

IBM Spectrum Scale Field Update

—
Achim Rehor
Spectrum Scale &
ESS Support EMEA



Agenda

- ***What do we do in support?***
- ESS upgrade
- OPC news
- Ganesha issues
- Surveys
- Survey: What can we do to improve in support?



Support Structure



Support for ESS and Spectrum Scale

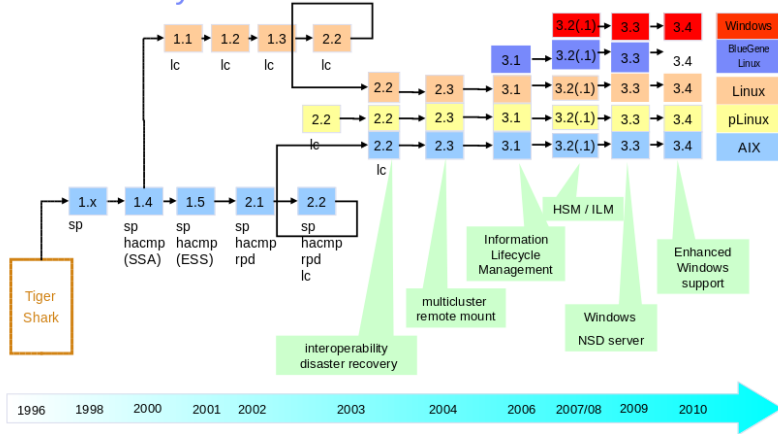
- FTS Severity 1 Cases
 - AP Beijing/Shanghai/Pune/Bangalore/Auckland
2 Teams : ESS and Scale
 - Americas Raleigh/Poughkeepsie/Toronto
2 Teams : ESS and Scale
 - EMEA Kelsterbach/Chemnitz.../Sofia
1 Team for ESS and Scale
- Severity 2-4 per Region

[IBM Spectrum Scale Support Reference Guide](#)

[IBM Elastic Storage Server \(ESS\) Support Reference Guide](#)

... a little bit of history

GPFS history and milestones



9076 SP – Scalable POWERparallel

Upto 512 AIX Nodes

HighSpeed Network (SP Switch)

GPFS Version 1.2 and up

... more recent

ESS 3000 Architecture

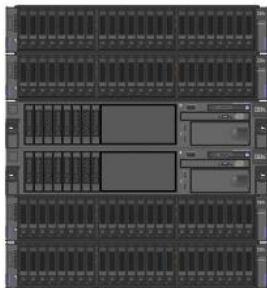


Front View
24 NVMe Drive Slots



Rear View
Two Canisters / Servers
Two Power Supplies

ESS GS4S



ESS GL6s



Model GL8C
8 Enclosures, 36U
846 NL-SAS, 2 SSD



11.8 PB raw

top500 back then and today



TOP500 List - November 1999

R_{max} and **R_{peak}** values are in GFlops. For more details about other fields, check the TOP500 description.

R_{peak} values are calculated using the advertised clock rate of the CPU. For the efficiency of the systems you should take into account the Turbo CPU clock rate where it applies.

previous 1 2 3 4 5 next

Rank	Site	System	Cores	R _{max} (GFlop/s)	R _{peak} (GFlop/s)	Power (kW)
1	Sandia National Laboratories United States	ASCI Red Intel	9,632	2,379.0	3,207.0	
2	Lawrence Livermore National Laboratory United States	ASCI Blue-Pacific SST, IBM SP 604e IBM	5,808	2,144.0	3,856.5	

TOP500 List - November 2019

R_{max} and **R_{peak}** values are in TFlops. For more details about other fields, check the TOP500 description.

R_{peak} values are calculated using the advertised clock rate of the CPU. For the efficiency of the systems you should take into account the Turbo CPU clock rate where it applies.

previous 1 2 3 4 5 next

Rank	Site	System	Cores	R _{max} (TFlop/s)	R _{peak} (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National Laboratory United States	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM	2,414,592	148,600.0	200,794.9	10,096
2	DOE/NNSA/LLNL United States	Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM / NVIDIA / Mellanox	1,572,480	94,640.0	125,712.0	7,438

Agenda

- What do we do in support?
- ***ESS upgrade***
- OPC news
- Ganesha issues
- Surveys
- Survey: What can we do to improve in support?



ESS Upgrade (and some pitfalls)

Current ESS5351(GA 23.2.2020)

- Quick Deployment Guide Overview

- EMS upgrade

- Prereq checking/software download for BE systems, check and potentially upgrade HMC
- Healthcheck
- Check and correct xcat settings (gssdeploy.cfg)
- Upgrade ems
Note: think about quorum first possibly move managers away first

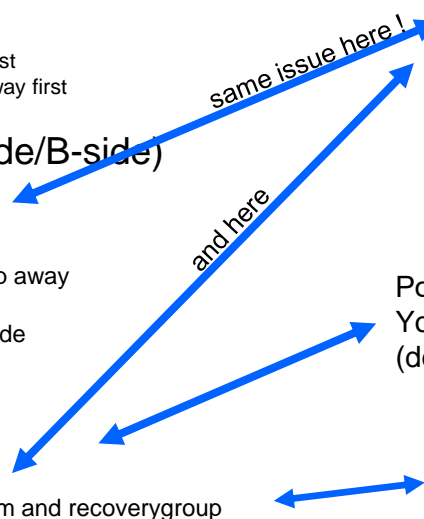
- IO Nodes upgrade (A-side/B-side)

- Move manager roles away
- Move locally owned recoverygroup away
- Umount, mmshtudown and upgrade
- Upgrade ofed and ipraid
- hostadapter-firmware update
- mmstartup and takeback of quorum and recoverygroup

Quorum change: Always check # mmgetstate -aLs

Node number	Node name	Quorum	Nodes up	Total nodes	GPFS state	Remarks
1	gssio1_9	2	3	3	active	quorum node
2	gssio2_9	2	3	3	active	quorum node
3	ems1	2	3	3	active	quorum node

Summary information	
Number of nodes defined in the cluster:	3
Number of local nodes active in the cluster:	3
Number of remote nodes joined in this cluster:	0
Number of quorum nodes defined in the cluster:	3
Number of quorum nodes active in the cluster:	3
Quorum = 2, Quorum achieved	



must match!

Potential pitfall: when coming from 'older' recoverygroup
You may potentially hit an assert, when mmfsd startup (defect 1083328)

Potential pitfall: there is a chance of running into a deadlock during node leave, if the clustermgr node is below 5.0.3.0 or 4.2.3.14 (defect 1074954)

more details on potential pitfalls :

- RG compatibility between old RGs (created before 4.2.2) and fresh ones > 5.0.2 (ESS 5.3.3 and up)
assert on the just updated node(s) due to “unhandled RGMaster state transition” (defect 1083328)
→ circumvention: update all nodes and run mmchconfig release=LATEST
fixed in ESS535(1)
- potential deadlock due to aclMsgFailureUpdate (defect 1074954)
→ mmshutdown of ‘pending clients’
fixed in 4.2.3.14 and higher, or 5.0.3.0 or higher
- kvm_cma reserved memory : [ESS Power LE node hang due to oom while free memory > vm.min free kbytes](#)
5% of TotalMem, reserved from 1st numa node

```
[root@gssio1 ~]# numactl -H
available: 4 nodes (0-1,16-17)
node 0 cpus: 0 1 8 9 16 17 24 25 32 33
node 0 size: 65536 MB
node 0 free: 30773 MB
node 1 cpus: 40 41 48 49 56 57 64 65 72 73
node 1 size: 65536 MB
node 1 free: 41780 MB
node 16 cpus: 80 81 88 89 96 97 104 105 112 113
node 16 size: 65536 MB
node 16 free: 27026 MB
node 17 cpus: 120 121 128 129 136 137 144 145 152 153
node 17 size: 65536 MB
node 17 free: 29825 MB
node distances:
```

```
[root@gssio1 ~]# cat /proc/vmstat | grep cma
nr_free_cma 0
```

ESS Upgrade (cont.)



more potential pitfalls :

- verbsPorts definition containing fabric portion is not preserved during upgrade

➔ use '-v verbsPort' flag to 'redefine'

12. Update the node configuration.

- **Online upgrade only**

```
/opt/ibm/gss/tools/samples/gssupgrade.sh -s CurrentIoServer-hs
```

This command is run from the EMS node.

- After the update a couple of nsdRaid parameters are complained about in mmfs.log.latest

```
[E] Unknown config parameter "nsdRAIDFlusherBuffersLowWatermarkPct" in /var/mmfs/gen/mmfs.cfg, line 22.  
[E] Unknown config parameter "nsdRAIDFlusherBuffersLimitPct" in /var/mmfs/gen/mmfs.cfg, line 23.  
[E] Unknown config parameter "nsdRAIDFlusherTracksLowWatermarkPct" in /var/mmfs/gen/mmfs.cfg, line 24.  
[E] Unknown config parameter "nsdRAIDFlusherFWLogLimitMB" in /var/mmfs/gen/mmfs.cfg, line 27.  
[E] Unknown config parameter "nsdRAIDFlusherThreadsLowWatermark" in /var/mmfs/gen/mmfs.cfg, line 28.  
[E] Unknown config parameter "nsdRAIDFlusherThreadsHighWatermark" in /var/mmfs/gen/mmfs.cfg, line 29.
```

➔ save current mmfs.cfg file

➔ mmchconfig <params>=DEFAULT

Note: unknown parameter, use the <999><Enter> exit

- Spectrum Scale Flashes

[Flashes, alerts and bulletins for IBM Spectrum Scale](#)

[IBM Spectrum Scale \(GPFS\) 5.0.4 levels: possible metadata or data corruption during file system log recovery](#) (fixed in ESS5351)

Finalize upgrade by updating enclosure and drive firmware

Note: depending on the number of drives this can take several hours (drive-by-drive update online)

Agenda

- What do we do in support?
- ESS upgrade
- ***OPC news***
- Ganesha issues
- Surveys
- Survey: What can we do to improve in support?



What is that?

https://de.wikipedia.org/wiki/Orthogonal_Defect_Classification

How do we do it?

PMR Classification

Mandatory field for Insert (Optional for query)

*PMR/Ticket Number:

*Initial Severity:

*Platform:

*Release:

*Product: [help](#)

*Scrum and Component(component optional for query): [help](#)

*Situation/Symptom: [help](#)

*Cause: [help](#)

What do we learn ...?

OPC News (cont.)

Based on that Classification sheet per ticket, we figure

- How severities are distributed over all tickets
- Which part of the product raises how many cases and which type
- Are these code defects, limitations, environment reasons or maybe caused by unclear or lacking documentation

What do we learn...?

- Continuous RAS improvements
- Focus on 'critical areas', higher volume
 - ganessa for example
- Proactive Services
 - SW CallHome findings

(Software) Call Home data evaluation

- ESS vs. Scale
- Which HW architecture is being used
- What versions are being used in the field
- Which features (protocols) are enabled/used

Basically, how do customers use our product(s)

2019 Product Quality Highlights

- Completed phase 1 of the Ganesha investment areas
- Established a Protocol Review Board to actively screen protocol deployment and support requests
- Developed a comprehensive memory leak test plan to use for each release

RAS Focus Items 2020

Proactive Services / Call Home

- Healthchecker
- Work with support team to establish process to proactively contact customers
- Add rules for upcoming best practice violations and upcoming Flashes
- Make Call Home more attractive to further increase Call Home adoption
- Start Call Home survey in the User Group mailing list

eFix test improvements

ESS simplified deployment (with ESS3000)

Healthchecker: rule-set to detect

- misconfiguration
- known issues (flashes)
- best practices not followed

Support staff identifies issues and adds ideas for rules

Call Home Data is checked against existing rules catalog

<p>IBM Spectrum Scale (GPFS) 5.0.4 levels: possible metadata or data corruption during file system log recovery</p> <p>IBM has identified a problem with the IBM Spectrum Scale parallel log recovery function at V5.0.4.0 - V5.0.4.1, which may result in metadata corruption or undetected data corruption during the course of a file system recovery.</p>	Core	Check for Spectrum Scale version $\geq 5.0.4.0$ and $\leq 5.0.4.1$ https://www.ibm.com/support/pages/node/1274428	No	Critical
---	------	---	----	----------

Agenda

- What do we do in support?
- ESS upgrade
- OPC news
- ***Ganesha issues***
- Surveys
- Survey: What can we do to improve in support?



Ganesha issues

GPFS release and delivered ganesha levels

- 4.2.3.x = nfs-ganesha-2.3.2-0(ibm38-ibm69)
- 5.0.1.x = nfs-ganesha-2.5.3(ibm020-ibm022)
- 5.0.2.x = nfs-ganesha-2.5.3(ibm026-ibm030)
- 5.0.3.x = nfs-ganesha-2.5.3(ibm036-ibm036.11)
plus efixes
- 5.0.4.0 = nfs-ganesha-2.7.5(ibm053)
- 5.0.4.1 = nfs-ganesha-2.7.5(ibm053.02)
- 5.0.4.2 = nfs-ganesha-2.7.5(ibm054.03)
- 5.0.4.3 = nfs-ganesha-2.7.5(ibm054.05)

Ganesha issues (cont.)



GPFS release and delivered ganesha levels

- 5.0.3.x = nfs-ganesha-2.5.3(ibm036-ibm036.11)
plus efixes

Major issues being fixed

- Fix memory leak when an xprt dies
- Loads of MDCACHE fixes
- Fix queuing issues of single clients filling up the queue (Dispatch_Max_Reqs_Xprt defaults to 512)
- Add caller ipaddr to a log message when we stall a transport
- Several memory leaks
- Several segmentation violations
- Change in handling of open fd's caused failure to close open fd's so running out of max_open_files which was fixed again with ibm036.14
- More crashes fixed in ibm036.15

Ganesha issues (cont.)



GPFS release and delivered ganesha levels

- 5.0.4.0 = nfs-ganesha-2.7.5(ibm053)
- 5.0.4.1 = nfs-ganesha-2.7.5(ibm053.02)
- 5.0.4.2 = nfs-ganesha-2.7.5(ibm054.03)
- 5.0.4.3 = nfs-ganesha-2.7.5(ibm054.05)
most current efix is (ibm054.06)

Major issues in the process of being fixed

- Crashes SIGSEGV in state_nfs4_state_wipe under high load
- Several SIGABRTs (crash in gsh_malloc__ crash_handler)
- fixed a crash (SIGSEGV) in mdcache_populate_dir_chunk
- fixed another crash in state_wipe_file when trying to to acquire the "obj->state_hdl->state_lock" again
- Issues in the mdcache area again, recommending efix in nfs-ganesha-2.7.5(ibm054.06)

NFS ganesha investment areas

Performance Hang	<ul style="list-style-type: none"> • Improve Performance statistics collection by implementing in code • Come up with Reference architecture for performance numbers 	<p>← Completed in 5.0.3</p> <p>← In progress with RealFast Team</p>
Memory Leak/Consumptions/Corruptions	<ul style="list-style-type: none"> • Uncover memory leaks in the code • Fix memory leak/corruption issues • Create framework for regular memory leak testing 	<p>← ASAN toolkit (Adress SANitizer)</p> <p>← Valgrind tool suite</p> <p>← Integrating memleak testing in regular test Cycle</p>
LOCKING	<ul style="list-style-type: none"> • Identify issues in Locking code base 	<p>...</p>

ACL	ACL tests improvements
AUTHENTICATION (KRB)	Authentication tests improvement
Readdir, Snapshots	Readdir tests and snapshot testing
Stability	Broader coverage of nfs testing
Miscellaneous tests	Miscellaneous tests to improve test coverage

Agenda

- Was machen wir ?
- ESS upgrade
- OPC news
- Ganesha issues
- **Surveys**
- Survey: What can we do to improve in support?



Surveys

Befragung nach closure des Tickets
Medallia Schema (<https://de.medallia.com/>)

NPS : Net Promoter Score



Wichtigste Frage :

Würden sie den IBM **Support** weiterempfehlen?

Agenda

- Was machen wir ?
- ESS upgrade
- OPC news
- Ganesha issues
- Surveys
- ***Survey: What can we do to improve in support?***



Support Improvements



Any ideas ?


Any complaints ?

Just drop me an eMail:

Achim.Rehor@de.ibm.com

Thank you!

Provide Feedback ×



Tell IBM What You Think

Let us know what you think about IBM Spectrum Scale. It takes only a couple of minutes for you to help us improve our service. [IBM Privacy Policy](#)

Please help us to improve Spectrum Scale with your feedback

- If you get a survey in email or a popup from the GUI, please respond
- We read every single reply