bw|HPC

KIT | NHR
National High-Performance Computing Center

File Systems Introduction

Peter Weisbrod
KIT, SCC

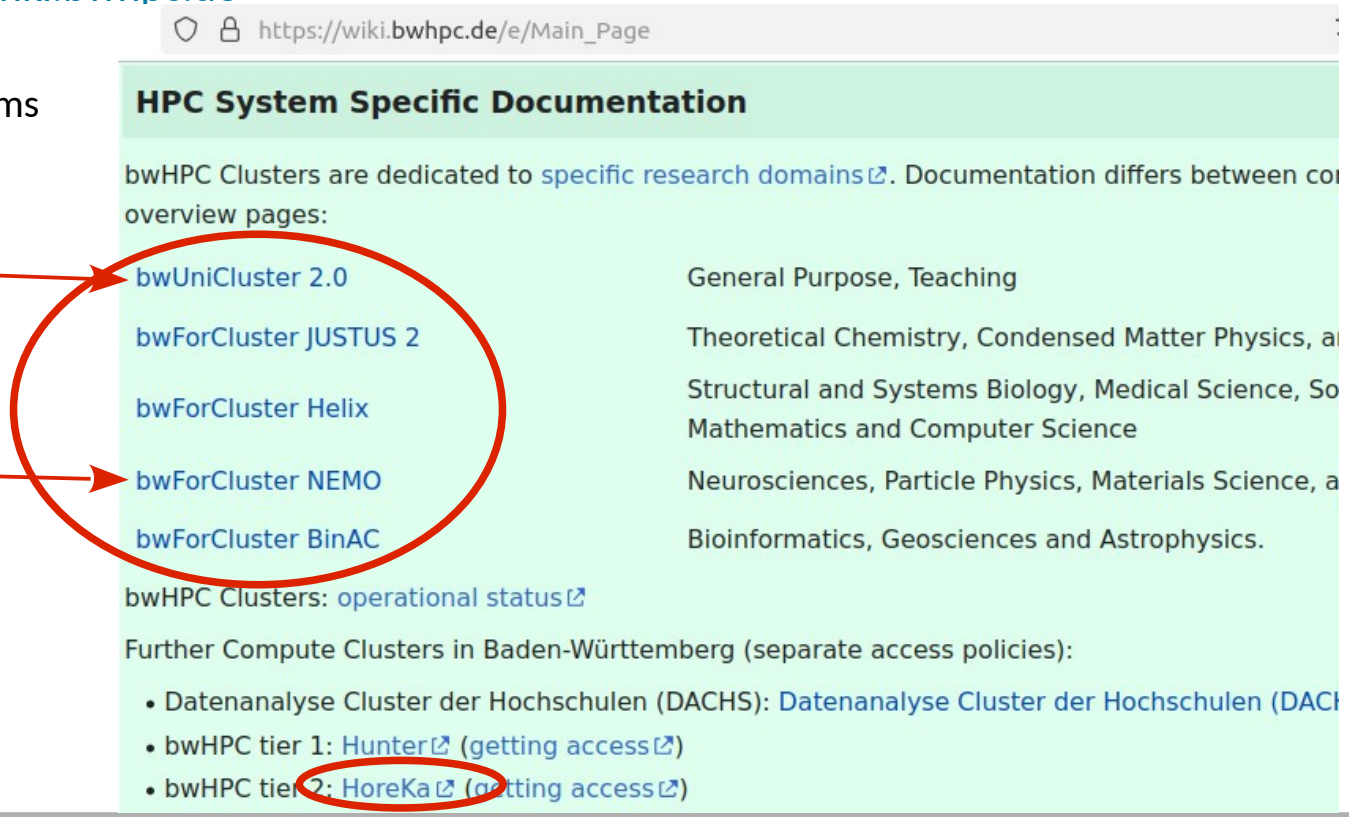
Funding: Baden-Württemberg MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST | Bundesministerium für Bildung und Forschung

Reference: bwHPC Wiki

- Most information given by this talk can be found at https://wiki.bwhpc.de
  - select cluster
  - then select File Systems

Replaced by bwUniCluster 3.0

Replaced by NEMO 2



Material: Slides & Scripts

- https://indico.scc.kit.edu/event/4937/
- BwUniCluster 2.0/3.0: /opt/bwhpc/common/workshops/2025-04-09/
- HoreKa: /software/all/workshop/2025-04-09/

How to read the following slides

Abbreviation/Colour code	Full meaning
`$ command -option value`	$ = prompt of the interactive shell The full prompt may look like: `user@machine:path$` The command has been entered in the interactive shell session
`<integer>` `<string>`	<> = Placeholder for integer, string etc
`foo, bar`	Metasyntactic variables

How to use each File System (1)

- $HOME = Home directory
  - → Software, configuration files, final results
  - → Omit heavy I/O
- Workspaces = Working directories with lifetime
  - → Intermediate results, huge input/output data sets
  - → Scratch data which needs to be shared between nodes
  - → Omit small files, tiny block sizes, lots of metadata operations

    If not possible to omit, KIT and HoreKa users can use Workspaces on flash storage
- $TMPDIR = Separate file system on each node using local disks
  - → Data is only available during job runtime on the local node
  - → Possibly transfer data here within a batch job
  - → All sorts of I/O allowed

How to use each File System (2)

- BeeOND = Private file system for batch job
  - → Data is only available during job runtime on the batch job nodes
  - → Possibly transfer data here within a batch job
  - → All sorts of I/O allowed, only available on bwUniCluster 3.0 and HoreKa
- External storage
  - → Archive scientific data, move data here when data sets become too large
  - → Each organization has different solutions, examples are RDA or LSDF at KIT
  - → Use huge files or compressed archives
- Summary
  - → Use $HOME for permanent data
  - → Use workspaces for huge files and sequential I/O
  - → Use $TMPDIR or BeeOND with many (> 10000) small files or random I/O

$HOME = Home directory

- $HOME is visible on all nodes of a cluster
- Properties of $HOME on different clusters

Cluster	Quota capacity limit	Quota file limit	Backup
JUSTUS 2	400 GB per user	2 mill. per user	Yes
Helix	200 GB per user	unlimited	No
NEMO 2	100 GB per user	unlimited	No
BinAC	40 GB per user	unlimited	Yes
BwUniCluster 3.0	500 GB per user also limit per organization	5 mill. per user	Yes
HoreKa	10 TB per project	10 mill. per project	Yes

$HOME on bwUniCluster 3.0

- HowTo goto:

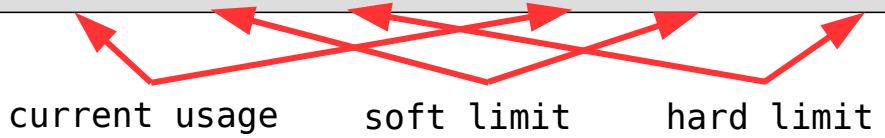
```
$ cd $HOME
```

```
$ cd
```

- User's quota usage and limits:

```
$ lfs quota -uh $(whoami) $HOME
```

```
Disk quotas for usr ka_ab1234 (uid 9999):
                Filesystem    used   quota   limit   grace   files   quota    limit   grace
/home/kit/ka_scc/ka_ab1234  22.59G    500G    550G       -   81408  5000000 5500000       -
```



  - Usage rate above soft limit for grace period (7 days) will only report a warning.

- Organization quota usage and limits:

```
$ lfs quota -ph $(grep $(echo $HOME | sed -e "s|/[^/]*/[^/]*$||") \
    /pfs/data6/project_ids.txt | cut -f 1 -d\ ) $HOME
```

$HOME / $PROJECT on HoreKa

- $HOME and $PROJECT are identical if your account is member of one project
  - Otherwise to change to another project which will also modify $PROJECT:

```
$ newgrp <another_project_group>
```

- Project group quota usage and limits:
  - First start an interactive job (sometimes this step is not needed):

```
$ salloc -p dev_cpuonly -n 1 -t 20 --mem=500
```

  - Show usage and limits of your project group on the $HOME file system:

```
$ /usr/lpp/mmfs/bin/mmlsquota -j $PROJECT_GROUP --block-size G -C hkn.scc.kit.edu hkfs-home
```

```
                     Block Limits        |    File Limits
Filesystem type       GB  quota   limit   |   files    quota    limit
home       FILESET  3467  10240   11264   | 4747501 10485760 11534336
```

current usage soft limit hard limit

Exercise 1: Show quotas

- Login to bwUnicluster 2.0 or HoreKa and show list of commands for exercises:
  - BwUniCluster:
    ```
    $ cat /opt/bwhpc/common/workshops/2025-04-09/pfs_commands.txt
    ```
  - HoreKa:
    ```
    $ cat /software/all/workshop/2025-04-09/pfs_commands.txt
    ```
- Use Cut & Paste to execute the first commands which show your quotas

Workspaces = Working directories with lifetime

- **Workspace**: lifetime on allocated folder
  - Available on all clusters, visible on all nodes of a cluster
  - HowTo:
    → https://wiki.bwhpc.de/e/Workspace

`$ ws_allocate foo 10`	Allocate workspace *foo* for 10 days
`$ ws_list`	List your workspaces
`$ ws_find foo`	Get absolute path of workspace *foo*
`$ ws_extend foo 5`	Extend lifetime of your workspace *foo* by 5 days from now. Number of extensions depends on cluster.
`$ ws_release foo`	Manually erase your workspace *foo*
`$ ws_… -F ffuc …`	Select non default workspace file system with -F (works for any command)

Properties of Workspaces on different clusters

Cluster	Capacity limit	File limit	Max lifetime	Max extensions
JUSTUS 2	20 TB per user	5 mill. per user	90 days	unlimited
Helix	10 TB per user	unlimited	30 days	10 times
NEMO 2	5 TB per user	1 mill. per user	100 days	99 times
BinAC	unlimited	unlimited	30 days	3 times
BwUniCluster 3.0	40 TB per user	20 mill. per user	60 days	3 times
HoreKa	250 TB per user	50 mill. per user	60 days	3 times

Exercise 2: Create workspace

- Allocate two workspaces

```
$ ws_allocate ws01 30
Info: creating workspace.
/pfs/work9/workspace/scratch/myuser-ws01
remaining extensions  : 3
remaining time in days: 30

$ ws_allocate -F ffuc ws_ssd 50
Info: creating workspace.
/pfs/work8/workspace/ffuc/scratch/myuser-ws_ssd
remaining extensions  : 3
remaining time in days: 50
```

Exercise 3: List workspace

- List workspaces

```
$ ws_list
id: ws_ssd
     workspace directory  : /pfs/work8/workspace/ffuc/scratch/myuser-ws_ssd
     remaining time       : 49 days 23 hours
     creation time        : Wed Oct  6 18:59:11 2021
     expiration date      : Thu Nov 25 17:59:11 2021
     filesystem name      : pfs6wor8
     available extensions : 3
id: ws01
     workspace directory  : /pfs/work9/workspace/scratch/myuser-ws01
     remaining time       : 29 days 23 hours
     creation time        : Wed Oct  6 18:55:17 2021
     expiration date      : Fri Nov  5 17:55:17 2021
     filesystem name      : pfs7wor9
     available extensions : 3
```

Exercise 4: Find workspace path

- Find workspace path and switch to it

```
$ ws_find ws01
/pfs/work9/workspace/scratch/myuser-ws01

$ ws_find -F ffuc ws_ssd
/pfs/work8/workspace/ffuc/scratch/myuser-ws_ssd

$ cd $(ws_find ws01)
$ pwd
/pfs/work9/workspace/scratch/myuser-ws01
```

Exercise 5: Release workspaces

- Release workspaces

```
$ ws_release ws01

$ ws_release -F ffuc ws_ssd

$ ws_list
```