



DATA MOVER

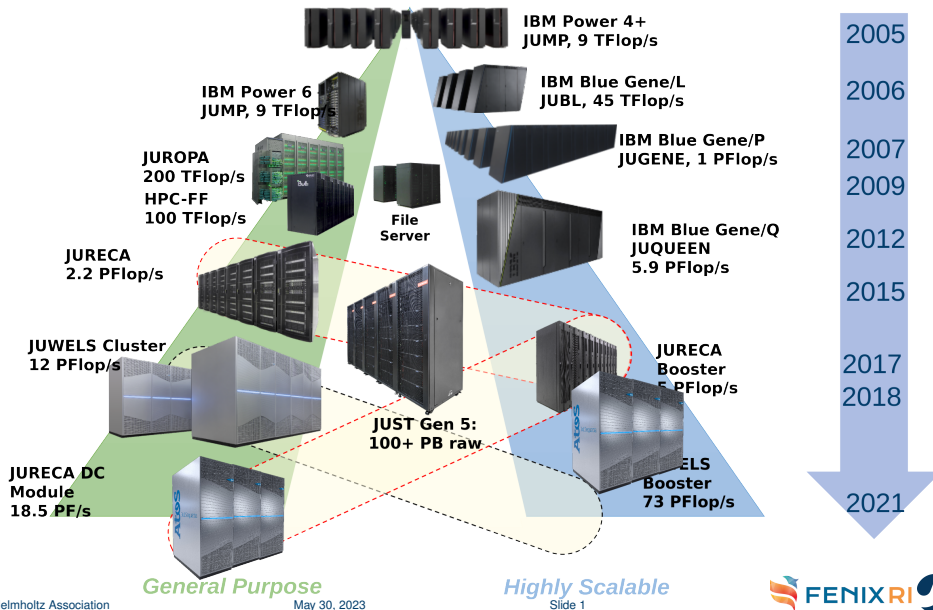
Bridging the gap between HPC and Cloud Storage

May 30, 2023 | Salem el Sayed, Martin Lischewski | Jülich Supercomputing Centre



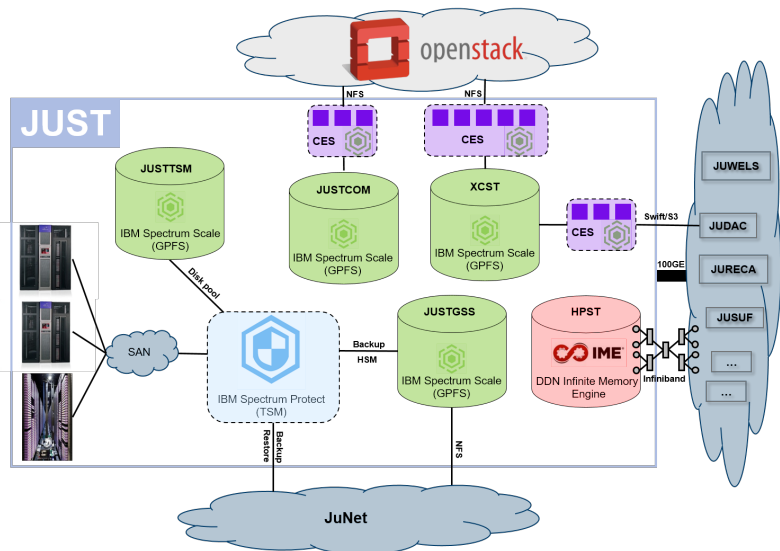
Motivation: Same site data movement

Data Movement history at the JSC



A single "File System"

Data is still in constant flux



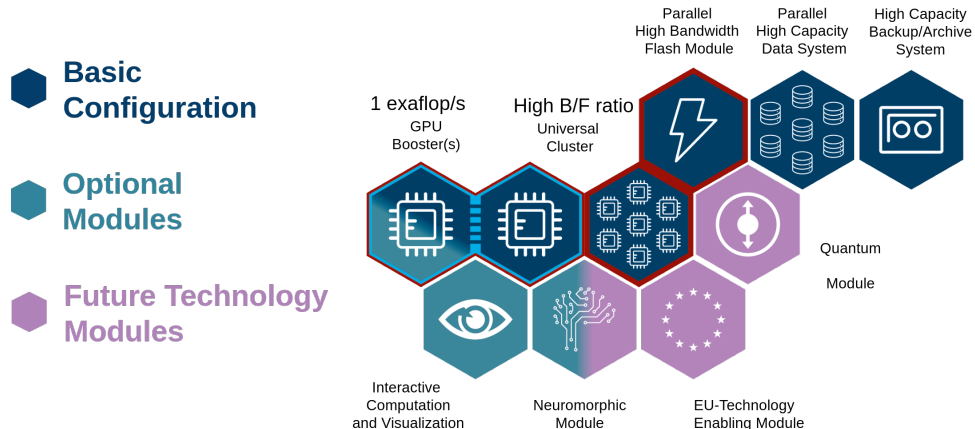
Moving

- between storage tiers (cp, rsync, ...)
- between technologies (IME, ...)
- between APIs (Swift/S3, ...)

JUPITER

European Exascale Supercomputer Coming to Jülich ¹

Back to separate File Systems, where the issue of movement is not yet clarified.



¹ <https://www.fz-juelich.de/en/news/archive/press-release/2022/first-european-exascale-supercomputer-coming-to-julich>

Is *rsync* and similar tools not enough

rsync

A fast, versatile, remote (and local) file-copying tool

Pros:

- Widely available
- Easy to use

Cons:

- No remote to remote (Direct access required to both sides)
- No simple transfer continuation
- No support for different APIs
- No parallel transfer (Maybe when combined with ex. *xargs*)
- Credential hocks done by users (Proxy jumps and similar to get from one system to another)
- ...



Data Mover

Data Mover - Objective

Move data between parallel HPC filesystems and object store

- Attributes
 - Programmable
 - High-Speed
 - Secure
 - Scalable
- One software for all partners
- Single point of authentication

Data Mover - Cooperation

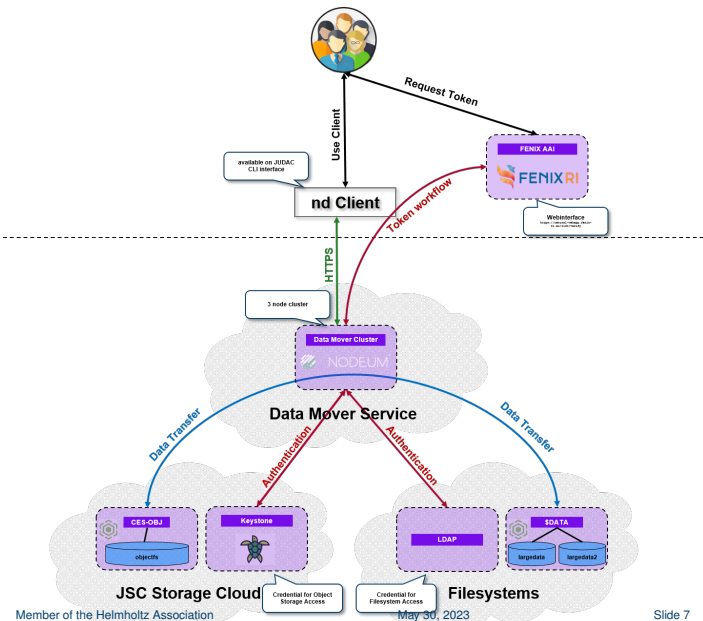
- Project Partners
 - MT-C/AKQUINET
 - <https://www.nodeum.io>
- ICEI Sites
 - BSC (Spain)
 - CEA (France)
 - CINECA (Italy)
 - CSCS (Switzerland)
 - JSC (Germany)

All sites at various stages of deployment.
At JSC service available for user access. ^a

^a <https://apps.fz-juelich.de/jsc/hps/judac/datamover.html>



Data Mover - Workflow



- 1 User initiates copy (ex. via command line)
- 2 User authenticates via browser on the FENIX AAI website (Token valid for 1h)
- 3 Data Mover authenticates transfer on both file systems
- 4 Data Mover initiates data transfer
- 5 Data Mover reports transfer progress and updates token (if needed) until concluded

Data Mover - Not a simple *rsync*

It is an orchestration of many components running on several nodes

ND Client Provides the user with a set of commands to execute the data movements (macOS, Linux and Windows compatible)

Dispatcher Request processing

Mover Discovers data and handles the actual data move as well as metadata updates

Finalizer Update status of data movement

...

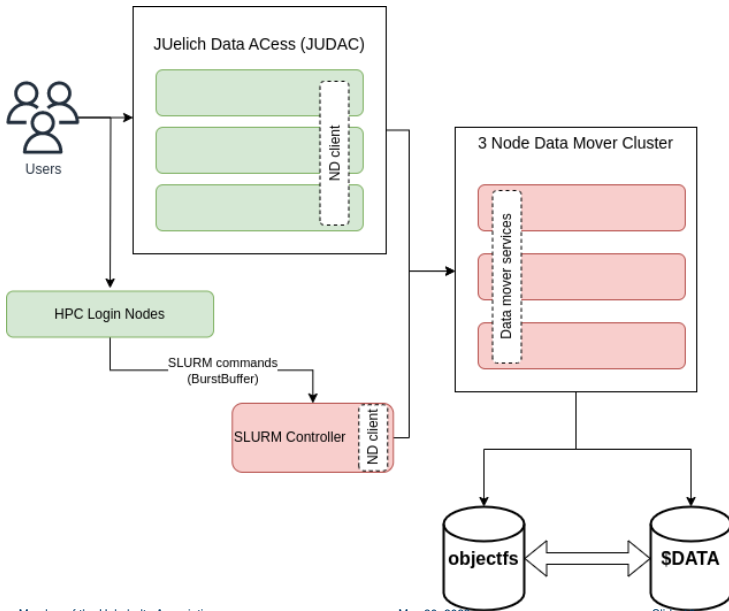
Horizontal Scalability by increasing movers and finalizers.

- More servers running movers
- More movers per servers
- More parallel streams per mover

Data Mover - Implemented features

- FENIX AAI Integration
 - Email as the main attribute
 - No root access of data
 - Automatic token renewal
- Interfaces
 - REST-API
 - Bash Client
 - Hooks for Python
- SLURM integration
 - Automated stage in / stage out
 - Under evaluation (SLURM has reduced support for multiple BB interfaces)

Data Mover - JSC Setup



- Deployed using Ansible
- Allows for scaling by expanding data mover cluster
- Login nodes are not loaded with copy operations
- Various degrees of resiliency level built into the Data Mover components



Data Mover: The user experience

ND Client - Bash

```
$ nd --help
NAME:
  nd - Nodeum CLI

USAGE:
  nd [global options] command [command options] [arguments...]

VERSION:
  2.0.6

COMMANDS:
  admin
  config  configure the Nodeum Client
  copy, cp  create copy task
  move, mv  create move task
  task
  help, h  Shows a list of commands or help for one command
```

ND Client - Status

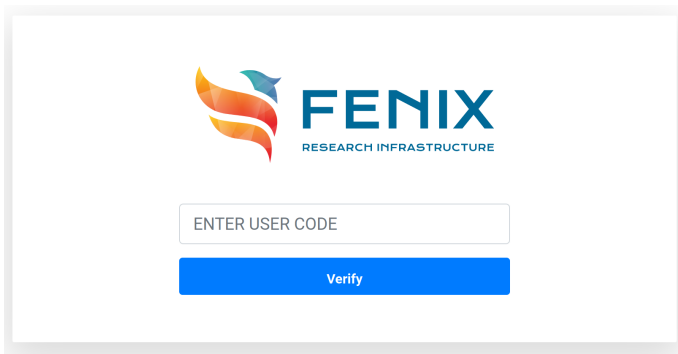
```
$ nd admin status  
INFO Connecting with device flow...  
Open https://central-webapp.fenix-ri.eu/oidc/verify in a browser and type I21-73Y-2NJ
```


ND Client - Status (Authenticate)

```
$ nd admin status  
INFO Connecting with device flow...  
Open https://central-webapp.fenix-ri.eu/oidc/verify in a browser and type I21-73Y-2NJ
```

ND Client - Status (Authenticate)

```
$ nd admin status  
INFO Connecting with device flow...  
Open https://central-webapp.fenix-ri.eu/oidc/verify in a browser and type I21-73Y-2NJ
```



ND Client - Status

```
$ nd admin status
INFO Connecting with device flow...
Open https://central-webapp.fenix-ri.eu/oidc/verify in a browser and type I21-73Y-2NJ
INFO Connected with user Mustermann
| | NAME                | VERSION          | ADDRESS          | UPTIME           | MEMORY           |
...
| | nodeum.dispatcher   | v2.0.3          | 10.2.2.83:50010 | 885h37m12s      | 22.75 MB        |
| | nodeum.dispatcher   | v2.0.3          | 10.2.2.84:50010 | 1029h48m37s    | 22.88 MB        |
| | nodeum.dispatcher   | v2.0.3          | 10.2.2.85:50010 | 1029h35m8s     | 23.71 MB        |
| | nodeum.finalizer    | v2.0.3          | 10.2.2.85:50030 | 1052h2m45s     | 25.62 MB        |
| | nodeum.finalizer    | v2.0.3          | 10.2.2.83:50030 | 1052h2m45s     | 22.85 MB        |
| | nodeum.finalizer    | v2.0.3          | 10.2.2.84:50030 | 1052h2m45s     | 24.63 MB        |
...
| | nodeum.mover        | v2.0.3          | 10.2.2.83:50020 | 885h40m2s      | 31.88 MB        |
| | nodeum.mover        | v2.0.3          | 10.2.2.85:50020 | 885h39m49s    | 23.32 MB        |
| | nodeum.mover        | v2.0.3          | 10.2.2.84:50020 | 885h39m45s    | 23.20 MB        |
...
```

ND Client - Copy

```
$ s3cmd ls s3://my_container  
2021-12-17 15:04 101 s3://my_container/test_data.txt
```

ND Client - Copy

```
$ s3cmd ls s3://my_container
2021-12-17 15:04 101 s3://my_container/test_data.txt
$ nd --config /etc/nd_client_config.json copy --name Transfer_file.test
--md project_name=storagetestdata \
nod://largedata2_pool/storagetestdata/file.test nod-cloud://object_pool/my_container/
INFO Connecting with device flow...
INFO Connected with user Mustermann
Processed size ... done! [1.07GB in 23s]
Processed items ... done! [1 in 23s]
      ID: 647098d21a72d36fa0785dae
      Task ID: 647098d2a91db0b0064b1bf8
      Name: Transfer_file.test
      Comment:
      Created by: Mustermann
      Nodes: 1 / 1
      Size: 1.07 GB / 1.07 GB
      Status: done
```

ND Client - Copy

```
$ s3cmd ls s3://my_container
2021-12-17 15:04 101 s3://my_container/test_data.txt

$ nd --config /etc/nd_client_config.json copy --name Transfer_file.test
--md project_name=storagetestdata \
nod://largedata2_pool/storagetestdata/file.test nod-cloud://object_pool/my_container/
INFO Connecting with device flow...
INFO Connected with user Mustermann
Processed size ... done! [1.07GB in 23s]
Processed items ... done! [1 in 23s]
      ID: 647098d21a72d36fa0785dae
      Task ID: 647098d2a91db0b0064b1bf8
      Name: Transfer_file.test
      Comment:
      Created by: Mustermann
      Nodes: 1 / 1
      Size: 1.07 GB / 1.07 GB
      Status: done

$ s3cmd ls s3://my_container
2023-05-26 11:32 1024M s3://my_container/file.test
2021-12-17 15:04 101 s3://my_container/test_data.txt
```

ND Client - Copy/Move options

```
$ nd copy --help
NAME:
  nd copy - create copy task
...
OPTIONS:
...
CALLBACKS
--callback type:./path/to/file add callback. Format is type:./path/to/file (accepts multiple inputs)
METADATA
--files-md key=value [ --files-md key=value ] set metadata on the files. Format is key=value.
--task-md key=value [ --task-md key=value ] set metadata on the task. Format is key=value.
--trigger-md key=value, --md key=value
  [ --trigger-md key=value, --md key=value ] set metadata on the trigger. Format is key=value.
OPTIONS
--context-gid value, --gid value perform copy in specific GID (default: unset)
--context-uid value, --uid value perform copy in specific UID (default: unset)
--ignore-hidden ignore hidden files and folders, starting with <dot> (default: false)
--overwrite overwrite existing entries (default: false)
--parallel value maximum number of process for mover (default: 1)
--priority value task priority [0..9] (default: 0)
--recursive, -R copy directories recursively (default: false)
--working-dir value, --wd value set working directory
```

ND Client - Task list

```
$ nd task list
INFO Connecting with device flow...
INFO Connected with user Mustermann
| TASK ID | TASK NAME | COMMENT | CREATED BY | LAST EXECUTION STATUS |
| 647098d2a91db0b0064b1bf8 | Transfer_file.test | | Mustermann | done |
| NUMBER OF TASK(S) | 1 | | | |
```

```
$ nd task --help
NAME:
  nd task
USAGE:
  nd task command [command options] [arguments...]
COMMANDS:
  list      list all tasks
  list-exec list tasks executions
  run       run a task
  status    get status of a task
  help, h   Shows a list of commands or help for one command
OPTIONS:
  --help, -h show help (default: false)
```


SLURM - An SBATCH example

Under evaluation,

```
#!/bin/bash
#SBATCH -N1
#SBATCH -A CompProj -p batch -t 30
#BB_LUA dm_common project_name=CompProj project_domain_name=xxx user_domain_name=xxx
#BB_LUA dm_common overwrite=true
#BB_LUA dm_common tokenx="\xxxx"
#BB_LUA dm_stage_in ops=copy src=nod-cloud://object_pool/container/ dst=nod://posix_pool/data/dataset
#BB_LUA dm_stage_out ops=copy src=nod://posix_pool/data/result dst=nod-cloud://object_pool/container/
srun my_commands
sleep 300
```



Thank You ! Questions?



Fenix has received funding from the European Union's Horizon 2020 research and innovation programme through the ICEI project under the grant agreement No. 800858.