Home > Storage architecture and strategy

DEFINITION

Remote Direct Memory Access (RDMA)

Robert Sheldon
James Alan Miller, Senior Executive Editor

What is Remote Direct Memory Access (RDMA)?

Remote Direct Memory Access is a technology that enables two networked computers to exchange data in main memory without relying on the processor, cache or operating system of either computer. Like locally based Direct Memory Access (DMA), RDMA improves throughput and performance because it frees up resources, resulting in faster data transfer rates and lower latency between RDMA-enabled systems. RDMA can benefit both networking and storage applications.

RDMA facilitates more direct and efficient data movement into and out of a server by implementing a transport protocol in the network interface card (NIC) located on each communicating device. For example, two networked computers can each be configured with a NIC that supports the RDMA over Converged Ethernet (RoCE) protocol, enabling the computers to carry out RoCE-based communications.

Integral to RDMA is the concept of zero-copy networking, which makes it possible to read data directly from the main memory of one computer and write that data directly to the main memory of another computer. RDMA data transfers bypass the kernel networking stack in both computers, improving network performance. As a result, the conversation between the two systems will complete much quicker than comparable non-RDMA networked systems.

At left is a standard network connection. At right is an RDMA connection. The initiator and the target must use the same type of RDMA technology -- RDMA over Converged Ethernet or InfiniBand, for example.


RDMA has proven useful in applications that require fast and massive parallel high-performance computing (HPC) clusters and data center networks. It is particularly useful when analyzing big data, in supercomputing environments that process applications, and for machine learning that requires low latencies and high transfer rates. RDMA is also used between nodes in compute clusters and with latency-sensitive database workloads.

Network protocols that support RDMA

An RDMA-enabled NIC must be installed on each device that participates in RDMA communications. Today's RDMA-enabled NICs typically support one or more of the following three network protocols:

- **RDMA over Converged Ethernet.** RoCE is a network protocol that enables RDMA communications over an [Ethernet](#) The latest version of the protocol -- RoCEv2 -- runs on top of User Datagram Protocol ([UDP](#)) and Internet Protocol ([IP](#)), versions 4 and 6. Unlike RoCEv1, RoCEv2 is routable, which makes it more scalable. RoCEv2 is currently the most popular protocol for implementing RDMA, with wide adoption and support.

- **Internet Wide Area RDMA Protocol.** iWARP leverages the Transmission Control Protocol ([TCP](#)) or Stream Control Transmission Protocol ([SCTP](#)) to transmit data. The Internet Engineering Task Force developed iWARP so applications on a server could read or write directly to applications running on another server without requiring OS support on either server.

- **InfiniBand.** [InfiniBand](#) provides native support for RDMA, which is the standard protocol for high-speed InfiniBand network connections. InfiniBand RDMA is often used for intersystem communication and was first popular in HPC environments. Because of its ability to speedily connect large computer clusters, InfiniBand has found its way into additional use cases such as [big data](#) environments, large [transactional](#) databases, highly virtualized settings and resource-demanding web applications.

Comparing traditional server-to-server (left) to RoCE server-to-server communications (right).
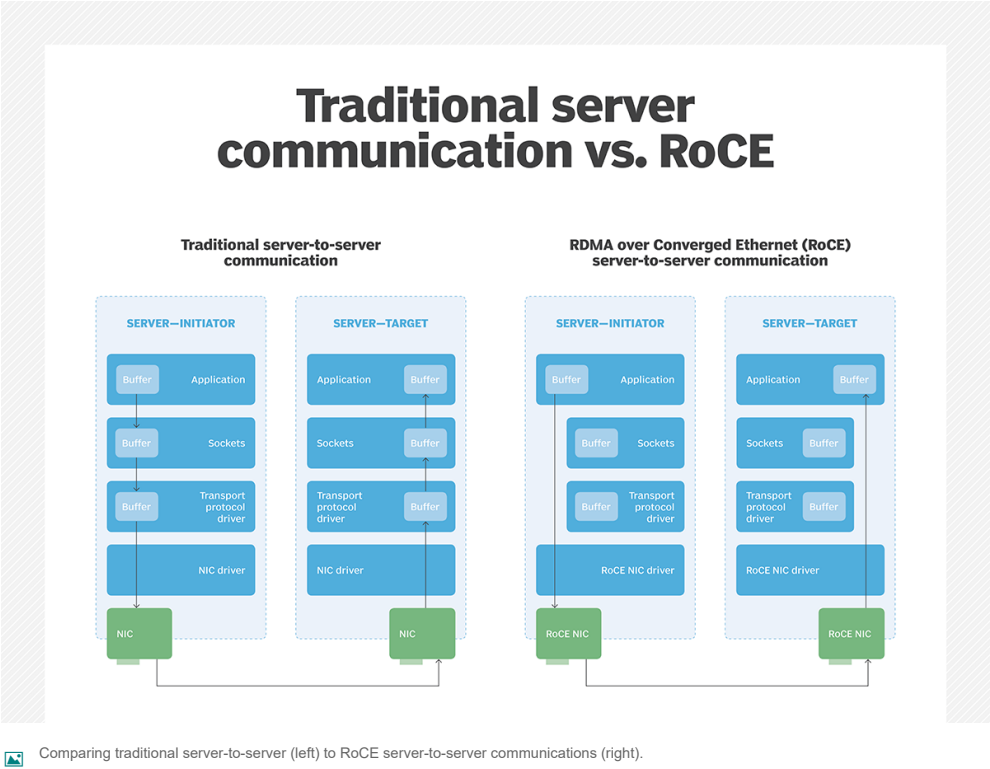
Products and vendors that support RDMA

A number of products and vendors support RDMA, among them:

- Apache Hadoop and Apache Spark big data analysis
- Broadcom and Emulex adapters
- Ceph object storage platform
- ChainerMN Python-based deep learning open source framework
- Chelsio Terminator 6 iWARP adapters
- Dell EMC PowerEdge servers
- FreeBSD operating system
- GlusterFS distributed files ystem
- Intel Xeon Scalable processors and Platform Controller Hub
- Marvell FastLinQ 45000 Series Ethernet NICs
- Nvidia ConnectX network adapters and InfiniBand switches
- Microsoft Windows Server (2012 and higher)
- Nvidia DGX deep learning systems
- Oracle Solaris and NFS over RDMA
- Red Hat Enterprise Linux
- SUSE Linux Enterprise Server
- TensorFlow open source software library for machine intelligence
- Torch scientific computing framework
- VMware ESXi

RDMA with flash-based SSDs and NVDIMMs

All-flash storage systems perform much faster than disk or hybrid arrays, resulting in significantly higher throughput and lower latency. However, a traditional software stack often can't keep up with flash storage and starts to act as a bottleneck, increasing overall latency. RDMA can help address this issue by improving the performance of network communications.

RDMA can also be used with non-volatile dual in-line memory modules (NVDIMMs). An NVDIMM device is a type of memory that acts like storage but provides memory-like speeds. For example, NVDIMM can improve database performance by as much as 100 times. It can also benefit virtual clusters and accelerate virtual storage area networks (VSANs).

What is a VSAN (Virtual Storage Area Network)?

| Storage

To get the most out of NVDIMM, organizations should use the fastest network possible when transmitting data between servers or throughout a virtual cluster. This is important in terms of both data integrity and performance. RDMA over Converged Ethernet can be a good fit in this scenario because it moves data directly between NVDIMM modules with little system overhead and low latency.
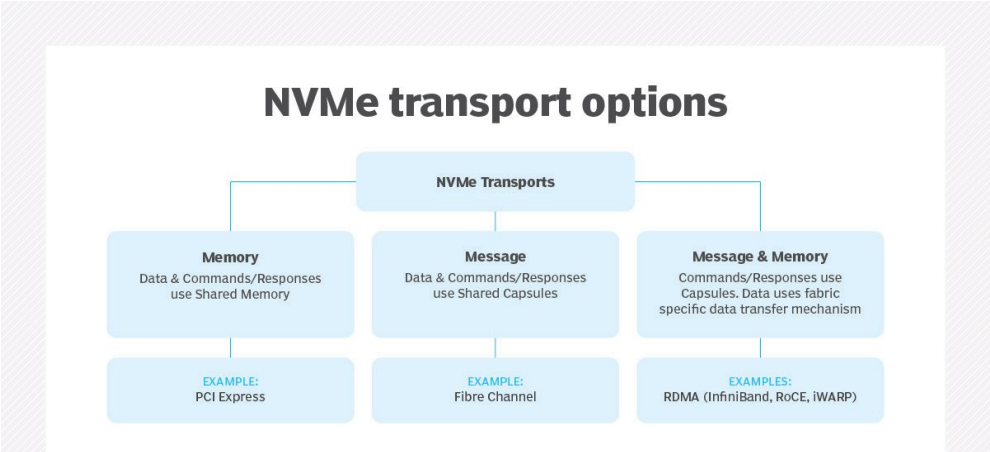
NVDIMM Types

NVDIMM-N	NVDIMM-F	NVDIMM-P
Standardized	**Vendor specific**	**Proposals in progress**
▪ Memory mapped DRAM. Flash is not system mapped. ▪ Access Methods -> byte- or block-oriented access to DRAM ▪ Capacity = DRAM DIMM (1's-10's GB) ▪ Latency = DRAM (10's of nanoseconds) ▪ Energy source for backup ▪ DIMM interface (HW & SW) defined by JEDEC	▪ Memory mapped Flash. DRAM is not system mapped. ▪ Access Method -> block-oriented access to NAND through a shared command buffer (i.e. a mounted drive) ▪ Capacity = NAND (100's GB-I's TB) ▪ Latency = NAND (10's of microseconds)	▪ Memory-mapped Flash and memory-mapped DRAM ▪ Two access mechanisms: persistent DRAM (–N) and block-oriented drive access (–F) ▪ Capacity = NVM (100's GB-I's TB) ▪ Latency = NVM (100's of nanoseconds)

RDMA is compatible with non-volatile dual in-line memory modules, which is a kind of memory -- with memory-like speeds -- that acts like storage.

NVM Express over Fabrics using RDMA

Organizations are increasingly storing their data on flash-based solid-state drives (SSDs). When that data is shared over a network, RDMA can help increase data-access performance, especially when used in conjunction with NVM Express over Fabrics (NVMe-oF).

The NVM Express organization published the first NVMe-oF specification on June 5, 2016, and has since revised it several times. The specification defines a common architecture for extending the NVMe protocol over a network fabric. Prior to NVMe-oF, the protocol was limited to devices that connected directly to a computer's PCI Express (PCIe) slots.

NVMe transport options

NVMe Transports

Memory	Message	Message & Memory
Data & Commands/Responses use Shared Memory	Data & Commands/Responses use Shared Capsules	Commands/Responses use Capsules. Data uses fabric specific data transfer mechanism
EXAMPLE: PCI Express	**EXAMPLE:** Fibre Channel	**EXAMPLES:** RDMA (InfiniBand, RoCE, iWARP)

RDMA is a transport -- via InfiniBand, RoCE and iWARP -- is transport option for NVMe fabrics.

The NVMe-oF specification supports multiple network transports, including RDMA. NVMe-oF with RDMA makes it possible for organizations to take fuller advantage of their NVMe storage devices when connecting over Ethernet or InfiniBand networks, resulting in faster performance and lower latency.

What is NVMe-oF (NVMe Over Fabrics)?

▶

This was last updated in November 2021

↘ **Continue Reading About Remote Direct Memory Access (RDMA)**

■ How to reduce latency issues with Remote Direct Memory Access

■ The benefits of using RDMA with virtualization

■ High-performance interconnects and storage performance

■ Introduction to RDMA

■ How RDMA technology works

Related Terms

Electronic Discovery Reference Model (EDRM)

The Electronic Discovery Reference Model (EDRM) is a conceptual framework that outlines activities for the recovery and discovery... See complete definition ℹ

hot plugging

Hot plugging is the addition of a component to a running computer system without significant interruption to the operation of the... See complete definition ℹ

NVMe over Fabrics (NVMe-oF)

NVMe over Fabrics, also known as NVMe-oF and non-volatile memory express over fabrics, is a protocol specification designed to ... See complete definition ℹ

↘ **Dig Deeper on Storage architecture and strategy**

NVMe over Fabrics (NVMe-oF)

8/31/23, 1:41 PM

What Is Remote Direct Memory Access (RDMA) Networking and How Does It Work?

By: Alexander Gillis

StorOne adds NVMe over Fabrics

By: Adam Armstrong

NVMe-over-fabrics: Five need-to-knows

By: Antony Adshead

NVMe-over Fabrics: How NVMe-oF revolutionises shared storage

By: Stephen Pritchard

DISASTER RECOVERY DATA BACKUP DATA CENTER SUSTAINABILITY AND ESG

Disaster Recovery

What are 5 good reasons to do yearly disaster recovery testing?

Do you think yearly disaster recovery testing is overkill? You're not alone, but you are missing out on some key ways DR testing ...

6 ways to support business resilience at your organization

Without enough support, resilience initiatives are unlikely to succeed. Here are six ways individuals and departments can support...

About Us   Editorial Ethics Policy   Meet The Editors   Contact Us   Advertisers   Partner with Us   Media Kit   Corporate Site

Contributors   Reprints   Answers   Definitions   E-Products   Events   Features

Guides   Opinions   Photo Stories   Quizzes   Tips   Tutorials   Videos