Storage & Data Transfer Service

dicos-support@twgrid.org

Academia Sinica Grid-computing Centre (ASGC)

2025/01/15 Alice Lin

Outline

- Storage space
- Introduction of Home Space
- Introduction of Group Working Space
- Subscribe to more Group Working Space
- Unsubscribe Space
- Working Space for Computing Jobs
- Using /tmp on the worker node as Intermediate Working Space
- Examples

Storage space

- Home Space (100GB for free, not extendable)

  /dicos_ui_home/{user_account}

- Group Working Space (3TB free /each group, pay for extension)

  /ceph/work/{group}

- [cryoEM group]

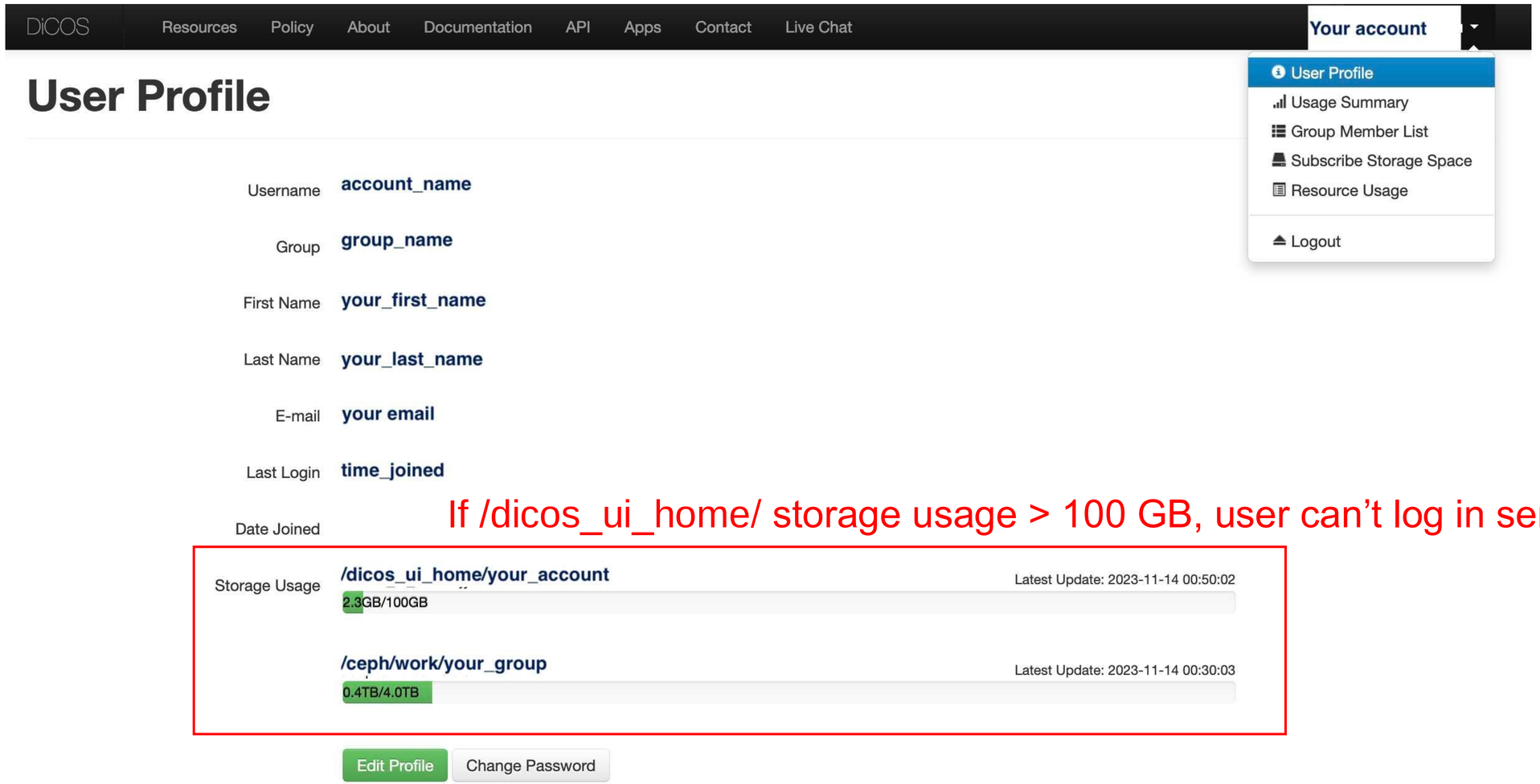
  /activeEM/data/{group}/{user}

- Backup Space (will be available in the future)

- Please **back up to secure your data** in these spaces.

Storage space

- Check your usage on https://dicos.grid.sinica.edu.tw/profile/info



If /dicos_ui_home/ storage usage > 100 GB, user can’t log in server.

Introduction of Home Space

/dicos_ui_home/{user_account}

- 100 GB for free, not extendable.
- Backup your important data in Home Space.
- /dicos_ui_home/ is a temporary space, backup your data in the /ceph/work/{group} or transfer to your own data storage.
- Home space will be cleared once your account is expired and deleted.
- **Avoid using the Home Space as the job’s working space**.

Introduction of Group Working Space

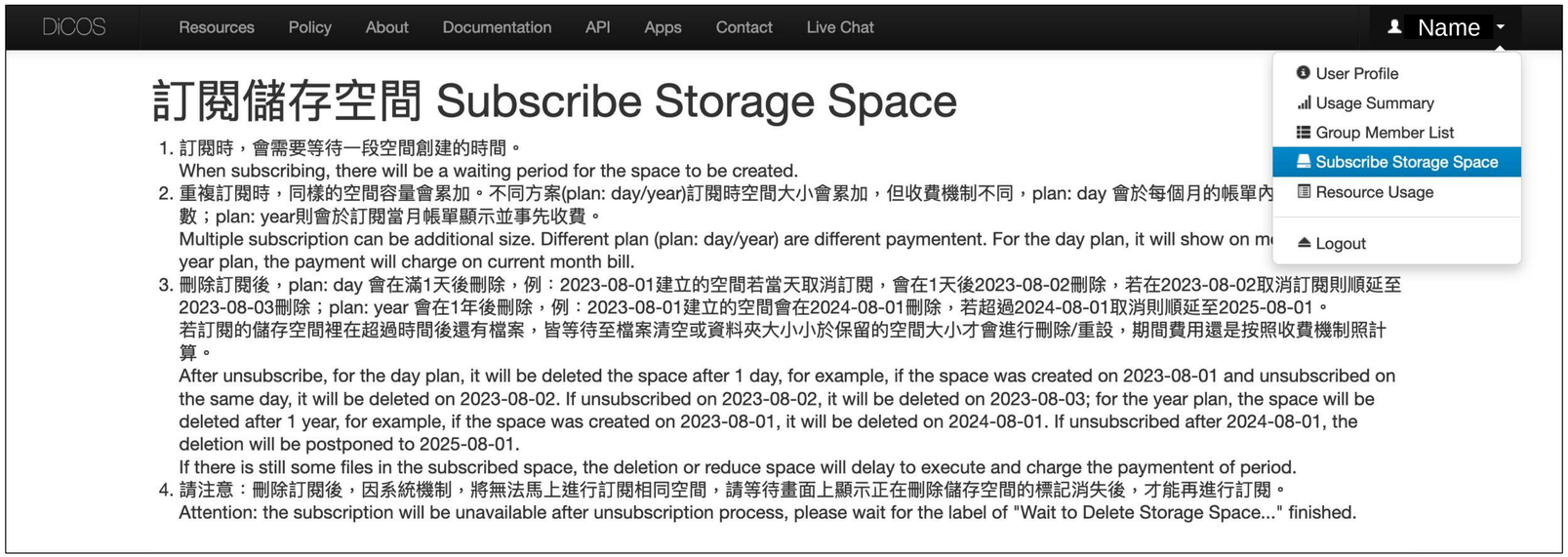
/ceph/work/{group}

- 3TB free space for each group.
- This space is shared among all members of the group.
- PI has full access and manage rights in this space.
- PI can subscribe & upgrade for more space via dicos.grid.sinica.edu.tw.
- Backup your important data in Group Working Space.

Subscribe to more Group Working Space

- PI can subscribe for upgrade storage space.

https://dicos.grid.sinica.edu.tw/accounting/subscribe-storage



1TB Storage: NT$3/day or NT$1000/year

More pricing information at https://dicos.grid.sinica.edu.tw/resources

Subscribe to more Group Working Space

Current Storage Space

Unsubscribe Space

- Clear your data and release your space before unsubscription.
- Unable to unsubscribe if the space is not released.

  → remove data until less or equal to target subscription.
- Unsubscription process will be blocked when

  space can not be released or cancel year plan.
- Contact info: dicos-support@twgrid.org

Working Space for Computing Space

- Group Working Space: /ceph/work/{group}

- If your job has high I/O wait output:

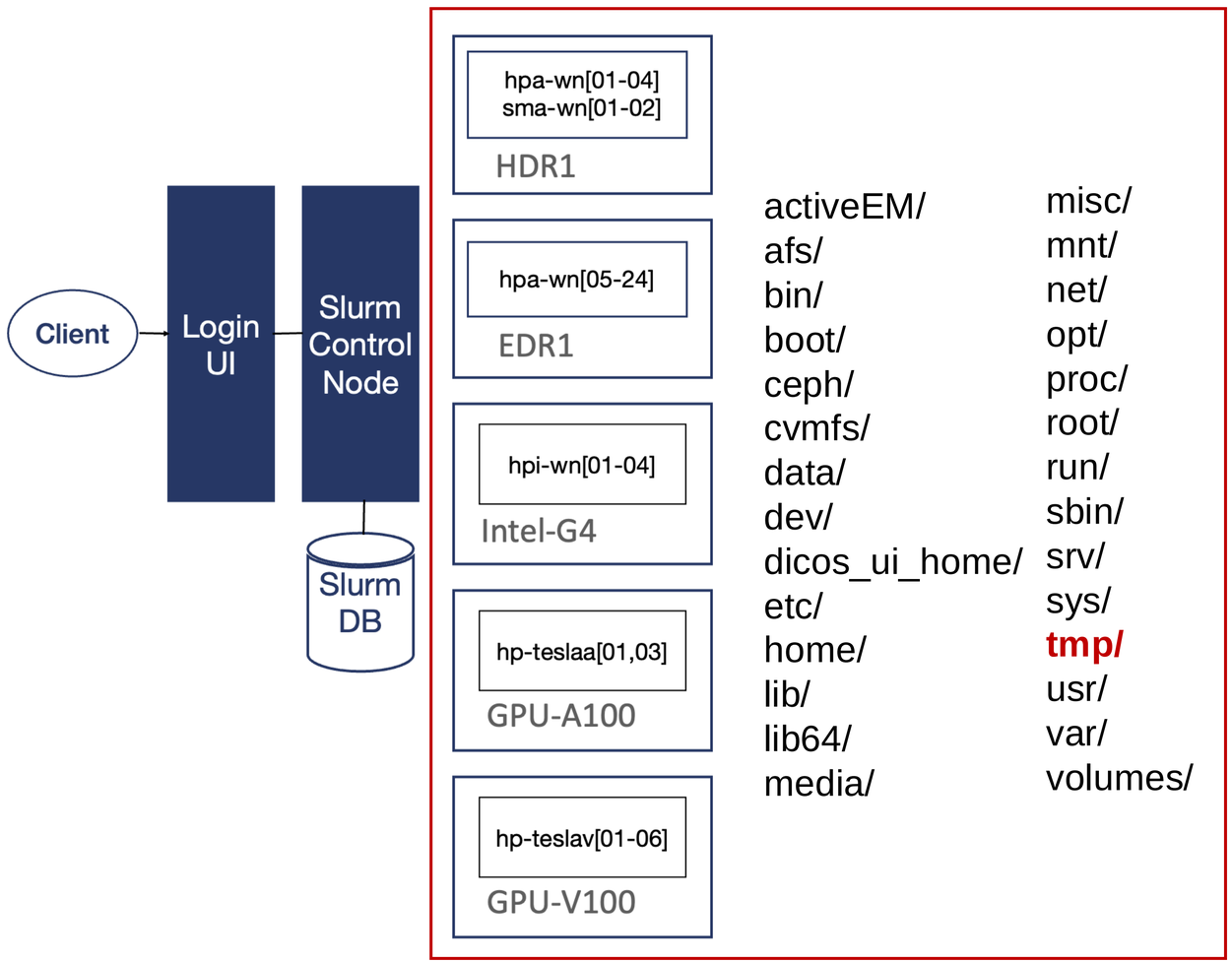
  You can also use **/tmp on the worker node** as an intermediate job working space.

  Copy intermediate data to /ceph/work/{group}.

Using /tmp on the worker node as Intermediate Working Space

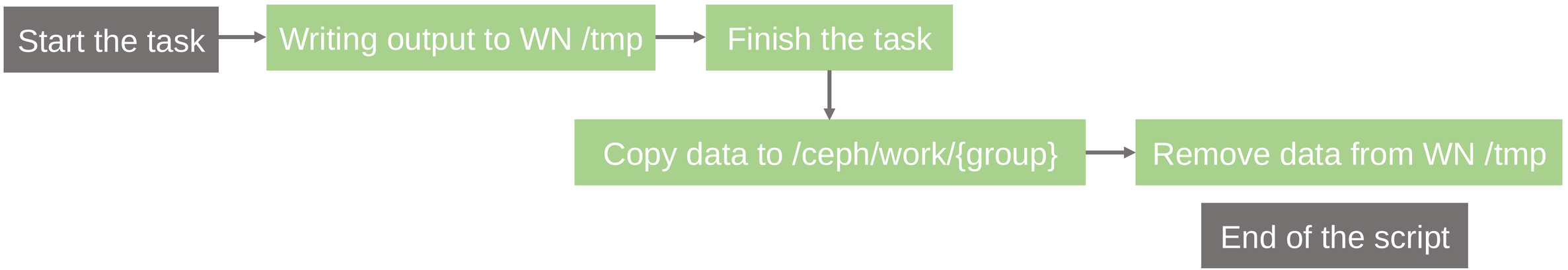
- Each worker node has a different spec (e.g., SSD) & capacity.

Cluster	Local Disk SSD
HDR1	1 TB
EDR1	2 TB
Intel-g4	2 TB



Using /tmp on the worker node as Intermediate Working Space

- Copy intermediate data to /ceph/work/{group}.
- Remove files from **/tmp on the worker node** when tasks finish.
- /tmp directory on the worker node will be cleaned anytime.



WN: worker node

Examples

Linux Basic Command

Using /tmp on the worker node as Intermediate Working Space

Data Transfer

Linux Basic Command

ssh username@slurm-ui.twgrid.org ssh username@dicos-sftp.twgrid.org	Login
pwd	Show current path
mv {src} {dest}	Move file
rm filename	Remove file
rm –r dirname	Remove directory
cat filename	View file content

Using /tmp on the worker node as Intermediate Working Space

```
# Create the temporary file
[username@slurm-ui01 ~]$ mktemp
/tmp/1mjAhAH3YM

[username@slurm-ui01 ~]$ mktemp /tmp/user-XXXXX
/tmp/user-WxWtv



# Create the temporary directory
[username@slurm-ui01 ~]$ mktemp –d
/tmp/tmp.1mjAhAH3YM

[username@slurm-ui01 ~]$ mktemp –d /tmp/user-XXXXX
/tmp/user-WxWtv
```

Using /tmp on the worker node as Intermediate Working Space

Job_TemporaryFile.sh

```
#!/bin/bash
#SBATCH --job-name=helloword               # Job name
#SBATCH --partition=intel-g4-al9_short       # Partiotion name
#SBATCH --nodes=1                       # Numbers of nodes
#SBATCH --ntasks-per-node=1               # Number of tasks per node
#SBATCH --cpus-per-task=1                # Number of CPUs per task
#SBATCH --output=%j.out                  # Standard output file (%j: Job ID)
#SBATCH --error=%j.err                   # Standard error file (%j: Job ID)
#SBATCH --time=00-01:00:00                # Time limit
#SBATCH --mail-type=ALL                  # Email notifications = BEGIN, END, FAIL, ALL
#SBATCH --mail-user=jennifer.chen@twgrid.org  # Email address to send notifications


# Create a temporary file with a unique name under /tmp/ directory
TMPFILE=$(mktemp /tmp/yiruchen0101-XXXXX)

# Get the filename (without the path) of the temporary file
filename=$(basename "$TMPFILE")

#your computing program
echo "Hello User" >> $TMPFILE
hostname >> $TMPFILE

# Copy the temporary file to the specified destination directory (/ceph/work/{Group}/{your directory})
cp $TMPFILE /ceph/work/ASGC/yiruchen0101/$filename
# Remove the temporary file from /tmp/ directory to clean up
rm $TMPFILE
```

Using /tmp on the worker node as Intermediate Working Space

```
# Submit slurm job
[yiruchen0101@slurm-ui01 yiruchen0101]$ sbatch Job_TemporaryFile.sh
 Submitted batch job 965001
[yiruchen0101@slurm-ui01 yiruchen0101]$ ls -l /ceph/work/ASGC/yiruchen0101/yiruchen0101-to7Sw
 -rw-------+ 1 yiruchen0101 ASGC 31 Jan 14 16:48 yiruchen0101-to7Sw


[yiruchen0101@slurm-ui01 yiruchen0101]$ cat yiruchen0101-to7Sw
 Hello User
 hpi-wn01.twgrid.org
```

Using /tmp on the worker node as Intermediate Working Space

Job_TemporaryDir.sh

```
#!/bin/bash
#SBATCH --job-name=helloword               # Job name
#SBATCH --partition=intel-g4-al9_short       # Partiotion name
#SBATCH --nodes=1                        # Numbers of nodes
#SBATCH --ntasks-per-node=1                # Number of tasks per node
#SBATCH --cpus-per-task=1                  # Number of CPUs per task
#SBATCH --output=%j.out                    # Standard output file (%j: Job ID)
#SBATCH --error=%j.err                     # Standard error file (%j: Job ID)
#SBATCH --time=00-01:00:00                 # Time limit
#SBATCH --mail-type=ALL                    # Email notifications = BEGIN, END, FAIL, ALL
#SBATCH --mail-user=jennifer.chen@twgrid.org  # Email address to send notifications


# Create a temporary directory with a unique name under /tmp/ directory
TMPDIR=$(mktemp –d /tmp/yiruchen0101-XXXXX)

# Get the directory name (without the path) of the temporary direcotry
filename=$(basename "$TMPDIR")

#your computing program
echo "Hello User" >> $TMPDIR/TMPD.txt
hostname >> $TMPDIR/TMPD.txt

# Recursively copy the entire temporary directory to the specified destination directory (/ceph/work/{Group}/{your directory})
cp –r $TMPDIR /ceph/work/ASGC/yiruchen0101/$filename
# Remove the temporary directory and its contents from /tmp/ directory to clean up
rm –r $TMPDIR
```

Using /tmp on the worker node as Intermediate Working Space

```
# Submit slurm job
[yiruchen0101@slurm-ui01 yiruchen0101]$ sbatch Job_TemporaryDir.sh
 Submitted batch job 409355
[yiruchen0101@slurm-ui01 yiruchen0101]$ ls -l /ceph/work/ASGC/yiruchen0101/yiruchen0101-HeycZ
 drwx------+  2 yiruchen0101 ASGC       1 Jan 14 17:32 yiruchen0101-HeycZ

[yiruchen0101@slurm-ui01 yiruchen0101]$ cat yiruchen0101-HeycZ/TMPD.txt
 Hello User
 hpi-wn01.twgrid.org
```

Data Transfer

- sftp
- scp
- FileZilla

Data Transfer using sftp

Node for Data Transfer: **slurm-ui.twgrid.org** or **dicos-sftp.twgrid.org**

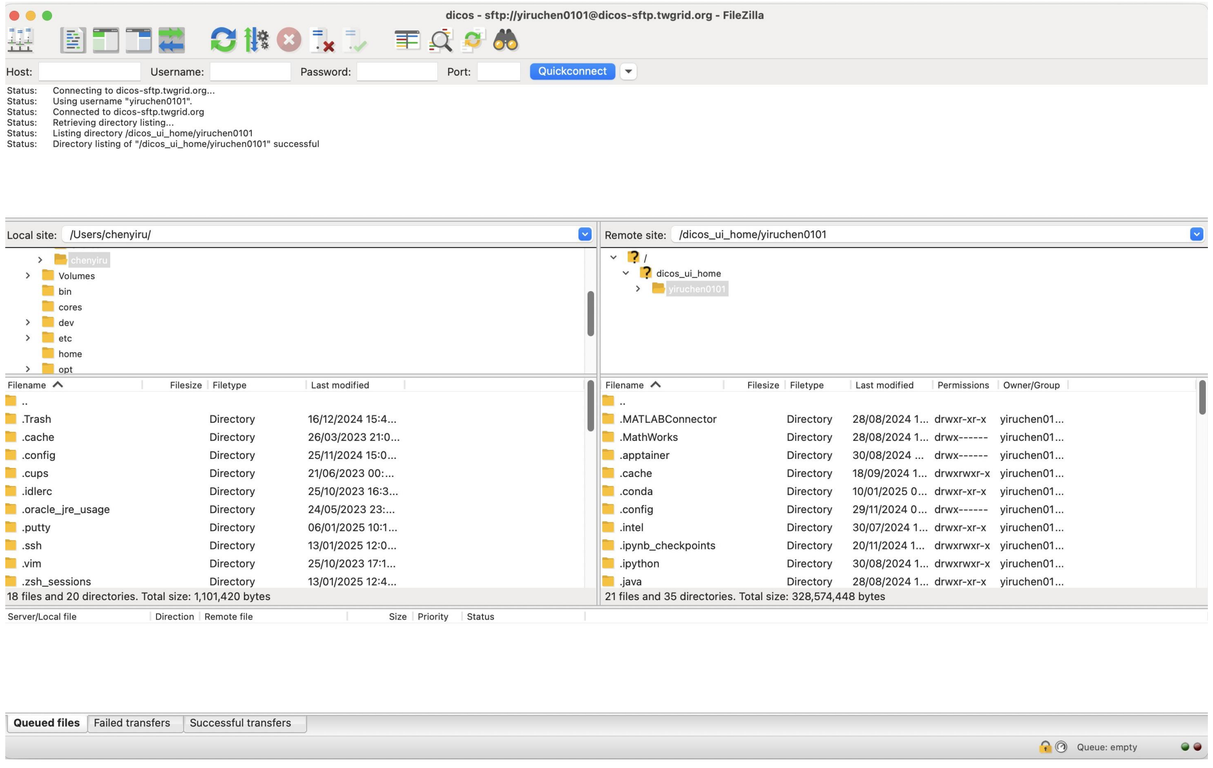
```
$ sftp username@slurm-ui.twgrid.org

(username@slurm-ui.twgrid.org) Password:
(username@slurm-ui.twgrid.org) Verification code:

Connected to slurm-ui.twgrid.org.

# Upload file

sftp> put filename
Uploading filename to /dicos_ui_home/username/

# Download file

sftp> get filename
Fetching  filepath to filename
```

Data Transfer using scp

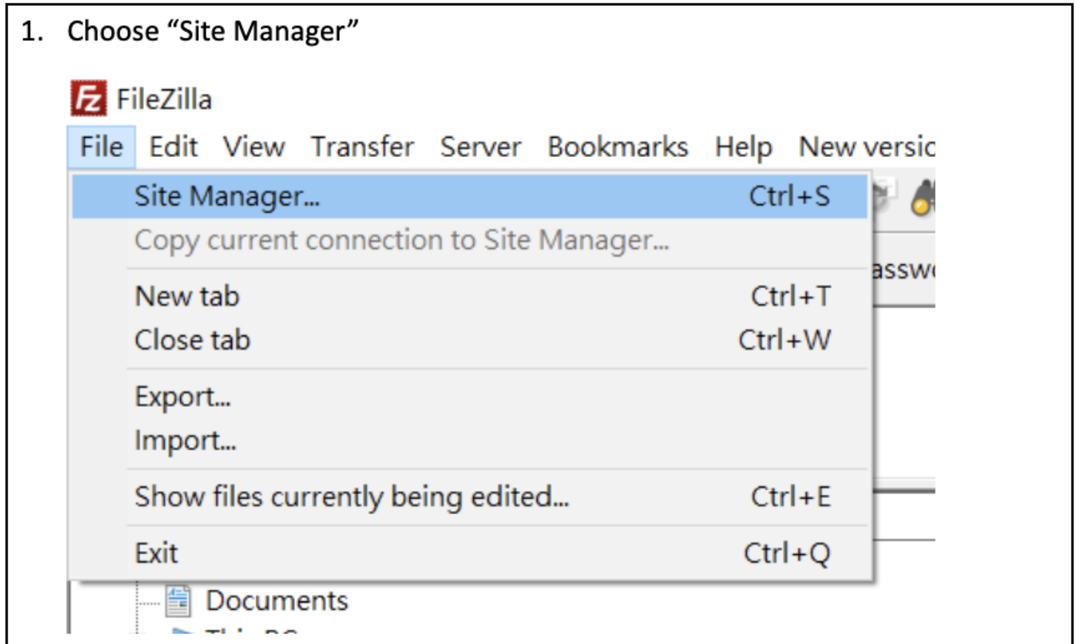
Node for Data Transfer: **slurm-ui.twgrid.org** or **dicos-sftp.twgrid.org**

```
# Copy specific file
$ scp username@slurm-ui.twgrid.org:/ceph/work/Others/mydata/this.file .

(username@slurm-ui.twgrid.org) Password:
(username@slurm-ui.twgrid.org) Verification code:


# Copy the whole directory within the path
$ scp –r username@slurm-ui.twgrid.org:/ceph/work/Others/mydata .

(username@slurm-ui.twgrid.org) Password:
(username@slurm-ui.twgrid.org) Verification code:
```

Data Transfer using FileZilla

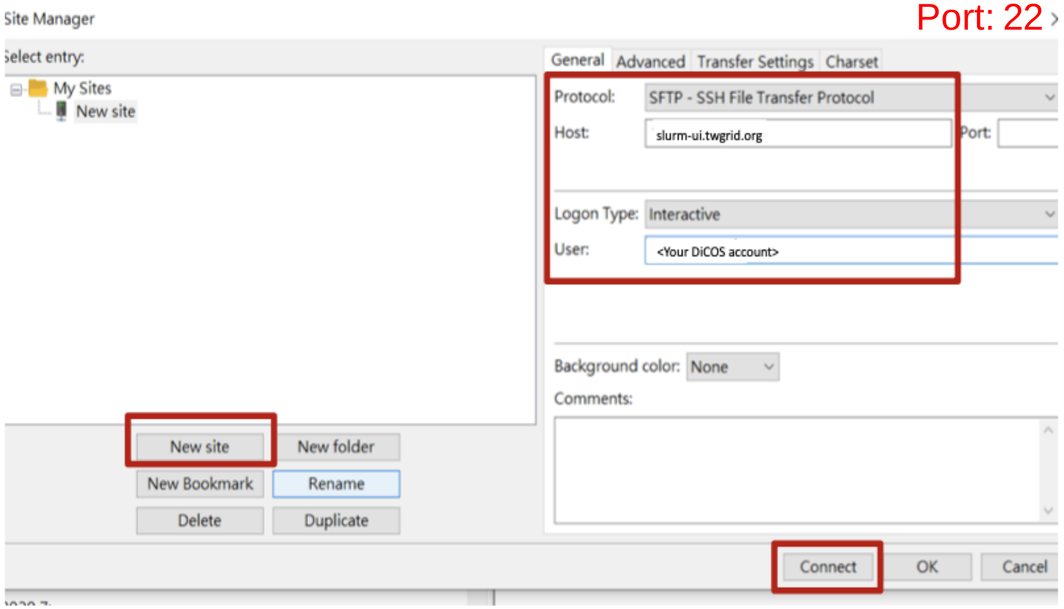


1. Choose “Site Manager”



Data Transfer using FileZilla

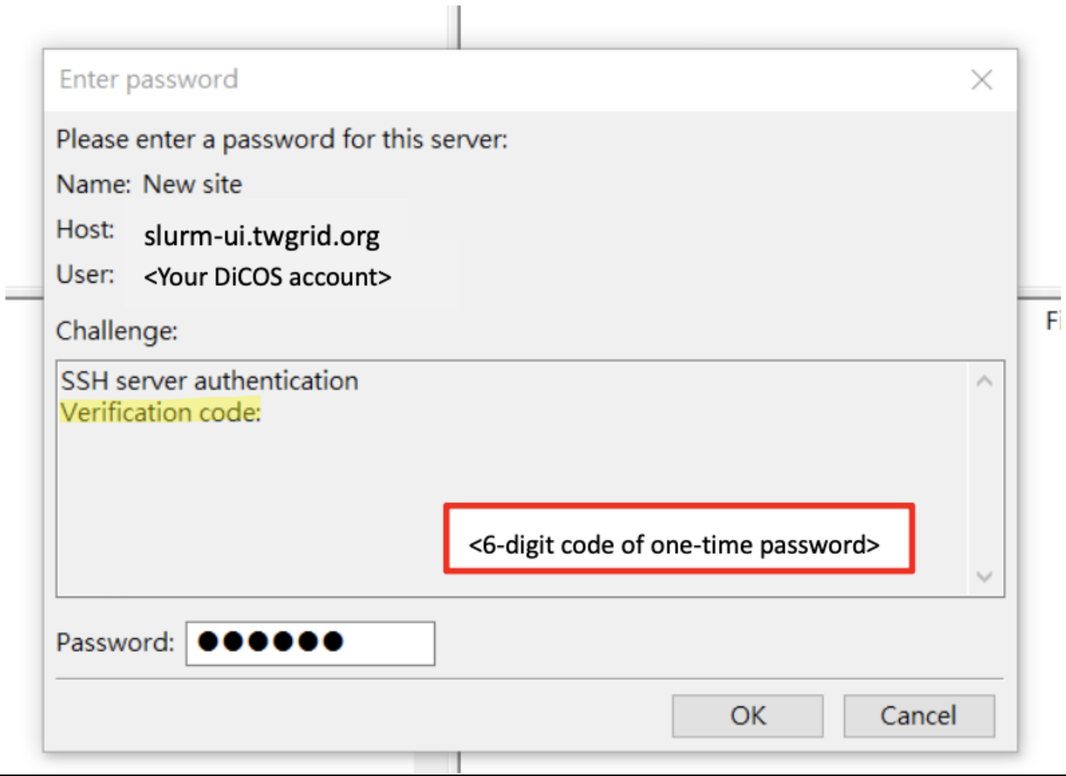
2. Click “New site” -> choose Protocol to “SFTP” -> enter Host “slurm-ui.twgrid.org” -> choose Logon Type to “Interactive” -> enter your DiCOS account to User field -> click “Connect”



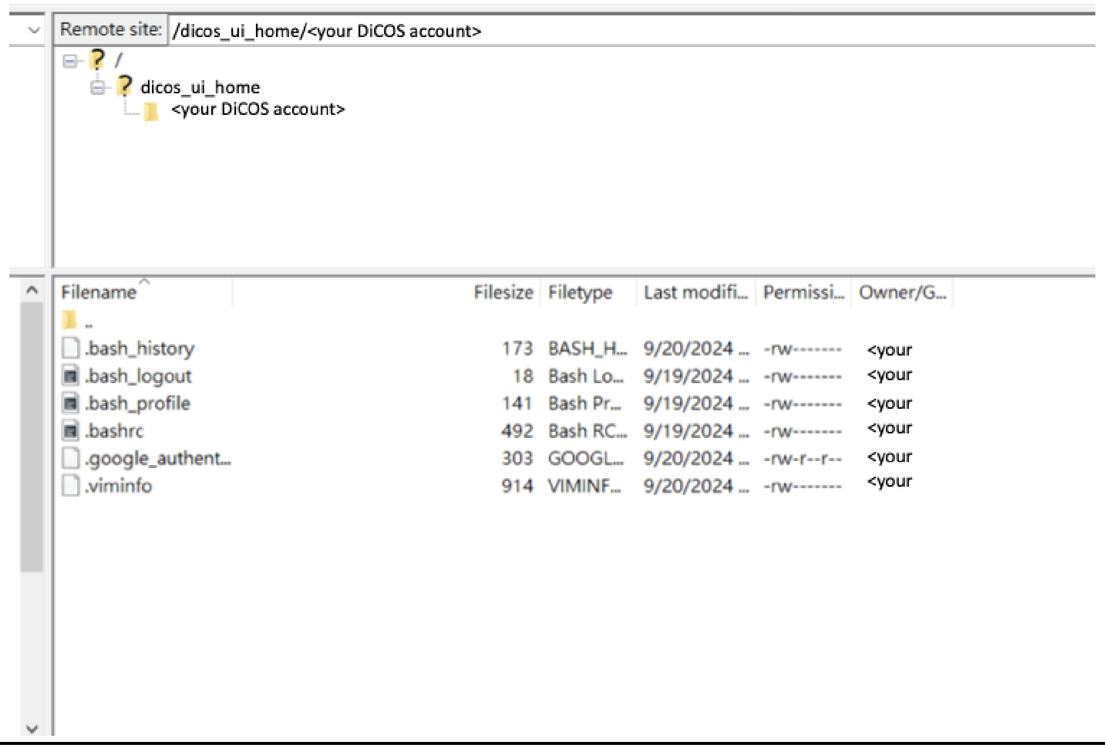
3. After connect, the first pop-up window please enter your DiCOS password

Data Transfer using FileZilla

4. The second pop-up window please enter 6-digit code of OTP shows on Google authenticator



5. After two authentication successed, you can see the right window list your DiCOS home directory



Thanks for listening

Q&A

dicos-support@twgrid.org

Academia Sinica Grid-computing Centre (ASGC)

2025/01/15 Alice Lin