ESXi Install

HW requirements: HCL compliant (see link below), 64bit x86 CPUs (minimum of 2 cores), 2GB RAM, 5GB disk, password 6-64 characters.
Set HW clock (BIOS) to UTC. VMFS only created on first disk. IDE/ATA drives not supported for VMFS.
ESXi Installable starts in evaluation mode (60 days) on first power-on even if host is licensed. If no DHCP at install, link local IP used 169.254.x.x/16. Disconnect Fibre Channel connections prior to installation.

Firewall

FW Port	Source	Destination	Protocol	Description
22	SSH client	ESXi	TCP	SSH server
53	ESXi	DNS server	UDP	DNS requests
80	Clients	ESXi	TCP	Redirects to HTTPS (443)
123	ESXi	NTP source	UDP	NTP (time) client
427	ESXi	CIM servers	UDP	CIM SLPv2 client to find server
443	Clients, vCenter	ESXi	TCP	HTTPS access
902	ESXi	ESXi	TCP/UDP	Migrate & provision
902	Client	ESXi	UDP	Access to VM console
902	ESXi	vCenter	TCP/UDP	Heartbeat
5900-5964	ESXi	ESXi	TCP	RFB for management tools-VNC
5988	CIM server	ESXi	TCP	CIM transactions over HTTP
5989	vCenter/ESXi	ESXi/vCenter	TCP	CIM XML over HTTPS
8000	ESXi	ESXi	TCP	vMotion requests

Possible extras:68(DHCP),161/162(SNMP),514(syslog), 1234/1235(HBR) & HA,FT,NFS,iSCSI traffic

ESXi Partitions: • 2 boot banks • 4GB VFAT scratch (system swap & vm-support info) – not required but uses ramdisk if not present, or can use remote NFS partition • locker • 110MB diagnostic for core dumps (can redirect to ESXi Dump Collector) • VMFS5 on each disk's free space.
Fresh install has GPT, upgraded keeps MBR style.
Not supported: • ESXi Installable & Embedded on same host • Booting multiple servers from 1 image.
Sources: • Boot – CD, USB, PXE boot, remote access mounted ISO (iLO, DRAC, RSA, etc) • Script – CD, USB, NFS, HTTP(S), FTP. Specify location in *kernelopts* line of boot.cfg, or `ks=` boot option (`Shift` + `O`) . ESXi5 cannot PXE boot from one image, then install different image. Boot image is always the installed image. Default ks file at /etc/vmware/weasel/ks.cfg (password is *mypassword*).
Destination: – SATA (considered remote – no scratch), SAS, SCSI disk, flash drive, FC or SW iSCSI (target IQN set in iBFT BIOS) SAN LUN, PXE stateless (Auto Deploy).
Image Builder: PowerCLI tool (server component & cmdlets) to create images (Image Profile) with customized updates & drivers. Deploy image as install CD or via Auto Deploy server. VIBs must pass dependency check & meet acceptance level – VMwareCertified, VMwareAccepted, PartnerSupported, CommunitySupported. -AcceptanceLevel parameter changes level. Include vmware-fdm VIB if host will be in HA cluster. Clone a published profile to create custom profile.
Auto Deploy (stateless): loads ESXi images across network into host's memory every boot. Can set configuration via Host Profile (see vCenter section). Server as Windows install or VCSA. 1-to-1 Auto Deploy registered to vCenter. Hosts need BIOS firmware (not UEFI). Host DHCP reservations recommended. VLAN tagged (trunked) boot NICs not recommended. Multiple hosts rebooting can cause boot storm for Auto Deploy server. Hosts require Dump Collector. Redirect logs to syslog server or NFS datastore. PowerCLI Bulk Licensing useful for Auto Deploy. Rebooted host stays in maintenance mode if vDS is used & vCenter is unavailable. Can use VIBs, Images Profiles & Software Depots (online - HTTP or offline - ZIP file) during install.
First boot • set DHCP for IP and point to TFTP server for gPXE, add rules to rules set • identify Image Profile • (optional) rule for Host Profile • apply Active Rule Set.
Re-provisioning (subsequent reboots) can change answer file, use different image or host profile.
Components: • Auto deploy server – manages state information, serves images & host profiles • Rules engine – manages rules & rule sets • Image Profile - matches sets of VIBs to host • Host Profile - machine specific information • Answer File - host specific information the user provided during first boot (only accessed via Host Profiles UI).
Rules engine: • Rules – assigns Image Profiles, Host Profiles, location within vCenter hierarchy, identifies host via MAC address, SMBIOS asset tag, BIOS UUID, or IP address • Active Rule Set – maps matching rules to hosts when image is requested • Working Rule Set - test rules before making active.
Deployment Information: • Image state - profile created by Image Builder PowerCLI tool. Contains executable software • Configuration state - from Host Profile • Dynamic state - runtime information in memory, lost during reboot • VM state - VM auto-start info, managed by vCenter but locally stored if HA is enabled • User Input – host profile set to require user supplied host specific information, stored in answer file.

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

ESXi Install

Post install: test cables are in correct VMNICs: `watch -n 1 'esxcli network nic list'`
Upgrade from ESX/ESXi 4.x: • vCenter Update Manager (needs 350MB free in /boot) • Interactive upgrade from CD or USB drive • Scripted upgrade. 5.0 upgrades to 5.x can also use • Auto Deploy (if used for 5.0 install) • `esxcli`.

Links

Installing ESXi 5.0 Best Practices http://kb.vmware.com/kb/2005099
Upgrading to ESXi 5.0 Best Practices http://kb.vmware.com/kb/2005102
Resolution Path – Troubleshooting Installs http://communities.vmware.com/docs/DOC-15789
Hardware Compatibility Guide ("*HCL*") www.vmware.com/go/hcl
Troubleshooting vSphere Auto Deploy http://kb.vmware.com/kb/2000988

ESXi Hosts

Maximums (per host): Logical CPUs (incl HT) = 160 RAM = 2TB VMs = 512
vCPUs = 2048 vCPUs per core = 25

Logs: All logs in /var/log/ directory (sym links from /var/run/log). View host logs via: • DCUI • ESXi Shell • Extract `vm-support` bundle • http://hostname/host • vCLI `vifs` • vSphere Client connected to host

auth.log	ESXi Shell authentication
esxupdate.log	ESXi patches/updates
fdm.log	HA logs
hostd.log	Host management (VM & host events; Client, vpxa, SDK connections)
shell.log	ESXi Shell usage (enable/disable & commands
sysboot.log	VMkernel & module startup
syslog.log	Management service initialization, watchdogs, scheduled tasks, DCUI
vmkernel.log	Core VMkernel logs (devices, storage/network device/driver events & VM startup)
vmkwarning.log	VMkernel warnings & alerts
vmksummary.log	ESXi startup/shutdown, uptime, VMs running, service usage
vpxa.log	vCenter vpxa agent

Shell Commands

`--help` for esxcli namespaces & commands relative to location. `localcli` bypasses hostd
Startup level for management services (& lists all services): `chkconfig --list`
Restart all management services: `/sbin/service.sh restart`
Restart single service (& start|stop|status available): `/etc/init.d/<service> restart`
Common services: • hostd (primary ESXi daemon) • vpxa (vCenter agent) • vmware-fdm (HA)
Backup host configuration: `vicfg-cfgbackup -s /tmp/<hostname>.cfgbak`
(restore `-l`, force restore to different build number `-f`)
Export detailed configuration file: `esxcfg-info > /tmp/esxcfg-info.txt`
Gather debugging report: `vm-support -w /tmp`
List running VMs (before maintenance): `esxcli vm process list`
Resource usage: `esxtop` (Shell) `resxtop` (vCLI). Customize & save: `W` (updates .esxtop50rc file)
List CPU details: `esxcli hardware cpu list`
Show CPU supported functions: `esxcli hardware cpu global get`
Show memory and NUMA nodes: `esxcli hardware memory get`
List free memory allocated to ramdisks: `esxcli system visorfs ramdisk list`
Show version information for ESXi: `esxcli system version get`
Show the host's acceptance level: `esxcli software acceptance get`
Show all the installed VIBs: `esxcli software vib list`
Detailed information on installed VIBs: `esxcli software vib get`
Show syslog configuration: `esxcli system syslog config get`
Show logging config for each log: `esxcli system syslog config logger get`
Show remote coredump config: `esxcli system coredump network get`
Lists firewall status & actions: `esxcli network firewall get`
Lists firewall rulesets: `esxcli network firewall ruleset list`
Refresh firewall after adding new ruleset: `esxcli network firewall refresh`
Show description of VMkernel error: `vmkerrcode <error_code_number>`
Lists drivers loaded at startup: `esxcli system module list`
List advanced options: `esxcli system settings advanced -l`

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

ESXi Hosts

CPU Power management policies: • Not Supported - no host support or disabled in BIOS • High Performance - only used when BIOS warning • Balanced (default) - conservative, shouldn't affect performance • Low Power - aggressive power management, can lower performance • Custom
Memory: Host reclaims memory from VM by: • TPS (Transparent Page Sharing) – "RAM dedupe", `PSHARE` in esxtop • Balloon driver (vmmemctl) - forces guest to use native algorithms (guest swap) • Memory compression • .vswp file (host level swapping). Local or networked SSD is tagged by VMkernel as optimal swap location to reduce impact. During contention, host memory allocated based on shares & working set size (recent activity). Idle memory is taxed progressively to prevent VM hoarding. Guest swap should be ≥ (vRAM - Reservation) x 65%, or balloon driver can cause guest kernel panic. Memory faults can be detected & quarantined to reduce chance of a PSOD (hardware dependent).
NUMA (Non-Uniform Memory Access): CPUs have localized memory. NUMA scheduler controls VM distribution across host memory to dynamically optimize CPU load & memory locality for VMs.
Firewall: Define service's port/protocol ruleset: /etc/vmware/firewall/service_*<name>*.xml (then refresh)
PAM (Pluggable Authentication Modules) plugins: /etc/pam.d/vmware-authd. Default password compliance plugin: pam_passwdqc.so. No restrictions on root password. Defaults for non-root users: password retries = 3, minimum password length = 8, shorter passwords if Characters Classes mixed (upper, lower, digits & other) 1 or 2 CC – min 8, 3 CC – min 7, 4 CC – min 6. First character as upper or last character as digit not counted.
DCUI (Direct Console UI): • Configures host defaults • Sets up administrative access • Troubleshooting. High contrast video mode `F4`. Can redirect DCUI to serial cable via Client or boot option (`Shift` + `O`).
Restarting Mgt agents effects /etc/init.d processes: hostd (mgmt-vmware), ntpd (time), sfcbd (CIM broker), slpd (discover/advertise services), wsman (share mgt info via SOAP), vobd (error reporting) & fdm (HA agent) if installed. To isolate ESXi host from DRS/HA cluster, disable management network.
Management Network Test: pings DG, primary DNS nameserver, secondary DNS, resolves hostname.
VIBs: can update image profiles or 3rd party extensions. Updates firewall ruleset & refreshes hostd.
Repair mode: On ESXi Installable CD, overwrites all configuration data. Serial number lost on repair, but restored when backup configuration applied. Configuration reset deletes root password, removes configuration & reboots host. Storage needs reconfigured & re-register VMs.
Recovery Mode: Invoked during boot with `Shift` + `R`. Reverts to previous image before last update.
SNMP agent embedded in hostd (disabled by default). Enable via `vicfg-snmp`. Can send traps & receive polling (GET) requests. **Syslog** service is vmsyslogd.
Host certificates: /etc/vmware/ssl/rui.crt (public key) & rui.key (private key).
Recreate: `/sbin/generate-certificates`
Lockdown mode: Forces operations via vCenter. Mode available only when host connected to vCenter. Enabling/disabling via DCUI wipes host permissions – set via vCenter. DCUI restricted to root, Shell & SSH disabled for all users, vSphere client & CIM monitoring only via vCenter not direct to host. Normal Mode: DCUI, Shell, SSH & CIM access allowed to root & Admin role users. vSphere Client access based on ESXi permissions. Total lockdown mode: also disables root access to the DCUI, if vCenter access is lost you must reinstall ESXi to regain control. root & vpxuser are only users not assigned No Access role on hosts by default, but have same rights as the Administrator role.

Links
Firewall Ports http://kb.vmware.com/kb/1012382
Location of ESXi 5.0 log files http://kb.vmware.com/kb/2004201
Video: Restarting management agents on an ESX/ESXi server http://kb.vmware.com/kb/1003490
Interpreting esxtop Statistics http://communities.vmware.com/docs/DOC-9279
Collecting diagnostic info using the vm-support command http://kb.vmware.com/kb/1010705
Decoding Machine Check Exception output after purple screen http://kb.vmware.com/kb/1005184

vCenter

Maxs
Maximums (per vCenter): Hosts = 1000 VMs = 15000 Running VMs = 10000 Clients = 100
MAC addresses = 65536 Datastore clusters = 256
Maximums (per datacenter): Hosts = 500
Maximums (Linked mode): vCenters = 10 VMs = 50000 Running VMs = 30000 Hosts = 3000

HW: Min - 2 CPU cores, 4GB RAM, 4GB disk space • Medium ≤50 hosts/500 VMs - 2 cores, 4GB RAM • Large ≤300 hosts/3000 VMs - 4 cores, 8GB RAM • Extra large ≤1000 hosts/10000 VMs - 8 cores, 16GB RAM
SW: • 64bit Win (2003 SP2/R2 SP1, 2008 SP2/R2) • 64bit DSN (*SQL Native* driver) • hostname ≤15 characters
Databases: • SQL 2008 Express (≤5 hosts/50 VMs) • SQL 2005 SP3 • SQL 2008 SP1 or R2• Oracle 10g R2 or 11g R1 • IBM DB2 9.5 fix pack 5 or 9.7 fix pack 2. VUM only supports Oracle & MS SQL.
User needs DBO rights. Default of max 50 simultaneous DB connections. MS SQL don't use master DB.

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

vCenter

vCenter Virtual Appliance (VCVA): Min 7GB disk, max 80GB. Supported DBs: • embedded (<5 hosts & <50 VMs) • Oracle. RAM sizing: ≥4GB for <10 hosts/100 VMs, ≥8GB for 10-100 hosts/100-1000 VMs, ≥13GB for 100-400 hosts/1000-4000 VMs, ≥17GB for >400 host/4000 VMs. Limits: no IPv6, Linked Mode, MS SQL, DB2. Default username: `root` default password: `vmware`

Extra vCenter DVD tools: • vSphere Update Manager (VUM) - needs 64bit OS but 32bit DSN • syslog server • ESXi Dump collector (no DVS support) - collects PSOD memory dump, useful for Auto Deploy host without local diagnostic partition • Authentication Proxy (see below) • Pre-Upgrade checker - checks for potential host issues • Auto Deploy server (see ESXi Install section) • Web Client server (see below)

Firewall

FW Port	Source	Destination	Protocol	Description
80	Clients	vCenter	TCP	Redirect to HTTPS
389	vCenter	Other vCenters	TCP	Linked Mode LDAP
443	Clients	vCenter	TCP	vSphere Client access
443	vCenter	ESXi	TCP	vCenter agent
902	ESXi	vCenter	UDP	Heartbeat
902	vCenter	ESXi	UDP	Host management, heartbeat
903	Clients	vCenter	TCP	VM Console

Possible extras: 25(SMTP), 53(DNS), 80/443/623(DPM), 88(AD), 161/162(SNMP), 636(Linked vCenters), 1433 (MSSQL), 1521(Oracle), 5988/5989(CIM), 6500/8000(Dump Collector), 8000(vMotion), 8080/8443/60099(webservices), 9443 (Web Client),10109/10111/10443(Inventory service),51915(Auth proxy)

Logs: DB upgrade: %TEMP%\VCDatabaseUpgrade.log vCenter agent: /var/log/vmware/vpx/vpxa.log
vCenter (Win XP, 2000, 2003): %ALLUSERSPROFILE%\Application Data\VMware\VMware VirtualCenter\Logs\
vCenter (Win 7, 2008): %ALLUSERSPROFILE%\VMware\VMware VirtualCenter\Logs\
(see KB in links below for description of different log files)
VCVA logs /var/log/vmware/vpx Windows Client Install %TEMP%\vmmsi.log
Windows Client Service %USERPROFILE%\Local Settings\Application Data\vpx\viclient-x.log (x=0-9)
Guest customization - Win: %WINDIR\temp\vmware-imc - Linux: /var/log/vmwareimc/toolsDeployPkg.log

Default roles (System roles - permanent, cannot edit privileges, ESXi & vCenter. Sample roles - just vCenter):

No access	System - Default except users in Admin Group. Cannot view or change.
Read only	System - View state & details except console tab.
Administrator	System - All privileges. Default for members of the Admin Group, & AD ESX Admins.
VM power user	Sample - Interact with, change VM HW settings, snapshots & schedule tasks.
VM user	Sample - Interact with, insert media & power ops. Cannot change VM HW settings.
Resource pool admin	Sample - Create, modify child pools & assign VMs, but not RP itself.
Datastore consumer	Sample - Allows space consumption of the datastore.
Network consumer	Sample - Allows hosts or VMs to be assigned to network.

Permissions: pair user/group with role & associate with object. Role - predefined set of privileges. Users initially granted *No Access* role on new objects including datastores/networks. Logged in users removed from domain keep permissions until next validation period (default 24 hrs). Tasks - activities that don't complete immediately. All roles allow schedule tasks by default. Can schedule tasks if user has permission when task created. vCenter Local Admins have Administrator role by default. Propagation is per permission, not universal. Child permissions override those propagated. User permissions override Group ones.
Use *No Access* role to mask areas from users. Moving objects needs permission on object, source & destination parent.

Licensing:	Essential	Essential+	Standard	Enterprise	Enterprise+
vRAM (per socket license)	----------	32GB	----------	64GB	96GB
vCPU	----------	8 way	----------		32 way
vpxa, Thin pro, VUM, VADP	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
vMotion, HA, vDR		**Yes**	**Yes**	**Yes**	**Yes**
SLES (SUSE Linux Ent Server) for VMware			**Yes**	**Yes**	**Yes**
DRS, DPM, Storage vMotion, FT, VAAI, Hot add, Linked mode, 3rd party MPP, Orchestrator, vShield Zones, Serial port concentrator	|>			**Yes**	**Yes**
DVS, NIOC, SIOC, Host Profiles, Auto Deploy, Policy-driven Storage, Storage DRS					**Yes**

vRAM - Memory configured on all powered-on VMs. Consumed vRAM capped at 96GB per VM. Only Essential & Essential+ has hard vRAM limit. CPU licenses from same license level are pooled across linked mode vCenters. Keys in vCenter not deployed add to entitlement. Add vRAM by adding licenses or upgrading existing. Consumed vRAM is 12 month average. Can create reports & alerts for consumed/entitled vRAM. Key assigned to host via vCenter is persistent. vSphere Hypervisor - free, no connection to vCenter , ≤32GB vRAM, only servers ≤32GB physical RAM, limited/read-only vCLI & PowerCLI support, no SNMP support.
vSphere Desktop - for VDI, functionality of Enterprise+ & unlimited vRAM. Per powered-on desktops.
Expiring licenses: vCenter - hosts are disconnected. ESXi - VMs continue to run, cannot power-on new VMs.

vCenter

Statistics: CPU, memory, datastore, disk, storage adapters/paths, network, power, DRS, HA, mgt agents, system & VM ops. Collection Intervals (time period stats archived in DB) frequency/retention is 5 mins - 1 day, 30 mins - 1 week, 2 hrs - 1 month, 1 day - 1 year. Real-time stats (just performance charts) flat file on hosts & vCenter memory (not in DB), frequency/retention is 20 secs - 30 mins, only powered-on hosts & VMs. Collection level 1-4 for each interval, most counters is 4 (default 1). Reports & Maps updated every 30 mins. VMware Tools adds perfmon objects to Windows guests.
Alarms: notifications of selected events, conditions & states. Composed of Trigger & Action. Triggers: condition, state or event. Actions: responses to triggered alarms. Can disable action without disabling alarm, but effects actions on all alarms. Disable for selected object, child continues. Default alarms not preconfigured with actions. Acknowledging alarm stops action, but alarm still visible. Reduce alarms with tolerance range & trigger frequency (default 5 mins). Disconnect hosts to suspend monitoring.
Linked mode: joins VCs. Global data: IP & ports, certificates. licensing, user roles. Uses ADAM (AD App Mode) to store & sync data. Instances can run under different domain accounts. Installed by domain user who is admin on both machines. Requirements: DNS, 2-way trust if different domains, time sync, DNS name matches hostname. Roles are replicated, assignments of roles are not.
Server settings: Licensing (vCenter & 3.x hosts), Statistics (intervals & DB size), Runtime Settings (unique ID, managed IP, name), AD (timeouts, query limit, validation period), Mail, SNMP receivers, Ports - http(s), client timeouts, Logging detail, DB connections (default 50), DB retention, SSL host verification, Advanced Settings
Host Profiles: Policy to centrally configure & check compliance of hosts. Set at host or cluster level. *Reference host* - host which created profile. Exported profile format .vpf. When profile is detached, settings persist on host/cluster. *Answer File* contains host specific input required by Auto Deploy - 1 per host. Host must be in Maintenance Mode to apply profile, Auto Deploy hosts need reboot.
Authentication Proxy: No AD credentials on ESXi, just domain name & proxy IP. Installer creates AD account prefixed with *CAM*. Authenticate proxy to ESXi by importing SSL certificate, or push via Host Profiles.
Web Client server: Alternative to Windows Client. Cross-platform & cross-browser (Adobe Flex plugin). Connects to vCenter (not to hosts directly), register Client server first. Subset of Windows Client functionality - Monitoring & VM deployment, no host/cluster/datastore/network configuration
Admin tool https://localhost:9443/admin-app Web client https://*<servername>*:9443/vsphere-client
Guest Customization: Guest OS must be on SCSI node 0:0. Requires Perl in Linux guests. Windows guest Admin password must be blank for customization to change it.
vService: Service dependency for vApps or VMs. vService Manager monitors health: • Red - issue needs fixed in solution (the extension) • Yellow - vService Manager is repairing • Green - OK

Links

Resolution Path Troubleshooting Licensing http://communities.vmware.com/docs/DOC-16082
Collecting diagnostic information for vCenter http://kb.vmware.com/kb/1011641
Location of vCenter Server log files http://kb.vmware.com/kb/1021804
Installing vCenter Server 5.0 best practices http://kb.vmware.com/kb/2003790
Upgrading to vCenter Server 5.0 best practices http://kb.vmware.com/kb/2003866
Sysprep file locations and versions http://kb.vmware.com/kb/1005593
Firewall Ports http://kb.vmware.com/kb/1012382
vCenter client shortcuts http://www.jume.nl/articles/vmware/143-vcenter-client-shortcuts
vSphere 5 Licensing, Pricing & Packaging http://www.vmware.com/files/pdf/vsphere_pricing.pdf

Cluster Resources

Max

Maximums (per DRS cluster): Hosts = 32 VMs (powered on) = 3000 (512 per host)
Maximums (per Resource Pool): Children = 1024 Tree depth = 8
Maximums (other): Hosts per datacenter = 500 RPs per host = 1600 RPs per cluster = 1600

Terminology: Datacenters - mark organizational & vMotion boundaries. Clusters - gather host resources. Resource Pools- apply policies to clusters. DRS cluster is implicitly a resource pool. Resources include CPU, memory, power, storage & networking. EVC (Enhanced vMotion) - masks CPU features that prevent vMotions. Storage DRS, Profiles & Datastore Clusters - see Storage section. NIOC & Network Resource Pools - see Networking section.

List resource group settings: `esxcfg-resgrp -l`

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

Cluster Resources

Resource pools: • Shares - low, normal, high & custom • Reservations - MHz(CPU)/MB(RAM) • Limits - MHz/MB • Expandable reservation - *yes* (can draw from parent's pool) - *no* (only from own pool)
Shares - only apply during contention. Shares are relative to siblings (VMs or Resource Pool). Reservations - guarantee a minimum, can be allocated more. Only checked when VM is powered on. Limits - upper bound, never exceeded; manage user expectations but can waste idle resources. Resource Pool Admission Control - prevents violations when VM is powered on or child pool created. Fixed reservations create strict isolation. Expandable reservations can borrow resources, don't automatically hunt upwards, but defines if admission control considers the reservation. More flexible but provides less protection. Child pools actively reserve resources from parent even if VMs are powered off. Hierarchical resource pools require DRS enabled.
DRS: • Manual • Partial (Initial VM placement) • Fully Automated (Initial VM placement & Dynamic balancing). Migration threshold slider sets allowable host load imbalance. *Current Host Load Standard Deviation* - load imbalance (higher number increases priority level). *Current* < *Target* unless recommendations unapplied. Priority levels 1-5 (1 is highest). *Grafted from* pools created when adding host to DRS cluster & keeping host's resource pool hierarchy. Maintenance mode only clears VMs off host if DRS cluster is fully automated. Disabling DRS deletes resource pools & affinity rules - set DRS to manual to keep settings. DRS can be overcommitted/yellow (host failure) or invalid/red (usually direct client changes).
Affinity Rules: VM-VM keep VMs together/apart. VM-Host keep VMs on/off specific hosts. *Should* rule is best effort. *Must* rule is mandatory (for licensing). Rule conflicts - older wins, newer rule disabled. Obeying anti-affinity ranks over affinity. Disabled rules ignored.
DPM: uses IPMI, iLO or WOL (in that order). DRS & DPM thresholds are independent. Verify host's *DPM Last Time Exited Standby*. DPM level: • Off • Manual (makes recommendations) • Automatic.

Links
DRS Deep Dive http://www.yellow-bricks.com/drs-deepdive
EVC (Enhanced vMotion Compatibility) FAQ http://kb.vmware.com/kb/1005764
EVC CPU compatibility http://kb.vmware.com/kb/1003212

VMs

Maxs
Maximums (per VM): vCPUs = 32 RAM = 1TB (64GB - FT VMs) Virtual swap file (.vswp) = 1TB
SCSI adapters = 4 Devices per SCSI adapter = 15 IDE devices (Disk/CD) = 4 VMDK = 2TB-512B
vNICs = 10 USB devices = 20 (USB 3.0 = 1) Floppy drives = 2 Parallel ports = 3 Serial ports = 4
Remote consoles = 40 VMDirectPath devices = 4 Video RAM =128MB

Files:

.hlog	vMotion log file	.vmsd	Snapshot metadata
.lck-XXX	Locking file on NFS datastore	.vmsn	Snapshot state file
.log	VM activity log	.vmss	Suspended state file
-#.log	Old VM log	.vmtx	Template header
.nvram	BIOS or EFI settings	.vmx	Primary configuration file
.rdm	RDM in Virtual Compatibility mode	.vmxf	Extra configuration file for VMs in a team
.rdmp	RDM in Physical Compatibility mode	.vswp	Swap file for overcommitted memory
.vmdk	Disk descriptor (also raw virtual disk for hosted products)		
-flat.vmdk	Raw pre-allocated virtual disk	-00000#.vmdk	Snapshot child disk
-ctk.vmdk	Changed Block Tracking file	-00000#-delta.vmdk	Snapshot differential file

Shell Commands
`--help` for esxcli namespaces & commands relative to location. `localcli` bypasses hostd
List running VMs: `esxcli vm process list`
List registered VMs (& displays <*vmid*>): `vim-cmd /vmsvc/getallvms`
Show VM's power state: `vim-cmd /vmsvc/power.getstate <vmid>`
Power on VM: `vim-cmd /vmsvc/power.on <vmid>`
Power off VM: `vim-cmd /vmsvc/power.off <vmid>`
Register a VM: `vim-cmd /solo/register /vmfs/volumes/vmname/vmname.vmx`
Unregister a VM: `vim-cmd /vmsvc/unregister <vmid>`
Forcibly kill VM: `esxcli vm process kill --type <soft|hard|force> --world-id <id>`
Create/Delete/Modify VMDKs, RDMs, VMFS volumes & storage devices: `vmkfstools`

Power Off = hard off • Shut Down = soft with VMware tools • Reset = hard • Restart = soft

VM HW: Memory/CPU Hotplug - VMware Tools required. Multicore requires HW v8. BIOS based VM min 4MB RAM, EFI min 96MB. Mac OS X VMs must run on Apple HW. CPU or Memory (NUMA) affinity not available in DRS clusters. vNUMA exposes host NUMA to guest OS. Guest swap ≥ (Configured vRAM – Reservation) x 65%, otherwise balloon driver could cause guest kernel panic.
HT sharing modes: • Any – vCPUs can share cores with other VMs • None – vCPUs have exclusive use when scheduled • Internal – can share core itself if VM has 2 vCPUs, not 2 vCPUs then same as None.
Disk types: • Thick Provision Lazy Zeroed - default, pre-allocates • Thick Provision Eager Zeroed - pre-allocates & zeros, better performance, slower creation • Thin Provision - allocates on-demand, monitor with "datastore usage" alarm. NFS - with HW acceleration supports all 3 types - without only Thin.

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

VMs

RDM: Can use SAN Snapshots, vMotion, SAN mgt agents & NPIV. Needs whole LUN. Physical RDMs no VM snapshots, clones, templates, only migrates mapping file. Virtual RDMs clones/templates copied into .vmdk
Snapshots: capture memory state, settings, disks. Can't snapshot physical RDMs or independent disks.
Independent Disk Modes: no snapshots. Persistent changes immediate & permanent. Nonpersistent changes lost on power-off or reset.
Snapshot Manager: Delete commits snapshot to parent. Delete all commits all snapshots before You are here. Go to reverts to that snapshot. Revert to snapshot back to parent's snapshot *You are here*.
VMDirectPath I/O: allows guest OS to access physical PCI/PCIe devices, sets VM memory reservation to vRAM. Requires VM HW v7 & Intel VT-d or AMD IOMMU. Restrictions no vMotion (can on Cisco UCS with Cisco DVS), FT, HA, DRS (cluster allowed, not VM), snapshots, hot add/remove, suspend, record/replay.
USB passthrough: Only 1 VM can connect to each device. Autoconnect uses physical port location. Supported: DRS, vMotion. Not Supported: DPM, FT. Initial connection when powering on/unsuspending must be local (pre-vMotion), to reconnect VM must be back on USB connected host.
SCSI controllers: • BusLogic Parallel • LSI Logic SAS • LSI Logic Parallel • PVSCSI (IDE is ATAPI)
PVSCSI (Paravirtual SCSI): at least HW v7, high-performance storage adapter. Not recommended for DAS. Guests: Win 2003, 2008, RHEL5. Not supported: Record/Replay, FT, MSCS, RHEL5 boot disks
NPIV (N-Port ID Virtualization): share FC HBA port as multiple virtual ports, each with unique IDs. VMs assigned 4 WWNs. Allows per-VM LUN access. Adds WWPN & WWNN to .vpx file. Limitations: requires NPIV enabled FC switch, only RDMs, Host HBA's WWNs also need access to LUN, NPIV capable HBAs, no Storage vMotion, VM can't power on if WWNs in use, vMotion requires all RDM files on same datastore.
vNICs: • Flexible - 32-bit guests, vlance without VMware Tools or vmxnet with VMware Tools • e1000 - Emulates E1000 NIC, default for 64-bit guests • vmxnet2 (Enhanced) - vmxnet with enhanced performance, requires VMware Tools • vmxnet3 - enhanced performance & networking features, requires VMware Tools & at least HW v7. WOL supported on vmxnet, vmxnet2 or vmxnet3.
MAC address: can manually assign in vmx: `ethernet<number>.addressType="static"` & `ethernet<number>.address=00:50:56:XX:YY:ZZ` (XX only 00-3F)
TSO (TCP Segmentation Offload): enabled in VMkernel by default, must be enabled at VM level. Needs enhanced vmxnet, might change the MAC. **Jumbo frames**: requires vmxnet2/3 or e1000.
OVF: templates imported from local file system or web server. OVF files are compressed. Client validates OVF file before importing. Can contain multiple VMs. OVA is single file version.
vApp: container containing one or more VMs, can power on & off, & be cloned. Metadata in VC's DB.
IP pool - network configuration assigned to network used by vApp. vCenter provides IPs to its VMs. Policies • Fixed – manual configuration • Transient – allocated from pool on vApp power on • DHCP

Links
Resolution Path – Troubleshooting VMs http://communities.vmware.com/docs/DOC-15963
Recreate missing virtual disk (VMDK) header/description file http://kb.vmware.com/kb/1002511
Consolidating snapshots in vSphere 5 http://kb.vmware.com/kb/2003638

Availability

Maximums (per HA cluster): Hosts = 32 VMs = 3000
Maximums (FT): Disks per VM =16 vCPUs per VM=1 RAM per VM = 64GB VMs per host = 4

Firewall Ports: HA interhost 2050-2250, 8042-8045 TCP/UDP, FT interhost 8100, 8200 TCP/UDP

HA: Single master, multiple slaves. If master fails, is shut down, or removed from cluster then an election occurs. All hosts not in standby, maintenance mode or disconnected participate in election. VMs on disconnected hosts are not protected. Master responsibilities: • monitors slaves • monitors VMs • restarts VMs on best host • reports cluster HA health to vCenter. Slaves: • monitors runtime state • reports state to master. Network heartbeat every second between master & slaves. Hosts can use all VMkernel networks for heartbeat except vMotion network (unless vMotion is on only available VMkernel network). If host stops responding to heartbeat, liveness checks datastore heartbeat & pings to management address. Host deemed Failed if both tests negative, if it passes liveness check it is Network Isolated (host cannot ping cluster isolation address) or Network Partitioned. Master monitors Network Isolated/Partitioned VMs and restarts them if they power off. Host Isolation Response: Action taken when Powered-on VMs are on isolated host. Options: • power off • leave powered on (default) • shut down (requires VMware tools). Cluster selects 2 datastores for heartbeat (*das.heartbeatdsperhost* can increase to 5). Datastores used for datastore heartbeat must be mounted by at least 2 hosts.
HA Admission Control: Can VMs power on when they violate availability constraints at HA failover. Actions that change a reservation must satisfy admission control. Secondary FT VMs added to calculation. Control policies: • Host Failures Cluster Tolerates - adds Advanced Runtime Info box showing slot size, total, used, available slots, total VMs on, hosts, good hosts • % of Cluster Resources (up to 50%) • Specify Failover Hosts (multiple allowed) - leaves host(s) unused until failure. Policy Factors: • resource fragmentation • flexibility • VM diversity. Slot size: represents VM CPU & memory resources needed for any powered on VM. Distorted by large VM reservations. Avoided with *das.slotCpuInMHz* or *das.slotMemInMB*.

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

Availability

FT: Uses anti-affinity rules to keep primary & secondary apart, checked when primary powers on. VM-VM affinity applies to primary only, VM-Host affinity applies to primary & secondary. Requires: HA & host monitoring, host certificate checking (on by default), dedicated logging NIC (not IPv6), compatible CPUs with same extensions & power mgt features, Hardware Virtualization (HV), thick disks on shared storage (VMDKs or vRDMs), supported guest OS, min 3 hosts in cluster.
Not supported: snapshots, Storage vMotion, Linked Clones, hotplugging, MSCS, VM backups, SMP, physical RDMs, PVSCSI, NPIV, VMDirectPath, 3D video, EPT/RVI is automatically disabled. DRS only if cluster has EVC. SiteSurvey can identify configuration issues. vLockstep Interval typically needs < 500ms.
vCenter Server Heartbeat: vCenter replication & clustering, monitors services including MSSQL, VUM & View Composer. Automated or Manual failover/failback. Uses Active/Passive nodes with heartbeat.
MSCS: • 2003 SP2 & 2008 (Failover Clustering) • 32 & 64bit • only 2 nodes clusters.
Not supported: DRS on VMs, vMotion, FT, NPIV, VMW_PSP_RR, FCoE/iSCSI/NFS based disks. SW iSCSI initiator in guest is supported.

	VMDK	Virtual RDM	Physical RDM
Cluster in a box (CIB)	Yes (zeroed)	Yes	No (not supported)
Cluster across boxes (CAB)	No	Only 2003	Yes (recommended)
Physical & VM (n+1)	No	No	Yes
Snapshots	Yes	Yes	No
SCSI target software	No	No	Yes

Configure all RDMs before configuring VM's network settings or initializing LUNs within windows. Add RDMs to 2nd SCSI controller i.e. SCSI(1:x). Set SCSI bus sharing: • CIB = Virtual • CAB or N+1 = Physical
NLB, Exchange CCR & DAG: does not use shared quorum disk, above restrictions not applicable.
SQL Mirroring: not considered clustering. Fully supported by VMware with no restrictions.

Links

HA Deep Dive http://www.yellow-bricks.com/vmware-high-availability-deepdiv/
HA and FT Error Messages http://kb.vmware.com/kb/1033634
CPUs & guest OSes that support FT http://kb.vmware.com/kb/1008027
MSCS Supported configurations (vSphere 4) http://kb.vmware.com/kb/1037959
MSCS Support on ESX/ESXi http://kb.vmware.com/kb/1004617

Networking

Maxs

Per host: 1GbE VMNICs = 2-32 dependent on HW | 10GbE VMNICs = 8 (or 6x10GbE & 4x1GbE)
PCI VMDirectPath devices = 8 | vSS/vDS ports = 4096 | Active ports (vSS/vDS) = 1016
Per vCenter: vDS switches = 32 | vDS port groups = 5,000(256 ephemeral) | vDS ports = 30,000
Per switch: Hosts (per vDS) = 350 | vSS port groups = 256 | vSS switch ports = 4,088

Terminology: VMNICs - logical name for physical server's NICs. vNICs - virtual NICs assigned to VMs. vSS - virtual Standard Switch. vDS - VMware's virtual Distributed Switch. DVS - can be vDS or 3rd party (e.g. Cisco 1000v). dvPort(Group) - port (group) on a vDS. dvUplink - uplink VMNICs on a vDS. Network vMotion - tracking of VM's network state on a DVS.

Shell Commands

`--help` for namespaces & commands relative to location. `localcli` bypasses hostd

List VMNICs:	`esxcli network nic list`
List vSwitches:	`esxcli network vswitch standard list`
List vDS:	`esxcli network vswitch dvs vmware list`
List vSwitch Port Groups:	`esxcli network vswitch standard portgroup list`
List VMkernel ports:	`esxcli network ip interface list`
List VMkernel interfaces:	`esxcli network ip interface ipv4 get`
List VMkernel Default Gateway:	`esxcfg-route`
List hostname:	`esxcli system hostname get`
List DNS servers:	`esxcli network ip dns server list`
List DNS search domain:	`esxcli network ip dns search list`

`esxcli` does not support configuring vDS dvPorts and dvUplinks: use `esxcfg-vmknic` with `dvs-name`, `dvport-id` & `esxcfg-vswitch` with `dvp-uplink` & `dvp` options

Ethernet tagging: • EST (External Switch Tagging) - Default. No trunking required. 1-1 relationship from VMNICs to physical (access) switch ports. Each VMNIC only sees 1 subnet. VLAN ID of 0 or blank. • VST (Virtual Switch Tagging) - Commonly used. VMNICs connected to a vSwitch can span several VLANs. Each Port Group has a VLAN ID of 1-4094. Set the VLAN ID to blank to use Native VLAN. • VGT (Virtual Guest Tagging) - Rarely used. Install 802.1Q trunking driver software in VMs, vSwitch keeps tags given by VMs. VLAN ID of 4095 on vSS, VLAN policy on vDS. Avoid VLAN ID of 1 - native Cisco VLAN ID. Use VLAN 4095 with promiscuous mode to sniff other port groups (IDS/packet sniffer).

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

Networking

Jumbo frames: MTU > 1500 up to 9000 bytes. Enable per vSS/vDS. vNIC must be vmxnet2/3 or e1000.
Link Discovery: vSS supports CDP (Cisco Discovery Protocol), vDS supports CDP or LLDP (Link Layer Discovery Protocol - 802.1AB). *Listen* (default), *Advertise* or *Both*.
PVLAN (Private VLAN): extension to VLAN standard to add further segmentation. Can reduce IP address wastage & solve VLAN ID limits. Not encapsulated. Primary PVLAN - Original VLAN divided into smaller groups. Secondary PVLAN - exists only within primary, has specific VLAN ID. Types: Primary is *Promiscuous* - connect with all VMs in primary. Secondary are *Community* - connect to themselves & VMs on promiscuous, or *Isolated* - connect with VMs on promiscuous.
NetFlow: Sends IP traffic records to collector for analysis. Traffic is intrahost, interhost or VM-physical.
Port Mirror: Mirror ports intrahost or interhost. Cisco's term is SPAN (Switch Port Analyzer).
NIOC (Network IO Control): prioritize egress traffic by type via dvUplink shares (low/normal/high-25/50/100) & host limits. Network Resource Pools: FT, iSCSI (not HW iSCSI), vMotion, Mgt, VR (SRM replication), NFS, VM, Custom (user defined). Supports 802.1p QoS priority tagging at MAC level.
TSO (TCP Segmentation Offload): enabled by default on VMkernel ports, allows very large frames (up to 64KB), even with smaller MTU. To enable VMs, use at least enhanced vmxnet vNIC.
NetQueue: enabled by default, allows certain VMNICs to spread processing across CPUs with multiple Rx queues, improves ingress performance.
vSS & vDS options: Options can be overridden on vSS & dvPortGroups. Individual dvPorts can override options, but dvPortGroups can disallow overrides.
Options nomenclature: • vSS - *Properties* • vDS/dvUplinks - *Settings* • dvPortGroups - *Policies*.
General • *Number of uplinks* (vDS only) • *Number of ports* - vSS default - 120, dvPortGroup - 128 • *Port Binding* (dvPortGroups only): *Static* - when initially connected, *Dynamic* - when connected/powered-on, *Ephemeral* - no binding. Host can assign port if vCenter is down. • *MTU* - default 1500 (cannot override on Port Groups) see Jumbo Frames above • *Discovery Protocol* (vDS only) see Link Discovery below • *VLAN ID* (vSS PGs only)
Network Adapters (vDS only) • Host to dvUplinks mapping
Private VLAN (vDS only) • Primary to Secondary mapping
Netflow (vDS only) • Collector *IP Address* & *Port* • *vDS IP Address* - so collector interacts with vDS not hosts • *Active flow export timeout* • *Idle flow export timeout* • *Sampling rate* - 1 packet collected per sampling rate • *Process internal flows only* - just intrahost traffic.
Port Mirroring (vDS only) Add session to mirror • *Allow normal IO on destination ports* - port to receive normal IO as well as mirrored traffic • *Encapsulate VLAN* - create VLAN ID to encapsulate all frames if destination is an uplink port. If *Preserve original VLAN* unselected then if VLAN is present then it's replaced not encapsulated • *Mirrored packet length* - limits size of mirrored frames • select *Ingress/Egress* • select *Port IDs* or *Uplink source & destination*.
Security • *Promiscuous mode* (default Reject) - only listens to traffic destined for its MAC address.
• *MAC Address Changes* (default Accept) - accepts inbound frames when VM changes MAC address.
• *Forged Transmits* (default Accept) - won't drop outbound frames if source MAC address is different
Traffic Shaping • *Status* (default Disabled) • *Average Bandwidth* (Kbps) • *Peak Bandwidth* (Kbps)
• *Burst size* (KB). vSS can shape outbound traffic, vDS can shape traffic in & out (Ingress/Egress)
VLAN (dvPortGroup only) • *None* - access port • *VLAN* - set ID • *Trunk range* - restrict IDs on trunked links • *PVLAN*
Teaming & Failover • Load Balancing - spreads outbound traffic from vNICs across VMNICs/dvUplinks, incoming traffic is load balanced by physical switch. *Originating port ID* (default) - uses uplink based on where traffic entered. *ip hash* - based on source & destination IP address of each packet (use if physical switch ports are etherchannel). *Source MAC hash* - based on source MAC address. *Route based on physical NIC load* (vDS only) - based on current loads on dvUplinks. *Use explicit failover order* - uses first active uplink in list. • Network Failover Detection - *Link status only* (default) - detects cable pulls & switch power failures, not misconfigurations. *Beacon Probing* – can also detect some misconfiguration, don't use with IP-hash load balancing & not supported with VGT. • Notify Switches - *No* or *Yes* (default) updates lookup tables. Disable for MS NLB in unicast mode. • Failback - *No* or *Yes* (default) uplink returns after recovering from failure. • Failover order - *Active*, *Standby* or *Unused* - Don't use standby uplinks with IP-hash load balancing.
Resource Allocation (dvPortGroup only) • select Network Resource Pool (see NIOC)
Monitoring (dvPortGroup & dvUplink only) - *Enable* or *Disable (default)* NetFlow (options on vDS)
Miscellaneous (dvPortGroup & dvUplink) • Port blocking - *No* (default) or *Yes* - shut down all ports.

Links: Troubleshooting Networking http://communities.vmware.com/docs/DOC-9876

Storage

Maxs

Maximums (per host): Virt disks = 2048 LUNs/Volumes = 256 Paths = 1024 NAS mounts = 256
FC HBAs=8 (ports=16) Targets per HBA=256 Paths per LUN=32 LUN size=64TB FCoE Adapters=4
iSCSI HW HBAs = 4 Targets per HBA = 62-128 (depends on card) Paths to each LUN = 8
iSCSI SW NICs = 8 Targets = 256 Paths to each LUN = 8
Maximums (per volume): Powered-on VMs = 2048 Hosts = 64
VMFS5 = 64TB vRDMs = 2TB (less 512B) pRDMs = 64TB File size = 2TB (less 512B)
Maximums (per datastore cluster):Virt disks=9000 Datastores=32 (datastore clusters per vCenter=256)

Firewall Ports: iSCSI – 3260 TCP, NFS – 111 TCP/UDP, 2049 TCP/UDP

Shell Commands

`--help` for esxcli namespaces & commands relative to location. `localcli` bypasses hostd
Rescan SCSI HBAs (devices, paths, claimrules, FS): `esxcli storage core adapter rescan`
List all SCSI paths: `esxcli storage core path list`
Map VMFS volumes to devices/partitions: `esxcli storage filesystem list`
List unresolved snapshot/replicas of volumes: `esxcli storage vmfs snapshot list`
SATP claiming rules: `esxcli storage nmp satp rule list`
List nmp devices with SATP & PSP: `esxcli storage nmp device list`
List all claim rules: `esxcli storage core claimrule list`
List storage devices with properties/filters: `esxcli storage core device list`
Lists HBA drivers & information: `esxcli storage core adapter list`
Show each device's VAAI support: `esxcli storage coredevice vaai status get`
List FCoE HBA adapters: `esxcli fcoe adapter list`
List FCoE CNAs: `esxcli fcoe nic list`
List iSCSI adapters: `esxcli iscsi adapter list`
Show current iSCSI session: `esxcli iscsi session list`
Discover iSCSI devices: `esxcli iscsi adapter discovery rediscover`
Check if software iSCSI is enabled: `esxcli iscsi software get`
List the NFS filesystems & mounts: `esxcli storage nfs list`
Test VMkernel connectivity: `vmkping [-s 9000] <ipaddress>`
SCSI performance statistic tool: `vscsiStats`
Create/Delete/Modify VMDKs, RDMs, VMFS volumes & storage devices: `vmkfstools`

Array types: Active-Active - IO to all LUNs simultaneously through all ports/SPs, without performance degradation. Active-Passive - one port actively provides access, others are backup for that LUN (but can be active for other LUNs). Path thrashing can occur. ALUA (Asymmetric Logical Unit Access) - on non Active-Active arrays, paths are not available or not optimized to every port on every SAN SP. ALUA support on a SAN helps the hosts find/manage the best paths for failover & load balancing. Virtual Port (iSCSI only) - SANs are Active-Active but mask all connections, ESXi accesses all LUNs via single virtual port. The SANs handle multipathing. **Multipathing**: path failover (redundancy) & load balancing.
Claim rules: specifies which MPP (MultiPathing Plugin), native or 3rd party, manages physical path. Rules based on SAN discovered, listed in /etc/vmware/esx.conf. Path evaluation every 5 mins.
NMP (Native MPP): Secondary rules applied for: • SATPs (Storage Array Type Plugins) - handles failovers for array type. Rules search order: drivers, then vendor/model, lastly transport, *VMW_SATP_DEFAULT_AA* default if array not recognized. • PSPs (Path Selection Plugins) - handles load-balancing for each device.
PSPs policies: *VMW_PSP_FIXED* - default for active/active, uses preferred path (marked with *) when available, default policy if device not recognized. *VMW_PSP_MRU* (Most Recently Used) - default for active/passive, if SATP is VMW_SATP_ALUA then active/optimized path used, if not ALUA then uses first working path found at boot. *VMW_PSP_RR* (Round Robin) - safe for all arrays, rotates through paths, not MSCS LUNs.
Resignaturing: VMFS datastores have unique UUID in file system metadata. Replicated or snapshotted disks keep same UUID. Can mount with existing signature or assign new signature. Resignaturing assigns new UUID & label, then mounts. If not resignaturing, original must be offline.
Rescans: datastore rescans are automatic for many operations. Manual rescans may be required: zoning changes, new SAN LUNs, path masking changes on host, cable reconnected, changed CHAP settings, adding /removing iSCSI addresses, re-adding hosts. Rescans LUN 0 to LUN 255. *Disk.MaxLUN* reduces number of LUNs scanned to increase boot times & rescans. If LUN IDs are sequential disable sparse LUN support.
Zoning: at the switch. **LUN masking**: mask certain LUN IDs at Array's SP or ESXi host using claim rules.
PDL (Permanent Device Loss): if LUN is being removed, detach it so volume is unmounted. Host may detect SCSI sense codes to determine LUN is offline & it is not a APD.
APD (All Paths Down): No active paths to storage device. Unexpected so host continually retries paths.
LUN queue depth: SCSI device driver parameter that limits number of commands a LUN can accept. Excess commands are queued in VMkernel. Increase queue depth if VMs' commands consistently exceeds queue depth. Procedure depends on host adapter. Setting higher than default can decrease number of LUNs supported. Change *Disk.SchedNumReqOutstanding* to match - it limits requests each VM can issue.

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie

Storage

LUN Device IDs: • SCSI inquiry – returned by device , unique across hosts, persistent, T10 standard e.g. naa.#, t10.# or eui.# • Path-based – not unique, not persistent, e.g. mpx.path • Legacy – created in addition to SCSI inquiry or Path-based, e.g. vml.# • Runtime name – host specific, not persistent, first path to device, adapter:channel:target:LUN, e.g. vmhba#:C#:T#:L# **FCoE**: Interfaces: - CNA (Converged Network Adapter) or NIC with partial offload & SW initiator. Disable STP (Spanning Tree Protocol) - might delay FIP (FCoE Initialization Protocol). Enable PFC (Priority-based Flow Control) & set to AUTO.
iSCSI: Interfaces: • iSCSI HBA (independent HW) • NIC with iBFT/iSCSI offload (dependent HW) & SW initiator • Regular NIC & SW initiator. Only 1 SW initiator per host. Independent HW: configured in Storage configuration. Non-independent configuration: 1 VMkernel interface to 1 active NIC, others unused, bind adapters. Set "MAC Address Changes" PG policy to *Accept*. SW initiator enabled by default. Boot from iSCSI SAN: only Independent HW LUN installs get diagnostic partition. iSCSI Nodes: • IP address • iSCSI name (IQN e.g. iqn.yyyy-mm.reversed_domain_name:string or EUI e.g. eui.string) • iSCSI alias – not unique, friendly name. iSCSI Discovery methods: • Dynamic - uses *SendTargets*, target responds with list. Targets added to Static list, removed targets can return after HBA rescan/reset or host reboot • Static - can manually add/remove items. iSCSI SAN access control: • Initiator name • IP addresses • CHAP. CHAP authentication: • 1-way (unidirectional) target authenticates initiator (set on SAN) • Mutual (bidirectional) target & initiator can authenticate each other – only SW iSCSI or dependent HW cards. CHAP character limits: • only alphanumeric (no symbols) • CHAP name ≤ 511, CHAP secret ≤ 255
NFS: Not supported: Access via non-root credentials (delegate user).
SIOC (Storage IO Control): shares VM's disk IO across datastore's hosts. Monitors latency, adjusts VM's host queue access. Can also enforce VM IOPS limits. Enable on datastore, set shares/limit on VM. Limit must be set on all VM's disks. Datastores must be managed by single vCenter. Not supported – RDMs, datastores with multiple extents. Auto-tiering arrays must be certified for SIOC. Enabled by default on Storage DRS enabled datastore clusters. Congestion Threshold is upper limit for datastore before SIOC allocates based on shares. Low threshold = lower device latency & strong VM IO isolation. High threshold = high throughput & weak isolation. Default 30ms, max 100ms, min 10ms.
Datastore Cluster: Manage multiple datastores as one logical resource. Can be VMFS or NFS, but not both in same cluster. Can't combine SRM replicated & non-replicated datastores in same cluster.
Recommendations: datastore of similar size & IO capabilities in cluster, don't mix HW accelerated & non accelerated backed datastores.
Datastore Maintenance Mode: automatically evacuates all VMs from a datastore. Affinity rules can prevent entering maintenance mode, to override use setting *IgnoreAffinityRulesForMaintenance* = 1.
Storage DRS: Load balancing & Initial placement of VMDKs to balance I/O & spread free space. By default, IO evaluated every 8 hrs, IO latency threshold is 15ms. Datastores used by multiple datacenters ignored by IO balancing (space balancing still used). All datastore connected hosts must be ESXi 5. Settings are preserved if feature is disabled. Storage DRS & SRM not aware of each other.
Storage DRS Automation Levels: • Manual (default) • Fully Automated • Disabled.
Aggressiveness thresholds: • Space Utilization • IP Latency. Advanced options: • Space Utilization difference – ensures minimum difference between source & destination • IO Load Balancing Invocation Interval • IO Imbalance Threshold. Create scheduled task to change automation or aggressiveness level so migrations more likely off-peak.
Storage DRS affinity rules: • VMDK Anti-Affinity (Intra-VM) - VM with multiple disks are split across datastores • VMDK Affinity - VM's disks kept together • VM Anti-Affinity (Inter-VM) - disks from specified VMs (only 2) are kept on apart. By default VM's disks kept together. Anti-affinity not enforced for user initiated migrations & not applicable to ISOs or swap files.
Storage Profiles: Set the requirements of the VM home files & disks.
HW acceleration: (enabled by default) offloads operations to supported arrays. Faster, consumes less host resources, reduces fabric bandwidth usage. Block: *Full copy* (or clone blocks or copy offload) - array copies data without host read/write. *Block zeroing* (or write same) – array zeros out blocks of newly allocated storage. *Hardware assisted locking* (or Atomic Test & Set (ATS)) - discrete VM locking avoiding SCSI reservations, allows disk locking per sector instead of LUN NFS: *File clone* - similar to VMFS File Copy except entire files are cloned instead of file segments. *Reserve space* - arrays can allocate space for thick format disks. *Extended file statistics* - accurate reporting of space usage of VMs.
Storage Capability: system or user defined. Outine of datastore's capacity, performance, availability, redundancy, etc.
Storage vMotion: supports vSphere snapshots and linked clones. Uses Mirror Mode driver.
SSD: Can be used for host's VM swapping, very high IO for increased consolidation, guest OS can recognize it as SSD. Use SATP claim rule to tag SSD if not recognized automatically.
NPIV (N-Port ID Virt): FC HBA port assigns dedicated VPORT to VM as WWN pair - see VM section

Links
Resolution Path – Troubleshooting Storage http://communities.vmware.com/docs/DOC-16708
NetApp vSphere Storage best practices whitepaper http://media.netapp.com/documents/tr-3749.pdf
File System alignment whitepaper (NetApp) http://media.netapp.com/documents/tr-3747.pdf
vSphere handling of LUNs detected as snapshot http://kb.vmware.com/kb/1011387

vReference vSphere 5.0 v1.0.3 for 5.0 Released: 15 Apr 2012
Forbes Guthrie ∞ Web: vReference.com ∞ Twitter: @forbesguthrie