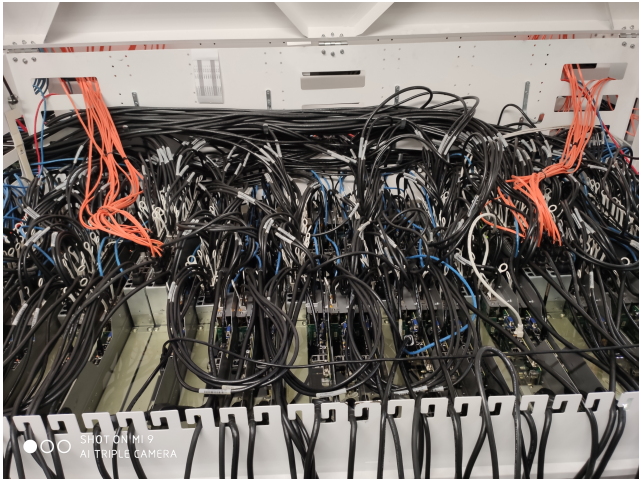
Impressions from 5 Years of Oil-Based Liquid Immersion Cooling

Siegfried Höfinger and Ernst Haunschmid

Vienna Scientific Cluster
A Joint HPC Facility of Austrian Universities
TU Wien

November 15, 2019

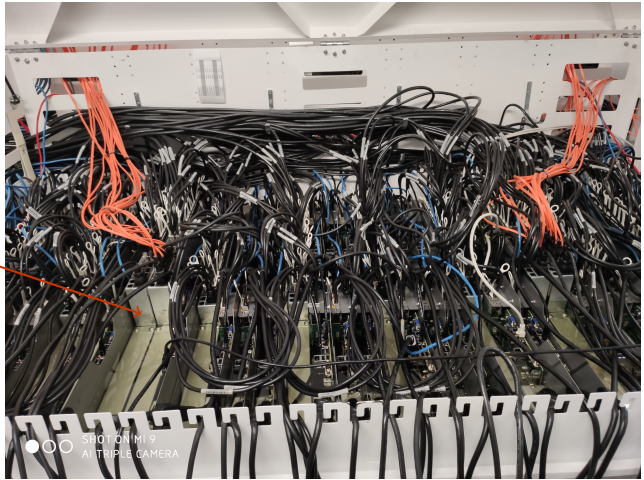
VSC-3 Container



©M Hickel

VSC-3 Container

Fully Immersed in White Mineral Oil



©M Hickel

Immersion Cooling

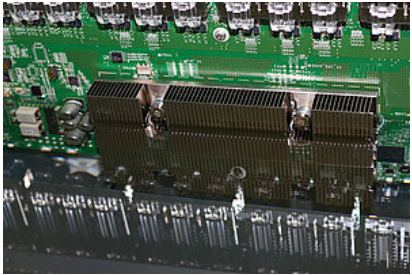
- Similar advantages to water cooling
- Lower costs
  - Very "simple" design
  - Infrastructure can be re-used for next generation system
  - Fewer components on the node level
- Very silent server room
- Oil cooling is widely used in industry
  - Transformers, train engines
- Cray-2 (1985) was based on immersion cooling (3M, perfluorocarbon)

VSC-3

- Contract was awarded to Clustervision
- Installation, Summer 2014
  - 2020 Supermicro based compute nodes
    - 2 x Intel E5-2650 v2 (8 cores, 2.6 GHz)
    - 64 GB main memory
    - 2 x Intel QDR Infiniband HCAs
  - Intel dual-rail QDR Infiniband
  - Fabric is composed of 8 islands
  - Each island includes about 4500 cores

VSC-3 Cooling

- Oil based immersion cooling
  - Provided by Green Revolution Cooling (GRC)
  - More than 32.000 litres of white mineral oil
    - In addition 5.000 litres backup
  - Pumps under the floor (raised)
- No air cooling in server room
  - Room temperature up to 45 °C
  - Only Infiniband core (single rack) is air-cooled
- Storage is located in nearby room



Containers

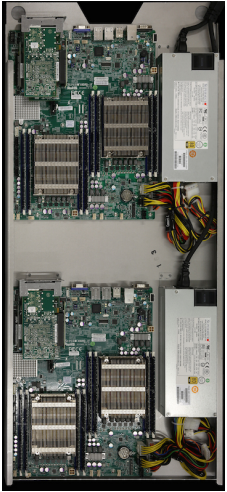
- 23 containers, each with
  - 24 U for nodes (96 nodes per container)
  - 14 U for switches (12 IB + 2 Ethernet)
- 32.5" depth
- Approx 1200 litres mineral oil
- Oil temperature up to 55 °C
- Both input output pipes at the top of the containers

Pump Modules

- Located below the floor level
- Heat transfer from oil to water
- Turbulent oil flow within the containers
- 2 independant and redundant oil pumps
- 2 heat exchangers in series
- 1 μ bypass filter removing particles from oil

Compute Nodes

- Provided by Supermicro (including assembly)
- 2 nodes per blade
  - Metal plate
  - 2 motherboards with 2 IB HCAs
  - 2 power supplies
  - Nothing else
  - Modifications:
    - Indium foil instead of thermal grease
    - Fans are disabled by firmware



Challenges

- White mineral oil is flammable
  - Flash point approx 180 °C
  - Fire prevention and detection system
  - Fire extinguishing system (Trigon 300)
- Oil leakage prevention
  - Metal compartments, foundation made of water-proof concrete
  - Soft barriers may be leaking
  - Temperature control / shutdown

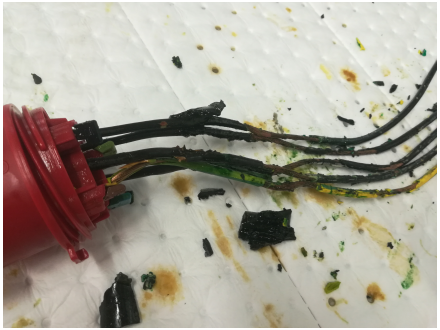
Challenges cont.

- Warranty
  - Full warranty for compute nodes assembled by Supermicro
- Serviceability
  - Full hardware support for three years by Clustervision
- Plasticizer elute into oil (long term process)
  - Stiff cables tend to break
- Optical cables may fail when fully submerged
  - Observation for different technologies:
    - ArrayVCSEL ✗
    - Silicon photonics ✓
  - Mid layer IB switches are raised above level of oil
  - IB core switches in air-cooled rack

Oil Leakage

- Defective filter
  - Filter element not properly replaced
  - More than 300 litres leaked
- Oil dripping via capillary effect through power cables
  - Floor slippery
  - Oil dripping through tiles
  - Defective electrical cabling
  - Additional tray to collect dripping oil
- All in all more than 1.700 litres lost in 5 years

Power Cable Damage



Particles in the Air

Serverraum mit Ölkühlung														Immersion Cooling
Stufe	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Di [µm]	0.010	0.022	0.040	0.071	0.12	0.19	0.31	0.48	0.75	1.2	2.0	3.0	4.4	7.3
Anzahl [cm^{-3}]	19575	4276	2872	1193	995	346	73	14	<1	<0.6	<0.3	<0.2	<0.1	<0.1

Serverraum mit Luftkühlung														Air Cooling
Stufe	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Di [µm]	0.010	0.022	0.040	0.071	0.12	0.19	0.31	0.48	0.75	1.2	2.0	3.0	4.4	7.3
Anzahl [cm^{-3}]	3354	1368	168	337	583	36*	136	0.2	<1	<0.6	n.a.	<0.2	0.6*	<0.1

*verkürzter Mittelungszeitraum

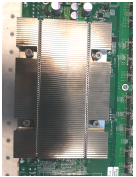
- Lots of very small sized oil particles
- Very thin layer of oil on surfaces of metals
- Safety regulations
  - 5 mg/m^3 monitoring limit for oil mist (detected <0.01 mg/m^3)

IB Switch Specs/Limitations

- Intel/Qlogic IB switches
  - Air input temperature up to 50 °C
    - Stable operation up to 60 °C
  - Immersion cooling
    - Stable operation up to 50 °C
    - High risk of failure at higher temperatures
    - But many switches start failing at much lower oil temperatures
- Compute nodes (for comparison)
  - Tested with oil temperatures up to 75 °C

IB Switch Failures

- Symptoms:
  - Some are really broken
  - Some become operational again when cooled down, but only at very low oil temperatures
- IB Switch ASIC (thermal interface material, TIM)
  - Soldered (sTIM)
    - Intel Xeon up to Broadwell
  - Polymer (pTIM)
    - Intel Skylake
    - Qlogic switch ASICs



Balance: IB Switch Failures

- Intel/Qlogic QDR switch 12200/12300
  - 280 edge switches immersion-cooled
  - 24 core switches air-cooled
  - 12 core switches air-cooled (storage room)
- More than 100 switches replaced within warranty (3 years)
- Only 2 switches failed in the last 20 months

Balance: Ethernet Switch Failures

- Supermicro 48 port 1G switch
  - 46 switches in total
  - No switch faiures within initial 4 years
  - 8 switches failed in the last 10 months

Balance: Node Failures

- More than 300 defective nodes in installation/acceptance period
  - Nodes assembled by Supermicro
  - No burn-in, no testing
- Failing components (in descending order of occurrence)
  - IB HCA related problems
    - HCA not properly connected to riser card/board
    - Broken HCA
  - ECC errors on memory modules
  - Power supply
  - 20 node failures within the last 18 months

Cooling Efficiency

Power consumption [kW] / Outdoor temperature	10°C	25°C	32°C
VSC-3 (compute nodes and network, without storage)	400,0	400,0	400,0
Water pumps	2,4	2,8	3,8
Oil pumps	< 0,4	4,0	6,0
Fans (free cooler, no wind)	< 0,1	< 1,0	21,0
Cooling overhead in percent of VSC-3 power consumption	**< 1%**	**< 2%**	**< 8%**

Remarks:

- Power consumption, 1h averages in kW
- Average temperature in Vienna over 30 years is 10,1 °C
- Cooling system configuration is still being optimized
- Overall goal is cost efficiency throughout total life cycle of HPC system

Lessons Learned

- Oil-based immersion cooling is still a viable option — very cost efficient

Lessons Learned

- Oil-based immersion cooling is still a viable option — very cost efficient
- Many improvements already implemented in newer generation immersion cooling systems

Lessons Learned

- Oil-based immersion cooling is still a viable option — very cost efficient
- Many improvements already implemented in newer generation immersion cooling systems
- Do not immerse IB switches

Lessons Learned

- Oil-based immersion cooling is still a viable option — very cost efficient
- Many improvements already implemented in newer generation immersion cooling systems
- Do not immerse IB switches
- Caution with power cable damage of electrical insulation — plasticizer extraction

Lessons Learned

- Oil-based immersion cooling is still a viable option — very cost efficient
- Many improvements already implemented in newer generation immersion cooling systems
- Do not immerse IB switches
- Caution with power cable damage of electrical insulation — plasticizer extraction
- Be prepared for the unforeseen

Lessons Learned

- Oil-based immersion cooling is still a viable option — very cost efficient
- Many improvements already implemented in newer generation immersion cooling systems
- Do not immerse IB switches
- Caution with power cable damage of electrical insulation — plasticizer extraction
- Be prepared for the unforeseen

... Thank you for your attention !