



LEIDEN UNIVERSITY MEDICAL CENTER

Connecting to other machines

Jeroen F. J. Laros

Leiden Genome Technology Center

Department of Human Genetics

Center for Human and Clinical Genetics



Servers

Remote machines can be very handy:

- One central machine for calculation.
- Cuts expenses.
- No one wants a cluster in their office.
- Specialised software only in one place.
- Multiple users can use it at the same time.
- ...

Logging in

There are lots of ways to connect to a server.

- HTTP – When visiting websites.
- IMAP – When fetching mail.
- ...

In order to execute commands, we need to *log in*.

We use a *secure* protocol to log in.

- Most plain text protocols are blocked by firewalls.
- When working with patient data, we don't want eavesdropping.
- The connection from your machine to the server is *encrypted*.

Secure Shell

Using Secure Shell (ssh):

```
ssh user@host
```

Keyword	Explanation.
user	Your <i>username</i> on the <i>server</i> .
host	Name of the <i>server</i> .

Example:

```
ssh course@ngs.fixedpoint.nl
```

Copying data

We frequently need to transfer data before and after we do an analysis.

- The input needs to be on the server.
- The output needs to be copied back.

We also use a secure protocol to copy.

- If Secure Shell works, then this will work too (same protocol).
- Two way traffic.
 - Copy data from your machine to the server (uploading).
 - Copy data from the server to your machine (downloading).

Secure Copy

Copying something to the server:

```
scp localfile user@host:/path/to/remotefile
```

Keyword	Explanation.
<code>localfile</code>	Name of the file on <i>your</i> computer.
<code>user</code>	Your <i>username</i> on the <i>server</i> .
<code>host</code>	Name of the <i>server</i> .
<code>/path/to/</code>	Directory on the <i>server</i> to store the file.
<code>remotefile</code>	Name of the file on the <i>server</i> .

Example:

```
scp myfile  
course@ngs.fixedpoint.nl:/tmp/yourfile
```

Secure Copy

Some defaults (when left empty):

```
scp localfile user@host:/path/to/remotefile
```

Keyword	Explanation.
<code>user</code>	The username that you use on your <i>local</i> machine.
<code>/path/to/</code>	The home directory of the user on the server.
<code>remotefile</code>	The same as the name of the local file.
<code>localfile</code>	May be replaced by a “.” when copying something from the server.

Example: `scp myfile ngs.fixedpoint.nl:`
`scp ngs.fixedpoint.nl:myfile .`

Windows

Windows does not have the `ssh` command, but there are programs that give the same functionality.

PuTTY – A Free Telnet/SSH Client.

A software package containing (amongst others):

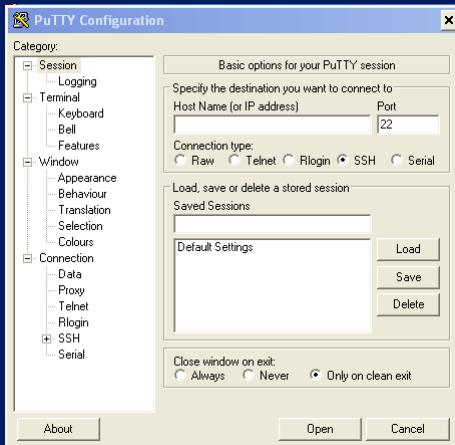
- PuTTY: Secure Shell client.
- PSCP: Secure Copy client.
- More related tools available on the website.

<http://www.chiark.greenend.org.uk/~sgtatham/putty>

Connecting from an other OS

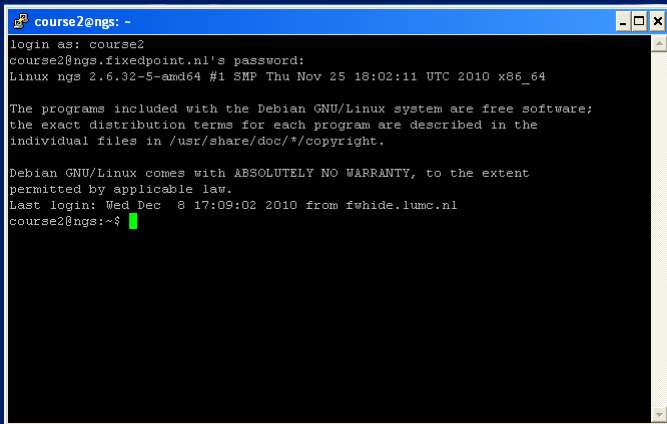
Windows

Connecting to a server using PuTTY.



Windows

A terminal when connected to a server.



```
course2@ngs: -  
login as: course2  
course2@ngs.fixedpoint.nl's password:  
Linux ngs 2.6.32-5-amd64 #1 SMP Thu Nov 25 18:02:11 UTC 2010 x86_64  
  
The programs included with the Debian GNU/Linux system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*/copyright.  
  
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent  
permitted by applicable law.  
Last login: Wed Dec  8 17:09:02 2010 from fwide.lumc.nl  
course2@ngs:~$
```

Typical workflow

When doing an analysis, the general workflow looks like this:

- First copy the input data to the server.
- Log on to the server.
- Run the analysis remotely.
- Copy the results from the server.
- Clean up the input data and the results on the server.
- Log out.

Typical workflow: an example

Step one: preparing the input.

On your machine, copy the raw data to the server, then log in on the server.

- `scp reads.fq course@ngs.fixedpoint.nl:`
- `ssh course@ngs.fixedpoint.nl`

Now the file `reads.fq` is available on the server.

Typical workflow: an example

Step two: The analysis.

On the server, you can do an analysis.

- `bwa aln ./indexes/chr17.fa reads.fq > reads.sai`
- `bwa samse ./indexes/chr17.fa reads.sai reads.fq > reads.sam`
- `samtools view -bt ./indexes/chr17.fa -o reads.bam reads.sam`
- `samtools sort reads.bam reads.bam.sorted`
- `samtools pileup -vcf ./indexes/chr17.fa reads.bam.sorted.bam > reads.pileup`

Typical workflow: an example

Step three: Retrieving the output.

Copy the output from the server back to your own machine.

- `scp course@ngs.fixedpoint.nl:reads.pileup .`

Step four: Cleaning up.

Clean up on the server and leave.

- `rm reads.*`
- `logout`

Why remote servers?

Clusters



Clusters

Massive parallel computing.

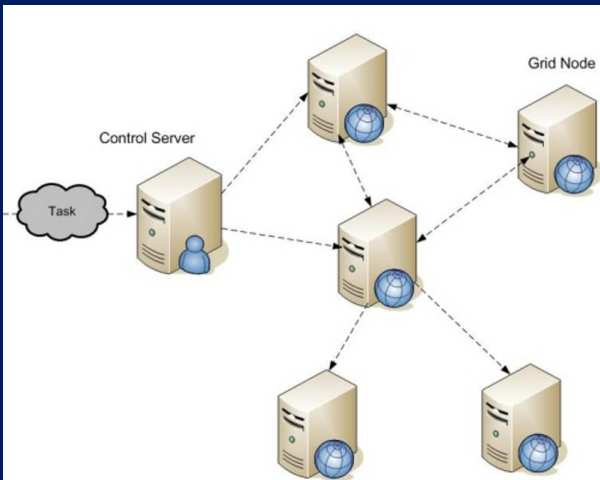
- A large number of computers working together.
- Analyse lots of samples at the same time.
- Sometimes a way to reduce memory requirements (if the problem permits it).
- Very suitable for NGS, especially alignment.

Cons:

- Not all problems are suitable for parallel computation.
- Programs must be adjusted to make use of a cluster.
 - Chop the problem up in parts / combine the results.

Why remote servers?

Clusters



Clusters

General characteristics of a cluster.

- Jobs are submitted to a *control node*.
- The control node dispatches a job to a free *worker node*.
- Jobs are monitored.
 - If a worker node doesn't finish for some reason, the job gets dispatched to an other worker node.
 - If all worker nodes are finished, the control node can alert the user that his jobs are finished.
- Jobs can be prioritised.
- ...



Michiel van Galen
Jeroen Laros

<https://humgenprojects.lumc.nl/trac/GAPSS3/wiki/course>