

White paper Tuxis cluster concept



The Tuxis cluster concept

A cluster is a group of at least three physical servers (*nodes*) that even in case of a hardware failure, will allow your business to continue. Ideal for building high end environments that require a high uptime en excellent performance.

The Tuxis Cluster gives you a scalable solution that can combine virtualization and storage into a single cluster. Start with three nodes and seamlessly grow to a 32 node virtualization cluster with a separate storage cluster for many petabytes.



Where you would normally minimize the impact on services as a result of hardware failure by using hardware with redundant power supplies, memory and complex inflexible RAID setups, the Tuxis cluster allows hardware to fail, and work around this failure. Although the cluster can be based on high end hardware with costly SLA's, we recommend commodity hardware and adding extra nodes instead. This will further reduce the impact of failing hardware while increasing resources significantly.

Growing the cluster

When more resources are required, we can increase the resources in the current nodes or add extra nodes. This goes for CPU, memory and storage.

High availability and storage

The Tuxis cluster delivers both virtualization and storage, with high availability. In traditional solutions a SAN is used as centralized storage for a virtualization cluster. The model of the SAN determines the speed and the maximum storage capacity. To build a redundant solution you usually need two of the same models. And then you need to decide on synchronization of data between the SAN's, the failover scenarios (when and how) and you might still end up with a split brain storage cluster due to a failed failover. And if you grow too fast, you might have to buy a new bigger, better and faster SAN. If you do not grow as planned, you might have bought a too expensive model. Doing it right is difficult.

The risks, choices and problems of choosing a SAN are eliminated in our integrated software defined storage solution. The software used is Ceph.

Ceph

In traditional architectures, clients communicate with a centralized component (e.g., a gateway, broker, API, facade, etc.), which acts as a single point of entry to a complex subsystem. This imposes a limit to both performance and scalability, while introducing a single point of failure (i.e., if the centralized component goes down, the whole system goes down, too).

Ceph eliminates the centralized gateway by enabling clients to interact with Ceph OSD Daemons directly. Ceph OSD Daemons create object replicas on other Ceph Nodes to ensure data safety and high availability. Ceph also uses a cluster of monitors to ensure high availability. To eliminate centralization, Ceph uses an algorithm called CRUSH (Controlled, Scalable, Decentralized Placement of Replicated Data).

To use CRUSH, a CRUSHmap is created. That map defines where the storage servers are physically located, and how many disks they have available. This map is then used to determine where data should be written or read, and how replicas of data should be distributed. This allows us to define that all data must be present in at least two different data centers.

Any server can be used as a storage server and increases throughput and capacity of the storage cluster. From a 1U server with room for 10 2,5" disks to a 4U server that can contain 72 disks, any server fits.



Keeping data safe

A Ceph pool is a gateway to your Ceph disks and is configured with a number of options, including replicas. If the pool is configured to store one replica, Ceph will make sure that the data in that pool is directly duplicated in case of failure of a disk or a node. Losing a disk or a server has no impact on the availability of the Ceph storage cluster.

Adding storage is a breeze and can be done on the fly. Expanding existing servers is an option, as well as adding complete new servers. Seamlessly and without any downtime.

Scalability

Although virtualization nodes and storage nodes can be separated, any node that is used for virtualization may also be used for storage.

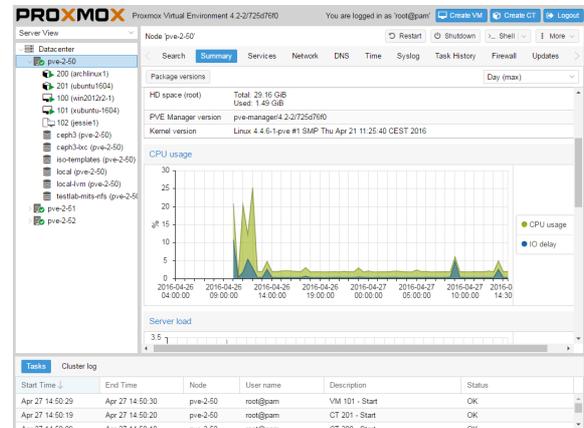
You can start with a small cluster of three nodes which can provide 15 Tb netto SSD storage, 1.5 Terabyte memory and 100+ cores and grow to a cluster with many Petabytes storage, 1100+ cores and 16+ Terabyte memory.

The hypervisor

The nodes used for virtualization need a hypervisor to be able to run virtual machines.

We install Proxmox VE, a complete open source server virtualization management software. It is based on KVM virtualization and container-based virtualization and manages KVM virtual machines, Linux containers (LXC), storage, virtualized networks, and HA clusters.

The enterprise-class features and the intuitive web interface are designed to help you increase the use of your existing resources and reduce hardware cost and administrating time - in business as well as home use. You can easily virtualize even the most demanding Linux and Windows application workloads.



Management console

Proxmox VE HA Manager

During deployment, the resource manager called Proxmox VE HA Manager monitors all virtual machines and containers on the whole cluster and automatically gets into action if one of them fails. The Proxmox VE HA Manager requires zero configuration, it works out of the box. Additionally, watchdog-based fencing simplifies deployments dramatically.

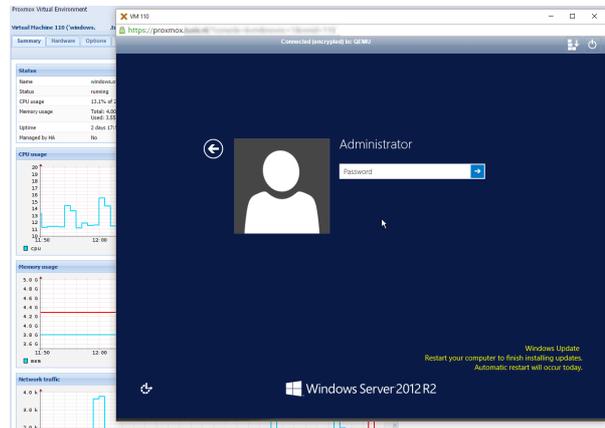
Rolling back-ups

Thanks to the Tuxis PMRB (ProxMox Rolling Backup) software, snapshots are created automatically on the fly. These snapshots can be used to restore VM's quickly to the point that they were working fine. Snapshots can also be created manually.

Functionality

Proxmox is the core of the cluster environment and supplies great tools to manage this environment:

- Remote management through a webbrowser
- Console access through the webbrowser without any plugins
- Live migration to other nodes
- Snapshots
- Cloning
- Statistics for CPU, Memory and disk load
- I/O throttling
- Resize disk, CPU and memory on the fly (when supported by the OS)
- High Availability
- Run Linux containers (LXC)
- Centrally managed build in firewall
- Authorization
- Two factor authentication



Console in webbrowser

Cluster summary

Number of virtualization nodes	Minimum: 3 Maximum: 32
Memory per node	Minimum: 64GB Maximum: 2TB
Cores per node	Minimum: 4 Maximum: 160
Guest operating system support	Windows Linux Other operating systems (community supported)
Number of storage servers	Minimum: 3 (or combined with virtualization node) Maximum: unlimited
Number of disks	Minimum: 3 disks (1 disks per node) Maximum: unlimited
Supported type of disks	- SSD - Spinning with SSD cache - Spinning
Network	Minimum: 4 Gbit/s Preferred: 10 Gbit/s
Licensing	Open Source

The data center

The availability of a cluster depends on the location and availability of power, cooling, infrastructure and physical security. Taking availability to the next level, the cluster can be distributed over three data centers while using a dedicated 10 Gbit/s fiber ring.

Construction	Data center
Height service area	Minimal 6 meters above NAP
Fire detection	Two independent, certified very early warning systems Automatic notification to fire brigade
Fire extinguishing	Certified Argonite (Ar+N ²) installation
Cooling installation	N+1 computairs N+1 cooling machines
Cooling power output	1500 W/m ²
Temperature	25 °C (+/- 2 °C) in cold paths
Lightning protection	Certified according to NEN norms
Electrical installation	Two incoming feeds into rack Separate UPS for every feed
Emergency power supply	N+1 diesel aggregates
Diesel supply	48 hours
Power per rack	Up to 96 A
Information security	ISO/IEC 27001 and NEN 7510 certified
Physical security	VEC certified safety class 4*
Alarm	Redundant connection to control center
Surveillance	Two independent surveillance services
Access control	Two factor authentication Biometric iris scanners RFID access passes
Internet Access	Router 1: 1 Gbit/s transit Joint Transit 1 Gbit/s peering NL-ix Router 2: 1 Gbit/s transit BIT
Support	24x7

Tuxis info

Tuxis Internet Engineering

Frankeneng 110B
6716 AA, Ede
Netherlands

e-mail: info@tuxis.nl

Website: <https://www.tuxis.nl/>

Tel: [0318-200208](tel:0318-200208)

Fax: 0318-830215

Chamber of commerce: [61527076](#)

VAT nr: NL854376999B01

Bank: ABN-AMRO

IBAN/SEPA Number: NL54ABNA0437576094

BIC Number: ABNANL2A