

Scale S3, NFS and SMB Development Update

IBM Storage Scale ISC 2026

June 23rd 2026 | Hamburg, Germany

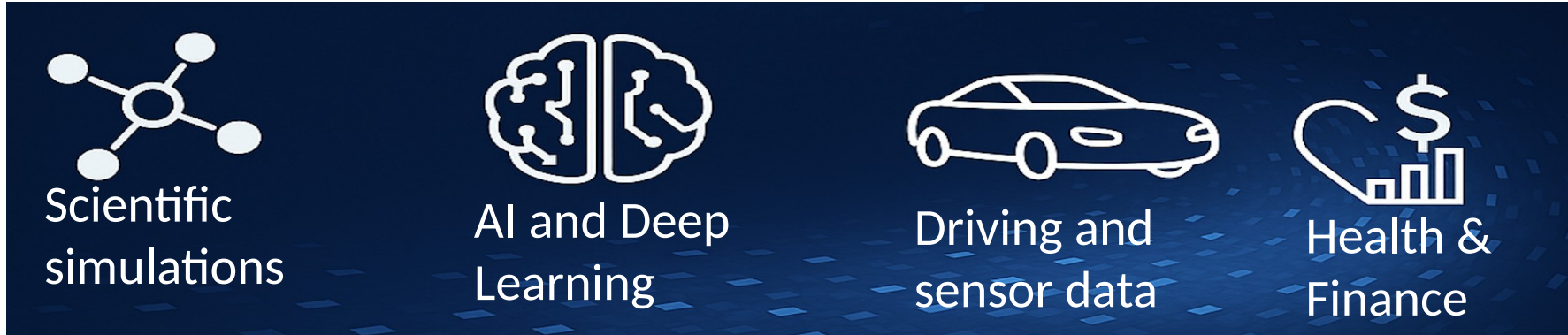
Michael Diederich

[mailto: diederich@de.ibm.com](mailto:diederich@de.ibm.com)



S3 over RDMA

AI and HPC applications generate massive amounts of data !



AI/ML or HPC applications processing require fast data access through protocols like S3 to access backend storage.

Benefits of S3 RDMA Data Access to Storage

- ✓ Offers incredible speed while still supporting the familiarity and flexibility of the S3 API.
- ✓ Ideal for AI/ML or HPC applications that require low-latency, high-throughput, and maximum efficiency from their storage.
- ✓ Control path and data path can be separated which helps boosts the AI/HPC data access performance
- ✓ Focuses on S3 operations that handle object data I/O between the S3 client and the S3 server – GetObject, PutObject, and PutPart (for multipart uploads).

IBM Storage Scale provides support for S3 over RDMA with Nvidia cuObj libraries.

Software and Hardware Supported

- ✓ Works with both CPUs and GPUs on Application nodes.
- ✓ Supported with RDMA enabled fabrics – includes both RoCE and Infiniband
- ✓ Plan to support with Nvidia cuObject server library integrated with Nooba S3 server on IBM Storage Scale CES nodes and Nvidia cuObj client library on client systems.

How it works ?

- S3 Client prepares a memory buffer for RDMA (which can be either a GPU memory or a CPU memory that is registered with the NIC) and then include a buffer descriptor in an HTTP header of the S3 operation
- Noobaa S3 Server running on CES node will use that remote buffer descriptor to get or put data directly into the client's memory
- No changes needed to the protocol's authentication, access-control, or any control flows

Performance Benchmark Results

	S3 RDMA with GPUs		S3 RDMA With CPUs	
Object Size	READ Bandwidth	WRITE Bandwidth	READ Bandwidth	WRITE Bandwidth
1 GB	~297 GB/s	~153 GB/s	~241 GB/s	~151 GB/s
128 MB	~308 GB/s	~149 GB/s	~260 GB/s	~146 GB/s
64 MB	~308 GB/s	~140 GB/s	~262 GB/s	~141 GB/s
16MB	~270 GB/s	~151 GB/s	~245 GB/s	~ 113 GB/s

* Performance was measured with single IBM Storage Scale 6000 cluster with 2 nodes

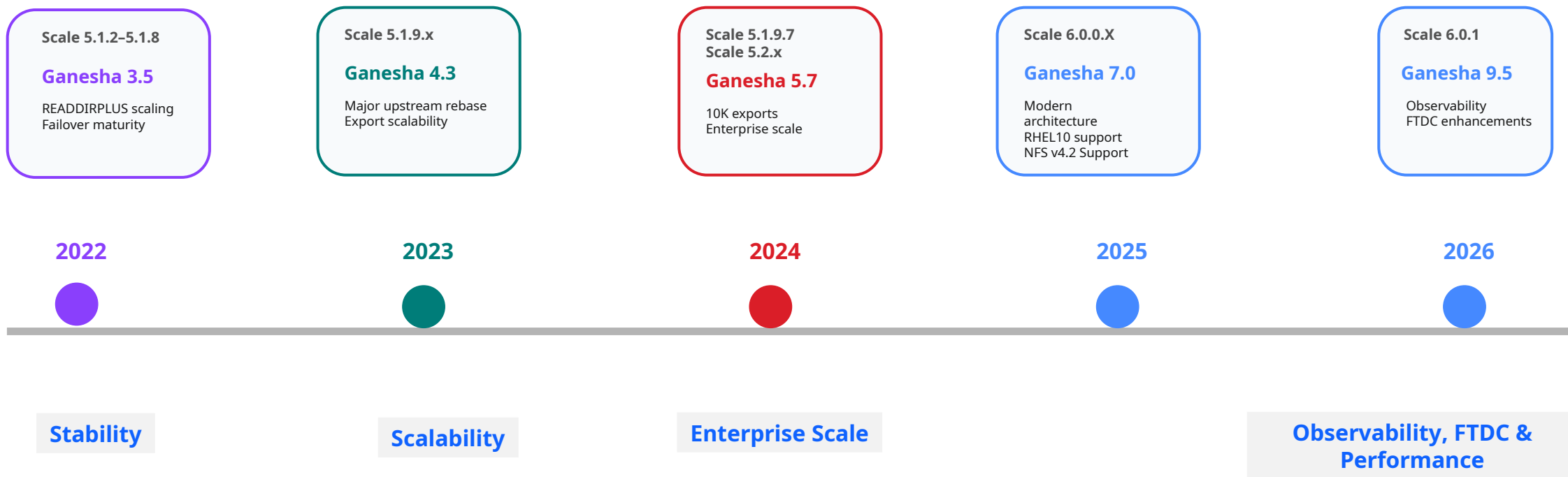
* Performance with POSIX for 1GB files in the same cluster: ~320 GB/s read, ~155 GB/s write

Conclusion:

Despite protocol overhead, **S3 over RDMA achieved 90%+ of POSIX performance**, showing its suitability for high-throughput AI and HPC workloads which prefer S3 access.

IBM Storage Scale CES NFS — Ganesha Evolution Timeline

From CES stabilization → enterprise scalability → observability, FTDC-driven diagnostics, Performance



Recommended Production Streams: 5.2.3.7+ → Ganesha 5.7 ibm028.21 LTS | 6.0.1.x → Ganesha 9.5 Observability & FTDC Generation

Key architectural transitions: Stability → Scalability → Enterprise Scale → Modern Observability

CES NFS FTDC Enhancements — IBM Storage Scale 6.0.1

From Manual, Reactive Diagnostics to Automated, On-Demand Visibility

 THE PROBLEM

Bridging Diagnostic Gaps in CES NFS Environments

- 

NFS-Ganesha issues span multiple layers
Protocol, filesystem, and network — difficult to correlate manually.
- 

Existing FTDC covered GPFS internals well, but NFS diagnostics were entirely manual.
Critical data were not collected in automated data collection.
- 

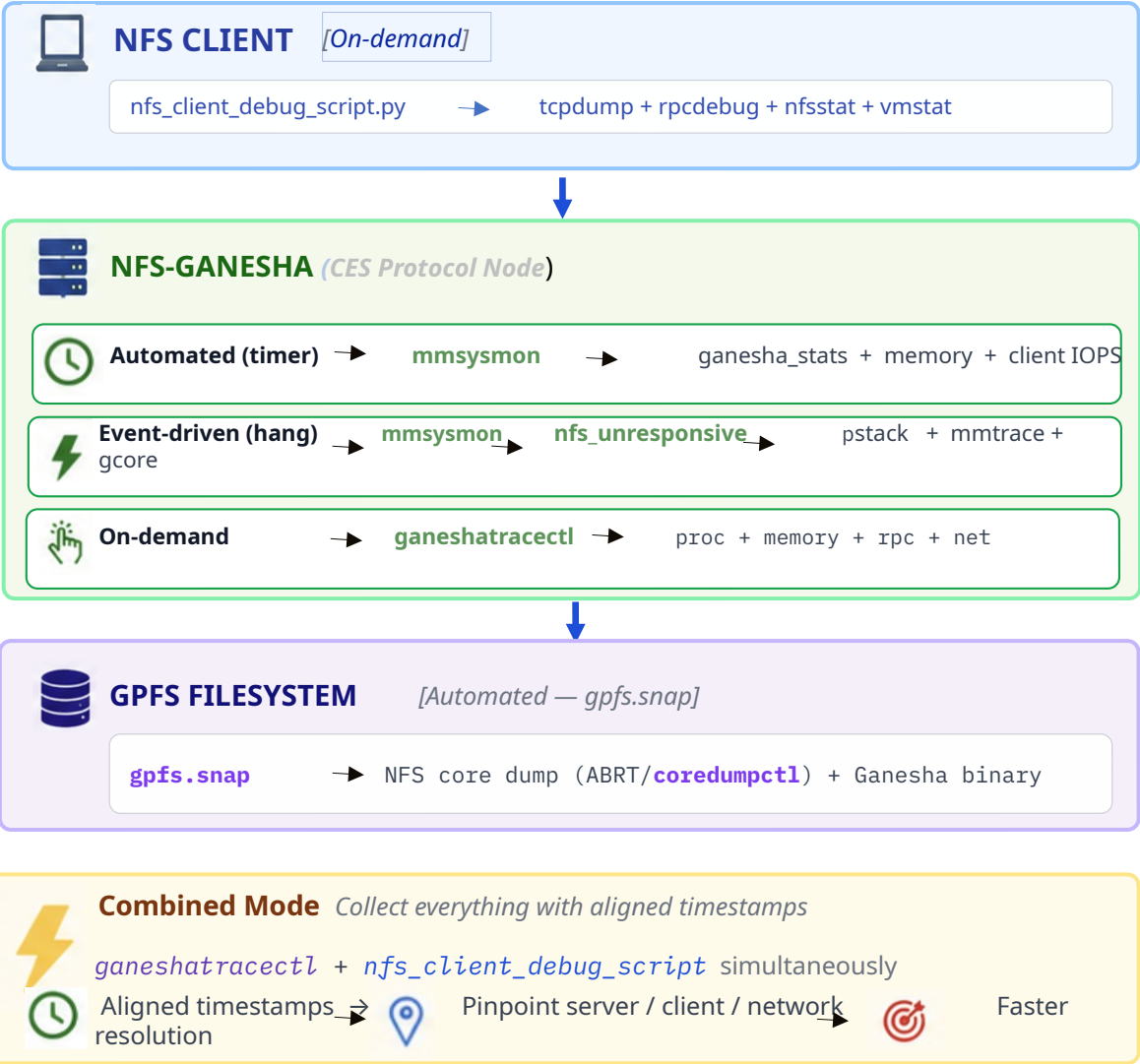
Critical diagnostic window often passed before data could be captured.
By the time collection started, the evidence was gone.

-  THE IMPACT
- ✓ End-to-end visibility across client, NFS-Ganesha, and GPFS
 - ✓ Automated collection: timer, event-driven, and on-demand
 - ✓ Faster root cause isolation and reduced time to resolution

“ From collect after the fact — to data is there when you need it ”



ARCHITECTURE



Client Layer Client-side Visibility



Protocol Layer NFS-Ganesha diagnostics



Filesystem Layer GPFS internal dump



Outcome Correlated actionable insights

NFS over RDMA — IBM Storage Scale

WHAT IS NFS OVER RDMA?

- NFS over RDMA (Remote Direct Memory Access) eliminates CPU involvement in data transfers — data moves directly between application memory and the NIC.
- Uses RDMA-over-Converged-Ethernet (RoCEv2). Both server and client must have RDMA-capable hardware.

SUPPORTED VERSION

Ganesha 9.5+

WIP

Scale 6.0.1+

STEPS TO ENABLE

1 Install Scale 6.0.1+ with Ganesha 9.5+

2 Enable RDMA transport

```
mmnfs config change transports=rdma,tcp
```

3 Verify — RDMA listening on port 20049

```
rdma_resource show cm_id
```

4 Mount from client

```
mount -t nfs -o rdma,port=20049 <srv>:<path> <mnt>
```

TEST SETUP

Single Client — Single Server - RoCEv2 200 Gbps
NFS benchmark tool : IOR
IO Block Size : 128 – 256G
Number of Threads. : 16

INITIAL PERFORMANCE NUMBERS – 128G & 256G BLOCK SIZE (PRELIMINARY)



WRITE Throughput

Block Size	RDMA (MiB/s)	TCP (MiB/s)	TCP nconnect=8 (MiB/s)
128G	2074	1787	1946
256G	2017	1799	1880

READ Throughput

Block Size	RDMA (MiB/s)	TCP (MiB/s)	TCP nconnect=8 (MiB/s)
128G	6617	1918	4304
256G	5606	1957	4276

Join Our NFS over RDMA Tech Preview

Contact IBM representative for enrolment details.

TECH PREVIEW PROGRAM
| Enroll H2 2026

NFS hangs - NFSv3 locking problem ?



SYMPTOMS

- Clients see no NFS responses in time – this can have many reasons – NFS is merely the conduit for IOs



IS IT A NFSv3 LOCKING PROBLEM ?

- Exclude node failover induced issues. If there are CES IP moves, this may not be a NFSv3 locking problem.



CHECK GANESHA LOG

Do we see messages like :

```
2024-10-10 11:29:34 : epoch 00040079 :  
ibmessprot1 :  
gpfs.ganesha.nfsd-370988[svc_33777]  
nsm_connect :NLM :EVENT :connect to statd  
failed:
```

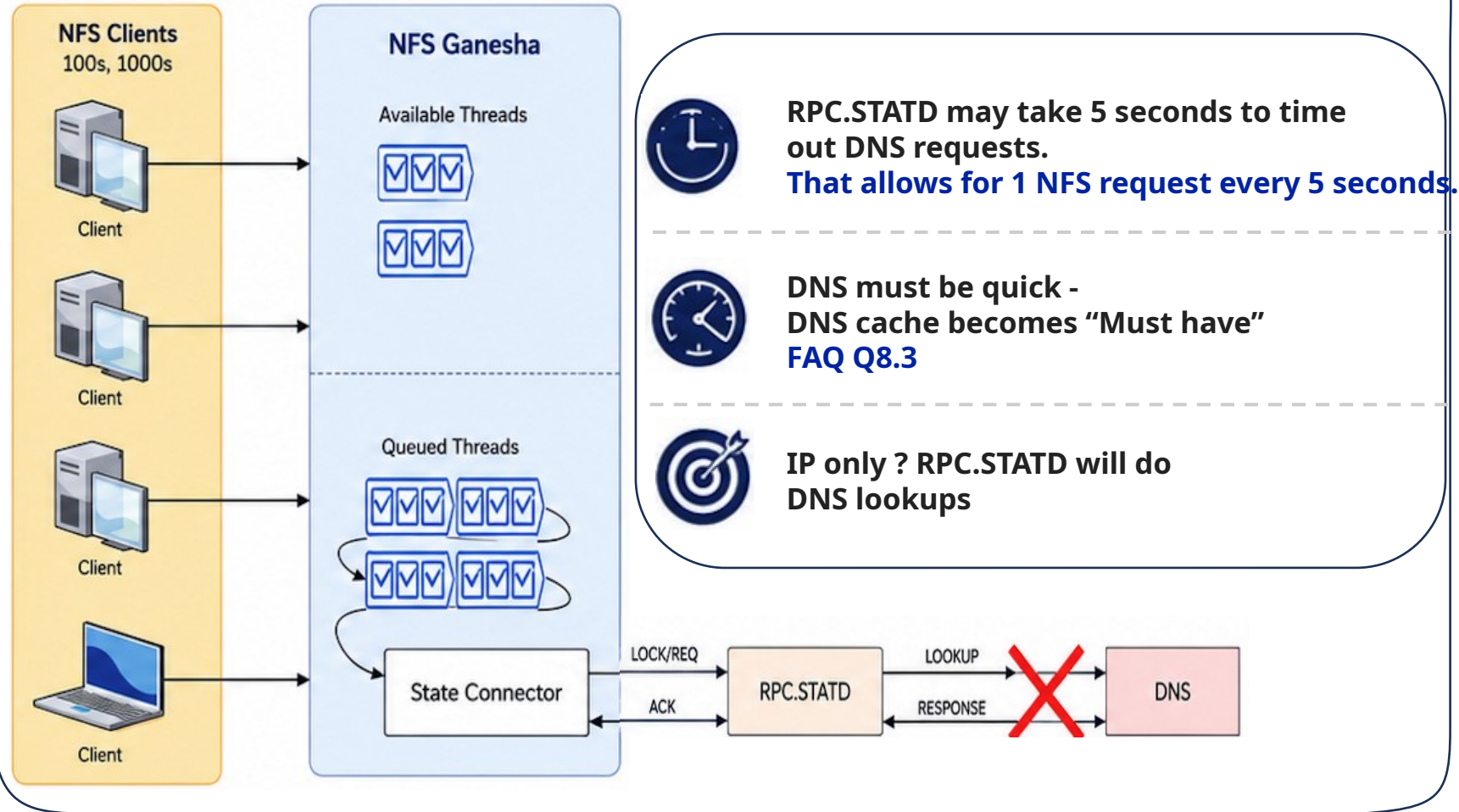
RPC: Unknown protocol

```
2024-10-10 11:29:34 : epoch 00040079 :  
ibmessprot1 :  
gpfs.ganesha.nfsd-370988[svc_33777]  
nsm_monitor_noretry :NLM :EVENT :Monitor  
10.141.131.64
```

nsm_connect failed "Clients see no NFS responses"



WHY NFSv3 LOCKING CAN HANG - RPC.STATD & DNS



KEY TAKEAWAY



Eliminate failover induced issues (CES IP moves).



Ensure fast, reliable DNS with proper caching.



Use Ganesha logs to confirm NFSv3 locking issues.

SMB Features — IBM Storage Scale 6.0.1

Diagnostics · DC Failover Reliability · Multichannel CLI · Samba Currency

01 SMB FTDC Improvements

New additions to gpfs.snap → smb.snap

- 1

testparm -sv
Captures default Samba config. Identifies if customer changed settings outside of Scale code.
- 2

smbcontrol smbd num-children
Number of SMB clients currently connected to each CES node.
- 3

CTDB data (ctdb listvars + mmlsperfdata hop counts)
Collects CTDB config variables and hop count / call latency performance metrics.

KEY BENEFIT
Faster SMB triage with rich diagnostic data captured automatically via gpfs.snap.

02 DC Failover Improvements

- Problem

DC affinity causes auth failures when a DC reboots. Node retries same DC through full TCP timeouts — all uncached users fail.
- Solution

New ldap_timeout parameter (default 10s). On LDAP timeout, immediately invalidates current DC and starts parallel DC probing.

- KEY IMPROVEMENTS

■

DC failover: minutes → ~10 seconds

■

No auth failures during DC maintenance

■

NFS/AD services remain unaffected

■

Enhanced winbind DC affinity management

03 Multichannel CLI Support

- Background

SMB Multichannel was introduced in Scale 6.0.0 but the CLI configuration commands were missing. Scale 6.0.1 adds full CLI control.

- Enable Multichannel

mmsmb config change \ --option "server multi channel support=yes"

- Disable Multichannel

mmsmb config change \ --option "server multi channel support=no"

- KEY BENEFIT

Enable/disable SMB multichannel via standard mmsmb CLI — no manual config edits.

04 Samba Currency Upgrade

- Version Upgrade

4.22.4

Previous

→

4.22.8

Scale 6.0.1

- What's included

■

Security fixes and CVE patches

■

Stability and reliability improvements

■

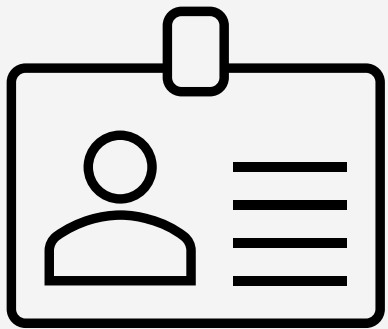
Bug fixes from upstream Samba community

- KEY BENEFIT

Latest upstream fixes and security hardening in Scale 6.0.1 Samba stack.

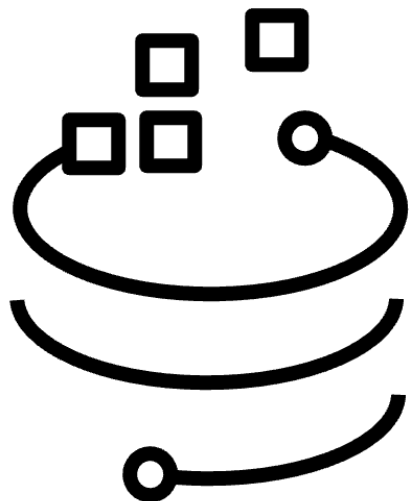
Recommended Production Streams: 5.2.3.7+ - samba 4.20.8 LTS | 6.0.1.x → 4.22.8 FTDC Generation

Thank you



Michael Diederich
diederich@de.ibm.com



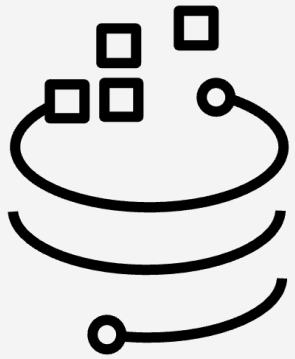


Storage Scale

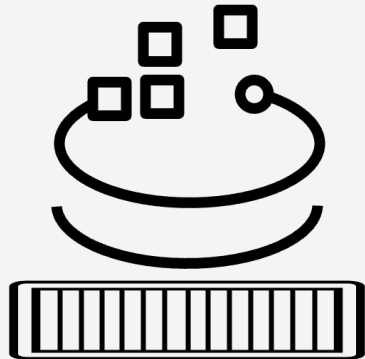
Disclaimer #2

- IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.
- IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Thank you for using



Storage Scale



Storage Scale
System

