StarWind NVMe over Fabrics (NVMe-oF)

Introduction

NVMe is one of the hottest topics in the world of storage these days. Expectations for this technology are so high that 2019 is sometimes called a year of NVMe. But when it comes to NVMe over Fabrics (NVMe-oF), we can say that it only takes its first faltering steps. During 2017, it was final approval of NVMe-oF standard when its support was added to the Linux kernel and a number of these OS distributions. But up to now, this standard has not been added to work with Microsoft Windows Server. For this very reason, StarWind takes the next step by adding NVMe-oF to StarWind Virtual SAN (VSAN) for reliable transfer of data, requests, and responses between the NVMe host and its subsystem in a Windows Server environment and for support of iSCSI, iSER, and so on.

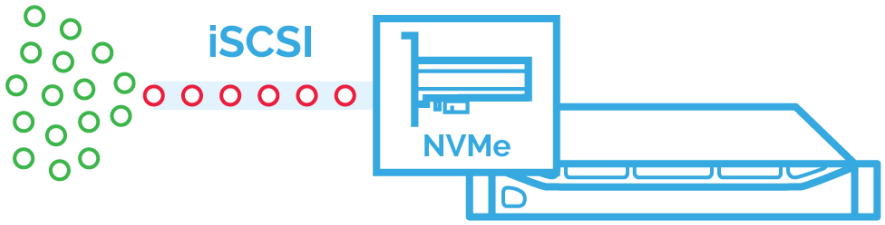
Problem

Nowadays, as flash memory becomes increasingly prevalent, adding NVMe drives to a SAN seems a really good idea, especially, if you run some IOPS-hungry applications in it. Unfortunately, even with those drives on board, the resulting infrastructure performance will still fall short from satisfactory because PCIe SSDs cannot be presented efficiently over the network. Wondering why most environments just cannot access a good part of PCIe SSD performance over the network? The answer is quite straightforward: traditional protocols.

By adding NVMe to your network fabric, the commonly used protocols like iSCSI and FC could not provide the desired effective communication between the NVMe host computer and a block-level storage device. They were designed to talk to cold storage media disks, not flash. Their single short command queue limits NVMe drive I/O so badly that applications do not get a good part of the underlying storage performance. In addition, they lead to high latency which makes a world of difference between local and remote storage. With a single short command queue, achieving a high level of parallelism for PCIe SSDs is out of the question.

Using traditional protocols like iSCSI, another problem also comes in, namely, the additional load on the server CPU. Instead of server processor cycles work one hundred percent to process applications, they are burdened with iSCSI and TCP/IP Stack processing.

Of course, you can use ISER to provide higher bandwidth for block storage transfers and lower CPU utilization. But you won't get the full performance for your SSDs with the remaining problem of single-queue iSCSI model.

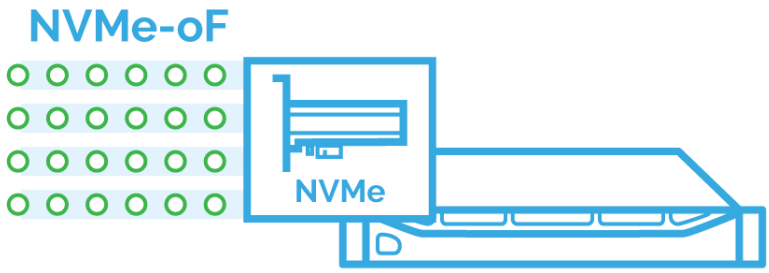


Serial Attached SCSI (SAS) – Single short command queue is a performance bottleneck

Solution

Are there any alternatives to iSCSI-derived protocols? Yes, there is one tailored to achieve the peak performance of your network fabric – StarWind NVMe-oF. It's used for communication between the NVMe host computer and a block-level storage device. Unlike its closest analogs, such as FC and iSCSI, NVM-oF provides much less latency compared to several microseconds, making the difference between local and remote storage almost imperceptible. The single short command queue typical for traditional protocols is replaced with 64 thousand command queues, 64 thousand commands each. You get a high level of parallelism in multicore processors when all I/O commands and further responses occur on the same processor core. Such design enables to reduce latency remarkably while NVMe SSDs are presented over the network, allowing to get all the IOPS that they can provide. Flash I/O bottleneck is eliminated, while iSCSI reintroduces into the I/O path.

StarWind NVMe-oF also deals with an issue of server CPU overload. NVMe-oF uses RDMA to transfer data over the network with RDMA over Converged Ethernet (RoCE) technology. By means of it, computers on the network can exchange data in the main memory without CPU, cache, and OS involvement. By excluding the CPU from the data movement process, NVMe-oF simultaneously reduces the latency and increases the performance and efficiency of the data transmission speed.



NVMe-oF – Networking is not a performance bottleneck anymore

Conclusion

When it comes to underlying storage performance, NVMe is the true king of the hill. However, it is still difficult to present PCIe SSDs over the network effectively since the well-known SCSI-based protocols do not work that good for flash drives. StarWind's implementation of NVMe-oF allows solving the issues of traditional protocol overlay, lack of parallelism support, and ineffective CPU usage. And the icing on the cake: a device can deliver up to 100% IOPS locally — get this number with StarWind NVMe-oF over the network.

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