

Scale-Up and Scale-Out

Introduction

Storage capacity is becoming a real challenge for IT departments because the amount of data most companies use is growing faster than expected. Choosing the right way to scale IT storage infrastructure is vital to building a balanced, robust and cost-effective production environment.

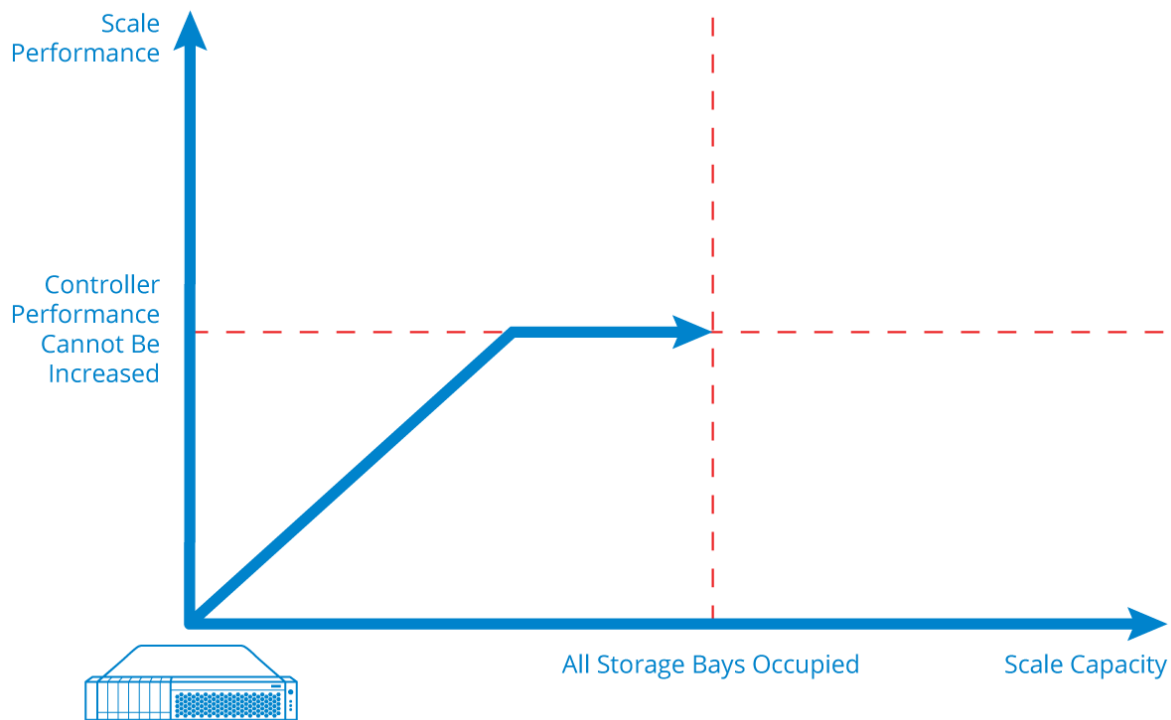
Problem

Scale-up is one approach to increasing the capacity or performance of a storage system or server.

Despite seeming simple it presents a number of problems:

- There are constraints to physical system upgrades: once all disk, CPU, PCI, and RAM slots are occupied it's impossible to improve performance by adding components.
- There will always come a point when a forklift upgrade is needed. Old hardware and parts have to be reused or sold, otherwise the TCO will go through the roof.
- Scaling up a small number of systems increases the performance impact on the configuration in case of a failure. In an environment relying on a few big systems, a failover causes a significant impact on the configuration: imagine hundreds of VMs and applications simultaneously starting on a system that is already running the same number of tasks.

This produces a powerful IO hit on the storage subsystem and often results in worse performance or application crashes due to slow IO. To avoid this, an adequate power reserve has to be maintained, however this is inefficient and causes a dramatic increase in TCO since the power reserve is mostly idle and can't be used for applications or VMs.

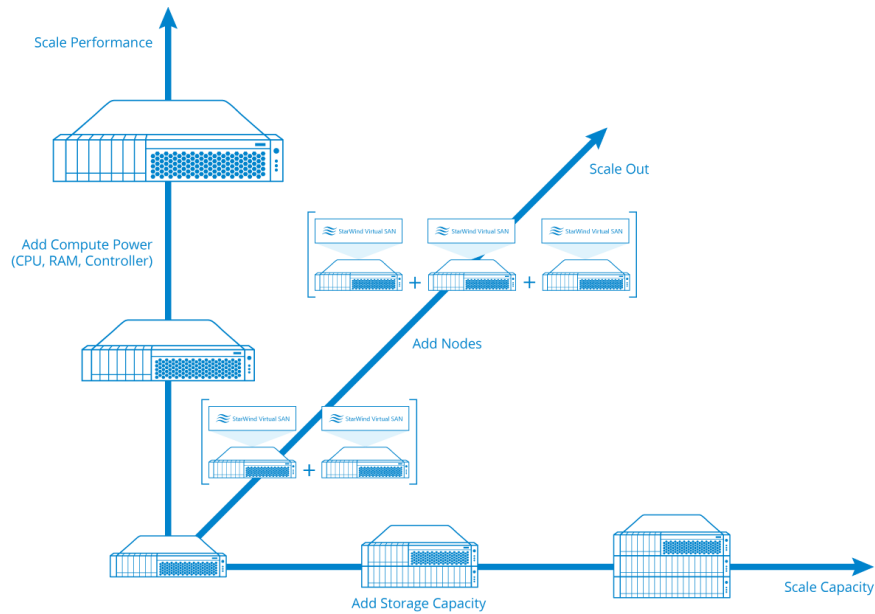


Scaling-Up has physical limits

Updating individual systems with more CPU, RAM and disks has been a typical approach for the last two decades. Servers mostly ran single applications so each server had to be upgraded. This caused many inconvenient issues like after-hours upgrades and significant amounts of downtime. Applying this model to a modern virtualized data center means we'd be powering off an entire infrastructure since one server or storage system could be hosting most of a company's applications and data. With today's business pace and uptime requirements it may be simply unacceptable for a business to go offline even for 5 minutes.

Solution

Scale-out by StarWind Virtual SAN is the approach with the modern virtualization trend in mind. It is a much more efficient and cost-effective alternative to scale-up. The idea behind scale-out is to grow both storage and compute power by adding additional nodes instead of adding disks, CPUs, NICs or RAM to individual systems. Increasing the number of nodes increases reliability by eliminating single points of failure. It also gives customers a more flexible computing environment by treating all servers as a unified compute and storage resource pool. Adding nodes with the same hardware configuration isn't necessary. Customers can add performance-tuned or capacity-tuned nodes depending on their IT environment requirements.



Scaling out by adding nodes

Conclusion

Implementing scale-out approach versus a scale up approach gives you a robust, well-balanced and easily scalable configuration. The scale-out approach gives IT administrators the following significant advantages:

- Managing a scale-out cluster is very easy and requires no downtime.
- The IT infrastructure is easily upgraded by adding nodes. Old hardware can be taken out of use without downtime.
- The performance impact in case of a failure is much lower compared to the scale-up approach. Should any node fail, all VMs and applications failover and re-balance between multiple nodes of the scale-out cluster thus, it doesn't cause an IO hit on the production environment.
- Implementing the scale-out approach will lower your OpEx and TCO because the configuration doesn't require time-consuming forklift upgrades or after-hours maintenance.

In 2016, Gartner named StarWind “Cool Vendor for Compute Platforms”.

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.