Course computing infrastructure

José-Maria Carazo

Biocomputing Unit, CNB-CSIC, Madrid

Instruct Image Processing Center, Madrid

Overview

5 brand new PC's have been presented to you with basic software and data.

However, for "heavy computation", we will remotely use resources at SurfSara

(Amazon computing services will also be introduced)

Each of you will have remote access to a SurfSata node with:

96 GB of memory + 16 vCPUs + 2 top GPU (Tesla K40)

12 TB storage

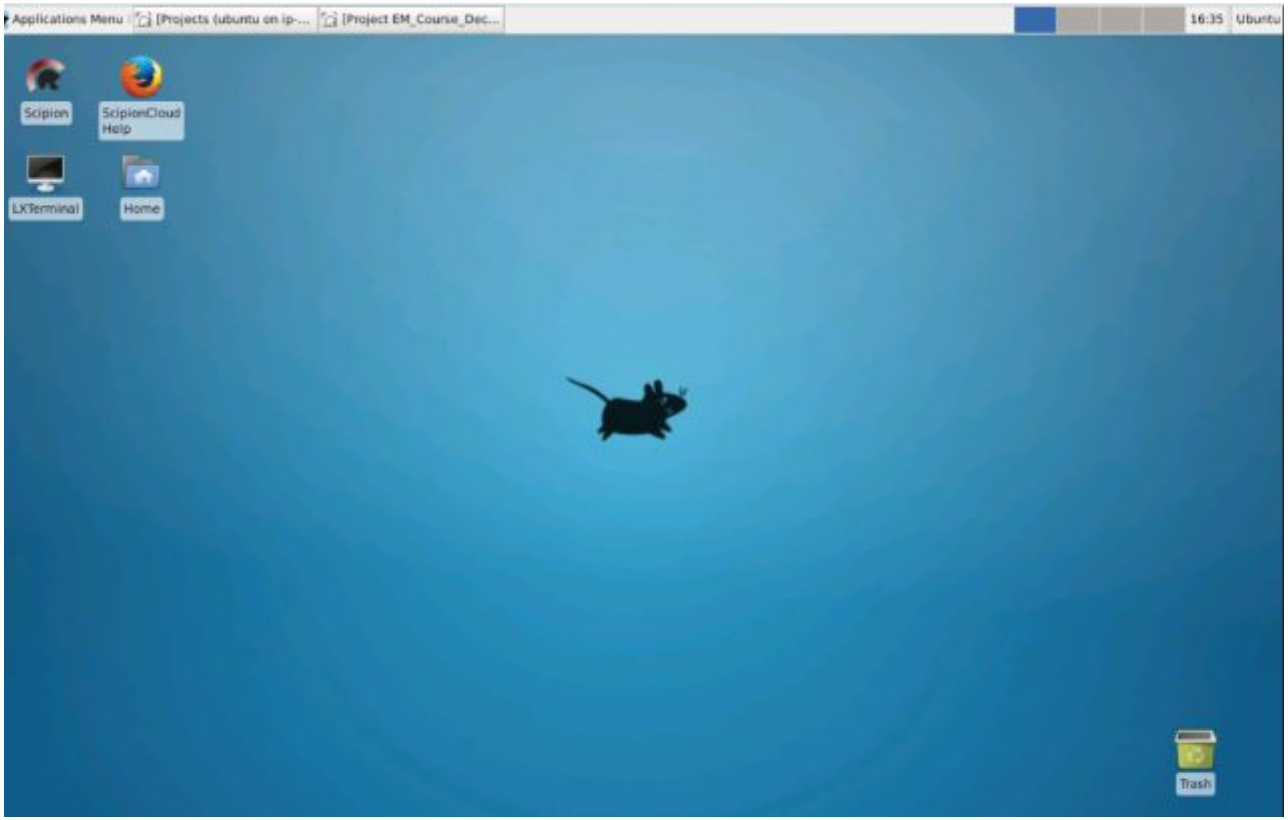
CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

What we wanted….. Cloud computing



https://54.194.70.152/scipion
Usr: scipion
Pass: awJX3rRVqe6eem

CSIC
instruct Integrating Biology
instruct image Processing Center

What we wanted….. Cloud computing

ScipionCloud: An integrative and interactive gateway for large scale cryo electron microscopy image processing on commercial and academic clouds

Jesús Cuenca-Alba[1,*], Laura del Cano[1,+], Pablo Conesa Mingo[1], José Miguel de la Rosa Trevín[1], Josué Gómez Blanco[1], and Jose-María Carazo[1]

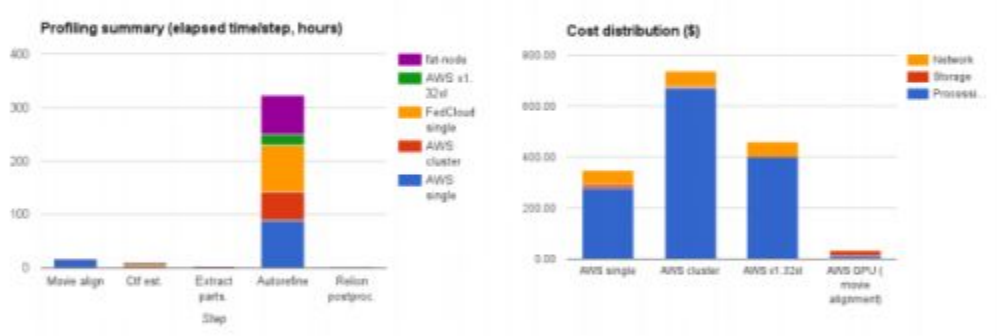
[1]Centro Nacional de Biotecnología (CNB-CSIC), Cantoblanco, Madrid, Spain, 28049

[*]Correspondence: jcuenca@cnb.csic.es (J.C.)

[+]Co-first author

September 30, 2016

What we wanted….. Cloud computing



Environment	Instance	vCPU	GPU	RAM(GB)	Cost ($/hour)
Amazon AWS	g2.2xlarge	8	1 Nvidia GRID K520 (4 GB VRAM)	16	0.702
	m4.4xlarge	16	-	64	1.056
	r3.8xlarge	32	-	244	2.66
	x1.32xlarge	128	-	1952	16.006
FedCloud	universe	40	-	240	-
Local	fat-node	32	-	512	-

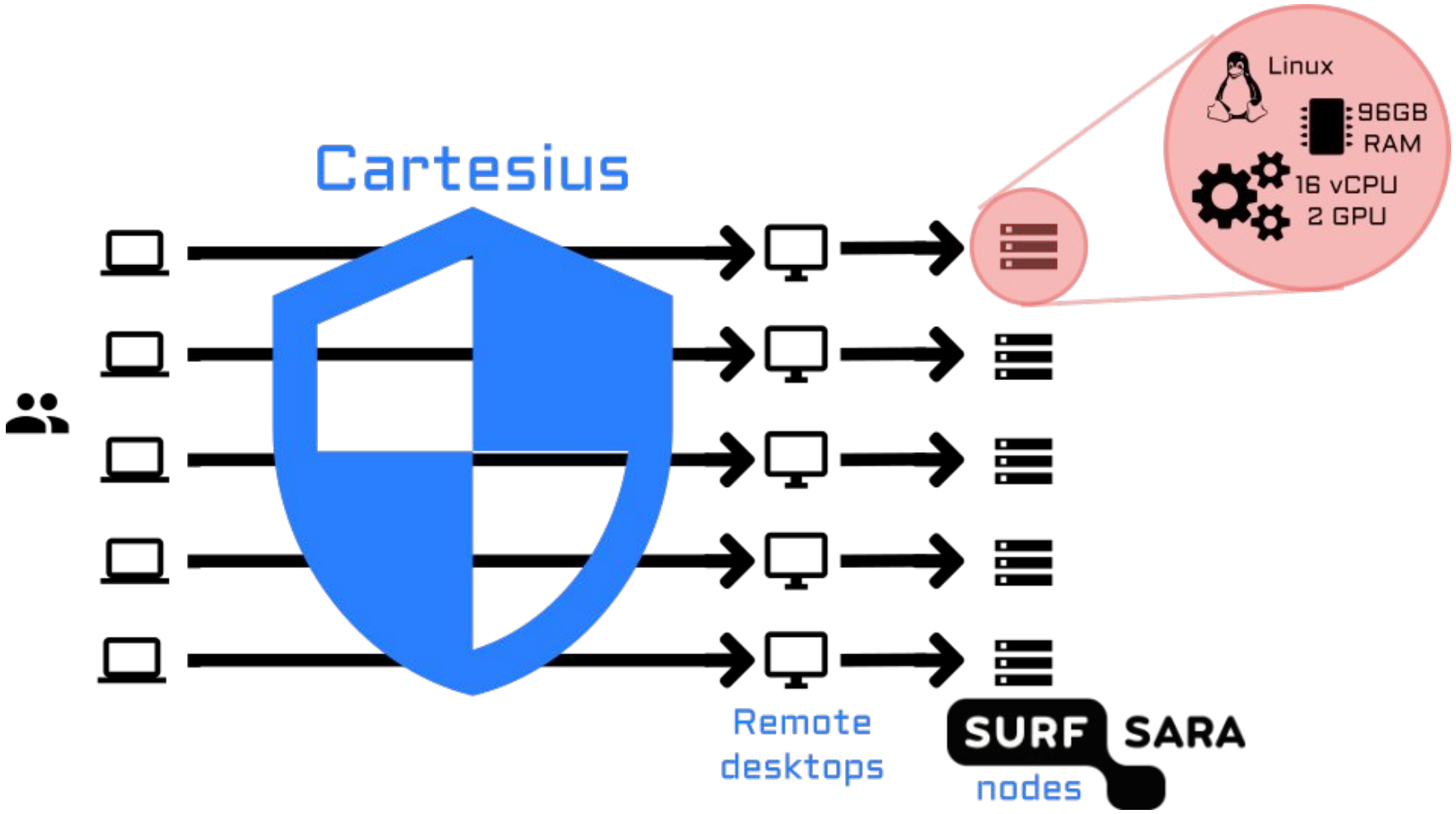
CSIC

instruct
Integrating
Biology

BIO COMPUTING Unit

instruct
image
Processing
Center

Overview (GPU forced)



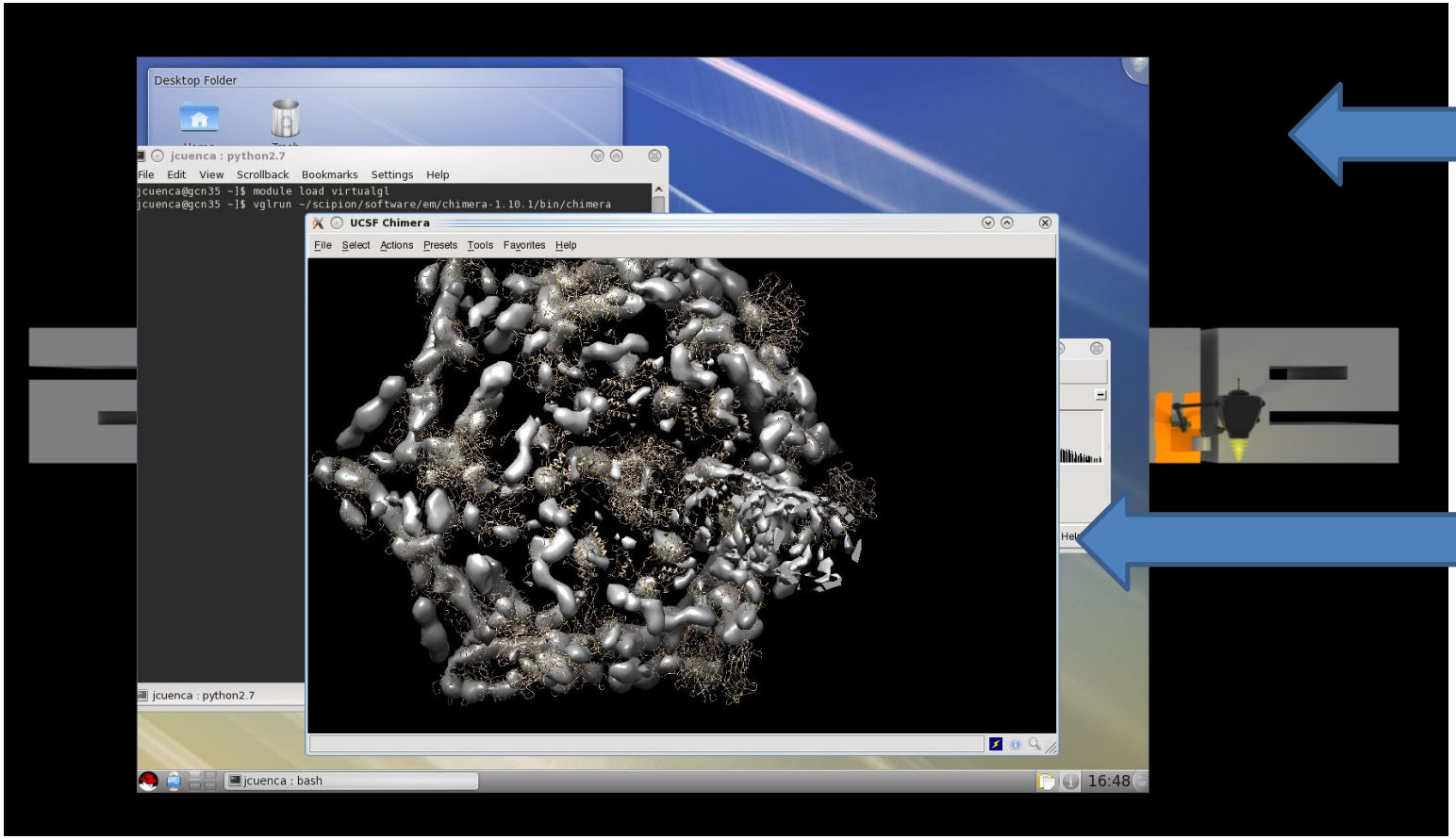
CSIC

instruct
Integrating
Biology

BO COMPUTING Unit

instruct
image
Processing
Center

The reality (GPU-forced): Remote Desktop Introduction



Local Desktop (running on Your laptop)

Remote Desktop (running at Surfsara)

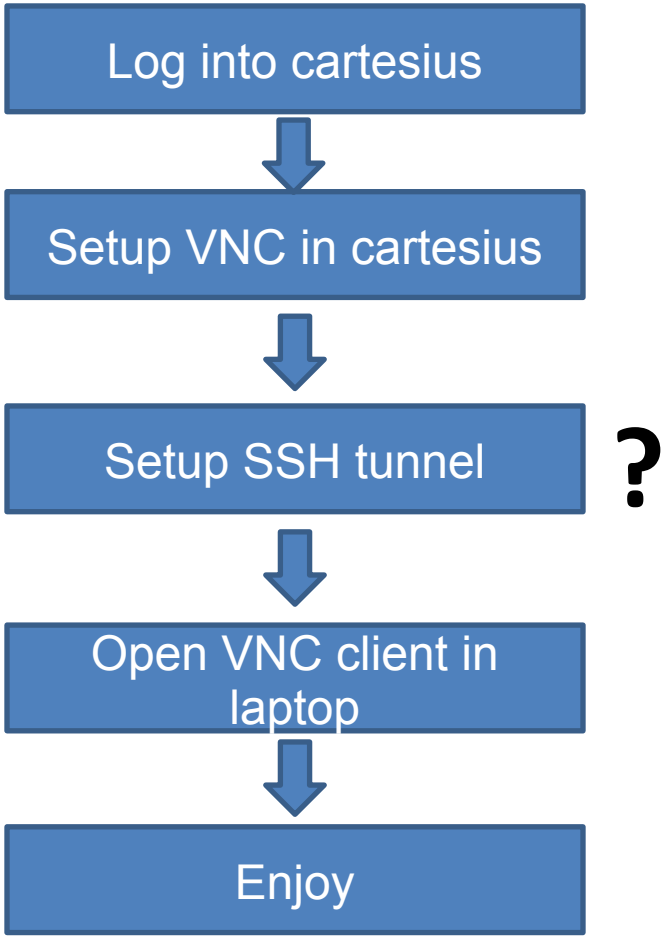
CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

Using Remote Desktop / Summary



CSIC instruct Integrating Biology

BO Computing Unit instruct image Processing Center

Introducing SSH tunnels

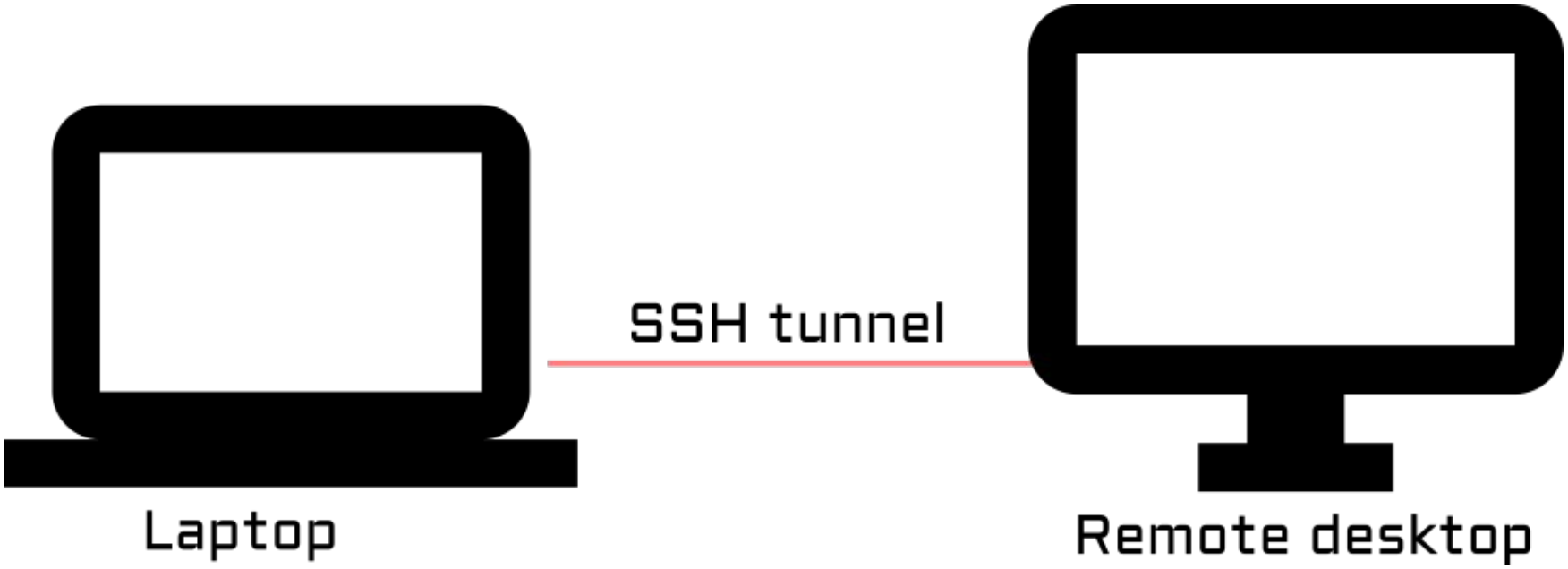
The TV metaphor

Imagine the antenna plug is too far away from your TV.

What to do?

Use an extension cable that connects the plug to your TV

Introducing SSH tunnels



1. Remote Desktop Setup / log into Surfsara

[in your **LAPTOP**] log in to Cartesius:

ssh **emstudXX**@cartesius.surfsara.nl

Now your are in CARTESIUS. You are going to setup remote visualization session for 5 days (120 hours) with the gnc_vnc (*see next slide*)

The default desktop expiration is 5 days, so steps 1&2 should only be neccessary every monday.

2. Remote Desktop Setup / gcn_vnc

gcn_vnc -p 120:00:00

```
[jcuenca@int2 ~]$ gcn_vnc -p 24:00:00                                    [85/1902

Note: format of -p <walltime> string "24:00:00" not recognized, passing value directly to sbatch as "-t 24:00:00"

Reserving a private GPU node (one of gcn2-66)

SLURM job ID is 2762025, waiting for VNC server to start running

We got assigned node gcn35

Waiting until VNC server is ready for connection...

VNC server gcn35:1 ready!

An SSH tunnel needs to be set up from the machine where you want to
run the VNC client, as the started VNC server isn't directly reachable
from outside of SURFsara.

There are two options to accomplish this:

OPTION 1 - MANUAL TUNNEL SETUP
------------------------------

On Linux and MacOS X:

  ssh -L 5901:gcn35:5901 jcuenca@vis.cartesius.surfsara.nl

On Windows:

  plink.exe -L 5901:gcn35:5901 jcuenca@vis.cartesius.surfsara.nl

The VNC server is then reachable as localhost:1, e.g. with

  vncviewer localhost:1
```

Copy this line (from your own terminal) to your clipboard

(obviously, it will differ from this example)

3. Remote Desktop Access / SSH tunnel

[**in your laptop**] setup a SSH tunnel

Paste the line you copied from gcn_vnc output

into a terminal of your laptop

ssh -L 5901:**my_cartesius_node**:5901 **emstudXX**@vis.cartesius.surfsara.nl

Then, get your **VNC password** from the file named vncpass:

cat **vncpass**

Note again that your particular command will be different from this example (cartesius node, mylogin…)

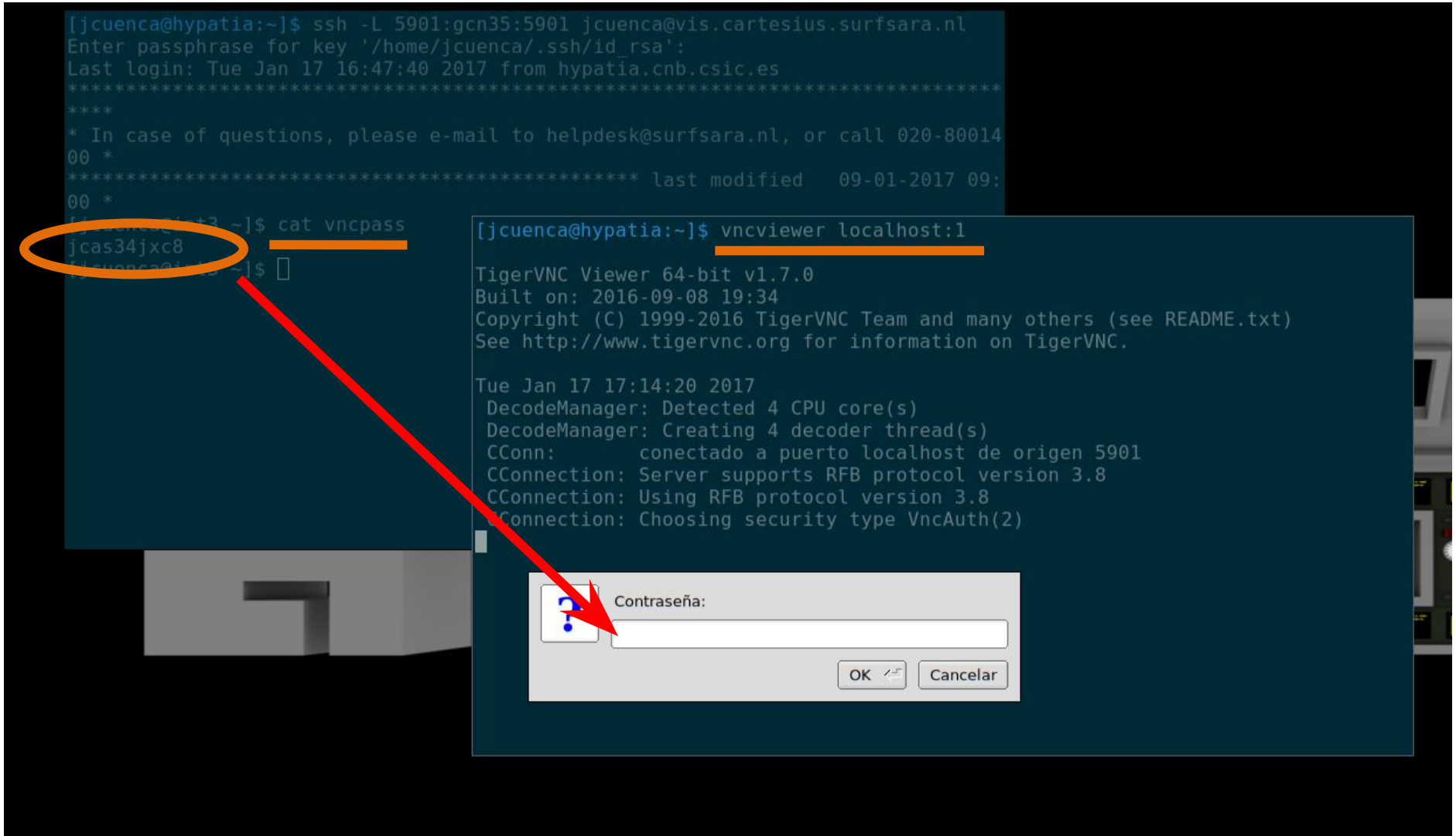
4. Remote Desktop Access / VNC

Open a new terminal **in your laptop**. You will run the VNC command in it (see next slide)

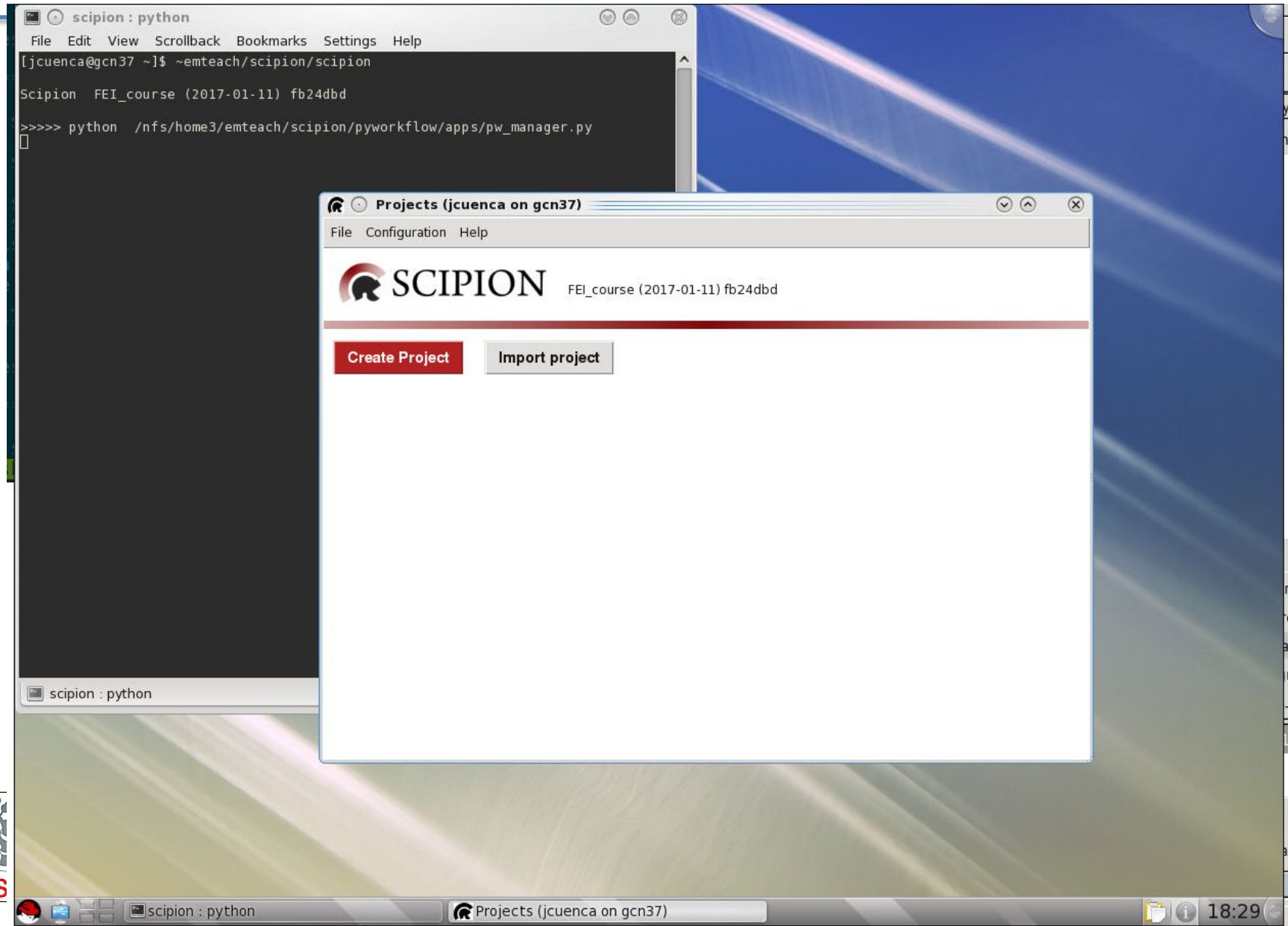
When the command is run, a dialog will appear asking for your VNC password. Use the password that you obtained from file vncpass

4. Remote Desktop Access / VNC

vncviewer localhost:1



5. Remote Desktop / Enjoy!



Data

Surfsara Cartesius offers different spaces to store information.

When you log in to cartesius, you appear in your **home** directory, which is space & performance **limited**.

From this home, you can access a directory called **big**, where you should collect all your data and results.

The directory **dropbox** can simplify your file transfers: in cartesius, you can copy the files to dropbox directory and then sync this directory to your laptop

CSIC instruct Integrating Biology BIO COMPUTING Unit instruct image Processing Center

Data transfer

During the course, you may need to
download files from Cartesius to your laptop, or
upload files from your laptop to Cartesius

We recommend using a simple command-line tool, called rsync

The basic syntax of rsync is simple:

1. rsync -av --progress
2. SOURCE
3. DESTINATION

Data transfer / Download **from** cartesius

How to refer to the directory dropbox in the account emstud04 of the computer cartesius at surfsara.nl ?

My laptop

~

Cartesius.surfsara.nl

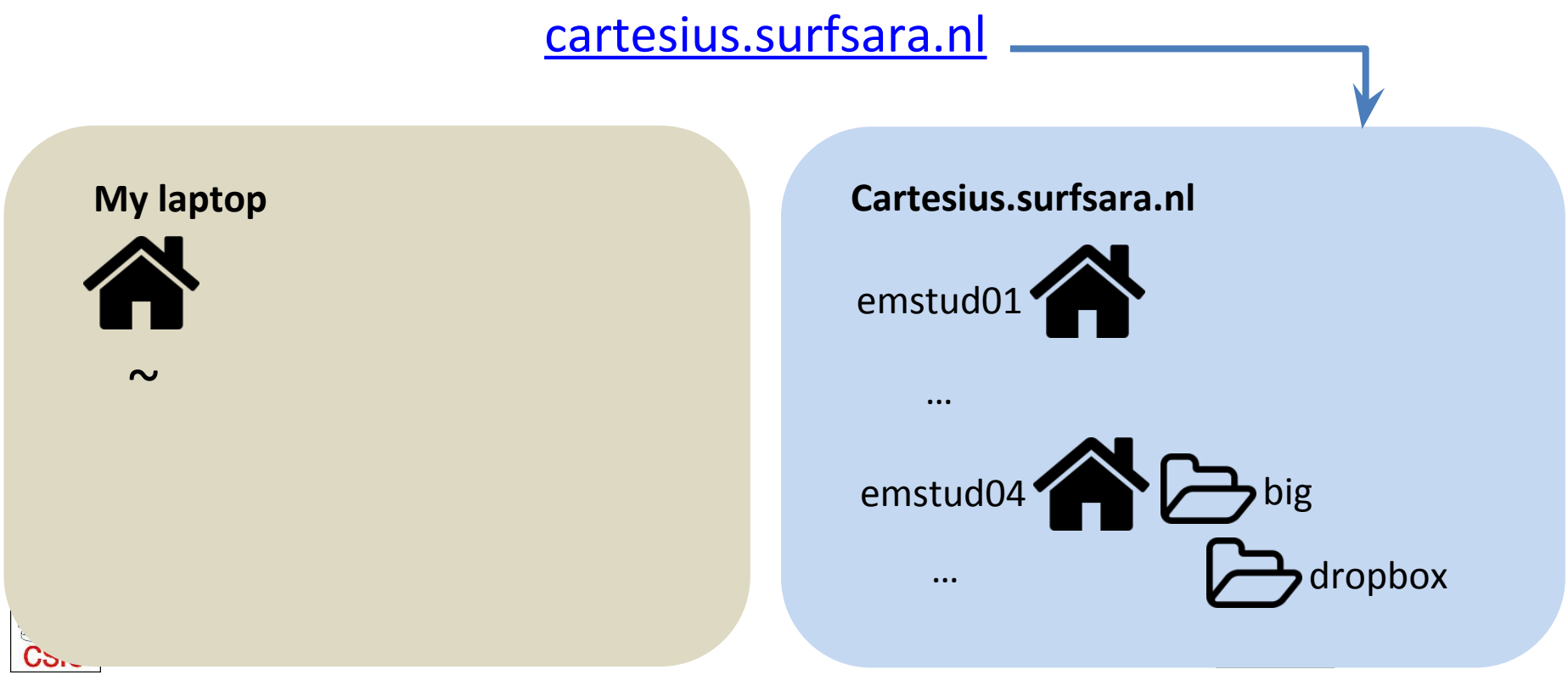
emstud01 

…

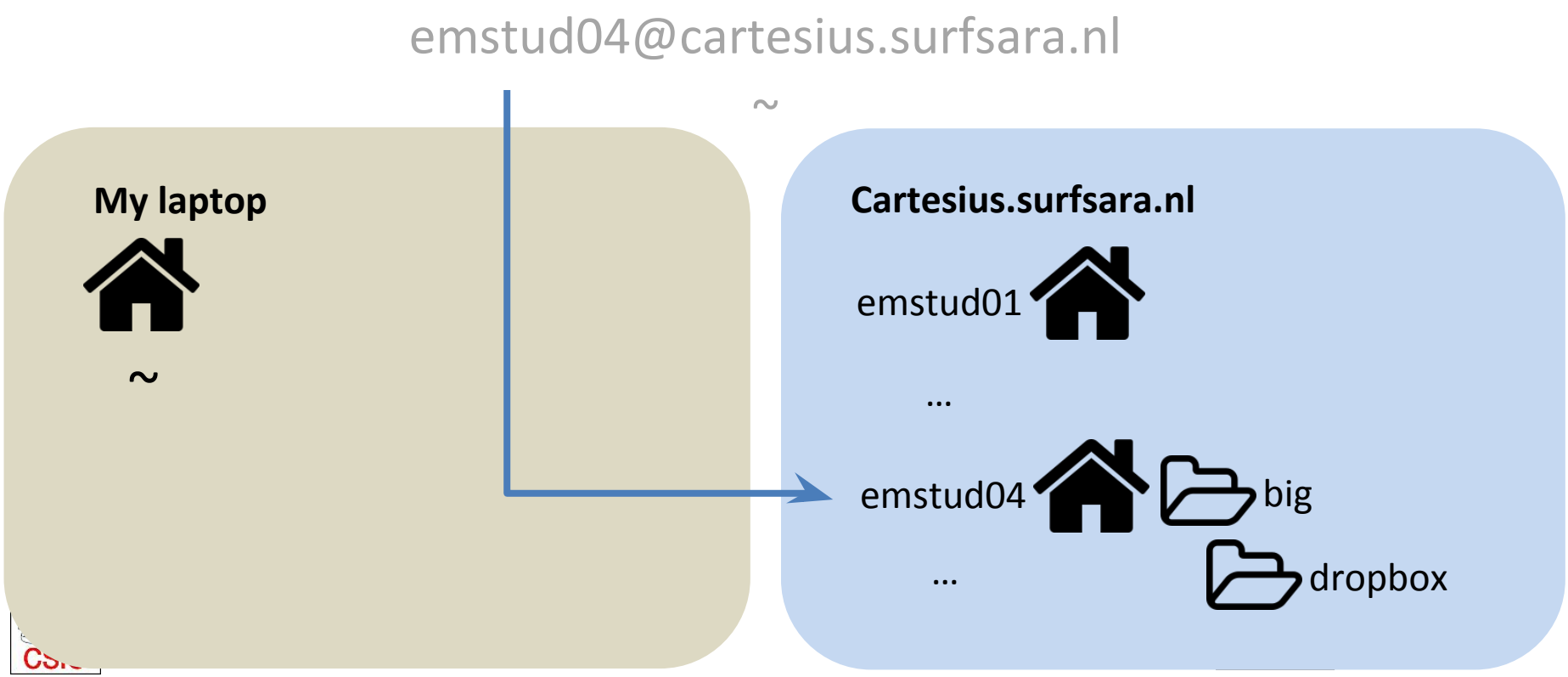
emstud04   big

…  dropbox

Data transfer / Download **from** cartesius

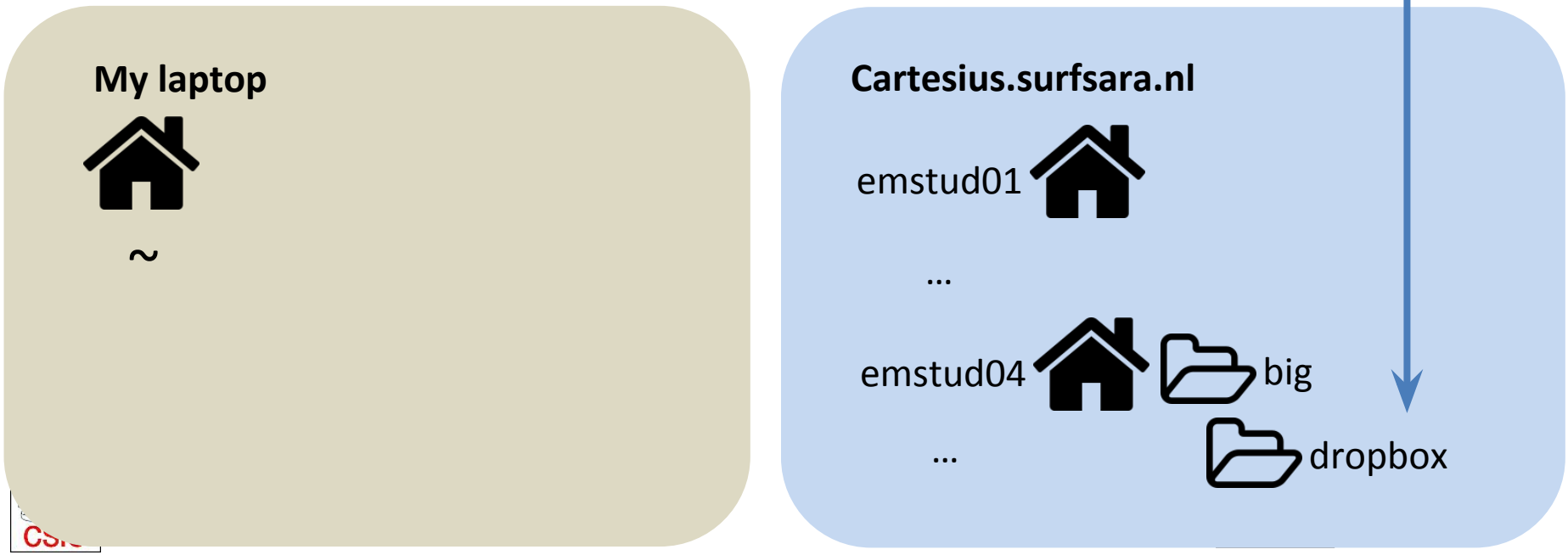


Data transfer / Download **from** cartesius



Data transfer / Download **from** cartesius

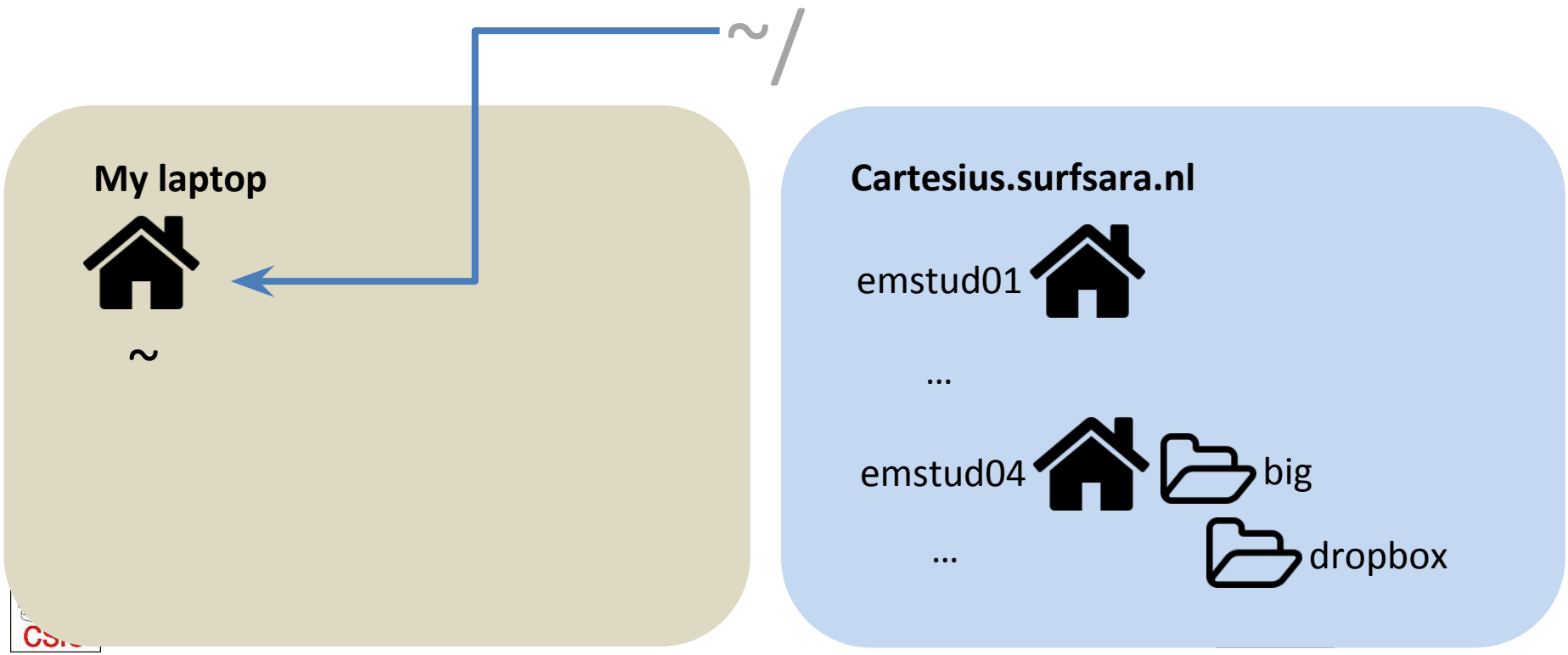
emstud04@cartesius.surfsara.nl:big/dropbox



Data transfer / Download **from** cartesius

My home directory

at my laptop: tilde symbol



Data transfer / Download **from** cartesius

General rsync syntax:

rsync -av –progress

SOURCE

DESTINATION

My laptop

~

Cartesius.surfsara.nl

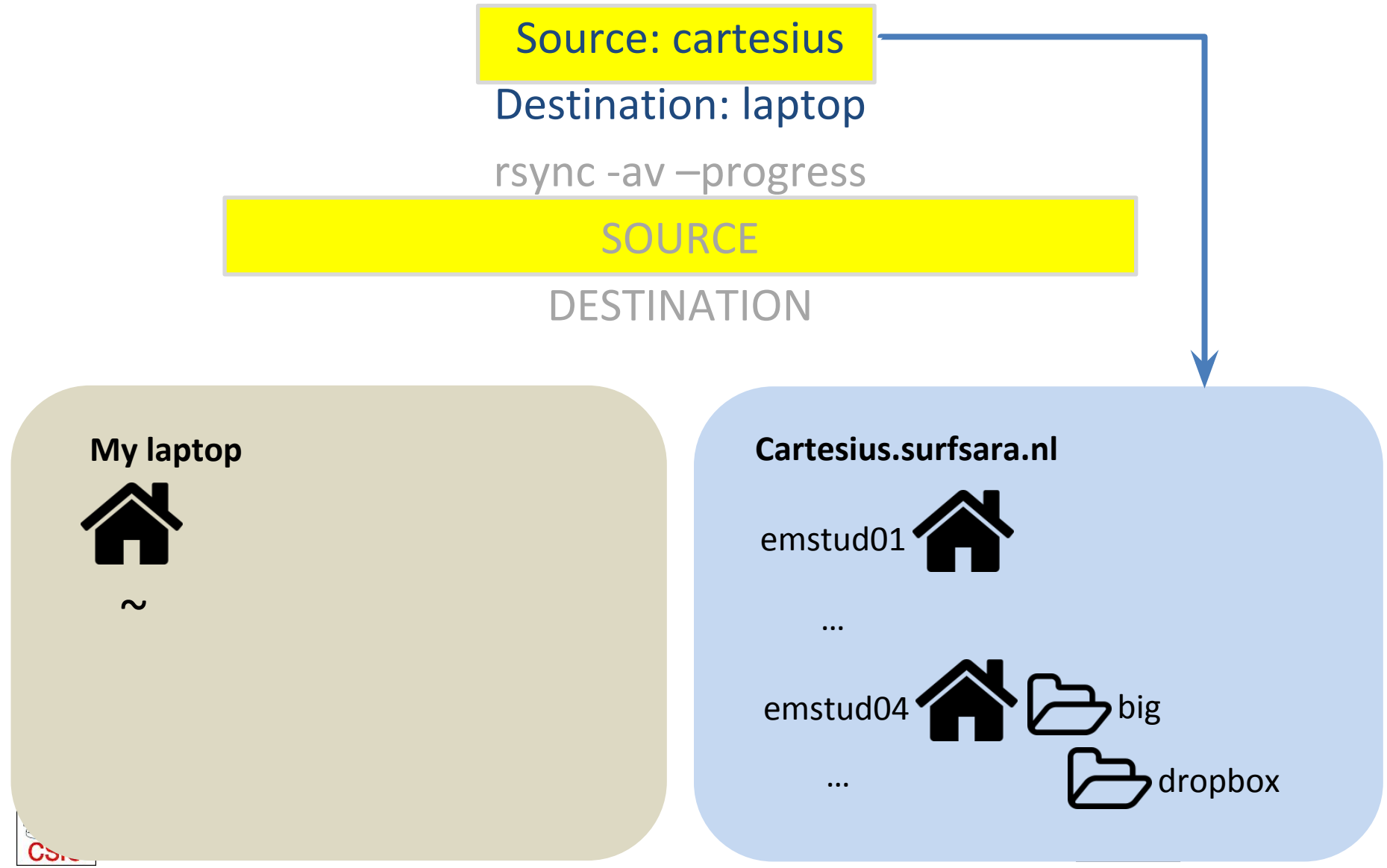
emstud01 

…

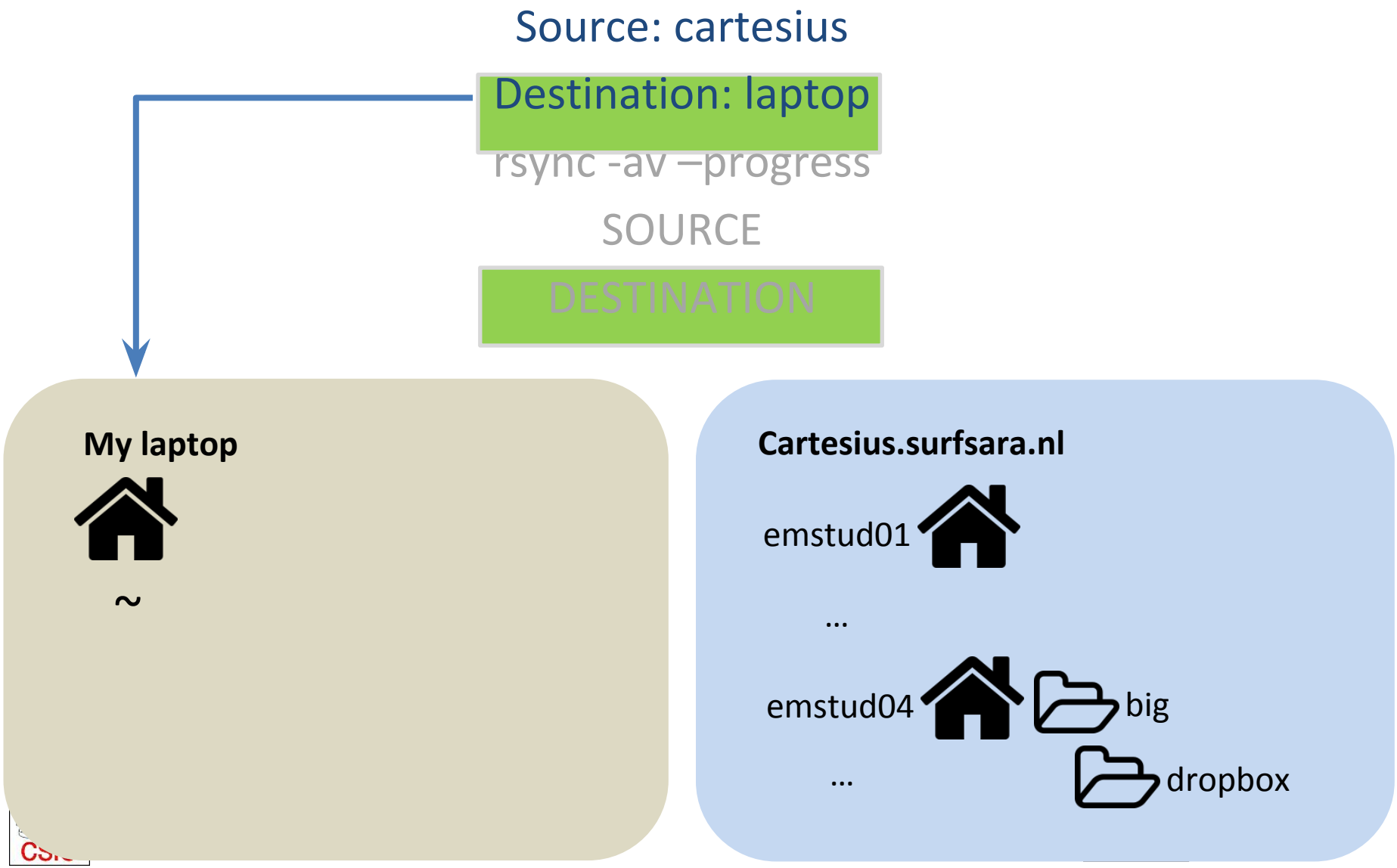
emstud04   big

…  dropbox

Data transfer / Download **from** cartesius



Data transfer / Download **from** cartesius



Data transfer / Download **from** cartesius

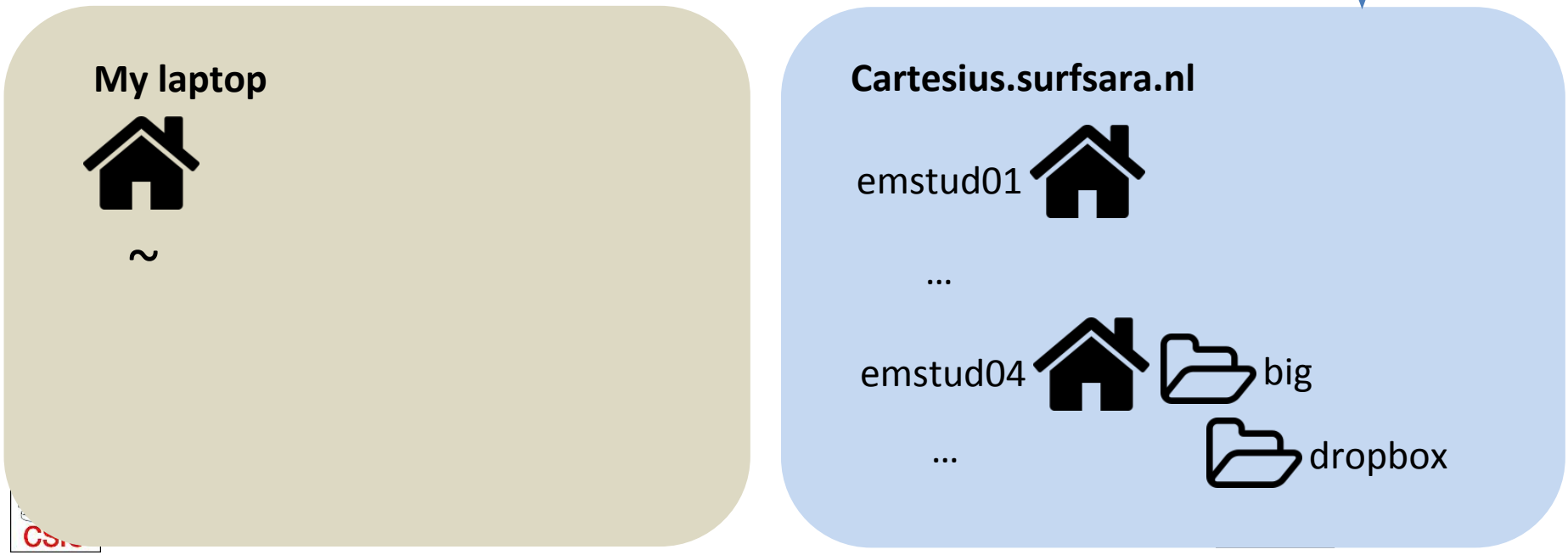
Source: cartesius

Destination: laptop

rsync -av –progress

emstud04@cartesius.surfsara.nl:big/dropbox

~/



Data transfer / Download **from** cartesius

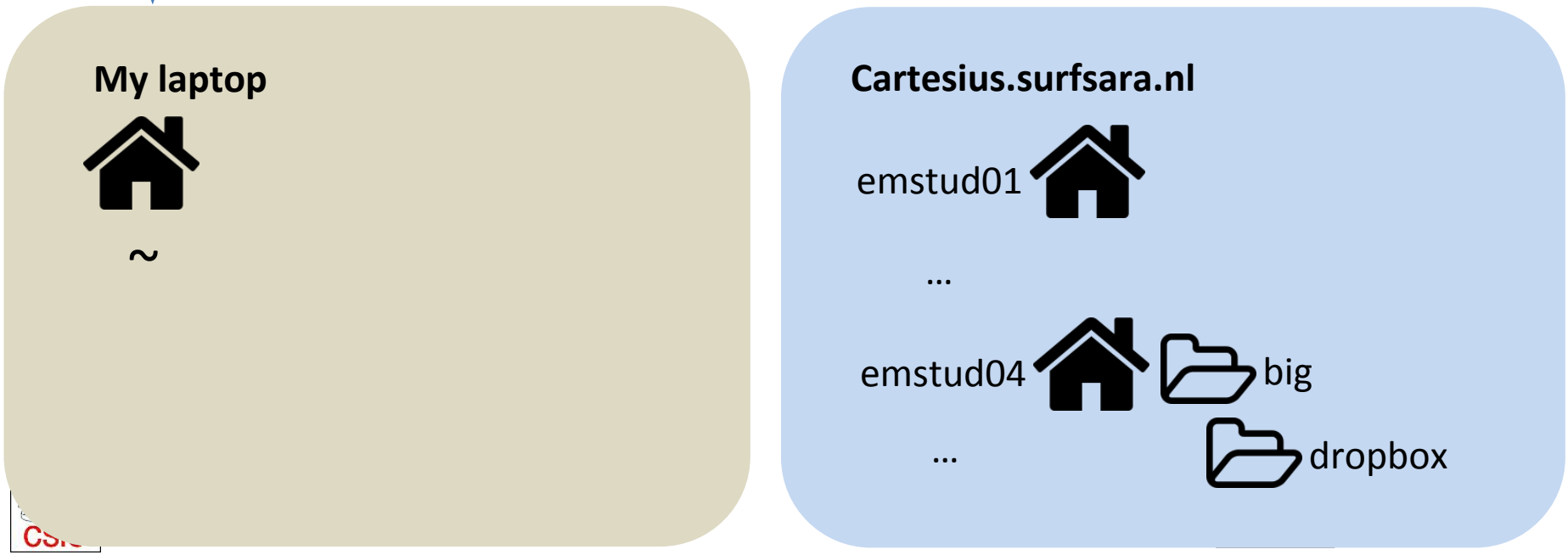
Source: cartesius

Destination: laptop

rsync -av –progress

emstud04@cartesius.surfsara.nl:big/dropbox

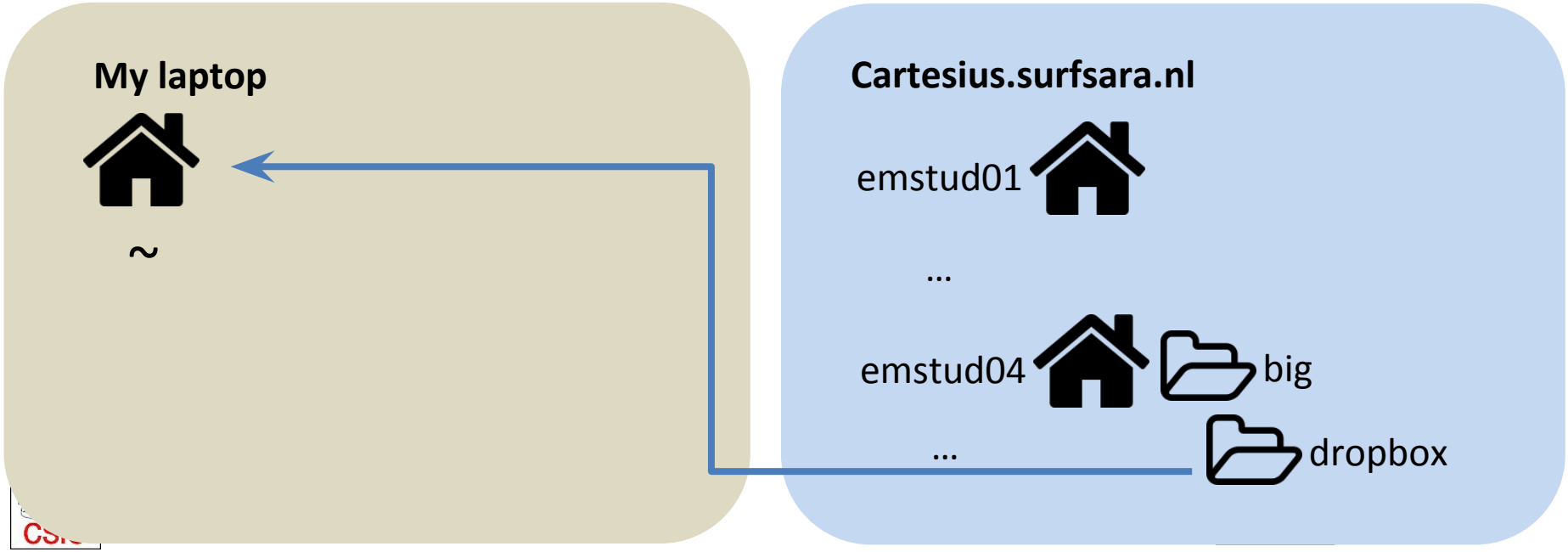
~/



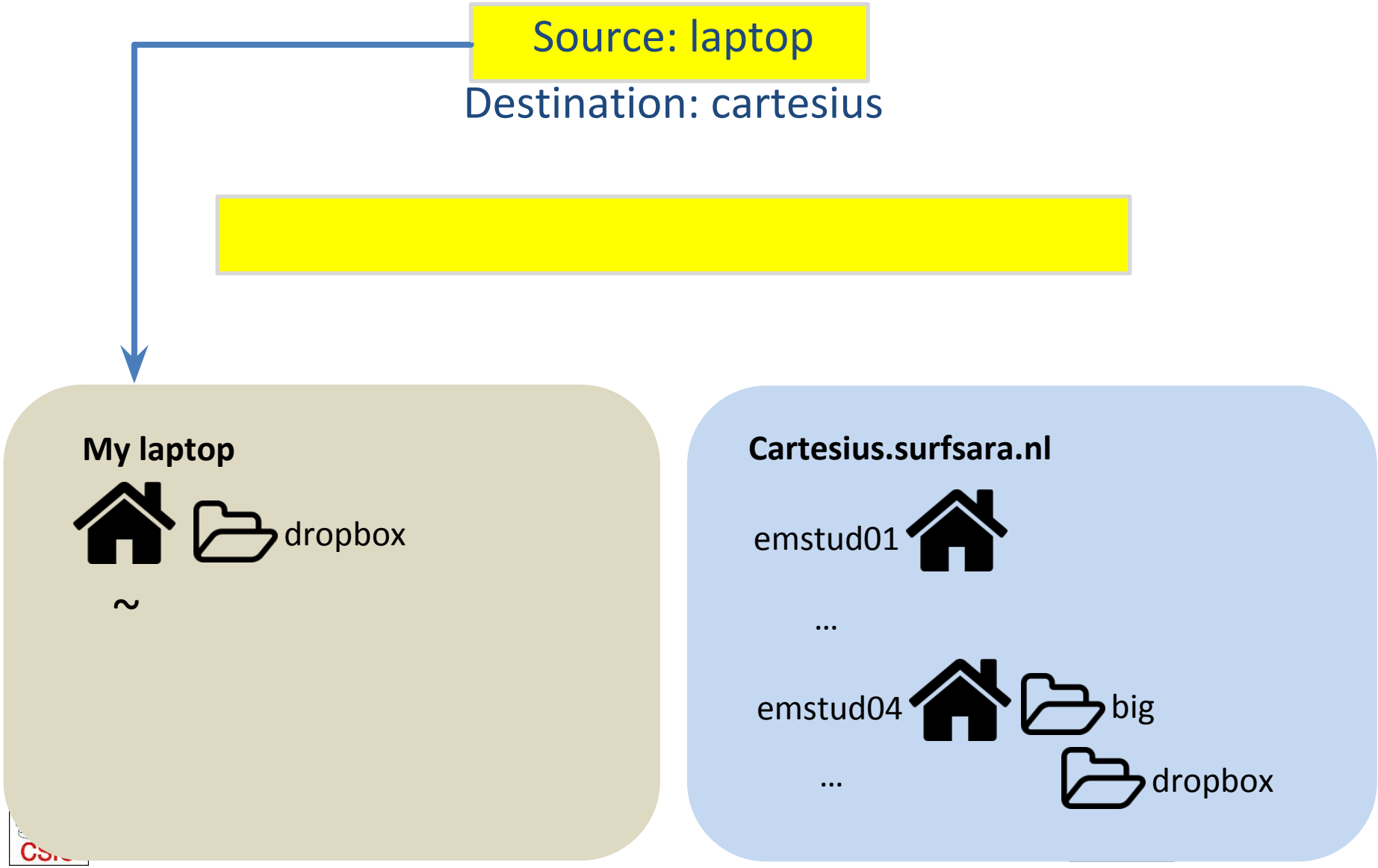
Data transfer / Download **from** cartesius

Then, the command is…

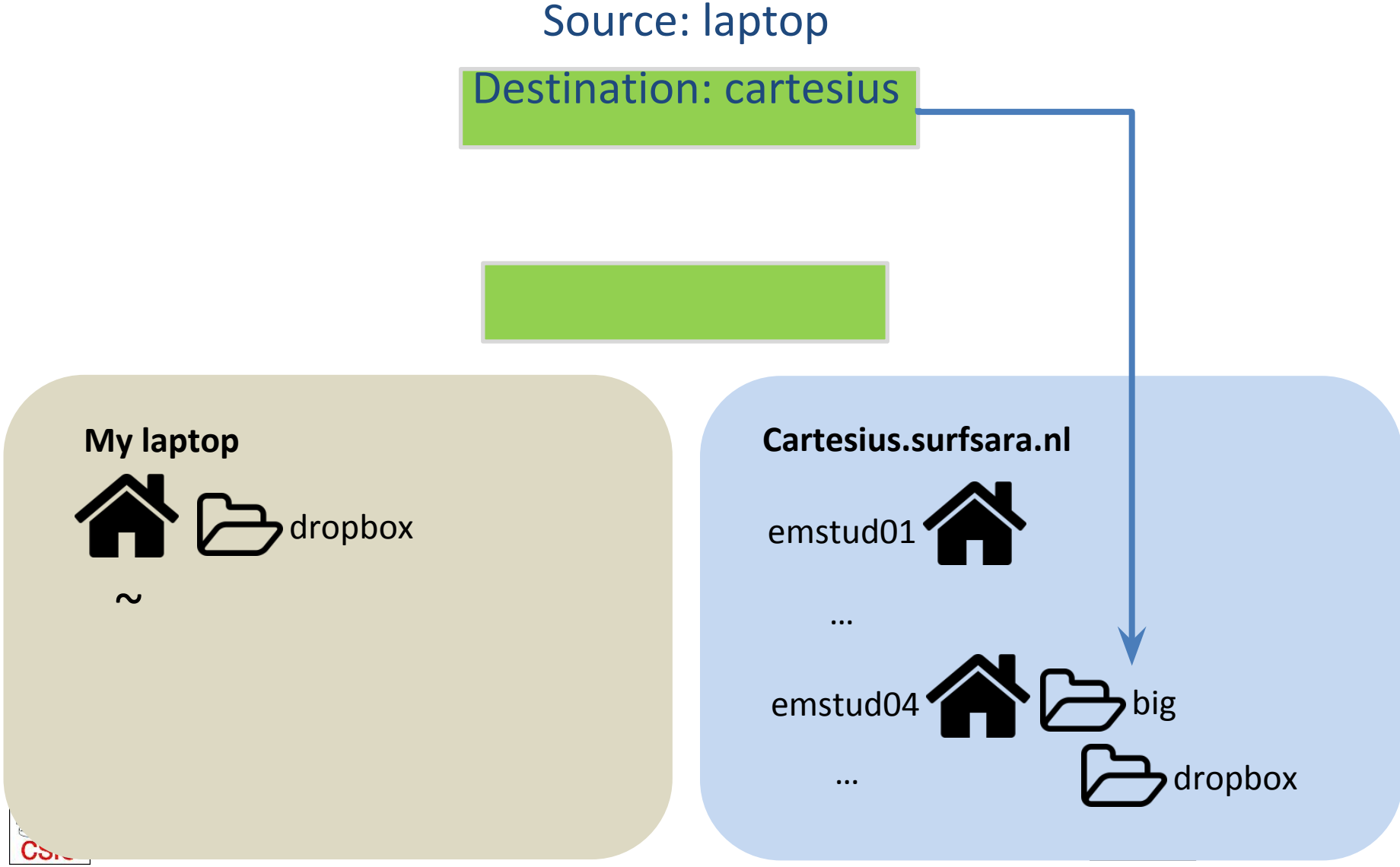
rsync -av --progress emstud04@cartesius.surfsara.nl:big/dropbox ~/



Data transfer / Upload **to** cartesius



Data transfer / Upload **to** cartesius



Data transfer / Upload **to** cartesius

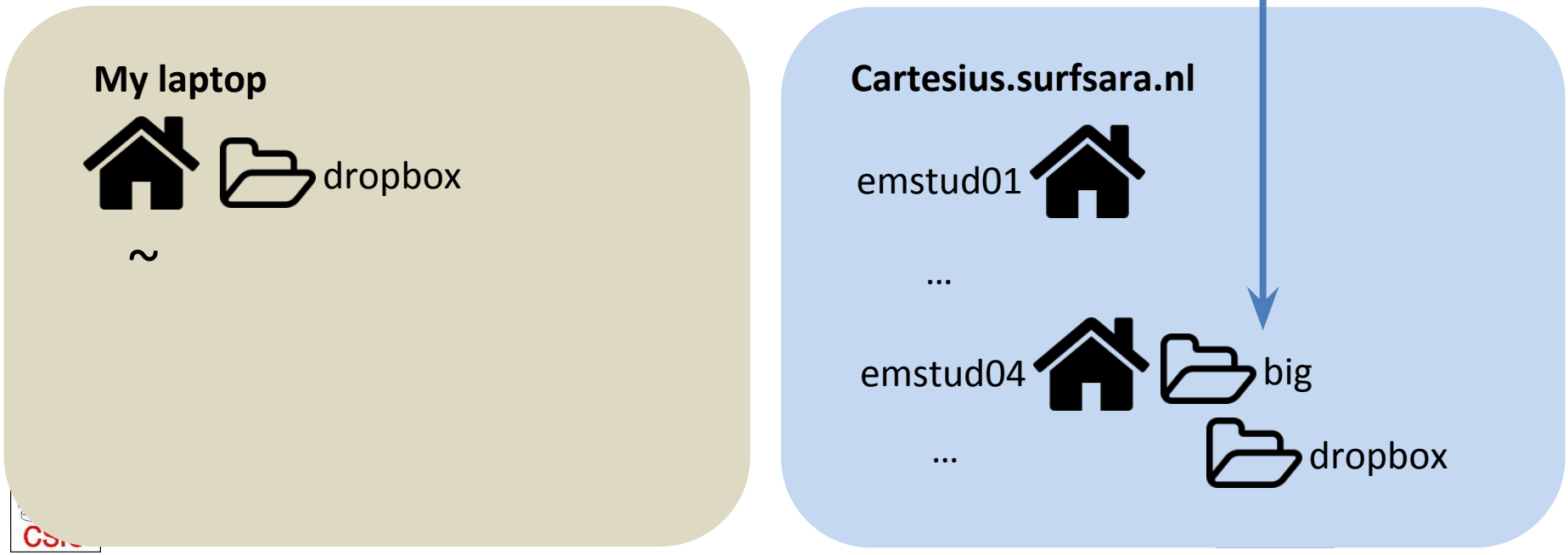
Source: laptop

Destination: cartesius

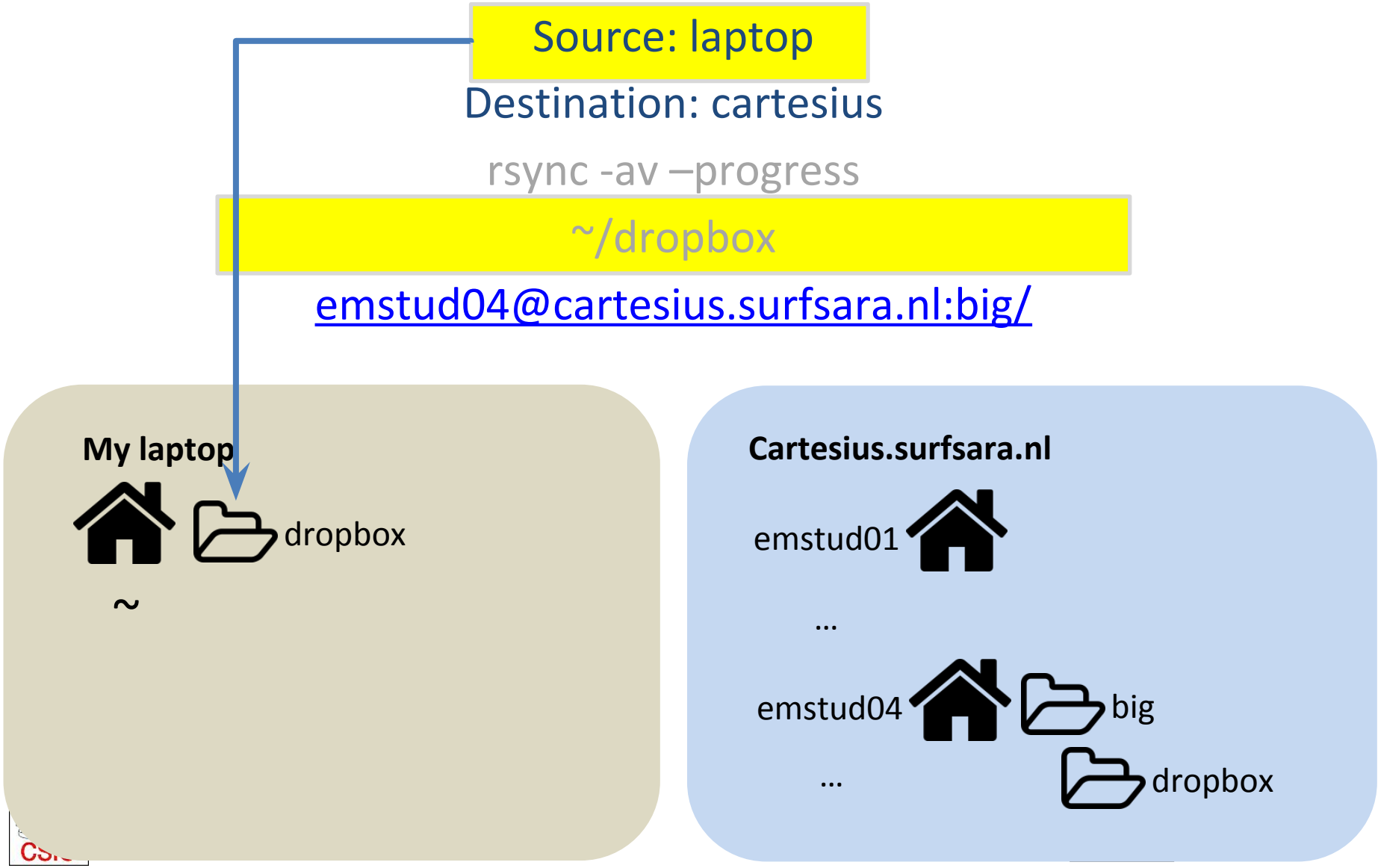
rsync -av –progress

SOURCE

emstud04@cartesius.surfsara.nl:big/



Data transfer / Upload **to** cartesius

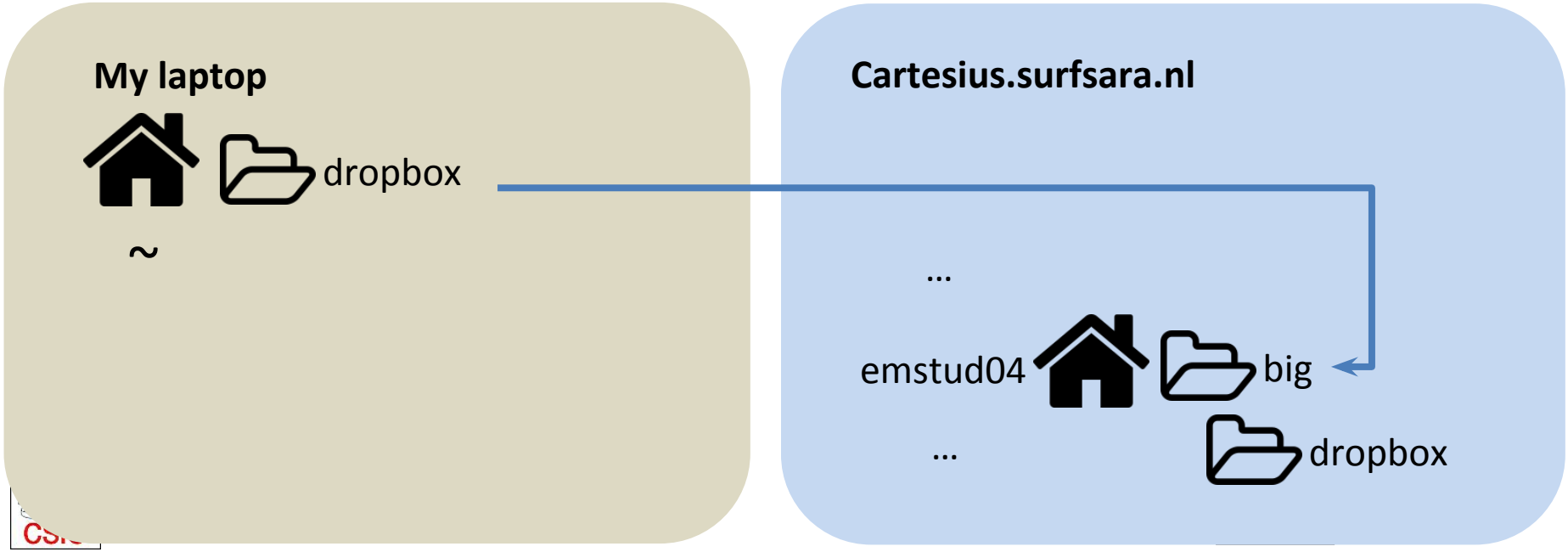


Data transfer / Upload **to** cartesius

Then, the command to upload directory dropbox to directoy big is…

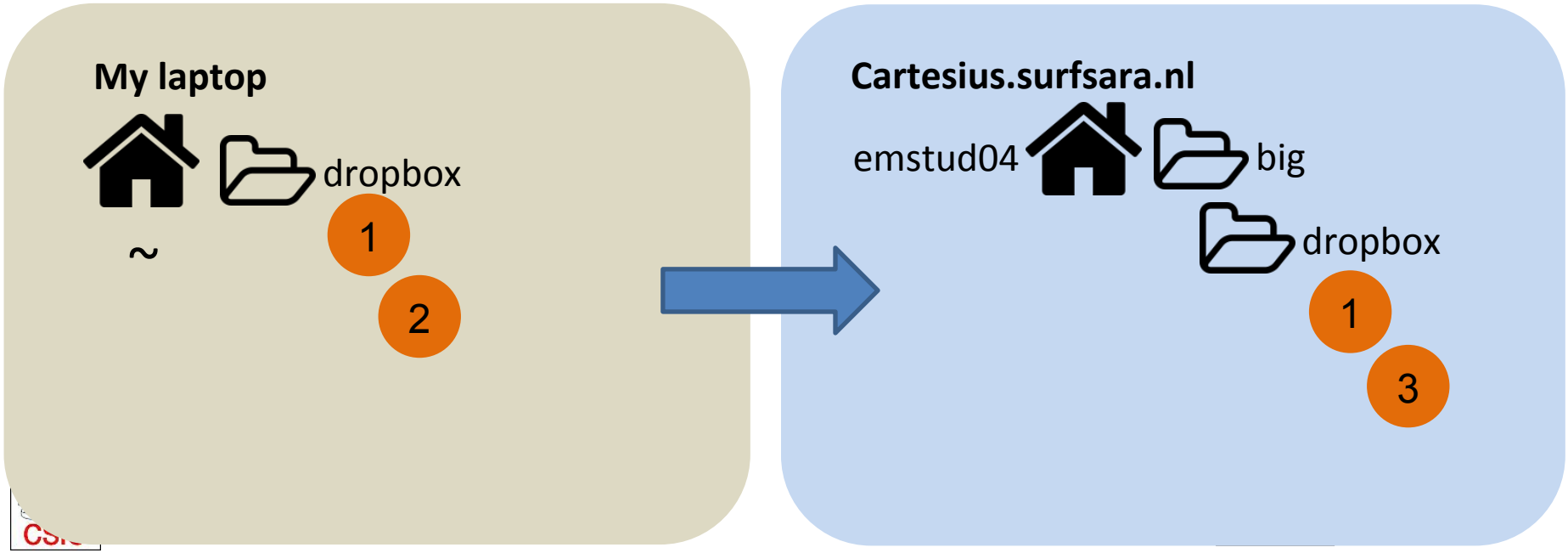
rsync -av –progress ~/dropbox emstud04@cartesius.surfsara.nl:big/

Since dropbox already exists in big, the rsync command will mix both



Data transfer / Mix (syncing)

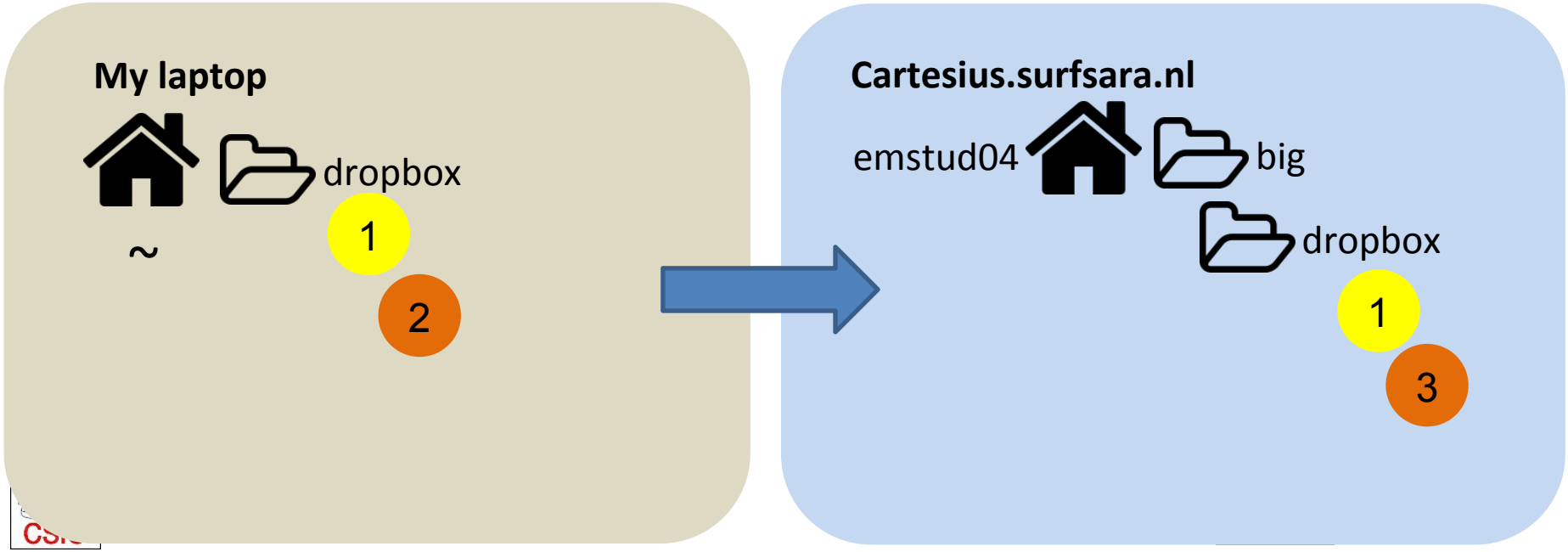
What does “mix” imply?



Data transfer / Mix (syncing)

File “1” exists in both the laptop and cartesius.

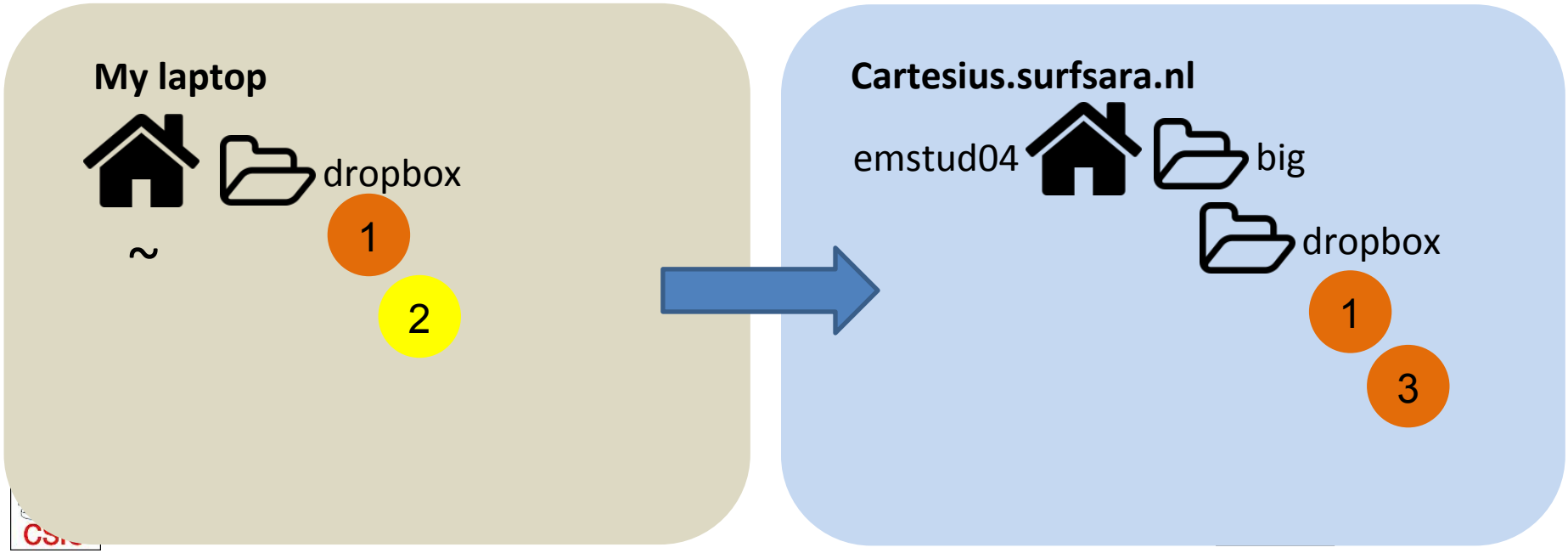
File “1” in cartesius will be **updated** so as to be equal to file “1” in the laptop



Data transfer / Mix (syncing)

File “2” only exists in the laptop.

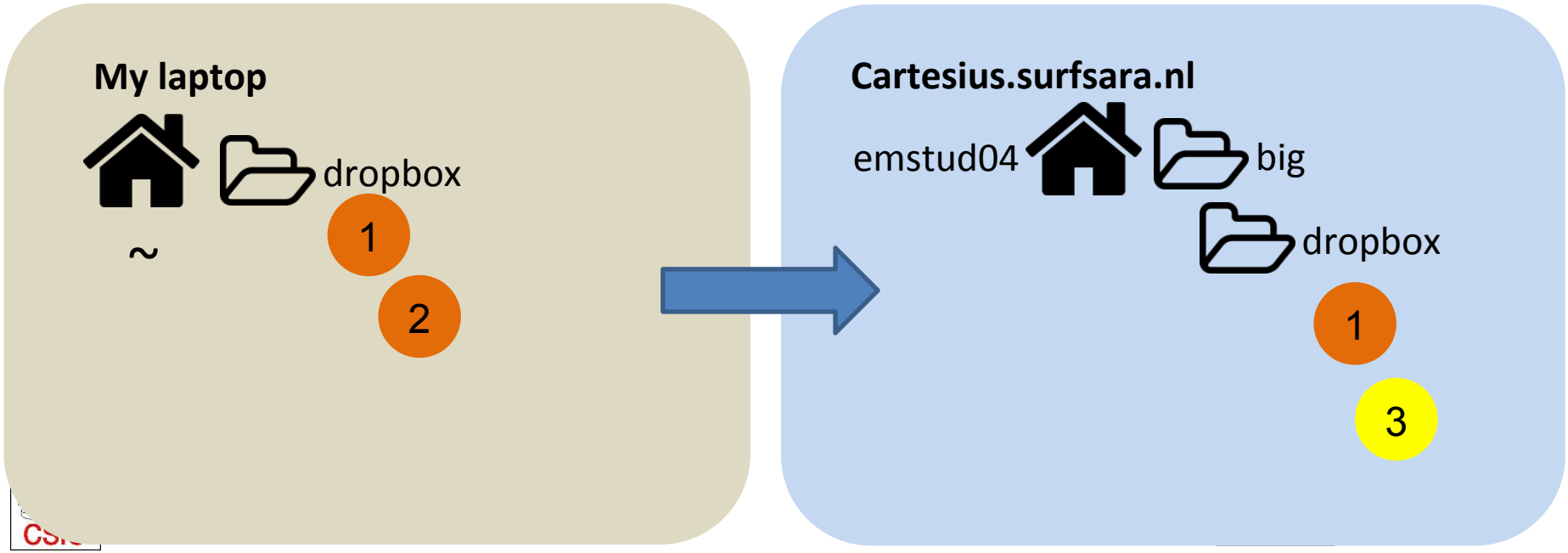
It will be added to cartesius



Data transfer / Mix (syncing)

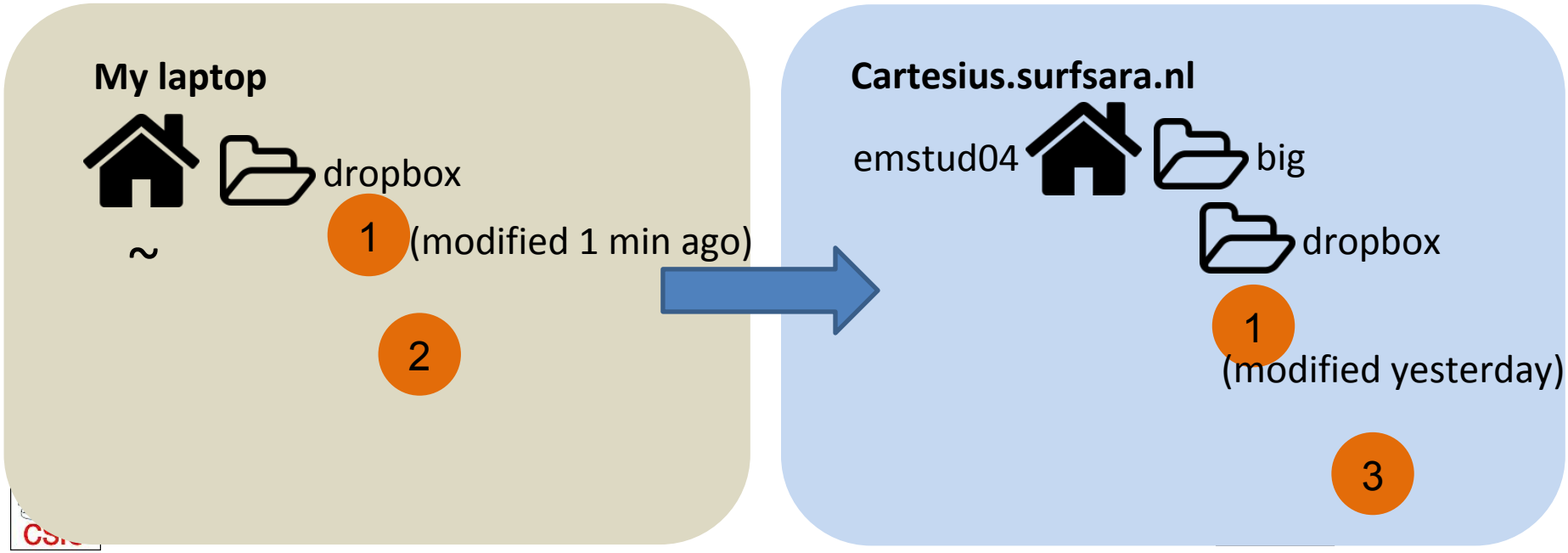
File “3” only exists in cartesius.

It will be left as is



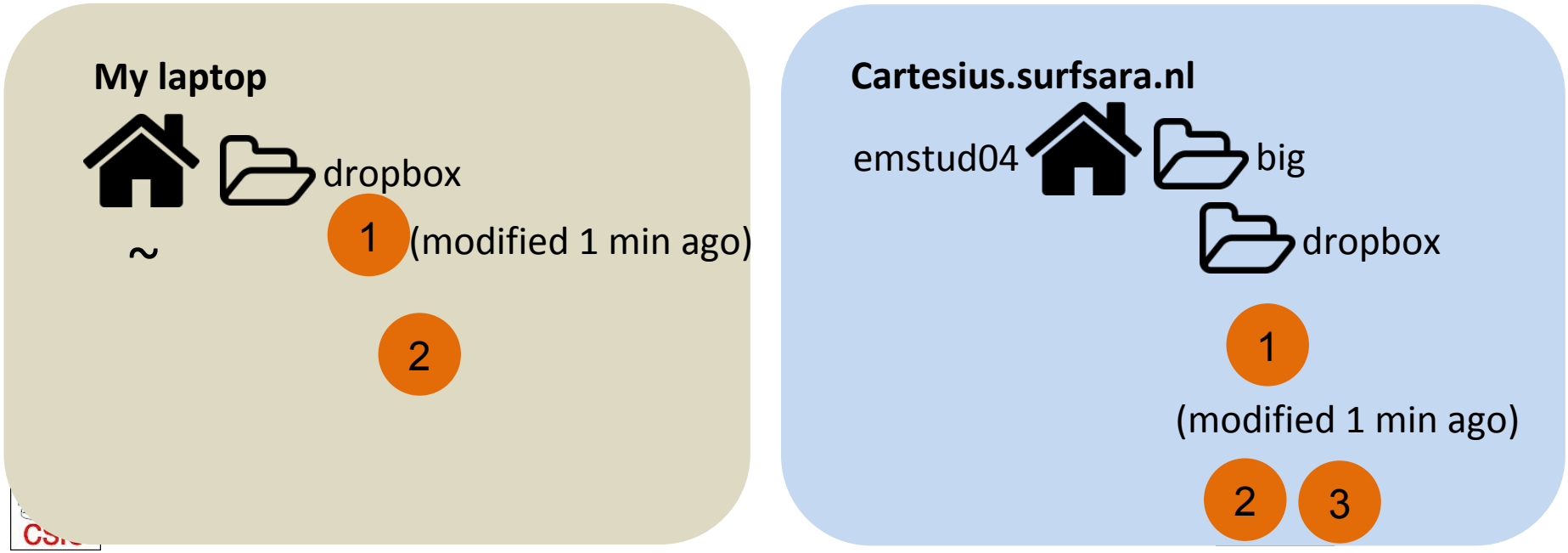
Data transfer / Mix (syncing)

Before running rsync…



Data transfer / Mix (syncing)

After running rsync…



5. Remote Desktop / Enjoy!