

Introduction to IJC computational infrastructure

Angelika Merkel (Head of Bioinformatics Core Facility)
10/04/2025



Angelika Merkel
Unit Manager
Bioinformatician



Emilio Lario
Software Engineer



Marta Meroño
Bioinformatician
(Student)

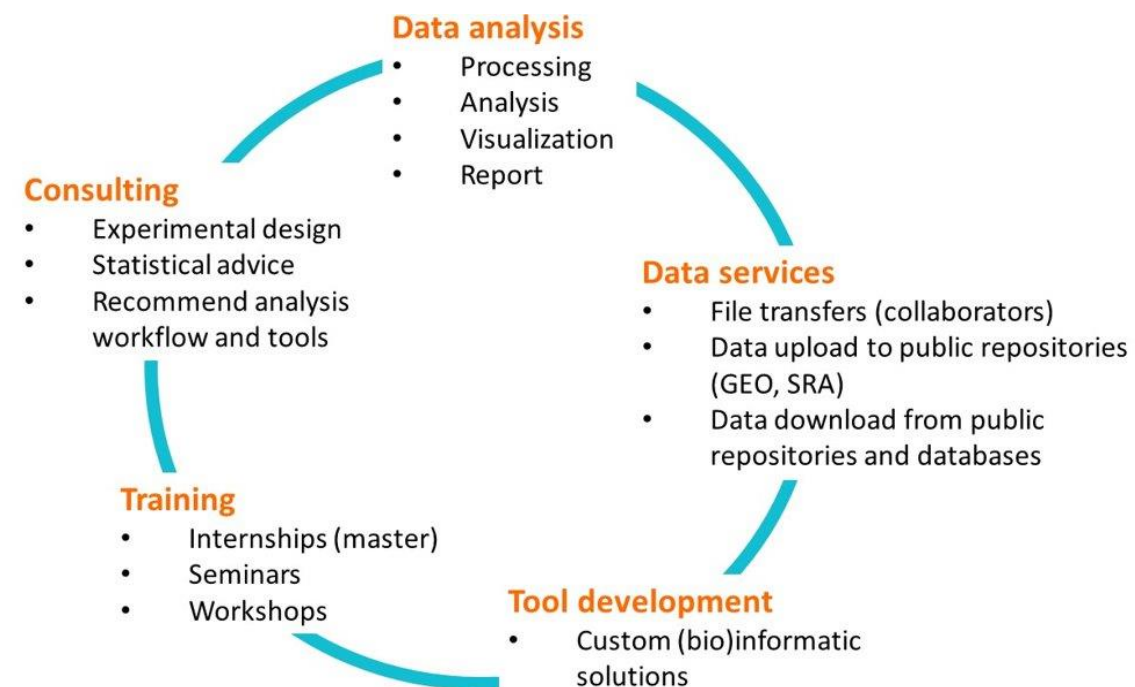


Paula Vela
Bioinformatician
(Student)

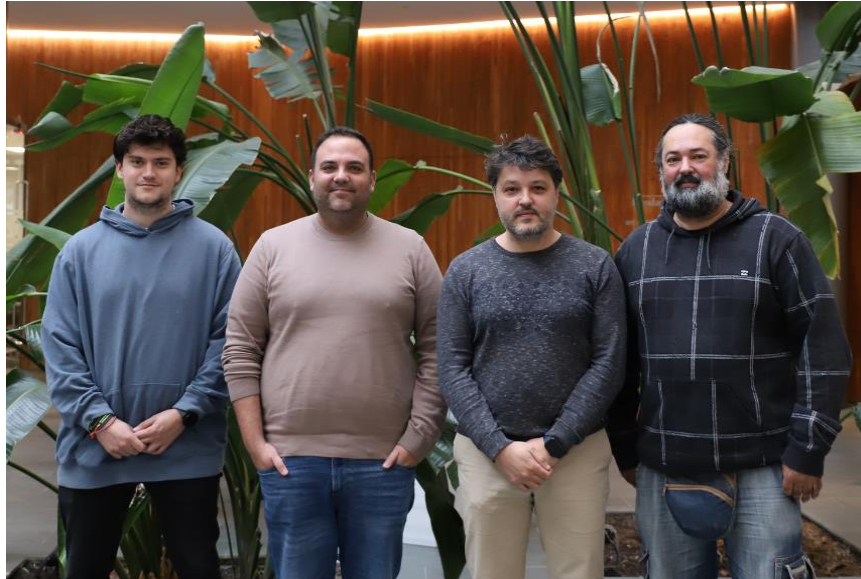
Office: Sala Prof. Albert Grañena (1st floor)(1st floor); phone: 4300

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

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



IT Team



Roger España
(Helpdesk)

José Monterde
(System
administrator,
Windows support)

Marc
Jubany
(Head of IT)

José Alcántara
(HPC administration/
Linux support)



Angel Borrás
(student)

Office: 0.00.04 (ground floor)

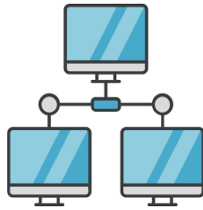
Email: helpdesk@carrerasresearch.org
<https://tickets.carrerasresearch.org/>

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)

SHARED FOLDER (300GB)



- Folders shared with specific groups or people. Created by the IT department upon request.
- Some default folders shared with all groups:
TRAVEL_UNIT, LABORATORIES, UNITAT_MICROSCOPIA, RRHH
- **For Office documents**
- **Back-up**



COMMON_EQUIPMENT

- Data produced by specific equipment
- Shared folder for each research group created upon request by platform manager
- **For raw data (temporary storage only!)**
- **Back-up**

High-performance computing (HPC) cluster



- Computational nodes
- Master node minastirith
- ISILON file storage system (default quota 5TB)
- **For Data, large files**
- **Back-up**



NAS BD*GROUP (default quota 5TB)

- Shared folder for each research group. Contains by default a private folder for the group leader and a shared folder accessible by anyone in the group
- **For Office documents, results and images**
- **Back-up (internal)**

Shared network resources (IJC only)

Access to shared network resources

1. IT administered machines:

- Windows (via file explorer):
 - SHARED_FOLDER; BD*GROUP, COMMON_EQUIPMENT are mapped automatically
 - Folders on the ISILON are mapped on request
- Linux (via file explorer or command line):
 - Mounting of any shared network resource upon request

2. Vpn portal: <https://vpn.carrerasresearch.org>

- Predefined bookmarks > Nextcloud_IJC > SHARED_FOLDER
- Predefined bookmarks > COMMON_EQUIPMENT
- Predefined bookmarks > Minas Tirith (access to ISILON via terminal)

File sharing (via web browser)

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



SHAREPOINT/ONE DRIVE (MICROSOFT)

- <https://carrerasresearch-my.sharepoint.com/>
- Personal access (default quota 1TB)
- **Pros:** Desktop compatible, integrated with office 365
- **Cons:** Data is located on MS servers off site



NEXTCLOUD

1. Core facilities:

- <https://nc1.carrerasresearch.org/>
- Share data files or results with collaborators

2. 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 SFTP



Transfer files from inside the IJC network to the outside and vice versa

SFTP

- transfer large data files
- 5TB temporary storage shared amongst users
- Host: ftpbios.carrerasresearch.org Port 4242
- Access from the inside and outside **only with credentials (only IJC users)**

Transfer files from and to the server with:

- 1) ftp client like Filezilla or
- 2) via Terminal

```
sftp -p 4242 username@ftpbios.carrerasresearch.org
> put myfile
> get myfile
```

File sharing via anonymous FTP



Sometimes there is a need to share files publicly without the use of credentials (e.g. for software such as the UCSC genome browser))

Anonymous (public) FTP

- Allows outside access via the internet **without credentials** (read only!)

URL: <http://homepages.carrerasresearch.org/~username/>

- Access from the inside with your IJC credentials
 - Host: bios.carrerasresearch.org
 - IP: 10.110.20.23 port 22
- 150GB temporary storage shared amongst users

Transfer files to the server with:

- 1) ftp client like Filezilla or
- 2) via Terminal

```
sftp bios.carrerasresearch.org
cd public_html          # if folder does not exist, please create it
> mput myfile
> get myfile
```

Filesharing tools

Interface	Tool	Access	Recommended for	File location
Web browser (office 365), file explorer	OneDrive	https://carrerasresearch-my.sharepoint.com/	Documents	Windows servers (off-site)
Web browser	NextCloud (core facilities only)	https://nc1.carrerasresearch.org/	Documents, data <1GB	IJC servers (on-site)
Filezilla	Anonymous ftp, sftp	Available for MacOS, Windows, Linux	Data > 1GB	IJC sftp server (on-site)
Terminal	Anonymous ftp	https://homepages.carrerasresearch.org/~username/ , or sftp username@bios.carrerasresearch.org > put my_file; > met my_file	Public data < 5GB	IJC sftp server (on-site)
Terminal	Sftp (secure)	sftp -P 4242 username@ftpbios.carrerasresearch.org > put my_file; > get my_file	Data > 1GB	IJC sftp server (on-site)

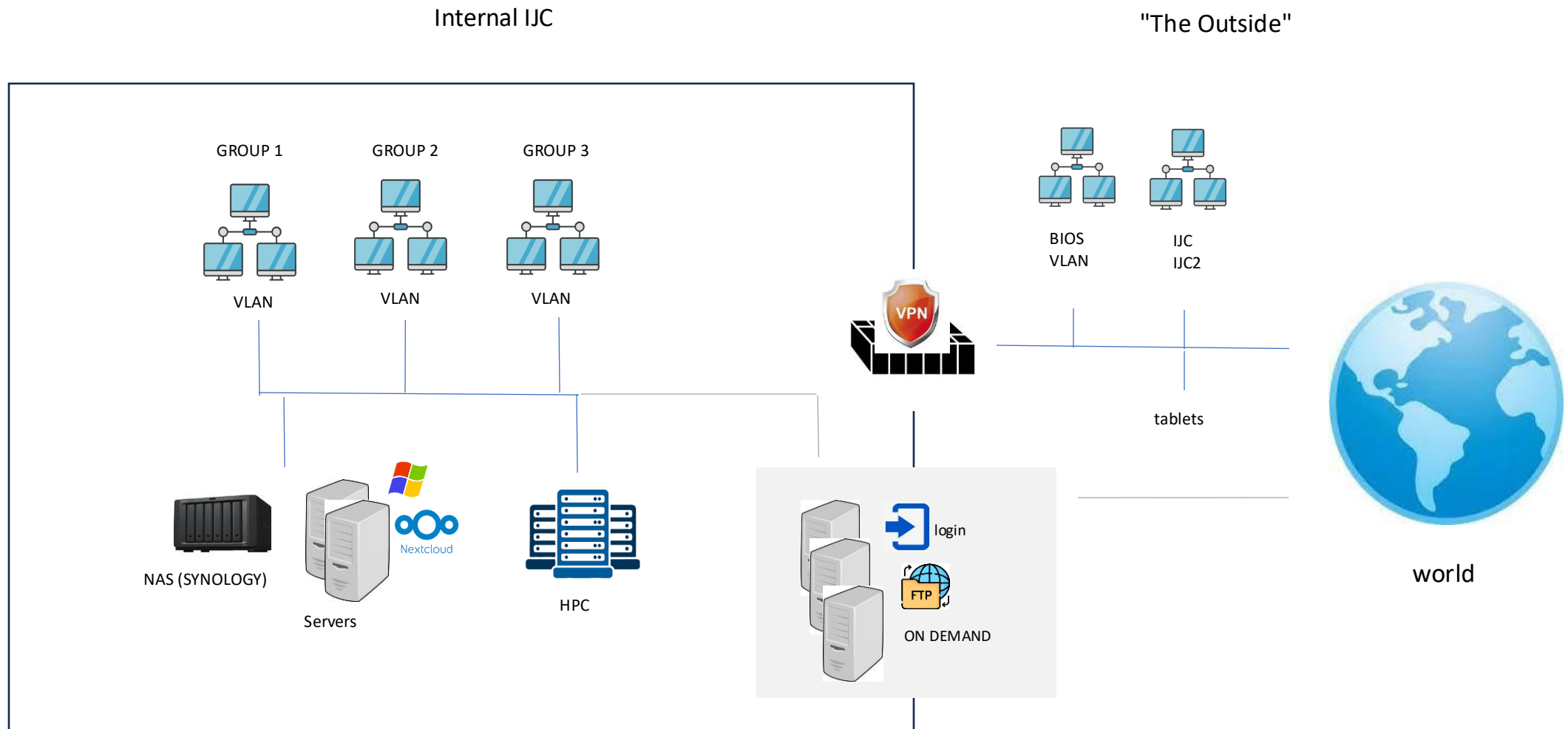
Data management best practices:

- Important **documents and data should be stored** with centralized management and frequent backups.
- **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 responsible (user/user group).
- **After the employee leaves, user data (personal folders) are deleted and the accounts is terminated** .(except email access which is valid 6months). Before leaving the institute ,the user shall remove irrelevant data, assure important data is kept accordingly and is well documented.
- **Bioinformaticians best practice:** Use RELATIVE PATHS in your scripts and avoid ABSOLUTE PATHS

Avoid multiple copies of data!

2. The IJC network & remote connections

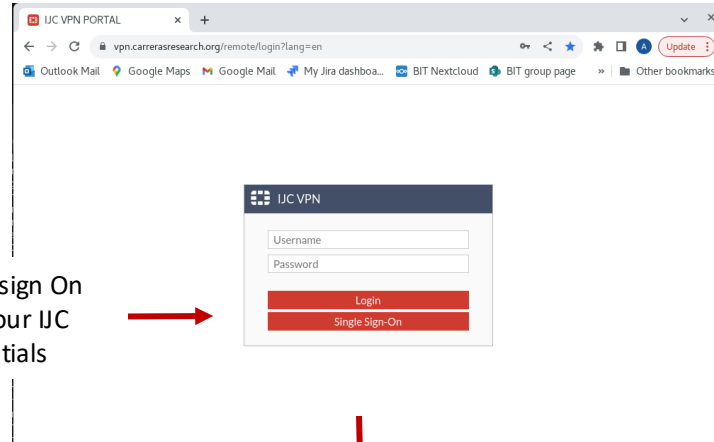
The network



The IJC VPN portal

Point your browser to: <https://vpn.carrerasresearch.org/>

Single-sign On
with your IJC
credentials

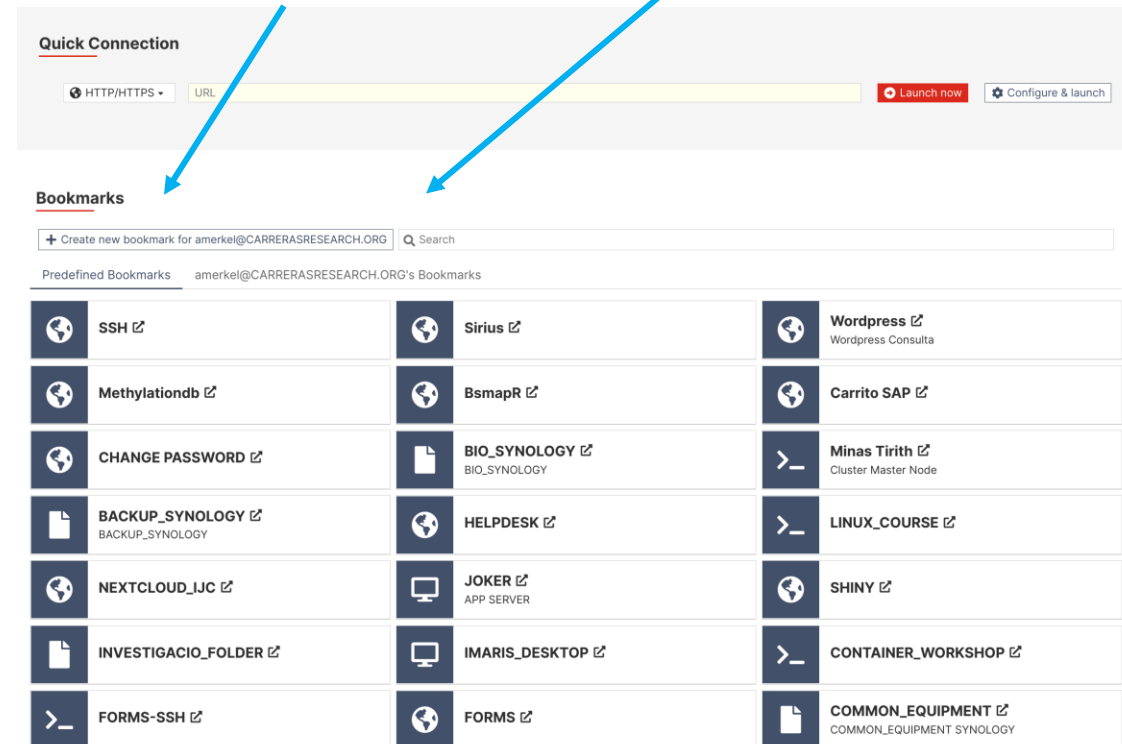


2 – Factor- Authentication



Shared bookmarks
(by IT)

Personal bookmarks (only
bioinformatics profile)



Predefined Bookmarks		
amerkel@CARRERASRESEARCH.ORG's Bookmarks		
SSH	Sirius	Wordpress
Methylationdb	BsmarP	Wordpress Consulta
CHANGE PASSWORD	BIO_SYNLOGY	Carrito SAP
BACKUP_SYNLOGY	BIO_SYNLOGY	Minas Tirth
NEXTCLOUD_IJC	HELPDESK	Cluster Master Node
INVESTIGACIO_FOLDER	JOKER	LINUX_COURSE
FORMS-SSH	IMARIS_DESKTOP	SHINY
	FORMS	CONTAINER_WORKSHOP
		COMMON_EQUIPMENT
		COMMON_EQUIPMENT SYNLOGY

The IJC VPN portal

Remotely access your workstation ('remote desktop') through a personal bookmark :
(requires that IT grants access)

SSL-VPN Portal BIOS
Powered by **Fortinet**


Quick Connection

HTTP/HTTPS URL

Bookmarks

+ Create new bookmark for amerkel@CARRERASRESEA

Predefined Bookmarks amerkel@CARRERASRES

 **IJCDesktop**

Edit Bookmark

Protocol type RDP

Name IJCDesktop

Host 10.100.106.122

Port 3389

Description

Use SSL-VPN credentials ☒

Username amerkel

Password Change

Color depth per pixel 16bits per pixel

Resolution type Preset Custom

Resolution 0 width 0 height

Keyboard layout English, United States.

Security Allow the server to choose the type of

Send preconnection ID ☒

Load balancing information

Restricted admin mode ☒

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, your workstation or any other machine you have access to (and know the IP adress):

```
ssh minastirith
```

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 and configured passcode (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, your workstation or any other machine you have access to (and know the IP address):

```
ssh minastirith
```

Or 'ssh jump' directly:

```
ssh -J username@loginnode.carrerasresearch.org:4242 username@minastirith
```

The login node

You can also copy file from one location to another via ssh using the command 'scp':

1) To a remote host:

```
scp originfile amerkel@minastirith:/destinationPath/
```

2) From a remote host:

```
scp amerkel@minastirith:/originfile /local/destinationPath/
```

Analogous to a 'ssh jump' you can also copy via an intermediate host:

1) To a remote host (via intermediate)

```
scp -o "ProxyJump=username@loginnode.carrerasresearch.org:4242" originfile amerkel@minastirith:/destinationPath/
```

2) From a remote host (via intermediate)

```
scp -o "ProxyJump=username@loginnode.carrerasresearch.org:4242" amerkel@minastirith:/originfile /local/destinationPath
```

IT administrated & non-IT administrated computers

IT

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

Pros:

- IJC network 600Mb/s CSUC internet
- BD*Group, COMMON_EQUIPMENT, SHARED_FOLDER can be mounted
- ISILON filesystem can be mounted
(-> allows file access with GUI applications)
- Connection from the outside directly to the IJC network with vpnclient (e.g. Linux via openvpn 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)

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, pass them to IT, they will transfer the content safely
- 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

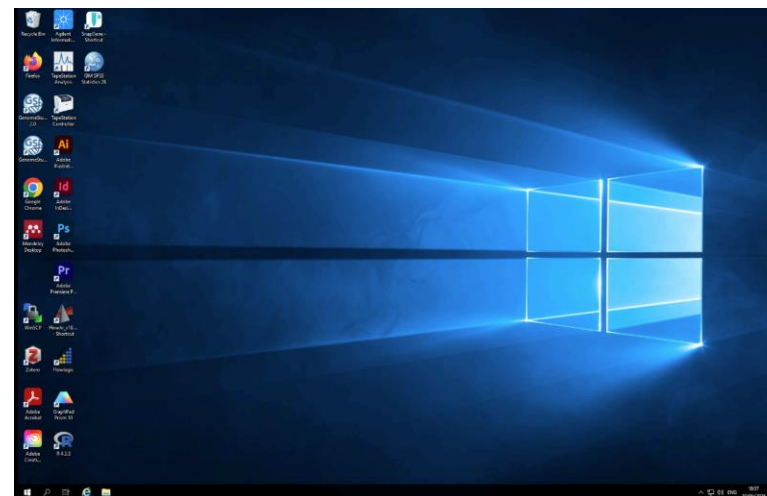
The Joker

Virtual windows desktop with licensed software:
(max 50 users at a time, after 30min off inactivity session is terminated)

- 1) ADOBE CREATIVE CLOUD *
- 2) GRAPHPAD PRISM 10 *
- 3) SNAPGENE
- 4) SPSS
- 5) Genome Studio
- 6) Flowlogic, FlowJo (1 user at a time)

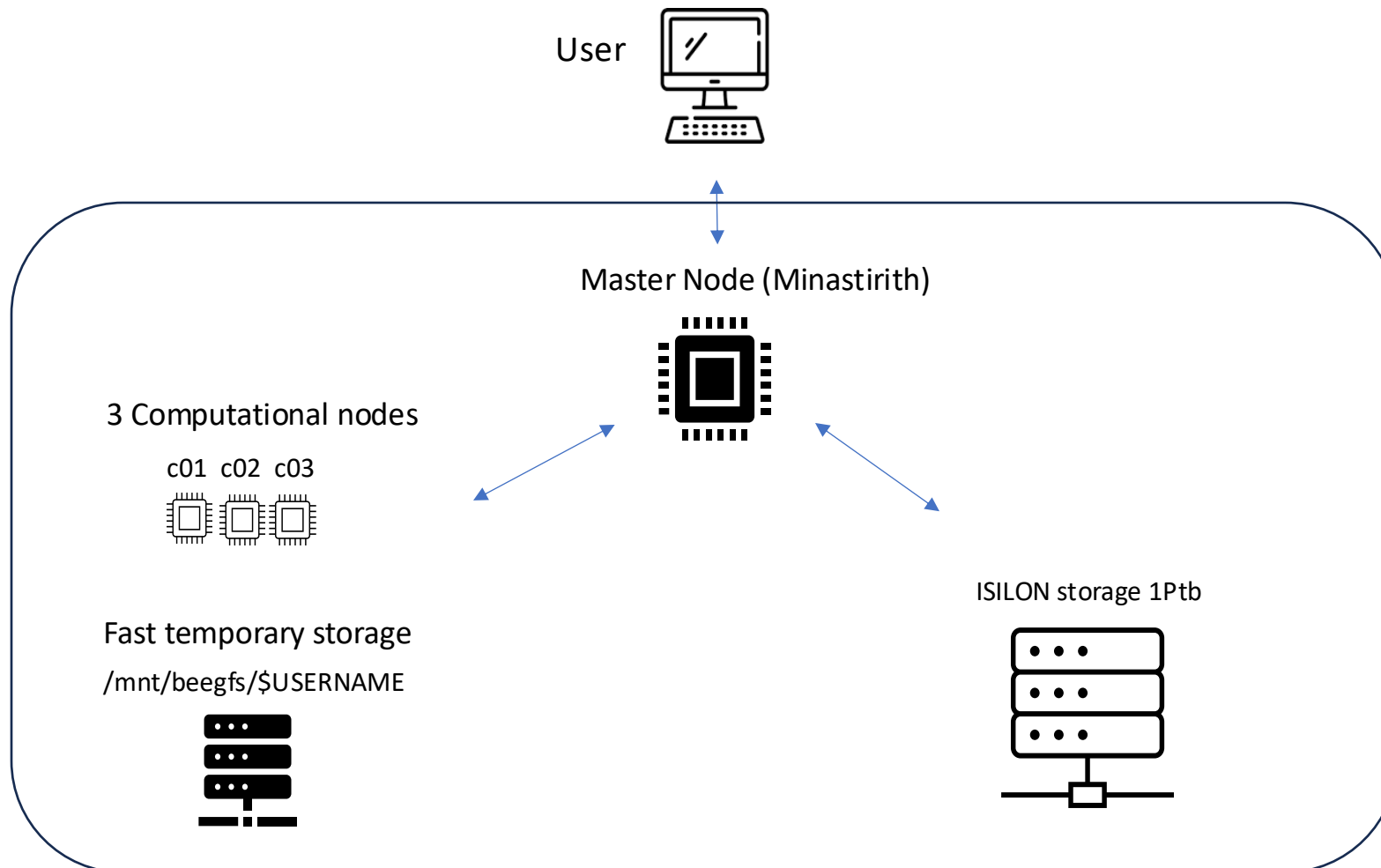
Access:

- 1) [Vpn.carrerasresearch.org](https://vpn.carrerasresearch.org) > predefiend bookmarks > joker
(Remote screen resolution can be configured in a connection through a personal bookmark)
- 2) Inside IJC: for Windows/Macs use 'remote desktop' app (> joker with your credentials)



3. High Performance Computing (HPC)

IJC HPC cluster



IJC cluster components



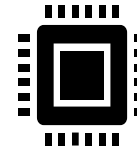
- Master Node (Minastirith)

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

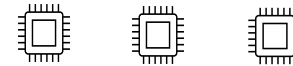
c02 128 cores 1Tb Ram

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 temporarily store data

-> Shared resource, use wisely!

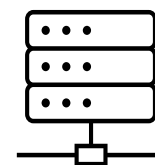


IJC cluster components



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) by request

/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?



BIT training materials are available [here](#):

Documentation on IT resource and procedures:
<https://howto.carrerasresearch.org/>