

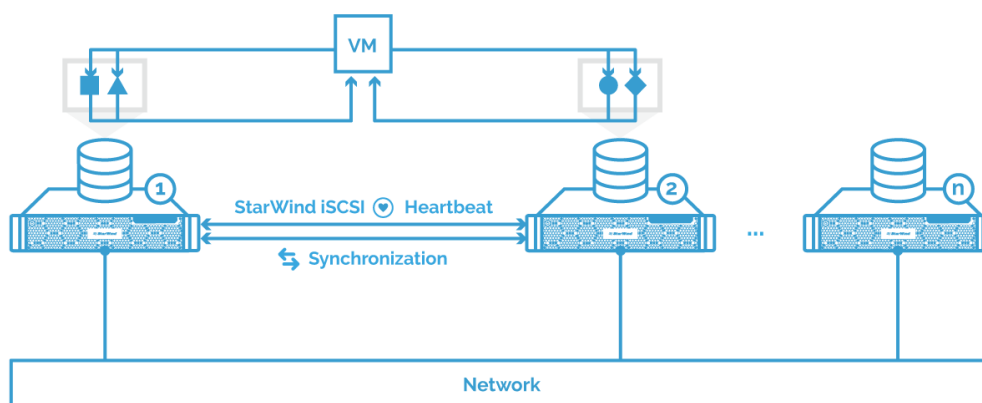
# Data Locality

## Introduction

Maintaining the balance of IT infrastructure components' throughput is crucial, so making sure that I/O occurs smoothly is vital for creating a powerful and stable environment. To build such an infrastructure, it is important to avoid any bottlenecks – slow-functioning components that bring down the performance of the entire system; no matter which hardware you use. Even a high-end-hardware-packed system will be brought down by one slow-functioning component called the “bottleneck”. I/O path may be a bottleneck since interconnection fabrics may be slower than storage. StarWind Virtual SAN features data locality – the principle which tunes the I/O path so that VMs access data in the fastest way possible.

## Problem of Unoptimized Data Access

When a VM migrates to another node within a highly available cluster (say, for load balancing concerns), it starts using the compute resources of that host while all its data may still be read and written from/to a remote (“old”) host. As a result, VMs have their performance degraded due to an unoptimized I/O path because data must be transferred over the network all the time. Networking, in its turn, may be much slower than high-speed local storage on that node, like DRAM or flash. So, no matter which formidable data processing characteristics a system has, it could be bottlenecked by interconnecting wires. As a result, the cost-efficiency of any environment, even “pumped” with a flash one, may degrade, leading to an unsatisfactory cost efficiency of the storage.

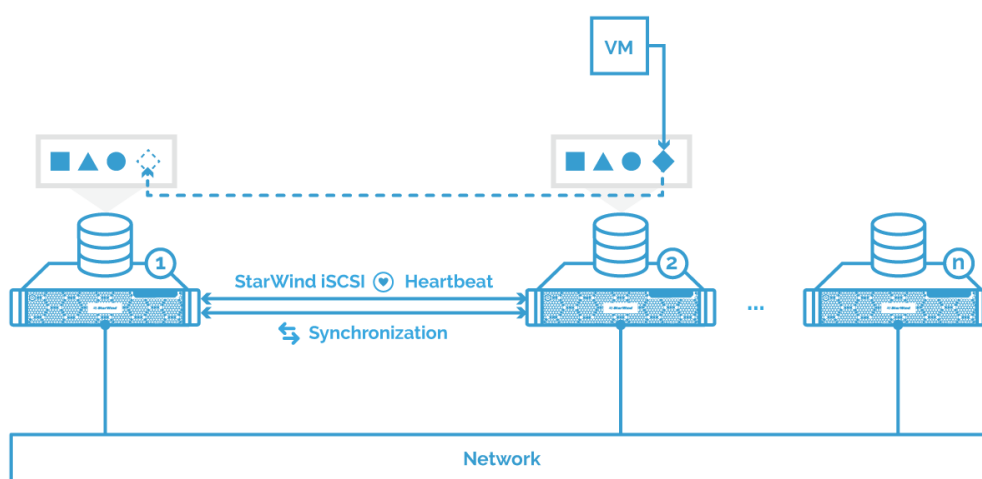


*Read & Write path: Distributing VM data across N nodes*

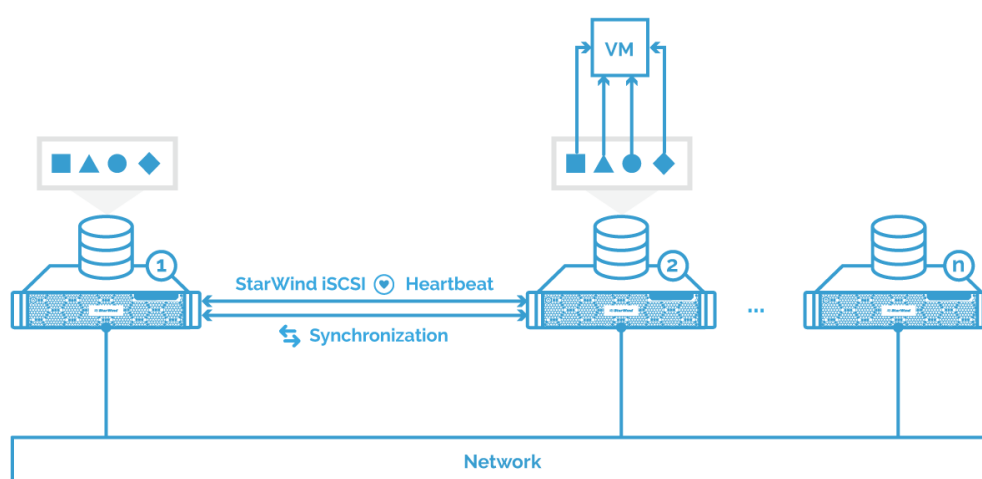
## Data Locality is a Solution

Keeping a copy of data locally and marking a local storage path as the optimal one may be a perfect solution to increase VM performance. Such an approach eliminates the latency induced by networking stack as VMs basically do not use wires to access their data anymore.

StarWind VSAN enables to synchronously replicate data between the hosts so that each one contains the most important data sets to spin up a production VM there; there's no need to read data from the neighbor hosts. With Asymmetric Logical Unit Access (ALUA) configuration, the local I/O path can be marked as a preferred one, which enables StarWind VSAN to sufficiently optimize logical unit access for each VM.



*VM data local: Write path*



*VM data local: Read path*

Latency is lower and performance is faster in the local I/O scenario.

## Conclusion

“Data locality” is a great approach to address the “bottleneck” problem of data having to travel through slow networks. Data do not need to circulate between a VM and remote hosts all the time as necessary data sets are available locally. Also, it keeps the I/O operations within a single physical node. Thus, “data locality” allows to avoid the processing overhead of the network stack, thereby considerably lowering latency because much fewer data require transferring via network.