IBM Storage for Data and AI - Solutions
High Performance SMB/CIFS

TUXERA Make it work.

Chris Maestas
IBM Chief Executive Architect
Storage Solutions for Data and AI
cdmaestas@us.ibm.com

TUXERA

Storage and networking technologies from micro-controllers to global public clouds

Who we are

Data management leader

17	top technology companies served
25+	years of open-source contributions
30+	awards
65%	CAGR since 2009

150	employees
300+	projects per year
300+	million product lines powered by Tuxera

Close partnership with Microsoft since 2009

TUXERA

Fusion File Share by Tuxera

World’s most advanced and scalable enterprise SMB server on Linux

Key advantages of Fusion File Share

FUSION FILE SHARE
By Tuxera

Our high-performance, highly-scalable, drop-in replacement for Samba.

- Highly threaded architecture
- High-performance – 2x to 60x faster than SAMBA
- 100% to 500% better scalability than SAMBA
- Fault tolerant with Transparent Failover and Continuous Availability
- Extensive SMB-protocol support – 3.1.1
- Scale-out (active-active)
- RDMA (SMB-Direct), Multichannel, and Compression
- Low CPU and memory usage
- Low latency
- Native GPFS support

Key advantages of Fusion File Share

FUSION FILE SHARE By Tuxera

Highly threaded architecture with adjustable settings for different workloads.

All configuration and tuning changes are applied runtime.

Each client connection is a thread, not a process:

- Data transport threads
- Meta data transport threads
- VFS data threads
- VFS meta data threads
- Minimized CPU & memory usage

Adjustable quality of service by tuning:

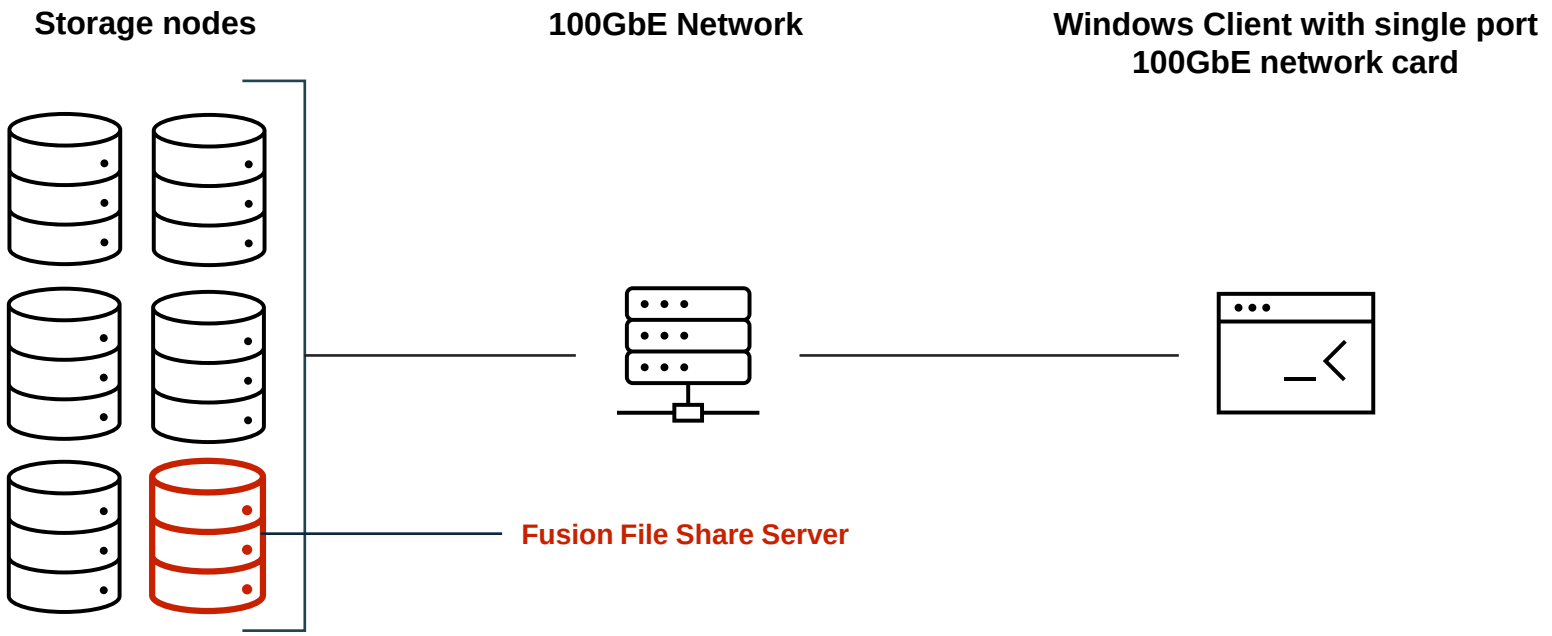
- Concurrent open files
- Concurrent client connections
- Concurrent open files per user-session
- Concurrent VFS threads per share

Enterprise features

- Windows Active Directory
- Advanced ACL handling
- Multiprotocol support: ACL, Shared access
- Custom VFS support
- Custom clustering support
- Persistent handles
- Continuous availability, with single, dual or multinode
- Transparent failover
- High availability
- Change notify
- Secure dialect negotiation
- Encryption: AES-256-CCM, AES-256-GCM
- Authentication: NTLM, Kerberos, LDAP
- Pre-authentication integrity
- Audit/logging support
- DFS support
- Dynamic configuration change
- Quota support
- Internal health monitoring
- Runtime statistics

Performance benchmarks

Single client performance test setup



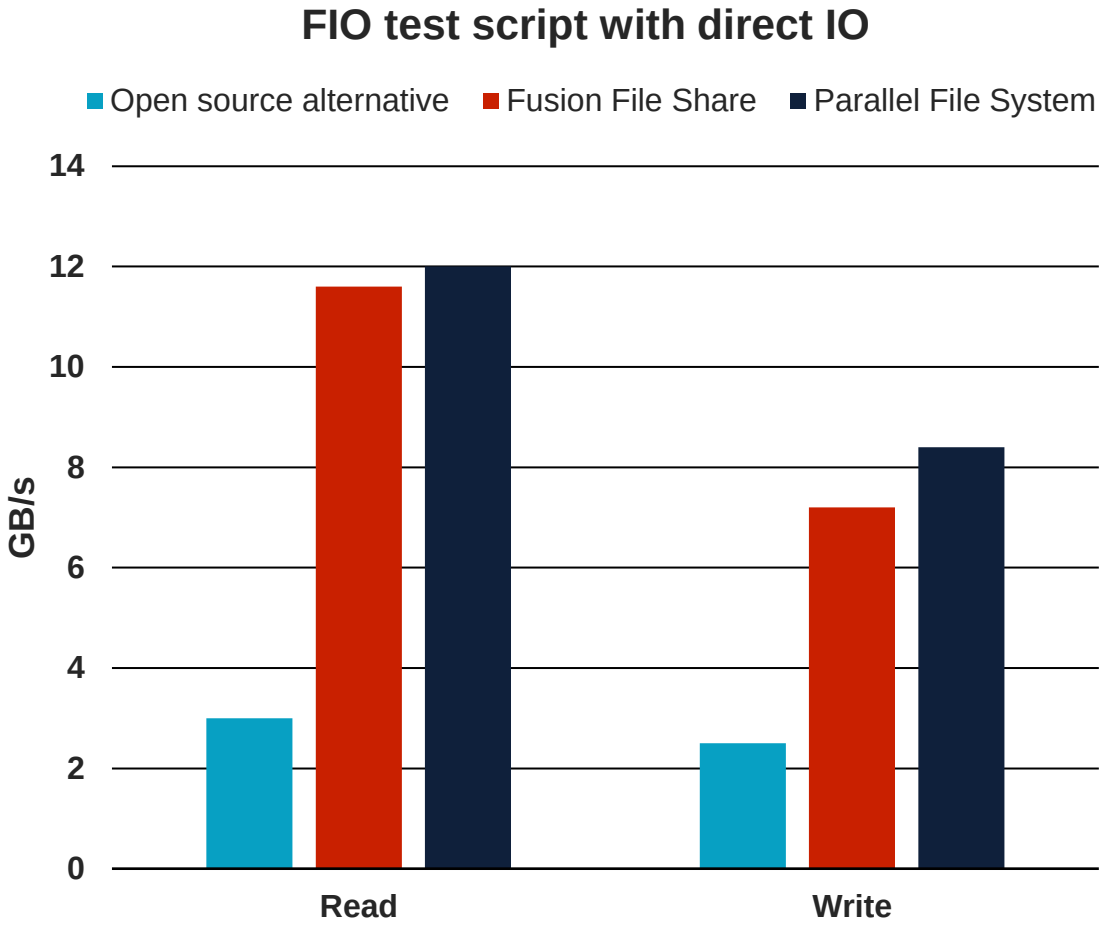
Single client performance

Fusion File Share contributes **over 85%** of the speed throughput for high-performance parallel file systems

Single client write performance		Single client read performance	
Fusion File Share	Parallel file system	Fusion File Share	Parallel file system
7.2 GB/s	8.4 GB/s	11.6 GB/s	12.0 GB/s

Test setup:

- Fusion File Share server: Active-passive, fault tolerant configuration used as the SMB gateway, running on a storage node.
- Parallel file system storage: 6 nodes of Supermicro architecture:
- Intel Xeon Gold 6226R, 192GB DDR4-2933 ECC REG SDRAM, Micron 9300 MAX 3.2TB NVMe PCIe 3.0 3D TLC U.2, Mellanox AOC-MCX555A-ECAT CX-5 VPI EDR IB adapter & 100GbE,1p, QSFP28, PCIe3x16
- Windows client: single port 100GbE network card with 2 x Xeon 4214 and 768 GB RAM
- Network is running 100GbE end-to-end, through a Mellanox 100GbE switch.



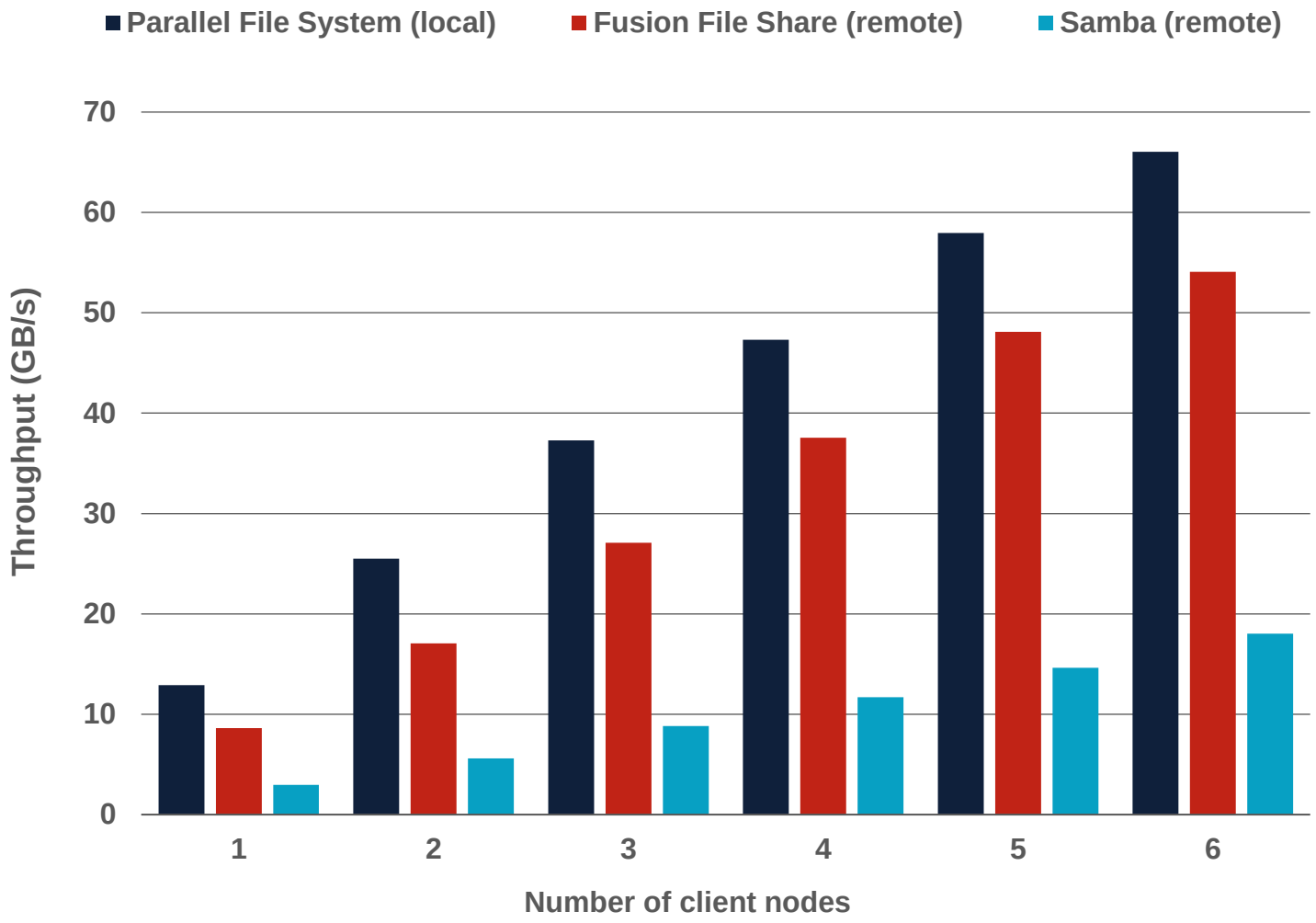
Actual performance may vary based on the hardware, software, and testing protocols used.

TUXERA

Maximize link speed potential with linear scaling

Samba is outperformed by Fusion with one client. As more clients are added, Samba continues to underperform compared to Fusion.

Scale-out sequential read performance comparison
Fusion File Share versus Samba using FIO



Actual performance may vary based on the hardware, software, and testing protocols used.

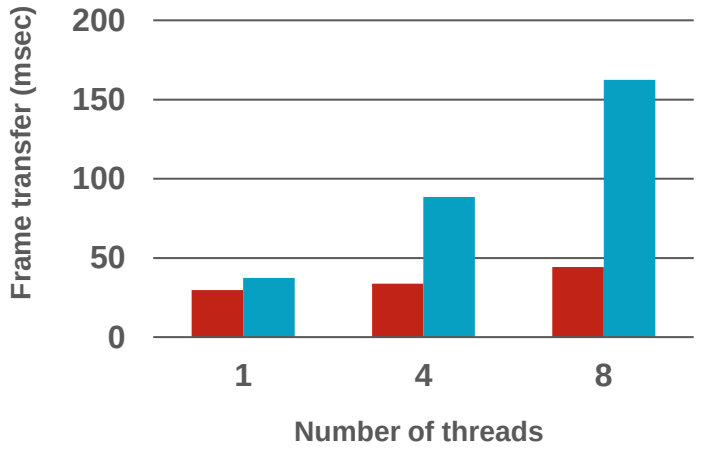
TUXERA

Up to 2.7x multi-threaded performance advantage over open source

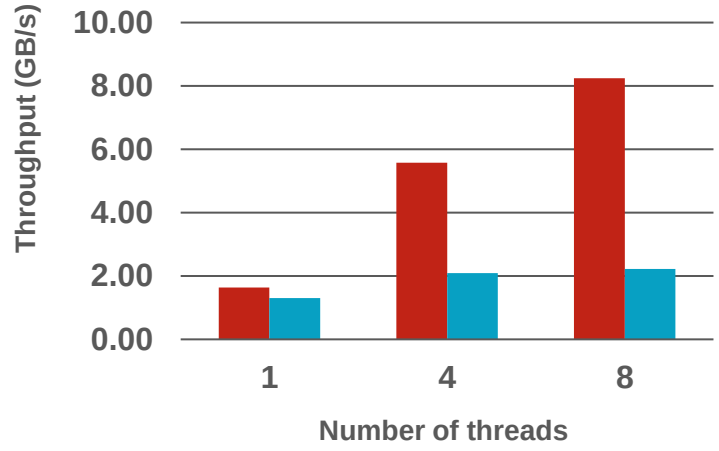
M&E workload performance comparison

Fusion File Share versus open source alternative using Frametest

Lower values are better.



Higher values are better.

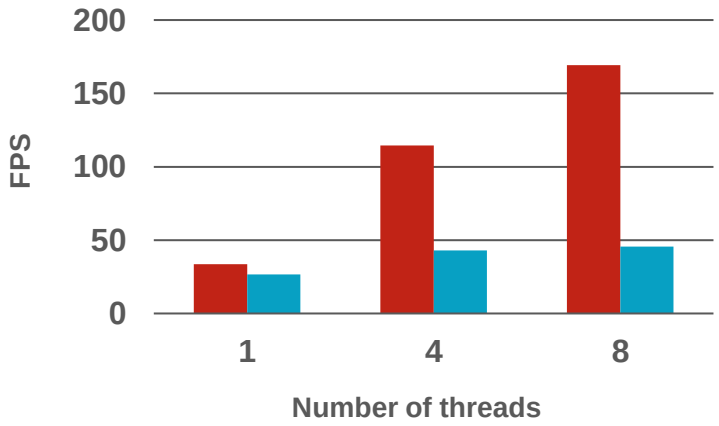


Fusion

Open source alternative

Frametest parameters:
4K // 2000 frames

Higher values are better.



Actual performance may vary based on the hardware, software, and testing protocols used.

TUXERA

M&E Customer Benchmark

- **Use case**
  - Concurrent video editing and color grading from shared storage

- **Customer issues**
  - Not getting enough frame rate to accommodate current workflow
  - Storage performance ~30GB/s (NVMe) not fully utilized
  - 100GB infiband network not fully utilized
  - RDMA not in use
  - Max performance of single node Samba ~2.8GB/s
  - 3 extra protocol nodes needed just for Samba to saturate network

M&E Customer Benchmark

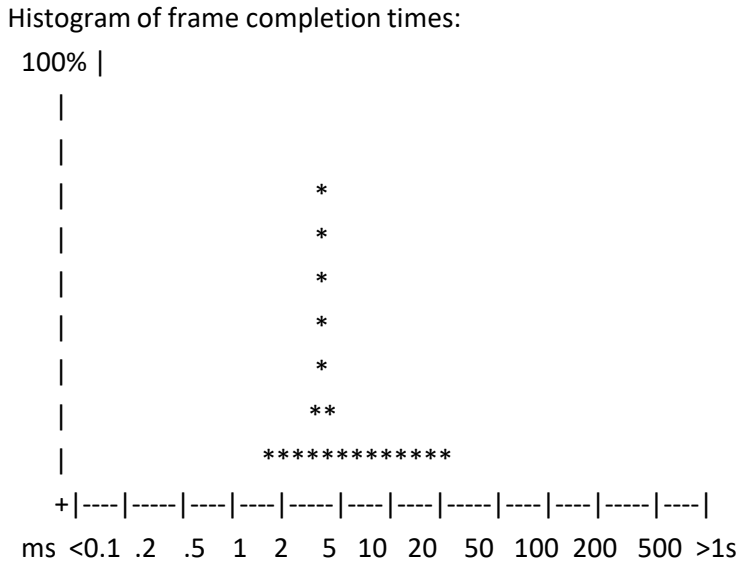
- Single Fusion File Share instance
- Running on storage node
  - No extra protocol nodes
- RDMA enabled

Benchmark - Write 4K, 9000 frames, 4 threads (RDMA)

	Open	I/O	Frame	Data rate	Frame rate
Last 1s:	2.978 ms	14.69 ms	4.96 ms	9819.44 MB/s	201.7 fps
5s:	1.547 ms	14.35 ms	4.95 ms	9833.37 MB/s	202.0 fps
30s:	1.883 ms	14.35 ms	4.96 ms	9819.71 MB/s	201.7 fps
Overall:	1.853 ms	14.38 ms	4.95 ms	9831.87 MB/s	201.9 fps

Frame Test

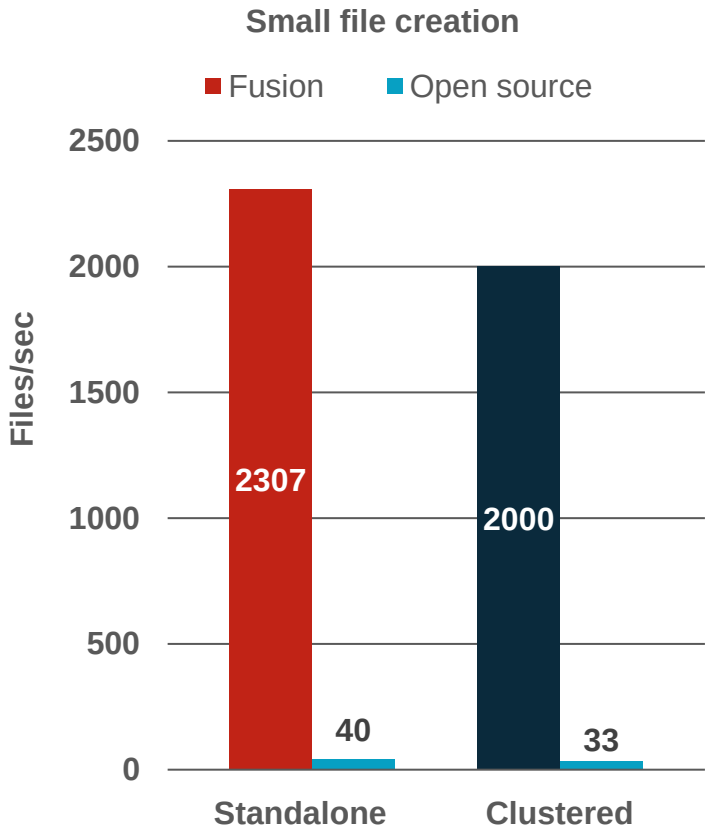
Test parameters: -w49856 -n9000 -t4 (4K, 9000 frames, 4 threads)
Test duration: 44 secs
Frames transferred: 8925 (434535.938 MB)
Fastest frame: 6.916 ms (7039.53 MB/s)
Slowest frame: 34.325 ms (1418.44 MB/s)



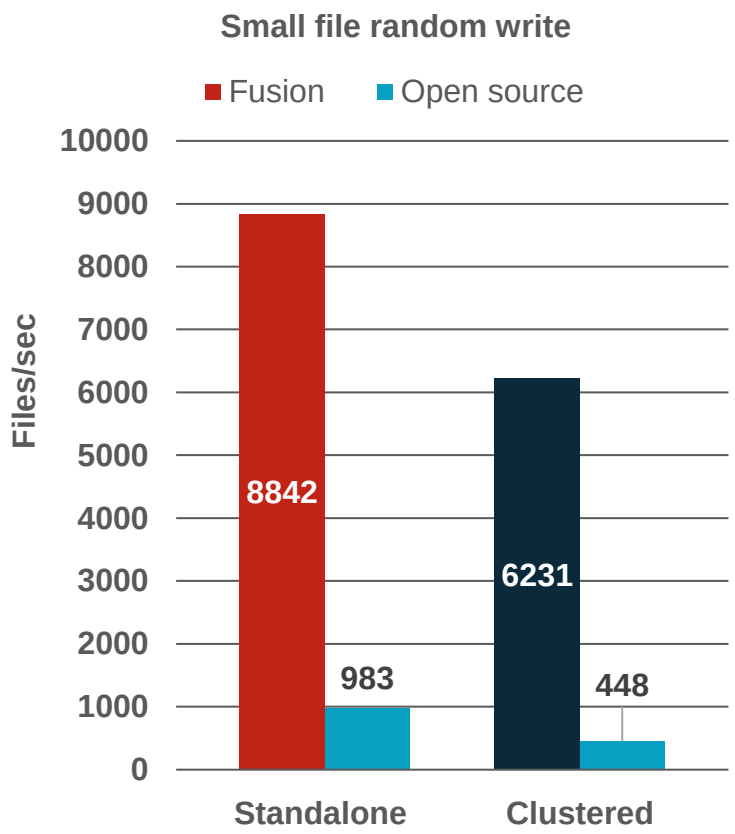
Up to 61x small file creation performance advantage over open source when clustered

Small file performance comparison

Standalone & clustered Fusion File Share vs open source using Oracle vdbench



Workload: create, write 1 kB, close 30,000 files in a single directory



Workload: randomly open, write 1 kB, close files in a directory with 30,000 files for a period of 30 seconds

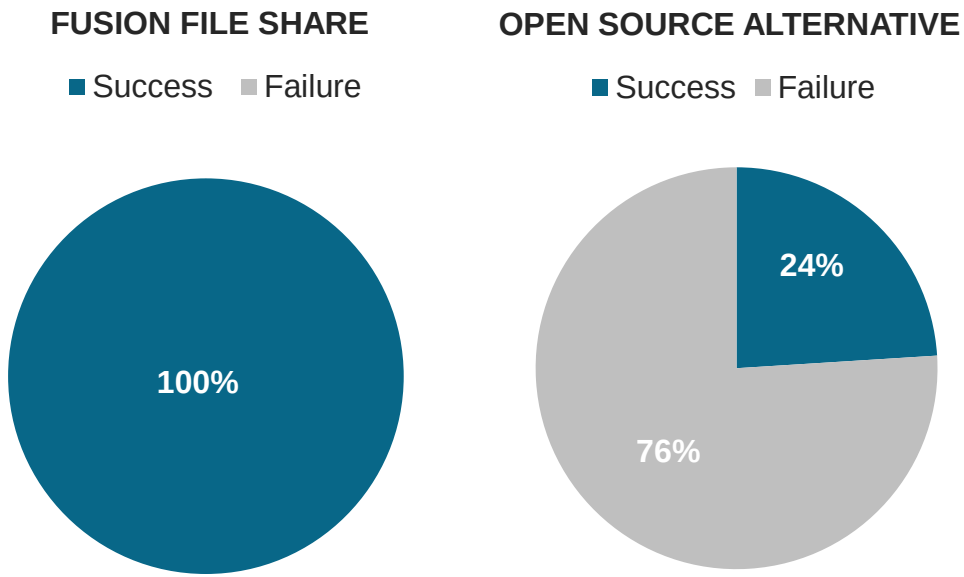
Actual performance may vary based on the hardware, software, and testing protocols used.

TUXERA

Fast, successful connections

The open source alternative failed to meet the required performance benchmark of connecting 200 clients per second at a rate of 76%

SMB connection rate
(200 new clients generated per second)



Test setup: Lenovo P52s Mobile Workstation // 8th Generation Intel® Core™ i7-8650U Processor with vPro® (1.90GHz, up to 4.20GHz with Turbo Boost, 8MB Cache) // Ubuntu Linux version 4.15.0-52-generic // 32 GB DDR4 (16 + 16) 2400MHz RAM // 1 TB Solid State Drive, PCIe-NVMe OPAL2.0 M.2 // 1 Gigabit Ethernet // Open source alternative

Actual performance may vary based on the hardware, software, and testing protocols used.

TUXERA

POC Environment

Test Cases

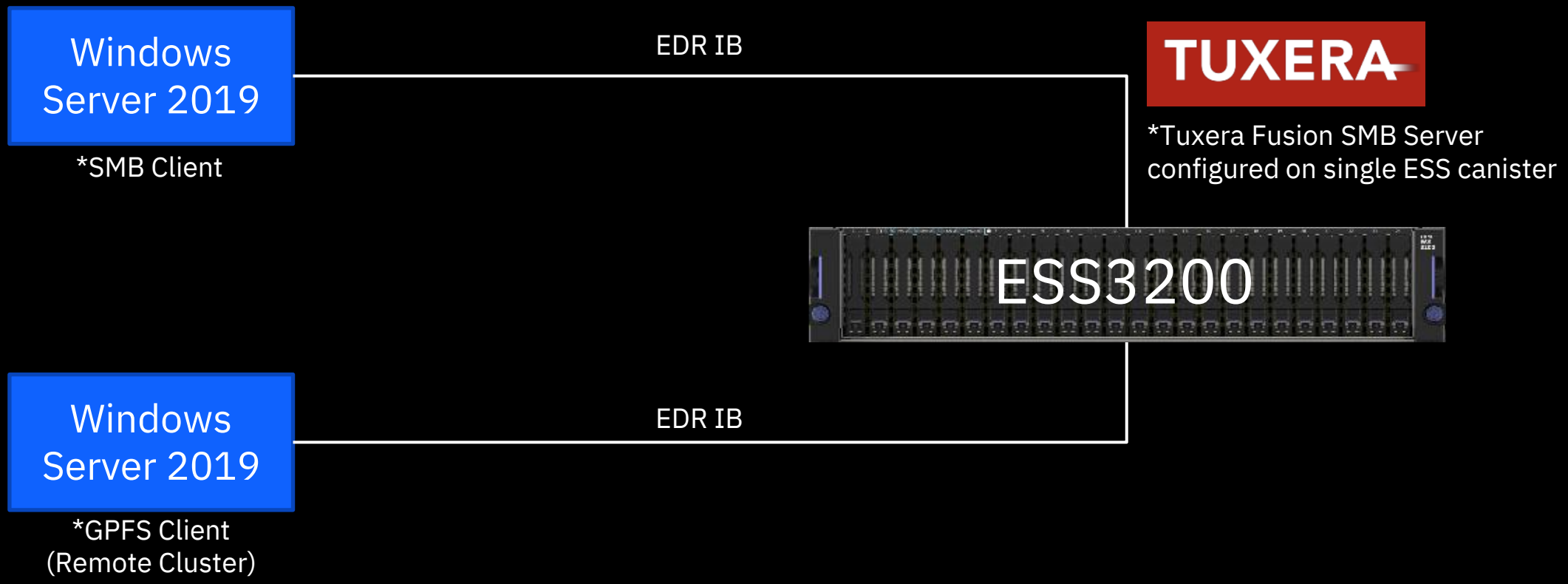
1- Single Client SMB2 Performance over TCP (IPoIB)

2- Single Client SMB3 Multi-Channel over TCP (IPoIB)

3- Single Client SMB3 Direct using RDMA (EDR IB)

4- Multi-Client SMB3 Direct using RDMA (EDR IB)

5- Single GPFS Client using RDMA (EDR IB)



POC/Benchmark Results – 4GB Filesize

FIO Write Test:

fio.exe --name=fiotest --directory=\\ESS32KSMB\ess32kshare\ --size=4G --rw=write --bs=4M --numjobs=24 --ioengine=windowsaio --iodepth=16 --group_reporting --runtime=60 --ramp_time=30 --direct=1

Test	Numjobs	xfersize	Avg MiB/s Write	Avg IOPs Write
Single Client SMB2 TCP	24	4M	2615	616
Single Client SMB3 Multi-Channel TCP	24	4M	9840	2519
Single Client SMB3 Direct RDMA	24	4M	9998	2499
Multi-Client SMB3 Direct RDMA	24	4M	TBD	TBD
Single Scale Client RDMA	24	4M	3039	685

FIO Read Test:

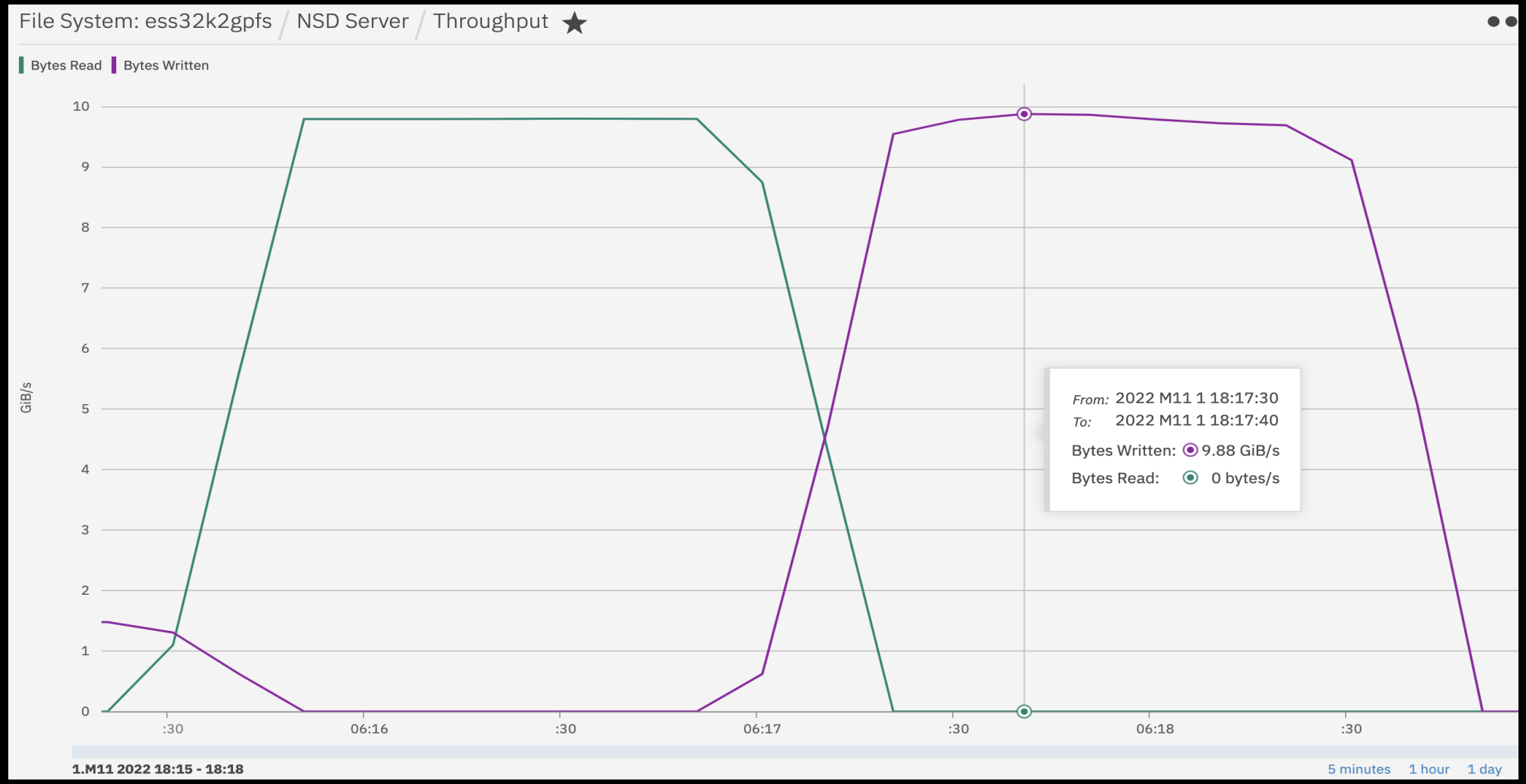
fio.exe --name=fiotest --directory=\\ESS32KSMB\ess32kshare\ --size=4G --rw=read --bs=4M --numjobs=24 --ioengine=windowsaio --iodepth=16 --group_reporting --runtime=60 --ramp_time=30 --direct=1

Test	Numjobs	xfersize	Avg MiB/s Read	Avg IOPs Read
Single Client SMB2 TCP	24	4M	3390	847
Single Client SMB3 Multi-Channel TCP	24	4M	10600	2718
Single Client SMB3 Direct RDMA	24	4M	11000	2816
Multi-Client SMB3 Direct RDMA	24	4M	19598	4898
Single Scale Client	24	4M	4972	1242

ESS Backend

FIO Test:

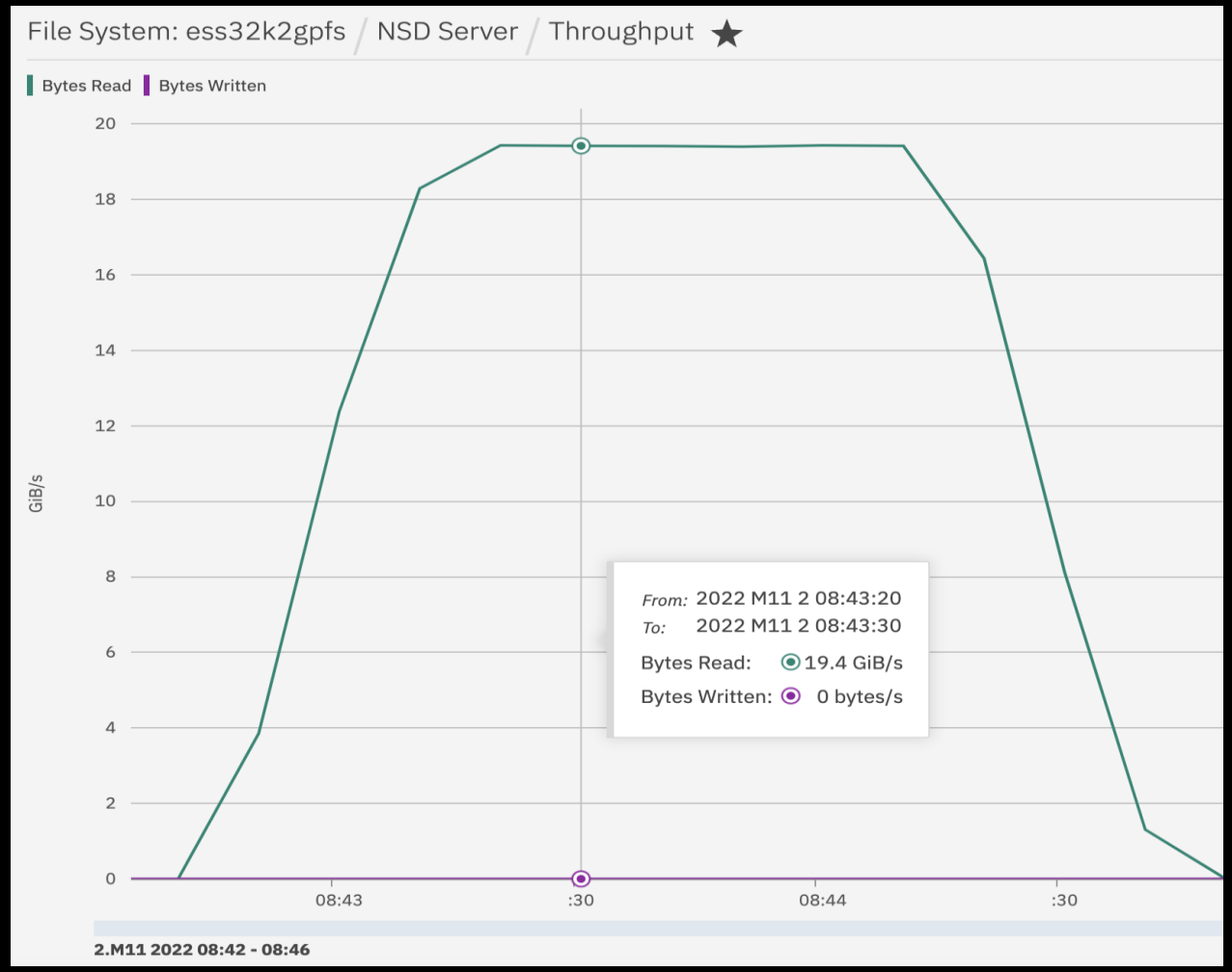
fio.exe --name=fiotest --directory=\\ESS32KSMB\ess32kshare\ --size=100G --rw=read --bs=4M --numjobs=24 --ioengine=windowsaio --iodepth=16 --group_reporting --runtime=60 --ramp_time=30 --direct=1



ESS Backend

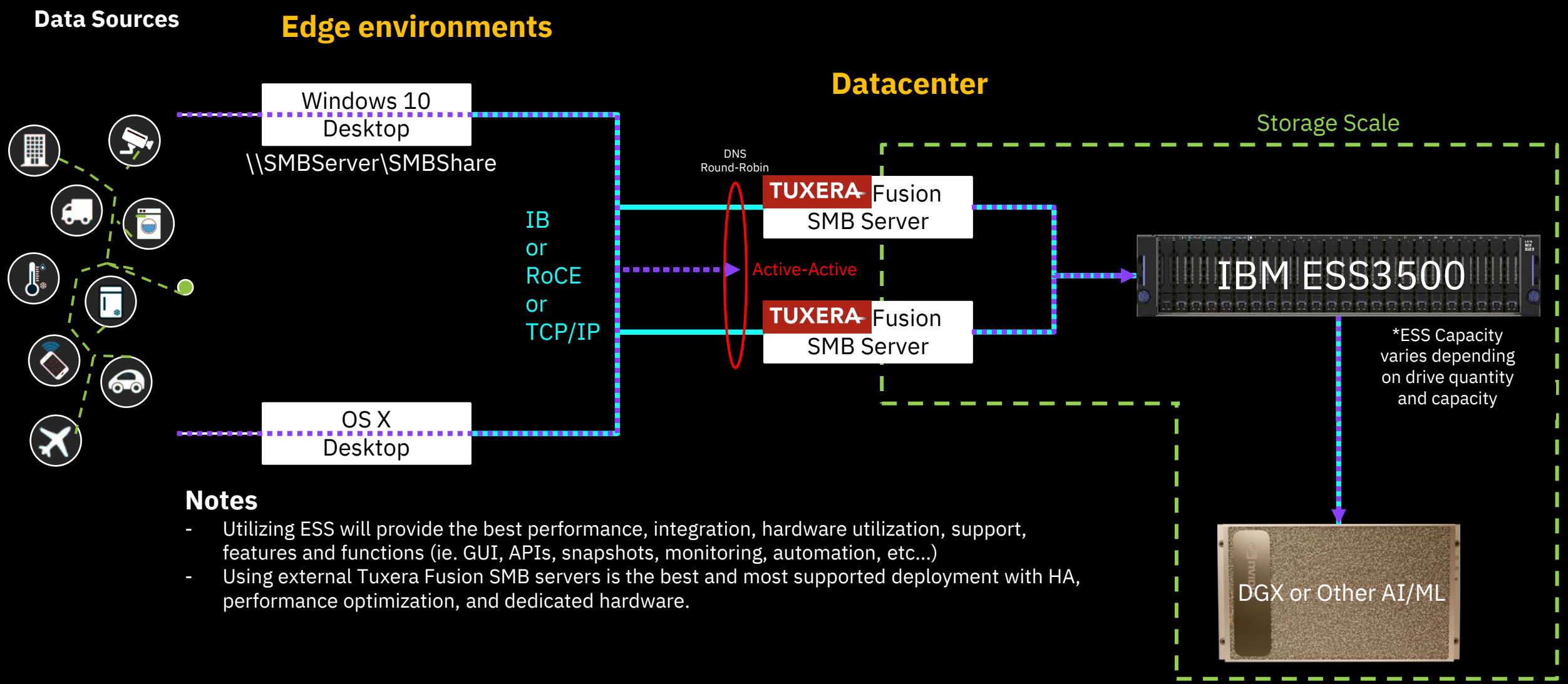
FIO Test:

fio.exe --name=fiotest --directory=\\ESS32KSMB\ess32kshare\d1 --size=100G --rw=read --bs=4M --numjobs=24 --ioengine=windowsaio --iodepth=16 --group_reporting --runtime=60 --ramp_time=30 --direct=1



Recommended Deployment

NSD+Tuxera Server Recommendations
- Dual CPU (2x 16C)
- 128GB+ Memory
- 2x Single port CX-6 HDR



Notes
- Utilizing ESS will provide the best performance, integration, hardware utilization, support, features and functions (ie. GUI, APIs, snapshots, monitoring, automation, etc...)
- Using external Tuxera Fusion SMB servers is the best and most supported deployment with HA, performance optimization, and dedicated hardware.


IBM

SAMBA has technical limitations

Not developed at the same pace as Microsoft's SMB
- Especially SMB 3.0 and up

License issues with GPLv3

Limited performance
- Process per connection
- Limited multichannel support
- No RDMA support
- No inline compression support

Low scalability
- Low number of concurrent opens
- Low number of concurrent connections
- Poor random workload support
- High CPU and memory usage

Lack of enterprise features
- No continuous availability
- No persistent handles
- No application transparent cluster support
- No Direct IO support