The Linux opportunity on Azure

Michael Li

Linux on Azure

Microsoft is committed to heterogenous datacenters

- You run Windows
- You run Linux
- You run FreeBSD
- You protect your freedom to run whatever is needed

Linux and FreeBSD work great on Hyper-V and Azure

- It's real
- It's robust
- It's available today
- It has the capabilities you want

Linux on Azure

Linux IaaS 8.7% MOM
- Continues to grow faster than the Windows counterpart
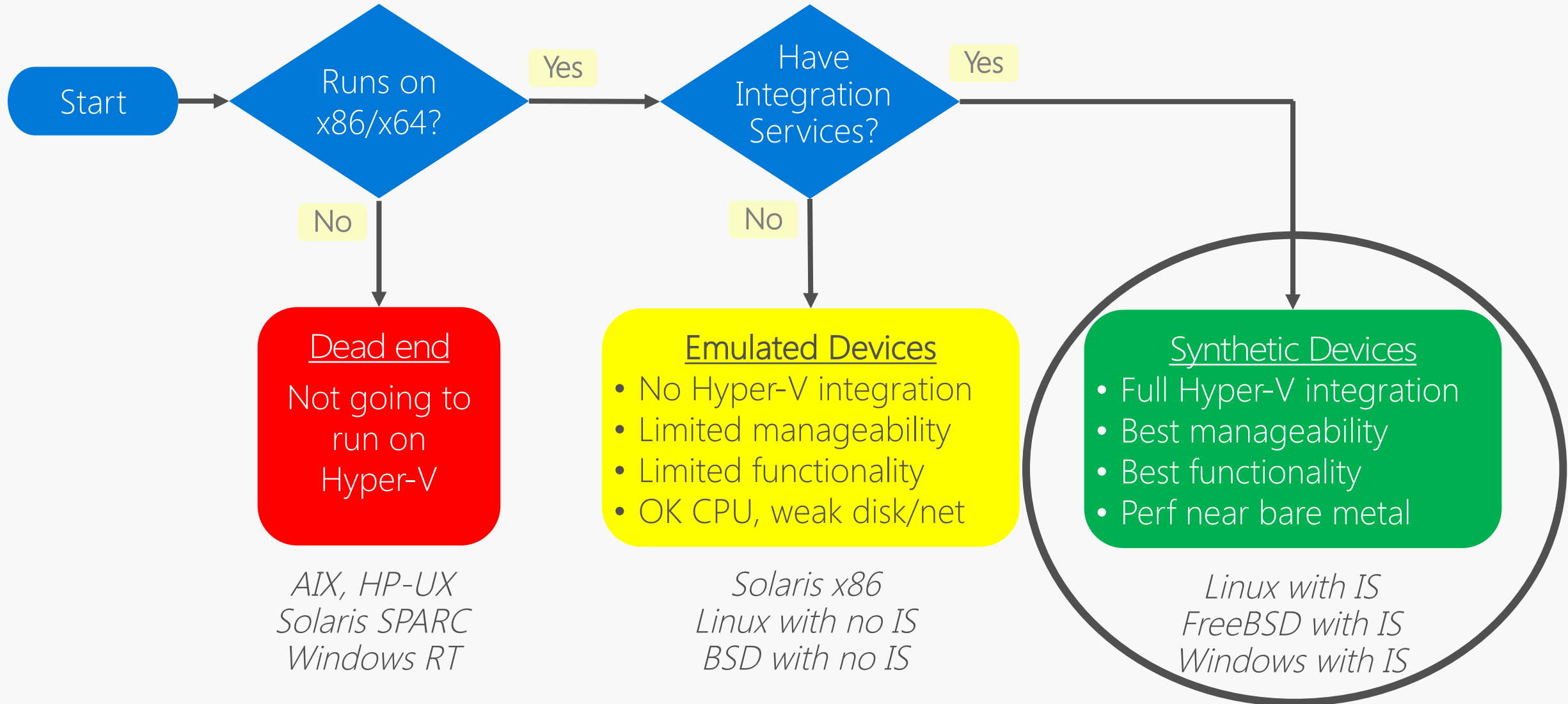
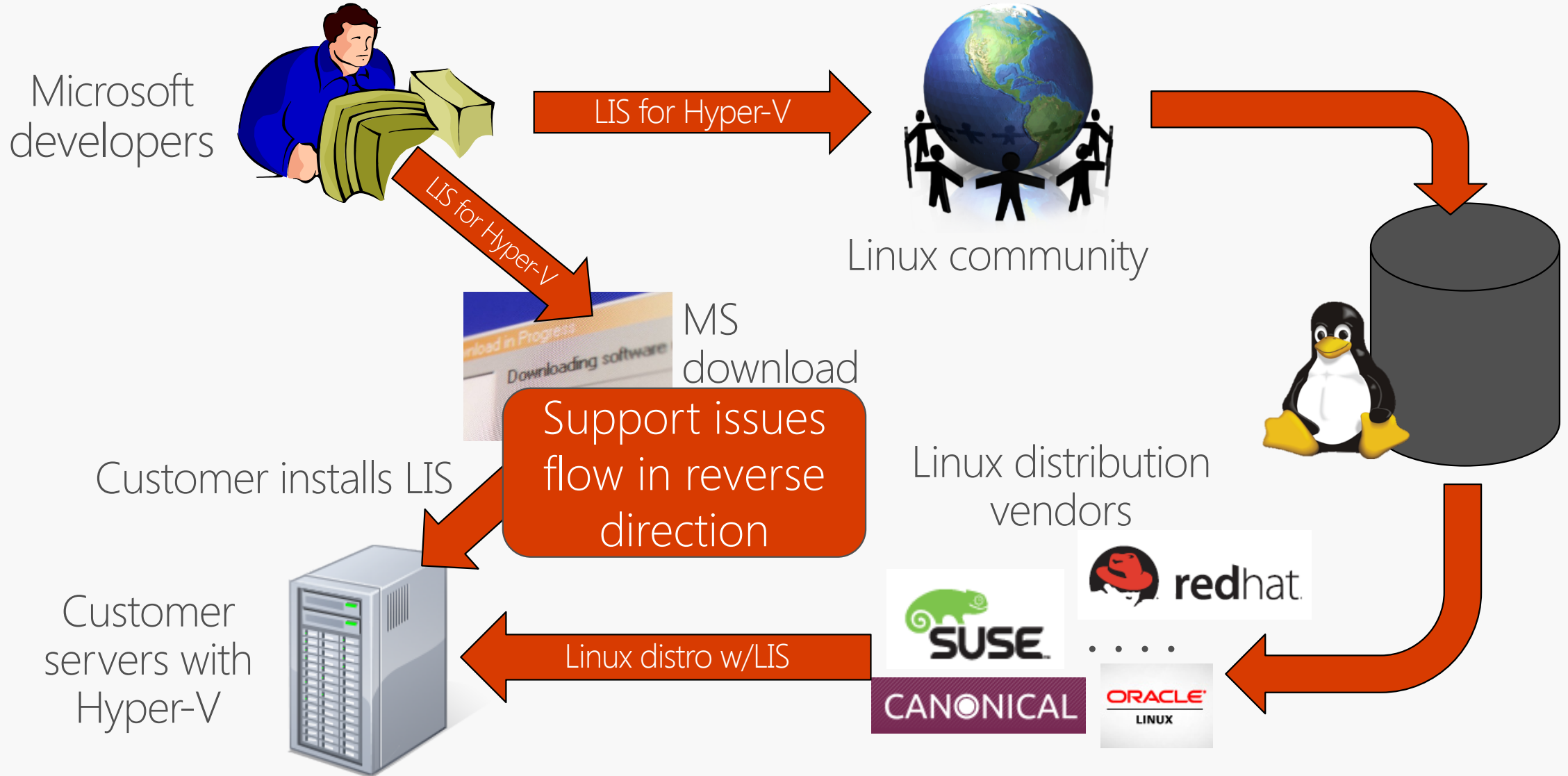Linux VMs ~33% of Azure VMs
- Goal is to trend towards 50%

Azure

♥

Linux

Running a Guest OS on Hyper-V

Start → **Runs on x86/x64?**

— Yes → **Have Integration Services?**

— Yes → **Synthetic Devices**

— No (from Runs on x86/x64?) → **Dead end**

— No (from Have Integration Services?) → **Emulated Devices**

Dead end
Not going to run on Hyper-V

AIX, HP-UX
Solaris SPARC
Windows RT

Emulated Devices
- No Hyper-V integration
- Limited manageability
- Limited functionality
- OK CPU, weak disk/net

Solaris x86
Linux with no IS
BSD with no IS

Synthetic Devices
- Full Hyper-V integration
- Best manageability
- Best functionality
- Perf near bare metal

Linux with IS
FreeBSD with IS
Windows with IS

Linux Hyper-V & Azure Lifecycle

Microsoft developers

LIS for Hyper-V

Linux community

LIS for Hyper-V

MS download

Support issues flow in reverse direction

Customer installs LIS

Linux distribution vendors

redhat

SUSE

Customer servers with Hyper-V

Linux distro w/LIS

CANONICAL

ORACLE LINUX

Linux on Azure

Distro	Number of VMs	Number of cores
Ubuntu	70%	61%
OpenLogic CentOS	19%	22%
Red Hat	6%	10%
Debian	3%	2%
CoreOS	1%	2%
SLES + openSUSE	1%	2%
Oracle Linux	1%	1%

RS3 Focus Areas

- Performance parity with Windows VMs on Azure
- GPU partitioning support for Linux
- Hyper-V Containers

Linux VM Network throughput issues

- Throughput against the Linux VMs was observed to be 10-30% worse than with Windows VM.
- This was root-caused to packet drops due to Linux guest issue, where the guest infrequently is taking significantly long amounts of time to complete receive indication, hence leading to NIC buffer exhaustion.
  - EOSG team will look further into these Linux VM specific issues.
  - Server Performance and Network Virtualization team would look into overall improvements of Hyper-V networking stack for Azure to improve VM Networking perf (both Windows and Linux VM).

VM Type	Linux (4.4 no NAPI/RSC)	Windows	Linux (4.12 with NAPI/RSC)
Throughput	17 - 21 Gbps	28 Gbps	28 Gbps
Retransmits/sec.	~10K/sec.	~300/sec.	~300/sec.
Host RSS CPU utilization	70-80% of 10 CPUs	100% of 10 RSS CPUs	100% of 10 RSS CPUs
Limiting factor	*Underutilized receiver due to packet loss.*	Host RSS Core utilization.	Host RSS Core utilization.
Guest VP usage (8VP)	40%	35%	35%
Hypervisor CPU usage	5.2%	1.7%	4.5% *

(*) Most of the Hypervisor difference can be attributed to much higher number of IPI activity in Linux Guest. Digging deeper will require Linux OS/App analysis. We suspect it is likely *TryLocal* is helping Windows to have less cross processor scheduling events.

- Next Steps
  - [Bug 12334157](): Root Cause Analysis of why Linux taking longer for completion with VMBus Send Completion in Multi-Connection case (long tail on Kernel 4.4)
  - [Bug 12334142](): Root Cause Analysis in to Packet Drop by Linux VM for single connection traffic

Low queue depth sync IO issue in Linux

Scheduler and Hyper-V cost differences

- Windows VM has DPC and IoMeter worker thread run in the same VP. Linux VM could run in two different VPs.
  - KY: To find out the reason and consider possible improvement (details [here](#)).
- ~1.3% HLT instructions cost in Linux VM, but not in Windows VM.
  - Windows VM uses a guest idle enlightenment.
  - KY: Should Linux VM take the enlightenment to remove the difference?

4KB IO Size, 1 QLEN	RS1 VM (IOPS)	Linux VM (IOPS)	Linux VM vs. RS1 VM (%)	IoMeter affinitized to Core1 (IOPS)	Linux VM with IoMeter Affinity vs. RS1 VM (%)	Fixed Linux VM (IOPS)	Fixed Linux VM vs. RS1 VM (%)
Seq Read	38,266	27,969	-27%	32,510	-15%	34,267	-10%
Seq Write	35,146	27,359	-22%	31,520	-10%	33,632	-4%
Rand Read	17,757	15,536	-13%	16,939	-5%	16,966	-4%
Rand Write	35,607	27,602	-22%	31,615	-11%	33,007	-7%

- ~350x number of MSR writes in Linux VM to HV_X64_MSR_STIMER0_COUNT
  - KY: Changed the MSR writes to use periodic timers; which dramatically improved the performance even further; but still needs further evaluation with other workloads before it can be considered as a mainline check-in.
  - Base: Any further optimizations for per timer cost?

Next Steps
- Agree on test collateral to use for continued monitoring of benchmark results (Alex),
- Discuss ongoing monitoring and strategy to identify future performance opportunities (KY/Mkaram)

References

Comparative Storage Performance on Low Queue depth between Linux & Windows VMs on Hyper-V - review deck:

https://microsoft.sharepoint.com/teams/SrvPerf/Shared%20Documents/Forms/AllItems.aspx?id=%2Fteams%2FSrvPerf%2FShared%20Documents%2FProjects%2FLinux%20VM%20Perf%2FStorage%2FLinuxVMStorageStatus%5FWrapUp%5F5%5F24%2Epptx&parent=%2Fteams%2FSrvPerf%2FShared%20Documents%2FProjects%2FLinux%20VM%20Perf%2FStorage

Single IO Performance of Linux VM on Hyper-V – Performance Report:

https://microsoft.sharepoint.com/teams/SrvPerf/Shared%20Documents/Forms/AllItems.aspx?id=%2Fteams%2FSrvPerf%2FShared%20Documents%2FProjects%2FLinux%20VM%20Perf%2FStorage%2FLinuxVMStoragePerfReport%5F0517%2Edocx&parent=%2Fteams%2FSrvPerf%2FShared%20Documents%2FProjects%2FLinux%20VM%20Perf%2FStorage.

Linux vs. Windows VM Networking Performance on Azure – Performance Report Working draft:

https://microsoft.sharepoint.com/teams/SrvPerf/_layouts/15/WopiFrame.aspx?sourcedoc={09cadc7a-5cd4-4274-8016-b6d6524aafa0}&action=edit

RS4 & Beyond

- Performance parity with Linux VMs on AWS
  - challenges on how to disambiguate infrastructure variance between Azure and AWS
- Continue work on Hyper-V Linux Containers
  - PODS support
- Large VM support

Linux Distro Status: RHEL and CentOS

Red Hat Enterprise Linux (RHEL)			
Version	32/64?	LIS Built-in?	LIS download?
5.2 thru 5.4*	32 & 64		LIS 4.2
5.5 thru 5.8	32 & 64		LIS 4.2
5.9 thru 5.11	32 & 64	Yes	LIS 4.2
6.0 thru 6.3	32 & 64		LIS 4.2
6.4 thru 6.7	32 & 64	Yes	LIS 4.2
6.8	32 & 64	Yes	LIS 4.2
6.9	32 & 64	Yes	
7.0 thru 7.3	64 only	Yes	LIS 4.2

CentOS			
Version	32/64?	LIS Built-in?	LIS download?
5.2 thru 5.4*	32 & 64		LIS 4.2
5.5 thru 5.8	32 & 64		LIS 4.2
5.9 thru 5.11	32 & 64	Yes	LIS 4.2
6.0 thru 6.3	32 & 64		LIS 4.2
6.4 thru 6.7	32 & 64	Yes	LIS 4.2
6.8	32 & 64	Yes	LIS 4.2
6.9	32 & 64	Yes	
7.0 thru 7.3	64 only	Yes	LIS 4.2

RHEL versions with LIS built-in are certified by Red Hat for running on Hyper-V, so you get the full benefits of your RHEL subscription.

Linux Distro Status: Ubuntu and SLES

Ubuntu Server

Version	32/64?	LIS
12.04	32 & 64	Built-in
12.10* 13.04* 13.10*	32 & 64	Built-in
14.04	32 & 64	Built-in
14.10* 15.04* 15.10*	32 & 64	Built-in
16.04	32 & 64	Built-in
16.10 17.04	32 & 64	Built-in

SUSE Linux Enterprise Server (SLES)

Version	32/64?	LIS
11 SP2*	32 & 64	Built-in
11 SP3*	32 & 64	Built-in
11 SP4	32 & 64	Built-in
12*	64 only	Built-in
12 SP1*	64 only	Built-in
12 SP2	64 only	Built-in

* "End Of Life" releases no longer in support

Linux Distro Status: Oracle Linux and Debian

Oracle Linux			
Version		*32/64?*	*LIS*
Red Hat Compatible Kernel	6.4-6.8	32 & 64	LIS 4.2
	6.9	32 & 64	
	7.0-7.2	64 only	LIS 4.2
	7.3	64 only	LIS 4.2
Unbreakable Enterprise Kernel	UEK3	32 & 64	Built-in
	UEK4	64 only	Built-in

Debian Server		
Version	*32/64?*	*LIS*
7.0 thru 7.11*	32 & 64	Built-in
8.0 thru 8.8	32 & 64	Built-in

Debian 8.3 is the first Debian with hyperv-daemons (KVP, etc.) in main