Karlsruher Institut für Technologie

Assistance in Lustre administration

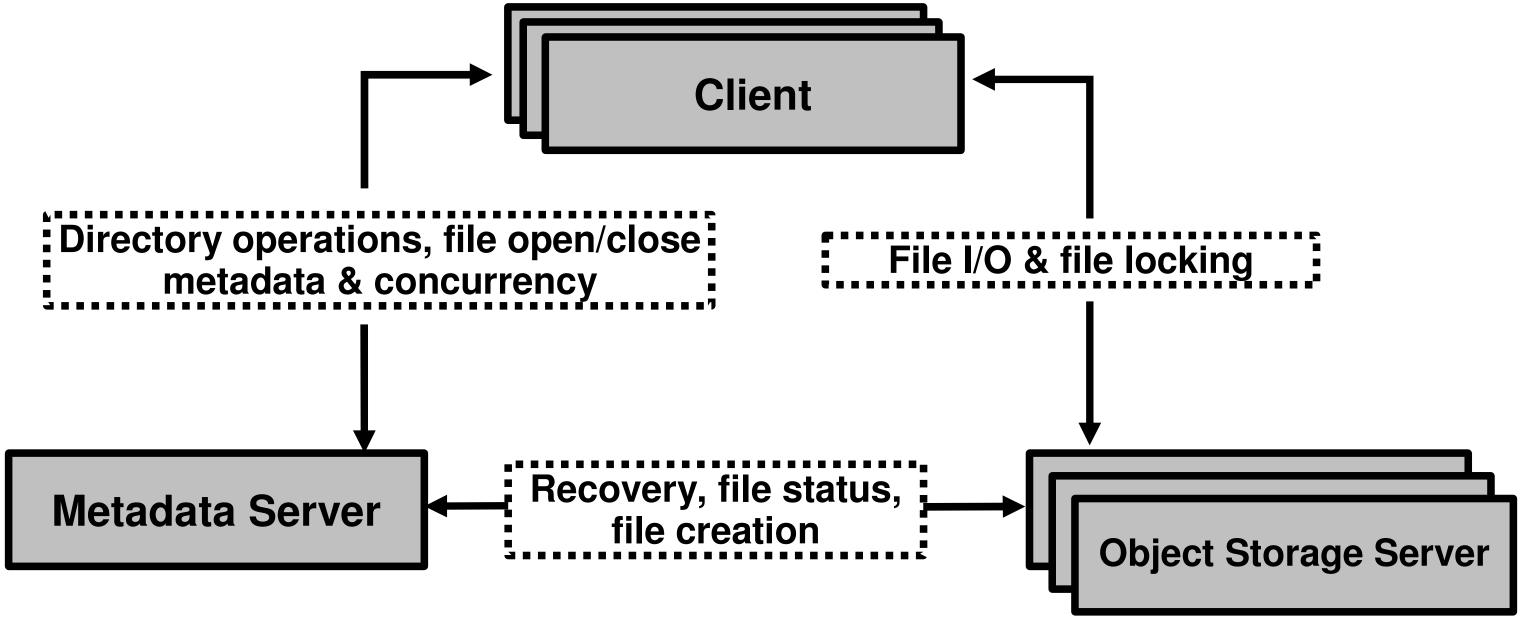
Roland Laifer

STEINBUCH CENTRE FOR COMPUTING - SCC

KIT – University of the State of Baden-Württemberg and
National Laboratory of the Helmholtz Association

www.kit.edu

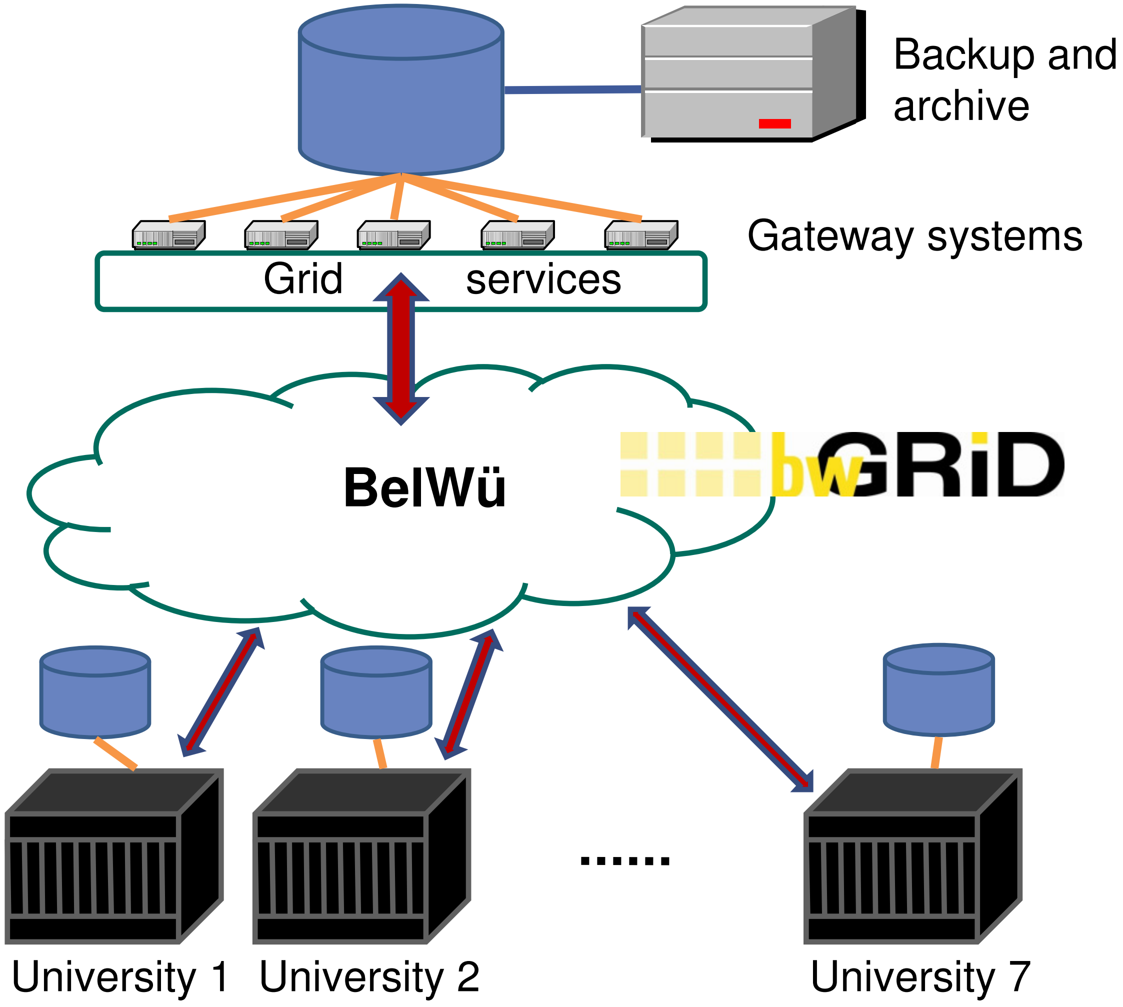
Basic Lustre concepts


Client
Directory operations, file open/close metadata & concurrency
File I/O & file locking
Metadata Server
Recovery, file status, file creation
Object Storage Server

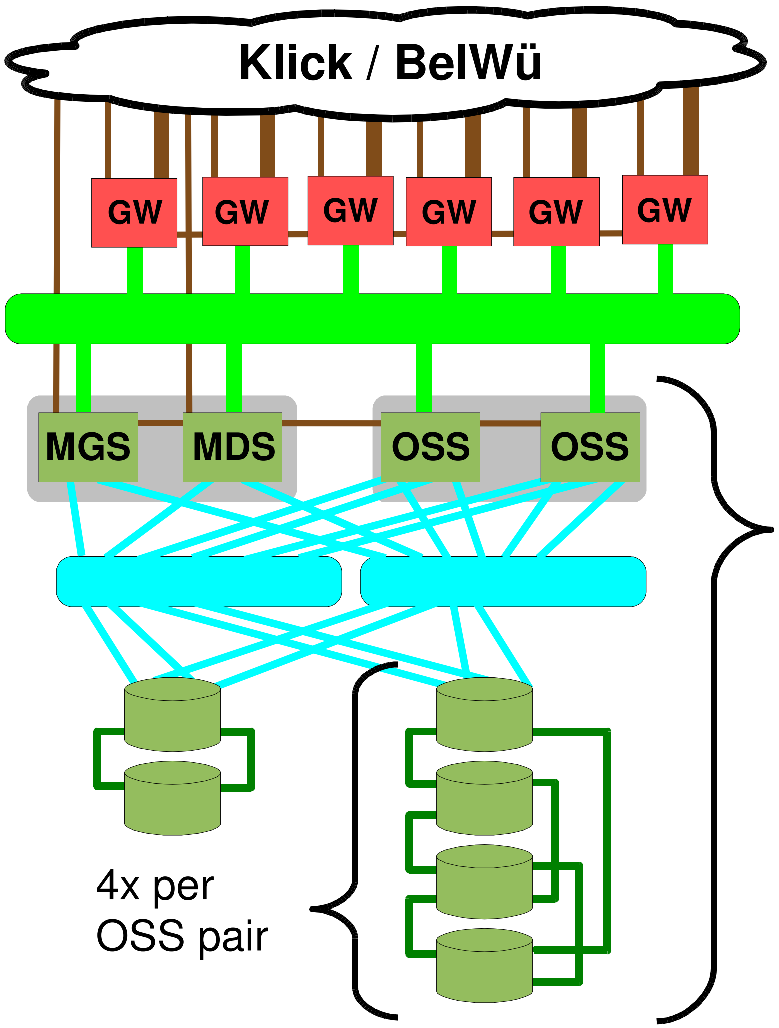
- Lustre componets:
  - Clients (C) offer standard file system API
  - Metadata servers (MDS) hold metadata, e.g. directory data
  - Object Storage Servers (OSS) hold file contents and store them on Object Storage Targets (OSTs)
  - All communicate efficiently over interconnects, e.g. with RDMA

bwGRiD storage system (bwfs) concept

- Grid middleware for user access and data exchange



bwGRiD storage system building blocks



- Gbit Ethernet
- 10 Gbit Ethernet
- 4X DDR InfiniBand
- 4 Gbit Fibre Channel
- 3 Gbit SAS

Central file system backup

Central file system work:
Same structure with 3 OSS pairs

Local file systems:
Same structure with 4x2 instead of
4x4 enclosures per OSS pair

bwGRiD storage system hardware in detail

File system:	128 TB central (backup)	256 TB central (work)	32 TB local (each site)	64 TB local (each site)
Location:	KIT CS	KIT CS	KIT CN, Freiburg, Mannh., Heidelb.	Stuttgart, Ulm, Tübingen
Metadata Server (MDS):				
# of servers	2	2	2	2
# of MSA2212fc	1	1	1	1
# of MSA2000 disk encl.	1	1	1	1
# of disks	24 * 146 GB SAS	24 * 146 GB SAS	24 * 146 GB SAS	24 * 146 GB SAS
Object Storage Server :				
# of servers	2	6	2	2
# of MSA2212fc	4	12	4	4
# of MSA2000 disk encl.	12	36	4	4
# of disks	192 * 1 TB SATA	576 * 1 TB SATA	96 * 750 GB SATA	96 * 1 TB SATA
Capacity	128 TB	256 TB	32 TB	64 TB
Throughput	1500 MB/s	3500 MB/s	1500 MB/s	1500 MB/s

Lustre administration challenges (1)

- Very complicated hardware
  - For example see building blocks slide
  - For each component driver, firmware or subcomponent may fail
    - Subcomponents are cables, adapters, caches, memory, …
  - With extreme performance new hardware bugs show up
    - Some are related to timing issues
- Complex Lustre software
  - Roughly 250,000 lines of code
    - Lustre error messages are still hard to understand
  - Focus on very high performance
    - e.g. requires low level Linux kernel interfaces
  - Distributed system at large scale
    - Not easy to find out which part is causing problems

Lustre administration challenges (2)

- Inefficient user applications can cause trouble
  - e.g. reading and writing to same file area from many nodes
  - Not easy to find out which user is causing the trouble
- Importance of the file system
  - Complete clusters are not usable if the file system hangs
  - Corrupted or deleted user data is very annoying
- This talk tries to help with most challenges

Best practices for MSA2000 storage systems

- Also see new documentation from HP
  - HP Scalable File Share G3 MSA2000fc How To
- Firmware upgrades and broken controller replacement
  - Either unmount clients and stop all servers
    - Requires full maintenance and waiting for jobs to complete
  - Or shutdown the affected server pairs which causes I/O to hang
    - Risk of some application I/O errors and follow-on problems
    - Might cause job aborts due to the batch system wall clock limit
  - Disable automatic partner firmware upgrade
    - i.e. upgrade each controller separately
- Multiple broken disks per enclosure
  - Up to 2 disks can be exchanged at the same time
  - Contact support with 3 or more broken disks per enclosure
- Enable email alerts

Check system status

- Lustre status
  - Check for LustreError messages in logs of servers and clients
    - pdsh -a grep LustreError: /var/log/messages
    - Without LustreError messages Lustre usually works fine
    - LustreError messages require further investigation, see next slides
  - Check if connections on all clients show status FULL
    - pdsh -a 'cat /proc/fs/lustre/*/*/*_server_uuid' | dshbak –c
- Overall system health
  - Use our script z20-hpsfs-check-health
    - Requires clean system status to create proper reference files
    - Understanding the output needs some experience
- Performance checks
  - e.g. with dd on each OST lun, see our upgrade documentation
  - Should be done before and after each maintenance

Understanding Lustre messages (1)

- *LBUG* means Lustre bug and indicates a software bug
  - Should be reported, could be searched on bugzilla.lustre.org
- String *evict* means aborted communication after timeout
  - Message on server shows client IP address and timestamp
    - Use batch system history to identify user job(s)
  - One possible reason is hardware failure, e.g. of InfiniBand adapter
  - Other possible reason are inefficient applications
    - timeout due to lots of conflicting requests from many clients
    - e.g. caused by many tasks writing to the same file area
- *Remounting filesystem read-only* could indicate fatal failure
  - Usually due to a storage subsystem problem
  - Also check for *SCSI error* messages

Understanding Lustre messages (2)

- Displayed error codes are standard Linux codes
  - Show explanation of their meaning
    - grep -hw <error number> /usr/include/*asm*/errno*.h
    - e.g. return code -122 means *Quota exceeded*
- Identify clients by UUID shown in logs
  - Example of such a log entry
    - Feb 21 14:35:08 pfs1n10 … LustreError: … timeout on bulk PUT … o3->**7fc2aa1f-6d70-21c3-4df3-fee6cf6676d1** …
  - Find out client IP address on corresponding server
    - pfs1n10# /usr/sbin/lctl get_param "*.*.exports.*.uuid" | \
      grep **7fc2aa1f-6d70-21c3-4df3-fee6cf6676d1**

Monitoring user activity

- Use collectl to monitor I/O usage
  - Installed on servers, installation on clients required
    - http://collectl.sourceforge.net/
  - Show read/write rates on clients or servers
    - `pdsh -a collectl -s l -i1 -c5`
  - Show metadata rates on client
    - `collectl -sl --lustopts M`
- Use Lustre statistics to find clients with high I/O usage
  - Useful script to execute on each server (Lustre >= 1.8.2 only!)
    - https://bugzilla.lustre.org/attachment.cgi?id=29248
- No easy way to check which user on client is doing I/O
  - Sorting open files by update time might give hints

    -
      ```
      ls -lrht `lsof | grep /bwfs/ | perl -e'while(<STDIN>) \
      {if ($_ =~ m|.*(/bwfs/\S*)| ) {if (-f $1) {$all .= " $1";}}}; print "$all\n";'`
      ```

Investigate reason for hanging file system (1)

- Check if client shows all services (OSTs and MDT)
  - lfs df
- Check if logs report frequent problems for a service
  - pdsh -a grep LustreError: /var/log/messages
- Check if servers are unhealthy or recovering and uptime
  - pdsh -a cat /proc/fs/lustre/health_check
  - pdsh -a 'lctl get_param "*.*.recovery_status"' | grep status:
  - pdsh -a uptime
- Check if all Lustre services are still mounted normally
  - pdsh -a 'mount | grep lustre | wc -l' | dshbak –c
  - pdsh -a grep read-only /var/log/messages

Investigate reason for hanging file system (2)

- Identify server and mount point for affected service
  - pdsh -a 'lctl get_param *.*.mntdev‘
  - Also compare /opt/hp/sfs/scripts/*filesystem*.csv
- Find out affected storage device
  - VOLID=`pdsh -w *server* grep *mpathnum* /var/lib/multipath/bindings \
    | perl -ne 'if (/(\w{16})\s*$/) {print "$1\n"}'`
  - *forallmsas* show volumes ; done | grep -B1 $VOLID
- Show critical and warning events on affected MSA2000
  - msa2000cmd.pl *msa* show events last 30 error
- Check LEDs on the storage system
- Do not forget to contact support at an early stage
- Carefully plan each step to repair the problem

Further information

- Detailed step by step information for upgrades
  - bwrepo.bfg.uni-freiburg.de below ~unikarlsruhe/repo/G3.2_upgrade
- More talks like this
  - http://www.scc.kit.edu/produkte/lustre
    - e.g. *Using file systems at HC3* for hints to use Lustre efficiently
- HP SFS documentation
  - http://www.hp.com/go/sfs-docs
- Lustre in general
  - http://www.lustre.org/
- Lustre future
  - http://www.opensfs.org/
  - http://www.whamcloud.com/