

Center for Research Computing Meeting

Dan Voss
Director, Center for Research Computing
dan.voss@ku.edu
12/20/2017



Agenda

10:00am – 10:45am

- Expansion and Replacement of KU's Research File Storage (RFS)
- New Tier of Storage: Cold Data Archive

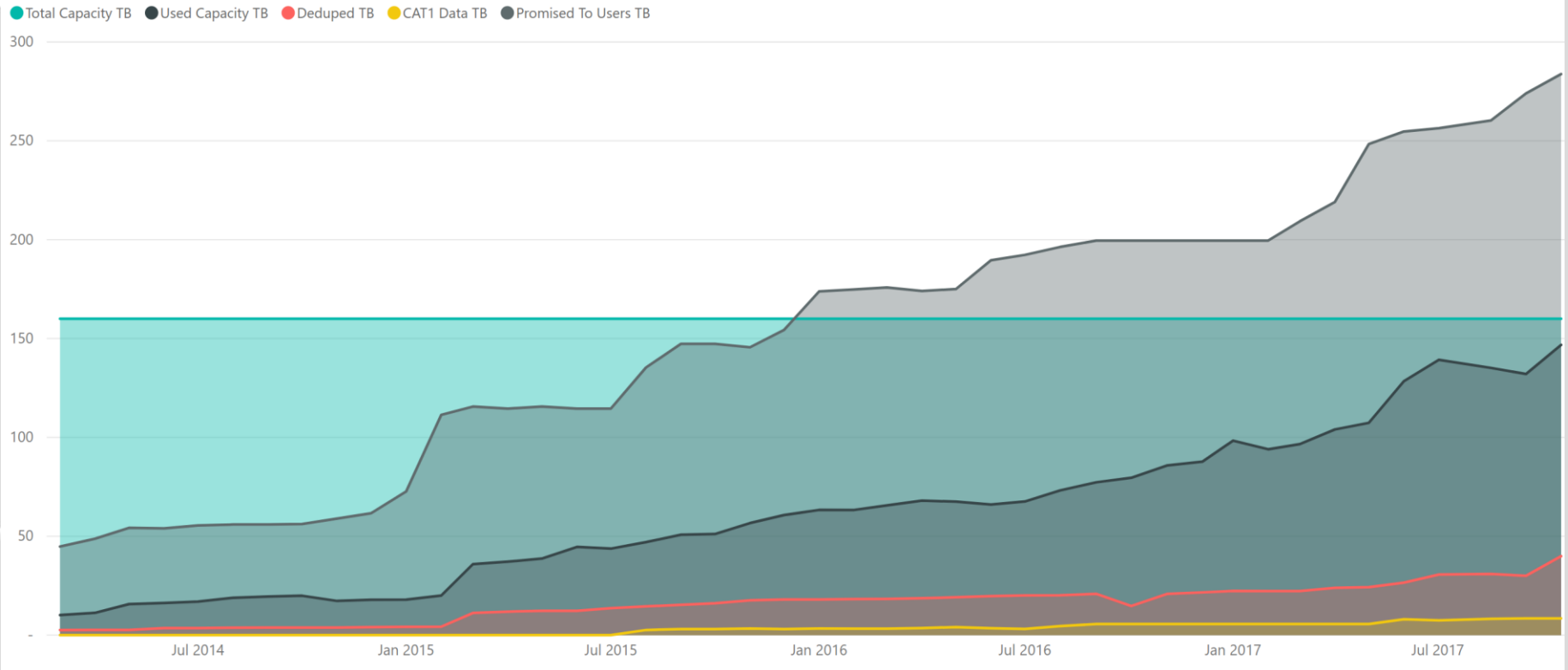
10:45am – 11:30am

- Configuration options for the 2018 expansion of the KU Community Cluster



Current RFS Usage

Total Capacity TB, Used Capacity TB, Deduped TB, CAT1 Data TB and Promised To Users TB by Date



New RFS

12 GB/s

- Two IBM High Performance GL-2S Elastic Storage Systems – ESS
- 166 10TB drives – 1.6 PBs RAW, 1.2 PBs Usable after Distributed RAID
- 5 year life cycle (January, 2018 – December, 2023)

New! Model GL2S:
2 Enclosures, 14U
166 NL-SAS, 2 SSD



Max: 1.6PB raw



New RFS

- Backups
- Billing?
 - Monthly
 - Quarterly
 - Yearly
- Access
 - Globus
 - CIFS
 - NFS (on KU Domain) Kerberos
- Price?
 - RESET BUTTON
 - Growth Projections

12 GB/s

New! Model GL2S:
2 Enclosures, 14U
166 NL-SAS, 2 SSD



Max: 1.6PB raw



New Tier of Storage: Cold Data Archive

- One IBM TS 4300 Library with Archiving Software
- 72 tape slots
- LTO8 tape media, 12.8 TB uncompressed, 32 TB compressed
- 922 TBs uncompressed – 2.3PB compressed
- Expansion to 232 slots with a capacity of 3PB uncompressed – 7.4PB compressed



New Tier of Storage: Cold Data Archive

Expectations?

- Stage data on RFS archive volume, data slowly migrated to tape.
 - Tar is your friend.
- File metadata is accessible on RFS archive volume without loading and reading tape (e.g., filenames, dates, size on tape, etc.).
- True **COLD** data - accessing data on tape might take days depending on how busy tape library is.



New Tier of Storage: Cold Data Archive

- 2 Copies (one in tape library, one offsite)
- Billing?
 - Upfront
 - Monthly
 - Quarterly
 - Yearly
- Access
 - RFS
- Price?
 - Best Buy Competitive!
 - Unit of Sale?



Agenda

10:00am – 10:45am

- Expansion and Replacement of KU's Research File Storage (RFS)
- Cold Data Archive

10:45am – 11:30am

- Configuration options for the 2018 expansion of the KU Community Cluster



KU Community Cluster

- 1) Standard Compute Unit (SCU) is a standard compute node with 600GB of global, single-copy, high performance storage.**
- 2) High-Memory SCU**
- 3) SCU + GPGPU**



KU Community Cluster

Configuration Options

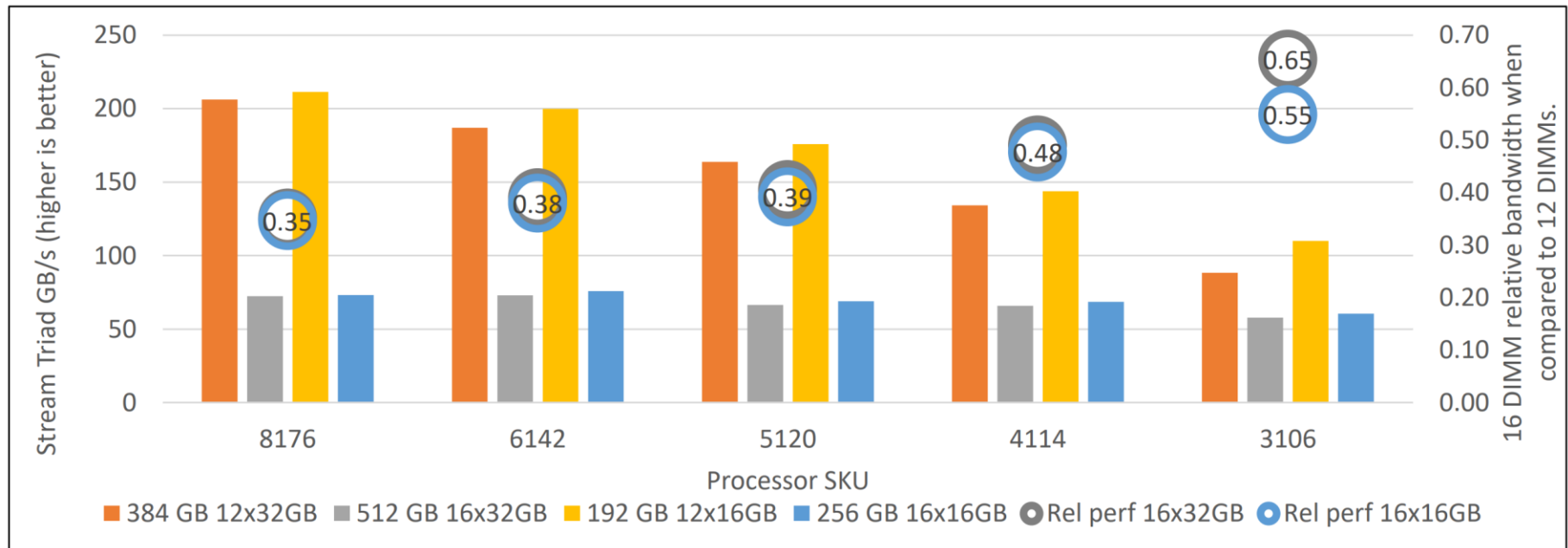
- **What's important to you?**
- **Our goal is to provide the best HPC system configuration based on domain benchmarks, performance, and value.**
- **Which benchmarks would you like us to compare?**
- **Intel or AMD?**
- **Intel ideal memory configuration: 192GB, 384GB, 768GB**
- **AMD ideal memory configurations: 128GB, 256GB, 512GB, etc.**



KU Community Cluster

Intel Unbalanced Memory Penalty

256 GB is a no-no. Impact of unbalanced memory



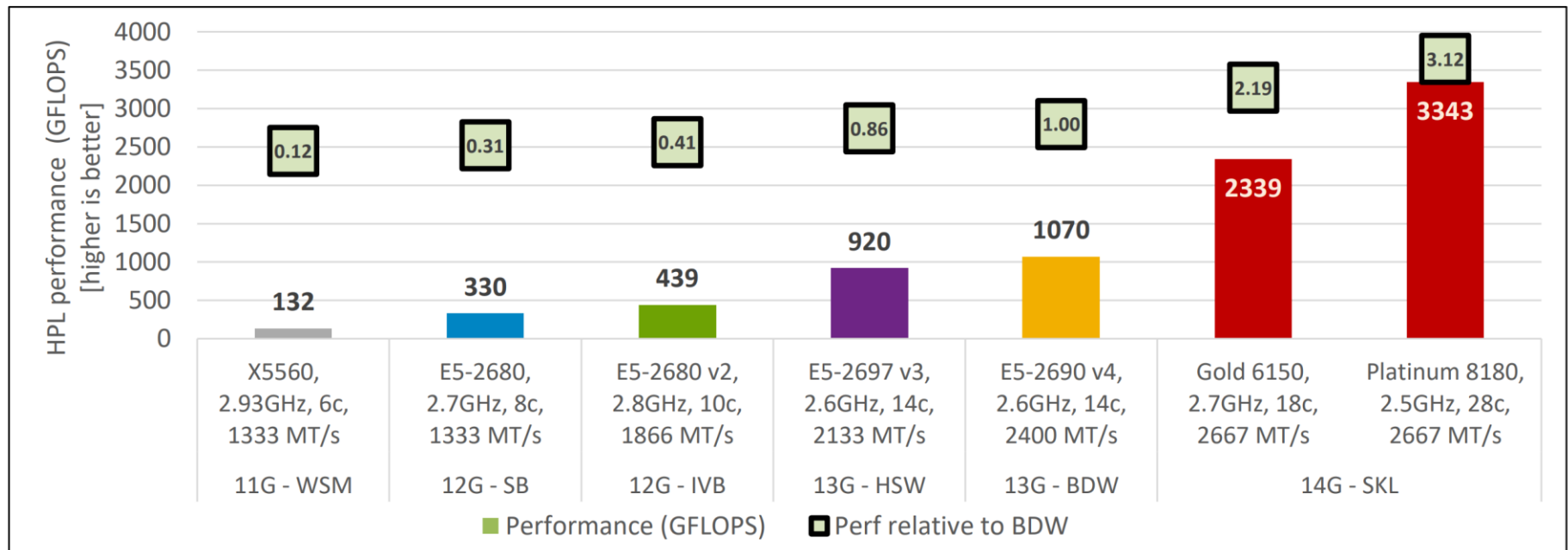
- The 16 DIMM unbalanced config provides only 35-65% bandwidth of the equivalent 12 DIMM balanced config. **Adding more DIMMs and more capacity DROPS performance by up to 65%, at least 35%.**



KU Community Cluster

Intel Benchmarks

HPL – memory bandwidth across generations



3.3 TF in a single server!

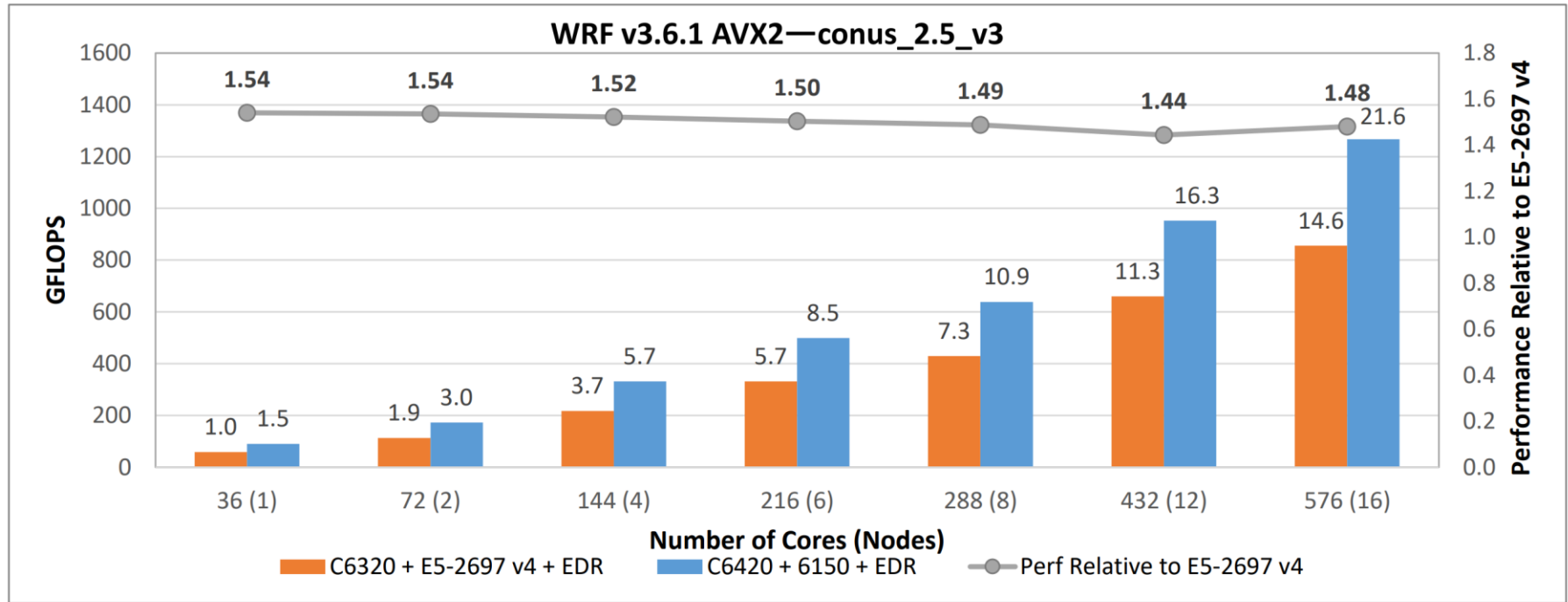
AVX512 with 32 DP FLOP/cycle makes the difference here. Plus more cores.



KU Community Cluster

Intel Benchmarks

WRF—Medium Model Scaling



- SKL performs 50% better than SKL at multiple node counts.
- Some reduction in performance advantage scaling past 8 nodes, small data set.



New Intel CPUs

Important changes available in Xeon Scalable Processor Family “Skylake”:

- **Up to 28 processor cores** per CPU
- New AVX-512 instructions double performance
- Up to two AVX-512 FMA units per CPU core
- Support for **DDR4 memory** speeds **up to 2666MHz**
- **Six-channel memory controller** on each CPU (up from four-channel on previous platforms)
- **Optional 1.5TB-per-socket system memory support** (only available on certain SKUs)



New NVIDIA V100 GPU, “Volta” Architecture

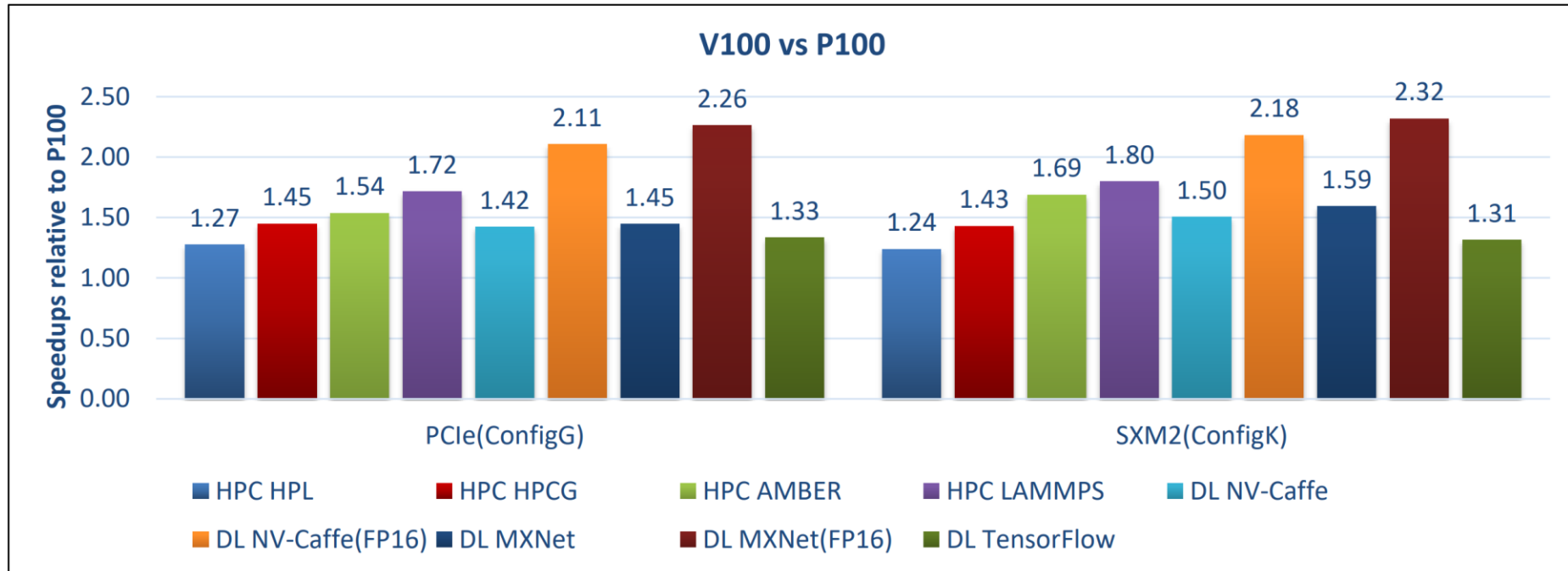
Performance of Tesla GPUs, Generation to Generation

	Tesla V100 with NVLink	Tesla V100 PCI-E	Tesla P100 with NVLink	Tesla P100 PCI-E	Ratio Tesla V100:P100
DP TFLOPS	7.5 TFLOPS	7.0 TFLOPS	5.3 TFLOPS	4.7 TFLOPS	~1.4-1.5X
SP TFLOPS	15 TFLOPS	14 TFLOPS	9.3 TFLOPS	8.74 TFLOPS	~1.4-1.5X
TensorFLOPS	120 TFLOPS	112 TFLOPS	21.2 TFLOPS 1/2 Precision	18.7 TFLOPS 1/2 Precision	~6X
Interface (bidirec. BW)	300GB/sec	32GB/sec	160GB/sec	32GB/sec	1.88X NVLink 9.38X PCI-E
Memory Bandwidth	900GB/sec	900GB/sec	720GB/sec	720GB/sec	1.25X
CUDA Cores (Tensor Cores)	5120 (640)	5120 (640)	3584	3584	



New NVIDIA V100 GPU, “Volta” Architecture

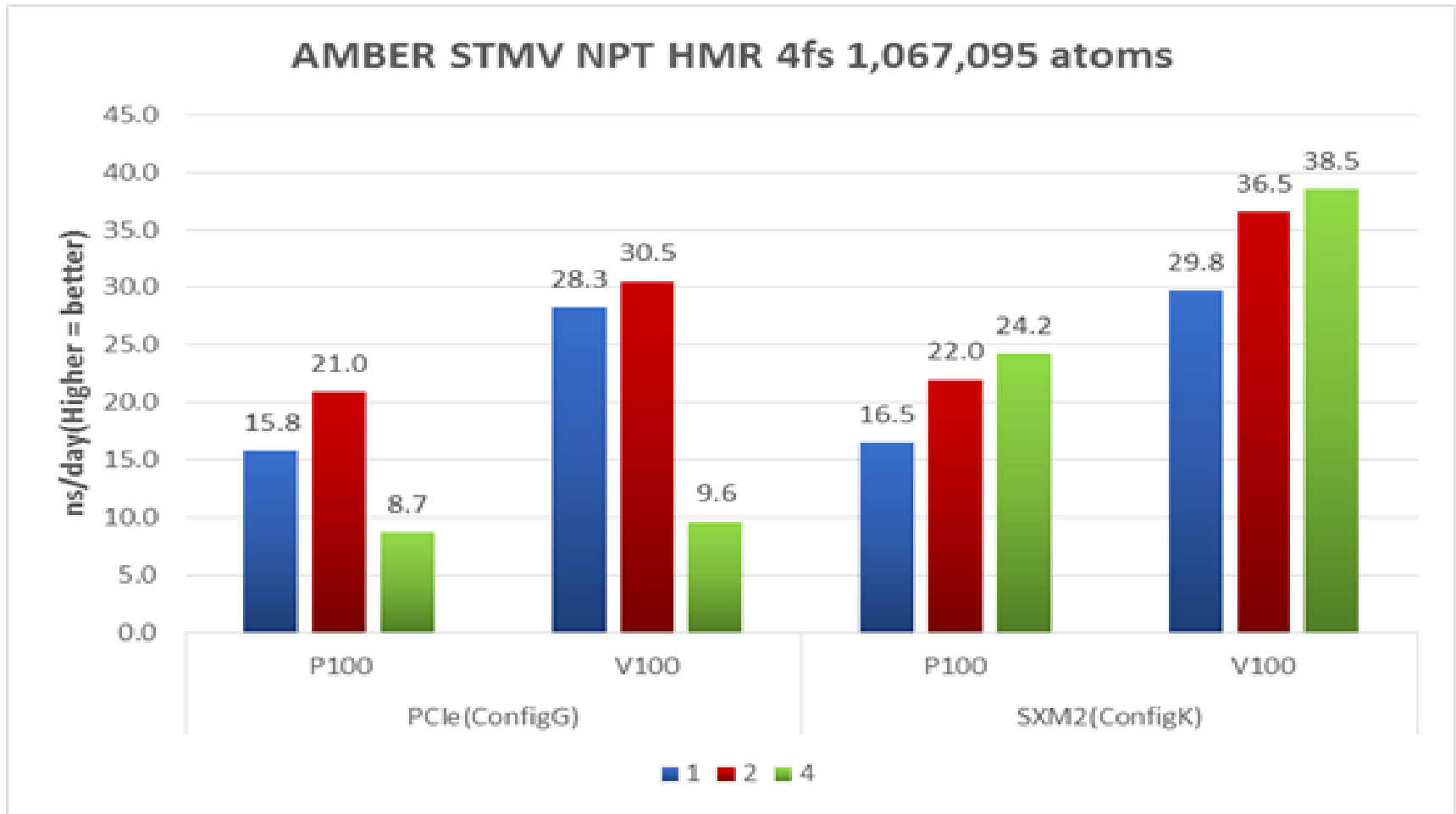
V100 performance across applications - summary



- V100 initial performance is impressive, 25% to 230% better than P100.
- Tests should performance of 4 GPU cards in a PowerEdge C4130



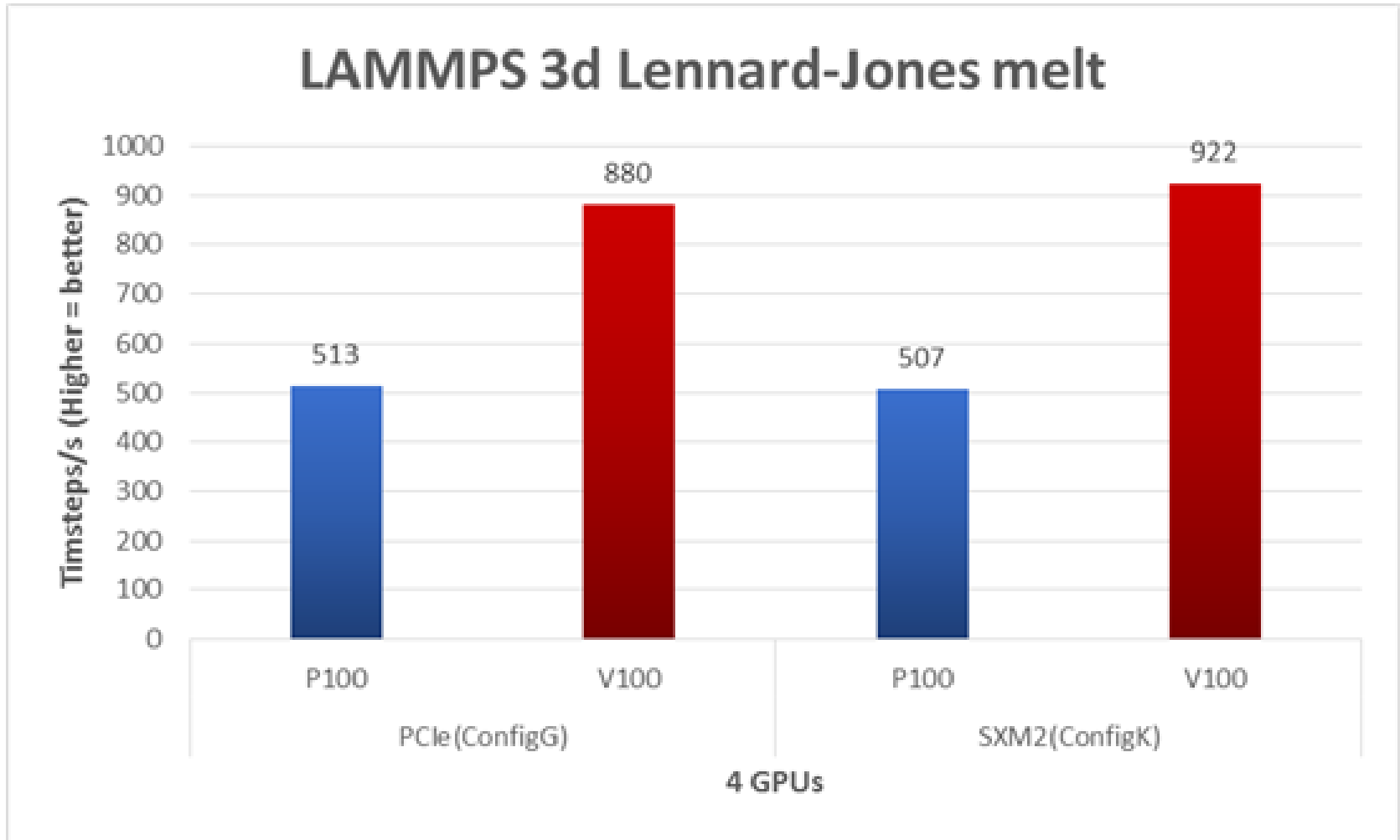
New NVIDIA V100 GPU, “Volta” Architecture



Even though the scaling is not strong, V100 has noticeable improvement than P100, giving ~78% increase in single card runs, and 1x V100 is actually 23% faster than 4x P100.



New NVIDIA V100 GPU, “Volta” Architecture



V100 is 71% and 81% faster on Config G and Config K respectively.



KU Community Cluster

Initial Pricing Estimate for an Intel CPU Node

	FY18	Price diff from FY16
SCU (192GB memory):	\$10,790	+ 12%/core
SCU + 192GB memory (384GB total):	\$13,710	
SCU + 576GB memory (768GB total):	\$21,360	
SCU + 2 NVIDIA V100 GPGPU's:	\$28,160	
SCU + 4 NVIDIA V100 GPGPU's:	\$47,435	
Additional Panasas Storage:	\$200/TB per year	

NOT final pricing
or hardware
configuration
(still testing CPUs).

This is initial Dell
pricing for a 40-core
high-end node.

SCU: Standard Compute Unit (SCU) is a standard compute node with 600GB of global, single-copy, high performance storage, 40-cores, 2.4GHz, 192GB memory, infiniband interconnect.



University of Kansas

Questions?

Comments?

dan.voss@ku.edu

