Accessing, connecting to, and transferring data to an HPC environment

First steps using a HPC

BU-ISCIII
29-04 de octubre de 2025
1ª edición

Outline

1. Requesting access and hpc rules
2. Introducing HPC xtutatis
3. Accessing the cluster (ssh)
4. Transferring data (cli and gui tools)
5. Organizing files and storage spaces
6. First steps once connected
7. Recap and hands-on exercises

How to request access

- An account is required to use the HPC service.
- Users must belong to ISCIII or associated projects.
- Access only for research or diagnostic purposes
- Apply through **sau.isciii.es**
- Must sign the **usage rules document (Normativa de Uso)**[1]

[1] *https://sau.isciii.es/front/helpdesk.faq.php?id=132*

Accepting the rules

- Users **must read and sign** the Normativa de Uso.
  - ***Commit to***:

    ✅Efficient use of resources (only research/diagnosis).

    ❌No long-term storage (HPC ≠ backup service).

    ✅Respect licenses of installed software.

    ✅Mention XTutatis in publications using its resources.
  - ***Obligations***:

    ❌Do not share accounts.

    ✅Use strong passwords.

    ✅Report anomalies to UTIC.

What is XTutatis?

- ISCIII HPC service for scientific research.
- Provides large-scale compute & storage.
- Managed by UTIC (infrastructure, software support).

Hardware overview

- 34 compute nodes
- 768 CPU cores total
- ~12.6 TB RAM
- 4× NVIDIA Tesla P100 GPUs
- 100 TB shared storage (scalable)
- 10 Gbps network interconnect

Node Types

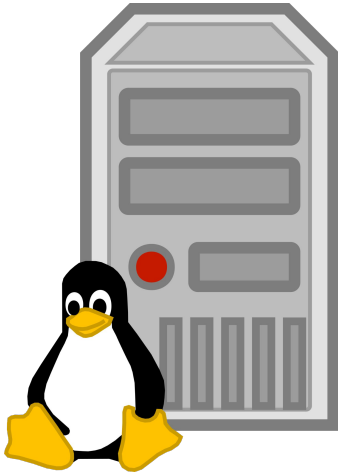
Ideafix (Cascade Lake, up to 2 TB RAM)

Numerobis (Skylake + GPUs)

Obelix (Ivy Bridge) -> no available at the moment

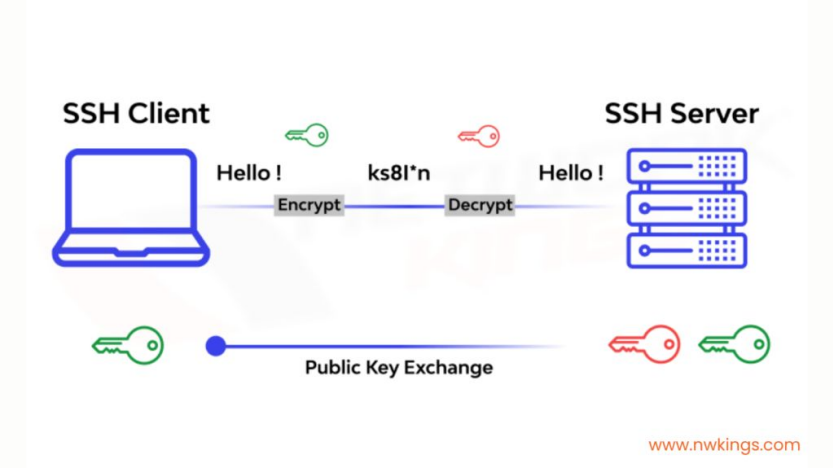
Operating system and Access node

- Centos 8 linux across all nodes
- login node (frontend) used for:
  - file management
  - job submission
  - never run heavy jobs on login node



What is SSH?

- Secure Shell protocol (encrypted communication)
- Used for login, file transfer, tunneling
- Available on Linux, macOS, Windows (10+ built-in)



- Communication is encrypted end to end
- Authentication with password or ssh keys
- Public key exchange ensures identity
- Prevents eavesdropping and tampering

Basic SSH command

SSH clients:

- Built-in CLI: ssh (Linux/Mac, Windows 10+)
- GUI options: PuTTY, MobaXterm (Windows)

```
>~$ ssh username@hostname
```

username = your cluster account
hostname = cluster login address (e.g., login.cluster.org)

Advanced SSH options

- -p -> port specify connection port (xtutatis uses 32122)
- -v ->verbose mode for troubleshooting
- Combine options in one command

Example xtutatis login

```
ssh -v -p 32122 username@portutatis.isciii.es
```

Always remember: *you're connecting from your local machine (laptop or workstation) → to the login node (portutatis.isciii.es) on port 32122*

Quiz/dicussion

Have you ever used ssh before?

- In which situations did you use it?
- Did you face any connection problems?
- Which operating system are you most comfortable using for ssh?

Troubleshooting connection

Troubleshooting with ssh -v

- Look at the last debug lines
- Common errors:
  - ***Connection timed out*** → wrong port, firewall, VPN
  - ***Connection refused*** → server reachable but SSH service not available
  - ***Permission denied*** → wrong password or key
  - ***Host not found*** → wrong hostname or DNS

Troubleshooting connection:Connection timed out

- Connection unavailable, forgot -p

```
$ ssh -v user@portutatis.isciii.es
OpenSSH_8.9p1 Ubuntu-3ubuntu0.13, OpenSSL 3.0.2 15 Mar 2022
debug1: Reading configuration data /etc/ssh/ssh_config
debug1: /etc/ssh/ssh_config line 19: include /etc/ssh/ssh_config.d/*.conf matched no
files
debug1: /etc/ssh/ssh_config line 21: Applying options for *
debug1: Connecting to portutatis.isciii.es port 22.
debug1: connect to address portutatis.isciii.es port 22: Connection timed out
ssh: connect to host portutatis.isciii.es port 22: Connection timed out
```

Troubleshooting connection:Connection refused

- wrong port or SSH not running

```
$ ssh -v -p 1234 user@portutatis.isciii.es
OpenSSH_8.9p1 Ubuntu-3ubuntu0.13, OpenSSL 3.0.2 15 Mar 2022
debug1: Reading configuration data /etc/ssh/ssh_config
debug1: /etc/ssh/ssh_config line 19: include /etc/ssh/ssh_config.d/*.conf
matched no files
debug1: /etc/ssh/ssh_config line 21: Applying options for *
debug1: Connecting to portutatis.isciii.es port 22.
debug1: Connecting to portutatis.isciii.es [10.22.140.230] port 1234.
ssh: connect to host portutatis.isciii.es port 1234: Connection refused
```

Troubleshooting connection:Permission denied

- wrong password or no key match

```
$ ssh -v -p 32122 user@portutatis.isciii.es
OpenSSH_8.9p1 Ubuntu-3ubuntu0.13, OpenSSL 3.0.2 15 Mar 2022
debug1: Reading configuration data /etc/ssh/ssh_config
debug1: /etc/ssh/ssh_config line 19: include /etc/ssh/ssh_config.d/*.conf matched no
files
debug1: /etc/ssh/ssh_config line 21: Applying options for *
debug1: Connecting to portutatis.isciii.es port 22.
debug1: Authentications that can continue: publickey,password
debug1: Next authentication method: password
Permission denied, please try again.
ssh: Permission denied (publickey,password).
```

Troubleshooting connection: Host not found

- typo in hostname

```
$ ssh -v user@portutatis.issciii.es
OpenSSH_8.9p1 Ubuntu-3ubuntu0.13, OpenSSL 3.0.2 15 Mar 2022
debug1: Reading configuration data /etc/ssh/ssh_config
debug1: /etc/ssh/ssh_config line 19: include /etc/ssh/ssh_config.d/*.conf matched no
files
debug1: /etc/ssh/ssh_config line 21: Applying options for *
debug1: Connecting to portutatis.issciii.es port 22.
ssh: Could not resolve hostname portutatis.issciii.es: Name or service not known
```

Troubleshooting connection: besides ssh -v

- Checking connectivity with ping and telnet

  a. **ping**: tests if the host is reachable

  b. **telnet**: tests if a specific port is open

- Helps distinguish network vs ssh service issues

Troubleshooting connection: ping host

- ✅ Success:

```
$ ping -c 2 portutatis.isciii.es

PING portutatis.isciii.es (10.22.140.230) 56(84) bytes of data.

64 bytes from 10.22.140.230: icmp_seq=1 ttl=61 time=12.3 ms

64 bytes from 10.22.140.230: icmp_seq=2 ttl=61 time=12.5 ms

--- portutatis.isciii.es ping statistics ---

2 packets transmitted, 2 received, 0% packet loss
```

Troubleshooting connection: ping host

- ❌ Failure:

```
$ ping -c 2 portutatis.isciii.es

ping: portutatis.isciii.es: Name or service not known
```

Troubleshooting connection: telnet host port

- ✅ Success:

```
$ telnet portutatis.isciii.es 32122

Trying 10.22.140.230...

Connected to portutatis.isciii.es.

Escape character is '^]'.

SSH-2.0-OpenSSH_8.0
```

Troubleshooting connection: ping host

❌ Failure (port closed):

```
$ telnet portutatis.isciii.es 22

Trying 10.22.140.230...

telnet: Unable to connect to remote
host: Connection refused
```

❌ Failure (no response):

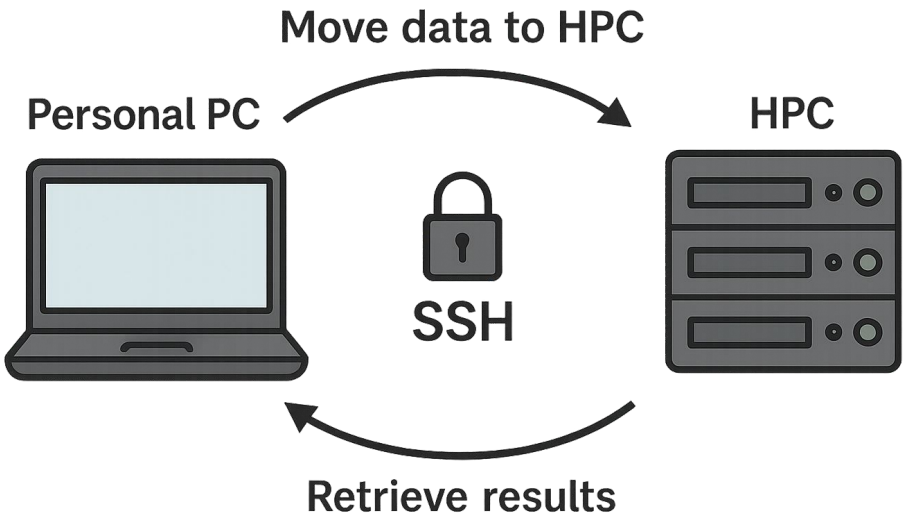
```
$ telnet portutatis.isciii.es 32122

Trying 10.22.140.230...

telnet: Unable to connect to remote
host: Connection timed out
```

Quiz/dicussion

What would you check first if ssh fails?

- *Wrong port or hostname*
- *Firewall / VPN connection*
- *Wrong password or missing key*
- *Network unreachable*

How to transfer data



Transfer options (CLI)

			
Tool	**scp**	**rsync**	**sftp**
Best for	Simple file or folder copies	Large directories, repeated syncs	Interactive transfers
Notes	Always copies full file, less efficient	Only transfers changes, supports resume	Command session similar to ftp, slower

Transfer data with scp

✅ Success

```
scp file.txt username@portutatis.isciii.es:~/DOC/
file.txt                    100%   12KB   2.3MB/s   00:00
```

❌ Common problems

- No such file or directory → wrong local or remote path
- Permission denied → no write access in remote directory
- Connection refused → forgot -p 32122
  a. Solution:

```
scp -P 32122 file.txt username@portutatis.isciii.es:~/DOC/
```

Transfer data with scp

✅ Success

```
scp file.txt username@portutatis.isciii.es:~/DOC/
file.txt                    100%   12KB   2.3MB/s   00:00
```

❌ Common problems

- No such file or directory → wrong local or remote path
- Permission denied → no write access in remote directory
- Connection refused → forgot -p 32122
  a. Solution:

```
scp -P 32122 file.txt username@portutatis.isciii.es:~/DOC/
```

Transfer data with rsync

✅ **Success**

```
rsync -av -e "ssh -p 32122" project/ username@portutatis.isciii.es:~/ANALYSIS/
sending incremental file list
project/file1.txt
project/file2.txt
sent 2.35K bytes  received 102 bytes  1.23K bytes/sec
total size is 1.2M  speedup is 2.0
```

Transfer data with rsync

❌ Common problems

- rsync: command not found on remote → not installed (rare on HPC)
- Permission denied → no access in target folder
- Connection stuck → VPN/firewall or wrong port

Transfer data with scp

✅ Success

```
scp file.txt username@portutatis.isciii.es:~/DOC/
file.txt                    100%   12KB   2.3MB/s   00:00
```

❌ Common problems

- No such file or directory → wrong local or remote path
- Permission denied → no write access in remote directory
- Connection refused → forgot -p 32122
  a. Solution:

```
scp -P 32122 file.txt username@portutatis.isciii.es:~/DOC/
```

Transfer data with scp

✅ Success

```
sftp -P 32122 username@portutatis.isciii.es
Connected to portutatis.isciii.es.
sftp> put file.txt
Uploading file.txt to /home/username/file.txt
file.txt                                    100%  12KB  2.3MB/s  00:00
```

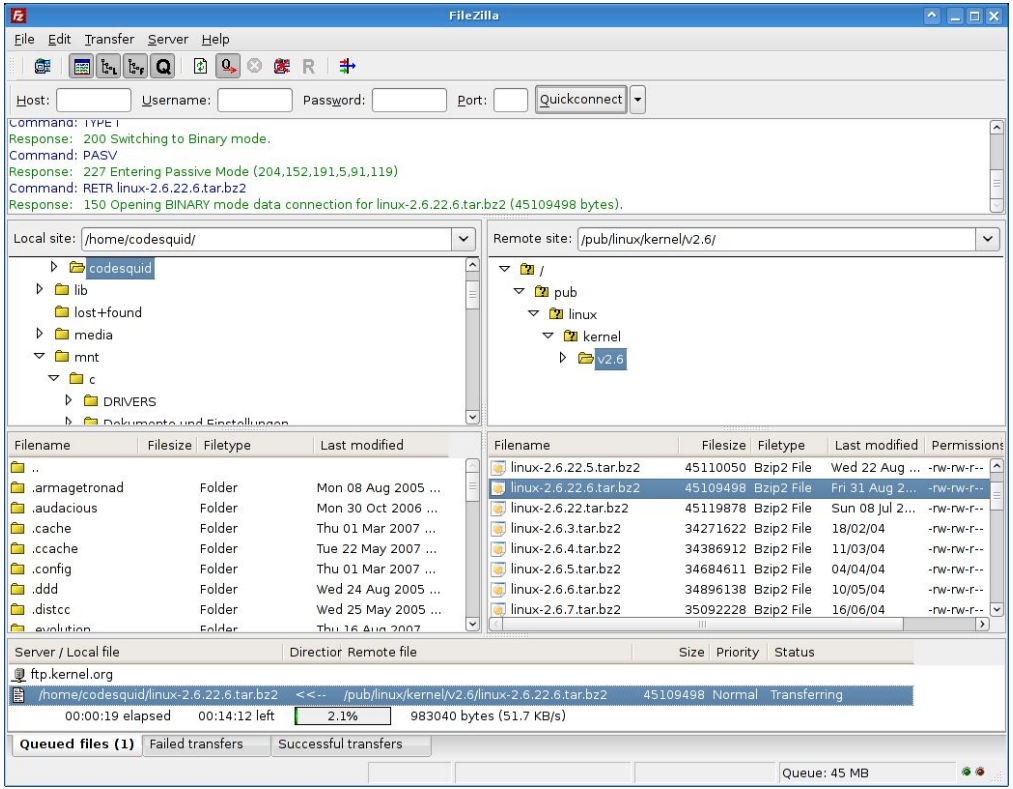
❌ Common problems

- Couldn't connect to server → wrong port or no network
- Permission denied → no access in target directory
- Upload fails mid-transfer → connection interrupted, retry

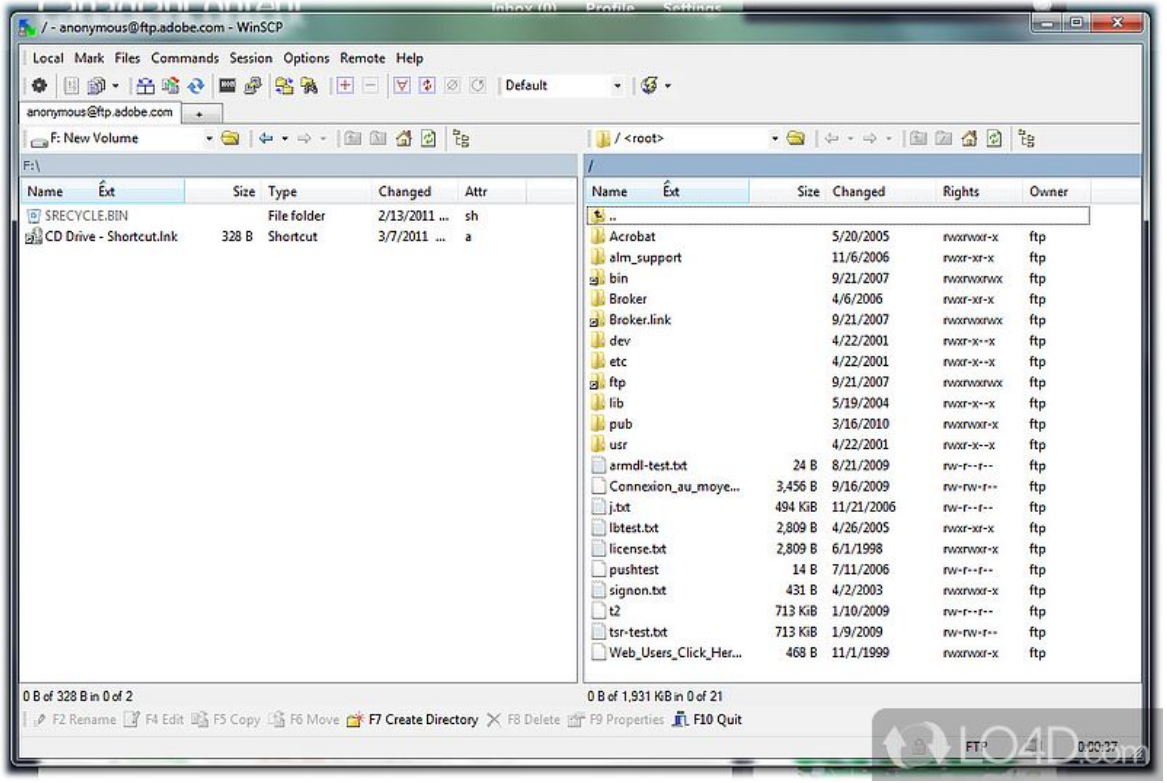
Transfer options (GUI)

Tool	Best for	Features	Limitations / Pricing
WinSCP / FileZilla 	Easy file transfers (drag & drop)	User-friendly interface, cross-platform, supports SCP/SFTP	Free to use
MobaXterm 	Integrated SSH + file management	Built-in SFTP browser, X11 server, all-in-one for Windows	Free (community version available)
Globus 	Very large data, fault-tolerant	Automated, reliable transfers, resume support	Free for researchers at non-profit institutions; paid subscriptions available for premium features

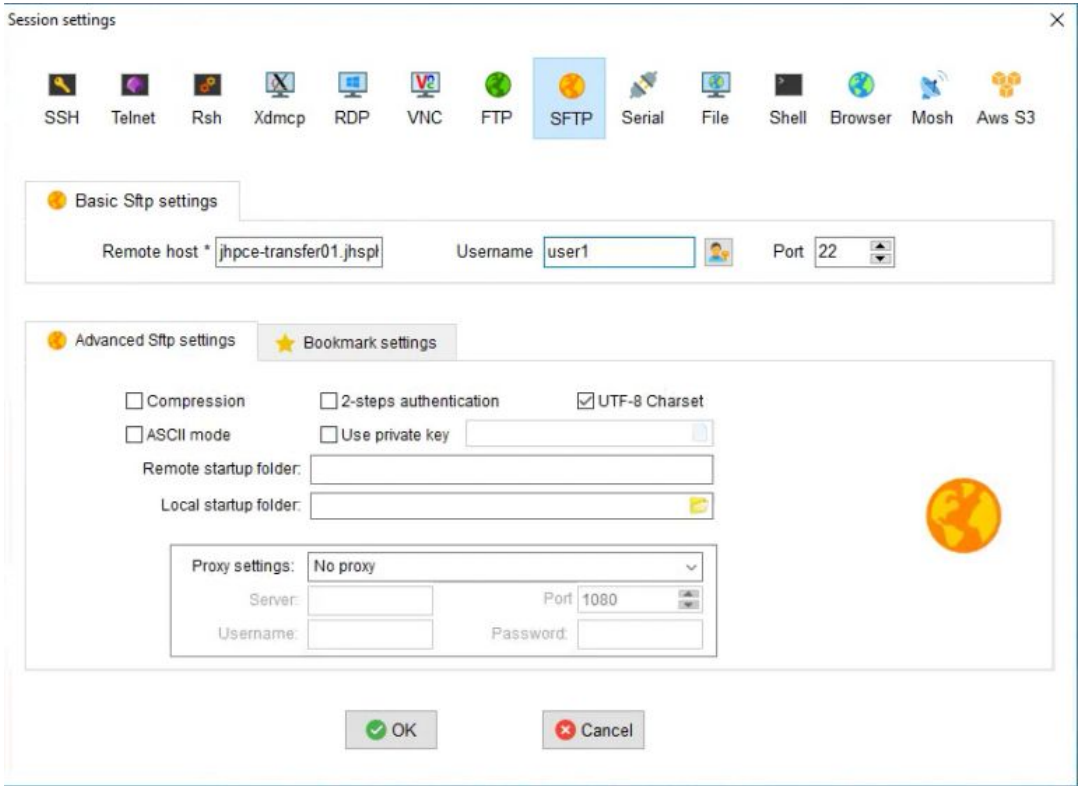
Transfer GUI option: Filezilla



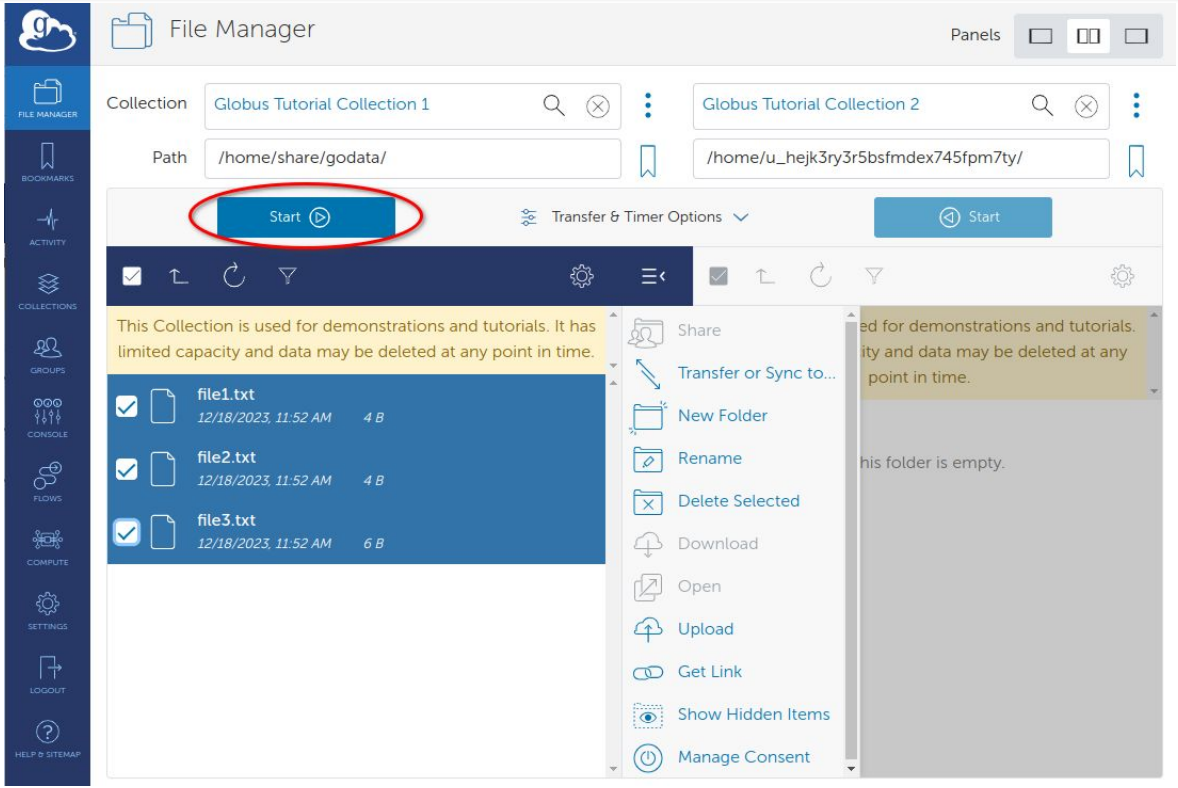
Transfer GUI option: WinSCP



Transfer GUI option: MobaXterm



Transfer GUI option: Globus



Quiz/dicussion

Which tool would you prefer for transferring your data?

- *CLI tools: scp, rsync, sftp*
- *GUI tools: winscp, filezilla, mobaxterm, globus*
- *Why would you choose one over another?*

Verify data integrity

- Use checksums
  a. A **checksum** is a digital fingerprint of a file
  b. Used to detect corruption or changes in transfer
  c. md5sum generates a hash value for a file (128-bit values)

- Compare checksum **before and after transfer**
- **Process:** generate hash → transfer file → regenerate hash → compare both. If they match, the file is intact.

Verify data integrity

Create checksum before transfer

```
$ md5sum file.txt
5d41402abc4b2a76b9719d911017c592  file.txt
```

Check after transfer

```
$ md5sum file.txt
5d41402abc4b2a76b9719d911017c592  file.txt
```

If corrupted

✅ Checksums match → file intact

❌ Different checksum → file altered or corrupted

```
$ md5sum file.txt
7c6a180b36896a0a8c02787eeafb0e4c  file.txt
```

Verify data integrity

It's common to store a file's hash in a .md5 file (e.g., md5sum file.txt > file.txt.md5) and later verify with:

```
$ md5sum *.fastq > checksums.md5

$ md5sum -c checksums.md5
sample1.fastq: OK
sample2.fastq: OK
sample3.fastq: FAILED
md5sum: WARNING: 1 computed checksum did NOT match
```

Quiz / Discussion

Have you ever consider corruption of files when moving data around?

- *Have you experienced integrity problems before?*
- *Do you think md5sum is something that can be extended to all type of users?*

Organizing files on HPC

- 🔍 Easy retrieval and collaboration
- 📁 Consistency across projects
- 📊 Prevents quota and storage issues
- 🔄 Improves reproducibility of results
- 👥 Simplifies sharing with collaborators
- 🛡️ Avoids accidental overwriting or data loss

Storage Spaces in XTutatis

- **/home:** personal, small (200 GB shared), scripts & configs
- **/data/unidad:** project results, shared per unit (20-30 TB)
- **/scratch/unidad:** main space for jobs, purged if inactive >5 days (7 TB SSD)
- **/local_scratch:** per-node temporary SSD (~800 GB), auto-cleaned
- **/srv/fastq_repo:** NGS data repository (1-year retention)

Suggested structure

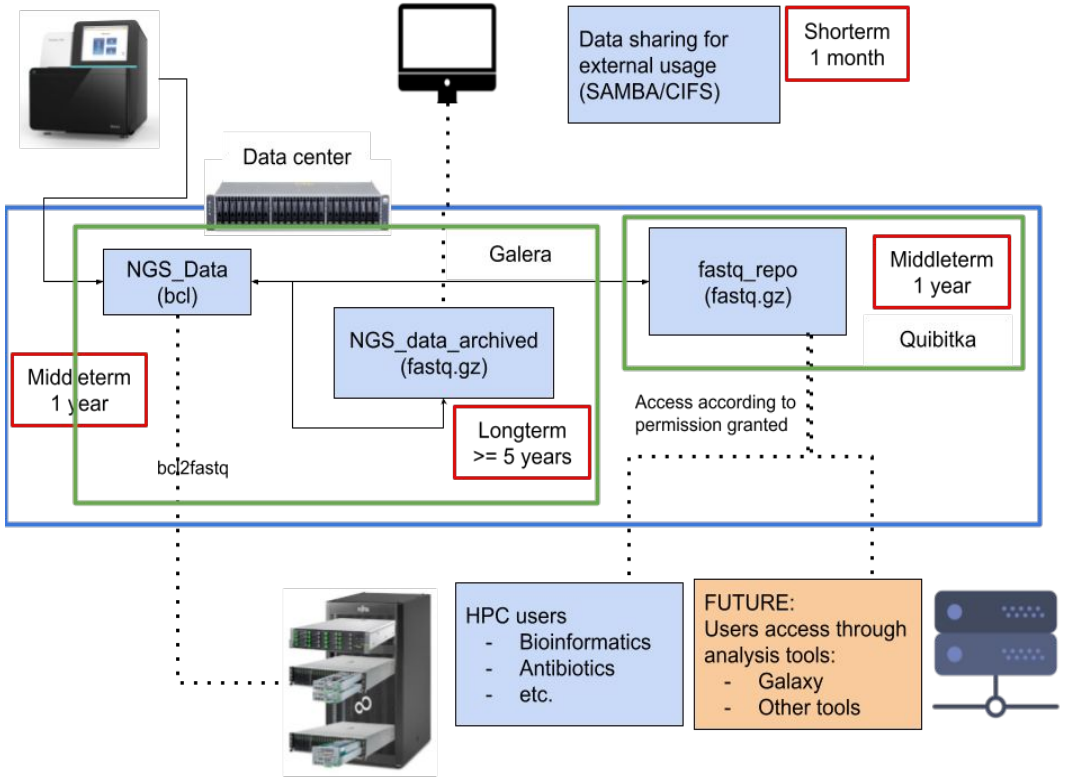
```
PROJECT/
├── RAW/                # raw sequencing data
├── ANALYSIS/           # processed results
├── DOC/                # documentation, scripts
├── TMP/                # temporary/intermediate
└── REFERENCES/ # reference genomes
```

Quizz / discussion

How would you organize your own hpc project?

- *What folders would you create first?*
- *How would you separate raw data from results?*
- *Would your structure be different for a short vs long project?*

Genomic data management protocol



Protocolo de organización y gestión de almacenamiento de datos genómicos en el ISCIII

Contenidos

Once connected

- **Login node:** lightweight tasks (setup, editing, transfers)
- **Compute nodes:** heavy jobs via scheduler (later session)
- **Who am I?** → whoami
- **What groups/resources do I have?** → id, groups
- **Disk quotas & storage** → df -h
- **Shell environment** → check .bashrc

Key takeaways

- Use SSH for secure access
- Multiple tools for file transfer (CLI & GUI)
- Organize your project logically
- Respect login node policies

Hands-on session

- Connect with SSH
- Explore environment
- Configure ssh key exchange
- Transfer data
- Verify with checksums
- Build project directory tree

Thank you for your attention

Questions?