# StarWind NVMe-oF Initiator

Introduction

More and more companies, including small and medium business (SMB), remote offices/branch offices (ROBO), and those that use Edge devices, turn to flash drives. This trend demonstrates how fast things change: some time ago, SSDs were a premium storage medium only large datacenters could use. Something similar may happen to PCIe SSDs someday as flash becomes increasingly affordable. The only problem with these drives aside from their price is lack of means for working with them. StarWind NVMe-oF Initiator is the only free software solution that after being deployed on the client side allows effectively working with underlying PCIe storage.

Problem

Locally, PCIe SSDs work like clockwork; problems show up once these drives are presented over the network. The traditional protocols – iSCSI, iSER, SMB3, and NFS – are inefficient for flash. Built to connect spindle drives, they are stuck in times when there were no drives faster than HDDs in mass usage. Their single short command queue does not work well for flash. Moneywise, presenting PCIe SSDs via those protocols means the low return on investment since users do not get the performance that they paid for: PCIe SSDs perform only 20% better than SAS HDDs, meaning that the performance vendors showcase is never seen in real setups.
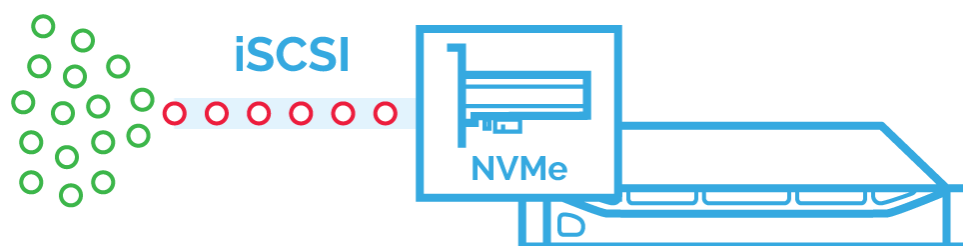
Serial Attached SCSI (SAS) – single short command queue is a performance bottleneck

Another problem with those protocols is CPU utilization. As a result, there's less compute resource left for applications.

Things like that are critical for SMB, ROBO, and Edge, the companies that are on snag IT budgets. Every time they invest in something, they typically try to squeeze most of that thing features. They try to utilize the hardware to the limit in order to deliver their customers the best service for a moderate price. Not the case for PCIe drives: speed gain is not commensurate with the investment.

Any alternatives to traditional protocols? Yes, NVMe over Fabrics (NVMe-oF) introduced in 2016 – the protocol for presenting PCIe SSDs over the network. This standard implies that you get all the IOPS from the datasheet while latency gain is less than 10ms. It can use connections like Fiber Channel (FC) or RDMA. Some serious work is done on NVMe over TCP too.

The only problem with NVMe-oF is that only very few OS support it. To use this protocol, you need a target and initiator. You can create a target with Intel SPDK Target or use Linux NVMe-oF Target but what about the initiator? And, here problems start as there are no software initiators these days.

If you are into storage industry, you know that [Chelsio](#) and [Mellanox](#) managed to create the network adapters that support NVMe-oF. The initiator is supplied with [hardware solutions](#), which means that you need to buy new hardware right after you boost servers with PCIe drives. That's overkill! And, it is the reason why StarWind NVMe-oF Initiator was invented.

Solution

StarWind introduces StarWind NVMe-oF Initiator – a free software solution that makes any client NVMe-oF-compatible. It is the only software NVMe-oF initiator available these days. The solution is implemented as a software module that connects a host (i.e., a client) to a block storage target (i.e., remote server). The solution [works](#) with both [Intel SPDK NVMe-oF Target](#) and Linux NVMe-oF Target – so we do not lock users in to StarWind solutions after all.

To understand StarWind NVMe-oF Initiator value, let's first discuss the inherent features of NVMe-oF. NVMe-oF enables multiple command queues that are independent of each other. For the best performance, their total number equals the number of physical CPU cores (by design, there can be up to 64K command queues). Each command queue, in turn, supports up to 64K commands (there are typically less than 512 commands per queue in real life).


NVMe-oF – performance is not bottlenecked by fabrics anymore

StarWind NVMe-oF Initiator talks to underlying storage via RDMA (RoCE). Working over RoCE, StarWind NVMe-oF implementation enables to minimize CPU participation in data transfer, freeing additional compute resource for your applications. The nice thing is that thanks to RoCE remote direct memory access is possible via Ethernet; the only thing you may still need to purchase are NICs that support

RDMA. RDMA supports zero-copy networking by enabling network adapters to transfer data from the wire directly to application memory and other way around. There's no need to copy data between application memory and data buffers, meaning that CPUs, caches, and context switches do not participate in data transfers. As a result, NVMe-oF dramatically reduces the latency of devices presented over the network.

Now, let's talk about how we brought NVMe-oF support to Windows. StarWind NVMe-oF Initiator consists of Kernel and NIC components. Installed on the client side, the solution can be connected to multiple namespaces (maximum 64K). Virtual devices that emerge on the controller represent the PCIe SSDs on remote namespaces. The controller is associated with one client at a time, whereas the port may be shared – NVMe allows hosts to connect to multiple controllers in the NVM subsystem through the same port or different ports.
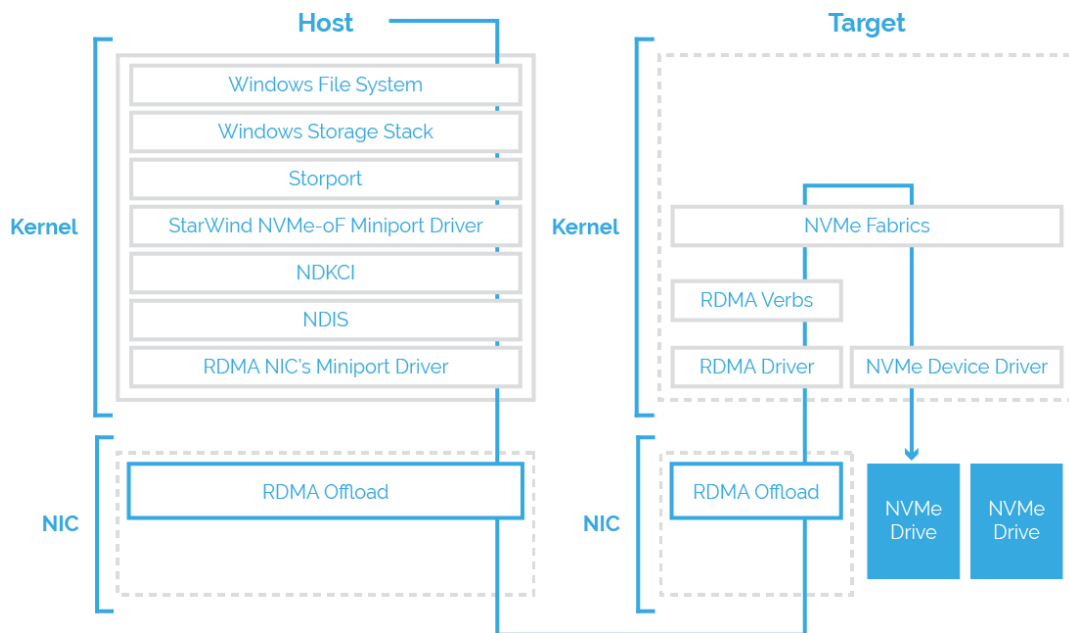
StarWind NVMe-oF Initiator implementation

How successful is StarWind NVMe-oF Initiator? See how we managed to get most of the underlying all-NVMe storage, [beating HCI performance records](). The software is free anyway, so why don't you just try it out for PoC?

Conclusion

StarWind NVMe-oF Initiator is a free solution allowing applications to work effectively with PCIe drives. This solution is the only software initiator these days. It can work with such targets as Intel SPDK NVMe-oF Target and Linux NVMe-oF Target.

To learn more, visit **www.starwind.com**