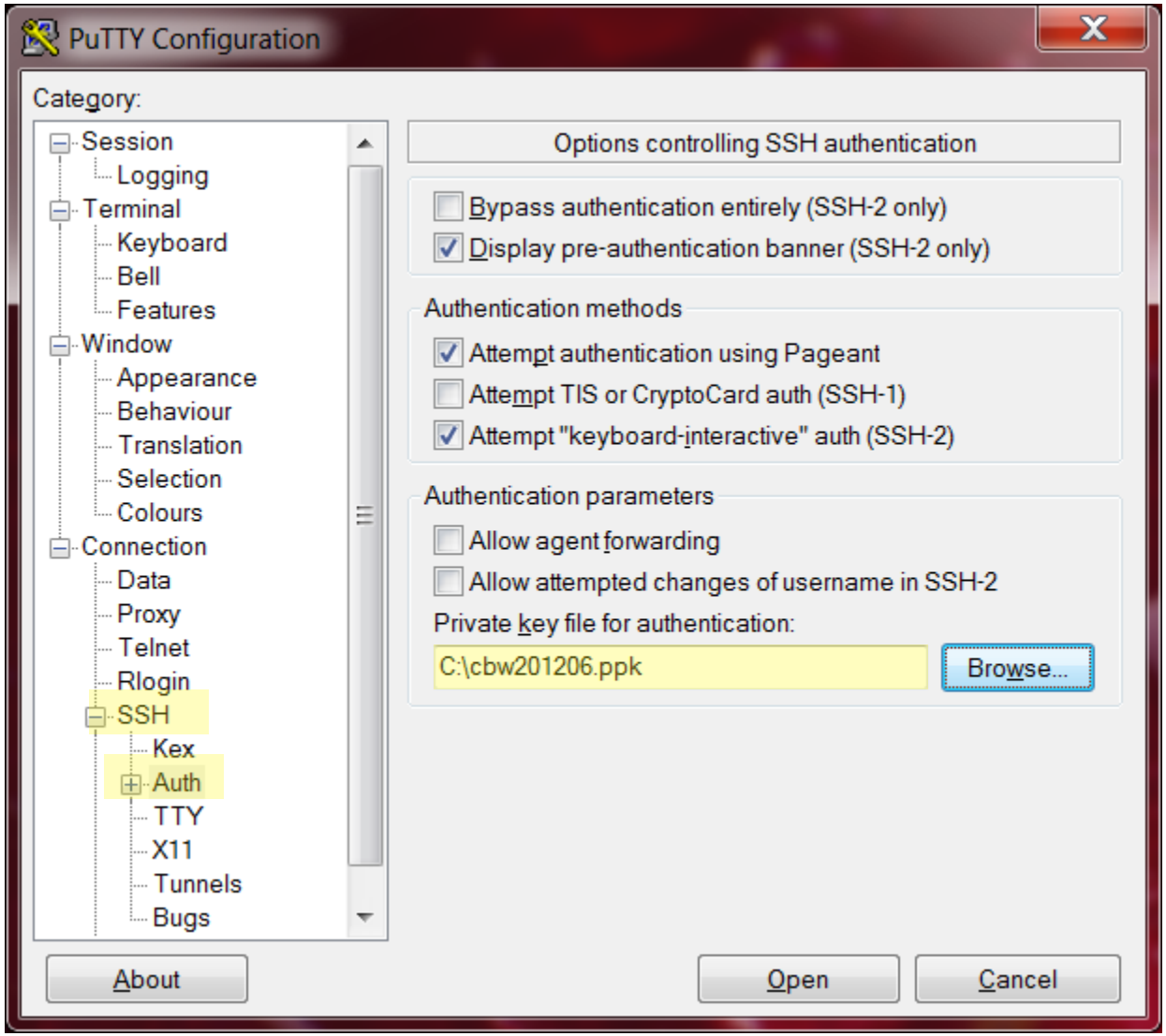
2



3

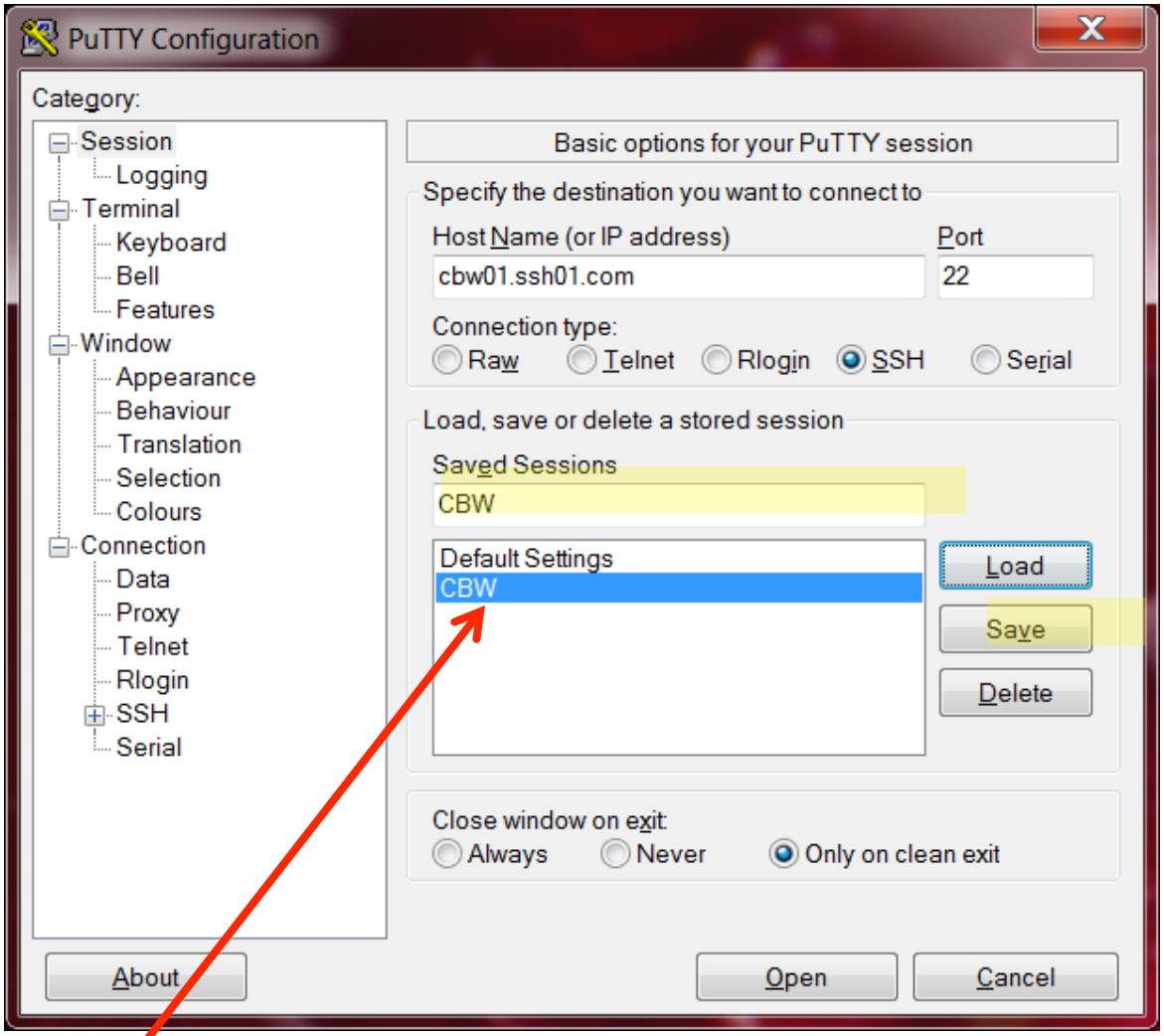
PuTTY configuration for windows users

4



5

PuTTY configuration for windows users



From now on, just double-click CBW to login.

So, at this point:

- Your laptop is ready for the workshop
- If it is not, you know where to get the information you need
- You know how to use the wiki for this workshop
- You know where all of the lectures are
- You have read all of the pre-lecture material
- If not, you know where the papers are, and you are a speed reader
- You know how to login to AWS

Break