VM & VPS

Virtualização, desempenho e escala.

Guia técnico e executivo para infraestrutura virtual

EnQ Digital

Edição 2026

Sumário

Da camada física ao serviço virtual: fundamentos, arquitetura, operação, segurança e decisão comercial.

1. VM, VPS e virtualização

Virtualização abstrai hardware físico para executar múltiplos sistemas isolados no mesmo host.

Conceitos

Termo	Definição
VM	Máquina completa com vCPU, RAM, disco, rede e sistema operacional
VPS	VM comercializada como servidor virtual privado, normalmente com plano e SLA
Host	Servidor físico que executa o hipervisor
Guest	Sistema operacional dentro da VM
Tenant	Cliente ou domínio lógico isolado

Diferença essencial

VM descreve a tecnologia. VPS descreve geralmente o serviço entregue ao cliente. Uma VPS é normalmente uma VM, mas uma VM interna de um cluster corporativo não é necessariamente vendida como VPS.

A qualidade de uma VPS depende do conjunto: host, política de overcommit, storage, rede, segurança, backup, suporte e transparência operacional.

2. Como a virtualização funciona

O hipervisor apresenta hardware virtual e arbitra o uso dos recursos físicos.

Caminho da execução

- vCPU é agendada sobre threads ou núcleos físicos.
- Memória virtual é mapeada para páginas de RAM do host.
- Discos virtuais usam arquivos, volumes ou blocos em storage local/compartilhado.
- vNICs conectam-se a bridges, vSwitches, port groups ou redes overlay.
- Drivers paravirtualizados reduzem overhead de I/O.

Extensões de hardware

Intel VT-x/VT-d e AMD-V/IOMMU permitem virtualização assistida, isolamento e passthrough de dispositivos. Secure Boot e TPM virtual apoiam cadeias de confiança.

Responsabilidade

O hipervisor isola e agenda; o sistema convidado ainda precisa de patch, hardening, monitoramento, backup e gestão de identidade.

3. Hipervisores tipo 1 e tipo 2

A escolha influencia disponibilidade, integração, licenciamento e operação.

Modelos

Modelo	Execução	Uso
Tipo 1	Direto no hardware	Data center e produção
Tipo 2	Sobre um sistema operacional	Laboratório e desktop

Plataformas comuns

- VMware ESXi/vSphere e VCF.
- Microsoft Hyper-V.
- KVM/QEMU em distribuições Linux e plataformas como Proxmox VE.
- Nutanix AHV e outras soluções integradas.

Critérios

Compatibilidade de hardware, recursos de HA, live migration, storage, rede, backup, APIs, ecossistema, capacitação e custo total devem ser avaliados em conjunto.

4. Hardware do host

A camada virtual não elimina limites físicos; ela os concentra.

Servidores rack utilizados como hosts de virtualização. Foto: Dsv/Wikimedia Commons, domínio público.

Elementos

Componente	Decisão
CPU	Sockets, cores, frequência, cache, TDP e geração
RAM ECC	Capacidade, canais, velocidade e expansão
PCIe	Lanes para NIC, HBA, NVMe e GPU
Storage	NVMe/SAS, RAID, SAN, NAS, Ceph ou HCI
Rede	10/25/40/100 GbE e redundância
BMC	Gestão fora de banda e automação

Resiliência

Fontes A/B, ventiladores hot-swap, interfaces redundantes e caminhos separados reduzem falhas locais. Para HA real, são necessários múltiplos hosts e dados acessíveis após a falha de um nó.

5. CPU, vCPU e overcommit

vCPU não equivale automaticamente a um núcleo físico dedicado.

Agendamento

O hipervisor agenda vCPUs em CPUs lógicas. Workloads que aguardam tempo de CPU acumulam CPU Ready; limites, reservas e shares influenciam a disputa.

Métricas

- Uso do host e da VM.
- CPU Ready/co-stop ou métrica equivalente.
- Frequência efetiva, steal time e throttling.
- Relação vCPU:pCPU por cluster e por perfil de carga.

Boa prática

Consolidar com base em percentis e simultaneidade. Bancos, aplicações licenciadas por core e baixa latência podem exigir menor overcommit, afinidade ou núcleos dedicados.

Mais vCPUs podem piorar desempenho se a VM precisar aguardar uma janela maior de escalonamento. Dimensione pela carga medida.

6. Memória, NUMA e ballooning

RAM costuma ser o primeiro limite de densidade de um cluster.

Mecanismos

• Reserva garante capacidade física.

• Limite restringe consumo máximo.

• Ballooning recupera memória com cooperação do guest.

• Compressão e swap são últimas linhas e aumentam latência.

• Transparent page sharing depende da plataforma e política.

NUMA

VMs grandes devem respeitar os limites dos nós NUMA. Acesso remoto à memória pode aumentar latência; vNUMA e posicionamento precisam estar alinhados ao sistema operacional e à aplicação.

Capacidade

Inclua RAM do hipervisor, cache, agentes e margem N+1. Não trate memória provisionada como consumo médio sem verificar picos e working set.

7. Storage virtual

Latência, IOPS e throughput são compartilhados entre VMs e atividades de infraestrutura.

Opções

Modelo	Vantagem	Atenção
Local	Baixa latência e simplicidade	Mobilidade e HA
SAN FC/iSCSI	Compartilhamento e recursos maduros	Fabrics, multipath e custo
NFS	Operação simples	Rede e servidor de arquivos
HCI/Ceph	Escala distribuída	Rede, quorum e rebuild
NVMe-oF	Alta performance	Projeto e compatibilidade

Controles

- Thin provisioning com alerta de capacidade.
- Multipath e caminhos físicos distintos.
- QoS para evitar noisy neighbor.
- TRIM/UNMAP, cache e alinhamento conforme plataforma.
- Monitorar latência no guest, host, datastore e array.

8. Rede virtual

A conectividade virtual combina NICs físicas, switches virtuais e políticas de segmentação.

Switches e patch panels de uma rede física que sustenta redes virtuais. Foto: Dsimic/Wikimedia Commons, CC BY-SA 4.0.

Planos de rede

- Gestão do hipervisor.
- Tráfego das VMs/tenants.
- Live migration.
- Storage.
- Backup e replicação.
- Cluster e gestão fora de banda.

Disponibilidade

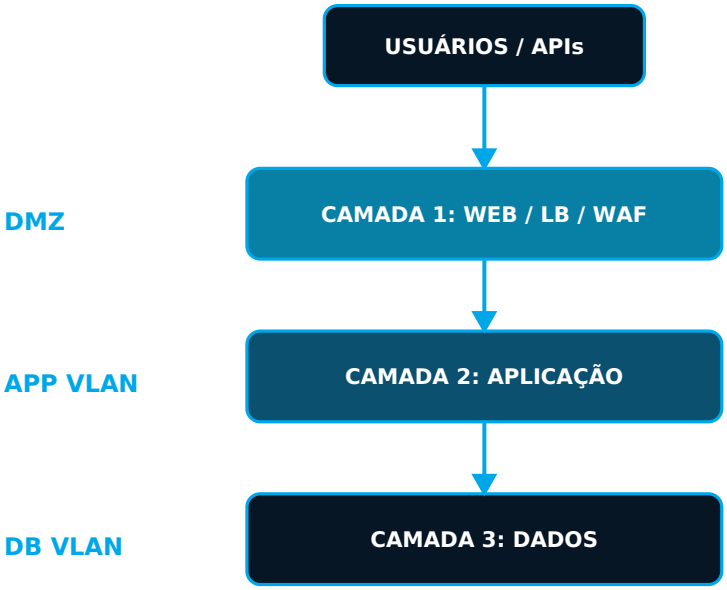
Distribua uplinks entre switches físicos, valide LACP/MLAG ou roteamento e mantenha MTU consistente. O desenho deve sobreviver à perda de uma NIC, switch ou caminho.

Virtualização de rede

VLANs, VRFs, VXLAN/EVPN, overlays e microsegmentação ampliam escala e isolamento, mas exigem observabilidade ponta a ponta.

9. Topologia em 3 camadas

Uma arquitetura básica separa borda, virtualização e dados.



Fluxo conceitual: Web/borda → Aplicação/VMs → Dados. Cada camada utiliza rede e regras próprias.

Exemplo

Camada	Componentes
1. Borda/DMZ	DDoS, roteadores, firewall HA, WAF e load balancer
2. Compute	Cluster de hipervisores e redes dos tenants
3. Dados	Storage, bancos, backup e replicação

Gestão, monitoramento e backup devem usar planos segregados e alcançar apenas os componentes necessários.

10. Segurança e isolamento

A segurança depende de controles no provedor, hipervisor, rede, guest e processo.

Controles

- MFA, RBAC e contas administrativas separadas.
- Rede de gestão isolada e acesso por VPN/bastion.
- Hardening e patch do host e dos guests.
- Microsegmentação e regras deny-by-default.
- EDR, vulnerabilidade, logs centralizados e SIEM.
- Criptografia em trânsito e em repouso conforme risco.

Noisy neighbor

QoS, limites, reservas e monitoramento reduzem impacto entre tenants. Isolamento lógico não substitui avaliação de criticidade; cargas com exigência máxima podem requerer host dedicado.

Administração

Acesso ao console do hipervisor é privilegiado. Registre ações, aplique segregação de funções e proteja backups e snapshots contra exclusão.

11. Templates, clones e snapshots

Esses recursos aceleram operação, mas têm propósitos diferentes.

Objetos

Recurso	Uso correto	Não substitui
Template	Imagem padrão para novas VMs	Gestão de patches
Clone	Cópia independente ou vinculada	Backup externo
Snapshot	Retorno curto antes de mudança	Backup de longo prazo
Golden image	Baseline testado e versionado	Hardening contínuo

Risco de snapshot

Snapshots longos crescem, aumentam cadeia de dependência e podem degradar I/O. Defina proprietário, validade, alerta e remoção controlada.

Cloud-init/Sysprep

Automatize hostname, rede, chaves e identidade na primeira inicialização; evite clonar segredos, IDs e agentes com estado.

12. Alta disponibilidade

HA reinicia ou mantém serviços após falhas, conforme capacidade e tecnologia.

Pré-requisitos

• Cluster com quorum confiável.

• Capacidade N+1 ou superior.

• Storage acessível ou replicado.

• Redes redundantes.

• Compatibilidade de CPU para migração.

• Health checks e políticas de reinício.

Live migration

Move uma VM entre hosts com pouca ou nenhuma interrupção, facilitando manutenção. Ela não substitui HA, backup ou proteção contra falha da aplicação.

Anti-affinity

Mantenha réplicas, controladores de domínio e nós do mesmo serviço em hosts ou domínios de falha diferentes.

Teste a perda de host, switch, datastore e alimentação. HA sem teste permanece uma suposição.

13. Backup e recuperação

Backup deve ser independente do ambiente primário e testado.

Estratégia

• Política 3-2-1 com cópia externa.

• Imutabilidade ou air gap contra ransomware.

• Backup consistente com aplicação para bancos e diretórios.

• Criptografia, retenção e controle de acesso separado.

• Testes regulares de restauração de arquivo, VM e serviço.

RPO e RTO

Frequência define potencial de perda; método e capacidade de restore influenciam tempo. Meça restauração completa, incluindo rede, DNS, certificados e dependências.

Agente x imagem

Backup em nível de imagem simplifica recuperação da VM; agentes podem oferecer granularidade e consistência específica. Ambientes críticos frequentemente combinam ambos.

14. Disaster Recovery

DR recupera serviços em outro domínio de falha.

Modelos

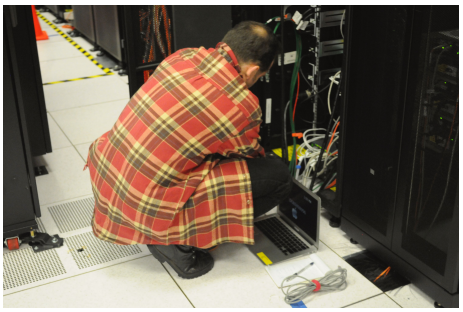
Modelo	Custo	Recuperação
Backup externo	Menor	Mais lenta
Cold standby	Baixo/médio	Horas
Warm standby	Médio	Minutos/horas
Hot/ativo-passivo	Alto	Minutos
Ativo-ativo	Muito alto	Baixa interrupção

Runbook

- Ordem de subida e dependências.
- Replicação, consistência e ponto de corte.
- Redes, DNS, IPs, certificados e firewall.
- Critérios de declaração e retorno ao site primário.
- Comunicação, responsáveis e evidências de teste.

15. Operação e observabilidade

O serviço virtual precisa ser observado do hardware à aplicação.



Operação técnica em rack de data center. Foto: Derrick Coetzee/Wikimedia Commons, CC0 1.0.

Indicadores

- Saúde de host, BMC, fontes, discos e temperatura.
- CPU Ready, memória ativa, ballooning e swap.
- Latência/IOPS de datastore e filas.
- Drops, erros, throughput e latência de rede.
- Disponibilidade de serviços e experiência do usuário.

Governança

Inventário, CMDB, capacity planning, patches, mudanças, incidentes, problemas e ciclo de vida devem ter responsáveis e evidências.

16. Automação com Ansible

Ansible padroniza configuração dentro das VMs e, quando suportado, integra plataformas de virtualização.

Automatizações

• Baseline Linux/Windows e hardening.

• Usuários, chaves, pacotes, serviços e certificados.

• Agentes de monitoramento, backup e EDR.

• Patching em lotes com health checks.

• Configuração de middleware e aplicações.

• Inventário dinâmico a partir da plataforma.

Segurança

Use vault ou secrets manager, contas de serviço, menor privilégio, logs, revisão e pipelines. Playbooks devem ser idempotentes e testados.

Fluxo seguro: selecionar lote → retirar do balanceador → atualizar → testar → reintegrar → avançar.

17. Terraform e OpenTofu

IaC torna o provisionamento de VMs e redes declarativo, versionado e revisável.

Recursos típicos

- VM, template, CPU, RAM e disco.
- Redes, VLANs, IPs e regras.
- Datastores, tags, pastas e políticas.
- DNS, load balancer e integrações.
- Outputs para inventário Ansible e CMDB.

State

Proteja o state com backend remoto, criptografia, lock, versionamento e controle de acesso. Segredos podem aparecer no state; trate-o como dado sensível.

Pipeline

Commit → lint/validate → plan → revisão → apply → cloud-init → Ansible → testes → observabilidade. Terraform e OpenTofu devem usar versões fixadas e detecção de drift.

Automação reduz variabilidade; governança reduz o raio de impacto da automação.

18. Containers x VMs

Containers compartilham o kernel; VMs virtualizam uma máquina completa.

Comparação

Critério	VM	Container
Kernel	Próprio	Compartilhado com host
Isolamento	Mais forte por padrão	Depende do runtime/política
Inicialização	Segundos/minutos	Milissegundos/segundos
Densidade	Média	Alta
SO diferente	Sim	Limitado ao kernel

Combinação

Kubernetes frequentemente roda sobre VMs para separar tenants, facilitar ciclo de vida e consumir recursos de clusters consolidados. Bare Metal pode ser preferido para desempenho máximo ou aceleradores.

19. Desempenho e benchmark

Benchmark deve reproduzir o perfil real e considerar disputa entre tenants.

Métricas

- CPU: tempo, ready, steal e frequência.
- RAM: working set, faults, ballooning e swap.
- Storage: IOPS, throughput, latência média e percentis.
- Rede: throughput, PPS, latência, jitter e perda.
- Aplicação: transações, filas, tempo de resposta e erro.

Teste

Use dataset, concorrência, duração e janela representativos. Teste em horário de carga, com backup e tarefas de infraestrutura ativos.

Expectativa

Desempenho de VPS não deve ser inferido apenas por vCPU e RAM. Geração da CPU, política de compartilhamento, storage e rede podem mudar o resultado.

20. Dimensionamento

Dimensione pelo consumo medido, crescimento, falhas e licenciamento.

Exemplo ilustrativo

Perfil	Configuração inicial	Atenção
Web leve	2 vCPU / 4 GB / 80 GB	Escala horizontal
Aplicação	4 vCPU / 8-16 GB / 150 GB	Latência e integrações
Banco médio	8 vCPU / 32 GB / NVMe	IOPS, backup e licença
VDI	Por usuário/concurrency	GPU, login storm e perfil

Cluster

- Somar picos simultâneos, não apenas médias.
- Reservar capacidade para falha de um host.
- Considerar manutenção, rebuild e crescimento.
- Separar perfis de alta I/O ou baixa latência.

Valores são exemplos didáticos; a proposta final deve nascer de telemetria ou prova de conceito.

21. Licenciamento

Virtualização pode alterar significativamente o custo de software.

Perguntas

- Licença por core físico, vCPU, socket, VM, host ou usuário?
- Há mínimo de cores por processador ou servidor?
- Mobilidade entre hosts exige licenciar todo o cluster?
- Ambiente de DR tem direito passivo ou licença adicional?
- Recursos de gestão, backup e segurança são separados?

Windows e aplicações

Edições e direitos de virtualização variam. Bancos, middleware e software comercial podem exigir regras próprias. Confirme sempre contrato, métrica e versão com o fabricante ou parceiro.

Open source

Ausência de licença proprietária não elimina custo de suporte, atualização, capacitação e operação.

22. VM x VPS x Bare Metal x Cloud

O melhor modelo depende da carga e da responsabilidade desejada.

Comparativo

Modelo	Controle	Elasticidade	Previsibilidade
VM privada	Alto	Média	Alta
VPS gerenciada	Médio/alto	Média	Alta
Bare Metal	Máximo	Baixa/média	Alta
Cloud pública	Variável	Alta	Variável
Container/PaaS	Menor infra	Alta	Depende do serviço

Escolha

VPS favorece rapidez e previsibilidade; VM privada favorece governança e integração; Bare Metal favorece isolamento e performance; cloud favorece elasticidade e serviços gerenciados.

23. Migração para VM/VPS

A migração deve mapear dependências, compatibilidade e janela.

Etapas

- Inventário de servidores, aplicações, dados e licenças.
- Assessment de CPU, RAM, I/O, rede e dependências.
- Desenho de destino, segurança, backup e monitoramento.
- Piloto, teste de carga e restauração.
- Replicação ou conversão P2V/V2V.
- Cutover com critérios de go/no-go e rollback.
- Hypercare, aceite e otimização.

Métodos

Backup/restore, replicação, migração de VM, export/import e reconstrução automatizada são opções. O menor downtime nem sempre é o método de menor risco.

Antes do corte: validar DNS, certificados, firewall, sincronismo de dados, desempenho, backup, observabilidade e plano de retorno.

24. VPS EnQ Digital

Recursos virtuais para aplicações corporativas, com integração a data center, conectividade e segurança.

Proposta

A EnQ Digital pode estruturar VPS e ambientes virtuais conforme capacidade, criticidade, sistema operacional, rede, backup, segurança e nível de gestão contratado.

Perfis comerciais de referência

Plano	vCPU	RAM	Indicação
Bronze	2	4 GB	Sites, serviços leves e homologação
Silver	4	8 GB	Aplicações corporativas médias
Gold	8	16 GB	Cargas mais intensivas e bancos pequenos

Possibilidades

- Linux ou Windows conforme disponibilidade e licença.
- IPs, VLANs, VPN e regras de firewall conforme projeto.
- Backup, monitoramento e suporte conforme SLA.
- Escala de CPU, RAM e storage mediante capacidade.
- Integração com Bare Metal, colocation, storage S3 e conectividade.
- Automação com Ansible e Terraform/OpenTofu conforme escopo.

Importante

Os perfis são referências de portfólio. Preços, storage, tráfego, sistema operacional, SLA e recursos adicionais devem constar na proposta comercial vigente.

25. Checklist e conclusão

Uma boa contratação transforma especificações em compromissos verificáveis.

Checklist

• Modelo e geração de CPU; política de vCPU e overcommit.

• RAM, reservas, limites e possibilidade de expansão.

• Storage: tipo, capacidade, IOPS, latência e proteção.

• Rede: banda, tráfego, IPs, DDoS, VLAN/VPN e firewall.

• SLA, suporte, manutenção e responsabilidade pelo guest.

• Backup: frequência, retenção, imutabilidade e restore.

• Localidade, compliance, portabilidade e processo de saída.

• Licenças, instalação, migração e custos adicionais.

Créditos

Dell PowerEdge Servers — Dsv, domínio público. 19-inch rackmount Ethernet switches and patch panels — Dsimic, CC BY-SA 4.0. Technician with laptop working on server rack at NERSC — Derrick Coetzee, CC0 1.0. Fontes: Wikimedia Commons; imagens redimensionadas para diagramação.

EnQ Digital — Cloud • IA • Data Center • Conectividade. Infraestrutura inteligente para empresas que não podem parar.