Hyperscale Heat Wave

Enormous Data Centers Creating Energy Hot Spots

By: Fred Moore President
www.horison.com

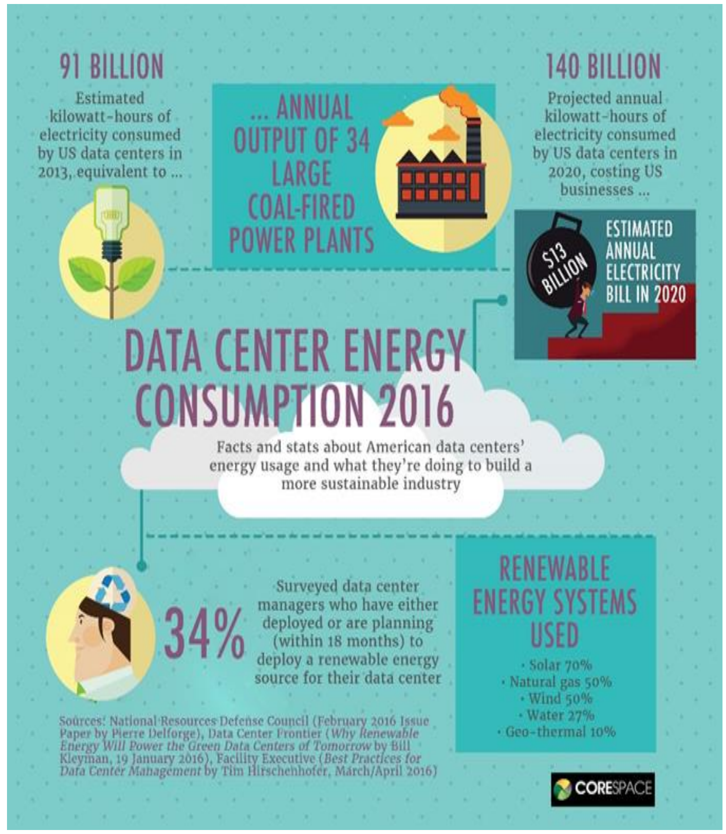
Introduction

Data centers have become the backbone of the global economy from the server rooms that power small organizations to the large enterprise data centers hosting enormous server and storage farms. The world's data centers now consume almost as much energy as Spain. The emerging and enormous hyper-scale data centers are relatively few and currently represent a relatively small fraction of all data center energy consumption in the United States, but they are quickly replacing thousands of smaller data centers.

The hyperscale data center market is comprised of cloud providers, colocation providers, and the largest enterprises and High Performance Computing (HPC) data centers. The HPC market segment is primarily focused on very high performance while the Hyperscale market is focused on massive scalability and both employ advanced cooling systems and redundant power. Heat is the major factor that can increase computer technology failure rates and lowers device reliability. Unfortunately, traditional data centers have generally made less progress in energy management than their new hyper-scale cloud counterparts.

Large, multi-tenant facilities are now seen as more energy efficient than small server rooms. When you move from onsite to larger colocation facilities, you're leveraging economies of scale and expertise and gain energy efficiency. Overall, electronics of all types are the fastest-growing consumer of electrical energy worldwide. Today its estimated about 5 percent of total global energy consumption is spent on electronics, and that's projected to grow to 40-50 percent by 2030 at the current pace if there are no major advances in the field that lead to lower energy consumption.

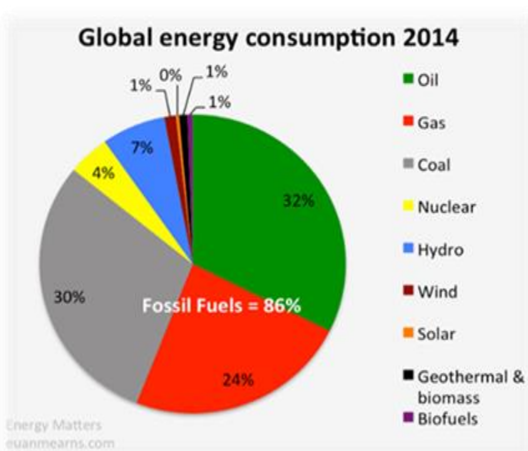
In 2013, U.S. data centers consumed an estimated 91 billion kilowatt-hours of electricity, equivalent to the annual output of 34 large (500-megawatt) coal-fired power plants. Data center electricity consumption is projected to increase to roughly 140 billion kilowatt-hours annually by 2020, the equivalent annual output of 50 power plants, costing American businesses $13 billion annually in electricity bills and emitting nearly 100 million metric tons of carbon pollution per year.

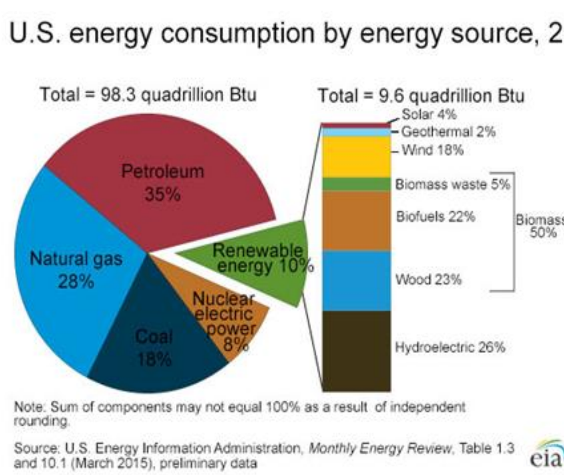


The explosion of digital content is setting the data center market on fire and has made data centers of all sizes one of the fastest-growing consumers of electricity in developed countries. Data center energy consumption has become one of the key criteria in the type of construction and geographic location of new utility company power plants. Energy consumption is one of the largest data center operating expenses and in some cases, can account for nearly 50 percent of total operating expenses.

Global and U.S. Energy Consumption by Source

Global energy consumption is the total energy used by all human civilization. Amazingly, it is estimated that 1.4 billion people in the world have no access to electricity. Typically measured annually by a variety of sources, it includes all energy harnessed from every energy source across every single industrial and technological sector and across every country. World energy consumption has deep and long-term implications for humanity's social, economic, and political sphere. In 2014, fossil fuels (coal, natural gas, and petroleum) accounted for 86% of the global and 81% of the U.S. energy consumption. Major energy sources and their percent share of total Global and U.S. energy consumption for 2014 are shown below.





Keep in mind that energy consumption and energy generation are not the same. Electricity is the world's fastest-growing form of end-use energy consumption, as it has been for many decades. A percent of all energy generated is normally lost in follow-on processes and all energy generated does not produce electricity. World net electricity generation is expected to increase by 69% in the IEO (International Energy Outlook) 2016 Reference case, from 21.6 trillion kilowatt-hours (kWh) in 2012 to 25.8 trillion kWh in 2020 and to 36.5 trillion kWh by 2040. Power systems have continued to evolve from isolated, noncompetitive grids to integrated national and even international markets. This is only the beginning of many challenges that lie ahead for the world's expanding electrical grid.

Electricity consumption is the form of energy consumption that uses electric energy. Electric energy consumption is the actual energy demand made on the existing electricity supply.

Electricity generation is the process of generating electric power from other sources of primary energy. Electricity is most often generated at a power station by electromechanical generators, normally driven by heat engines fueled by combustion or nuclear fission, and by other means such as the kinetic energy from flowing water and wind. Other energy generation sources include solar and geothermal power. Coal-fired power plants are being retired creating a growing demand for natural gas and are making natural gas the fuel of choice for electricity generation. In any case, electricity is the lifeline for today's data centers.

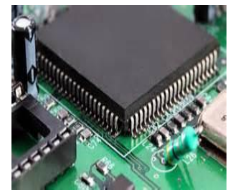
What is Hyperscale Computing?

Hyperscale solutions de-couple application computing, software and storage resources, enabling each to scale either performance or capacity independently depending on business needs. As storage needs grow, companies can add servers running software defined storage (SDS) to enable policy-based provisioning of data storage independent of the underlying hardware and to expand capacity independently of the application tier. For hyperscale systems, data is automatically distributed across the entire cluster of storage servers as new nodes are added to the system. Conversely, as performance needs grow, companies can add servers to increase compute power, independent of the storage tier.

The Hyperscale Shift scenario involves the consolidation of many of the servers in smaller non-hyperscale data centers into enormous hyperscale data centers. Hyperscale datacenters are used by enterprises that maintain thousands of servers and store vast amounts of data and some will approach the Exascale ($1x10^{18}$ bytes of total storage capacity) levels by 2020. Consider data centers with thousands of servers of different ages and efficiency levels. When a server is plugged in, it consumes electricity non-stop over a 24-hour period totaling 8,760 hours in a year. Such datacenters require continued advancements in server rack design and storage systems to cope with major growth in energy consumption, number of users, the amount of data and number of devices cost-effectively.

The Hyperscale shift is driving a massive amount of investment in enormous, new data centers that power cloud, mobile, online retail, social and analytic (Big Data) type workloads. Hyperscale data centers are often associated with cloud computing and were pioneered by the very large data centers owned by Amazon, Apple, Facebook, Google, and Microsoft, who are literally spending billions of dollars on new data centers and advanced energy systems to power their workloads.

Hyperscale data centers operate servers at higher utilizations in infrastructure-efficient spaces so that consolidating IT services from many small individual data centers into a hyperscale data center can yield significant overall energy savings. The variety, efficiency and complexity of servers are improving due to pressure from these hyperscale IT companies and thus, there will be numerous new and improved technologies available for servers.

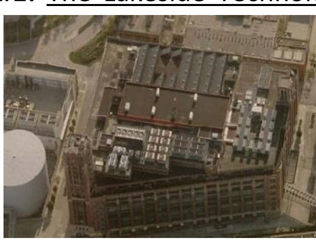


For example, there's a growing need for low-power servers in higher-density data centers. One of those low-power server options relies on further advancements with the System-on-a-Chip (SoC). These SoCs will include integrating networking controllers, storage controllers, co-processors and memory into a single chip. SoCs are very common in the mobile electronics market because of their low power-consumption.

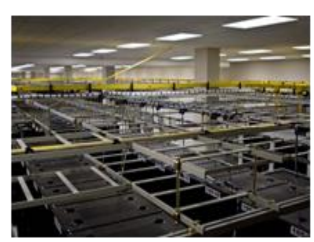
Hyperscale data centers, which are typically built with stripped down commercial off the shelf servers, storage and network equipment, can have millions of virtual servers and often accommodate increased computing demands without requiring net additional physical space, cooling or electrical power. The hyperscale servers are customized for data center needs and the racks are wider than the standard, traditional 19-inch racks. This accommodates more components across the motherboard as these PCs are assembled from common components that can be easily swapped out by the user when failures occur.

There is a broad interest throughout the IT industry in hyperscale computing now because the open source software that such organizations have developed to run their huge data centers is expected to trickle down to smaller organizations, helping them to become more efficient, use less power and respond more quickly to their own user's needs.

Let's Look at Some of the World's Largest Traditional and Hyperscale Data Centers. For reference, the perimeter of a football field is 1,040 feet and the area is 57,600 ft^2.



#1: The Lakeside Technology Center in Chicago is a 1.1 million square foot multi-tenant data center hub owned by Digital Realty Trust. Originally developed by the R.R. Donnelley Co. to house the printing presses for the Yellow Book and Sears Catalog, the location was converted to telecom use in 1999, and today it is one of the world's largest telecom carrier hotels and the nerve center for Chicago's commodity markets, housing data centers for financial firms among the 70 tenants.



#2: The QTS Metro Data Center in Atlanta is a 990,000-square foot building, and when complete will yield 560,000 square feet of data center space. The huge facility is supported by 80 megawatts of power capacity via an on-site Georgia Power substation, and has 19 two-megawatt diesel generators to provide backup power.

#3: The NAP of the Americas in Miami is a massive Terremark Worldwide data fortress that was completed in June 2001, offering 750,000 square feet of data center footprint just as the dot-com bust was under way. The six-story facility not only survived the downturn, but has become a key connectivity hub for the Southeastern US and Latin America, providing critical infrastructure to the U.S. military and the global domain name system.

#4: The Next Generation Data Europe is a new huge facility in Newport, Wales recently opened its doors with British Telecom and Logica occupying two large data halls at the 75,000-square meter (750,000 square foot) center.

#5: Microsoft- Chicago, Illinois 700,000 square feet

One of the largest data centers ever build so far but also being the most unusual. The two-story building uses 40-foot shipping containers to house their web servers and uses the second story for the traditional raised-floor data center space. Microsoft's data centers store more than 30 trillion individual pieces of data. Read More . . .

#6: Microsoft- Dublin, Ireland 550,000 square feet

Located in Ireland, Microsoft has another major facility for their endless amount of data and helps to power their global cloud computing operation. This data center operates at a Power Usage Effectiveness (PUE) of 1.25 and gains an advantage from the facility’s ability to use outside air to cool the data center. Microsoft's data centers process over 1.5 million requests every single second. Read More . . .

#7: PHOENIX ONE, PHOENIX, AZ 538,000 square feet

i/o Data Centers uses the 538,000-square foot Phoenix ONE facility site as both a data center and its corporate headquarters. The huge data center features a number of design innovations, including an enormous rooftop array of solar panels that will eventually generate as much as 4.5 megawatts of power for the data center, and a thermal storage system that will allow i/o Data Centers to run chillers for its cooling systems at night when power rates are lower.

#8: Apple- Maiden, North Carolina 514,246 square feet

The Apple data center itself sits on 183 acres of land that Apple purchased to help run the center and rumor has it that they are looking to buy another 75 acres right across the road. The largest private solar array was built by Apple in North Carolina to power this amazing data center. Read More . . .

#9: Google- Lenoir, North Carolina 500,000 square feet

No one know the exact number but North Carolina media indicate that the two buildings that make up the data center total more than 500,000 square feet. This is equal to 8.7 football fields. Google built a 20-megawatt solar farm next to the Lenoir facility to meet power requirements. A second 20 megawatts solar farm is expected late this year. Read More . . .

#10: Microsoft- Quincy, Washington 500,000 square feet

The Microsoft data center in Quincy Washington takes advantage of hydro-electric power from nearby dams on the Columbia River. They are planning to expand this site and when Microsoft completes the development the facility will be the largest in the world. The number of times Microsoft's fiber optic network, one of North America's largest, could stretch to the moon and back is three. Read More . . .

#11: DuPONT FABROS- **CH1, Elk Grove Village, Ill.** 485,000 square feet

This huge facility in the Chicago suburbs marked the first major expansion beyond the northern Virginia market for **DuPont Fabros Technology** (DFT), a real estate investment trust specializing in data centers. The facility is supported by 32 huge 2 megawatt diesel backup generators with 200,000 gallons of diesel storage, as well as 32 rotary (flywheel) UPS systems.

#12: Microsoft- San Antonio, Texas 477,000 square feet

This enormous Microsoft data center took almost 1,000 full time employees to build. The enormous 477,000 square foot data center is a 1.3 mile walk around to entire building. Read More . . .

#13: Facebook- Altoona, Iowa 475,000 square feet

The Altoona, Iowa data center of 475,000 square feet is one of four facilities that Facebook owns and operates. This enormous center is vital to the companies efficiently because Facebook currently accounts for 9 percent of all internet traffic. Read More . . .

#14. THE SUPERNAP**, Las Vegas, NV** 407,000 square feet

The 407,000 square foot data center in Las Vegas was built to accommodate up to 7,000 cabinets, with 250 megawatts of power capacity. Others have since built larger facilities as the data center arms race accelerated, but the SuperNAP is perhaps most notable not for its size but its ability to support power densities of up to 1,500 watts a square foot.

#15: Facebook- Prineville, Oregon 334,000 square feet

Coming in at number seven is Facebook for the second time. The data center was built in two phases and sits on 120 acres of land. Over the development of the site more than 1,300 people worked on the center and an average of 250 workers a day. Read More . . .

#16: Facebook- Forest City, North Carolina 300,000 square feet

The data center is Forest City is almost as big as Microsoft's but arranged must differently. The 300,000-square foot building looks like 3 aircraft carriers place front to back and dropped in the middle of the landscape. Read More . . .

#17: Yahoo- Lockport, New York 275,000 square feet

Yahoo's Lockport data center recently experienced a 170-million-dollar expansion to its facility based in upstate New York. The 150,000 square-foot add on will provide additional capacity and a new customer service center. The data center was built in two phases on a 30-acre property. Read More . . .

What's on the Horizon? The United States is home to the world's largest data center and data center campuses, but for how much longer? According to Forbes, China is scheduled to a complete a data center in Langfang by 2016 that will rival the size of the Pentagon. The Chinese may only be able to be the largest data center for three to five years. Microsoft recently announced a new data center initiative near Des Moines, Iowa that will cost it $1.13 billion, and will also be more than 6 million square feet. It is due to be completed between 2019 and 2021. Microsoft has more than 100 data centers in its global portfolio.

What About Google?

The specific size of all Google’s data enters is not made available but Google has fifteen extremely large data centers around the world. Google operates eight data centers in North America, one in South America, two in Asia and four in Europe and has several others under construction. [How does Google decide where to build its data centers?] Here are the factors that are known to influence Google’s data center site location process:

- The availability of large volumes of cheap electricity to power the data centers.
- Google’s commitment to carbon neutrality, which has sharpened its focus on renewable power sources such as wind power and hydro power. The Dalles, Oregon was chosen primarily for the availability of hydro power from the Columbia River, while the local utility’s wind power program influenced the selection of Council Bluffs, Iowa.
- The presence of a large supply of water to support the chillers and water towers used to cool Google’s data centers. Several recent Google data center sites have been next to rivers or lakes.
- Large parcels of land, which allow for large buffer zones between the data center and nearby roads. This makes the facilities easier to secure, and is consistent with Google’s focus on data center secrecy. Google purchased 215 acres in Lenoir, NC, 520 acres for the Goose Creek, SC project, 800 acres of land in Pryor, OK and more than 1,200 acres in Council Bluffs, Iowa. The extra land may also be used for windmill farms to provide supplemental power at some facilities.
- Distance to other Google data centers. Google needs lightning-fast response time for its searches, and prizes fast connections between its data centers. While big pipes can help address this requirement, some observers believe Google carefully spaces its data centers to preserve low latency in connections between facilities.
- Tax incentives. Legislators in North Carolina, South Carolina, Oklahoma and Iowa have all passed measures to provide tax relief to Google.

Google’s data centers can often use 50% less energy than the typical data center. Google raises the temperature to 80°F, uses outside air for cooling, and builds custom servers. Per various estimates, Google has around 900,000 servers in all its data centers based around the world. Google’s data centers currently use around 260 million watts of power which accounts to 0.01% of global energy consumption!

Water vapor rises above the cooling towers in The Dalles data center in Oregon. These plumes of water vapor create quite a mist at dusk.

Google shares detailed performance data to help move the entire industry forward. Google's massive global data center infrastructure supports Internet services such as Google Search, Gmail and YouTube, but its servers and storage farms generate massive amounts of heat that must be removed to keep the equipment running reliably. Nowhere is the hyperscale heat wave and its implications more visible than at Google.

A Look at Amazon's Data Centers The [Amazon Web Services (AWS)] doesn't aggressively brand, provide specific details or call attention to their data centers. The AWS Cloud currently operates 38 Availability Zones within 14 geographic Regions around the world, with nine more Availability Zones and four more Regions planned to come online. Availability Zones consist of one or more discrete data centers, each with redundant power, networking and connectivity, housed in separate facilities. Amazon wind and solar farms are gaining momentum to meet the hyperscale energy demand for Amazon's data centers.

What's Going on Inside the Data Center? The simple fact is that today's data centers are extremely hungry for electricity, with the Uptime Institute estimating that they collectively consume as much as three percent of all global electricity production. Data centers invest in complex building management systems that can monitor and provide a tremendous amount of information on energy consumption and develop strategies for energy reduction.

Power Usage Effectiveness (**PUE**) is a measure of how efficiently a computer data center uses energy; specifically, how much power is used by the computing equipment (in contrast to cooling and other overhead). PUE was developed by a consortium called The Green Grid. PUE is the ratio of total amount of energy used by a computer data center facility to the energy delivered to computing equipment.

Data Center Infrastructure Efficiency (DCIE) is the inverse of PUE. DCIE is the percentage value derived by dividing computing equipment power by total facility power. The ideal PUE would be 1.0 and DCIE 100%. Anything that isn't considered a computing device in a data center (i.e. surveillance, signs, lighting, cooling, etc.) falls into the category of facility energy consumption.

PUE = Total Facility Power / IT Equipment Power

DCIE = IT Equipment Power / Total Facility Power

Power (Watts or W) $P = V \times I$ or $I^2 \times R$
Voltage (Volts or V) $V = I \times R$
Current (Amps or I) $I = V \div R$
Resistance (Ohm) $R = V \div I$
The kilowatt hour (kWh) is a unit of energy equal to 1,000 watt-hours

PUE	DCIE	Level of Efficiency
3.0	33%	Very inefficient
2.5	40%	Inefficient
2.0	50%	Average
1.5	67%	Efficient
1.2	83%	Very efficient
1.0	**100%**	**Maximum efficiency**

Advanced cooling strategies such as hot aisle isolation and liquid cooling help make the cooling process less energy intensive. Servers designed to sharply reduce their energy consumption during off-peak periods are another key reason for the efficiency gains, as is the increasing prevalence of cloud and virtualization. The Industry-wide weighted average PUE remains at 1.89.

What about Wasted Electricity? Server utilization rates are quite low, typically averaging under 25% percent busy, yet they consume energy 24 hours per day. The next biggest energy consumer, spinning disk drives (HDDs) average 50% or less used capacity while normally drawing power 24 hours per day. When you consider 1U servers (U or RU is a rack unit of measure defined as 1.75 inches) can use about 100 to 200 W, even while idle, it becomes obvious that a significant percent of year-round energy costs goes to underutilized computing equipment.

Intelligent power management solutions, some beginning to implement artificial intelligence (AI), allow data centers to track seasonal, daily, and even hourly fluctuations in demand, and then dynamically adjust the power state of idle server in response and even shut off unused resources. These solutions include everything from virtualization, to Data Center Infrastructure Management (DCIM) products, to Application-Aware Power Management (AAPM) software. As more data centers adopt these tools in the future, they will have much deeper insight into managing their energy consumption, server and application performance, and energy productivity.

What else could add to an energy efficient data center infrastructure? Think about rooftop arrays of solar panels generating as much as 4.5 megawatts of power, lightning prevention systems that detract strikes from a 300-foot radius, flood-aware exteriors, cool outside air intake systems and structures designed to withstand a category 5 earthquake.

Who Builds These Enormous Data Centers? For example, [Sabey Data Centers] manages or has under development more than 3 million square feet of building space specifically built to house efficient data centers. Those centers represent 65 percent of the value of a Sabey enterprise currently worth more than $600 million. Sabey recently paid $120 million for a windowless Verizon building in lower Manhattan; the site will likely prove to be the city's largest special purpose data center. Sabey and others focus on better ways to provide more energy for less cost, low PUE ratios, utility incentives and industry recognition for superior design and management. [Switch SuperNAP] builds large-scale, highly energy efficient data centers that can run on 100% renewable energy. While many data centers are built, and operated by real estate developers, SuperNAP's passion is evolved ecosystems and efficient data centers.

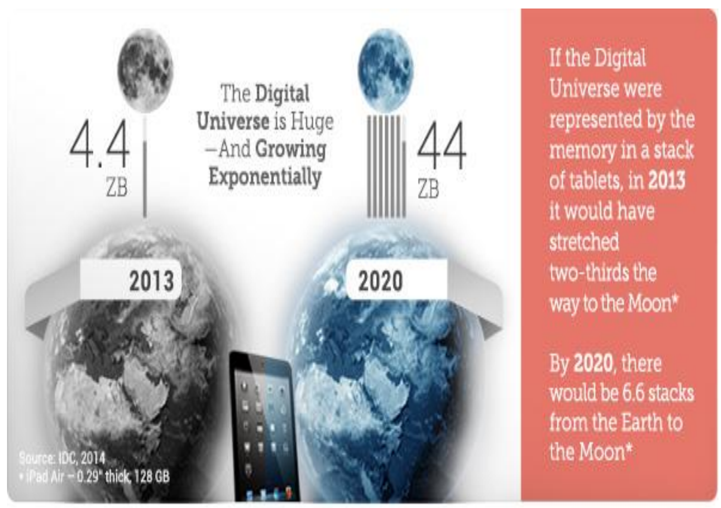
Artificial Intelligence for Energy Arrives - Google tapped into the superior intelligence of its [DeepMind Neural Network] to find ways to vastly reduce the energy it uses in its data centers, which account for 40% of the worldwide Internet activity. Artificial intelligence (AI) is, in simple terms, the science of doing by computer the things that people can do. DeepMind, a London-based artificial intelligence company that [Google acquired in 2014], is a neural network inspired by the human central nervous system that can actively learn about an environment in order to solve complex tasks.

An Underwater Data Center? - The ocean's cooling properties, and untapped energy potential led Microsoft to successfully run a cloud center there for 105 days. [Project Natick](http://www.citylab.com/tech/2016/02/microsoft-cloud-ocean-project-natick/459318/) submerged a 38,000-pound, 10-by-seven-foot steel tube off the coast of California for three months to see if the servers inside it continued to function with the power of 300 desktop PCs. The underwater data center was equipped with 100 sensors that tracked humidity, motion, and pressure, and didn't experience and hardware failures or leaks. The data center ran projects from Microsoft's Azure, a cloud computing platform. The company says this is the first time a data center has successfully operated under the sea. That achievement opens the door for future data-storage infrastructure to take advantage of the natural cooling properties and renewable energy of underwater locations.

Source: http://www.citylab.com/tech/2016/02/microsoft-cloud-ocean-project-natick/459318/

What's Fueling the Hyperscale Heat Wave?

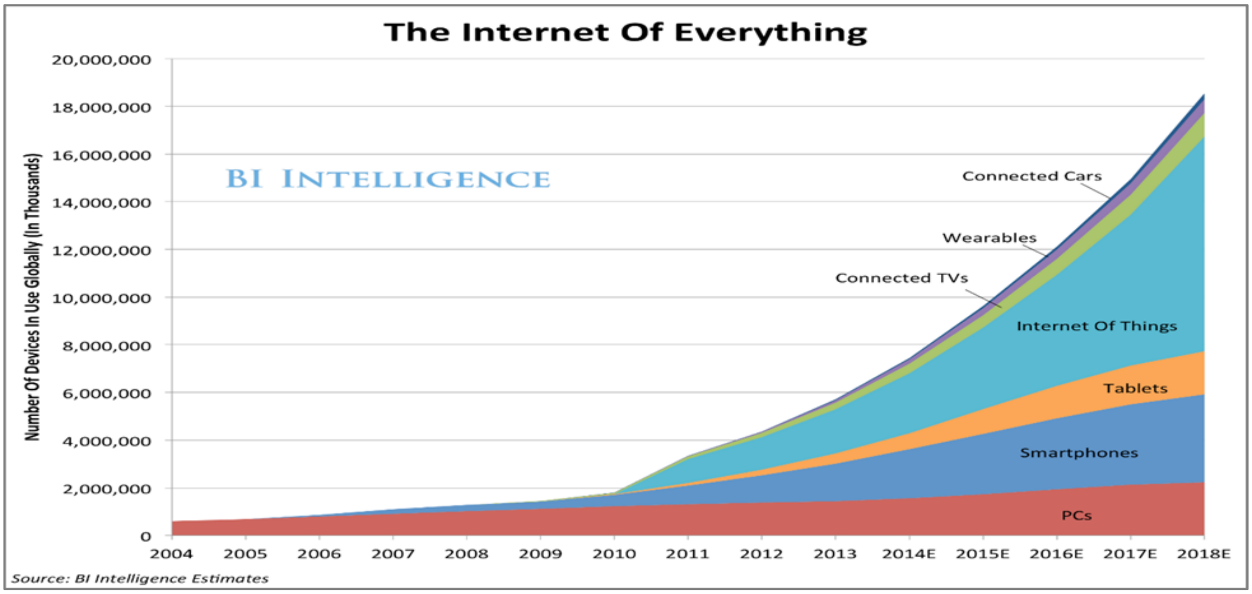
The answer is the exponential data growth of the Internet is driving an enormous appetite for network capacity and data storage, creating a new class of data centers that can scale along with the Internet. **The Digital Universe,** or the amount of digital data, which is growing exponentially into the next decade, is expanding to include not only the increasing number of people and enterprises doing everything online, but also all the IoT (Internet of Things) – smart devices – connected to the Internet, unleashing a new wave of opportunities for businesses and people around the world. The billions of connected devices will trigger a greater demand for electricity to power and cool the technology.



Overall data center workloads are projected to nearly double from 2013 to 2018. However, cloud workloads will nearly triple over the same period. Workload density (workloads per physical server) for cloud and hyperscale data centers will grow from 5.2 in 2013 to 7.5 by 2018 as multi-core servers get more dense and powerful. Comparatively for traditional data centers, workload density is expected to grow from 2.2 in 2013 to 2.5 by 2018.

Keep in mind that the energy provider or utility company's energy rates always increase making net energy expense reduction difficult. If you reduce your consumption 15% and the rates go up 15%, you don't reduce your energy bill! You essentially end up paying more for less energy consumption. Lowering your energy consumption and lowering your energy bill are not the same.

The Internet of Things (IoT) Various industry sources estimate that the IoT, or the Internet of Everything, will connect over 25 billion "things" to the Internet by year 2020. If it has an IP address and connects to the network, it will be a candidate for a node in the IoT. Clearly the true value of connecting everything imaginable to the internet is questionable. In the future, IoT sensors collect data that could run a dishwasher based on how cold or hot it is outside, or even switch off and balance appliances such as air conditioners or wind turbines when the other devices are utilizing greater amounts of power.

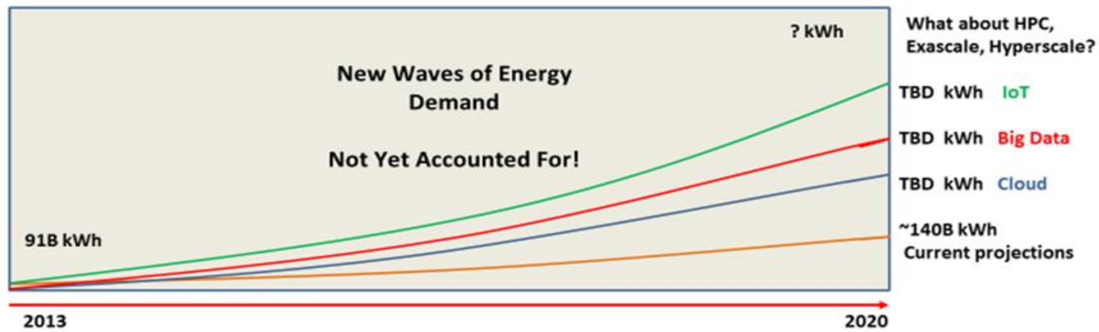


Intelligent power-management software is one opportunity for IoT, but it promises to provide relief to this seemingly endless cycle of increased energy consumption. By 2020, IoT analytics are expected to provide $60+ billion in annual savings to the U.S. enterprises that are equipped. However, IoT technology will itself require increased power, significantly more data storage capacity and more bandwidth to contain the data to analyze and make the intelligent decisions.

Cloud computing and hyperscale data centers are inseparably linked now and for the foreseeable future. For the next five to 10 years, expect to see huge increases in storage and bandwidth to get to the cloud from appliances, smart phones and mobile applications. Streaming video has exceeded expectations for bandwidth and storage and will require more servers to process all the new data, more storage capacity, more energy, and more bandwidth to move the data from the points of collection to the point of analysis, often in the cloud. This insatiable chain reaction for electricity demand seems undeniable.

A Solid Granite Cloud? - The Perpetual Storage, Inc. (PSI) data vault is excavated into a solid granite mountain in Salt Lake City's Little Cottonwood Canyon. The vault is encased in granite with a galvanized steel inner lining and should withstand any force known to man. The physically secure PSI vault is one of safest offsite storage facilities in the world. PSI provides high-security data transport and storage of back-up, disaster recovery (DR) and archival data through various fiber optic carriers into a scalable facility within the granite mountain.

IT Energy Consumption to Accelerate



- **Several New Waves of Storage Demand Will Increase Data Center Energy Consumption.**
- **In 2013, U.S. data centers consumed an estimated 91 billion kilowatt-hours of electricity, equivalent to the annual output of 34 large (500-megawatt) coal-fired power plants.**
- **Data center electricity consumption is projected to increase to roughly 140 billion kWh annually by 2020 - based on current trends**
- **But does not include the impact of IoT, Big Data and Cloud demand.**

Source: Horison, Inc.

Managing Data Center Energy Consumption

By implementing aggressive off-peak and load-balancing management strategies, the typical data center can save 10-20% or more of its annual energy consumption. However, the largest opportunities for IT energy consumption savings are linked to the surprising under-utilization of data center equipment, especially electricity-hungry servers and HDDs.

In large scale data centers, up to 30 percent of the servers may be “comatose” and no longer needed. The average non-mainframe server averages less than 30% busy and the average HDD seldom allocates over 50% of its total capacity with the rest being unused (wasted) space. HDDs range from 5,400 to 15,000 RPMs and their typical operating power ranges from 6-9 watts of power to spin it and to cool it as well. Flash SSDs use about 1/3rd the watts of HDDs. Tape cartridges spend most of their life in a library slot or on a shelf and consume no energy when not mounted on a tape drive. Thus, energy costs for an equivalent amount of tape capacity are less than 5% of the equivalent amount of disk capacity.

The overall server install base is set to grow by 5% and digital data creation is growing at 30-35% annually further driving energy requirements. Servers in hyperscale data centers use more power than those in internal or service provider due to higher utilizations but they are also becoming more efficient. See 2014 Average PUE chart below.

2014 Average PUE by Space Type						Source: U.S. Data Center Energy Usage Report June 2016
Space Type	IT	Transformer	UPS	Cooling	Lighting	Total PUE
Closet	1	0.05	-	0.93	0.02	2.0
Room	1	0.05	0.2	1.23	0.02	2.5
Localized	1	0.05	0.2	0.73	0.02	2.0
Mid-tier	1	0.05	0.2	0.63	0.02	1.9
High-end	1	0.03	0.1	0.55	0.02	1.7
Hyperscale	1	0.02	-	0.16	0.02	1.2

Managing Data Center Energy Consumption

Data Center Use 150 watts/sq. ft. ~ $700-950/sq. ft.
Office Buildings Use 3-5 watts/sq. ft. ~ $60-90/sq. ft.

What Can the Data Center Do?

- Virtualize Servers/Reduce Server Count.
- Implement Tiered Storage, Archive to Tape.
- Retire Old and Idle Technology.
- Re-evaluate Cooling Strategies; Consider Hot/Cold. Aisles or Liquid Submersion Cooling.
- Use Energy Consulting Group and Software Tools.
- Take Advantage of Utility Efficiency Incentives.

What Can the Electrical Utility Do?

- Offer alternative energy Sources.
- Offer energy management assistance.
- Help data centers implement cost-effective energy. efficiency measures.
- Provide rebates and other incentives for efficiency.

Average Data Center Power Distribution	
IT Infrastructure – External Consumption	Avg.
Chillers, cooling, air-conditioning, pumps, water	24%
Uninterruptible power supply	8%
Air movement, circulation, fans etc.	10%
Misc. lighting, security, surveillance, appliances	3%
Total External	45%
IT Hardware – Internal Consumption	
Servers	>30%
Disk drives and control units	~15%
Tape drives, robotic libraries	<2%
Network gear, SAN switches and other devices…	8%
Total Internal	55%

Source: E Source, Horison, Inc.

Best practices for using less energy in the data center directly focus on the two highest areas of consumption – servers and disk storage. Consolidation of servers by the virtualization of lightly utilized servers, decommissioning unused (orphaned) servers, and purchasing more energy-efficient servers provide the biggest benefit. Upgrading to more efficient UPSs, and PDUs can provide significant benefits.

While server consolidation is relatively straightforward, attaining an optimized storage infrastructure can be more involved, but presents an enormous opportunity for energy reduction. Studies consistently show that during an average life cycle, most data becomes stale (inactive or archival) 30 to 90 days after creation date. In that case, it makes sense to move that data from expensive, primary disk storage, to more economical tiers of storage such as low cost capacity disk and optimally to modern tape storage systems. The progress of tape technology in recent years has redefined the role of tape going forward.

Today's most efficient storage systems integrate SSDs, disk drives, modern tape, compression, thin provisioning, load balancing, and de-duplication with virtualization capabilities to manage the data explosion and reduce complexity. Data centers are capitalized and built on long time horizons. Thus, data center designed, built and commissioned using traditional, power and cooling models can be difficult to upgrade in any significant way for several years. Disruptive power and cooling technologies cannot be easily adopted within an existing facility if the original design did not call for them.

How About a 100% Renewable Energy Data Center in Iceland?

ICELAND DATA CENTER

Iceland's doesn't get as cold as one might imagine as it has a very narrow range of temperatures, and stays cool all year. The country relies entirely on two sources of emissions-free renewable power: geothermal and hydroelectric. Data center operators can just let the outside air in! Iceland's power generation is predominantly based on hydropower, but they are increasingly building geothermal turbines and wind farms as well. Iceland (pop. 330,000) is well connected with multiple fiber cables to both Europe and the U.S. and is located half-way between each continent thus location isn't an issue.

Faster processors and 7x24 spinning disks consume most of the data center energy and generate most of the heat. Racks now consume 10 times the power they did a decade ago as packaging densities have increased. The chart below summarizes several data center server and storage optimization techniques.

Improving Server and Storage Energy Efficiency

Server Challenge	Server Efficiency Solution	Avg. Server/Energy reduction	Technologies Used
Too many servers, or lightly used servers	Consolidate by virtualizing servers	20-50% ...or more	Virtualization software
Unused or orphaned servers	Decommission and remove unused servers	Varies widely	Management decision
Storage Challenge	**Storage Efficiency Solution**	**Avg. Storage/Energy Reduction**	**Technologies Used**
Tiered storage	Move less active and archival data from disk to tape	~20 - 50% of GB can be moved to tape	Tape is the optimal archive solution
RAID 1 creates a duplicate copy of disk data but also doubles the storage and power consumption	For storage that is not mission critical, use RAID 5 as it guards against a single disc drive	Expect to reduce disk energy by 30-45%	Disk
Reduce volume of backup data (backup window)	Incremental, differential, and deduplication	20 - 80%	Tape, disk
Eliminate duplicate data (Primarily backup data)	Deduplication *(Must re-hydrate data first)	Up to 80% of GB Varies by degree of duplicity	Disk
Too many devices	Consolidate/ virtual storage	Reduction varies	SSD, Disk and Tape
Data Compression (reduce file sizes)	Compression ratio (2.5 to 1)	50 - 66% GB	Tape, Disk, Flash SSD
Over-allocation of disk	Thin provisioning	20 - 30% GB	Disk
Lower Speed Disk Drives When Possible Drives range from 5,400 to 15,000 RPM	Use slower spin speeds as power use is proportional to the cube of disc spin speed	20-40%	Lower cost SATA disk versus faster but more expensive SAS or FC
Flash SSD	SSD are much faster than SSD with no moving parts	Uses typically 1/3rd the energy of disk drives	Flash SSD

There are numerous technologies listed in the table above that storage administrators can implement to significantly improve storage efficiency. Storage efficiency is the ability to store and manage data that consumes the least amount of space with little to no impact on performance; resulting in a lower total operational cost. Storage efficiency addresses the real-world demands of managing costs, reducing complexity and limiting risk. Without major changes, lifetime electrical cost will exceed the cost of IT equipment. The Storage Industry Networking Association (SNIA) defines storage efficiency in the SNIA Dictionary as follows:

$$\textbf{Storage efficiency} = \frac{(\textbf{effective capacity} + \textbf{free capacity})}{\textbf{raw capacity}}$$

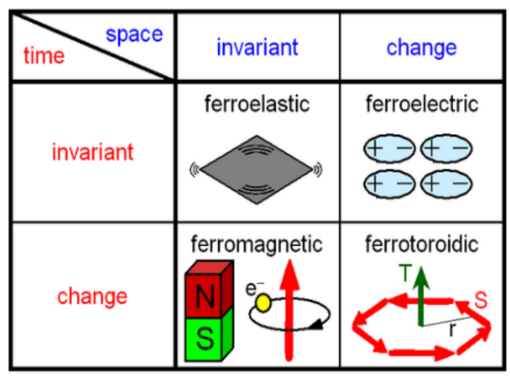
New and Future Energy Solutions and Projections

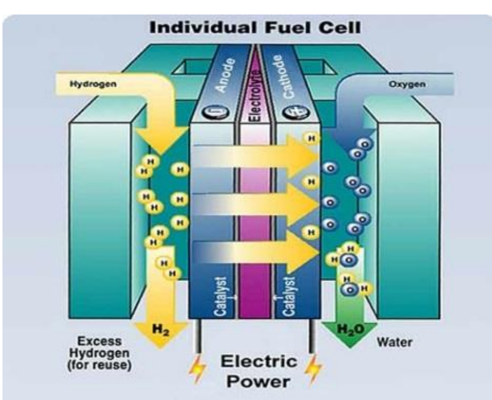
The Tesla Powerwall is another energy savings solution for homes and will address energy availability for millions of homes and smaller businesses. With many of the IoT connected "things" being dependent on electricity in the home, the availability requirement of electricity in the case of local power failures will go well beyond the data center. The Powerwall home battery charges either using electricity generated from solar panels, or when utility rates are low, and powers your home in the evening. It also protects the home against power outages by providing a backup electricity supply. Automated, compact and simple to install, Powerwall offers independence from the utility grid and the security of an emergency backup.

[Tesla Powerpack Battery Storage Station](#) Pacific Gas and Electric has gone live with its first utility-scale, lithium-ion battery storage system. The storage substation, located about 50 miles north of Sacramento, is made of 22 Tesla Powerpacks, each the size of a refrigerator. In all, the 500-kilowatt (KW) substation can store up to 2000KW hours (KWh) of power, enough to power up to 380 typical homes. The batteries can discharge at full power for four hours.

[Floating Solar Panels](#) The world's largest ongoing floating solar panel project on the Yamakura Dam reservoir in Japan, has a generating capacity of 13.7MW. Currently, there is less than 50MW of floating solar power installed globally, but that could double to 100MW this year. With space limited in some parts of the world and real estate getting pricey companies are looking to install PV panels on lakes, canals, bays and perhaps even out on the ocean. The world's largest floating solar panel farm is currently the Yamakura Dam Reservoir in Japan. Completed in 2016, it has a generating capacity of 13.7MW and can power more than 5,000 households.

[Multi-ferroics](#) operate at room-temperature and hold much promise. They are becoming a hotly pursued goal in the electronics field because they require much less power to read and write data than today's power hungry semiconductor-based devices. In addition, they are non-volatile meaning data doesn't vanish when the power is shut off. Those properties could power devices that require only brief pulses of electricity instead of the constant stream that's needed for current electronics, using an estimated 100 times less energy.





[Fuel cells](#) have long offered great promise and have been proposed as a good way for data centers to become more independent of the local electricity grid in the future. Fuel cells convert chemical energy into electrical energy and most use hydrogen and oxygen as the chemicals. Although hydrogen is the most abundant element in the universe, it is difficult to store and distribute. The biggest remaining hurdle for fuel cells today is cost. Fuel cells cannot yet compete economically with more traditional energy technologies, though steady technical advances are being made. Patience may be the key to the arrival of fuel cells.

Summary

Along with food, water and oxygen, electricity may be the fourth most critical element of global survival. Imagine what the world would look like without electricity. Banks, airlines, hospitals, communications, the military defense systems would cease to function. The global food supply would dramatically change without refrigeration. It’s nearly unthinkable but a cyber-attack or complete failure of the power grids could be catastrophic. Securing the power grid should be one of the most critical national priorities - is it? Without electricity, there is no IT industry!

Fortunately, both single-company data centers and multi-tenant hyperscale data centers are making significant advancements in solving the energy efficiency equation. At a minimum, the two biggest steps to reduce data center electricity consumption are led by server virtualization and moving low-activity and archival data from disk to modern tape subsystems. Growth in hyperscale computing has just begun and is providing valuable insights into what the future of what computing will look like. As more and more data centers migrate to the cloud, it’s possible one day that servers and storage will be removed from all data centers except the cloud providers. Maybe then computing would eventually be like the electricity utility – an information utility where you just plug in and pay for what you use?

The Hyperscale Data Center Market Size is estimated to grow from $71.2 billion by 2022, a CAGR of 20.7% from 2016 to 2022. The ensuing hyperscale heat wave will provide countless new opportunities along with complex energy management and infrastructure challenges, are you prepared for the what lies ahead?

Fred Moore began a 21-year career with StorageTek as the company’s first systems engineer and concluded as Corporate Vice President of Strategic Planning and Marketing. In 1998, Fred founded Horison Information Strategies in Boulder, Colorado, a data storage industry firm specializing in executive briefings, marketing strategy, identifying emerging technology trends for end-users, storage suppliers, and startup ventures. A sought-after motivator and frequent keynote speaker at storage conferences and IT events worldwide, he currently serves on a few select boards in the storage networking industry. www.horison.com