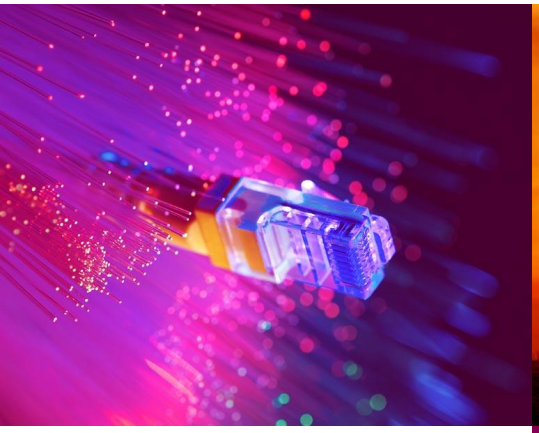

Webinar: Moving data between CSC and local environment

Ari-Matti Sarén, Application specialist, CSC

CSC – Finnish research, education, culture and public administration ICT knowledge center

Some brief generalizations:

- Directories containing tens of thousands of files are often problematic
  - use subdirectories and/or aggregation
  - data should always be packaged for saving in HPC Archive server or IDA
- It's usually faster to move one large file than many small ones
  - depends on transfer protocol
- On the other hand you should avoid too large files
  - it's nicer to re-send one 100 GB chunk than the whole 1 TB file
- Consider compression
  - text based file formats (*e.g.* .csv, .tsv etc) usually compress well

Check your files after transfer:

Quick checks:

size:

```
ls –l example.file
```

number of lines:

```
wc –l example.file
```

Checksums:

Calculate checksum:

```
md5sum example.file > example.file.md5
```

Check file:

```
md5sum -c example.file.md5
```

Commands used during the webinar

sshfs

Mount folder over sshfs

```
sshfs username@taito.csc.fi:/wrk/username ./taito_wrkdir
```

Unmount

```
fusermount -u taito_wrkdir
```

scp

Copy file with the same name

```
scp file1 username@taito.csc.fi:/wrk/username/webinar/
```

Copy file with a new name

```
scp file1 username@taito.csc.fi:/wrk/username/webinar/new.gtf
```

Copy many files

```
scp file1 file2 file3 username@taito.csc.fi:/wrk/username/webinar
```

```
scp *.txt username@taito.csc.fi:/wrk/username/webinar
```

Copy a folder recursively

```
scp -r test username@taito.csc.fi:/wrk/username/webinar
```

rsync

syncing folders

```
rsync -avz -e ssh example_folder username@taito.csc.fi:/wrk/username/webinar
```

common options

-a	Use archive mode: copy files and directories recursively and preserve access permissions and time stamps.
-v	Verbose mode.
-z	Compress.
-e	Specify the remote shell to use.
-u	update
--delete	delete extraneous files from the receiving side (*i.e.* those that do not exist on sending side)
-n	dry run

rsync

large files

```
rsync -P example.file username@taito.csc.fi:/wrk/username/webinar/example.file
```

common options

--partial	Keep partially transferred files.
--progress	Show progress during transfer.
-P	same as --partial --progress

wget

copy a file from an URL

```
wget ftp://example.org/abc.gz
```

Use wildcard

```
wget "ftp://example.org/*.gz"
```

continue interrupted transfer

```
wget --continue ftp://example.org/abc.gz
```

More information

- CSC Computing environment user guide :

  https://research.csc.fi/csc-guide

- Especially Chapter 5:

  https://research.csc.fi/csc-guide-moving-data-between-csc-and-local-environment

CSC – IT Center for Science Ltd

+3589 457 2821 (service requests)
+3589 457 2001 (call center/ contacts)

servicedesk@csc.fi

www.csc.fi

https://www.facebook.com/CSCfi

https://twitter.com/CSCfi

https://www.youtube.com/c/CSCfi

https://www.linkedin.com/company/csc---it-center-for-science