Overview of RDMA, InfiniBand, and RoCE Networks

Antonio Perez

July 30, 2025

Introduction

RDMA (Remote Direct Memory Access) is a network technology that allows a device to access the memory of another device directly, without involving the CPU, operating system, or network stack of the remote host. This leads to extremely low latency, high throughput, and minimal CPU utilization.

1. What is RDMA?

- RDMA stands for *Remote Direct Memory Access*.
- Enables zero-copy data transfers directly between system memories across nodes.
- Offloads networking work from the CPU to the NIC.
- Reduces both latency and CPU overhead.
- Common in high-performance computing (HPC), database clusters, and virtualization environments.

2. What is InfiniBand?

- InfiniBand is a high-speed, low-latency interconnect architecture.
- Designed specifically for HPC clusters and data-intensive environments.
- Uses RDMA natively, independent of TCP/IP.
- Requires specialized hardware such as InfiniBand switches and Host Channel Adapters (HCAs).
- Supports link speeds from FDR (56 Gbps) to HDR (200 Gbps) and beyond.

3. What is RoCE?

RoCE (RDMA over Converged Ethernet) brings RDMA capabilities over standard Ethernet infrastructure.

- **RoCE v1:** Layer 2 implementation (not routable).
- **RoCE v2:** Works over UDP/IP at Layer 3 (routable across subnets).
- Allows RDMA performance without replacing Ethernet with InfiniBand.
- Requires Data Center Bridging (DCB) for lossless Ethernet transport.

Feature	InfiniBand	RoCE v2
Network Type	Proprietary	Standard Ethernet
Latency	Ultra-low	Low
RDMA Stack	Native	Over IP/UDP
Routing	Native	IP Routable
Hardware Requirement	InfiniBand adapters	Ethernet NICs with RDMA
Cost	High	Lower
Common Use Cases	HPC, AI, ML	Data centers, NVMe-oF

Table 1: Comparison of InfiniBand and RoCE v2

4. Comparison Table

Conclusion

RDMA-based networks like InfiniBand and RoCE offer advanced performance for demanding applications, significantly improving latency and throughput. Choosing between InfiniBand and RoCE depends on infrastructure cost, existing Ethernet deployments, and latency requirements.