PUE: Power usage effectiveness

The measure of accountability in data center power use

Executive summary

How to save 63 million kilowatt hours of electricity per year

This white paper examines energy efficiency in data center design and shows how those design choices produced a data center that is 80 percent more energy efficient than traditionally-built data centers.

Data centers use a tremendous amount of energy; world-wide use is equivalent to the output of 30 large nuclear power plants, highlighting the need to use that energy wisely. This need for accountability in data center power use brought about power usage effectiveness as a measure of data center energy efficiency. Power usage effectiveness, generally referenced as PUE, is calculated by dividing the total energy used in a data center by the total energy used by the IT equipment (servers, storage, switches, etc.) within the data center.

Data centers exist to support the computing infrastructure within them, so other energy consumed in the data center to support operations such as for cooling, security and power distribution, is “overhead” that should be minimized. Power usage effectiveness is reported as a number and a lower number denotes a more energy efficient data center. Theoretically, the lowest PUE is 1.0, which indicates that all of the energy going into the data center is going into the servers. Of course, this is not possible in the real world, where energy is also needed for lights, cooling and other necessary functions.

$$\text{PUE} = \frac{\text{Total facility energy}}{\text{IT equipment enegy}}$$

PUE background

Power usage effectiveness was created by a number of data center industry organizations including The Green Grid, ASHRAE and the Uptime Institute. There are four categories of PUE measurement, moving from lower accuracy in PUE Category 0, to higher accuracy in PUE Category 3. The categories differ in what is measured (instantaneous power vs. cumulative energy) and where it is measured, as shown below.

	PUE category 0	PUE category 1	PUE category 2	PUE category 3
IT energy measurement location	UPS output	UPS output	PDU output	IT Equipment input
Definition of IT Energy	Peak IT electric demand	IT annual energy	IT annual energy	IT annual energy
Definition of total Energy	Peak total electric demand	Total annual energy	Total annual energy	Total annual energy

Flexential measures PUE in a way as close to Category 3 as is practical. Since Flexential does not control the IT equipment within a customer cage or rack, we measure IT equipment input of energy at each power circuit entering the cage or rack. This measurement method, which measures annual energy at both the IT equipment and for the total data center, results in a highly reliable PUE measurement that is virtually the same as a true Category 3 measurement. The difference between the measurements is statistically insignificant.

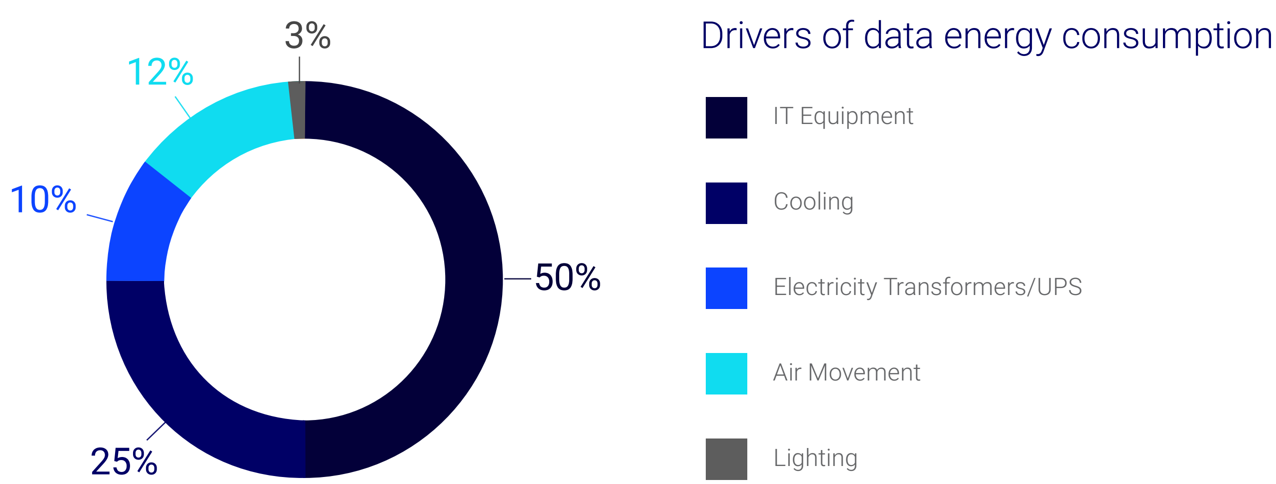
Power usage effectiveness varies widely in data centers. The Uptime Institute estimates that typical data centers have a PUE of 2.5, with small corporate data centers having PUE measures as poor as 3.0 or higher. Larger data centers average a PUE rating of 1.8 according to a survey of more than 500 data centers conducted by The Uptime Institute. PUE is the leading metric for tracking the energy efficiency of data centers. The average PUE is a subject of some debate in the industry. The Uptime Institute, which tracks the operations of a group of enterprise data centers, has previously said that the average PUE was around 2.5. Others cite an average of 2.0, and the EPA's Energy Star program reported a 1.91 average PUE in 2009 data it collected from more than 100 data centers.

Data center energy consumption

In a PUE measurement, by definition, the energy going into the IT equipment totals 1.0 of the PUE number. The remainder of the PUE number is energy powering "support" systems such as cooling, security, lighting, etc. The Environmental Protection Agency studied a number of data centers and found the breakdown of energy use in large data centers that have a PUE of approximately 2.0. Small amounts of energy are consumed by support functions such as lighting, security, monitoring and controls systems.

Electrical conditioning (uninterruptible power supplies, or UPSs), voltage transformation (transformers) and distribution account for 10 percent of the total energy use, or 20 percent of the non-IT equipment use. But the majority of non-IT equipment energy use goes to cooling, (labeled cooling and air movement in the pie chart that follows).

Variations in overall PUE are driven overwhelmingly by cooling systems' efficiency, and to a lesser extent, by losses in the electrical conditioning, transformation and distribution systems. The only other system that contributes in a measurable way to PUE is lighting.



PUE variability

Energy efficiency can be improved significantly across the following areas of the data center, improving PUE dramatically:

Lighting – Use of LED lights or low-power fluorescent lights in the data center can reduce PUE by a few hundredths. However, keeping the lights off has a larger impact. Use of occupancy sensors, with a high degree of sensor granularity and a well-designed delay until time-off, can provide sufficient visibility in the data center while keeping the lights off most of the time. Lighting should be designed to produce specific brightness (lumens) at the data center floor, even when IT racks and equipment are installed.

Electrical losses – Electrical system energy efficiency can be improved significantly by using energy efficient equipment and also by deploying the equipment in energy efficient configurations.

Uninterruptible power supplies, transformers, power distribution units and other electrical equipment have inherent losses in use. Even electrical wire itself has small losses when energy flows through it. The selection and use of highly efficient (97+ percent) electrical equipment can significantly improve PUE, especially if the equipment is still efficient at partial loading. Keeping power conditioning and transformation in power rooms and off the raised floor can also improve PUE since it takes more energy to cool equipment to raised floor temperature levels than to the higher temperatures that are acceptable in a power room.

Configuration of the electrical systems can also have significant PUE implications. Since most electrical equipment works more efficiently at higher percentage loads, moving from a "brute force" 2N or 2(N+1) UPS configuration to a "4 to make 3" N+1 configuration can raise normal equipment utilization from 50 percent to 75 percent, increasing energy efficiency substantially, without compromising redundancy or reliability. Careful selection of electrical equipment can reduce PUE by a couple of tenths from the previous pie chart.

Cooling losses – Energy efficiency in a data center is a combination of hundreds of small improvements, and then one area that really matters: cooling. On average, even in well-run data centers, cooling is by far the largest energy use outside of IT equipment. Increasing cooling efficiency is where large PUE improvements can be made. However, cooling efficiency improvements should be balanced by reliability and consistency considerations. It is far easier to introduce large amounts of power into a data center than it is to remove the resultant heat.

Cooling efficiency methods

Cooling systems in data centers are complex and precise, due to the need to provide constant cooling to IT equipment. Traditionally data center cooling has been over-delivered (colder air and a larger volume of air than is required) due to caution, ignorance and poor controls. Areas for improvement in cooling efficiency include:

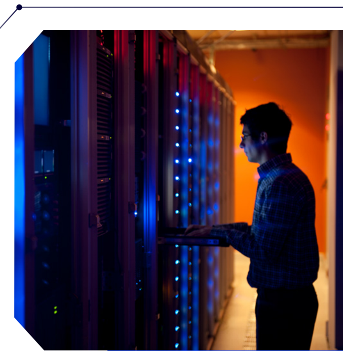
Humidification – Humidification and de-humidification can use a significant amount of energy, especially if air handlers "fight" each other, with some units humidifying and some units de-humidifying. Traditional steam humidification is also costly from an energy perspective, since creating steam takes significant energy, and then the heat in the steam needs cooled, using additional energy. Ultrasonic and other non-heat-based humidification can significantly reduce energy use. Also, modern IT equipment no longer requires tight ranges of acceptable humidity, generally accepting humidity levels from 20 to 80 percent relative humidity. Allowing humidity to float across a wider range reduces the energy needed to humidify and de-humidify.

Hot aisle/cold aisle – Organizing IT equipment in the data center so that it all draws cold air from cold aisles and exhausts hot air into hot aisles significantly improves energy efficiency by reducing the movement of air. Hot aisle/cold aisle organization is evaluated in a number of papers and is fundamental to an energy-efficient data center layout.

Air Movement – Fan horsepower is a significant energy user in data centers, since IT equipment takes in and exhausts a tremendous amount of air. Fan horsepower is required to move air under the raised floor, move air from the servers back to the air handlers, and move air across the cooling coils in the air handlers. The air can be delivered from the air handlers to the servers via velocity (throwing air) or via static pressure (flowing air). Both methods are energy efficient, although static pressure is more flexible, especially in data centers where the hot aisle/cold aisle layout is not fixed over time. Keeping cold air temperature to a designed level, for example, 72-degree cold air, and not over-cooling the air reduces cooling energy use significantly. Also, delivering only the amount of air (cfm) needed by the IT equipment reduces fan horsepower and makes the cooling equipment more energy efficient (higher delta-T across the cooling equipment).

Cooling efficiency – The energy efficiency of cooling is measured by kW/Ton, the amount of energy required to "generate" a ton of cooling. Mechanical cooling devices that use compressors to cool, such as chillers and direct expansion units vary in their energy efficiency from approximately .25kW/Ton for the most efficient large centrifugal chillers to 1.5kW/Ton or higher or less for smaller direct expansion units. The most energy efficient equipment is more expensive up front than the less efficient equipment, but results in a lower PUE and lower total costs over the life of the equipment.

Free cooling – No matter how efficient the compressorbased (mechanical) cooling equipment is, it is still very costly from an energy perspective to use compressors for cooling. Free cooling, or compressor-less cooling, has the ability to dramatically lower the energy used in cooling, and is the single largest opportunity to lower overall PUE. Free cooling refers to using ambient outside air, water or ground temperatures to produce cool data center air vs. using a compressor to cool that air. While free cooling is not "free" from an energy perspective, due to the continued need to move air with fan horsepower, it is "free" from mechanical cooling, which is the largest contributor to high PUEs.



Types of free cooling – While free cooling is sometimes obtained by using cold water sources, such as deep lake or ocean water, or ground sources such as permafrost or other cold land, almost all free cooling is derived from air. There are four major ways to turn outside air into free cooling:

1. Direct air – Introducing cold outside air directly into the data center to cool the IT equipment.
2. Indirect air – Using outside air to drive an air-to-air heat exchanger, such as a heat wheel or Kyoto wheel, which produces cold inside air from the cold outside air.
3. Direct evaporative – Air (commonly outside air but sometimes recirculated inside air) is passed through liquid (most commonly water); the resulting evaporation directly cools the air.
4. Indirect evaporative – Outside air is evaporated to cool a loop of condenser water similar to direct evaporative. The resulting cold condenser water is run through a water-to-air heat exchanger (similar to your car radiator), cooling the inside air.

Pros and cons of different free cooling methods

Each of the methods of air-based free cooling has pros and cons, and is applicable to different local ambient conditions. Close examination of psychrometric charts, with local hourly temperature and humidity plotted, will show which cooling method has the best impact based on local conditions. For example, consistently low ambient temperatures (dry bulb temperature, or how cold the air is at a certain point, think of the Arctic) would yield a large number of direct or indirect air free cooling hours. Consistently low ambient humidity (wet bulb temperature, or how cold the air is made with evaporation, think of Las Vegas) would yield a large number of direct or indirect evaporative free cooling hours.

Use of indirect air free cooling is limited in large applications such as data centers. Air-to-air heat exchange is not nearly as efficient as water-to-water or water-to-air, so air-to-air heat exchangers are not commonly used in large volume data centers.

Use of direct air and direct evaporative free cooling have two significant downsides. First is the need to introduce large volumes of outside air (necessary in direct air, common in direct evaporative). In most locations, outside air has contaminants such as dust, dirt, smoke, chemicals, exhaust, etc. that are not healthy for IT equipment. No matter how much filtering and scrubbing is applied, some of those contaminants will penetrate into the data center, potentially negatively affecting IT equipment. Another side effect of filtering and scrubbing the air is increased fan horsepower, since pushing air past all the filtering equipment requires more horsepower and therefore more energy.

Direct air and indirect air free cooling are also called air side economizing. Direct evaporative and indirect evaporative are also call water side economizing.

The second downside of direct air and direct evaporative free cooling is humidity control. Direct evaporative, by definition, introduces large amounts of humidity into the air entering the data center, and high humidity is not healthy for IT equipment. Direct air introduces humidity that is present in the outside air, which can vary from low to high humidity, even over a short period of time. Changes in humidity are even worse for IT equipment. While humidity can be added to and removed from the air after it is cooled, this process is energy intensive, reducing the benefits of those direct air and direct evaporative.

Free cooling refers to using ambient outside air, water or ground temperatures to produce cool data center air vs. using a compressor to cool that air.

This leaves indirect evaporative as the best free cooling choice for dry climates. The inefficiency in the water-to-air heat exchanger approach is offset by less need for filtering or humidity control, and results in a very predictable air quality since the air from the data center can be recirculated constantly. However, indirect evaporative, like all evaporative methods, does not work well or at all in very humid environments. In very humid environments, direct air free cooling is the most efficient choice, if ambient temperature is low enough.

Possible PUEs

Since the PUE measure was established, many data centers claim low PUE; some supported by facts and some completely fabricated, including some claims of sub-1.0 PUE, which is not possible unless measurements or calculations are done incorrectly. Generally speaking, the most energy efficient data centers today will run at about a 1.5 PUE, in which 0.5 kWh is consumed by support functions for every 1.0 kWh consumed by the IT equipment. While a drop from an average PUE of 2.0 to a best-in-class PUE of 1.5 does not sound like much, remember that 1.0 of both figures is not changeable. Therefore, the 1.5 PUE is twice as efficient as the 2.0 PUE, using only 50 percent of the energy in non-IT equipment functions.

Specialty data centers built by large search and social media companies advertise, and actually achieve, very low PUEs in the 1.1 to 1.2 range. However, those companies can choose to "break" traditional data center constraints which drive energy use. Some of the traditional constraints broken by specialty data centers include:

1. Eliminating humidity control
2. Increasing ranges of acceptable cold air supply temperature, including exceeding 95 degrees Fahrenheit
3. Eliminating power conditioning, including UPS support
4. Eliminating air filtering
5. Using custom-built servers and storage that can operate at higher temperatures

Indirect evaporative, like all evaporative methods, does not work well or at all in very humid environments. In very humid environments, direct air free cooling is the most efficient choice, if ambient temperature is low enough.

Flexential: everything you need to improve the customer experience

At Flexential, we understand there is no one-size-fits-all approach to optimizing the benefits of a hybrid IT strategy. We provide colocation, private cloud and unparalleled network connectivity together with a comprehensive suite of professional services to empower you with the agility, scalability, security and control you need to deliver a superior customer experience.

Learn more