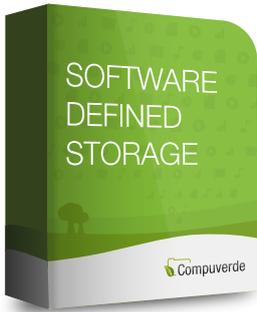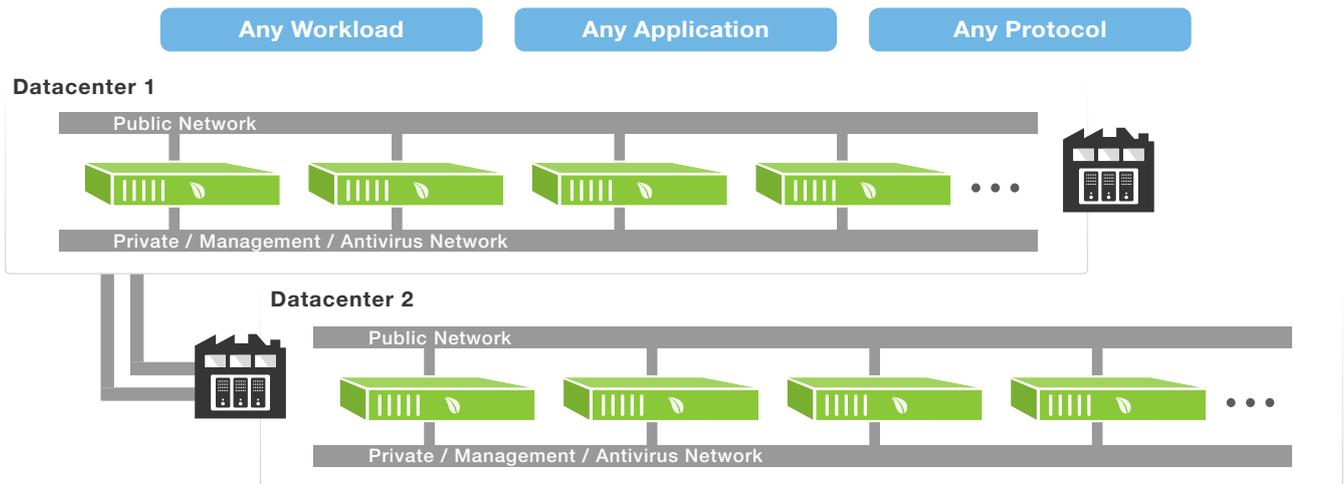# Product Overview: Compuverde Metro Cluster

Compuverde Metro Cluster is an enterprise-grade, absolute high-availability solution designed to tolerate the loss of an entire datacenter, perfect for mission-critical applications that demand zero downtime.

SOFTWARE DEFINED STORAGE

Compuverde

Compuverde Metro Cluster is a simple, multi-site configuration that offers complete redundancy to support mission-critical data and applications. Data is synchronously mirrored and distributed to two independent physical locations. The Metro configuration can be utilized with both Compuverde vNAS and Hyperconverged setups.

- ✔ Active-active nodes
- ✔ No central orchestration
- ✔ No single point of failure
- ✔ High availability

Any Workload	Any Application	Any Protocol

Datacenter 1

Public Network

Private / Management / Antivirus Network

Datacenter 2

Public Network

Private / Management / Antivirus Network

Private network
For synchronization and communication between the nodes. Use high-speed connections for low latency and increased performance. Optionally, there can be separate networks for management and antivirus.

Public network
All the nodes on both the primary and secondary datacenter are available to the clients through the public network.

Interconnection requirements
Connection between sub-clusters should be low latency, 2 ms round-trip delay time or better, and with minimum the same bandwidth as the private network.

Management Tool
Designed to help perform administrative tasks, monitor the system, change configuration features, access logs and roll out new firmware updates. Connected to the management network, it is not part of the storage cluster or node-to-node communication.

One single file system spans both datacenters and is available to applications and users inside and to the outside through Ethernet. The storage keeps itself internally synchronized, horizontally throughout the cluster and vertically down to the physical storage so that all nodes will have the same consistent view of the unified file system for all reads, writes and updates.

The two datacenters are classified as primary and secondary where the clients connect to the primary side. All nodes on both the primary and secondary side will have the same view of the files. With Metro, all data changes and updates are synchronously mirrorored and distributed to both parts of the cluster, the primary and secondary. This gives you the ability to lose one entire location without losing data.

Metro cache replication is implemented to protect all writes and updates, including the gateway state and any write operations not yet flushed to storage. As a result, some latency is introduced when writes are being mirrored to and confirmed by the secondary side, comparable to the round-trip delay between the sub-clusters (equal to 2 ms or better). Reads are unaffected. Latency can be reduced by excluding metro cache replication but at the risk of data loss if the primary side should go down (i.e. the gateway state and writes not yet distributed to the storage layer on the secondary side).

Each location is secured with erasure coding so that node failure will not affect the operation. Should the entire secondary side go down, the primary will continue to operate without

Compuverde
Software-defined storage

any loss. Should the primary side go down, the secondary is still available but in order to continue operation, it will have to be temporarily promoted to become the primary. This is to avoid any risk of "split brain", should the connection between the two locations be lost. Automatic transfer to the secondary side is on the roadmap and will be introduced soon.

Deployment: Using a license with metro enabled, it is possible to start with the primary side first, which will then operate as a Compuverde scale-out cluster, and then add the nodes for the secondary part later for full metro functionality.

Setup

1. Install the software on each node
2. Activate Metro
3. Create the storage cluster

Everything after step 1 is done through setup wizards and GUI inside the Compuverde management tool.

Scaling: In a Compuverde Metro Cluster, scaling is simply done by installing new nodes and adding them to your primary and secondary location. You will see a linear increase both in capacity, performance and IOPS as new nodes are added. With Auto Rebalancing, data is automatically moved to the new nodes to keep the cluster in balance. Make sure that the storage size on both the primary and secondary location is approximately equal in size.

Erasure Coding: In addition to being mirrored to two locations, data is erasure coded and distributed within the sub-cluster for additional protection. This ensures that each datacenter can tolerate the loss of one or more disks or nodes (depending on the chosen resilience level). The self-healing mechanism on all remaining nodes automatically kicks in to replicate the missing data simultaneously.

Virtual IP provides a failover mechanism for nodes, ensuring that all nodes appear online at all times even in the case that one goes down.

You can read more about this and other Compuverde features in the Compuverde Technical Overview.

Network: Each datacenter is interconnected through a private, high speed, low latency local area network that deals with synchronization between the nodes. Optionally, use one additional network for management and one for antivirus servers. One or more public networks deliver access for clients through any of the available protocols, e.g. SMB and NFS.

Switches: Make sure that IGMP is properly configured for all switches that are affected by private and management network traffic. For each network, all the switches should have IGMP Snooping enabled and only one switch in the group has IGMP Querier enabled.

Storage disks: Increasing the number of storage disks within each node contributes to higher maximum throughput by spreading the load. Likewise, increasing the number of nodes increases the maximum throughput for the entire cluster.

Cache devices: For performance and throughput, one or more NVMe devices are recommended for each node. Use cache replication to keep write cache safe in case the node goes down or a cache device malfunctions. Durability must be considered; enterprise class or similar (DWPD > 9).

Requirements
• A minimum of eight nodes are recommended, four for each location
• Start with equal capacity on both locations
• Low latency (2 ms or better) connection between the two locations, bandwith on par with the private network
• Metro feature is enabled in your license

Recommended hardware specifications, each node

CPU	x64 (4+ cores) SSE4.1
RAM	32+ GB
Boot disk	60+ GB
Cache disk	100+ GB NVMe SSD (one or more)
Storage disk	SAS/SATA disks
Network	10 Gigabit or more
Network switch	2 x switches
Number of nodes	Scale out from 4 nodes, each location

The Compuverde Management Tool requires a Windows computer (64-bit Windows 7 or later, 100 MB for installing the Management software) on the same subnet as the storage cluster back-end interface.

Compuverde
Software-defined storage