Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Linux Cluster Hardware

ITS Meeting

Monday, 6th July 2015 11:15-12:15

BFH – Conference Room C.O.R.A. (Dammweg 3, 1. 1st Floor, Bern)

Daniel Baumann <daniel.baumann@bfh.ch>

IT System Engineer, Infrastructure Team

Overview

- System Overview

- Vendor Bids
- Buy Decision

- Upgrade Options
- Networking Options

- Totals
- Further Steps

System Overview

Ceph OSD Nodes (6x)

Parts

- 1x Supermicro 5028R-E1CR12L (2U Barebone, 1x CPU, 12x 3.5“ front HDD, 2x 2.5“ read HDD, 4x 1Gbit)
- 1x Intel E5-2620V3 (6x 2.4GHz, 85W TDP)
- 1x Kingston KVR21R15D4K4/64 (4x 16GB DDR4 reg. ECC RAM)
- 2x Intel DC S3700 100GB (OS SSDs, SATA3)
- 3x Seagate ST4000NM0033 4TB (Data HDDs, SATA3)
- 3x Western Digital WD4001FFSX 4TB (Data HDDs, SATA3)
- 1x Technogroup 5Y 5x11 NBD (Support Pack)

Ceph MON Nodes (3x)

Parts

- 1x Supermicro 6018R-MTR (1U Barebone, 2x CPU, 4x 3.5“ front HDD, 2x 1Gbit)
- 2x Intel E5-2620V3 (6x 2.4GHz, 85W TDP)
- 1x Kingston KVR21R15D4K4/64 (4x 16GB DDR4 reg. ECC RAM)
- 2x Intel DC S3700 100GB (OS SSDs, SATA3)
- 1x Technogroup 5Y 5x11 NBD (Support Pack)

OpenStack Compute Nodes (6x)

Parts

- 1x Supermicro 6018R-MTR (1U Barebone, 2x CPU, 4x 3.5“ front HDD, 2x 1Gbit)
- 2x Intel E5-2620V3 (6x 2.4GHz, 85W TDP)
- 2x Kingston KVR21R15D4K4/64 (8x 16GB DDR4 reg. ECC RAM)
- 2x Intel DC S3700 100GB (OS SSDs, SATA3)
- 1x Technogroup 5Y 5x11 NBD (Support Pack)

OpenStack Admin Nodes (2x)

Parts

- 1x Supermicro 5018R-MR (1U Barebone, 1x CPU, 4x 3.5“ front HDD, 2x 1Gbit)
- 1x Intel E5-2620V3 (6x 2.4GHz, 85W TDP)
- 1x Kingston KVR21R15D4K4/64 (4x 16GB DDR4 reg. ECC RAM)
- 2x Intel DC S3700 100GB (OS SSDs, SATA3)
- 1x Technogroup 5Y 5x11 NBD (Support Pack)

Monitoring Node (1x)

Parts

- 1x Supermicro 5018R-MR (1U Barebone, 1x CPU, 4x 3.5“ front HDD, 2x 1Gbit)
- 1x Intel E5-2620V3 (6x 2.4GHz, 85W TDP)
- 1x Kingston KVR21R15D4K4/64 (4x 16GB DDR4 reg. ECC RAM)
- 2x Intel DC S3700 100GB (OS SSDs, SATA3)
- 1x Seagate ST4000NM0033 4TB (Data HDDs, SATA3)
- 1x Western Digital WD4001FFSX 4TB (Data HDDs, SATA3)
- 1x Technogroup 5Y 5x11 NBD (Support Pack)

Vendor Bids

Notes

General

- guiding principle: optimally balanced system for 2x 50'000.- budget
- all prices are including VAT
- all systems equal for the base specification, differences in upgradeability

Dell/HP/Lenovo

- no declaration which HDDs are used exactly (cluster risk)
- no declaration which SSDs are used exactly (deal breaker for Ceph Write Journal)

Lenovo

- using small memory modules in some systems is not nice for future RAM upgrades

Personal Experience

- Brack is uncomplicated and fast
- Lenovo does not get the Bid right and takes unusually long

Prices: Supermicro

System	#	Item Price	Total Price
Ceph OSD	6	5'928.55 CHF	35'571.30 CHF
Ceph MON	3	4'217.70 CHF	12'653.10 CHF
Ceph Total	**9**		**48'224.40 CHF**
OpenStack Compute	6	4'853.70 CHF	29'122.20 CHF
OpenStack Admin	2	3'483.70 CHF	6'967.40 CHF
Monitoring	1	3'925.05 CHF	3'925.05 CHF
OpenStack Total	**9**		**40'014.65 CHF**
Total	**18**		**88'239.05 CHF**

Prices: Dell

System	#	Item Price	Total Price	Supermicro	%
Ceph OSD	6	11'529.73 CHF	69'178.38 CHF	35'571.30 CHF	194.48%
Ceph MON	3	7'089.46 CHF	21'268.38 CHF	12'653.10 CHF	168.09%
Ceph Total	**9**		**90'446.76 CHF**	**48'224.40 CHF**	**187.55%**
OpenStack Compute	6	8'229.61 CHF	49'377.66 CHF	29'122.20 CHF	169.55%
OpenStack Admin	2	6'425.74 CHF	12'851.48 CHF	6'967.40 CHF	184.45%
Monitoring	1	7'416.26 CHF	7'416.26 CHF	3'925.05 CHF	188.95%
OpenStack Total	**9**		**69'645.40 CHF**	**40'014.65 CHF**	**174.05%**
Total	**18**		**160'092.16 CHF**	**88'239.05 CHF**	**181.43%**

Prices: Lenovo

System	#	Item Price	Total Price	Supermicro	%
Ceph OSD	6	7'530.57 CHF	45'183.42 CHF	35'571.30 CHF	127.02%
Ceph MON	3	5'503.03 CHF	16'509.09 CHF	12'653.10 CHF	130.47%
Ceph Total	**9**		**61'692.51 CHF**	**48'224.40 CHF**	**127.93%**
OpenStack Compute	6	6'333.37 CHF	38'000.22 CHF	29'122.20 CHF	130.49%
OpenStack Admin	2	4'901.70 CHF	9'803.40 CHF	6'967.40 CHF	140.70%
Monitoring	1	4'893.03 CHF	4'893.03 CHF	3'925.05 CHF	124.66%
OpenStack Total	**9**		**52'696.65 CHF**	**40'014.65 CHF**	**131.69%**
Total	**18**		**114'389.16 CHF**	**88'239.05 CHF**	**129.64%**

Prices: HP

System	#	Item Price	Total Price	Supermicro	%
Ceph OSD	6	9'239.40 CHF	55'436.40 CHF	35'571.30 CHF	155.85%
Ceph MON	3	5'866.56 CHF	17'599.68 CHF	12'653.10 CHF	139.09%
Ceph Total	**9**		**73'036.08 CHF**	**48'224.40 CHF**	**151.45%**
OpenStack Compute	6	6'435.72 CHF	38'614.32 CHF	29'122.20 CHF	132.59%
OpenStack Admin	2	5'088.96 CHF	10'177.92 CHF	6'967.40 CHF	146.08%
Monitoring	1	5'952.96 CHF	5'952.96 CHF	3'925.05 CHF	151.67%
OpenStack Total	**9**		**54'745.20 CHF**	**40'014.65 CHF**	**136.81%**
Total	**18**		**127'781.28 CHF**	**88'239.05 CHF**	**144.81%**

Buy Decision

Supermicro

System	#	Item Price	Total Price
Ceph OSD	6	5'928.55 CHF	35'571.30 CHF
Ceph MON	3	4'217.70 CHF	12'653.10 CHF
Spare Parts	1	1'324.05 CHF	1'324.05 CHF
Ceph Total	**9**		**49'548.45 CHF**
OpenStack Compute	6	4'853.70 CHF	29'122.20 CHF
OpenStack Admin	2	3'483.70 CHF	6'967.40 CHF
Monitoring	1	3'925.05 CHF	3'925.05 CHF
OpenStack Total	**9**		**40'014.65 CHF**
Total	**18**		**89'563.10 CHF**

Upgrade Options

Upgrade Options

Ceph

- Capacity: use all OSD disk trays (from 144TB/72TB to 288TB brutto/144TB netto): **CHF +7'944.30 (later possible)**
- Performance: use more RAM (from 64GB/node to 128GB/node): **CHF +3'816.00 (later possible)**
- Performance: use more CPU (from 6x 2.4GHz to 8x 2.4GHz): **CHF +1'512 (now or never) [yes]**
- Performance: use more SSD cache (from 4GB/OSD to 8GB/OSD): **CHF +2'526.00 (now or never) [yes]**
- Scale out with additional OSDs (48TB brutto/24TB netto): **CHF +7'252.60/node (later possible)**

Upgrade Options

Ceph

- Capacity: use all OSD disk trays (from 144TB/72TB to 288TB brutto/144TB netto): **CHF +7'944.30 (later possible)**
- Performance: use more RAM (from 64GB/node to 128GB/node): **CHF +3'816.00 (later possible)**
- Performance: use more CPU (from 6x 2.4GHz to 8x 2.4GHz): **CHF +1'512 (now or never) [yes]**
- Performance: use more SSD cache (from 4GB/OSD to 8GB/OSD): **CHF +2'526.00 (now or never) [yes]**
- Scale out with additional OSDs (48TB brutto/24TB netto): **CHF +7'252.60/node (later possible)**

OpenStack

- Performance: use more RAM (from 128GB/node to 256GB/node): **CHF +7'632.00 (later possible)**
- Performance: use more CPU Cores (from 6x 2.4GHz to 8x 2.4GHz): **CHF +3'024 (now or never) [yes]**
- Scale out with additional Compute Nodes: **CHF +4'853.70/node (later possible)**

Upgrade Options

Ceph

- Capacity: use all OSD disk trays (from 144TB/72TB to 288TB brutto/144TB netto): **CHF +7'944.30 (later possible)**
- Performance: use more RAM (from 64GB/node to 128GB/node): **CHF +3'816.00 (later possible)**
- Performance: use more CPU (from 6x 2.4GHz to 8x 2.4GHz): **CHF +1'512 (now or never) [yes]**
- Performance: use more SSD cache (from 4GB/OSD to 8GB/OSD): **CHF +2'526.00 (now or never) [yes]**
- Scale out with additional OSDs (48TB brutto/24TB netto): **CHF +7'252.60/node (later possible)**

OpenStack

- Performance: use more RAM (from 128GB/node to 256GB/node): **CHF +7'632.00 (later possible)**
- Performance: use more CPU Cores (from 6x 2.4GHz to 8x 2.4GHz): **CHF +3'024 (now or never) [yes]**
- Scale out with additional Compute Nodes: **CHF +4'853.70/node (later possible)**

Other

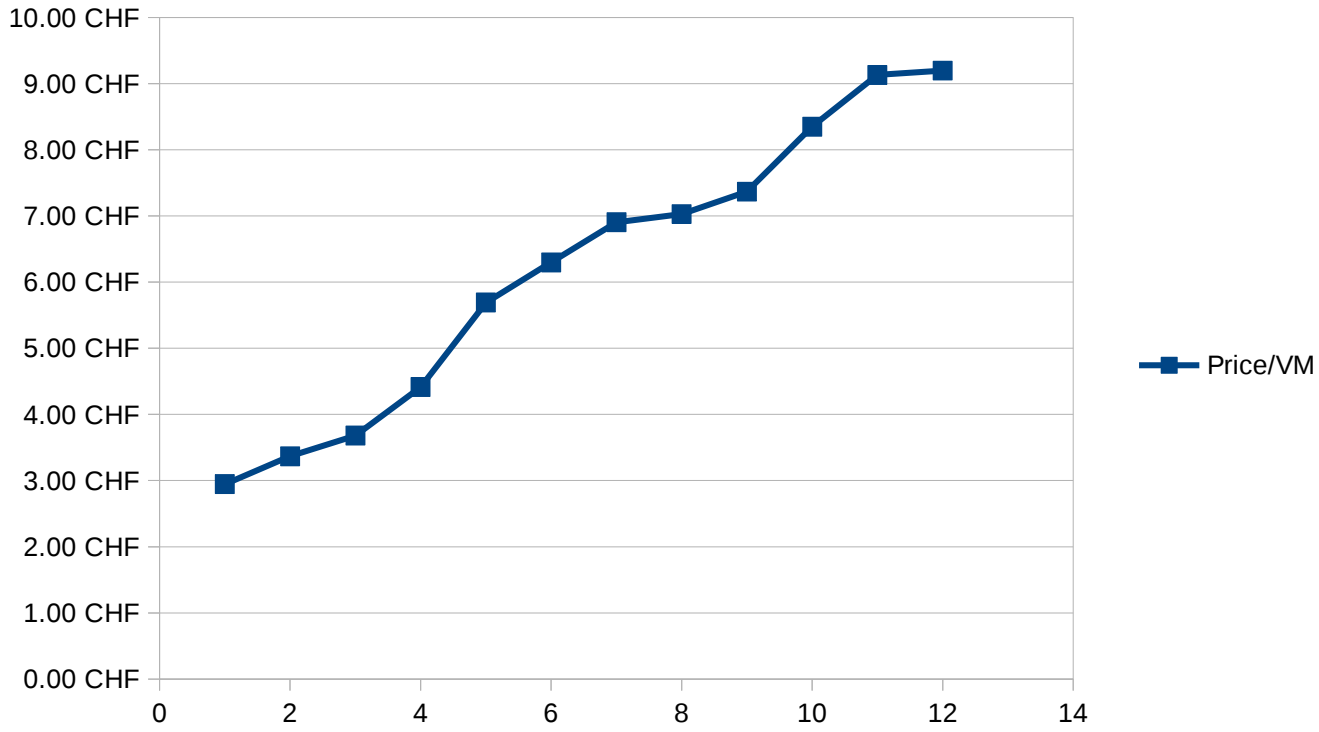
- OneRNG Entropy Source: CHF 50/node, **CHF +750 (~September 2015) [yes]**

Off-Topic: CPU Core Price Ratio (1/2)

Model	CPU #	Frequency (Ghz)	VM #	Item Price	Price/Core	Price/VM
E5-2603v3	6	1.6	76.8	226.25 CHF	37.71 CHF	**2.95 CHF**
E5-2609v3	6	1.9	91.2	307.00 CHF	51.17 CHF	**3.37 CHF**
E5-2620v3	**6**	**2.4**	115.2	**424.00 CHF**	**70.67 CHF**	**3.68 CHF**
E5-2630v3	8	2.4	153.6	678.00 CHF	84.75 CHF	**4.41 CHF**
E5-2640v3	8	2.6	166.4	947.00 CHF	118.38 CHF	**5.69 CHF**
E5-2650v3	10	2.3	184.0	1'159.00 CHF	115.90 CHF	**6.30 CHF**
E5-2660v3	10	2.6	208.0	1'436.00 CHF	143.60 CHF	**6.90 CHF**
E5-2670v3	12	2.3	220.8	1'552.00 CHF	129.33 CHF	**7.03 CHF**
E5-2680v3	12	2.5	240.0	1'768.00 CHF	147.33 CHF	**7.37 CHF**
E5-2690v3	12	2.6	249.6	2'084.00 CHF	173.67 CHF	**8.35 CHF**
E5-2695v3	14	2.3	257.6	2'352.00 CHF	168.00 CHF	**9.13 CHF**
E5-2697v3	14	2.6	291.2	2'678.00 CHF	191.29 CHF	**9.20 CHF**

Note: end-user prices, but ratio is the same

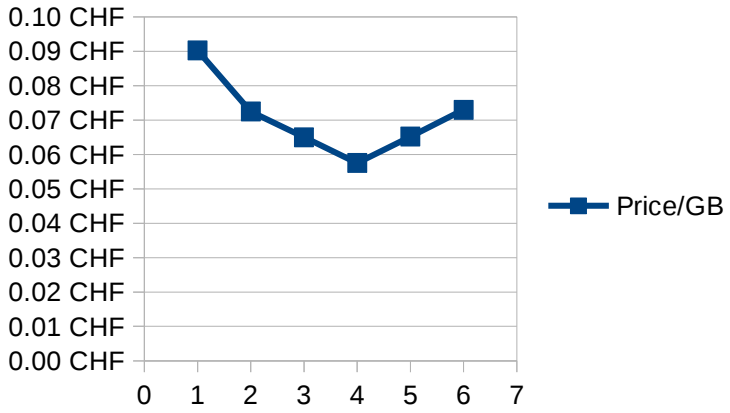
Off-Topic: CPU Core Price Ratio (2/2)



Off-Topic: Disk Price Ratio

Model	TB #	Item Price	Price/GB
ST1000NM0033	1	90.25 CHF	0.09 CHF
ST2000NM0033	2	145.00 CHF	0.07 CHF
ST3000NM0033	3	195.00 CHF	0.07 CHF
ST4000NM0033	**4**	**230.05 CHF**	**0.06 CHF**
ST5000NM0024	5	326.00 CHF	0.07 CHF
ST6000NM0024	6	438.00 CHF	0.07 CHF

Note: end-user prices, but ratio is the same



Networking Options

Notes

On SDN

- in an ideal world we would buy two SDN switches
- in reality they are too expensive and the whole area is not mature enough
- no real gain yet, we are just starting with a few VLANs anyway
- using bridges first, going for open-vswitch instead/prefered

Notes

On SDN

- in an ideal world we would buy two SDN switches
- in reality they are too expensive and the whole area is not mature enough
- no real gain yet, we are just starting with a few VLANs anyway
- using bridges first, going for open-vswitch instead/prefered

On Ports numbers

System	#	Backend	Frontend	Admin	Total w/o Admin	Total w/ Admin
Ceph OSD	6	2	2	1	**24**	**30**
Ceph MON	3	0	2	1	**6**	**9**
OpenStack Compute	6	0	2	1	**12**	**18**
OpenStack Admin	2	0	2	1	**4**	**6**
Monitoring	1	0	2	1	**2**	**3**
Total					**48**	**66**
Logging	1	0	2	1	**2**	**3**

Network Options (1/2)

1Gbit Copper w/o PoE

Item	#	Item Prize	Total Prize
Cisco WS-C2960X-48TD-L	2	2'533.00 CHF	5'066.00 CHF
Cisco 2960X-STACK	2	665.00 CHF	1'330.00 CHF
Flexoptix P.8596.02-MSA01	8	75.45 CHF	603.60 CHF
Total	**12**		**6'999.60 CHF**

1Gbit Copper w/ PoE+

Item	#	Item Prize	Total Prize
Cisco WC2960X-48FPD-L	2	3'564.00 CHF	7'128.00 CHF
Cisco 2960X-STACK	2	665.00 CHF	1'330.00 CHF
Flexoptix P.8596.02-MSA01	8	75.45 CHF	603.60 CHF
Total	**12**		**9'061.60 CHF**

Network Options (2/2)

10Gbit Fiber: Ceph only

Item	#	Item Price	Total Price
Cisco WS-C4500X-24X-ES	2	16'465.10 CHF	32'930.20 CHF
Cisco WS-C2960X-48TD-L	1	2'533.00 CHF	2'533.00 CHF
Supermicro AOC-STGN-i2S	15	433.00 CHF	6'495.00 CHF
Flexoptix P.8596.02-MSA01	40	75.45 CHF	3'018.00 CHF
Total	**18**		**44'976.20 CHF**

10Gbit Fiber: Ceph and OpenStack

Item	#	Item Price	Total Price
Cisco WS-C4500X-32SFP+	2	19'482.45 CHF	38'964.90 CHF
Cisco WS-C2960X-48TD-L	1	2'533.00 CHF	2'533.00 CHF
Supermicro AOC-STGN-i2S	23	433.00 CHF	9'959.00 CHF
Flexoptix P.8596.02-MSA01	56	75.45 CHF	4'225.20 CHF
Total	**26**		**55'682.10 CHF**

Why not use a Nexus 2000 (FEX)?

- **too expensive**
  - Copper FEX is **5'000 CHF/piece**, Fiber FEX is **7'500 CHF/piece**; we would need 2

Why not use a Nexus 2000 (FEX)?

- **too expensive**
  - Copper FEX is **5'000 CHF/piece**, Fiber FEX is **7'500 CHF/piece**; we would need 2
- **too slow**
  - Port to Port traffic on the **same FEX** goes through the Nexus
  - Port to Port traffic from **one FEX to the other** goes through the Nexus
  - Copper FEX←→ Nexus is only 40GBit (**1.2:1 oversubscription**)
  - Fiber FEX ←→ Nexus is only 240Gbit (**2:1 oversubscription**)

Why not use a Nexus 2000 (FEX)?

- **too expensive**
  - Copper FEX is **5'000 CHF/piece**, Fiber FEX is **7'500 CHF/piece**; we would need 2
- **too slow**
  - Port to Port traffic on the **same FEX** goes through the Nexus
  - Port to Port traffic from **one FEX to the other** goes through the Nexus
  - Copper FEX←→ Nexus is only 40GBit (**1.2:1 oversubscription**)
  - Fiber FEX ←→ Nexus is only 240Gbit (**2:1 oversubscription**)
- **not worth it**
  - administration overhead is not so much of a problem (all ports same config)

Why not use a Nexus 2000 (FEX)?

- **too expensive**
  - Copper FEX is **5'000 CHF/piece**, Fiber FEX is **7'500 CHF/piece**; we would need 2
- **too slow**
  - Port to Port traffic on the **same FEX** goes through the Nexus
  - Port to Port traffic from **one FEX to the other** goes through the Nexus
  - Copper FEX←→ Nexus is only 40GBit (**1.2:1 oversubscription**)
  - Fiber FEX ←→ Nexus is only 240Gbit (**2:1 oversubscription**)
- **not worth it**
  - administration overhead is not so much of a problem (all ports same config)
- **unusable for bridges/open-vswitch**
  - FEX is for end-host connectivity only, it blocks BPDU

- For more information: Cisco Nexus 2000 Series Fabric Extender Software Configuration Guide

Totals

Finally.. :)

System	#	Item Price	Total Price
Ceph OSD	6	5,928.55 CHF	35'571.30 CHF
Ceph MON	3	4,217.70 CHF	12'653.10 CHF
Spare Parts	1	1,324.05 CHF	1'324.05 CHF
Ceph Total	**9**		**49'548.45 CHF**
OpenStack Compute	6	4,853.70 CHF	29'122.20 CHF
OpenStack Admin	2	3,483.70 CHF	6'967.40 CHF
Monitoring	1	3,925.05 CHF	3'925.05 CHF
OpenStack Total	**9**		**40'014.65 CHF**
Cisco WS-C4500X-32SFP+	2	19,482.45 CHF	38'964.90 CHF
Cisco WS-C2960X-48TD-L	1	2,533.00 CHF	2'533.00 CHF
Supermicro AOC-STGN-i2S	23	433.00 CHF	9'959.00 CHF
Flexoptix P.8596.02-MSA01	56	75.45 CHF	4'225.20 CHF
Networking Total	**26**		**55'682.10 CHF**
Ceph 8 Core Option	1	1,512.00 CHF	1'512.00 CHF
Ceph 200GB SSD Option	1	2,526.00 CHF	2'526.00 CHF
OpenStack 8 Core Option	1	3,025.00 CHF	3'025.00 CHF
OneRNG Option	1	750.00 CHF	750.00 CHF
Options Total	**4**		**7'813.00 CHF**
Total	**48**		**153'058.20 CHF**

Total costs including everything (except network cables)

Further Steps

Further Steps

- Buying Hardware now
- Roadmap and Timeline after Bootstrap Meeting (IS+SSB) on 2015-07-17

Proposed Upgrades in 2016

- Ceph: use all OSD disk trays (from 144TB/72TB to 288TB brutto/144TB netto): **CHF +7'944.30**
- Ceph: use more RAM (from 64GB/node to 128GB/node): **CHF +3'816.00**
- OpenStack: use more RAM (from 128GB/node to 256GB/node): **CHF +7'632.00**

Thank You for Your Attention.

❤ Source Code is freely available

git clone git://git.bfh.ch/git/staff/bad9/other/talks.git