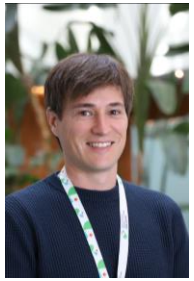
Introduction to IJC computational infrastructure

Angelika Merkel (Head of Bioinformatics)
11/06/2024

Supported by IT & BIT

IJC Bioinformatics

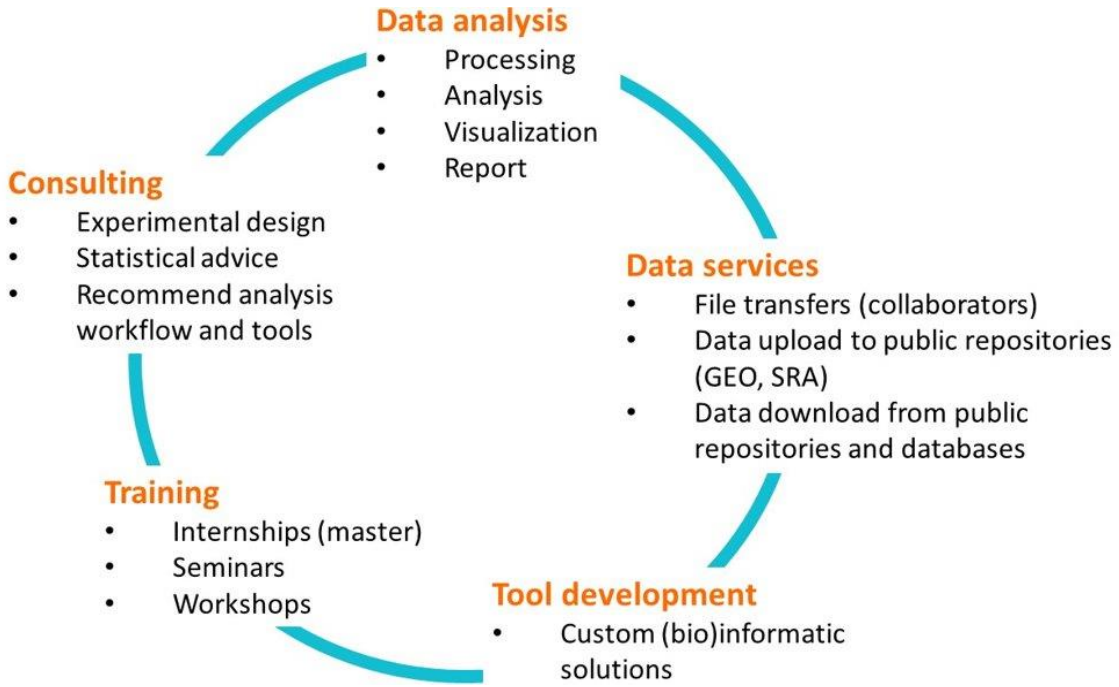
Angelika Merkel
Unit Manager
Bioinformatician



Emilio Lario
Software Engineer

Marina Vilardell
Bioinformatician

Marta Meroño
Bioinformatician
(Student)



Office: Sala Prof. Albert Grañena (1st floor)(1st floor); phone: 4300

https://carrerasresearch.sharepoint.com/sites/BIT

https://www.carrerasresearch.org/en/bioinformatics-unit

IT Team

Marc Jubany

(Head of IT)

José Alcántara

(HPC administration/ Linux support)

Roger España

(support)

Office: 0.00.04 (ground floor)

Email: helpdesk@carrerasresearch.org
https://tickets.carrerasresearch.org/

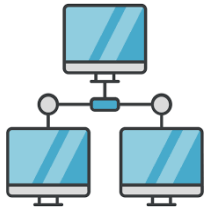
Overview:

Resources

Shared resources

Filesharing

Data management



IJC network **& working remotely**

Network

Connecting from the outside

Cybersecurity

High performance computing (HPC)

Architecture

Scheduling

1. Resources & file sharing

Shared network resources (IJC only)

INVESTIGACIO

- Shared folder for research group or core facility. Contains by default a private folder for the group leader and a shared folder accessible by anyone in the group
- **For Office documents and results only!** No RAW DATA or big images.
- **Back-up**

SHARED FOLDER

- Folders shared with specific groups or people. They are created by the IT department upon request.
- Some default folders shared with all groups: TRAVEL_UNIT, LABORATORIES, UNITAT_MICROSCOPIA
- **For Office documents and results only!** No RAW DATA or big images.
- **Back-up**

High-performance computing (HPC) cluster

- Computational nodes
- ISILON file storage system (**RAW DATA, large files**)
- **Back-up**

NAS SYNOLOGY BIODATA (OLD!!!!)

- Network attached storage (100TB) for data storage
- Shared folder for each group \BDgroupleadername, (IP: 10.110.20.7) set up by IT
- Predates the current HPC and will become obsolete in the future
- **NO back-up!**

File sharing (via web browser)

Access your files from everywhere through a browser and share them easily via link or email

ONE DRIVE/ SHAREPOINT (MICROSOFT)

- https://carrerasresearch-my.sharepoint.com/

- Pros: Desktop compatible, integrated with office 365
- Cons: Data is located on MS servers off site

NEXTCLOUD

1. Core facilities:
    - https://nextcloud_micro.carrerasresearch.org/
    - Share data files or results with collaborators
2. Personal:
    - Accessible through VPN portal
    - Access to > INVESTIGACIO, > SHARED_FOLDER, > BD$GROUP
3. On request
    - Request a custom link from IT to share files

- Pros: Highly customizable access (time, groups, etc.)
- Cons: limited file viewing properties for office documents

File sharing via FTP

Move files from inside the IJC network to the internet and vice versa

Secure FTP

- transfer large data files
- 5TB temporary storage shared amongst users
- Host: ftpbios.carrerasresearch.org Port 4242
- Access over the internet **only with credentials**

Public FTP

- e.g. for software requiring public URLs to upload data (genome browser)
- account needs to be configured by IT
- Host: bios.carrerasresearch.org Port 22
- https://username.carrerasresearch.org
- Allows access over the internet **without credentials**

Transfer files with:

1) ftp client like Filezilla or
2) via Terminal

```
sftp –P 4242 username@ftpbios.carrerasresearch.org (secure)
     > mput myfile
     > get myfile
```

```
sftp bios.carrerasresearch.org  (public)
```

External filesharing tools

Interface	Tool	Access	Recommended for	File location
Web browser	OneDrive	https://carrerasresearch-my.sharepoint.com/	Documents	Windows servers (off-site)
Web browser	NextCloud	https://nextcloud_ijc.carrerasresearch.org	Documents, data <1GB	IJC servers (on-site)
Software	Filezilla	Available for MacOS, Windows, Linux	Data > 1GB	IJC sftp server (on-site)
Terminal	Sftp (public)	https://username.carrerasresearch.org sftp username@bios.carrerasresearch.org > mput my_file; > mget my_file	Public data < 5GB	IJC sftp server (on-site)
Terminal	Sftp (secure)	sftp – P 4242 username@ftpbios.carrerasresearch.org > mput my_file; > mget my_file	Data > 1GB	IJC sftp server (on-site)

Data management best practices:

- Important **documents and data should be stored on IJC servers**, with centralized management and frequent backups. (Intellectual property concerns, IJC servers are managed by IT).

- **Local workstations serve to interact with IJC file servers,** for development and temporary storage. There is **no back-up** (Workstations supervised by IT but are still managed by the user).

- **No back-up** on external hard drives.

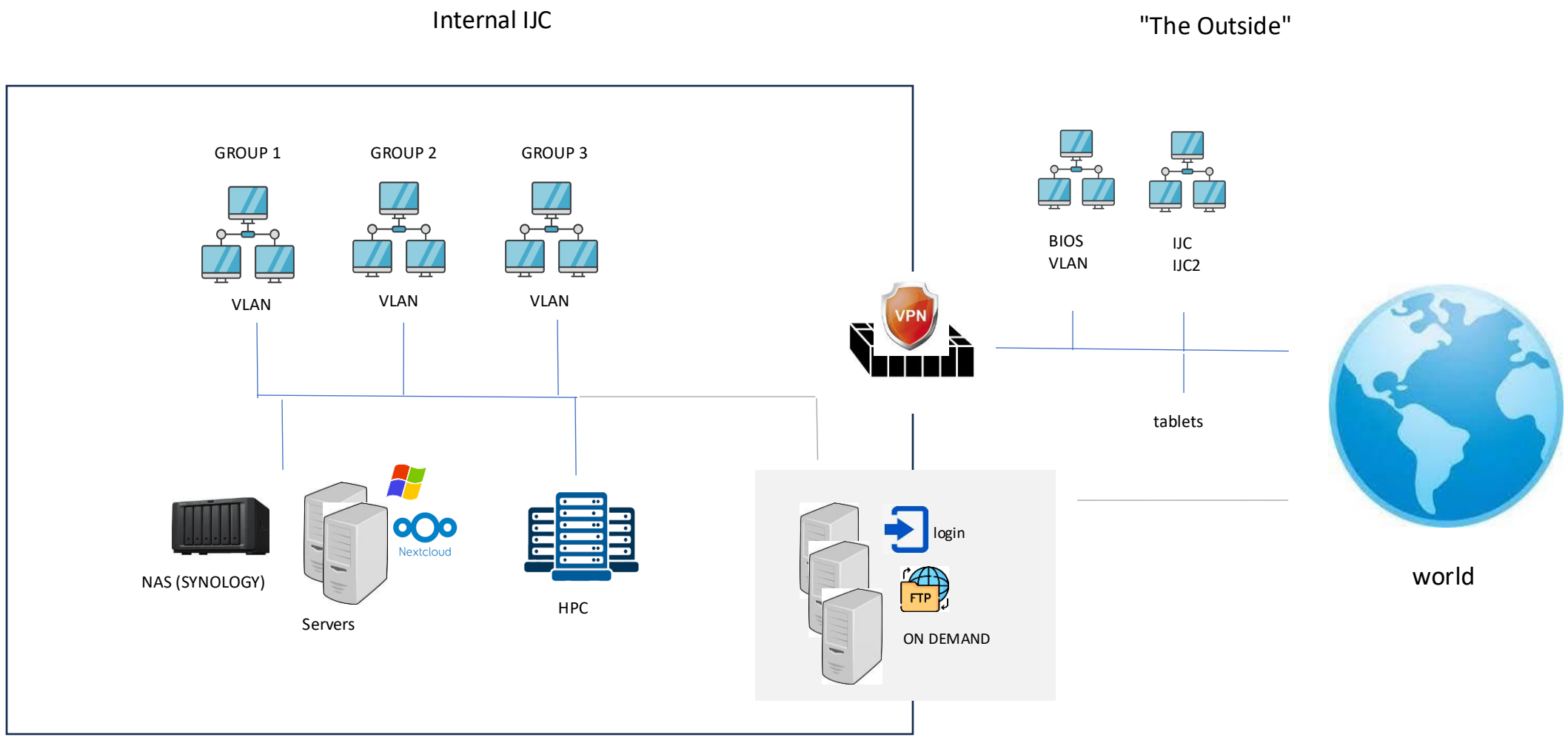
- Repositories of databases and software should be centralized (when possible) and associated with a user/user group. The repository should be revised when the associated user or user group changes or leaves.

- User data and accounts are terminated after the employee leaves.(email valid 6months). Before leaving the institute the user shall remove irrelevant data, assure important data is kept accordingly and is well documented. (@Bioinformaticians: Use RELATIVE PATHS in your scripts avoid ABSOLUTE PATHS!)

Avoid multiple copies of data!

2. The IJC network & remote connections

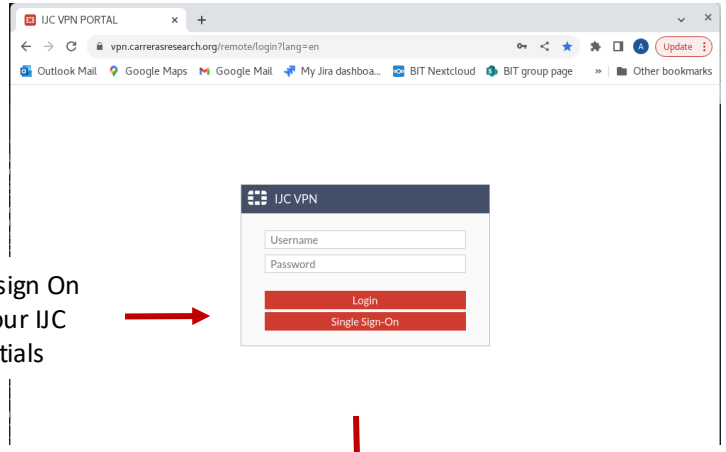
The network


Internal IJC
"The Outside"
GROUP 1
GROUP 2
GROUP 3
VLAN
VLAN
VLAN
VPN
BIOS
VLAN
IJC
IJC2
tablets
world
NAS (SYNOLOGY)
Nextcloud
Servers
HPC
login
FTP
ON DEMAND

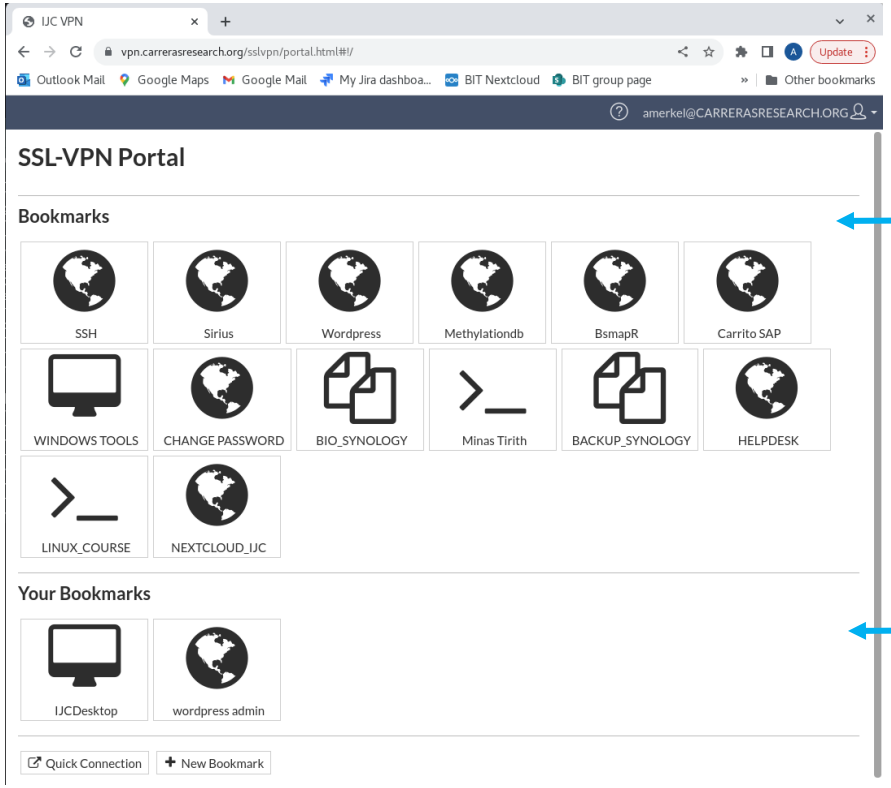
The IJC VPN portal

https://vpn.carrerasresearch.org/

Single-sign On with your IJC credentials



2 – Factor- Authentication



Shared bookmarks (by IT)

Personal bookmarks (only bioinformatics profile)

The login node

If you are connecting from outside the IJC (or you are using a computer that is not administrated by IT) you can connect through the **'login node'**.

You need:

1. an SSH client (bash for Linux/zsh for Mac, or putty for windows)
2. an active domain user
3. your smartphone with a 2 Factor Authentication app (different from MS single-sign-on and configured by IT!)

You can access the machine with the following command:

```
ssh -P 4242 username@loginnode.carrerasresearch.org
Password: (your windows/mail password)
Verification code: (your 2FA app code)
```

Now you can ssh to minastirith or your workstation

IT administrated & non-IT administrated computers

IT

- Software installation and maintenance by IT
- No 'sudo' rights

Pros:

- IJC network 600Mb/s CSUC internet
- ISILON filesystem is permanently mounted
  (-> allows file access with GUI applications)
- Connection from the outside directly to the IJC network with vpnclient (Linux/Mac via biosvpn with IT certificate)
- As-you-go printing (all IJC printer)

Cons:

- (-> limited control, IT dependent)

Non-IT

- Software installation and maintenance by user
- 'sudo' rights

Pros:

- full control

Cons:

- Isolated bios network (without backup), 600Mb/s Movistar internet
- ISILON access only via ssh client through minastirith, no mount
  (-> no file access for GUI applications)
- Connection from the outside to the IJC network only via VPN portal or login node
- Printing only on guest printers (next auditorium, 2nd floor)

IT administrated & non-IT administrated computers

IT

- Software installation and maintenance by IT
- No 'sudo' rights

Pros:

- IJC network 600Mb/s internet
- ISILON filesystem is permanently mounted
  (-> allows file access with GUI applications)
- Connection from the outside directly to the IJC network with vpnclient (Linux/Mac via biosvpn with IT certificate)
- As-you-go printing (all IJC printer)

Cons:

- (-> limited control, IT dependent)

Non-IT

- Software installation and maintenance by user
- 'sudo' rights

Pros:

- full control

Cons:

- Isolated bios network (without backup), 100Mb/s internet
- ISILON access only via ssh client through minastirith, no mount
  (-> no file access for GUI applications)
- Connection from the outside to the IJC network only via VPN portal or login node
- Printing only on guest printers (next auditorium, 2nd floor)

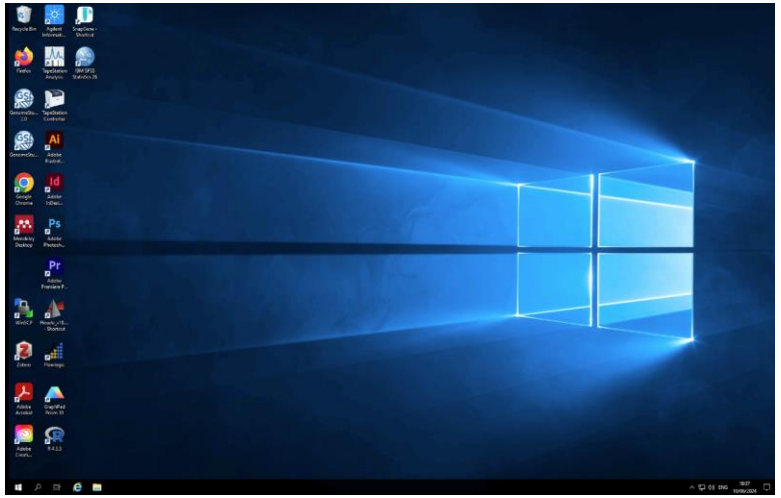
Cybersecurity best practices

- Do not open email with suspicious attachments or from unknown senders (if in doubt verify with IT)
- Do not plugin external hard disks with unverified content
- Do not create/ maintain unsanctioned network connections between the IJC network and the outside world
- Do not share your IJC credentials (IT or your PI never will/ should ask for your credentials)
- Do not approve 2FA if you are not accessing your account at the time

The Joker

Virtual windows desktop with licensed software:

1) ADOBE CREATIVE CLOUD *
2) GRAPHPAD PRISM 10 *
3) SNAPGENE
4) SPSS
5) Genome Studio

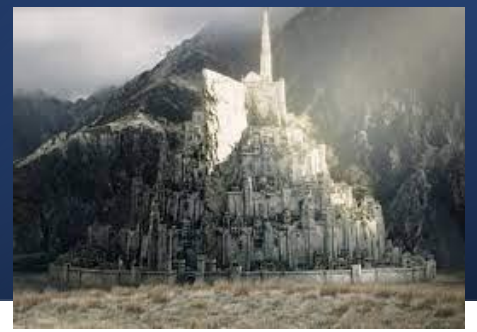


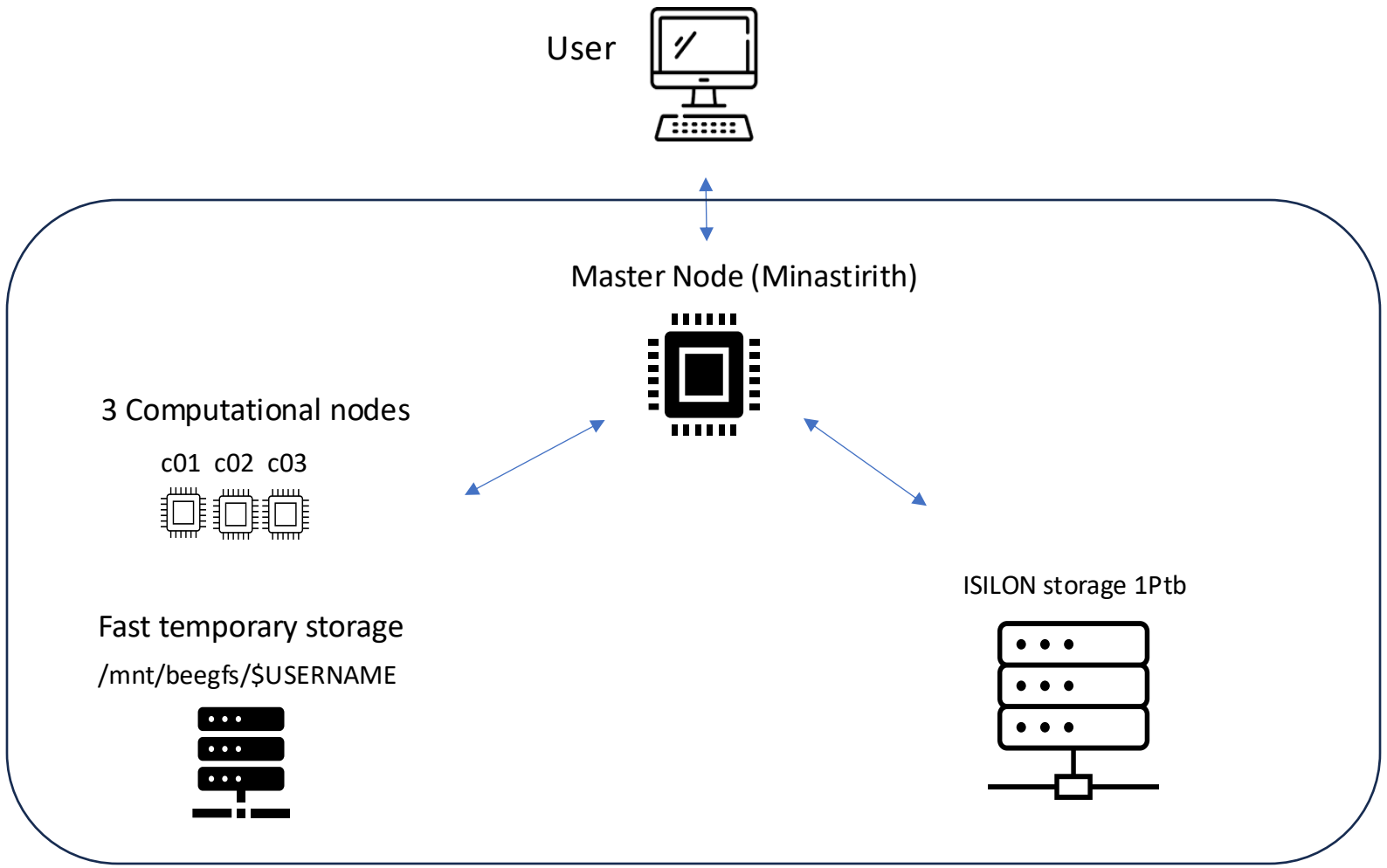
Access:

1) vpn.carrerasresearch.org
2) Windows/Macs 'remote desktop' app (> joker with your credentials)

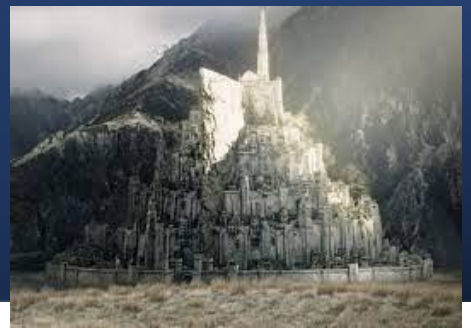
3. High Performance Computing (HPC)

IJC HPC cluster





IJC cluster components



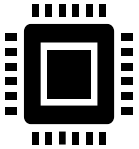
- Master Node (Minastirith)

  12 cores 128Gb Ram

  = connects user, computational nodes (+ flash storage), ISILON

  USE: to connect, submit slurm jobs

  -> Shared resource, do not run your jobs here!



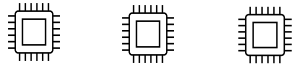
- Computational nodes

  c01 128 cores 1Tb Ram
  c02 128 cores 1Tb Ram
  c03 128 cores 1Tb Ram GPU

  = computing (job execution via SLURM)

  USE: to execute parallel or high-mem jobs

  -> Shared resource, use wisely!



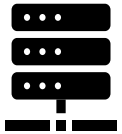
- Flash storage '/mnt/beegfs/'

  100Tb SSD memory

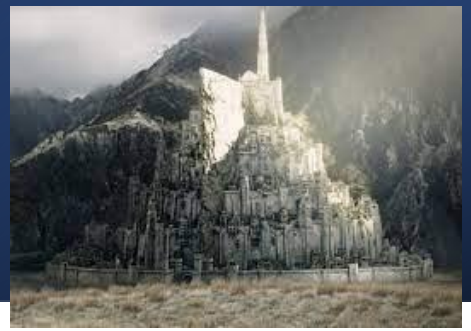
  = flash memory attached to computational nodes for fast read/write access

  USE: to temporarly store data

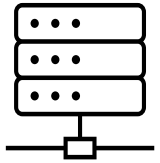
  -> Shared resource, use wisely!



IJC cluster components



ISILON DELL



= centralized, backed-up storage, distributed filesystem

/ijc/LABS/$GROUP/RAW = 45TB (raw data, static)
/DATA = 5TB (analysis data, dynamic)
/LTS (long term storage, archive data for tape back-up)

/ijc/USERS/$USER = 500GB (user directory)

/ijc/PROJECTS/$PROJECT = on request (project directory with special permissions created by IT upon request)

NOTE: If you your computer is IT-administered, IT can mount the ISILON in your computer (This allows you to utilize graphical applications installed on your local machine with data, that is located on the ISILON.)

Questions?