HPC TCO: Cooling and Computer Room Efficiency

Route Plan

- Motivation (Why do we care?)
- HPC Building Blocks: Compuer Hardware (What‘s inside my dataroom? What needs to be cooled?)
- HPC Building Blocks: Subsystems (What is around my data room and what do I need it for?)
- Traditional Air Cooling (What we have been doing for decades)
- Chillers (So where does all the heat go?)
- Cooling Towers (Can‘t we do that for free?)
- Warm water cooling (Well thats sounds like an oxymoron)

Route Plan

- Liquid summersion coling (taking the servers for a bath)
- Modular Data Centers
- Efficiency Metrics (Where this gets boring)
  - PUE
  - ERE
  - Space Utilization
  - computation
- Low vs High Density DC (Tight is better)
- Liquid Cooling vs. Air Cooling (Wet is better)

Motivation

- If HPCs are not constantly cooled, they overheat and, in the worst case scenario, may be damaged.
- Most of the energy consumed by an HPC or DC that is not used for computation is used in keeping the equipment at an appropriate temperature.
- The cooling system used in an HPCs has mayor implications in the design of the data room and the facility as a whole and will significantly affect the capital and operational costs.

Motivation

- Power consumption and density have been rising steadily in the past and are expected to keep on rising.
- Keeping density high is in our interest because it let us operate very powerful machine with low space requirements.
- Archiving high density in a reliable and safe way is not a trivial task..

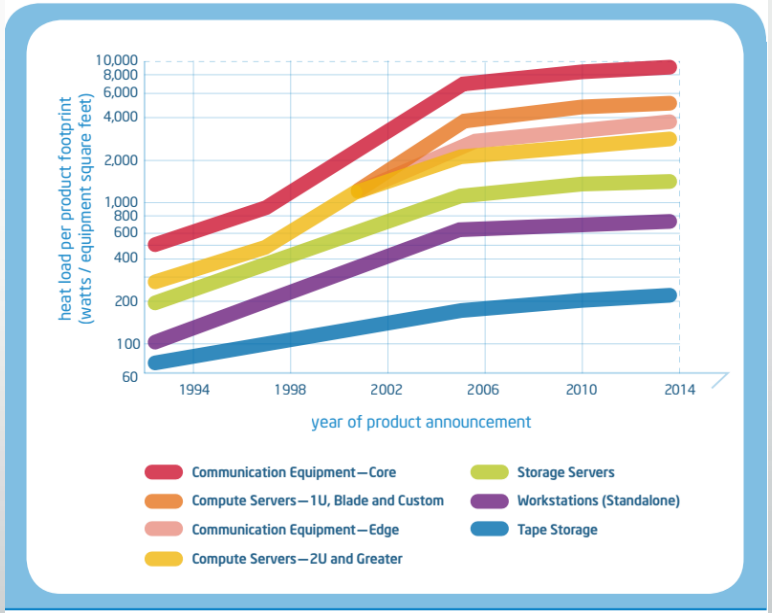


Figure 1. ASHRAE Datacom Trend Chart showing increasing density over time

HPC Building Blocks: Compuer Hardware

- Current HPC systems are built from commodity server processors
- Modular servers like 1U or blades are put into racks and are are interconected though some kind of network fabric.
- The more servers/blades we put into a single rack, the higher the heat concentration.
- The layout of the data room depends greatly on how we plan to cool the system.

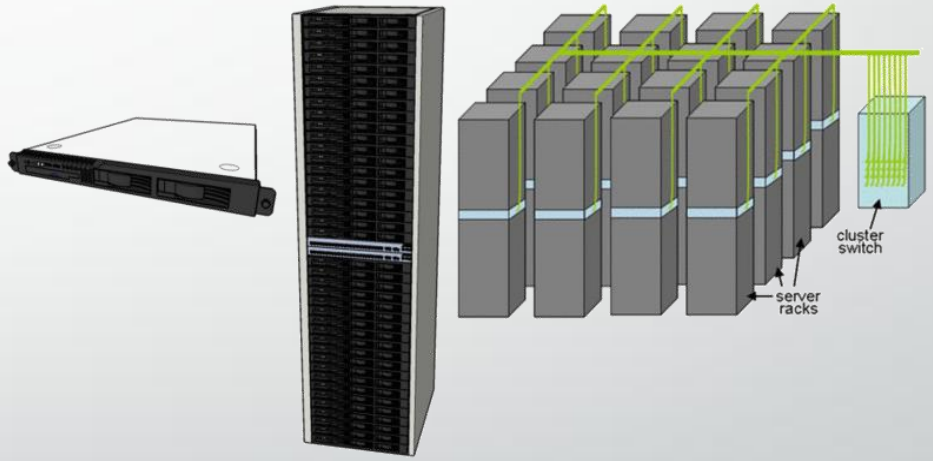


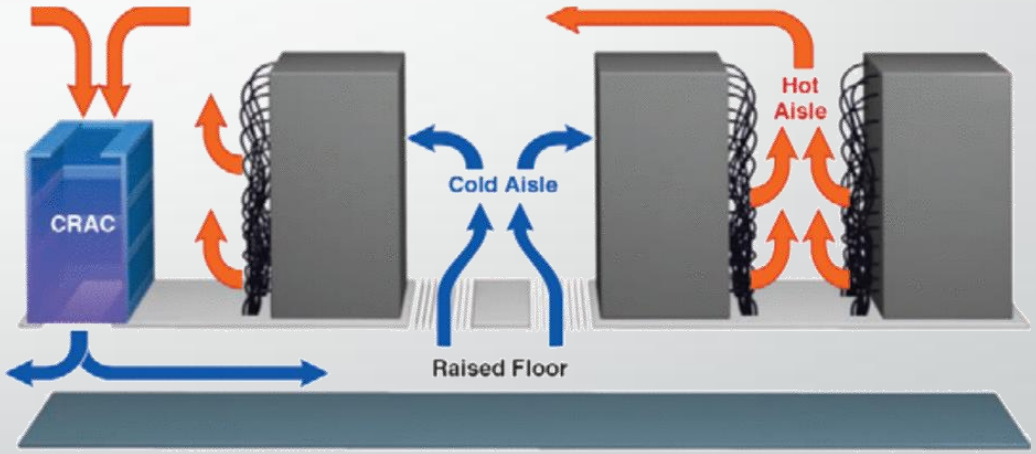
FIGURE 1.1: Typical elements in warehouse-scale systems: 1U server (left), 7′ rack with Ethernet switch (middle), and diagram of a small cluster with a cluster-level Ethernet switch/router (right).

HPC Building Blocks: Subsystems

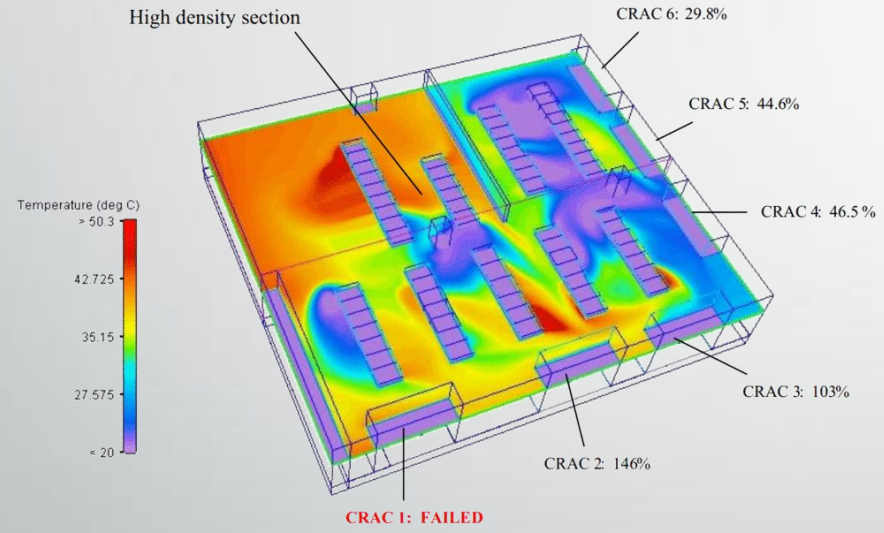
- Electricity is supplied either from the grid or an on-site generator; this is then conditioned on-site before delivery to the computer room.
- Central plant chillers provide continuous supply of cold water for use in the computer room air-conditioning (CRAC) units.
- Additionally, apart from the rack switches, network connectivity must be provided for enabling data transmission within the HPC.

Traditional Air Cooling

- The most common datacenter layout in today's data centers is a repeating of rows of racks side-by-side with alternating cold aisles and hot aisles.
- Computer Room Air Conditioning units (CRACs) pull hot air across chillers and distribute the cool air beneath the raised floor.



Traditional Air Cooling

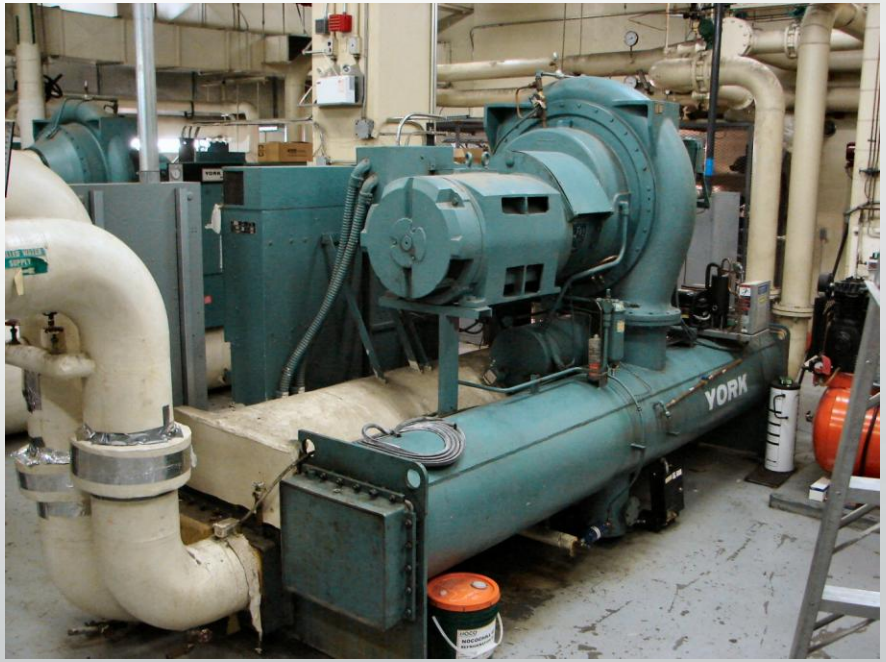


A CFD simulation by HP from 2005

- The computer room needs to be carefully designed
- Possible equipment failures have to be taken into account
- In high density setups computational fluid dynamics (CFD) simulations might be needed to achieve an efficient design.

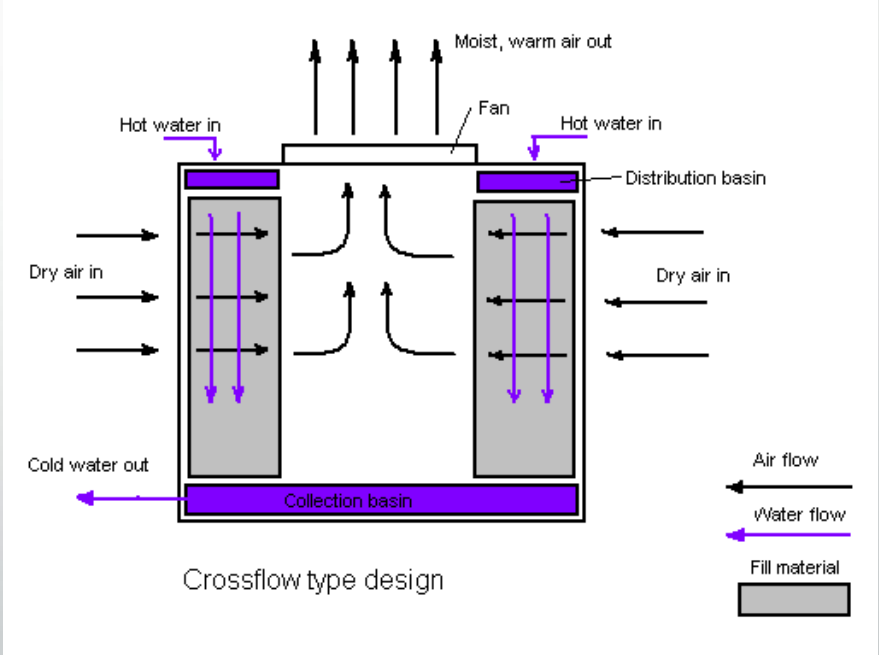
Chillers

- CRAC units usually need cold water or some kind of liquid coolant for heat exchange with the hot air.
- After the heat exchange the water or liquid coolant has to be cooled down for reuse.
- This is usually done in a chiller.
- A chiller is a machine that removes heat from a liquid via a vapor-compression or absorption refrigeration cycle.
- Chillers can consume a lot of energy.



Cooling Towers (Free Cooling)

- Hot water comes at the top of the towers
- Waters fall down the tower and cools down mainly through evaporation.
- Much cheaper operational costs than traditional chillers.
- Redundancy can be costly.
- It's tricky to prevent freezing in cold climates.



Crossflow type design

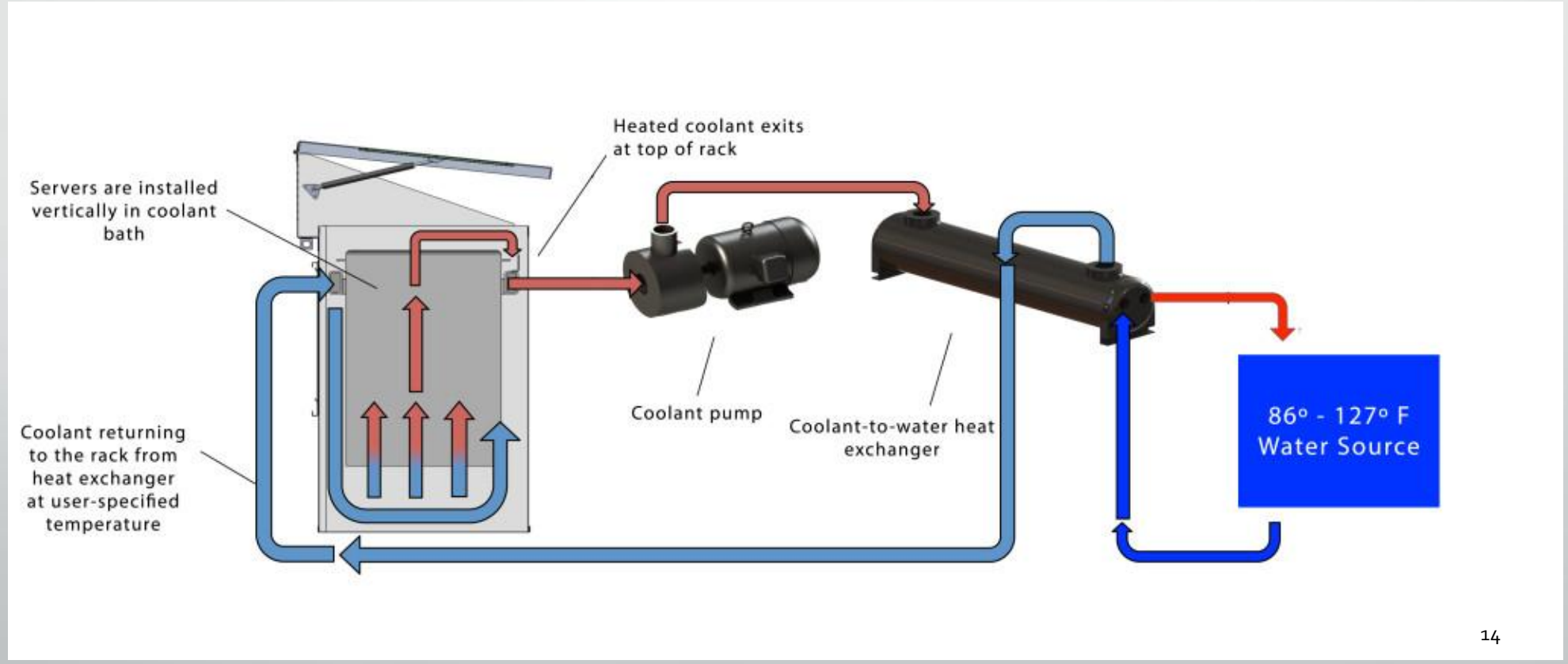
Warm Water Cooling

- Water is circulated as close to the computer hardware as possible in order to allow heat exchange from the hardware into the water even at relatively high temperatures .
- Keeping the water temperature above the wet-bulb temperature of the ambient air means that free cooling is possible all the time completely removing the need for chillers.
- It allows for energy reuse for heating purposes (SuperMUC) or to drive an adsorption chiller (iDataCool).

Liquid Submersion Cooling

- The servers are immersed in a non conductive coolant.
- The direct contact of the coolant with the hardware allows for a more efficient heat exchange.
- The servers are placed vertically in the racks
- The racks have a very "exotic" form factor that uses a wider area than conventional racks
- Like warm water cooling, it allows for higher densities than air cooling.

Liquid Submersion Cooling



Modular Data Centers

- Portable modular data centers, fits data center equipment (servers, storage and networking equipment) into a standard shipping container, which is then transported to a desired location.
- Containerized data centers typically come outfitted with their own cooling systems.
- Modular data centers typically consist of standardized components, making them easier and cheaper to build
- Their main advantage is their rapid deployment.
- Still need considerable infrastructure around them

Efficiency Metrics: PUE

- $PUE = \frac{Total\ Facility\ Energy\ Consumption}{IT\ Equipment\ Energy\ Consumption}$
- PUE stands for power usage effectiveness. It measures how much of the electrical power entering a data center is effectively used for the IT load.
- The perfect theoretical PUE is equal to 1, that means all of the energy entering the data center is used to feed IT equipment and nothing is wasted.

Efficiency metrics: ERE

- ERE stands for energy reuse efficiency and it is the ratio between the energy balance of the data center and the energy absorbed by the IT equipment.
- $$ERE = \frac{Total\ Facility\ Energy\ Consumption\text{-}Recovered\ Energy}{IT\ Equipment\ Energy\ Consumption}$$
- It was common habit to factor the energy recovery into the PUE, talking about a PUE lower than 1, which makes a mathematical non-sense.
- The ERE can range between 0 and the PUE. (not really)

Efficiency Metrics: Space Utilization

- **Watts/rack** Measures the energy consumption of a single rack
- **Watts/sq ft** is a common but ambiguous metric because it doesn't specify the nature of the area in the denominator.
- **Watts/sq ft** of work cell is the preferred metric for data center to (air cooled) data center benchmarking , or in infrastructure discussions.
- **Layout efficiency** measures the data center square footage utilization and the efficiency of the data center layout. This is defined as racks per thousand square feet and so it is a measure of space efficiency related to the electrically active floor in a data center.

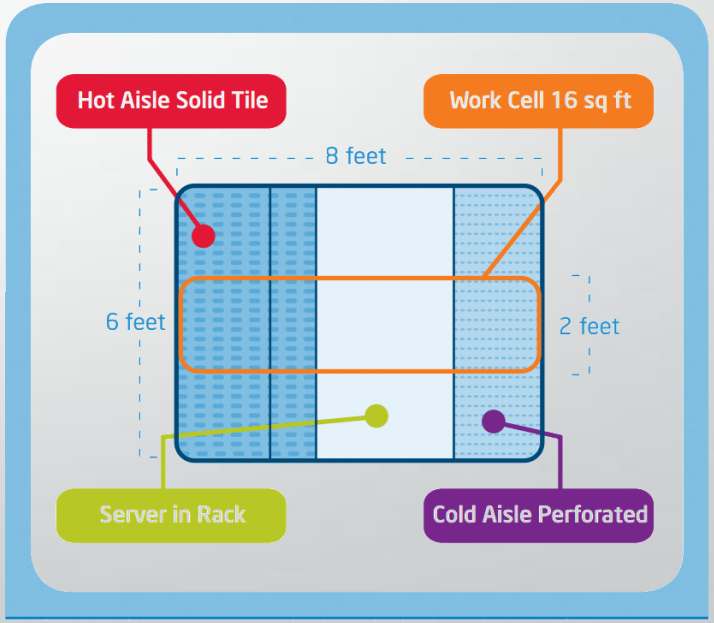


Figure 2. Single work cell in a row of servers in a Data Center, looking down from above

Efficiency Metrics: Computation

- **MFlops/Watt** and **GFlops/Node** are important metrics that should be taken into consideration specially when choosing the processors.
- Together with the space utilization metrics this helps us to get an idea of how much space we will need for a certain computational requirement.

Low vs High Density DC (Comparison by Intel in 2007)

	Low-Density Data Center	High-Density Data Center
# of servers	10,000	10,000
Watts / server	400	400
CFM / server	39	39
kW/rack	6.6	17
Servers / rack	16	42
Total racks	625	238
Sq ft / work cell	16	16
Layout Efficiency (rack/Ksf)	~22	~22
Sq ft of raised floor needed	28,571	10,880

Low vs High Density DC (Comparison by Intel in 2007)

	Low-Density Data Center	High-Density Data Center
Total Airflow CFM	468,000	468,000
Raised floor height	18 inches	30 inches
CFM / rack	749	1966
Total UPS power needed	4 MW	4 MW
Cost of power	10¢/kW-hr	10¢/kW-hr

Low vs High Density DC (Comparison by Intel in 2007)

	Low-Density	High-Density	Notes
Capital Cost – Building	$6,285,620	$2,393,600	$220/sq ft for CSA
Design Cost for CFD	$0	$54,440	Assumes $5/sq ft;
Capital cost taller DC	$0	$239,360	Assumes +10%
Capita cost for 30" RF.	$0	$10,880	$1 sq ft
Lighting	$126,000	$0	NPV(5 yr, i=5%)
IT Equipment (Racks)	$1,875,000	$714,000	$1.5K/ea + $1.5K/install
Oper Cost – Cooling	$1,091,000	$736,000	NPV of 5 yr with i=5%
Total Cost Delta	$9,377,620	$4,148,240	$5.2 M savings

Low vs High Density DC (Comparison by Intel in 2007)

- High density data centers require higher raised floors and a more careful design but the reduced area and improved cooling efficiency make up for it by significantly lowering the capital and operational costs.
- As a result high density datacenters have a lower TCO.

Liquid Cooling vs. Air Cooling (A Comparison by Eurotech)

	Air cooled low density HPC	Air cooled high density HPC	Liquid cooled high density HPC
TFLOP/S	500	500	500
Architecture	1U servers, Intel CPUs	Blades, Intel CPUs	Blades, Intel CPUs
Processors per node	2 x Intel Xeon X5690 6C 3.46 GHz	2 x Intel Xeon X5690 6C 3.46 GHz	2 x Intel Xeon X5690 6C 3.46 GHz
Layout efficiency	22	22	40
GFLOPS/node	165	165	165
Total nodes needed	3030	3030	3030
Servers/blades per rack	35	90	256
cores per rack	420	1080	3072
Server racks needed	87	34	12
Total network equipment	379	379	379

Liquid Cooling vs. Air Cooling (A Comparison by Eurotech)

Total racks for network equipment	9	9	9
Total racks needed	96	43	21
Occupancy (ft2) of electrical active floor	4345	1940	521
Total DC space (in ft2)	8691	3881	1221
(In m2)	800	357	112
Energy consumption			
mflops/w	450	450	450
Total IT Power (Kw)	1,111	1,111	1,111
Total DC power (Kw)	2556	2000	1167
Reliability			
MTBF per node (hours)	560,000	560,000	672,000
MTBF per system (hours)	185	185	222
price of outage per h ($)	5000	5000	5000

Liquid Cooling vs. Air Cooling (A Comparison by Eurotech)

	Air cooled low density	Air cooled high density	Liquid cooled high density
Initial investment costs			
Cost of IT (HW)	$7,500	$7,500	$7,500
Building permits and local taxes	$600	$270	$80
CSA capital costs	$1,910	$850	$260
Capital cost of taller DC	$190	$80	$0
Design cost for CFD	$30	$10	$0
Delta cost raised floor	$20	$10	$0
Fire suppression and detection	$60	$30	$20
Racks	$290	$130	$70
Rack management hardware	$290	$130	$70
Liquid cooling	$0	$0	$810
Total for network equipment	$750	$750	$750
Cooling infrastructure/plumbing	$3,330	$3,330	$500
Electrical	$4,440	$4,440	$4,440
Annual costs (3 years)			
Cost of energy	$5,390	$4,980	$2,900
Retuning and additional CFD	$40	$20	$0
Reactive maintenance (cost of outages)	$720	$720	$600
Preventive maintenance	$450	$450	$450
Facility and infrastructure maintenance	$1,360	$1,180	$730
Lighting	$40	$20	$10
TOTAL TCO	$27,410	$24,900	$18,990

Liquid Cooling vs. Air Cooling (A Comparison by Eurotech)

	Air cooled low density	Air cooled high density	Liquid cooled high density
Cost of energy	$1,690	$1,560	$910
Retuning and additional CFD	$20	$10	$0
Reactive maintenance (cost of outages)	$230	$230	$190
Preventive maintenance	$150	$150	$150
Facility and infrastructure maintenance	$450	$390	$240
Lighting	$20	$10	$5
Annualized 3 years capital costs	$3,390	$3,270	$3,220
Annualized 10 years capital costs	$1,100	$1,100	$820
Annualized 15 years capital costs	$300	$130	$40
ANNUALIZED TCO	$7,350	$6,850	$5,575

Liquid Cooling vs. Air Cooling (A Comparison by Eurotech)

- Liquid cooling allows for higher density DCs.
- It makes possible a higher Watts/Racks ratio and layout efficiency.
- Removes the need for CRACS and other electrically active equipment not used for computation
- Incurs in additional capital costs for plumbing and pipes.
- Achieves a higher reliability by removing vibration and moving parts.

Final Notes

- It's important to make considerations about space needed for our data room when planning a DC or HPC.
- The cooling system is a decisive factor to calculate the needed space and to anticipate the kind of infrastructure that is going to be needed to support the systems.
- Traditional air cooling approaches underperform in most benchmarks when compared with technologies like warm-water cooling or liquid submersion cooling.
- There are other possibilities for cooling not mentioned in this presentation like using naturally occurring cold water (CSCS Swiss National Supercomputing Centre)
- There have been efforts in the last years to create ARM based HPC which could theoretically achieve much better Mflops/watt rations than current systems.