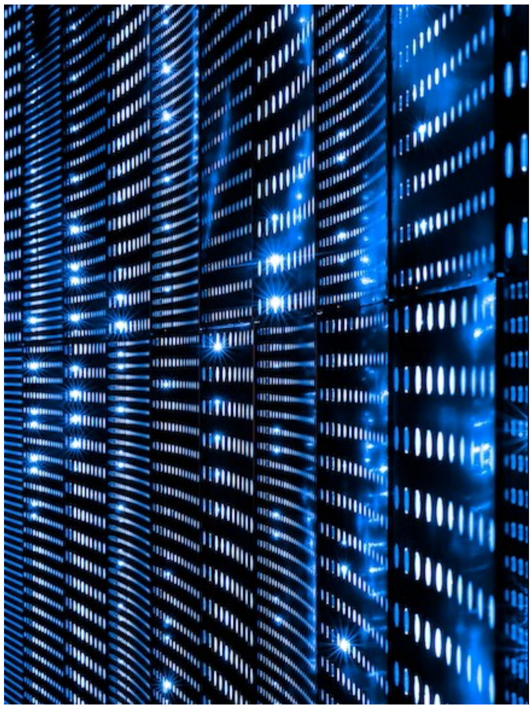
Rotterdam School of Management
Erasmus University

Case Development Centre
FOUNDED BY
Rotterdam School of Management
Erasmus University
RSM Erasmus



No Heat Goes to Waste: Redefining Blockheating's Technological Strategy

RSM - a force for positive change

RSM Erasmus

Introduction

It was an exceptional sunny February afternoon when Jeroen Burks, the founder of Blockheating just got back home to his Delft apartment after a meeting with a potential new client. While going through the memo he wrote during the meeting, he was thinking about what the future of Blockheating could look like for the following 5 or 10 years. He envisioned a bright future that – besides commercial success – could create significant value by reducing energy waste.

The company was created in 2018, almost exactly 5 years before, with the mission of providing data hosting services to clients who are looking to rent computational power and doing so in a way that has a smaller footprint than others on the market by reselling the heat produced by the servers that would otherwise be wasted. Jeroen himself has always been passionate about reducing the waste he creates and finding smart solutions to do so. Having a background in applied physics, he understands the technology behind datacentres but decided to take it a step further and focus on how to create business value with a datacentre that is also beneficial for the environment and society.

Blockheating was ready to start scaling by increasing its capacity and finding more customers. However, Jeroen was wondering if they truly had the best possible technological solution in place, or if they should change and improve the cooling system in their datacentres before growing bigger.

About the Industry

Technology

A datacentre's purpose is to provide external computational power, storage facilities and network facilities. In order to do so, it consists of many servers that need to operate for 24 hours per day. These servers are pieces of computer hardware with all vital components like the processor and RAM, mounted on a motherboard (**Exhibit 1**).

All these components require energy to function and by doing so, they produce heat. Generally, the more power a component uses, the more heat is produced. The working mechanism is the same as with personal computers. When streaming a heavy file or when running multiple things at once, a laptop usually starts to heat up. As a result, one can feel hot air emerging from the sides. This is produced by the cooling system of the laptop, so that it does not overheat.

This teaching case was written by Adrienn Tóth under the supervision of Dr. Sandra Langeveld of the Rotterdam School of Management (RSM), Erasmus University. The authors would like to thank Jeroen Burks, founder and CEO of Blockheating for his time and input, as well as Tao Yue, Managing Editor at the RSM's Case Development Centre for her feedback.

This case was created using field research data and is intended for use in class discussion rather than to illustrate either effective or ineffective handling of a management situation. Copyright © 2024 RSM Case Development Centre, Erasmus University. No part of this publication may be copied, stored, transmitted, reproduced or distributed in any form or medium whatsoever without the permission of the copyright owner. Please address all correspondence to cdc@rsm.nl.

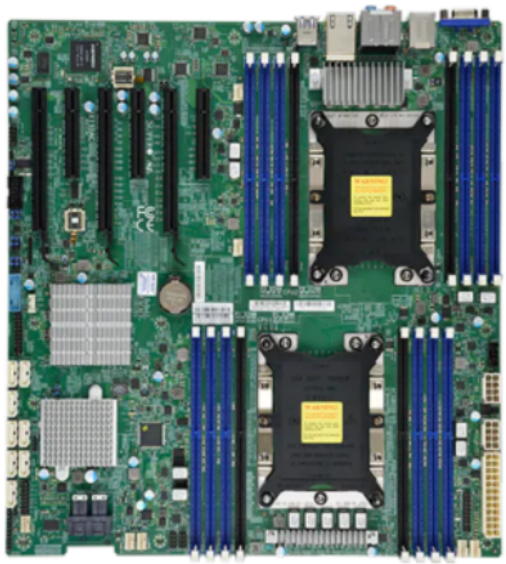


Exhibit 1: Figure of a motherboard

Exhibit 2: Figure of a server

Since the components are designed to operate within a certain temperature region, a cooling system is required to prevent system failure. Combined with the required energy for the servers themselves and all other electrical components present in the datacentre, datacentres are responsible for 1.5% of the total global energy consumption. Looking at the still increasing demand, it is crucial to look for energy efficient cooling mechanisms.

There are three main cooling mechanisms used for the cooling system of datacentres. They vary in terms of mechanical complexity, energy need and efficiency, so each company chooses a method based on their custom needs. The three methods are the following:

Air cooling

The air cooling technique is one of the most basic and common cooling techniques. The cool air used for cooling the servers can come in two ways: directly or indirectly.

For direct free air cooling the outside air is directly distributed to the datacentre. Before entering the datacentre, the air is filtered to obtain the necessary air quality and mixed with warm air to obtain the right temperature. The humidity of the air is also controlled. For indirect free air cooling the outside air never reaches the datacentre. This method is based on an air/water heat exchanger, which does exactly what the name indicates; it exchanges heat. The setup uses a chiller, which can be seen as a refrigerating system that uses some sort of heat medium. In this case the heat medium is water. The heat exchanger is placed on the hot water return of the chiller. The outside air is blown past this heat exchanger and so the heat is removed before it reaches the receiving side of the chiller. The chiller cools the air that is already present in the datacentre.

The main advantage of this technology is the ease of operation, as it does not require a complicated setup since the air can flow freely through the system. That is why this solution is widely used, so maintaining it through common risks and failures is a skill that many possess in the industry. However, as hardware technology develops, server racks are becoming more high performing which results in the usage of more and more power. Keeping these new systems on the required temperature can sometimes prove to be too difficult for a traditional air cooling system. However, improving the system – for example by adding more and stronger fans – can result in significantly increased noise pollution and power consumption that can be harmful for workers and require them to wear protective equipment.

Direct-to-chip cooling

The second technology is direct-to-chip cooling. Rather than air, the cooling is based on liquid cooling, in particular water cooling in this case.

As the name suggests, the cooling is done directly on certain components. While with air cooling, the air flows freely through the entire server, in the case of direct-to-chip cooling, liquid is guided directly to the components that need cooling. To implement this technique, a so-called cooling plate is mounted directly above the server part, to which the server delivers all its heat. A cooling fluid flows over this plate, directly transporting the heat away from the server. Important to note that there is no direct contact between the liquid and the computer elements. However, to optimize the cooling properties of the water the distance between the cold plate and processor is as small as possible.

A main advantage of this technology is that heat is directly available for processing in this case as the cooling system is connected to the specific components that need cooling. Moreover, as this solution uses water instead of air for cooling, two important properties change:

1) The specific heat c_h of water at a temperature of 60 $^\circ$C at atmospheric pressure is 4.1875 kJ kg^{-1} K^{-1}. This is almost 4 times as big as the specific heat of air. This means that given the same mass flow and temperature increase, we can carry away 4 times as much heat.

2) As we know, water is a lot heavier compared to air. The density ρ is a thousand times higher. Therefore, we can transport almost 4 thousand times more heat with the same amount of volume.

However, using water brings about some challenges. To guide the water past the server components, small pipes are used. One of the main challenges in this type of liquid cooling is for that reason, corrosion. Corrosion makes the possibility of leakage greater and so the structure needs more maintenance. Also, by guiding water to every CPU in every server, a lot of components are added to the process. This operational complexity increases the likelihood of failure.

Immersion (liquid) cooling

With immersion cooling the whole server rack is placed in a tank filled with a dielectric liquid, much like someone immersing in a cold bath after the sauna. Instead of the cold water in the bath, dielectric fluid is used as a coolant. A dielectric fluid is non-conductive, which means that it has a very high resistance to electric breakdown. However, the fluid is heat conductive, meaning it can transfer heat. The dielectric fluids that are suitable for this technology are synthetic or mineral oils. As it is harmful to expose the components directly to the fluid in the tank, the server components are sealed in a readily accessible way.

Because the components are submerged in the dielectric fluid, the heat released by server components is directly transmitted to the coolant. The tank in which the server rack is placed is connected to a pump. This way the coolant can be pumped out of the tank towards a heat exchanger. There the fluid transfers the absorbed heat to the cool water in the heat exchanger. The cooled fluid is then pumped back into the tank again.

Immersion cooling offers several advantages. The cooling setup is relatively simple, and the components don't take up much space. Therefore, the infrastructure needed to implement this technology is relatively simple. Another advantage lies within the maintenance of the coolant. With good sealing of the components, coolant loss is low and so there is no demand for frequent coolant replenishment. However, the sealing is also where the challenge of this technology lies. Since the complete server is emerged in the liquid, a leak in the sealing can have disastrous consequences. Moreover, like direct-to-chip cooling, the liquid is pumped through pipes. The smaller the pipes used; the faster corrosion occurs.

Cooling technology providers

Datacentre providers on the market decide to use different versions of the mentioned technologies to provide their services, or even a mix of technologies. The competitors of Blockheating also represent users of varying cooling technologies described above, and some of them even have green solutions to reduce waste, similar to that of Blockheating.

Company	HQ	Description
Asetek	Denmark	Similarly to Blockheating, Asetek also focuses on direct-to-chip water cooling to take advantage of its many benefits, like effectiveness and quietness. They even took the technology a step further and created their own solution that they call the All-In-One liquid cooler. This solution made setting up a liquid cooling system easier and cheaper and operating the system more reliable. The company has operations in China, Taiwan and the United States and in terms of customers, they are focused on the gaming industry.
ZutaCore	San Jose, CA, USA	ZutaCore is using a technology they call HyperCool which is a two-phase boiling and condensation process. It uses the direct-to-chip cooling method, however, there is no water in the system, so equipment is protected from corrosion and other water-related threats. The company offers its cooling system to cloud datacentres, edge datacentres and hotspots alike. Moreover, they are reusing the heat generated by the system for the heating of office buildings, schools, or homes.
Asperitas	Amsterdam, Netherlands	Asperitas is another company providing the cooling technology itself to datacentres and other hardware owners. They use immersion cooling for their services which they provide to enterprise-level customers.
Submer	Barcelona, Spain	Submer is using a single-phase immersion cooling solution for which they have created their own dielectric fluid, the SmartCoolant. In terms of customers, they are focused on the Cloud, Edge, HPC and AI industries.
CoolIT	Calgary, Canada and Connecticut, USA	CoolIT Systems provides immersion cooling solutions for datacentres. Their solution is called Direct Liquid Cooling (DLC) that uses warm liquid rather than cold air to dissipate heat from computer and server components. By capturing component heat in a liquid path, DLC allows for higher component performance and reliability, higher densities and decreased datacentre operating expenses.
Qarnot	Montrouge, France	Qarnot offers low-carbon and sovereign high-performance computing power. Their strategy is essentially based on the valorisation of computer waste heat and a distributed approach: they

		mainly deploy their servers in existing buildings capable of utilising the heat released by the computers. Their hardware solutions also enable their customers to reduce their own carbon footprint within their IT infrastructures. Industry-wise, they are mainly focused on HPC (High-Performance-Computing) industry, especially the finance and 3D animation sectors.
LeafCloud	Amsterdam, Netherlands	LeafCloud is another company that puts the reduction of energy waste to its focus. They provide cloud server capacity to their customers, during which they use their energy twice: first, to power their servers, and second, to substitute the use of fossil fuel for heating. They do this by using air cooling technology but recycle the wate heat with a heat pump. LeafCloud servers are installed at locations such as apartment buildings, hotels, nursing homes, swimming pools, and other structures that use hot water year-round. At these sites, their servers deliver heat directly into the local water and heating system. This replaces heating from sources such as natural gas, thereby drastically reducing CO2 emissions. Because they use existing structures, they also make significant savings by not building large structures - and the required infrastructure - in the middle of nowhere.

Blockheating

History

The story of Blockheating dates back to five years before, when Jeroen, the founder of the company had an idea for how to reuse heat generated by datacentres as a result of a coincidence. By that time, Jeroen had already been interested in ways to use the latest technologies to create value for society and the environment. That is why after his studies in applied physics, he decided to pursue a business degree as well to understand more about how to manage these technologies.

One day early in his career, Jeroen took home some data servers that would have otherwise been thrown out to see how he could reuse them. The idea of recycling the heat generated by the cooling system already existed by then, but Jeroen did not see the business case for it yet. Together with a friend, they have deployed a small datacentre where the generated heat was used for the heating of a building. That was the moment where Jeroen saw that the idea could indeed work in practice, so he decided to try it on a larger scale, thus Blockheating was born. Now, instead of apartments, they provide the recycled heat from their datacentres to greenhouses.

As of 2023, the company was still operated by a small team. Apart from Jeroen, the CEO, there were four other roles in the organisation. The technology was handled from three different aspects by three people: first, the Chief Product Officer oversaw the hardware technology; then, the Chief Technology Officer was responsible for the IT infrastructure within the datacentre to make sure the system as a whole operates properly; finally, the IT network engineer's job was to take care of external connections, for example to connect the datacentre to the Internet. Additionally, as the team was expected to grow further that would also bring about more administrative tasks, an office manager was also part of the team to facilitate these processes.

Technology

The system by Blockheating utilises direct-to-chip cooling, using a combination of water and air to cool. In total 65% of the system is cooled by water and 35% by air. The water enters the device at 55 °C and exits at 62 °C. Air enters at 35 °C and exits at 45 °C. This way, the required mass flow is substantially lower when compared to air cooling, resulting in less stress on the systems.

Because Blockheating is reusing the heat generated instead of just releasing it into the air, it is crucial for them to have a high exit temperature of water which they can use in the greenhouse afterwards. On the other hand, in the case of air cooling, a lot of air needs to be pushed through the datacentre and then a heat pump which heats up the water, creating lower exit temperatures. Similarly, a heat pump would be needed in the case of immersion cooling as well to produce hot water that can be used by the greenhouses. The ideal technology for Blockheating has been direct-to-chip as it was the most feasible to implement, however they need to keep in mind the considerations for customer value creation and reduction of environmental impact as well.

With the described setup, Blockheating's datacentre had to be directly connected to the greenhouse it served. Therefore, one datacentre could only provide heat to 1-2 Ha of greenhouse. Because of this, if the company wanted to serve multiple datacentres in the future, they were going to need to set up datacentres on various locations, which is not the usual industry practice. A regular datacentre with no efforts to reduce energy waste would create larger server parks where they can take advantage of economies of scale as much as possible, so the recycling of heat added an extra level of complexity for Blockheating.

Business model

There are two types of customers that Blockheating serves: the first is the greenhouse that pays for and uses the heat, and the second are the companies who rent computational power from the datacentre. Greenhouses are important stakeholders because in order to actually use the waste heat there needs to be a party that is in need of heating almost nonstop. Tomato, bell pepper and cucumber greenhouses have a heat demand of about 3 kWh/week/m2 in the summer, so those are the most suitable to make use of the generated heat. While both the greenhouse and the customers pay for the service provided, the revenue coming from the greenhouse is marginal in the current application, so the business model is dependent on the second customer group.

As of 2023, Blockheating had one operating datacentre with preparations in progress to set up the second one. Their customers can be located anywhere, there is no need for physical proximity in order to provide the service. At the time, Blockheating served 27 customers from this location, each renting between 15 (1/3rd of a server rack) and 400 servers.

In terms of pricing, the company used a fixed monthly pricing for renting the servers. This meant that a customer could decide when signing the contract the amount of server capacity they would like to rent and pay the fee based on that, regardless of the amount of power they actually ended up using. Meanwhile, a common practice in the industry was to charge after the energy consumed by a client, which is something Blockheating was also considering in the long term. The plan was to switch to a model where the customers pay for the power consumed separately from the monthly rent price. With the prices charged then, renting one server for a month cost between 110 – 170 EUR depending on how much memory there was in the server, and the minimum number of servers Blockheating rented out to a certain customer was 15.

In terms of service offering, there are two common practices in the industry: colocation and cloud. In colocation, a customer brings their own servers, and the datacentre only provides space, cooling and electricity. In the case of cloud, the supplier provides the dacacentre, the servers and a software layer to rent out small pieces of a server. Blockheating's model is between these, as they do provide the datacentre and the servers, but leave the software layer up to customers.

The most important suppliers of Blockheating can be put into three categories:

- Server supplier: Blockheating needs servers to provide their service to the IT customers. They prefer to cooperate with companies that offer refurbished servers in order to lower emissions and operate even more sustainably.
- Cooling production providers: the company designs the cooling itself, but for production they rely on external providers.
- Connectivity providers: in order to use the server, a customer should be able to connect to it. There are 2 options to do this, a direct connection between Blockheating's server and a customer server or over the Internet to a user.

Strategic directions

On the short term, Blockheating was focused on opening its further datacentre locations. Their goal was to open three datacentres close to each other (in a 2-kilometer radius), so the connection between them would be easy, as well as heat and fibre connections should be easily available. The company was ready to scale and bring their offering to the market as their first datacentre proved that the system was stable and able to reliably serve its customers.

Regarding the technology, Blockheating was satisfied with their solution in place, however it has not been tried out yet on a bigger scale. Before they grew significantly bigger, the management team needed to make sure that the technology also served their business needs well in terms of value creation for customers and the greenhouses as well as cost and operational efficiency. Therefore, they were considering a potential switch to a different technology in the case if it would allow them to perform significantly better in at least one of these dimensions without being worse in any other one.

Next Steps

Jeroen made another coffee and sat back down to his laptop to continue working. He was confident Blockheating had the potential to grow big while also creating positive environmental impact. However, he wanted to be sure they had the perfect technology in place before implementing their growth strategy.

Which of the available technologies would be the most cost-efficient for Blockheating to apply? And which one would allow them to provide the highest quality service from a customer perspective? Would this solution also be capable to save enough heat to use in the greenhouses? If they were to change the technology, what would be the implications for their business model? These were the questions nonstop circulating in Jeroen's head.

Appendix 1: Physics of cooling

When looking at the different cooling mechanisms for the various server parts, they boil down to the same idea; the mechanism is comparable to cooling someone's dinner by blowing at it.

On the one hand, we have a hot object that emits heat; the food or the processor. On the other hand, we have some sort of cooling fluid that flows along this object; the air. By doing so, it absorbs some of the heat coming from the hot object and, as a result, the temperature of the fluid increases.

This is also described by the second law of thermodynamics. Simply said, this law states that heat transfer occurs from higher to lower temperature bodies. As we all intuitively know, placing ice in the Sahara makes the ice melt. The final goal is to keep the processor at a certain operating temperature, therefore, all excess heat must be carried away by the cooling fluid.

The different cooling technologies use variations on the same mechanism. In order to compare these variations, we need to be able to quantify the situation using some variables. Note that in physics, it is common to place a unit within brackets next to the described quantity to simplify the description.

Let's state that the processor emits a certain amount of heat per second Q in [kJ s^{-1}] or [kW]. This heat is absorbed by the cooling fluid which flows along the processor. At the beginning, the fluid has a temperature T_{in} [K] and after the processor the temperature has increased to T_{out} [K]. The fluid has absorbed the heat from the processor.

Now there are two important quantities that determine the amount of heat that is absorbed by our fluid. 1) The flow rate $\dot{m}$. The flow rate [kg s^{-1}] indicates the amount of fluid that travels along the processor per second. You can imagine that if you blow very powerfully along your food, it cools more.

The flow rate can be calculated by the following formula.

$$\dot{m} = \rho v A,$$

where ρ is the density of the fluid in kg m^{-3}, v the velocity in m s^{-1} and A the cross section of the part through which it flows; the cross section of a pipe for example.

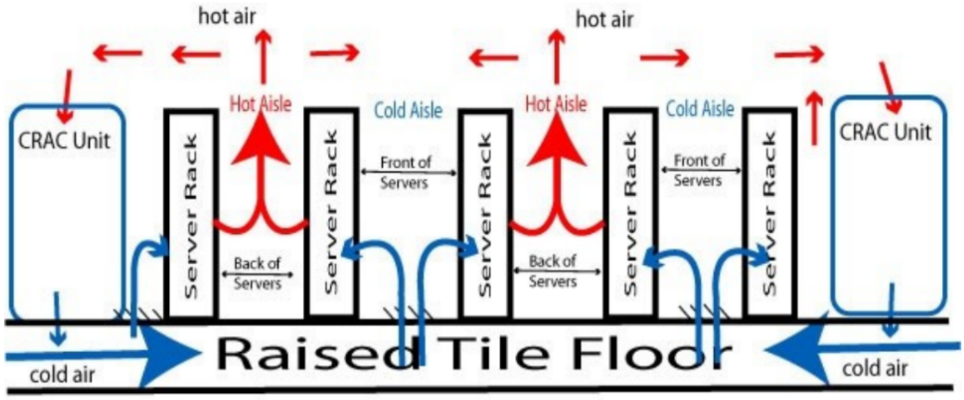
2) The specific heat capacity c_v [kJ kg^{-1} K^{-1}]. This quantity indicates how much heat a kilogram of a certain liquid must absorb in order to raise its temperature by one degree.

This entire story can be summarised using one equation.

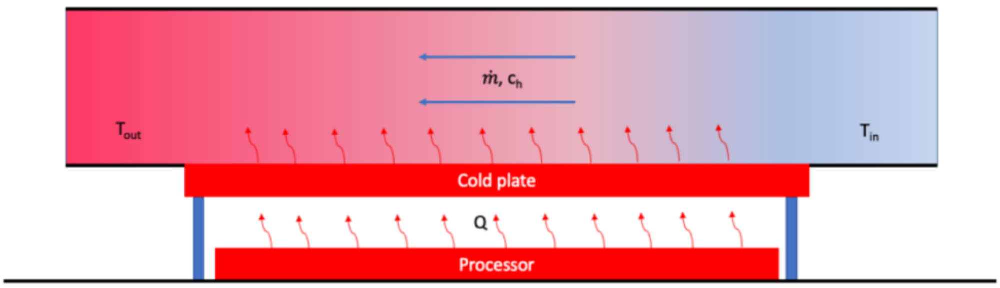
$$Q = c_v \dot{m} (T_{out} - T_{in})$$

This equation states that the amount of heat that is absorbed or released by the processor is equal to the specific heat, times the amount of fluid that flows along it times the total increase in temperature.

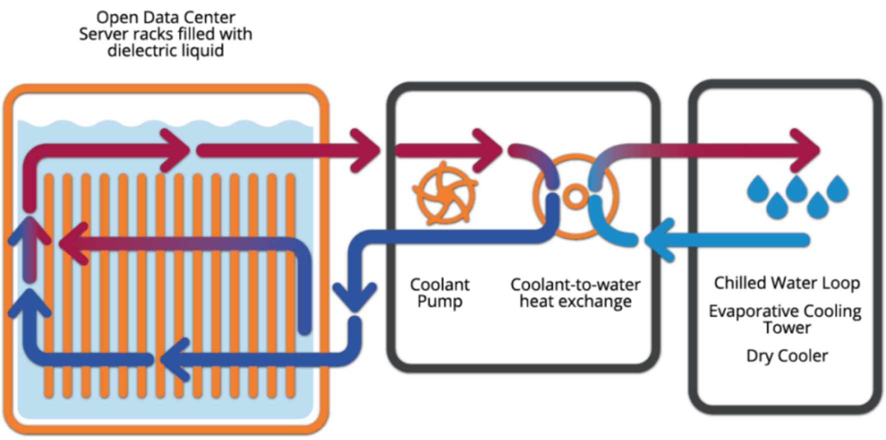
Appendix 2: Schematic description of heat being transferred in an air cooling system using a hot isle/ cold isle layout and a raised tile floor



Appendix 3: A schematic representation of direct-to-chip cooling for a server processor



Appendix 4: Schematic description of immersion technology. Here the server is placed in the tank. The dielectric fluid is cooled with use of a heat exchanger



Appendix 5: Useful terminology used in the datacentre industry

- PUE (Power Use Effectiveness): Total facility power / total IT power. Note: this does not take the power consumption of cooling systems (e.g. fans) in the server into account, as that is counted as "IT power"
- ERF (Energy Reuse Fraction): The proportion of the heat that is reused
- WUE (Water Use Efficiency): In indirect or direct air cooling, most datacentres use chillers or heat pumps to get rid of the waste heat. As a "green" alternative, some companies evaporate (drinking) water for this cooling process. To see this in action, take a room temperature drink, wrap it in a soaked towel and put it in the shade. As the water evaporates, your drink will cool.