QSAN Official Document

# GDS Integration Guide

## Application Note

June 2025

# ANNOUNCEMENT

## Copyright

© Copyright 2025 QSAN Technology, Inc. All rights reserved. No part of this document may be reproduced or transmitted without written permission from QSAN Technology, Inc.

QSAN believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

## Trademarks

- QSAN, the QSAN logo, and QSAN.com are trademarks or registered trademarks of QSAN Technology, Inc.
- Microsoft, Windows, Windows Server, and Hyper-V are trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries.
- Linux is a trademark of Linus Torvalds in the United States and/or other countries.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Mac and OS X are trademarks of Apple Inc., registered in the U.S. and other countries.
- Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
- VMware, ESXi, and vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other countries.
- Citrix and Xen are registered trademarks or trademarks of Citrix Systems, Inc. in the United States and/or other countries.
- Other trademarks and trade names used in this document to refer to either the entities claiming the marks and names or their products are the property of their respective owners.

# TABLE OF CONTENTS




©2025 QSAN Technology, Inc. All rights reserved.

# FIGURES





**Announcement**
©2025 QSAN Technology, Inc. All rights reserved.
www.qsan.com

Official Document

# TABLES




Announcement
©2025 QSAN Technology, Inc. All rights reserved.
www.qsan.com
Official Document

# NOTICES

The information contained in this document has been reviewed for accuracy. But it could include typographical errors or technical inaccuracies. Changes are made to the document periodically. These changes will be incorporated in new editions of the publication. QSAN may make improvements or changes in the products. All features, functionality, and product specifications are subject to change without prior notice or obligation. All statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.


©2025 QSAN Technology, Inc. All rights reserved.

# PREFACE

## Technical Support

Do you have any questions or need help trouble-shooting a problem? Please contact QSAN Support, we will reply to you as soon as possible.

- Via the Web: https://www.qsan.com/technical_support
- Via Telephone: +886-2-77206355
- (Service hours: 09:30 - 18:00, Monday - Friday, UTC+8)
- Via Skype Chat, Teams / Skype ID: qsan.support
- (Service hours: 09:30 - 02:00, Monday - Friday, UTC+8, Summer time: 09:30 - 01:00)
- Via Email: support@qsan.com

## Information, Tip, and Caution

This document uses the following symbols to draw attention to important safety and operational information.

### INFORMATION

INFORMATION provides useful knowledge, definition, or terminology for reference.

### TIP

TIP provides helpful suggestions for performing tasks more effectively.


©2025 QSAN Technology, Inc. All rights reserved.

## CAUTION

CAUTION indicates that failure to take a specified action could result in damage to the system.


©2025 QSAN Technology, Inc. All rights reserved.

# 1. INTRODUCTION TO GDS

Welcome to this document that highlights the seamless integration of GDS (GPUDirect Storage) with XN series storage solution. In today's data-intensive world, especially in AI training, deep learning, and HPC (High-Performance Computing) applications, the need for high-performance, low-latency data access is critical.

GDS enables GPUs to directly access storage devices without CPU intervention, significantly reducing bottlenecks and improving overall system performance. Combined with the powerful infrastructure of QSAN XN series storage, GDS provides a strong foundation for enterprises to accelerate data processing workflows.

## 1.1. What is RDMA

RDMA (Remote Direct Memory Access) is a network technology that allows one computer to access the memory of another computer without CPU or operating system involvement. It is widely used in HPC and distributed storage environments.

**Key Operating Principles of RDMA**

- Direct memory-to-memory data transfer between nodes (zero copy)
- No CPU involvement in data movement
- Transfers over InfiniBand, RoCE (RDMA over Converged Ethernet), or iWARP


©2025 QSAN Technology, Inc. All rights reserved.

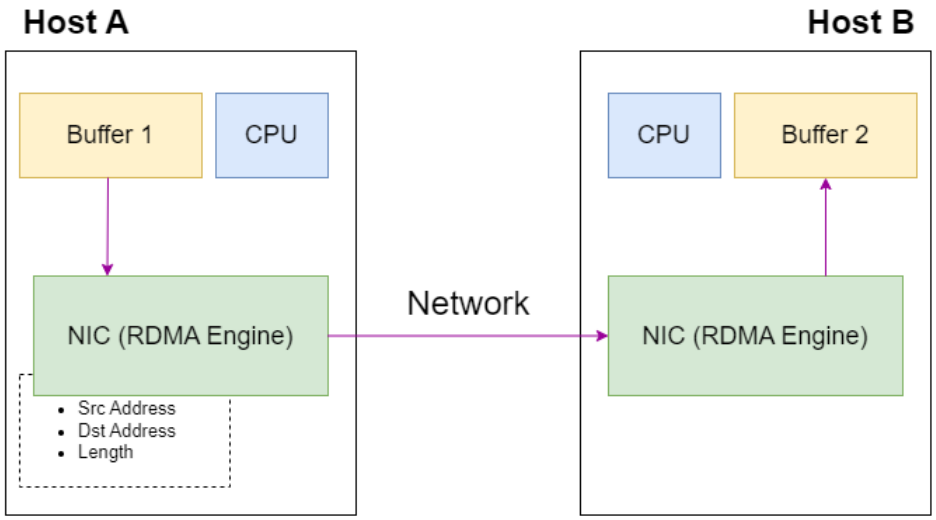


*Figure 1-1 RDMA Operation Diagram*

**RDMA Benefits**

- Low latency, eliminating CPU and OS data copy overhead
- High throughput, leveraging hardware-accelerated transfers
- Low CPU load, CPU initiates transfers; DMA engine handles movement

## 1.2. What is GDS

GDS (GPUDirect Storage) is an NVIDIA technology designed to optimize GPU direct access to storage systems, reduce CPU overhead, and improve data transfer efficiency.

**Traditional GPU Data Flow**

1. Application reads data from NVMe SSD or NFS storage to CPU memory
2. CPU copies data to system memory
3. CPU transfers data to GPU memory via PCIe
4. GPU starts processing

**GDS Optimized Flow**

1. GPU issues DMA request
2. Data is transferred directly from NVMe SSD or NFS storage to GPU memory via RDMA


©2025 QSAN Technology, Inc. All rights reserved.

3. No CPU memory cache required
4. Faster data availability and lower CPU load

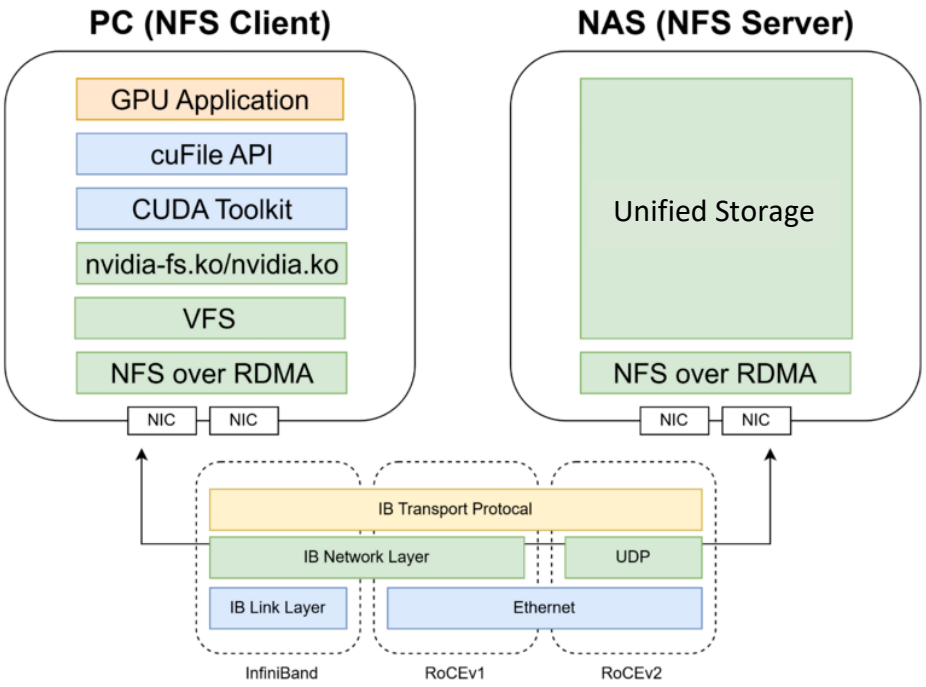


*Figure 1-2 GDS Operation Diagram*

**GDS Benefits**

Reference - NVIDIA GDS Benefits

- Reduced latency
- Reduced CPU overhead
- Increased data throughput
- Improved overall application performance

---

## INFORMATION

For more information, see NVIDIA GDS Benefits.

---


©2025 QSAN Technology, Inc. All rights reserved.

# 2. CONFIGURATION SETTINGS AND TEST RESULTS

This chapter will introduce the core concepts of RDMA (Remote Direct Memory Access) and GDS configuration settings, test results, and real-world use cases to guide enterprises to maximize their data-driven workloads through efficient and scalable solutions.

## 2.1. Environment and Topology

**Demonstration Environment**

- Server

  CPU: 16 cores
  Memory: 64 GB (4 x 16 GB)
  GPU: NVIDIA RTX A2000 (GDS enabled)
  NIC: Mellanox ConnectX-4 Lx RoCE NIC
  OS: Ubuntu 22.04.5 LTS

  - Software:

    NVIDIA Driver (checked via nvidia-smi)
    CUDA 12.6 Toolkit
    NVIDIA GDS libraries
    rdma-core utilities

**TIP**

- For best performance, IOMMU must be disabled before installing GDS.
- Ensure CUDA version matches NVIDIA driver to ensure compatibility.

- Storage

  - Model: XN5112S
    Memory: 16 GB
    Firmware: 4.1.5


©2025 QSAN Technology, Inc. All rights reserved.

HDD: Seagate Cheetah 15K.7, ST3300657SS, 300 GB, SAS 6 Gb/s
NIC: Intel E810 (iRDMA driver)

- NFS Server Settings:
  Export tmpfs-based NFS folder to avoid backend disk I/O bottlenecks
  Use NFS v4.0 with RDMA (RoCE)Demonstration Topology

### Topology

The connection topology is simple, direct high-speed RDMA network (preferably RoCE v2) between the GPU server and the unified storage system.

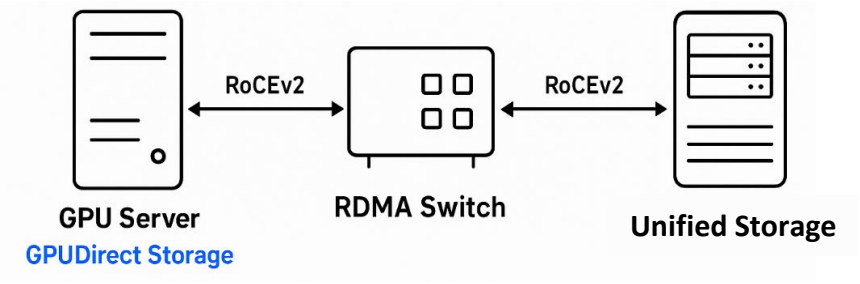


*Figure 2-1 Demonstration Topology*

## 2.2. Configure Steps

### GDS Installation Overview

1. Install NVIDIA drivers and CUDA Toolkit 12.6.
2. Verify GDS installation using

```
> /usr/local/cuda-12.6/gds/tools/gdscheck.py -p
```

3. (Optional) Disable PCIe ACS for best performance.

### RDMA Settings

1. Ensure RDMA network card drivers are properly installed. Install MLX5 driver on Mellanox ConnectX-4 Lx and iRDMA driver on Intel E810.


©2025 QSAN Technology, Inc. All rights reserved.

2. Install rdma-core utility for debugging.
3. Enable PFC (Priority Flow Control) or use global pause control to create a lossless Ethernet environment.
4. Confirm RDMA paths using the following commands.

```
> rdma link show
> rdma statistic show
> ethtool -a <interface>
```

**NFS over RDMA Mount**

1. Mount NFS exported from unified storage shared

```
> sudo mount -t nfs4 -o rdma,vers=4.0 <Unified Storage_IP>:/share/nfs_tmp /mnt/nfs
```

# 2.3. Use Cases

## 2.3.1. Case 1: Streaming Data to GPU for AI Training

**Scenario**

Stream training datasets directly from unified storage to GPU memory using GPUDirect Storage over NFS RDMA.

**Steps**

1. Mount NFS via RDMA share
2. Load dataset from mount path using PyTorch or TensorFlow
3. Verify CPU usage remains low and throughput remains high

**Expected Results**

- Data bypasses CPU memory
- Significantly reduced training startup time


©2025 QSAN Technology, Inc. All rights reserved.

### 2.3.2. Case 2: High Performance Analytics

**Scenario**

Analyze massive datasets (e.g. Parquet, CSV) stored on the unified storage directly through GPU memory.

**Steps**

1. Preload dataset to tmpfs-enabled NFS share
2. Use RAPIDS cuDF/Dask-cuDF for analysis
3. Monitor IO and CPU usage using mpstat, iostat, and iftop

**Expected Results**

- Near real-time analysis with minimal CPU overhead

## 2.4. Performance Evaluation

**Test Tools**

- fio: Standard I/O benchmark
- gdsio: Dedicated GDS performance test tool

**Test Command**

```
> taskset -c 0 gdsio -d 0 -I 0 -i 64K -s 128M -w 4 -x 0 -D /mnt/nfs -T 10
-x 0 = GDS path (storage→external to GPU)
-x 1 = storage→CPU
-x 2 = storage→CPU→GPU
```


©2025 QSAN Technology, Inc. All rights reserved.

**Performance Insights: GPUDirect Storage vs. Traditional Path**

To evaluate the effectiveness of GDS, we compared its performance (transfer type = 0) with traditional CPU-mediated data transfers (transfer type = 2). These two modes represent distinct I/O architectures:

- Transfer Type 0 (GDS)

  Storage → GPU memory via RDMA (zero copy, bypassing the CPU)

- Transfer Type 2 (Legacy)

  Storage → Host memory → GPU memory (copy via PCIe)

By focusing on these two modes, we can accurately assess the benefits of GDS for GPU workloads.

*Table 2-1 GPUDirect Storage vs. Traditional Path Test Result - 1*

**RW Mode: random read**

| Tool | xfer type | Block Size | File Size | Transmission | Workers | Bandwidth (MiB/s) | Avg Latency (us) | Usr CPU Util. | Sys CPU Util. | CPU Utilization (XN5112) |
|---|---|---|---|---|---|---|---|---|---|---|
| gdsio | 0 | 4KiB | 128MiB | single port | 4 | 130 | 117 | 5% (1 CPU) | 25% (1 CPU) | 8% (Total Avg) |
| gdsio | 0 | 64KiB | 128MiB | single port | 4 | 1060 | 230 | 3% (1 CPU) | 22% (1 CPU) | 5% (Total Avg) |
| gdsio | 0 | 64KiB | 128MiB | single port | 8 | 1067 | 457 | 3% (1 CPU) | 21% (1 CPU) | 5% (Total Avg) |
| gdsio | 0 | 64KiB | 128MiB | dual port | 8 (4+4) | 1059+1062 | 230 | 10% (1 CPU) | 45% (1 CPU) | 12% (Total Avg) |
| gdsio | 2 | 4KiB | 128MiB | single port | 4 | 135 | 112 | 25% (1 CPU) | 25% (1 CPU) | 8% (Total Avg) |
| gdsio | 2 | 64KiB | 128MiB | single port | 4 | 1058 | 230 | 15% (1 CPU) | 20% (1 CPU) | 8% (Total Avg) |
| gdsio | 2 | 64KiB | 128MiB | single port | 8 | 1068 | 457 | 13% (1 CPU) | 20% (1 CPU) | 10% (Total Avg) |
| gdsio | 2 | 64KiB | 128MiB | dual port | 8 (4+4) | 1056+1054 | 230 | 39% (1 CPU) | 39% (1 CPU) | 12% (Total Avg) |


©2025 QSAN Technology, Inc. All rights reserved.

*Table 2-2 GPUDirect Storage vs. Traditional Path Test Result - 2e*

**RW Mode: random write**

| Tool | xfer type | Block Size | File Size | Transmission | Workers | Bandwidth (MiB/s) | Avg Latency (us) | Usr CPU Util. | Sys CPU Util. | CPU Utilization (XN5112) |
|---|---|---|---|---|---|---|---|---|---|---|
| gdsio | 0 | 4KiB | 128MiB | single port | 4 | 113 | 137 | 5% (1 CPU) | 20% (1 CPU) | 12% (Total Avg) |
| gdsio | 0 | 64KiB | 128MiB | single port | 4 | 938 | 260 | 3% (1 CPU) | 23% (1 CPU) | 7% (Total Avg) |
| gdsio | 0 | 64KiB | 128MiB | single port | 8 | 1048 | 465 | 3% (1 CPU) | 23% (1 CPU) | 10% (Total Avg) |
| gdsio | 0 | 64KiB | 128MiB | dual port | 8 (4+4) | 922+928 | 263 | 7% (1 CPU) | 45% (1 CPU) | 20% (Total Avg) |
| gdsio | 2 | 4KiB | 128MiB | single port | 4 | 112 | 135 | 20% (1 CPU) | 20% (1 CPU) | 12% (Total Avg) |
| gdsio | 2 | 64KiB | 128MiB | single port | 4 | 863 | 282 | 15% (1 CPU) | 17% (1 CPU) | 8% (Total Avg) |
| gdsio | 2 | 64KiB | 128MiB | single port | 8 | 1038 | 470 | 18% (1 CPU) | 20% (1 CPU) | 10% (Total Avg) |
| gdsio | 2 | 64KiB | 128MiB | dual port | 8 (4+4) | 937+929 | 260 | 38% (1 CPU) | 40% (1 CPU) | 22% (Total Avg) |

**Test Summary**

1. Block size matters: Benchmarks show that a 64 KiB block size provides the highest bandwidth and most consistent latency regardless of transfer type.
2. Reduced CPU Utilization: When using Transfer Type 0 (GDS enabled), user CPU utilization is reduced by approximately 80 ~ 85% compared to Transfer Type 2 (Legacy). This significant reduction frees up host CPU resources for other concurrent operations, thereby improving system efficiency.
3. Throughput remains stable: Despite bypassing host memory, Transfer Type 0 (GDS enabled) still achieves similar or better bandwidth than traditional methods, indicating that it is ready for production use in AI and data analytics workloads.

## 2.5. Troubleshooting

1. **Question:** My GDS installation is complete, but gdscheck.py fails. What should I verify?

   **Answer:**

   - Make sure IOMMU is disabled in the kernel boot parameters.
   - Verify that the NVIDIA driver and CUDA versions match.


©2025 QSAN Technology, Inc. All rights reserved.

- Check that the nvidia-fs kernel module is loaded (lsmod | grep nvidia_fs).
- Verify that your GPU is listed as GDS compatible by NVIDIA.

2. **Question:** NFS over RDMA mount succeeds, but why is performance poor?

   **Answer:**

   - Verify that the mount is using proto=rdma and vers=4.0.
   - Verify that tmpfs is used on the server side to eliminate backend disk I/O.
   - Verify that direct I/O (for example, via fio or gdsio) is enabled during testing.
   - Check CPU utilization, high CPU percentage may indicate fallback to CPU path instead of GDS.

3. **Question:** gdsio is working fine but bandwidth is very low or zero, what could be the reason?

   **Answer:**

   - Make sure the GDS data path is specified using -x 0.
   - Confirm that GPU index (-d) is valid and corresponds to nvidia-smi.
   - Make sure the mount point is an RDMA mounted NFS folder.
   - If security modules (such as AppArmor / SELinux) are interfering with device access, try disabling them.

4. **Question:** RDMA traffic is not detected. How can I confirm that RDMA is working properly?

   **Answer:**

   - Use rdma link show and rdma statistic show to confirm active links.
   - Check ethtool -a <interface> to ensure pause frame (PFC) is enabled.
   - Confirm port=20049 and proto=rdma in the mount output.
   - On Mellanox network cards, run show_gids and mlxconfig to verify RoCE settings.

5. **Question:** CPU usage is high during gdsio testing. Is GDS not working properly?

   **Answer:**

   - Compare xfer_type=0 (GDS) to xfer_type=1 or 2 (CPU path):
   - GDS should show significantly lower CPU usage.
   - If CPU usage is similar for all modes, GDS may not be working properly.
   - Check if GPU memory is used during I/O via nvidia-smi.


©2025 QSAN Technology, Inc. All rights reserved.

6. **Question:** RDMA mounts fail with NFS v4.1 or v4.2. Why?

   **Answer:**

   - GDS currently only works reliably on NFS v3 and v4.0.
   - NFS v4.1 and v4.2 introduced protocol changes that are incompatible with RDMA on most Linux distributions.
   - Explicitly use -o vers=4.0,proto=rdma during mounts.

7. **Question:** I see "nfs: server not responding" errors during high load testing?

   **Answer:**

   - Check for network congestion or missing PFC configuration.
   - Verify that the RDMA switch supports lossless Ethernet (DCQCN / PFC).
   - Monitor bandwidth, disk, and network card health using tools such as iftop, ibstat, and iostat.

8. **Question:** Why does RDMA / NFS mount not work automatically after a system reboot?

   **Answer:**

   - The RDMA module may not load automatically. Add the rdma_ucm, ib_ipoib, and driver modules to /etc/modules.
   - If using systemd mounts, make sure network-online.target is set as a dependency.

9. **Question:** When using tmpfs, gdsio reports O_DIRECT errors, what is the reason?

   **Answer:**

   - By default, tmpfs does not support O_DIRECT.
   - Either
     Use fio without direct=1, or
     Test with RAM-backed block devices (e.g. zram, ramfs), or
     Accept buffered I/O for comprehensive performance measurements.


©2025 QSAN Technology, Inc. All rights reserved.

# 3. CONCLUSION

The integration of GDS (GPUDirect Storage) with QSAN unified storage solutions provides a powerful architecture tailored for data-intensive workloads such as AI training, real-time analytics, and high-performance computing. By enabling direct data transfer between storage and GPU memory via RDMA, GDS effectively eliminates CPU bottlenecks, reduces latency, and increases data throughput.

**Key Takeaways**

- **Seamless Compatibility:** QSAN unified storage supports NFS over RDMA, making it easy to deploy GDS with minimal configuration overhead.
- **Performance Efficiency:** The data path directly accessing GPU memory reduces CPU load, increases training speed, and improves GPU utilization.
- **Proven Architecture:** Real-world test results confirm the robustness of this solution, with GDS consistently delivering higher throughput and lower latency compared to traditional I/O paths.
- **Flexible Deployment:** This architecture supports a variety of use cases, from model training to large-scale data analytics, while remaining scalable and cost-effective.

Enterprises that adopt GDS and QSAN unified storage are able to gain competitive advantage by accelerating data pipelines and optimizing infrastructure performance. The solution not only meets today's demand for AI scalability and speed, but also lays the foundation for future growth in GPU-accelerated environments. By reducing data movement overhead and unlocking the full potential of GPUs, enterprises can streamline innovation and shorten time to insight for mission-critical applications.


©2025 QSAN Technology, Inc. All rights reserved.

# 4. APPENDIX

## 4.1. Apply To

- QSM firmware 4.1.5 and later

## 4.2. Reference

Product Page

- XCubeNXT 5100 Series

Document

- QSM 4 Software Manual


©2025 QSAN Technology, Inc. All rights reserved.