

The background is a vibrant, abstract composition of various geometric shapes and patterns. It features a dark blue base color with lighter blue, green, orange, and yellow accents. The shapes include circles, ovals, and irregular polygons, some with concentric lines or dots inside them, creating a dynamic and modern aesthetic.

# Hello And Welcome VMWare Refugees\*

\*Let's Talk about Open Source Virtualization Stuff!

# Introduction - About Me



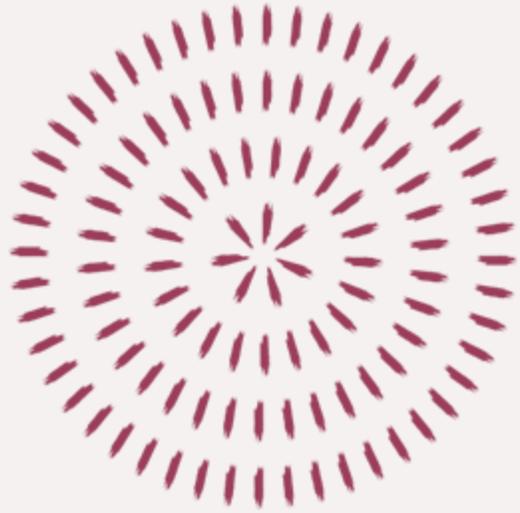


LEVELONETECHS



# CONSULTING

IF YOU'RE NOT A PART OF THE SOLUTION,  
THERE'S GOOD MONEY TO BE MADE IN PROLONGING THE PROBLEM.



## **Act 1**

**Why we're here (trust,  
"social contract", mental  
model shift)**



# Hypervisors, Virtualization and Open Source

- And the "social contract"
- Where do you want to go today?
- - “This is not a product pitch. It’s a *thinking shift*—from brand names to primitives.”

# Landscape of Open Source Virtualization Options

- RedHat OpenShift – Kubernetes and containers.
- RedHat OpenStack – Virtualization Stack (can put Shift on Stack)
- RedHat still wants you to think about the Product mindset when really innovation, going forward, isn't product oriented

# XCP-NG

XCP-NG IS PART OF THE **VATES VIRTUALIZATION STACK**



[News](#)

[Documentation](#)

[Community](#)

[Pro Support](#)



## Your seamless Hypervisor

XCP-ng: the user-friendly, high-performance virtualization solution, developed collaboratively for unrestricted features and open-source accessibility.

[Download](#)

[Get Started](#)



1500000

Downloads



85000

Forum posts

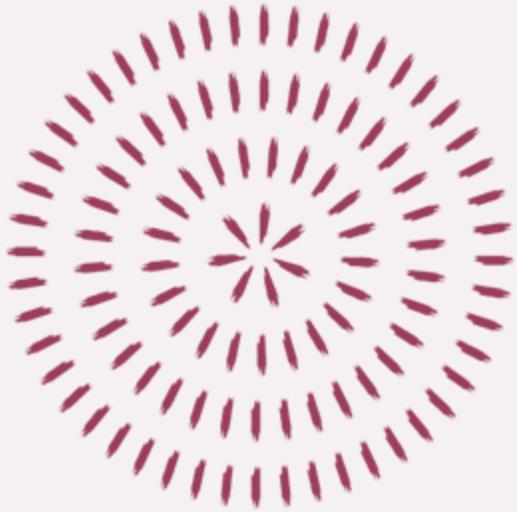


12000

Forum contributors

# Proxmox

- Debian Underneath
- QEMU/KVM for
- Proxmox is "control plane" not "product"
- European Market Aspect of Proxmox, VM, social contract "rug pull"



## **Act 2**

The open-source/commercial  
open-source stack view  
("cathedral vs bazaar")



- **Cathedral vs Bazaar (the mental model shift)**
  - “Cathedral”: one vendor, one vision, soup-to-nuts
  - “Bazaar”: assemble primitives + automation + support where needed
  - Commercial open-source = paying for **risk transfer + update channels + expertise**

# Open source ≠ “DIY forever”

You can buy:

- support SLAs
- stable repos
- escalation paths
- training

You should still demand:

- interoperability
- data portability
- open interfaces

# Proxmox: what it *is*

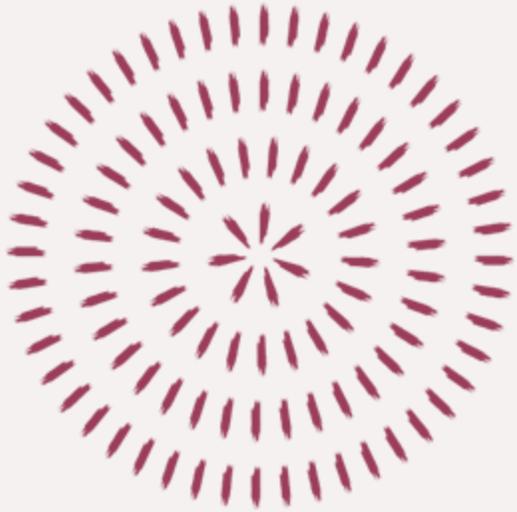
- Debian-based platform for VMs + containers
- KVM/QEMU for VMs, LXC for containers
- Multi-node clustering + web UI + REST API

# Proxmox: what it *is not*

- Not “Horizon out of the box”
- Not a single-vendor, single-brand universe
- Not magic: you still own design decisions

# Some Proxmox Components

- Storage: ZFS, Ceph, LVM, iscsi
- Networking: SR-IOV, RedHat's Paravirtualized Drivers, SDN
- SDN?!?!??
- Sync: Corosync
- Debian: Netdata, Integrit, FileMon
- LXC for Lightweight Containers
- PBS - Backups



## **Act 3**

Proxmox as a control plane  
over mature primitives  
(compute/storage/network/  
backup)



# Storage + backup: Janitorial Perspective

- Same fundamentals: latency, IOPS, throughput, failure domains
- What changes: more *choice*, more *responsibility*
- Design goal: “boring, observable, recoverable”  
... introducing ZFS

# ZFS (safe, but opinionated)

- Copy-on-write + checksums + snapshots
- Great replication story (send/receive)
- Not a clustered FS (design accordingly)
- **VMFS mindset:** “My array is sacred; ESXi is a consumer.”
- **ZFS mindset:** “My hypervisor is now part of the storage trust boundary.”

# ZFS and Clusters

- Replication for small (3 node)
- Hyperconverged means something different here than VMWare
- CEPH for larger clusters?
- How does clustering work?
- Live Migration okay?

# ZFS Gotchas

- Avoid stacking copy-on-write on copy-on-write (e.g., qcow2-on-ZFS) unless *you mean it*
- Alignment / ashift choices matter
- Watch write amplification in real workloads

# NFS/iSCSI/etc: you have options

- iSCSI and NFS are first-class in many designs
- SAN patterns still apply (multipath, zoning, failure testing)
- You can get eye-watering performance with modern transports



Recycle Bin



asdf



atto-disk-b...

Here  
1/3.r

Untitled - ATTO Disk Benchmark 4.01.0f1

File Help

Drive: [C:] Local Disk  Direct I/O

I/O Size: 16 MB to 16 MB  Bypass Write Cache

File Size: 256 MB  Verify Data

Queue Depth: 32

<< Description >>

Test Results

Write 6.50 GB/s Read 10.46 GB/s

16 MB

Transfer Rate - GB/s

ATTO Storage and Network Connectivity Experts  
www.atto.com

Task Manager

File Options View

Processes Performance App history Startup Users Details Services

CPU 3% 2.65 GHz

Memory 7.8/16.0 GB (49%)

Disk 0 (C:) SSD 100%

Ethernet VirtIO VLAN3 S: 0.9 R: 4.4 Mbps

Disk 0 (C:) QEMU QEMU HARDDISK SCSI Disk ...

Active time 100%

60 seconds

Disk transfer rate 10 GB/s

60 seconds

Active time Average response time Capacity: 120 GB

100% 22.0 ms Formatted: 120 GB

System disk: Yes

Read speed Write speed Page file: Yes

12.0 GB/s 10.8 KB/s Type: SSD

Open Resource Monitor

# CEPH

- Distributed storage system (scale + resilience tradeoffs)
- Proxmox integrates Ceph management in GUI
- Not always the right answer—depends on failure domains & ops maturity

# Proxmox Backup Server

- “Snapshots aren’t backups”
- Proxmox Backup Server (PBS): incremental + dedup + encryption options
- Restore speed and *confidence* are the feature
- ZFS can be a big player here

# Acceleration

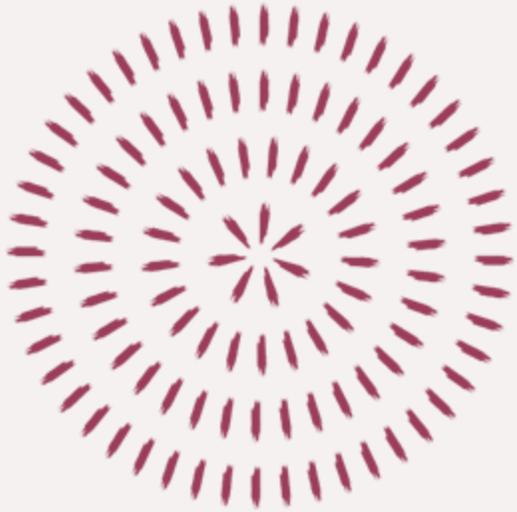
- For VMs, VirtIO drivers for both networking and storage
- SR-IOV / RDMA / offloads can slash CPU and jitter
- Observable design beats "it should be fast"

# Proxmox 9: SDN

- Open vSwitch (OVS) + SDN: “power tools”
  - Cite OVS overview + Linux Foundation hosting.
- OVS exists to enable network automation + SDN features
- Proxmox has SDN features and is evolving (“fabrics” concept in VE 9)
- Reality: don’t build a science fair—start with a few repeatable patterns

# Gotchas

- Talk about ZFS encrypted Volumes
  - vTPM Windows 11
  - RDMA o IB still RedHat Ronly, for now
  - But remember ZFS is not an option for RedHat
- Networking: bonding/VLAN design can accidentally block the fast path



## **Act 5**

VDI: the user-experience  
physics (encode/decode, GPU  
slicing, SR-IOV/MIG)



# VDI: Why it needs GPU Acceleration

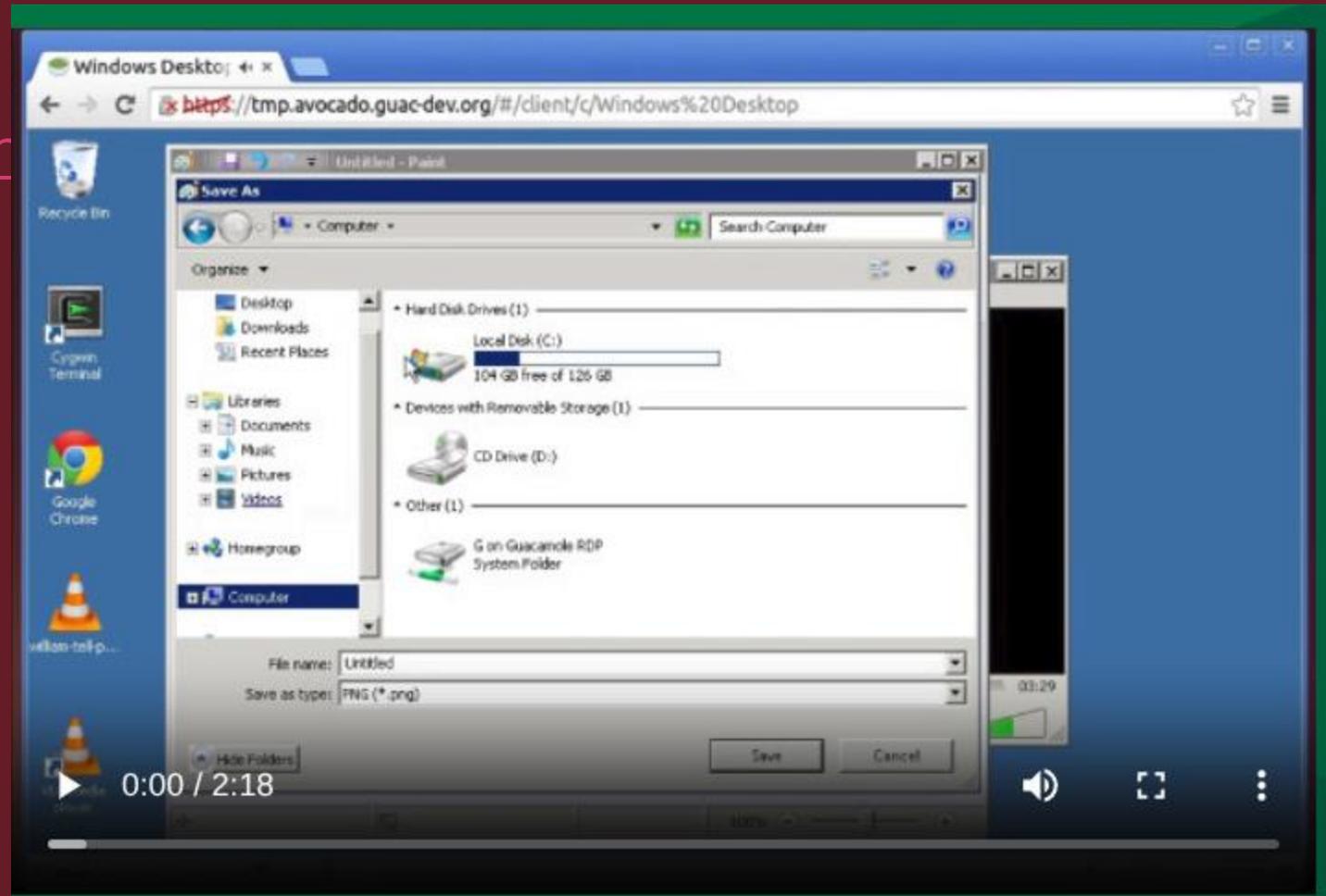
- VDI = centralized desktops + remote display protocols
- UX is dominated by:
  - encode/decode
  - network latency/jitter
  - storage latency
- “More vCPU” often loses to “proper acceleration”

# VDI Demo

- A **2-core VM with GPU encode/decode** often *feels better* than an 8-core VM without it

# No Horizon you say?

- Apache Guacamole  
<https://guacamole.apache.org/demos/demo.mp4>

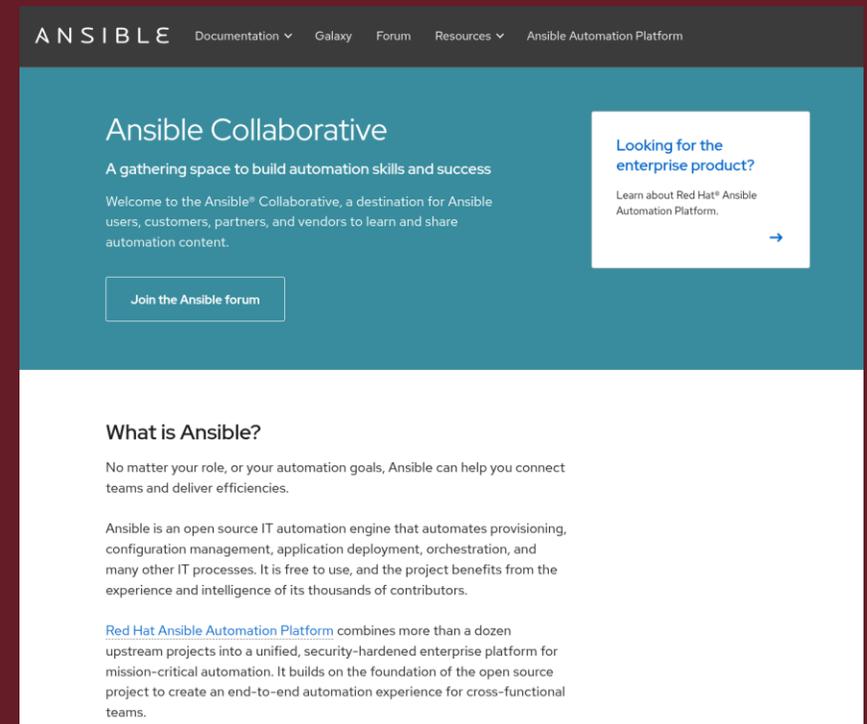


# How do you do automation?

- Remember, this is a Bazaar, not a cathedral
- Ansible! (wait.. More RedHat?!? Kinda)

No Horizon-like turnkey suite by default

- You build an experience using:
  - automation (Ansible)
  - golden images
  - identity + policy
  - a remote display stack
- This is the bazaar trade: **choice > branding**



The screenshot shows the Ansible website's 'Ansible Collaborative' section. The navigation bar at the top includes 'ANSIBLE', 'Documentation', 'Galaxy', 'Forum', 'Resources', and 'Ansible Automation Platform'. The main content area features a teal background with the heading 'Ansible Collaborative' and the subtext 'A gathering space to build automation skills and success'. Below this is a paragraph welcoming users to the collaborative space. A 'Join the Ansible forum' button is prominently displayed. To the right, a white box asks 'Looking for the enterprise product?' and provides a link to learn about the Red Hat Ansible Automation Platform. The lower section, on a white background, is titled 'What is Ansible?' and contains two paragraphs describing the platform's capabilities and its open-source nature. A link to the 'Red Hat Ansible Automation Platform' is provided at the end of the text.

# Ansible



# GPU virtualization menu

- **Passthrough (VFIO):** whole GPU to one VM
- **SR-IOV:** one device exposes multiple VFs to multiple VMs
- **MIG:** hard partitions (on supported NVIDIA GPUs)
- **vGPU desktop (licensed):** shared scheduling model, density-oriented



# NVIDIA MIG: the subscription avoidance “keyhole”

- MIG partitions supported GPUs into **isolated instances**
- RTX PRO 6000 Blackwell Server Edition: **up to 4 MIGs @ 24GB**
- Hardware slices: predictable QoS, strong isolation

# MIG caveats

- Hard partitions are great for predictable interactive workloads
- Not for “let’s pretend one card is a multi-GPU training cluster”
- If your goal is multi-worker training efficiency: we should talk architecture

# Intel as the challenger (SR-IOV path)

- SR-IOV on Intel Arc Pro is real but still maturing
- Firmware profiles can trade **density vs stability**
- Linux enablement is improving, but early adopter tax exists

# IOMMU: why SR-IOV isn't "YOLO passthrough"

- IOMMU maps device DMA into correct guest memory
- Enables isolation for VF assignment
- Still requires careful platform + BIOS + firmware choices
- IOMMU Device Group Hints (new)

# "Future" Here

- AMD SEV-SNP: VM memory confidentiality + integrity protections
- Intel TDX: confidential VMs via TEE-style model
- These exist because hyperscalers demanded stronger tenant isolation

vmware support here is already pretty mature; FANG's DIY solutions also pretty mature.

- CXL in the winter of DRAM availability
- PCIe Fabrics / NVMe over fabric

# What you'll miss from VMWare

- One-brand, end-to-end suite polish
- A huge certified ecosystem
- Some "magic" defaults (until they aren't)

## Gains

- Interoperability and stack transparency
- Choice in storage/network/backup strategies
- Automation-first operations
- Less hostage risk