

Taking Advantage of Technology Innovations in the Next Generation of NeSI HPC Infrastructure



11 February, 2021

Jeff Zais



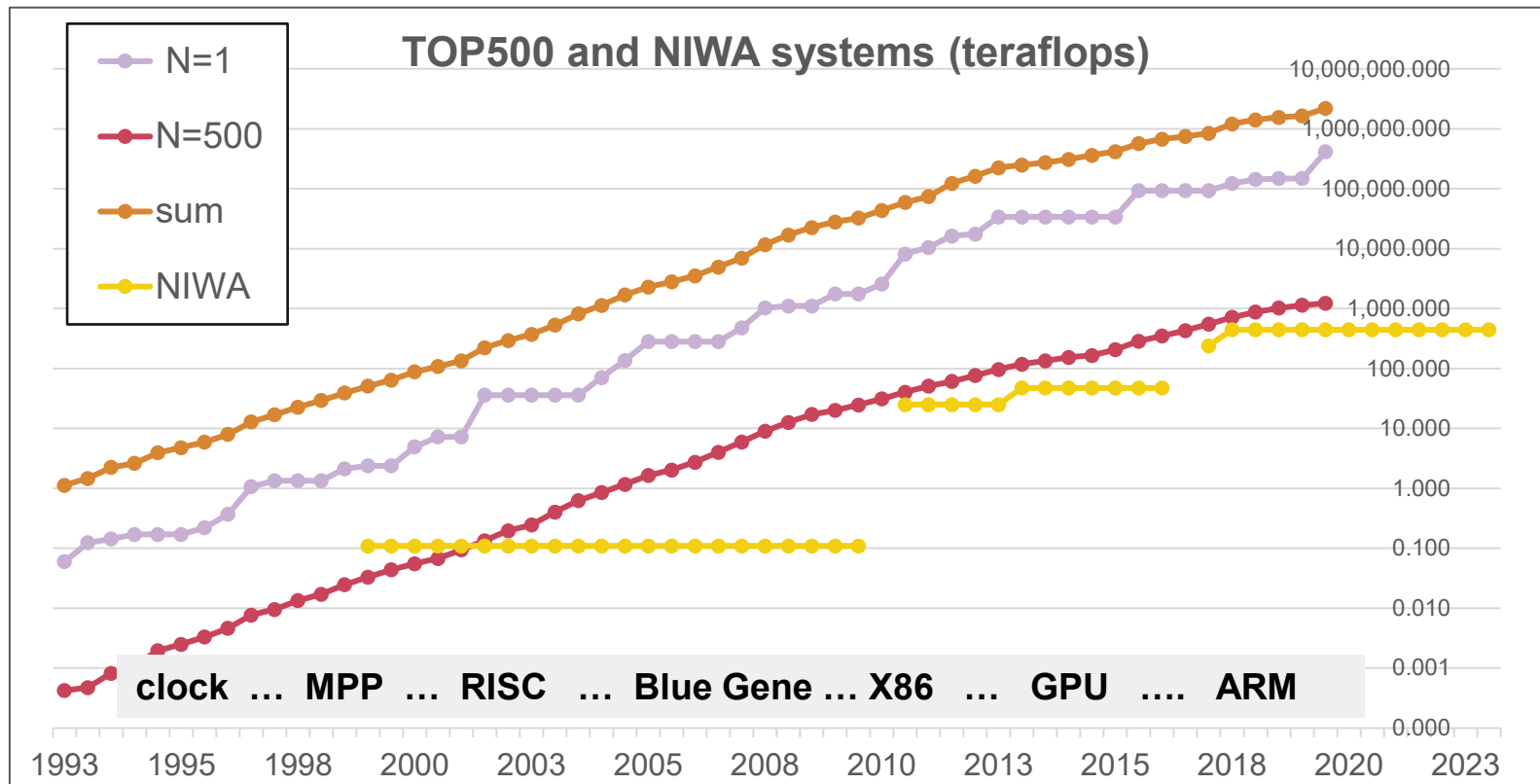
Today's exercise

$$\frac{64 \text{ (5.83)}}{18 \text{ (3.86)}} \times \frac{3200 \text{ (8)}}{2400 \text{ (4)}} \times \frac{64}{234} \times \frac{40 \text{ (200)}}{36 \text{ (100)}} \times \frac{200}{120} = ?$$

Benefits of Natural Advantages



The history of HPC in a single chart



Capacity & Capability vision becomes reality with Mahuika & Maui

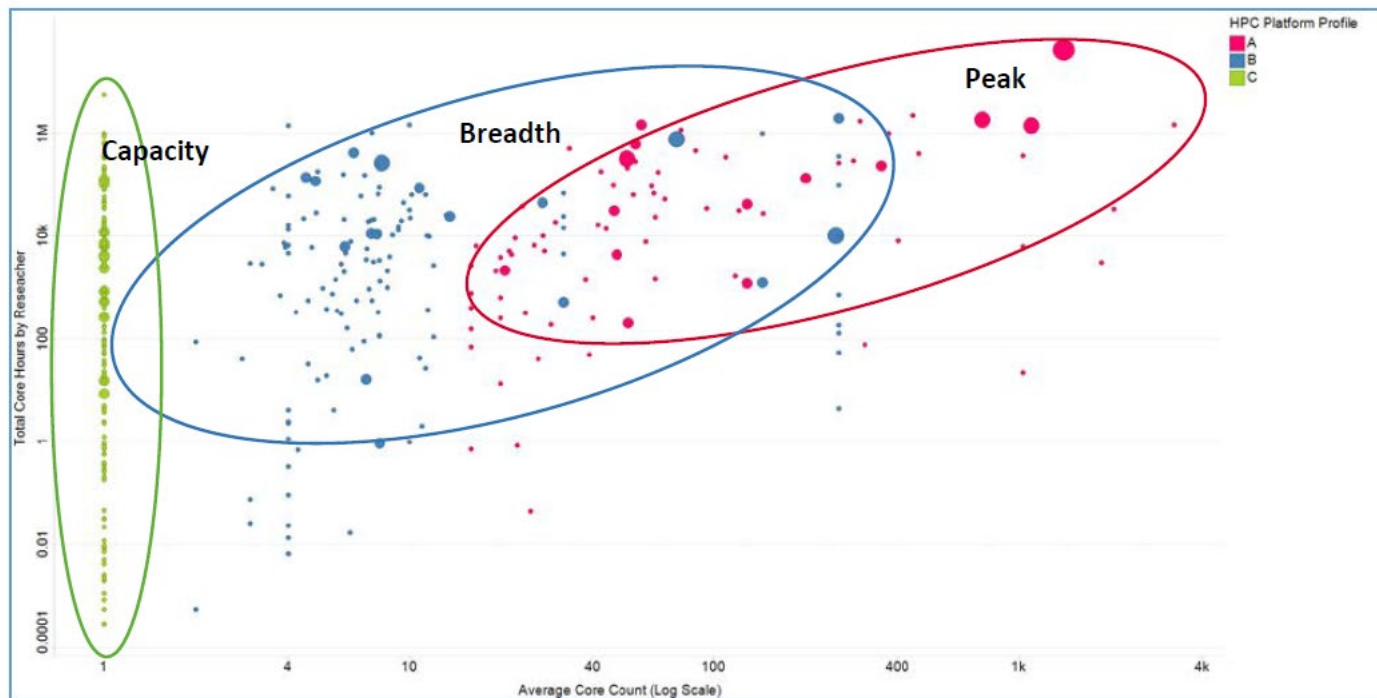


Figure 5 – Profiles of HPC Use, 2012 – Q1 2014

NeSI HPC equipment in a single photo





MAHUIKA



MAHUIKA

Mahuika is the Maori goddess of fire. In customary lore, Pahi, sourced fire for his people by convincing his grandfather Mahuika to hand over flames burning beneath her wings. Mahuika is now inspiring researchers throughout New Zealand in a range of fields, from geomatics to the analysis of complex structures.

- Mahuika Configuration:**
- Core: 640 x 2 (Intel Core Broadwell) cores
 - 32.0 Terabytes of memory
 - Core: 640 x 2 (Intel Core Broadwell) cores
 - 640 x 2 (Intel Core Broadwell) cores
 - 32.0 Terabytes of memory
 - 640 x 2 (Intel Core Broadwell) cores

Mahuika & Maui Usage



HPC Innovations 2017-2021



Technology Innovations

item	2017	2021
processors	Intel "Broadwell" (18-core) Intel "Skylake" (20-core)	Intel "Ice Lake" (32-core ++) AMD "Milan" (64-core)
memory	4 channels @2400 MHz 4 channels @2667 MHz	8 channels @3200 MHz
InfiniBand	36-port EDR (100 Gb/s)	40-port QDR (200 Gb/s)
GPU	NVIDIA P100	NVIDIA A100 AMD MI100

Processor Innovations - Detail

item	2017	2021
processors	Intel "Broadwell" (18-core)	AMD "Milan" (64-core)
memory	4 channels @2400 MHz 16GB each = 64GB 3.6GB per core	8 channels @3200 MHz 32GB each = 256GB 4.0GB per core
clock speed	2.1 GHz	2.0 GHz ?
TDP (watts)	120	200 ?
SPECfp_rate <i>per core</i>	3.7	4.5 ?

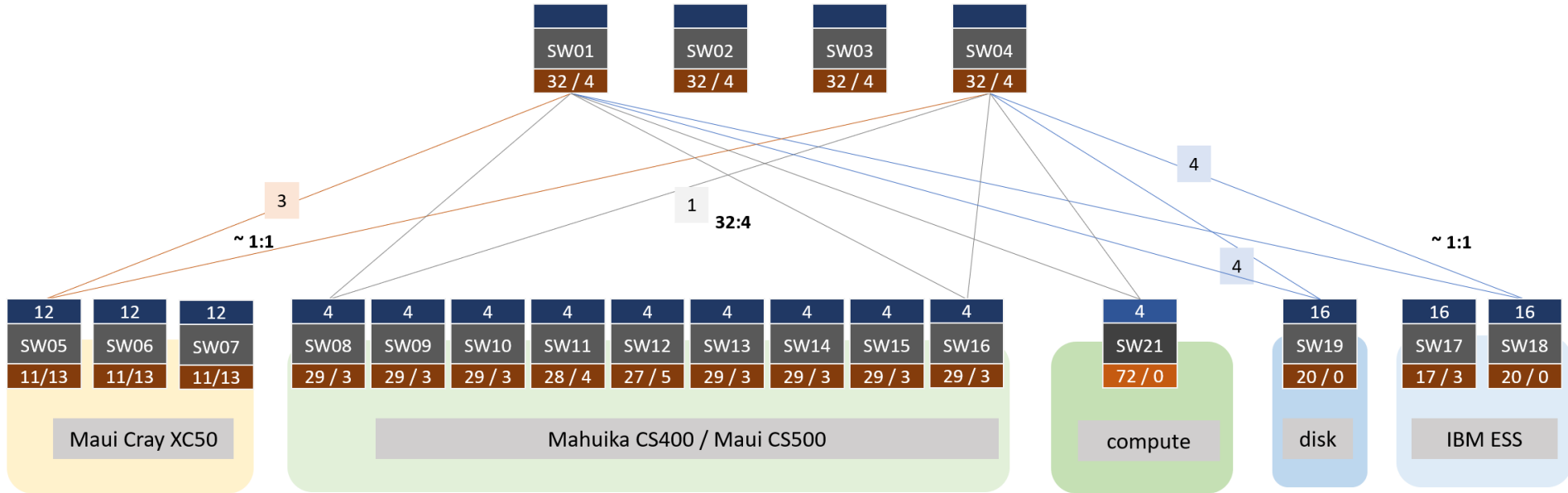
Interconnect Innovations - Detail

item	2017	2021
connection speed	FDR 56Gb/s EDR 100 Gb/s	HDR 200 Gb/s HDR100 100 Gb/s
ports per switch	36	40 – at 200 Gb/s 80 - at 100 Gb/s
latency (microsec)	1.10 ?	1.05 ?

GPU Innovations - Detail

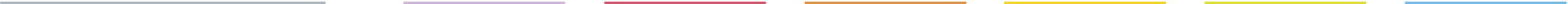
NVIDIA Accelerator Specification Comparison				
	A100 (PCIe)	A100 (SXM4)	V100 (PCIe)	P100 (PCIe)
FP32 CUDA Cores	6912	6912	5120	3584
Boost Clock	1.41GHz	1.41GHz	1.38GHz	1.3GHz
Memory Clock	2.43Gbps HBM2	2.43Gbps HBM2	1.75Gbps HBM2	1.4Gbps HBM2
Memory Bus Width	5120-bit	5120-bit	4096-bit	4096-bit
Memory Bandwidth	1.6TB/sec (1555GB/sec)	1.6TB/sec (1555GB/sec)	900GB/sec	720GB/sec
VRAM	40GB	40GB	16GB/32GB	16GB
Single Precision	19.5 TFLOPs	19.5 TFLOPs	14.1 TFLOPs	9.3 TFLOPs
Double Precision	9.7 TFLOPs (1/2 FP32 rate)	9.7 TFLOPs (1/2 FP32 rate)	7 TFLOPs (1/2 FP32 rate)	4.7 TFLOPs (1/2 FP32 rate)
INT8 Tensor	624 TOPs	624 TOPs	N/A	N/A
FP16 Tensor	312 TFLOPs	312 TFLOPs	112 TFLOPs	N/A
TF32 Tensor	156 TFLOPs	156 TFLOPs	N/A	N/A
Relative Performance (SXM Version)	90%	100%	N/A	N/A
Interconnect	NVLink 3 12 Links (600GB/sec)	NVLink 3 12 Links (600GB/sec)	NVLink 2 4 Links (200GB/sec)	NVLink 1 4 Links (160GB/sec)
GPU	GA100 (826mm2)	GA100 (826mm2)	GV100 (815mm2)	GP100 (610mm2)
Transistor Count	54.2B	54.2B	21.1B	15.3B
TDP	250W	400W	250W	300W
Manufacturing Process	TSMC 7N	TSMC 7N	TSMC 12nm FFN	TSMC 16nm FinFET
Interface	PCIe 4.0	SXM4	PCIe 3.0	SXM
Architecture	Ampere	Ampere	Volta	Pascal

Mahuika with the Extended Compute – and Future Disk



Pitfalls





Pitfall? Migration from Intel to AMD?

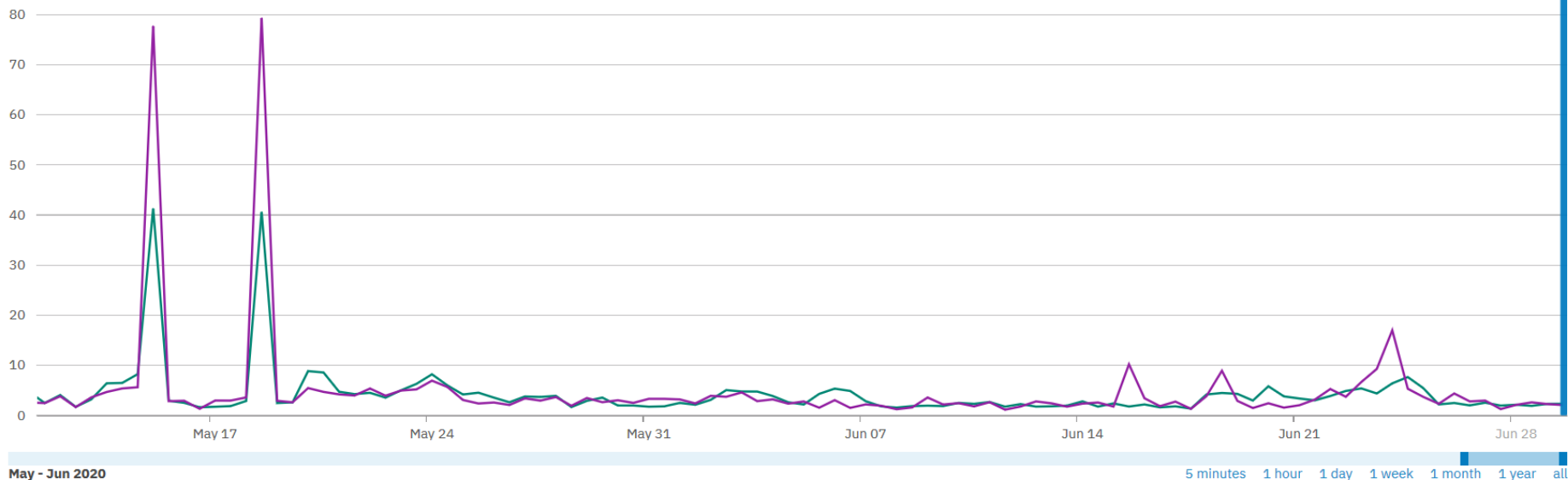
how much work to recompile and run applications?

Pitfall? Storage Bandwidth

4 connections to storage, each 100 Gb/s, 400 Gb/s = 50 GB/s

File System: scale_wlg_nobackup / NSD Server / Throughput ★

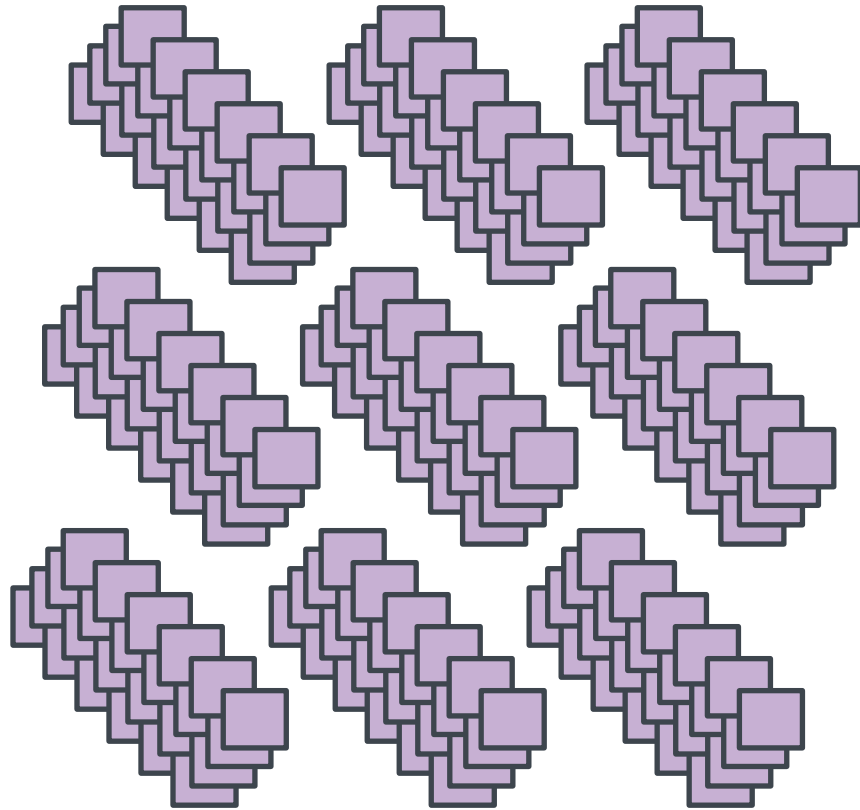
Bytes Written Bytes Read



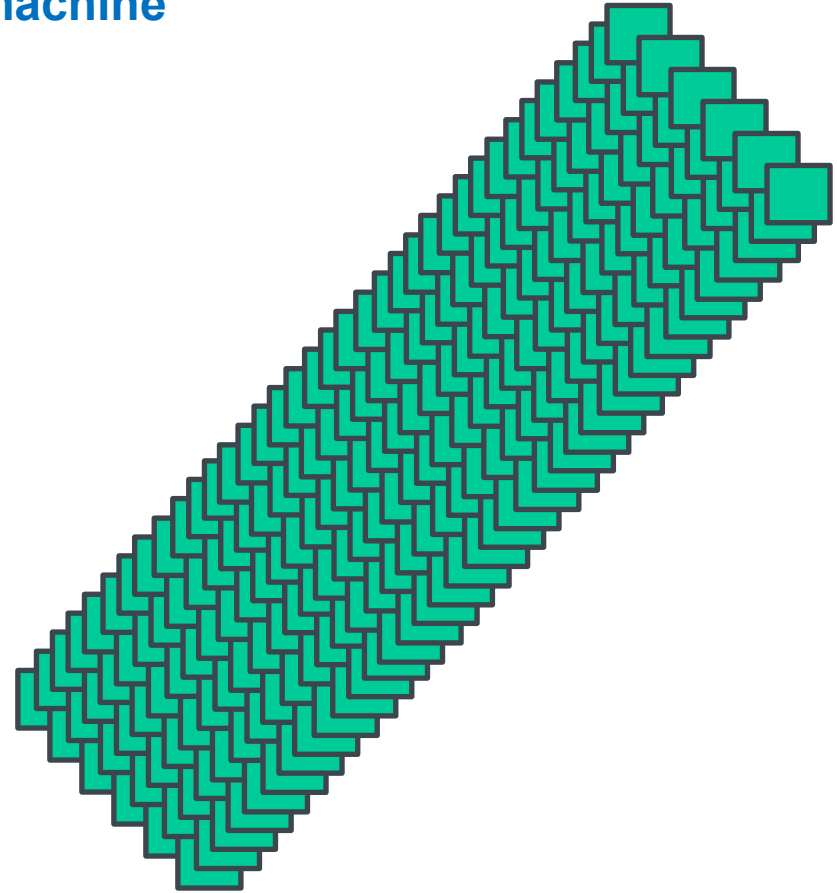
Bonus



Bonus – the extension is a capability machine



8424 cores loosely coupled 8:1



9216 cores tightly coupled 1:1

Return to today's equation



New Zealand eScience Infrastructure



$$\frac{64 (5.83)}{18 (3.86)} \times \frac{3200 (8)}{2400 (4)} \times \frac{64}{234} \times \frac{40 (200)}{36 (100)} \times \frac{200}{120} = 1.23$$

3 Mahuika	4 Mahuika	5 Mahuika	6 Mahuika	7 Mahuika
21	59	59	59	59
EP501	48 port GBE switch		48 port GBE switch	
48 port GBE switch	48 port GBE switch	48 port GBE switch	48 port GBE switch	48 port GBE switch
Brocade 7750 switch	EDR IBsw01	EDR IBsw02	EDR IBsw03	EDR IBsw04
Mellanox EDR-GB.com	EDR IBsw08	EDR IBsw10	EDR IBsw12	EDR IBsw15
Brocade 7750 switch	EDR IBsw09	EDR IBsw11	EDR IBsw13	EDR IBsw16
Brocade 7750 switch			EDR IBsw14	
wb001 (1.5 TB memory node)	1 1 1	wb001 2U Supermicro 6 TB		
0 0	1 1	1 1	1 1	1 1
1 1	1 1	1 1	1 1	1 1
1 1	1 1	1 1	1 1	1 1
1 1	1 1	1 1	1 1	1 1
1 1	1 1	1 1	1 1	1 1
1 1	1 1	1 1	1 1	1 1
Login Node 01	1 1	1 1	1 1	1 1
"wb005"	1 1	1 1	1 1	1 1
Build Node 02	1 1	1 1	1 1	1 1
"mahuika02"	1 1	1 1	1 1	1 1
Build Node 01	1 1	1 1	1 1	1 1
"mahuika01"	1 1	1 1	1 1	1 1
KVM	1 1	1 1	1 1	1 1
Bright Cluster Manager 02	1 1	1 1	1 1	1 1
Bright Cluster Manager disk	1 1	1 1	1 1	1 1
Bright Cluster Manager 01	1 1	1 1	1 1	1 1
4 - wbg004 - with 2x P100	1 1	1 1	1 1	1 1
3 - wbg003 - with 2x P100	1 1	1 1	1 1	1 1
2 - wbg002 - with 2x P100	1 1	1 1	1 1	1 1
1 - wbg001 - with 2x P100	1 1	1 1	1 1	1 1
huge memory node 8 sockets wbl001	3 - wbl003	6	9	12 - now wbg005 - with 1x P100
6 - wbl006	2 - wbl002	5	8	11 - wbl011
	1 - wbl001	4	7	10 - wbl010





New Zealand eScience Infrastructure

Thank you.

Kia ora koutou.

