DataCenterBoom!

How is water used in data centers?

A brief guide

November 2025.

I. Introduction

This document complements the information on the Data Center Boom! Website with additional technical details. The purpose is to explain how water is used, the most common cooling methods for data centers, their disadvantages, and the challenges they create. The aim is to demonstrate that no single technological miracle will eliminate water use. Instead, a series of technical and political decisions—each with associated costs—must be considered to achieve optimal water efficiency.

II. Overview

Unlike agriculture, whose water footprint is mainly green (i.e., water stored in the soil and used by plants), most of the water footprint of data centers is blue water extracted from rivers, lakes, or groundwater, which is directly accessible for human use but often more limited in availability. However, water can come from various sources, including blue sources, piped sources such as municipal water, and gray water sources (e.g., recovered purified water). The use of recycled or non-potable water to cool data centers is a well-established practice for conserving limited potable water resources, especially in dry or drought-prone areas.

Water consumption in data centers should be understood by differentiating between *water withdrawal* and *water consumption*. Withdrawal corresponds to the amount of water extracted from blue or gray sources minus the water discharged by the centers (mainly hot water left over from cooling IT racks) that is returned to the environment. At the same time, consumption is the fraction that evaporates or does not return to the local ecosystem. In data centers, a significant proportion—between 60% and 80% of the water withdrawn—is consumed through cooling processes, especially in cooling towers that release steam into the atmosphere (Pengfei Li et al, 2025). The remaining water is discharged to municipal wastewater facilities. The large volume of wastewater from data centers can overwhelm existing local facilities, which were not designed to handle such high volumes.

Researchers at The Green Grid, a non-profit industry consortium, developed a metric called Water Use Efficiency (WUE) to measure the water consumption of data centers. Similar to the Power Usage Effectiveness (PUE) metric, which measures the energy efficiency of a data center, the WUE metric assesses its water usage efficiency.

WUE is expressed in liters per kilowatt-hour (kWh): the total water consumption of a data center, measured in liters, is divided by the total energy consumed by that data center in kilowatt-hours during the same period of time.

Although "0" is the ideal WUE score, it can only be achieved in air-cooled data centers, and most data centers cannot meet this target due to the climatic conditions of their location. The average WUE for data centers is 1.9 liters per kWh, which is a significant target to exceed.

Data center water use varies based on factors like location, climate, water supply, size, and chip density of IT racks. In warmer regions, such as the southwestern United States, data centers require more water to cool the facilities and equipment.

Data center water consumption covers three areas:

- Scope 1: *On-site* water uses for data center cooling. This is because the components of IT racks within a data center generate a lot of heat when operating constantly. Therefore, it is imperative to regulate the data center's temperature and humidity to keep the computer equipment in optimal condition.

- Scope 2: water outside the facilities needed for electricity generation. Powering data centers with renewable energy sources, such as solar or wind, requires significantly less water consumption than obtaining energy from fossil fuel power plants.

- Scope 3: water used outside the facilities that is consumed in the supply chain for manufacturing servers used in data centers.

Summary of data center water consumption		
Based on information from Pengfei Li, Jianyi Yang, Mohammad A. Islam, and Shaolei Ren. 2025. Making AI Less 'Thirsty'. Commun. ACM 68, 7 (July 2025), 54–61. https://doi.org/10.1145/3724499; and Yañez-Barnuevo, Miguel. 2025. Data Centers and Water Consumption. Environmental and Energy Study Institute. https://www.eesi.org/		
Scope 1: On-site water for cooling	**Scope 2: Water from electricity generation**	**Scope 3: Supply chain water**
This process involves two sequential stages: server-level cooling followed by facility-level cooling. **Server level:** Heat is transferred from the servers to the facility or to a heat exchanger, typically using air or liquid cooling methods (e.g., direct chip cooling or immersion cooling, which do not evaporate or consume as much water). With the increase in the number of centers supporting AI applications, chip density is also growing, causing an increase in ambient temperature. **Facility level**: Heat is expelled from the data center facility into the outside environment. Although there are several cooling methods, water-intensive cooling towers and water evaporation-assisted air cooling are two common approaches used in many data centers.	In many countries, thermoelectric power is among the top sectors in terms of water withdrawal and consumption. Therefore, as with Scope 2 carbon emissions, data centers are responsible for Scope 2 water use outside their facilities associated with electricity consumption. Different power plants use different amounts of water per kWh generated, depending on cooling techniques. Typically, water extraction due to hydroelectric power generation is excluded, but water consumption due to increased water evaporation rates from hydroelectric power generation is included. For electricity generation, the national average water withdrawal and consumption in the United States is estimated at approximately 43.8 L/kWh and 3.1 L/kWh, respectively. Meta's reported Scope 2 water consumption for its global data center fleet was 3.7 L/kWh (i.e., 55,475 megaliters divided by 14,975,435 MWh) in 2023.	The manufacture of AI chips and servers consumes a large amount of water. For example, ultrapure water is needed for wafer manufacturing, and water is also needed to keep semiconductor plants cool. It is important to note that, at this stage, the water discharged may contain toxic chemicals and/or hazardous waste. Although water recycling in semiconductor plants can effectively reduce water extraction, the recycling rate remains low in many cases. Although largely unknown, the use of Scope 3 water is likely to be significant. For example, Apple reports that its supply chain accounts for 99% of its total water footprint.
More details at www.datacenterboom.net		

III. Scope 1 Specifications

In a data center, cooling operates on two interconnected hierarchical levels: the facility level and the IT or server level.

The facility level houses all the building's thermal infrastructure—such as cooling towers, *chillers*, heat exchangers, and *free cooling* or adiabatic systems. Their function is to produce and distribute cold air or water, dissipating accumulated heat to the outside. This level acts as the data center's "cooling plant" and keeps the overall ambient temperature stable.

In contrast, the server level is located within the IT area and is responsible for extracting heat directly from electronic equipment and transferring it to the installation-level systems. This involves air cooling methods (internal fans, cold and hot aisles, rack cooling) and the most advanced liquid or immersion cooling technologies.

Both levels are closely linked: the heat generated by the chips flows to the rack, then to the internal thermal distribution system, and is finally evacuated to the environment through the building's cooling infrastructure.

1. Facility level

At the facility level, a data center's cooling systems are responsible for collecting heat from the entire server room and releasing it outside. Large industrial equipment used includes *chillers,* cooling towers, heat exchangers, and systems that utilize outside air or water evaporation. The two primary methods are air cooling and water cooling, although they are almost always combined in hybrid schemes.

a) Air cooling:

- **Air cooling (dry) systems:**

In air (dry) cooling, heat from the interior is expelled into the environment using large condensers that use fans and dry coils. This system is completely "dry": it does not use water, only the heat exchange between the ambient air and the refrigerant gas or fluid. It is simpler and does not depend on water resources, but it loses efficiency when outside temperatures are high, as hot air cannot absorb as much heat. This is why it is associated with higher electricity consumption, which can exacerbate the overall pressure on water resources due to higher water consumption in scope 2.

- **Free cooling or free cooling by outside air:**

This system directly exploits the climatic conditions of the environment to reduce the heat generated inside a data center without the need for compressors or mechanical refrigerants. The principle is simple: when the temperature and humidity of the outside air are low enough, that air can be used to cool the flow of air or hot water coming from the servers. In direct free cooling systems, outside air is filtered and introduced into the server room, replacing the hot air expelled. In indirect free cooling systems, ambient air cools a closed circuit (for example, using an air-to-air or air-to-water exchanger) without entering the interior of the data center.

The great advantage of this method is its energy efficiency, as it reduces or eliminates the use of compressors and refrigerants, potentially lowering the facility's total electricity consumption by up to 70% in cold or temperate climates.

Furthermore, as it does not require water, *free cooling* is considered more environmentally sustainable than evaporative or adiabatic systems. However, its effectiveness depends entirely on the climate: it can only work when the outside air is below certain temperature values (generally below 27°C) and relative humidity values (below 60%). In hot or humid regions, its use becomes limited or requires mechanical cooling support.

In general, outdoor air cooling is more water-efficient than cooling towers. However, hot weather increases the demand for water evaporation and peak water consumption, which can place a strain on local water supplies during peak demand on hot days.

- **Evaporative-assisted to adiabatic systems**

There are also evaporative-assisted or adiabatic systems that add an intermediate step: the air that cools the condensers or exchangers first passes through a water mist or wet medium. As it evaporates, the water reduces the air temperature, improving cooling capacity and reducing the electrical energy required. This type of system, known as adiabatic or assisted evaporative cooling, has become very popular because it can save between 20% and 40% of energy compared to totally dry systems.

b) Water cooling systems

- **Cooling tower:**

Some of the water evaporates (i.e., is "consumed") in the cooling tower to dissipate heat into the environment, while the remaining water travels through an open circuit to the heat exchanger to further absorb heat from the server.

Incidentally, non-evaporated water can only be recycled a few times (typically between 3 and 10 cycles, depending on water quality) before discharge, requiring continuous replenishment of clean fresh water to prevent mineral and salt buildup. Therefore, for the cooling tower to continue operating, new water must be constantly added to compensate for evaporated water and discharged water. Clean fresh water (potable water in many cases) is needed here to prevent clogging of pipes and/or bacterial growth.

In the case of cooling towers, water withdrawal refers to the total amount of water added, including both evaporated water and discharged water, whereas water consumption refers exclusively to the water that is evaporated.

With good water quality, approximately 80% of water withdrawn evaporates and is considered “consumption.” On average, depending on climate conditions and operational settings, data centers can evaporate approximately 1 to 9 l/kWh of server energy: 1 l/kWh for Google’s annualized global water efficiency at its facilities and 9 l/kWh for a large commercial data center during the summer in Arizona.

- **Chillers**

In this scheme, *chillers* cool the water circulating to data center equipment (such as AHUs or chilled back doors) and transfer heat to the water, which is then dissipated in the cooling tower. This type of system combines power, thermal stability, and efficiency, although it requires costly infrastructure and indirect water use due to the operation of the tower. In contrast, air-cooled chillers eliminate the use of a tower: they dissipate heat directly into the environment using coils and fans, maintaining an internal water loop. They do not consume water, but their efficiency decreases in hot climates, and their electrical demand increases.

- **Closed-loop adiabatic systems**

These systems function as closed water cooling systems but include evaporative modules to boost efficiency when outdoor temperatures are high. Under normal conditions, they operate “dry” and only use water during thermal peaks, which lowers their water usage. However, they still require maintenance of the wet system and precise controls to prevent leaks or biological contamination.

- **District cooling**

Data centers are connected to an urban network that supplies them with cold water from a central plant. This model, used in cities such as Singapore and Dubai, optimizes collective energy efficiency and can integrate waste heat recovery for heating or industrial uses. Finally, some data centers implement thermal recovery and reuse systems, which harness the heat from the hot water generated by the servers for other purposes, improving the overall energy balance and reducing associated emissions.

2. Server level

At the server level, water chillers cool IT rooms to maintain optimal temperatures and prevent damage to chips. With the advent of artificial intelligence, the central issue in this debate is, in fact, GPU cooling methods. Two methods can achieve this:

- **Air cooling using water evaporation:**

Data centers have traditionally used air to cool servers, which is an open-loop method that consumes more water. They use a hot and cold aisle configuration (Awati & Kirvan, 2025), with servers located in the cold aisle, where they are surrounded by cold air. The heated air is then vented into a hot aisle. By extracting heat from the cold aisles, the cold is maintained, allowing air conditioning to meet the demand for traditional workloads.

This system is the most widespread due to its simplicity, low cost, and standardization. Still, it has limitations: as the thermal density of the racks increases (for example, in artificial intelligence or HPC applications), the air loses its ability to transport heat efficiently, creating areas of overheating or "hot spots."

- **Liquid server cooling:**

Server cooling is a more expensive approach that supplies liquid coolant directly to graphics processing units (GPUs) and central processing units (CPUs). Direct-to-chip liquid cooling and immersive liquid cooling are two standard server liquid cooling technologies that dissipate heat and significantly reduce water consumption.

During immersion cooling, water or specialized synthetic liquids flood the chips, absorbing heat. In areas with limited water availability, liquid cooling of servers is the best option, as it requires minimal water consumption. This technique achieves very high thermal efficiency, virtually eliminates the need for air conditioning, and allows waste heat to be recovered for other uses. Its main limitations are high initial costs, lack of industry standardization, and complexity in maintenance, since equipment must be handled within the liquid. Conversely, in regions with an overloaded power grid, an evaporative cooling tower is a suitable building design, as it requires minimal energy consumption.

However, regardless of the chosen approach, a heat exchanger is necessary to capture the hot air or hot water produced as a by-product of the cooling process. The hot water from the servers is cooled using water from the air-cooled chiller or a cooling tower. Similarly, the hot air is replaced with cooler air. A heat exchanger transfers heat from the server room to the building's cooling system.

3. There are no technical miracles; it's all a balancing act

We constantly hear that "closed-loop liquid cooling does not consume water and therefore water consumption should no longer be a concern for affected communities." That argument needs careful review, especially when discussing AI workloads.

For AI researcher Masheika Allgood (2025), while closed-loop systems significantly reduce water consumption compared to open-loop systems, they still require trace amounts of water for replenishment and evaporation. Therefore, there is no such thing as zero water consumption.

The researcher explains that AI runs on GPUs (Graphics Processing Units), which cause these electronic circuits to get very hot. For example, the chip that powers Grok3 typically operates at around 87°C but can reach a maximum operating temperature of 98°C. The chips are designed to slow down their processing when they reach 95°C and shut off at 98°C. Running at these high temperatures for extended periods can cause wear and reduce their lifespan.

Therefore, for the chips to operate as efficiently as possible and maximize the investment, data center operators must implement systems to cool them. AI workloads run on chips that require fairly extreme cooling. And they don't run on just a few chips. Grok3 runs on a cluster of 200,000 chips. Each of them needs to be cooled from 95°C/203°F to 87°C/188°F. Air conditioning simply cannot keep up. That's why data centers are switching to liquid cooling to support AI workloads.

The most common liquid cooling system is the closed-loop system. In a closed-loop system, cold liquid circulates through the server, which heats it. The heated water then circulates through a heat exchanger, where it is cooled, and finally, the cold water circulates back through the server. That loop is closed. No liquid is gained, no liquid is lost.

But the trick that is sometimes overlooked is that this is not the only circuit. The key component of the closed-loop system is a second external circuit: the open heat exchanger circuit. The external circuit circulates cold water through the heat exchanger and hot water to the water cooling tower. This circuit is responsible for data centers consuming 60% of the water they extract.

There are two main technologies for cooling liquids in a closed-loop system: chillers and water cooling towers. Water cooling towers are preferred for industrial-scale cooling because of their energy efficiency. Although both systems can use air for cooling, air cooling is only effective in very cold environments. In most cases, both systems utilize evaporative cooling.

Evaporative cooling systems pass heated water through a cold water spray or over a medium soaked in cold water. The heat evaporates some of the water into the air, which carries the heat away and cools the remaining water.

Water towers require a constant flow of water to operate, so they are constantly evaporating. Water-cooled chillers require less water to operate but are often used in conjunction with water towers in large-scale operations.

This is why data centers are such large consumers of water, says Masheika Allgood (2025): because only one of the circuits is closed. The other is in a constant battle to cool large quantities of super-hot GPUs, evaporating millions of liters of water per day to meet demand.

Furthermore, as Shaolei Ren and Amy Luers (2025) mention, alternatives to evaporative cooling, such as systems that recycle cooling water in a closed loop, can lower the need to use local drinking water supplies. However, many of these systems increase electricity demands, which can, in turn, raise indirect water use.

For these latter authors, the reality of intertwined water and electricity systems forces data center operators to make difficult trade-offs between global climate goals and local water needs. For example, in water-scarce regions, the priority should be to use cooling systems that consume little or zero-water-consumption cooling systems to reduce direct water use, while investing in adding renewable energy to local grids to curb indirect water use and minimize carbon emissions from increased electricity demand. In wetter regions with carbon-intensive electricity grids, priority should be on reducing energy consumption to lower overall water use, even if that means continuing to use evaporative cooling, which has higher on-site water consumption.

4. Efficient water use by data centers

Based on Shaolei Ren and Amy Luers (2025), Leila Karimi et al. (2025), and Nuoa Lei et al. (2025), to make a more sustainable decision for nearby communities and ecosystems, from the perspective of the developers and designers of these infrastructures, the following must be taken into account:

- Water must be included as a resource as critical as energy in design and operation: decisions on location, thermal design, and choice of cooling technologies must also be evaluated under water sustainability criteria, not just energy efficiency.
- In regions with water stress,
    - The priority should be to use cooling systems with low or zero water consumption to reduce direct use, even if this increases electricity use.
    - It is essential to apply hourly models (not annual averages) to plan infrastructure, as impacts vary throughout the year.
    - Policy and design makers should consider the cross-effects between energy, water, and local climate to minimize environmental stresses.
    - Renewable energy should be added to local grids to curb indirect water use and minimize carbon emissions from increased electricity demand.
- In wetter regions with carbon-intensive grids, reducing energy consumption should be prioritized to decrease total water consumption, even if it means continuing to use evaporative cooling, which has higher on-site water consumption.
- Many water efficiency strategies adopted in practice may have limited benefits if they are not adapted to the actual workload profile within data centers, as some extreme thermal loads or operational peaks require the use of more intensive cooling methods. Data center operators should therefore adopt more refined metrics, in addition to WUE, that disaggregate water consumption by load type, density level, time of day, etc.

Concerning Scope 2 water consumption, i.e., from energy generation, Shaolei Ren and Amy Luers (2025) recommend changing the composition of the energy matrix that powers data centers:

- Favor energy sources with low water use, such as solar photovoltaic and wind power, which require virtually no water to generate electricity (unlike thermal power plants).
- Develop "firm" technologies with low water demand, such as sources that can generate electricity constantly (like nuclear or geothermal), but with technologies that

use less water or have more efficient recirculation systems.

- Encourage innovation and integrated energy policies so that investment decisions in AI and data centers also consider the source of electricity supply and its total water impact.

***