

Book of vSphere - How vSphere on Nutanix Works

[PDF generated August 17 2023. For all recent updates please see the Nutanix Bible releases notes located at https://nutanixbible.com/release_notes.html. Disclaimer: Downloaded PDFs may not always contain the latest information.]

Array Offloads – VAAI

The Nutanix platform supports the VMware APIs for Array Integration (VAAI), which allows the hypervisor to offload certain tasks to the array. This is much more efficient as the hypervisor doesn't need to be the 'man in the middle'. Nutanix currently supports the VAAI primitives for NAS, including the 'full file clone', 'fast file clone', and 'reserve space' primitives. Here's a good article explaining the various primitives: <http://cormachogan.com/2012/11/08/vaa-comparison-block-versus-nas/>.

For both the full and fast file clones, a DSF 'fast clone' is done, meaning a writable snapshot (using re-direct on write) for each clone that is created. Each of these clones has its own block map, meaning that chain depth isn't anything to worry about. The following will determine whether or not VAAI will be used for specific scenarios:

- Clone VM with Snapshot → VAAI will NOT be used
- Clone VM without Snapshot which is Powered Off → VAAI WILL be used
- Clone VM to a different Datastore/Container → VAAI will NOT be used
- Clone VM which is Powered On → VAAI will NOT be used

These scenarios apply to VMware View:

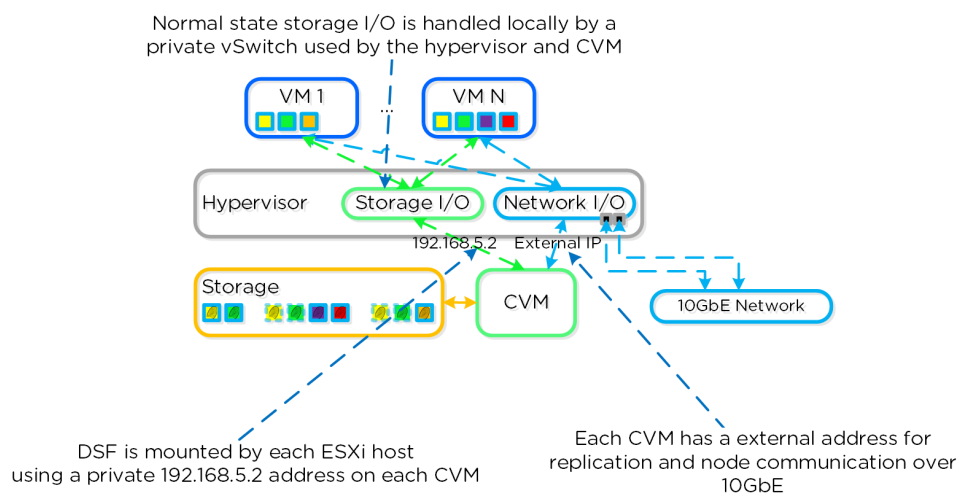
- View Full Clone (Template with Snapshot) → VAAI will NOT be used
- View Full Clone (Template w/o Snapshot) → VAAI WILL be used
- View Linked Clone (VCAI) → VAAI WILL be used

You can validate VAAI operations are taking place by using the 'NFS Adapter' Activity Traces page.

CVM Autopathing aka Ha.py

In this section, I'll cover how CVM 'failures' are handled (I'll cover how we handle component failures in future update). A CVM 'failure' could include a user powering down the CVM, a CVM rolling upgrade, or any event which might bring down the CVM. DSF has a feature called autopathing where when a local CVM becomes unavailable, the I/Os are then transparently handled by other CVMs in the cluster. The hypervisor and CVM communicate using a private 192.168.5.0 network on a dedicated vSwitch (more on this above). This means that for all storage I/Os, these are happening to the internal IP addresses on the CVM (192.168.5.2). The external IP address of the CVM is used for remote replication and for CVM communication.

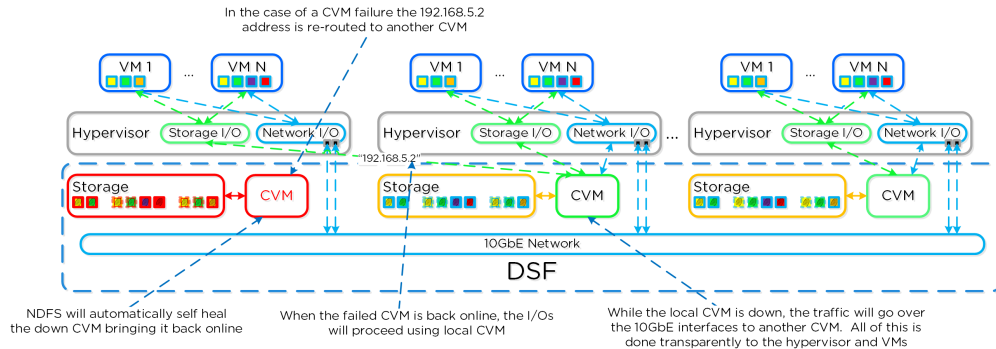
The following figure shows an example of what this looks like:



ESXi Host Networking

In the event of a local CVM failure, the local 192.168.5.2 addresses previously hosted by the local CVM are unavailable. DSF will automatically detect this outage and will redirect these I/Os to another CVM in the cluster over 10GbE. The re-routing is done transparently to the hypervisor and VMs running on the host. This means that even if a CVM is powered down, the VMs will still continue to be able to perform I/Os to DSF. Once the local CVM is back up and available, traffic will then seamlessly be transferred back and served by the local CVM.

The following figure shows a graphical representation of how this looks for a failed CVM:



ESXi Host Networking - Local CVM Down