



VM Worker Node Performance Testing

Alden Stradling
11 October 2011

Choice Point

- Many options for VM hypervisors and client machine creation
 - Hypervisors like XEN, KVM, VMware, VirtualBox
 - VMs can be of various provenance, including homegrown of various flavors and CernVM
- Choice for a way forward in virtualized WNs is complex



Simplify the Space

- Parameters: Performance, cost and support load.
- Based on the conclusions of Yushu Yao in April 2010, focus on file-based VMs instead of paravirtualized
- Focus on KVM and VMware as primary options
- Test performance, then weigh against support load and TCO.



Performance Metrics

- Athena jobs
- Analysis code runs
- Disk performance
- Network IO performance



Caveats

- As with all multi-core systems, the comparative disk and bus speeds will vary by the number of cores/disk and cores/northbridge.
- These tests were done on 1-4 cores to *avoid* seeing this effect. As with all systems, measures will need to be taken to mitigate these problems at a hardware level.
- Multiple disks and SSDs are possible solutions to avoid these bottlenecks, but the VM itself will offer limited solutions to this issue.



VM Configurations

- Giant resources unnecessary for the VM test. All on cvmfs
 - CernVM 2.40 + KVM: 2 GB RAM, 4 cores, 2.27 GHz (R410)
 - CernVM 2.40 + VMware Enterprise Server 4.1u1 (vSphere ESX host), 4GB RAM, 4 cores, Intel X5670 @ 2.93GHz
 - Homebrew + KVM: 2 GB RAM, 4 cores, 2.27 GHz (R410), SLC5.7
 - Native R410: 24 GB RAM, 16 cores, Intel E5520 @ 2.27 GHz, SL5.4



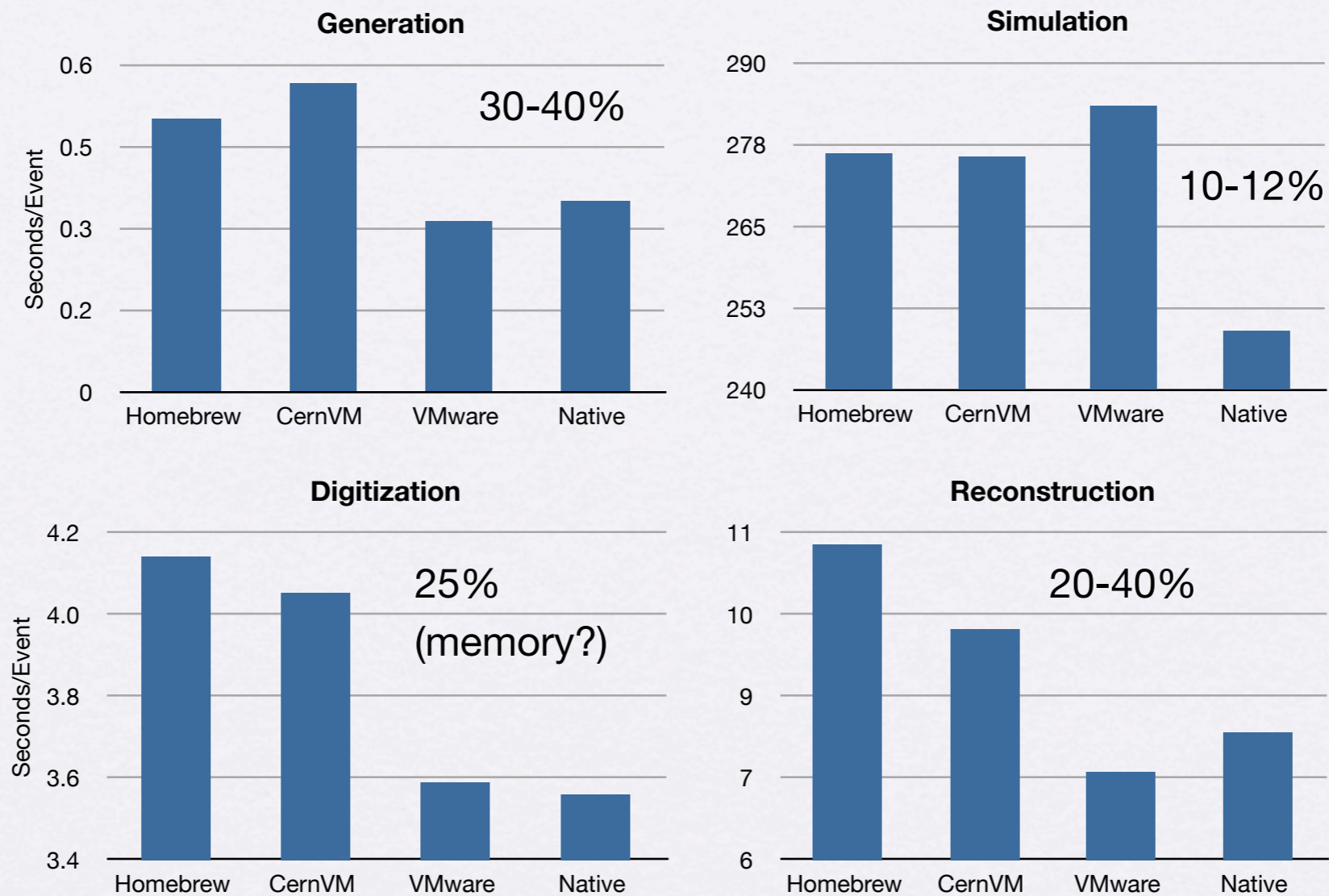
Athena

- Used 17.0.4 as a basis for the testing
- Generation, simulation, digitization, reconstruction comparisons for 1, 2, 3 and 4 cores
- Used the Atlas standard recipes from the workbook. More or less exotic tests were not attempted at this time
- Used a basis of 124 events, 3 trials, averaged. Multiple simultaneous jobs (for simultaneous CPU loads) also averaged
- Simple correction factor applied for the VMware CPU difference
- cvmfs was pre-charged with a 1-event run to remove its caching as a factor



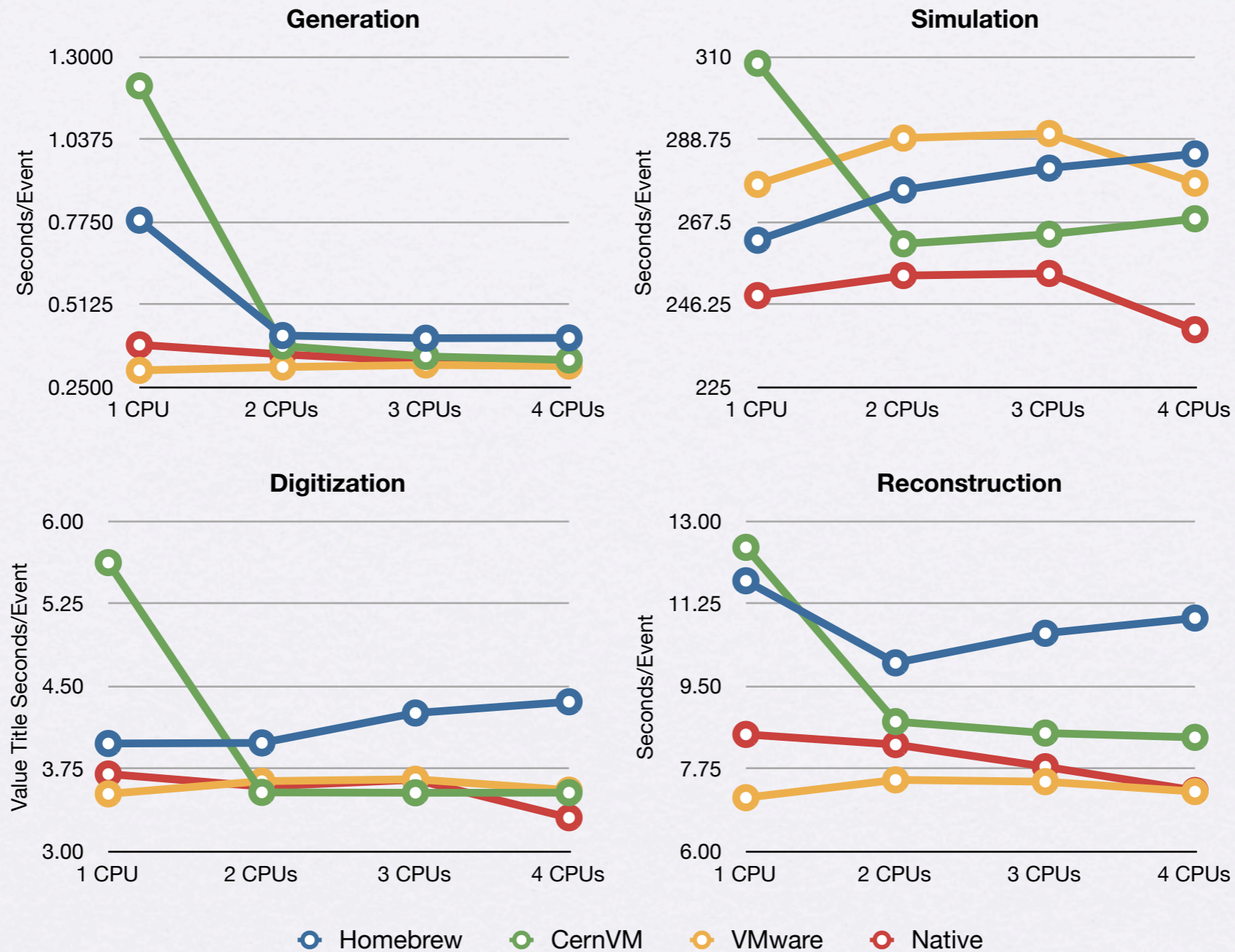
Athena Overall

Event Processing Time By System Type (smaller is better)



Athena by CPUs

Event Processing Time By Number of CPUs (smaller is better)

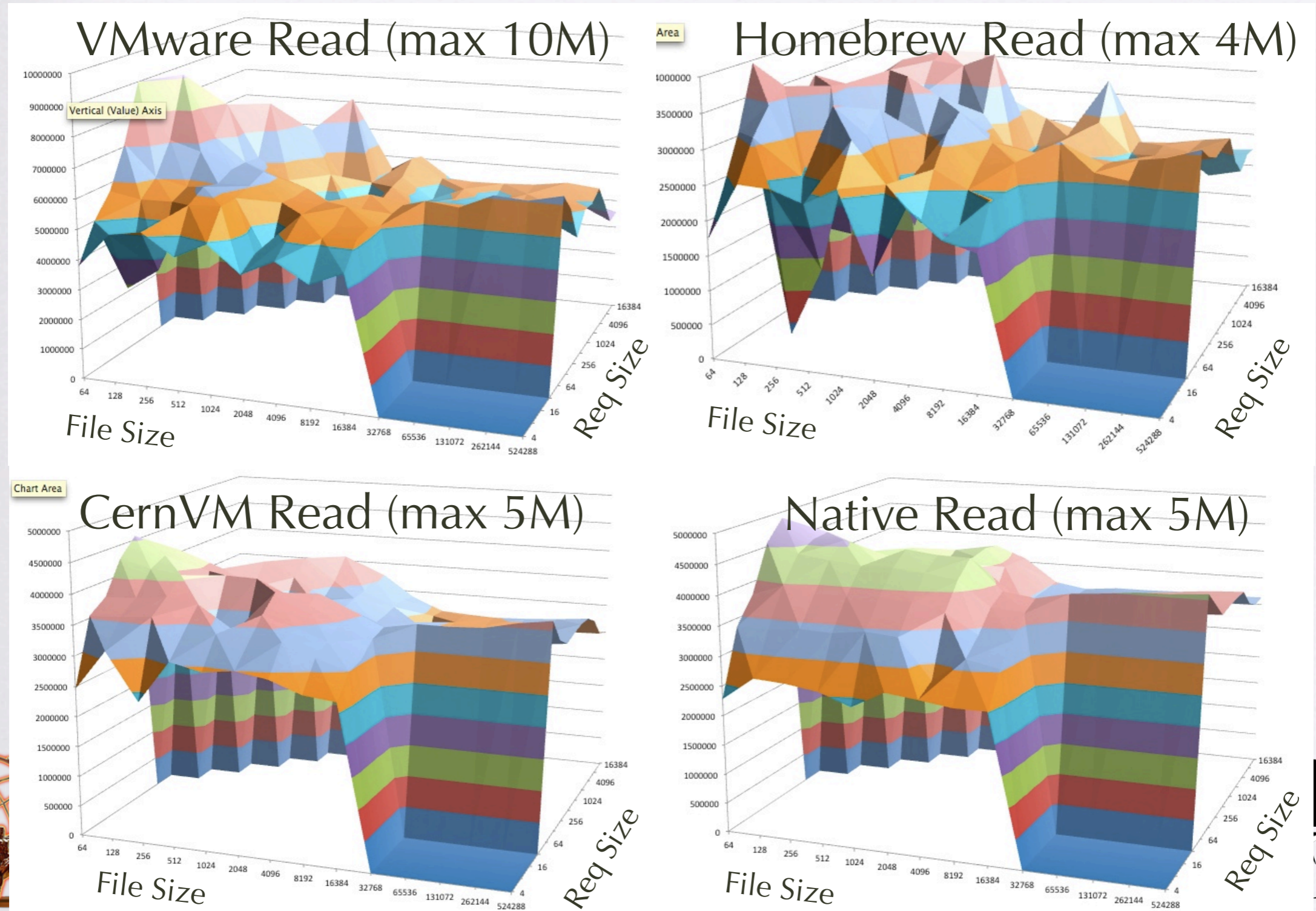


Disk Tests

- IOzone3_397
 - Very broad and deep set of results
 - Describes a landscape
- Bonnie++-1.96
 - Complementary approach
 - Average of three runs



IOzone Performance



Bonnie++ Performance

Version 1.96		Sequential Output					Sequential Input					Random Seeks		Sequential Create						Random Create						
	Size	Per Char		Block		Rewrite		Per Char		Block				Num Files	Create		Read		Delete		Create		Read		Delete	
		K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	/sec	% CPU		/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU
Homebrew_c26	4G	701	99	44128	45	30548	24	1849	98	74238	10	161.4	27	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	12599us		1440ms		678ms		24586us		109ms		372ms		Latency	236us		638us		666us		159us		141us		149us	
Homebrew_c26	4G	681	99	55163	60	29091	22	1814	98	61444	5	155.3	23	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	12181us		2232ms		564ms		14935us		92877us		413ms		Latency	289us		643us		725us		208us		82us		117us	

Version 1.96		Sequential Output					Sequential Input					Random Seeks		Sequential Create						Random Create						
	Size	Per Char		Block		Rewrite		Per Char		Block				Num Files	Create		Read		Delete		Create		Read		Delete	
		K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	/sec	% CPU		/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU
CernVM_c23	4G	728	99	108958	26	80109	28	2425	99	668625	45	2403	122	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	17439us		816ms		2498ms		5347us		832us		406ms		Latency	17112us		895us		950us		207us		25us		195us	
CernVM_c23	4G	726	99	111588	27	78118	27	2156	99	458970	37	2016	62	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	18170us		1081ms		2383ms		6145us		1014us		404ms		Latency	960us		950us		993us		211us		25us		147us	
CernVM_c23	4G	728	99	110527	28	75982	27	2302	99	453959	39	2129	47	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	17204us		1175ms		1605ms		6108us		8786us		404ms		Latency	930us		944us		964us		216us		21us		61us	

Version 1.96		Sequential Output					Sequential Input					Random Seeks		Sequential Create						Random Create						
	Size	Per Char		Block		Rewrite		Per Char		Block				Num Files	Create		Read		Delete		Create		Read		Delete	
		K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	K/sec	% CPU	/sec	% CPU		/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU	/sec	% CPU
VMware2	8G	657	61	81550	9	49157	2	2072	55	136895	6	135.6	1	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	318ms		2084ms		977ms		340ms		93150us		934ms		Latency	118ms		451us		504us		704us		66us		81us	
VMware2	8G	929	94	157484	19	50451	5	3528	96	71416	2	346.9	5	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	12688us		695ms		976ms		46877us		2436ms		575ms		Latency	25964us		483us		485us		1339us		30us		55us	
VMware2	8G	1041	93	167549	19	50872	3	2902	75	142159	5	1139	11	16	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++	+++++	+++
	Latency	9628us		1000ms		890ms		99509us		360ms		84525us		Latency	7691us		478us		483us		1260us		46us		187us	

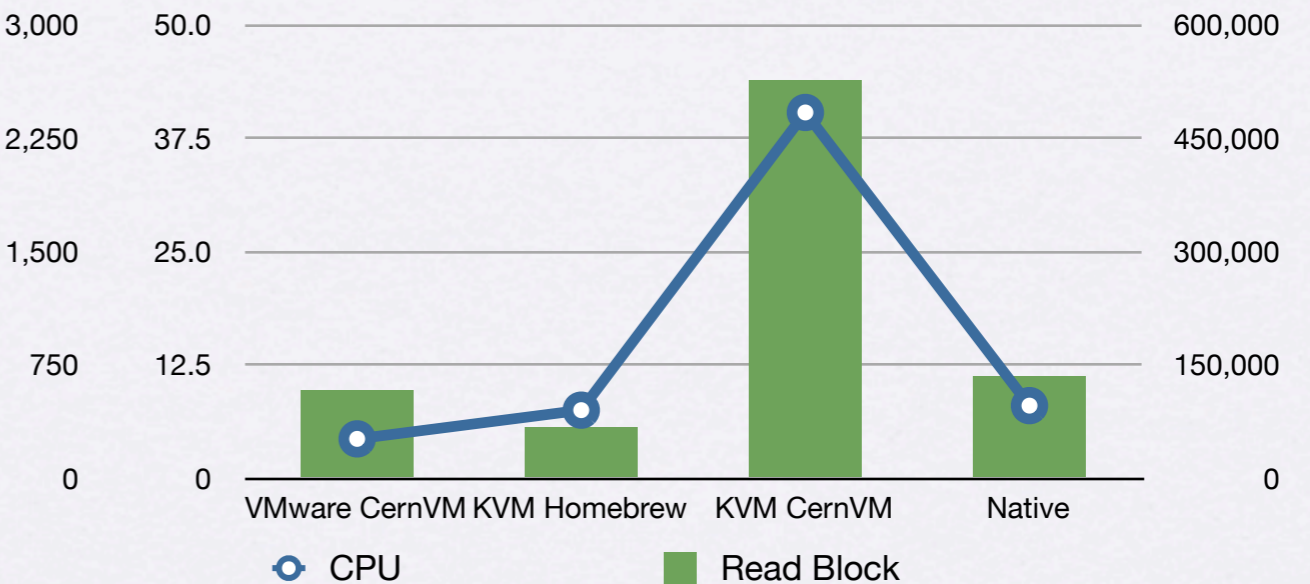
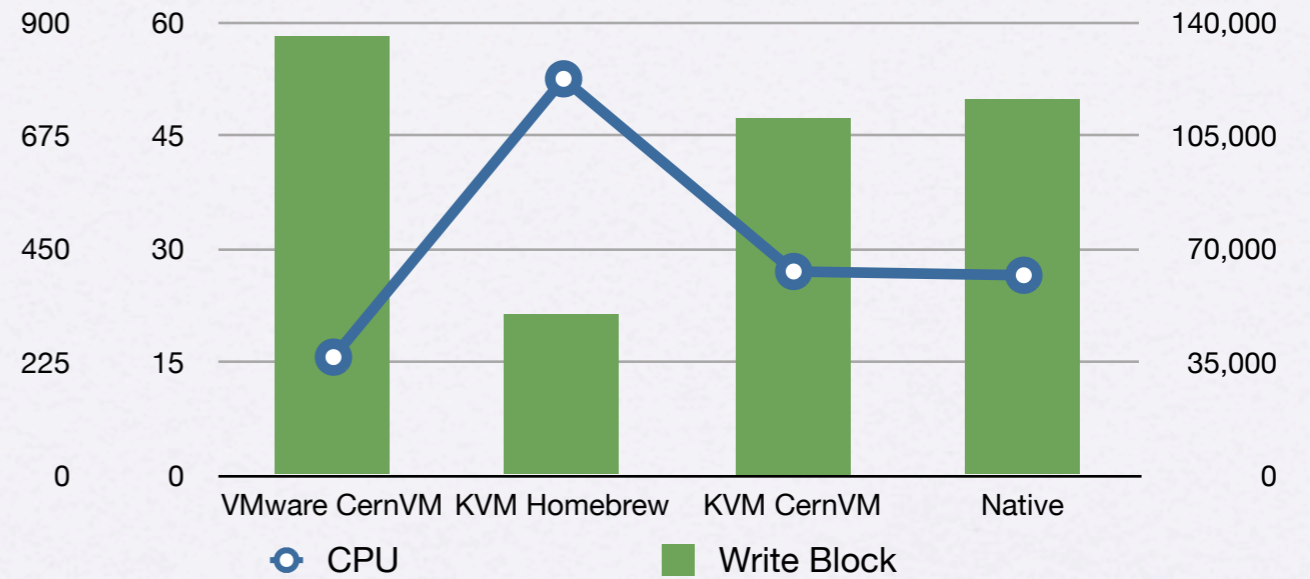
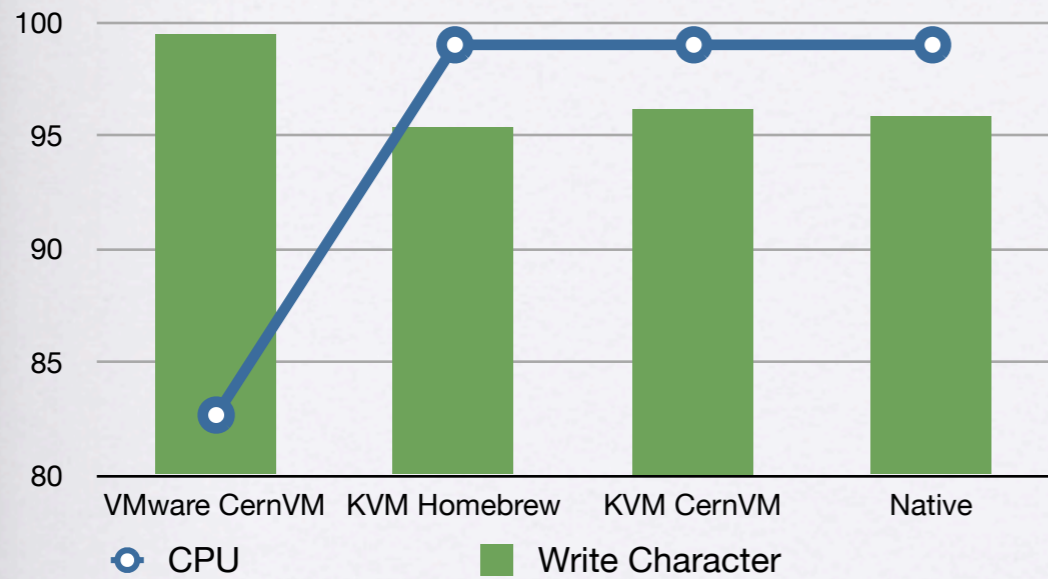


Bonnie++ Performance

- RAM determines file size
 - CernVM KVM: 4 GB filesize
 - CernVM VMware: 8 GB filesize
 - Homebrew KVM: 4 GB filesize
 - Native: 48GB filesize



Bonnie++ Performance

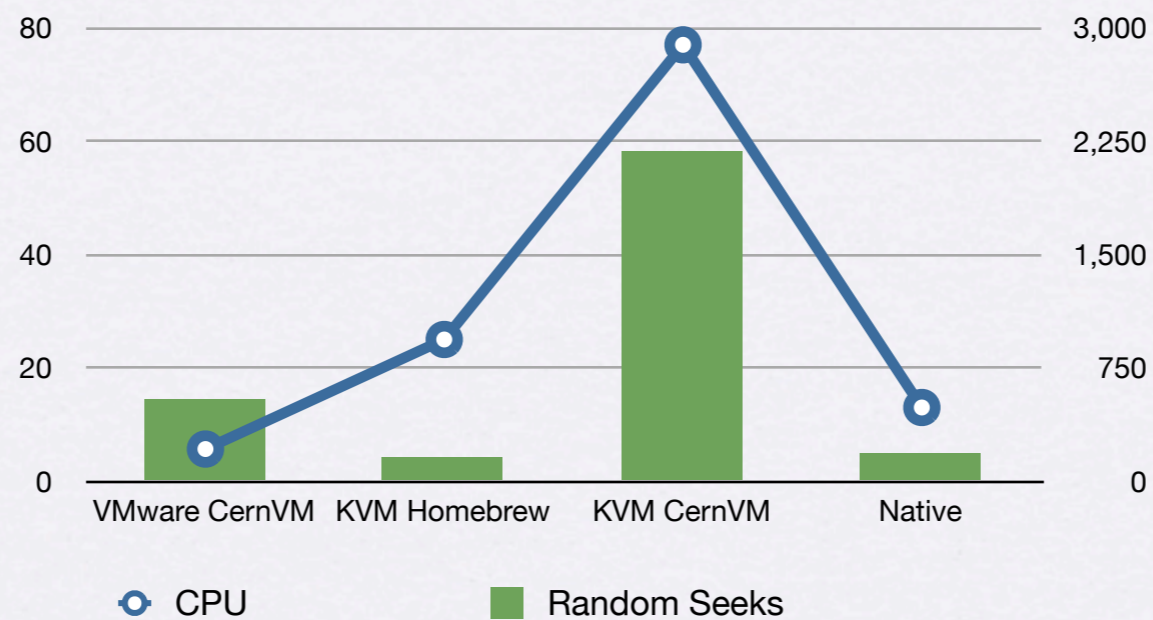
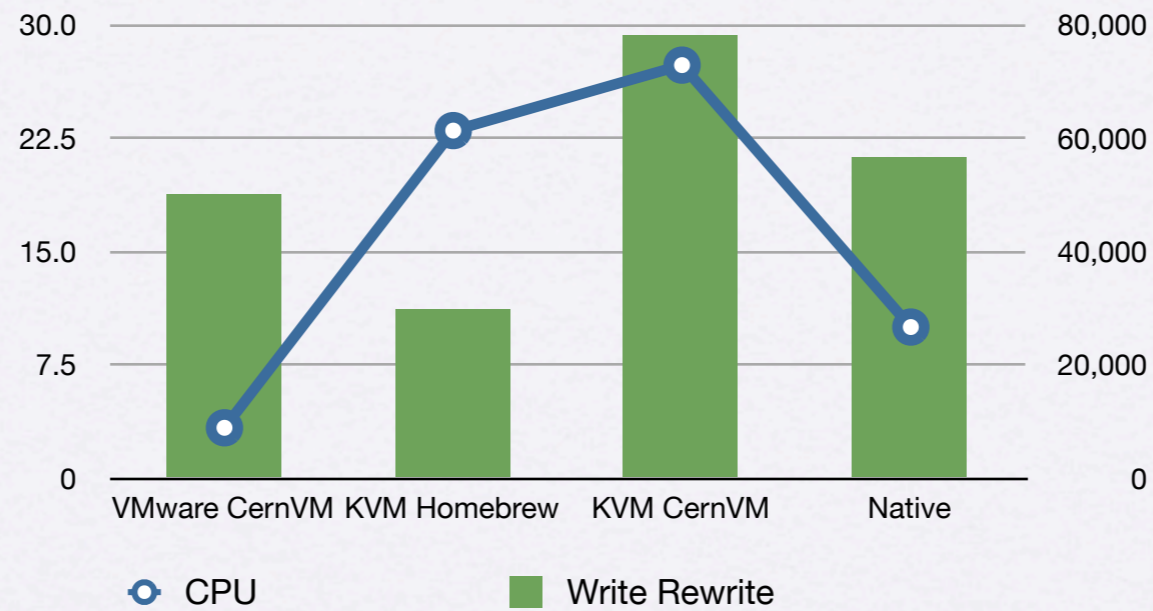


Kilobytes per second (larger is better)



Bonnie++ Performance

Kilobytes per second (larger is better)



UT ARLINGTON™

Management Interfaces

- KVM's native interface (virt-manage) is functional, pre-installed, and X11-based. It is also poorly laid out, slow and annoying
 - Doesn't matter much – most work from the command line
- VMware's tools are (reputedly) very functional, flexible and refined. Command line tools are also more robust and usable.
 - Not a shock – that's why you pay the big bucks



Ease of setup

- CernVM/KVM was terribly difficult... until Predrag found and corrected a bug in the contextualization system.
- CernVM/VMware was quick and easy. Just had to watch out for the `/etc/resolv.conf`
- Homebrew/KVM was quite doable... but unnecessary
- CVMFS was the key to making this work well, and it comes preconfigured in CernVM



Interim Conclusions for Athena

- VMware results with CernVM were very impressive – in bandwidth and disk-related tasks. Nearly native, within error bars.
 - Seems likely to be hardware-related
- CernVM equivalent to homebrew – so why maintain your own when it's nicely done already?
- Variations wrt. 1-CPU results probably involves averaging and run conditions.



Interim Conclusions for Disk

- KVM virtio seems to have done REALLY well on the **synthetic** benchmarks
 - No special modifications made to CernVM image used under VMware – worth looking into further optimization

-



Further Work

next couple of weeks

- Xrootd Comparison
 - Looking for ways to compare apples to apples
 - Plans are to compare KVM vs. native, and CernVM vs. Homebrew. Normalize for network... somehow.
- Same applies to NFS
- Go up to N(max) CPUs and see what slowdowns occur
- Repeat Bonnie and IOmax benchmarks with different settings



Interim Recommend

- Not completely sold on some of the results (especially the bonnie++ for KVM CernVM).
- Significant Athena performance improvements in VMware reflect really good hardware/hypervisor interactions. Shawn knows more about this.
- Sim performance on CernVM/VMware was unexpectedly low
- For price, I'd go with KVM *at the moment*. Perhaps it would be good to have a more apples/apples benchmark for VMware.

