



THE CARBON EMISSIONS OF SERVER COMPUTING FOR SMALL- TO MEDIUM-SIZED ORGANIZATIONS

A Performance Study of
On-Premise vs. The Cloud

October 2012



Key Findings	02
Background	03
Methodology and Model	05
Analysis and Findings	07
The Key Variables and Their Relative Impacts	08
PUE	09
Server Utilization and Virtualization	09
Carbon Emission Factors of Electricity Source	10
Considerations for SMOs.....	11
The Relative Performance of Applications	12
Appendix and Assumptions	13
Key Input Parameter Definitions	15
Footnotes	16

Key Findings

As more Information Technology (IT) managers move or consider moving some of their IT applications to the cloud, understanding the environmental impacts of that choice is becoming increasingly important.

While cloud computing providers are touting the environmental benefits of their platforms, is cloud computing always greener than on-premise computing? Are there scenarios when on-premise computing may have a lower environmental impact than cloud computing?

NRDC and WSP Environment & Energy, LLC (WSP) launched this study to uncover the key factors that determine how on-premise and cloud computing compare from a carbon emissions and an energy efficiency perspective. The study compares the estimated efficiency of four commonly used small-to-medium-sized organizations (SMO) software applications against five computing deployment scenarios.

The key variables in the analysis include the following:

- Power Usage Effectiveness (PUE) of the server room or data center which hosts the servers running the applications;
- Hardware utilization factor, i.e., how much of the server's maximum processing capacity is effectively utilized; and
- The carbon emission factor (kg CO₂e per kWh) of the electricity used to power the server room or data center.

The results confirm claims that running a business application in the cloud is generally more energy and carbon efficient than running it on-premise because cloud computing offers greater diversity and can serve more customers at the same time, achieving better economies of scale than small and medium organizations.

However, this study demonstrates that this is not always the case and exceptions exist. Many factors contribute to the energy- and carbon-efficiency of both on-premise server rooms and cloud services. The most efficient server rooms can have significantly lower environmental impacts than the least efficient cloud services. Therefore, it is essential for IT managers to consider all the key variables when making application delivery choices based on environmental criteria.

Additional key findings include:

- The single largest factor to increase the efficiency of typical server rooms is higher server utilization, for example through virtualization. Sharing servers across applications, and across customers in the case of cloud computing, can greatly increase average server utilization.
- The second most important variable is the carbon emissions factor of the electricity powering the servers. Two identically sized and designed data

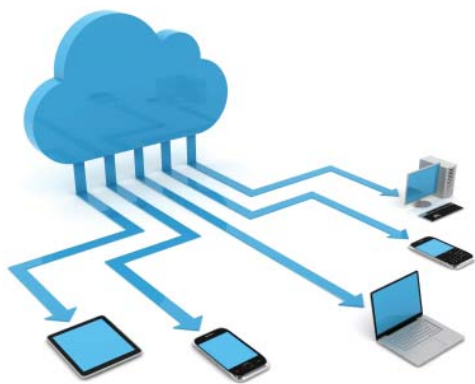
centers using power from high-carbon sources such as coal, or from lower-carbon sources such as renewable energy, will have a very different carbon footprint (varying by a factor of nearly four depending on which region in the U.S. they are located in).

- The third most critical variable is the efficiency of the server room infrastructure, including cooling and power distribution, as measured by the PUE metric. While PUE remains a key opportunity for efficiency improvements in server rooms, the potential emissions savings from improving PUE may be less than what can be achieved from increasing server utilization or using cleaner electricity.
- Upgrading server equipment to newer models is another way to reduce overall energy consumption, given that computing efficiency reportedly is doubling every 1.5 years.¹ Replacing or "refreshing" outdated equipment saves energy by taking advantage of higher efficiency and lower idle power consumption in the newest equipment. However, the promise of more efficient hardware is not fully realized if server utilization levels are not increased as well. For example, an application using just 5% of an older server may run at 1% utilization on a newer one, not fully using the increased performance capabilities of the new hardware.
- A private cloud, which consolidates servers and applications accessed across a company's intranet, offers similar benefits and limitations to a public cloud. The main difference is potentially lower server utilization levels due to smaller economies of scale and diversity of users.
- Off-premise "colocation" is also a popular alternative to an on-premise server room, and can provide more efficient cooling and power distribution. However, if the servers are run at low utilization levels and/or powered by dirty electricity, colocation is only marginally better than an on-premise, non-virtualized server room when it comes to carbon emissions.
- All four of the applications investigated present similar results, indicating that application type matters less than the three key variables identified.



Background

The cloud computing model is growing rapidly, both in the types of applications available and the range of users interested in leveraging its flexibility.



Accordingly, data center electricity consumption continues to grow and now represents roughly 2% of total electricity consumption in the US.² Recently a number of large cloud computing providers such as Google®, Microsoft®, and Salesforce.com have demonstrated the clear energy efficiency benefits of their services, but the question of whether a typical cloud platform is significantly more energy- and carbon-efficient than an equivalent on-premise deployment remains open. More specifically, what key factors make cloud computing potentially more or less energy- and carbon-efficient than on-premise computing?

Using a range of publicly available and best practice sources, NRDC and WSP completed a study to uncover the key factors that determine how on-premise and cloud computing compare from a carbon emissions perspective. The analysis focuses on the SMO: while much media attention has been dedicated to the giant internet companies and their efficient server farms, half of the servers in the United States still reside in smaller server rooms and closets which are typically managed less efficiently than large data centers.³ Furthermore, the scale of large cloud computing service providers already gives them a strong incentive to optimize energy management as energy represents a significant component of their operating expense – this is not necessarily the case with smaller cloud providers or a SMO's on-premise solution. By investigating the drivers that affect the carbon efficiency of a computing scenario, NRDC and WSP aim to enable IT managers to more easily compare the overall carbon emissions performance of some common deployment scenarios and be better equipped to integrate aspects of sustainability into their computing deployment and procurement decision-making processes.

Definitions

In order to structure the analysis, it is important to define the different computing scenarios that are modeled as well as the application types.

Deployment Types

On-Premise Not Virtualized (OPNV):

The SMO hosts their own servers for their business applications on-site in an IT closet or server room. The business owns these servers and the software applications. The servers are often underutilized because they are provisioned to manage the highest levels of processing and for the purposes of the study; one server is allocated to run a single application and the servers only serve the SMO. The servers tend to be older and less efficient, and to be housed in facilities with sub-optimal air flow and cooling.

Colocation/ Outsourcing: In some cases the SMO may choose to outsource their servers to a colocation data center; where although the infrastructure might belong to a third party, the fundamentals of how the computing works to serve the customer are the same as OPNV. In this study, the scope of the colocation scenario is limited to “unmanaged” colocation where the customer retains ownership of the equipment and responsibility for its administration. It does not include “managed” colocation which can be seen as a case of private cloud deployment.

On-Premise with Virtualization (OPV): This scenario assumes the same infrastructure efficiency as OPNV, except that the SMO has employed virtualization to reduce the number of physical servers and the corresponding server power consumption.

Cloud Computing

Cloud computing enables the SMO access to computing and data storage capabilities without requiring they invest in their own infrastructure or train new personnel. It delivers computing services through any virtual server which is available to support a workload, rather than being limited to dedicated servers for each workload. The energy efficiency associated with cloud computing is usually attributed to the fact that the servers can potentially be more efficiently managed by a third party than they are on-premise and achieve an overall higher utilization than on-premise computing by virtue of the shared workloads. From a carbon accounting perspective, the customer must only consider their fraction of server usage in the cloud, while if they own and operate their own hardware they must take into account the entire carbon emissions associated with that server.

Background

Private Cloud vs. Public Cloud

The physical setup of “the cloud” may encompass a broad range of scenarios; the servers might be in a data center that serves multiple customers who share the same hardware, or the servers could be hosted on-site behind an organization’s firewall and work together to support multiple business units in the same organization. In this study the difference between private cloud and public cloud is defined as whether or not the cloud infrastructure serves multiple organizations or a single organization, not where the servers are physically located.

Private Cloud: The term “private cloud” describes a cloud environment which serves a single organization. In some cases an SMO may consolidate their servers and applications into an organization-wide system which is maintained behind a firewall and can be accessed across an intranet. These servers may have a diversity of applications on them and serve multiple business units, but still only serve one organization. An SMO may also purchase their own portion of a public cloud; providers such as Amazon® or Rackspace® can provision a private, isolated section of a cloud platform to launch resources in a virtual network which supports a single customer.

Public Cloud: In this case the SMO purchases a desired service from a cloud provider, which can be accessed from anywhere via the internet. The cloud provider may deliver their services through their own data centers, or through outsourced servers in colocation facilities or other public cloud facility. The hardware usage is shared by multiple customers and the individual customer only has to pay for the capacity they need.⁴

Examples of public cloud computing services may include:

- Google® Apps or other office applications for their business;
- Customer Relationship Management (CRM) web-based service such as Salesforce.com to manage its contacts and sales leads;
- Cloud-based business administration applications such as accounting, expense management, employee travel management, and meeting scheduling (e.g., Doodle); and

- Specialized business applications delivered as cloud services to a small number of customers, for example a membership management application for a large non-profit organization.

Application Types

There are many factors which influence an IT manager’s decision in how to deploy an application, including but not limited to, latency tolerance, scalability, growth of the organization, security, and access to hardware. While it is generally acknowledged that “transient” and “elastic” apps – those with frequent provisioning, cloning, and need for adjustment over time – may be more readily transferred to the cloud,⁵ NRDC and WSP selected the applications below based upon their ability to be easily deployed in either an on-premise or cloud environment, and their relevance for SMOs.

- *Office Productivity Apps:* Email is perhaps the most ubiquitous of business applications, accessed from desktops, laptops and mobile devices. It has a high data storage requirement because of file attachments and a high user access requirement, but given the variability across user groups, is most often licensed and managed on a per user basis.
- *Content Management & Database Apps:* CRM software and Content Management Software (CMS) are widely used application types that rely upon back-end databases to store and retrieve information. Similarly, website hosting and website development

services use database layers to store content, and in the case of blogging services like WordPress, posts (articles), web pages, and visitors’ comments. Each time the database is accessed or a button in a website is clicked, a “transaction” of information is processed; augmenting data to existing fields or adding new ones. These types of applications usually have low data storage requirements because large files are not typically uploaded or associated with a particular database, but have a high user access profile similar to email.

- *Business Administration Apps:* This type of application is used to manage information to support the backbone of an organization, no matter the industry. Human resources software, accounting software, and financial management software are common examples. These applications are categorized and assessed for performance based upon the number of transactions that are processed by the application.
- *Utility Apps:* File storage and file sharing software may be as critical as email, but these types of applications are not necessarily accessed as frequently. File storage and sharing software typically have a much higher storage requirement depending on the nature of the files (for instance, an engineering firm might store large Autocad® or Revit® files, while a publishing organization would have less intensive word processing documents) and are, therefore, deployed based upon the file storage and sharing requirements.

Table I –
Characteristics of the Application
Types Selected for the Study

Application Type	Metric	Characteristics
Office Productivity (Email)	kg CO ₂ e per user	Typically high data storage requirements and high user access
Content Management & Database (CRM & Web)	kg CO ₂ e per transaction	Typically low data storage requirements and high user access
Business Administration (Finance & Accounting)	kg CO ₂ e per user	Typically high data storage requirements and high user access
Utility (File Storage & Sharing)	kg CO ₂ e per Gigabyte stored	Typically high data storage requirements and low user access

Methodology and Model

NRDC and WSP developed a model that considers four application types relevant to the SMO against five potential deployment scenarios:

- an on-premise server room, not using virtualization;
- servers hosted externally in a colocation deployment, not using virtualization;
- an on-premise server room, using virtualization;
- a private cloud; and
- a public cloud.

Leveraging previous studies, a quantitative analysis of the energy efficiency and carbon emissions impact by deployment and application type was developed to assess the potential impact of businesses shifting their computing requirements to a cloud service.

Throughout the modeling and analysis process, NRDC and WSP used publicly available data gathered through primary and secondary research and engaged with industry experts to validate key data inputs, assumptions, and directional findings. The methodology used to develop this independent model is aligned to the assessment methodology developed by the Global e-Sustainability Initiative (GeSI)⁶ and follows the process being developed by the Greenhouse Gas (GHG) Protocol's forthcoming standard for the Information

Communications & Technology (ICT) Sector Guidance for Cloud Computing and Data Center Services.

The analysis represents directional performance trends and ranges that organizations can expect to realize based upon their given environment and potential constraints. The results also specify key performance indicators per application type, which are being recommended as the suggested metrics to manage by the GHG Protocol.

Key Input Parameters and Variables

The model uses the following key inputs and variables (Table 2) to assess performance under a variety of deployment scenarios (Table 3) for a given application type. Data sources for these inputs, definitions, and model assumptions are presented in the appendix.

Table 2 –
Key Input Parameters and Variables
for the Five Deployment Scenarios

	PUE (IT/ Total DC)	Server Utilization Rate (CPU)	Server Refresh Period (yrs)	Virtualization Ratio	Grid Emission Factor (kg CO ₂ e/ kWh)
On-Premise Not Virtualized					
Worst Case	3.0	5%	5	n/a	0.819
Average	2.0	10%	3	n/a	0.554
Best Practice	1.5	25%	2	n/a	0.268
Colocation Not Virtualized					
Worst Case	2.5	5%	5	n/a	0.819
Average	1.8	10%	3	n/a	0.554
Best Practice	1.3	25%	2	n/a	0.268
On-Premise with Virtualization					
Worst Case	3.0	6%	5	3 1	0.819
Average	2.0	30%	3	5 1	0.554
Best Practice	1.5	60%	2	10 1	0.268
Private Cloud					
Worst Case	2.5	7%	5	3 1	0.819
Average	1.8	30%	3	5 1	0.554
Best Practice	1.3	60%	2	10 1	0.268
Public Cloud					
Worst Case	2.0	7%	3	5 1	0.819
Average	1.5	40%	2	8 1	0.554
Best Practice	1.1	70%	1	12 1	0.268

Table 3 –
Description of Typical Scenarios Corresponding
to the Key Input and Variables in Table 2

	On-Premise Not Virtualized		Colocation/Outsourcing		On-Premise Virtualized		Private Cloud		Public Cloud	
	Worst Case	Best Case	Worst Case	Best Case	Worst Case	Best Case	Worst Case	Best Case	Worst Case	Best Case
Design	Individual servers running single applications hosted on-site.		Individual servers running single applications hosted in an off-site data center (same computing configuration as OPNV).		Virtualization has been employed to decrease the number of physical servers required to run an application.		Cloud service is provided to a single customer but hosted in inefficiently managed data centers, or a low-utilized on-site model.	Cloud service is provided to a single customer and hosted in efficiently managed data centers, or well-managed on-site infrastructure.	A large-scale cloud service providing computing services to high numbers of customers; servers housed in large, efficiently-managed data centers.	
PUE (Cooling and Power Distribution Equipment)	The servers are not cooled effectively; there may be cooling which is oversized for the load, with poor air flow, constant speed fans, and a cooling set-point set too low.	The servers are cooled effectively without wasting energy.	The servers are not cooled effectively; there may be cooling which is oversized for the load, with poor air flow, constant speed fans, and a cooling set-point set too low.	The servers are cooled effectively without wasting energy. The PUE may be better than the best case OPNV.	Some servers have been eliminated and the remaining servers have higher utilization. However, the degree of virtualization is less than the best case scenario.	Some servers have been eliminated and the remaining servers have higher utilization than the OPNV or Colocation example.	The servers are not cooled effectively; there may be cooling which is oversized for the load, with poor air flow, constant speed fans, and a cooling set-point set too low.	Servers are cooled effectively in the cloud provider's data center.	The servers are not cooled effectively; there may be cooling which is oversized for the load, with poor air flow, constant speed fans, and a cooling set-point set too low.	Servers are cooled effectively in the cloud provider's data center.
Utilization	Low	Low	Low	Low	Low	High	Low	High	Low	High
Hardware	Outdated	Efficient	Outdated	Efficient	Outdated	Efficient	Outdated	Efficient; custom built, or top of the line commodity	Outdated	Efficient; custom built, or top of the line commodity

Power Usage Effectiveness (PUE): PUE has become the preferred energy efficiency metric used by the industry to measure data center infrastructure efficiency relative to the electrical load of its IT equipment. It is commonly used as an overall indicator of performance for energy consumption. As companies strive to reduce their PUE to the theoretical limit of 1.0, it forces a focus on the efficiency of the data center infrastructure.

Server Utilization Rate: As the engines of the data center, servers operate in either an idle or an active mode, but even in the idle mode they are consuming a meaningful amount of electricity. As average server utilization rates hover in the 5-15% range,⁷ there is a substantial amount of wasted energy in traditional operation. The underutilization of IT assets also takes an additional toll because IT manufacturing is a resource-intensive process that uses scarce materials which are damaging to the environment when extracted or processed.

Server Refresh Rate: Recent research by Jonathan Koomey suggests that computing efficiency is doubling every 1.5 years, so

the rate at which equipment is replaced can have an impact on overall energy consumption.⁸ Most SMOs and even many data center operators are using outdated equipment which, if replaced with newer equipment, could result in reduced energy consumption. Although increasing the refresh rate contributes to the total carbon footprint of the data center due to the imbedded carbon emissions impact associated with the hardware, the emissions associated with the use phase of the server typically far outweigh the imbedded footprint for servers,⁹ and the reduction in use phase emissions and the increase in imbedded emissions can be expected to be in a similar proportion.

Virtualization: Virtualization and other related techniques can address the inefficiencies of low server utilization by allowing multiple applications to share a single server, thereby reducing the number of physical servers required to run workloads. Fewer servers, even when running at higher utilization levels, consume less electricity, which also reduces the need for cooling. Virtualization, if deployed effectively, is one of the key drivers of efficiency within data centers.

Emission factor of server room/data center electricity source: While the other variables identified focus on the energy efficiency of a data center, the carbon emissions factor of the purchased electricity required to run a data center is one of the ultimate drivers of total carbon impact, and is dictated by the location of the data center and/or the electricity procurement decisions in markets where customers can choose between multiple providers. A carbon emission factor of purchased electricity is required to calculate the carbon footprint of a particular data center. Two identically sized and designed data centers using electricity produced by different generation means, such as renewables, hydropower, natural gas or coal, will have a potentially very different carbon footprint (factors can vary by a factor of nearly four depending on the region in the US).

While an on-premise deployment does not typically have the flexibility to move to a lower carbon intensive region, an IT manager deciding between two cloud service providers can use this variable as a criterion for selection, as can cloud service providers in selecting the location of their individual data centers.

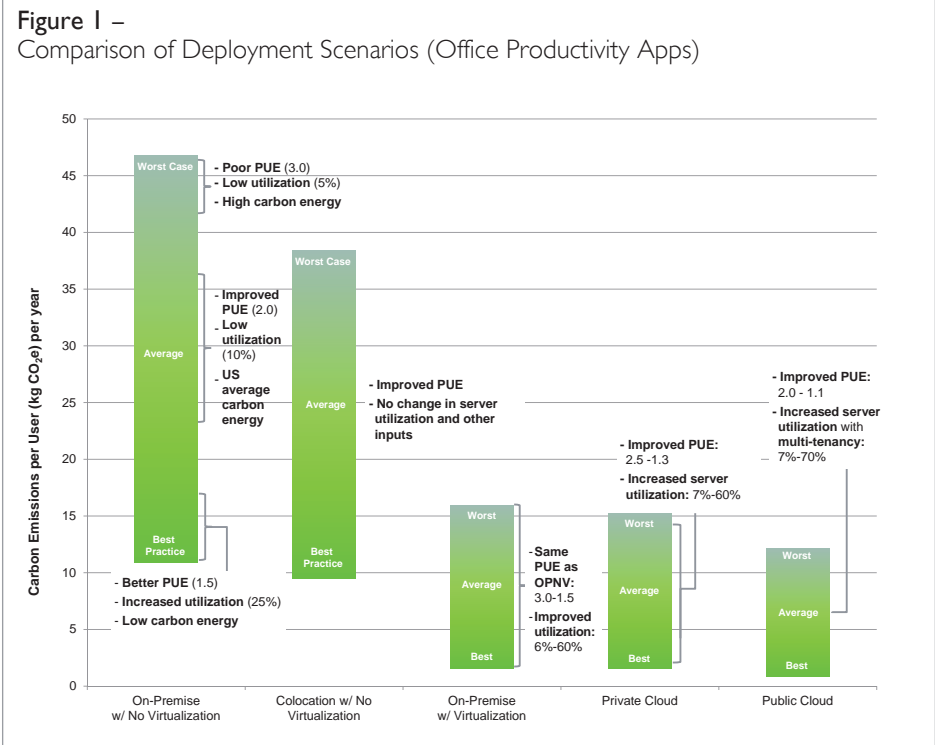
Analysis and Findings

A Range of Performance

The findings of the study indicate the impacts of the previously described variables on computing performance. NRDC and WSP's model focused on those variables which contribute to the largest impact and include:

- Effective PUE of the server room or data centers which are hosting the servers;
- Effective utilization of the servers which are running a given application and corresponding number of physical servers; and
- Carbon emission factor of purchased electricity based on the location of the server room or data centers.

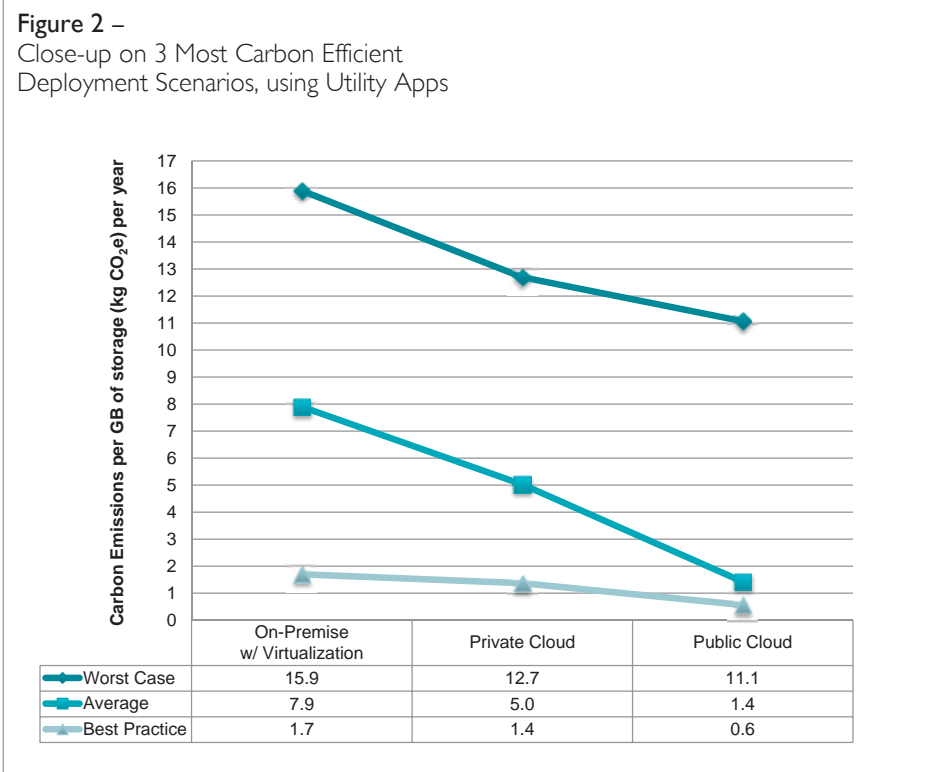
Using publicly available data for the key inputs of worst case, average, and best practice, Figure 1 highlights the range of performance realized in a given deployment scenario, and the relative impacts of changing variables.



The OPNV scenario is on average responsible for the most carbon emissions per user, closely followed by colocation which features improved PUE, but similar utilization and electricity source. The other three scenarios all use virtualization, resulting in significantly lower emissions than non-virtualized deployments. Public cloud is on average the most efficient type of deployment, although worst-case public clouds can be responsible for much higher carbon emissions per user than best case OPV and private cloud, depending on the degree of implementation of best-practices regarding utilization levels, PUE and clean electricity procurement.

The overlap across each deployment type indicates that at the margin, a deployment type that may be better on average is not necessarily so; for example, a well-managed on-premise server room with no virtualization may be equivalent to a poorly managed cloud service.

Figure 2 focuses on the three most carbon efficient deployment types only and shows the average values from worst case to best case for all variables. This indicates that the average carbon efficiency gains realized by the public cloud are significant.



The Key Variables and Their Relative Impacts

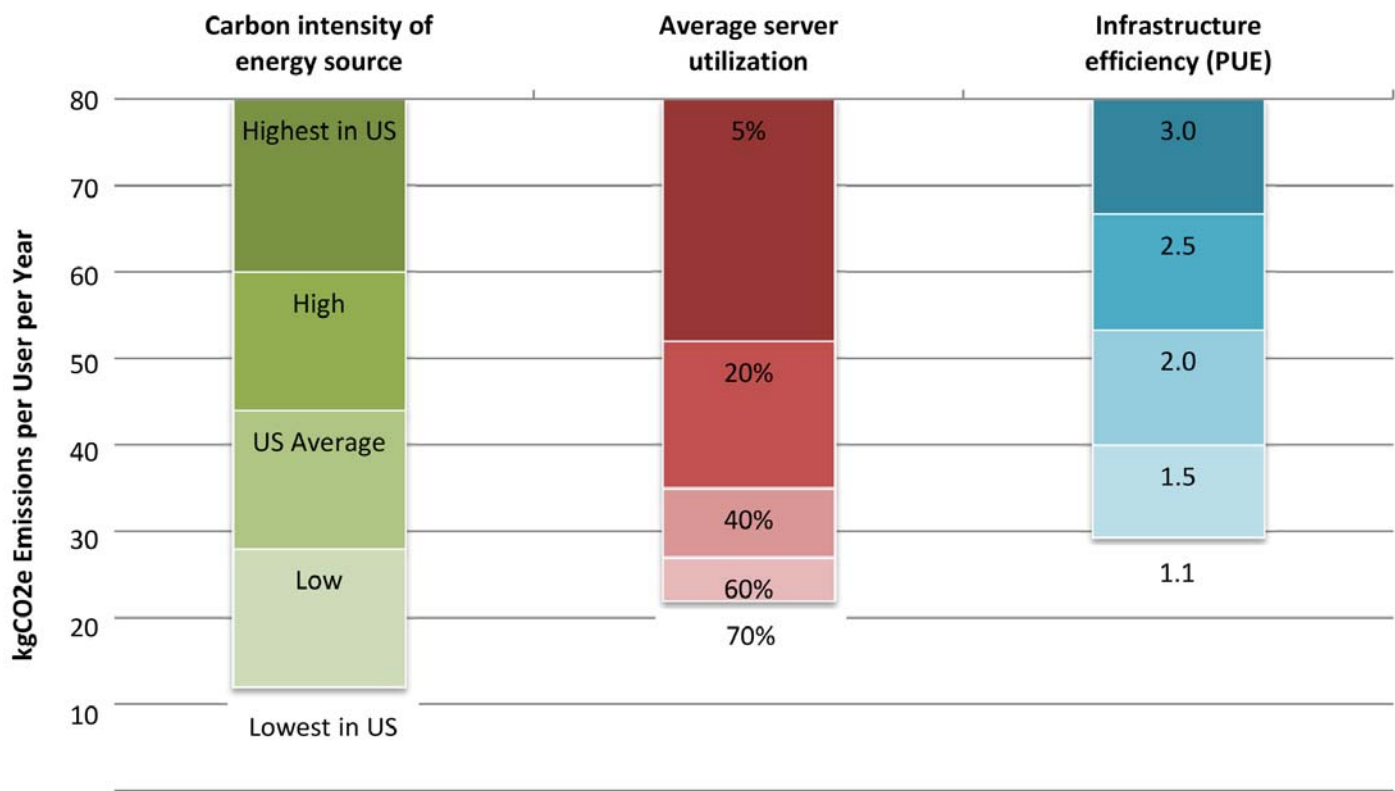
A discussion around each variable can help IT managers better understand how and why the variables selected can impact performance so that they can prioritize their efforts at improving their organization's IT carbon efficiency. Figure 3 shows the change in carbon emissions when starting from a worst case OPNV scenario and changing each key variable in isolation.

When comparing the carbon efficiency between a virtualized and a non-virtualized environment, the OPV scenario realizes improved performance as servers are consolidated and the remaining servers are more highly utilized. Since it does not take significantly more energy to run a fully-utilized server than a partially loaded one, the overall number of servers is diminished.

Therefore, if a SMO wants to lower its overall carbon footprint, the most efficient strategy may be to virtualize its servers and IT platform. Or as the results reinforce and other studies have shown,¹⁰ moving to either a private cloud or a public cloud - even if the average carbon emission factor of that cloud's power source is high - would be more efficient than focusing only on reducing the PUE of the on-premise server room. Figure 3 illustrates this difference in efficiency between the three primary strategies.

While a focus on utilization provides a quick reduction in emissions, those savings are diminished at high rates.

Figure 3 –
Relative Impact of Key Data Center Variables*



PUE

PUE is a critical metric to consider in the actual delivery of computing services. According to the Uptime Institute, the average PUE across the United States data center pool is 1.8, slightly lower than the Environmental Protection Agency's (EPA's) initial survey showing 1.91 in 2009.¹¹ This means that although data centers are slowly becoming more efficiently cooled and older facilities are taken offline, the stock of legacy data centers will continue to have an impact on national data center efficiency, even as data centers built by Google®, Microsoft®, eBay®, and Facebook® are lauded for their impressive efficiencies.¹²

Small SMO server rooms are often managed inefficiently and suffer from classic economies of scale issues that cloud service providers are directly incentivized to manage. Large cloud providers can rationalize investments in efficiency improvements because they can spread the costs over a larger server base and they usually own and operate the data centers – a huge overhead cost. Most SMOs are unable to achieve the low PUEs that cloud server providers are realizing in their best-in-class data centers. Regardless, the proportional contribution that PUE can have on the overall carbon emissions footprint is less than that of focusing on increasing server utilization through virtualization or the location of data centers.

Server Utilization and Virtualization

It is estimated that, on average, servers running a single application may be about 5% to 15% utilized,¹³ meaning that for most of the time the majority of the server is drawing a substantial amount of power, but performing little or no useful computing. Even when the server is idle or underutilized it may consume between 60% and 90% of its full power draw.¹⁴ This inefficiency causes tremendous energy waste, and according to recent research from Microsoft®, for every 100 watts of power consumed in a typical data center, only three watts are used for useful computing. In other words, a typical data center consumes more than 30 times the energy that it uses to perform computations.¹⁵ Virtualization and other related techniques offer a solution to address the primary challenge of hardware under-utilization and, ultimately, the inefficiencies inherent in traditional on-premise computing. Because virtualization can result in a 10 to 1 server consolidation ratio to run identical workloads,¹⁶ its implementation can realize significant energy reductions. Virtualization effectively increases the energy efficiency of the system because that energy is actually going to computing rather than powering idle or near-idle servers.

Public cloud providers can realize the potential of very high utilization rates because workloads may be transferred between servers and data centers as the hardware is accessed by multiple users at the same time, smoothing out load peaks and troughs. In this scenario, the number of servers required becomes less relevant as thousands of transactions are processed simultaneously using the same hardware.¹⁷ However, the true performance and utilization of the cloud must be determined on a case-by-case basis; despite the opportunity for this increased efficiency, various studies have indicated that in actuality the utilization of cloud environments may be well less than perceived, with measured examples indicating utilization rates as low as 7% on average.¹⁸ Nonetheless, it's clear that in better managed public cloud environments, the aspects of diversity and aggregation where thousands of unique users in different geographies help to spread the computing loads across a provider's platform, allowing for increased utilization and lower electricity consumption.



Carbon Emission Factors of Electricity Source

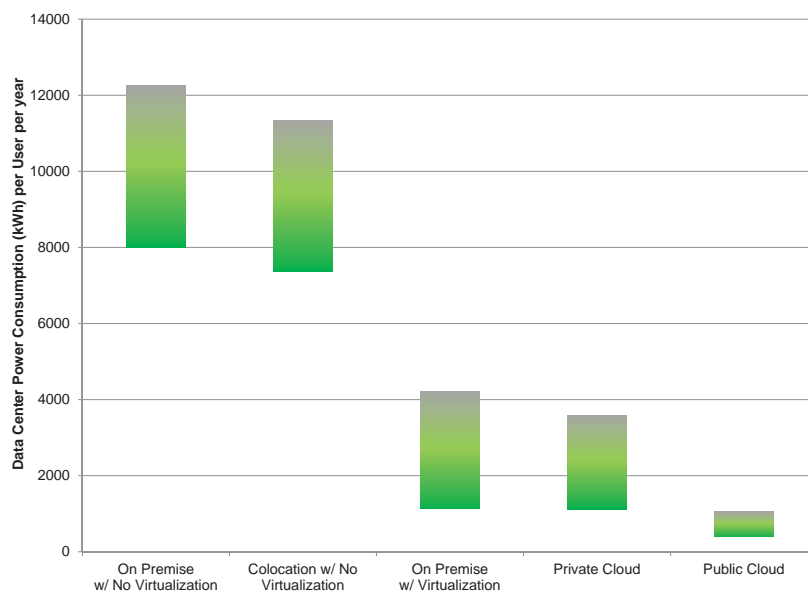
Despite server consolidation and virtualization trends, the physical installed server base continues to expand¹⁹ and ultimately, no matter how efficient a data center is, it still requires electricity to operate. That energy source is associated with a carbon emissions factor based upon the type of fuel used. Consequently, the power source of electricity can significantly alter the overall carbon efficiency of a data center. Corporate reporting practices dictated by the GHG Protocol require that an organization report carbon emissions associated with their operations by using EPA's eGRID emission factors, which are calculated in sub-regions across the United States, rather than using emission factors from their specific local utility. This is in some cases a burden for companies that invest in renewable energy directly to power their data centers or have access to utilities that can offer a high percentage of renewables even though the regional average is higher.

While an SMO is not likely to be able to move their location solely because of carbon emission factors, cloud providers are increasingly considering this variable amongst the other key factors such as the adequate availability of reliable and affordable electricity, telecommunications infrastructure and accessibility, tax incentives, climate, and proximity to end users.

Even though it may be difficult to pin down a real emissions average for any cloud provider's architecture unless they have published this data transparently and are actively engaged in tracking it, the model suggests that no matter how efficient a cloud provider's IT architecture or how effectively the hardware is utilized and cooled, the location of the data center and the composition of the fuel mix will make a significant impact on the overall carbon emissions of a scenario. This is best illustrated by removing the carbon emissions factor as a variable in Figure 4, which considers only energy efficiency performance of the four deployment scenarios. The results indicate that the public cloud, in either a worst or best case, significantly outperforms the other deployments types.

However, when the carbon emissions factor of purchased electricity is layered over these results, the range of impact and performance is considerably increased (as demonstrated in Figure 1).

Figure 4 –
Comparison of Deployment Scenarios for
Energy Efficiency: Office Productivity Apps



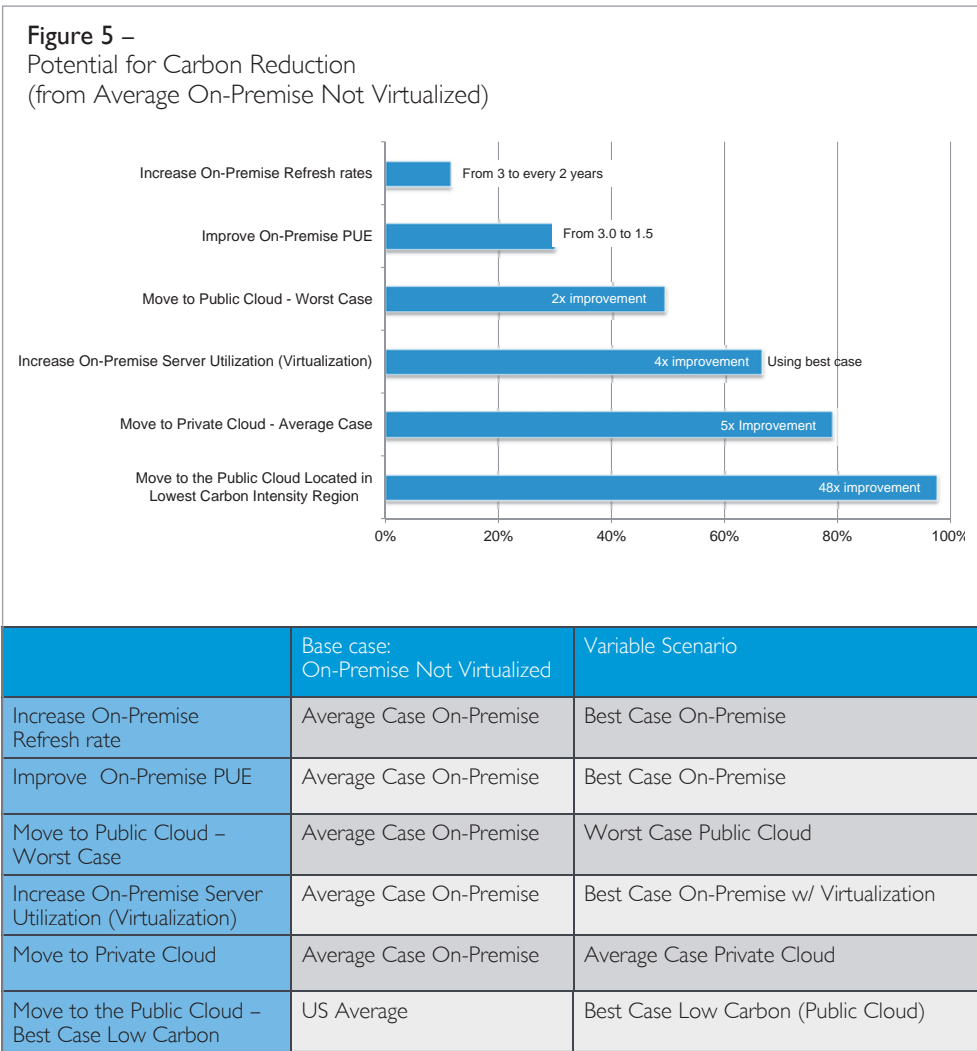
Considerations for SMOs

If your server room is a typical OPNV, with average values for PUE, utilization and carbon-intensity, Figure 5 compares the potential for savings of common alternatives.

While increasing refresh rates and improving PUE do offer savings between 10% and 30%, the percentage impact and order of magnitude that a focus on increasing server utilization and virtualization or a move to a private or public cloud deployment, is a considerable advantage for a typical on-premise deployment.

Finally, it's worth comparing the carbon efficiency offered by the public cloud but with the variable of the carbon emission grid factor. In a worst case scenario, where public cloud data centers may be located in areas with high carbon emission factors, such as some Midwest US and Mid-Atlantic states, the cloud remains more carbon-efficient by a factor of 2 or a reduction in emissions of roughly 50%. But when a data center provider's public cloud is located in areas of low carbon intensity, such as in the Pacific Northwest, the savings increase dramatically to nearly a 48 times improvement.

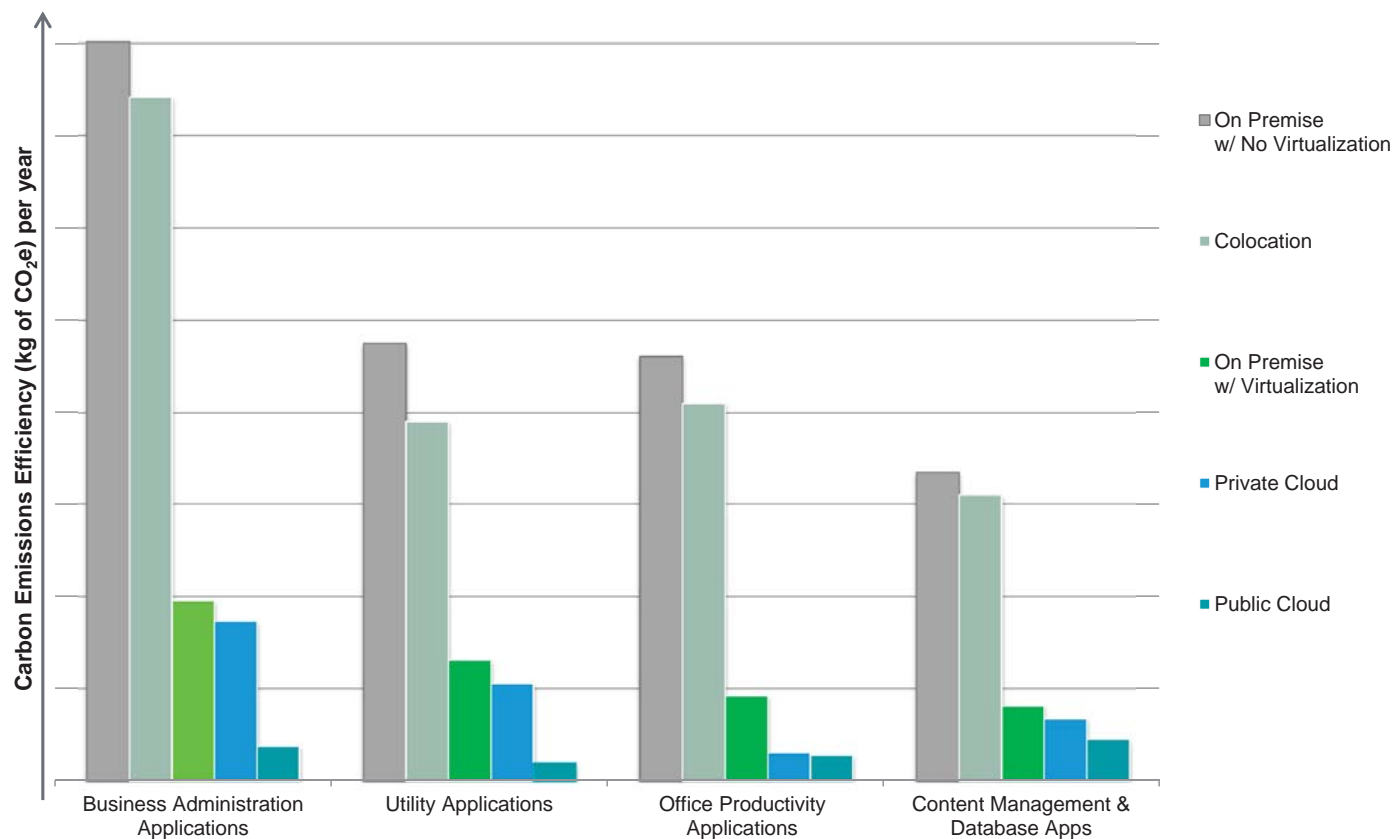
Thus, while cloud computing can generally be shown to be a more energy and carbon efficient solution when compared to on-premise computing, all of the key variables identified in this paper contribute widely to the overall carbon efficiency of a given computing scenario. Each variable must be considered when making application delivery choices and vendor selection with cloud providers if environmental and/or sustainability criteria are important performance indicators.



The Relative Performance of Applications

Figure 6 models the four application types based on the “Average” inputs for the previously described variables. It is clear that while carbon emissions vary per application type, all applications behave similarly when compared across the deployment types; in other words, they all demonstrate performance gains by implementing virtualization, or by moving to a private or public cloud. As the physical limitations of hosting applications in a variety of ways disappear, the differences between the server requirements for different applications begin to look immaterial.

Figure 6 –
Carbon Emissions Ratio per Application
Type & Deployment Scenario
(using average values for other variables)



Appendix and Assumptions

Model Overview

- The study quantified four application types against five deployment scenarios by dividing the total energy consumption and resulting carbon footprint against the number of active users, transactions, or gigabytes of delivered data (as relevant) for a given application.
- The model was independently developed based on ISO 14044 guidelines for Life Cycle Assessment, BSI PAS 2050 Specifications for the Assessment of GHG Emissions of Goods and Services, and the WRI/WBCSD GHG Product & Supply Chain Protocol, including the forthcoming ICT Sector Supplement.
- The aggregated results in this report have been calculated based on a scope limited to the United States. Organizations operating in different regions will be subject to different carbon emission factors and specific data center utilization rates that could affect the findings of a similar study.
- Secondary data inputs were derived from a variety of data sources referenced throughout the paper.
- Time period considered: a one-year application use, licensing, or subscription agreement.
- GHG emissions ("carbon emissions") included are stated as carbon dioxide equivalent (CO₂e) emissions and take into account the six primary GHG gases including carbon dioxide (CO₂), sulphur hexafluoride (SF₆), methane (CH₄), nitrous oxide (N₂O), hydrofluorocarbons (HFCs), and perfluorocarbons (PFCs).

- The study includes the use phase of the product by the customer. While use is assumed to be the same rate for cloud and on-premise, the efficiency and energy consumption associated with the two scenarios are different.

Materials

- Primary materials included in the study consisted of servers and related network equipment used to host an application. Embodied emissions from physical hardware were estimated based on the weight and composition of each component.
- Embodied emissions from physical infrastructure included servers, but not facilities and other equipment.
- Emissions related to the material manufacture, assembly, and recovery of servers and networking equipment are based on a 3.5-year refresh rate for data center hardware and allocated according to a prescribed specific scenario refresh rate. Life Cycle Inventory of a server derived from three different published studies from IBM®, Dell®, and Apple®.

Process Energy for IT Infrastructure

- Estimated power consumption of servers is based on industry-standard figures provided by Hewlett Packard® and Dell®, and verified by experts using specific server configuration sizing calculations. A mixture of different vendors' systems was assessed, rather than any single server product.
- The model includes essential power for critical IT environment and utilizes publicly available data center-specific

PUE ratios covering worst to best practice PUE's based upon EPA, Green Grid, Microsoft®, Salesforce.com, and Google® published reports and Jonathan Koomey research.

- Appropriate carbon emissions factors were applied to the energy consumption of specific data center locations from the U.S.EPA's eGRID2012 Version 1.0, May 2012.
- A storage consumption and network usage efficiency ratio were also applied based upon primary data provided by salesforce.com, and referenced from secondary data from the EPA, Green Grid, and research by Jonathan Koomey, PhD.
- Research from the National Laboratory for Applied Network Research (NLANR) Project informed the path of data transfer (from a data center to a business customer).

Model Data Input Exclusions

- Energy consumed during software development.
- Tertiary suppliers and process materials which are not significant (i.e., do not constitute an input to 95% of the product or process).
- Refrigerants (except where used in primary production of raw inputs).
- Embedded energy of capital equipment, transportation vehicles, buildings, and their energy use not directly related to servers and associated equipment.
- Maintenance of capital equipment.

Appendix and Assumptions

Table 4 -
Model Data Inputs for Server Utilization²⁰

On-Premise Not Virtualized		
Worst Case	5%	The Power of Incentives on Data Center Efficiency. Corporate Eco-Forum. March 25, 2012.
Average	10%	The Business Value of Virtualization. Forrester Research. July 2009. Best Practices Guide for Energy-Efficient Data Center Design. National Renewable Energy Laboratory. February 2010.
Best Practice	25%	Cole, Arthur. Striking the Right Balance in Server Utilization. IT Business Edge. November 16, 2009.
Colocation – Data inputs and assumptions for Colocation are the same as OPNV with an improved PUE based on scale and 3rd party financial incentives to better manage cooling loads		
On-Premise with Virtualization		
Worst Case	6%	Kaplan, Forrest, Kindler. Revolutionizing Data Center Energy Efficiency. McKinsey & CO. July 2008.
Average	30%	Best Practices for Leading Edge Data Centers. Corporate Eco-Forum. May 22, 2012.
Best Practice	60%	How VMware Virtualization Right-sizes IT Infrastructure to Reduce Power Consumption. VMware. 2008.
Private Cloud		
Worst Case	7%	Miller, Rich. "Kundra: Fed Data Centers 7 Percent Utilized". Data Center Knowledge. April 9th 2010.
Average	30%	Best Practices for Leading Edge Data Centers. Corporate Eco-Forum. May 22, 2012.
Best Practice	60%	Koomey, Jonathan. 4 reasons why cloud computing is efficient. GigaOM. July 25, 2011.
Public Cloud		
Worst Case	7%	A Measurement Study of Server Utilization in Public Clouds. 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing. Huan Liu, Accenture Technology Labs. 978-0-7695-4612-4/11 © 2011 IEEE
Average	40%	Best Practices for Leading Edge Data Centers. Corporate Eco-Forum. May 22, 2012.
Best Practice	70%	Koomey, Jonathan. 4 reasons why cloud computing is efficient. GigaOM. July 25, 2011. How VMware Virtualization Right-sizes IT Infrastructure to Reduce Power Consumption. VMware. 2008.

Key Input Parameter Definitions

Server Count: Number of servers required to support a platform or the specific business applications, inclusive of business continuity and disaster recovery ($n+1$).

Users: The number of active users for the application.

Device Utilization: Computational load that a device (server, network, or storage array) is handling relative to the specified peak load.

Virtualization: Technology which allows numerous “virtual machines” to run on one physical server. These virtual machines can run different operating systems and multiple applications on the same physical hardware, which effectively increases the utilization of that server.

Power Consumption per Type of IT

Hardware: Calculated energy consumed by a server at a given rate of device utilization and estimated power for networking and storage equipment.

Power Usage Effectiveness: The total facility power divided by the power used to run the IT equipment. The lower the PUE (with a minimum of 1.0 possible), the more efficiently the data center is run. However, because the PUE is relative to the IT power, it makes no indication of how efficiently the power is used for the IT equipment itself.

Multi-tenant/Multi-tenancy: Multi-tenancy defines IT architectures that let multiple customers (tenants) share the same applications and/or compute resources with security, reliability, and consistent performance. It is through multi-tenant architectures that cloud services achieve high cost efficiencies, increased energy efficiency for a given load of work, and can deliver low costs because the infrastructure and services are shared by many customers.

Internet Transmission: Energy required for data transfer through access technologies and the internet backbone for cloud deployments.

Embedded Carbon Emissions for IT

Hardware: Life-cycle estimates of embedded carbon emissions on a per-server basis.

Emissions Factor of Server Room/Data

Center Electricity Source: The amount of carbon emitted to generate the electrical power consumed by the server room or data center (in kilograms of carbon emissions per kilowatt-hour). The figure dramatically varies between regions and is based primarily upon the mix of primary energy sources (coal, hydro, nuclear, wind, etc.).

Footnotes

- 1 Key Findings: Koomey, Jonathan G., Stephen Berard, Marla Sanchez, and Henry Wong. 2011. "Implications of Historical Trends in The Electrical Efficiency of Computing." *IEEE Annals of the History of Computing*. vol. 33, no. 3. July-September. pp. 46-54. <http://doi.ieeecomputersociety.org/10.1109/MAHC.2010.28>
- 2 Background: Koomey, Jonathan. 2011. *Growth in Data center electricity use 2005 to 2010*. Oakland, CA: Analytics Press. August 1. <http://www.analyticspress.com/datacenters.html>
- 3 Small Server Rooms, Big Energy Savings Opportunities and Barriers to Energy Efficiency on the Small Server Room Market. NRDC. February 2012. <http://www.nrdc.org/energy/saving-energy-in-server-rooms.asp>
- 4 The NIST Definition of Cloud Computing. Special Publication 800-145. September 2011. <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- 5 VMware and Cloud Computing. 2010. <http://www.vmware.com/files/pdf/vmware-cloud-computing-BR-EN.pdf>
- 6 Global e-Sustainability Initiative (GeSI). *An Assessment Methodology*. 2010.
- 7 Methodology and Model: US EPA. 2007. *Report to Congress on Server and Data Center Energy Efficiency*, Public Law 109-431. Prepared for the U.S. Environmental Protection Agency, ENERGY STAR Program, by Lawrence Berkeley National Laboratory. LBNL-363E. August 2. <http://www.energystar.gov/datacenters>
- 8 Koomey, Jonathan G., Stephen Berard, Marla Sanchez, and Henry Wong. 2011. "Implications of Historical Trends in The Electrical Efficiency of Computing." *IEEE Annals of the History of Computing*. vol. 33, no. 3. July-September. pp. 46-54. <http://doi.ieeecomputersociety.org/10.1109/MAHC.2010.28>
- 9 Uncertainty and Variability in Carbon Footprinting for Electronics: Case Study of an IBM Rack-mount Server
- 10 Analysis and Findings: Cloud Computing – The IT Solution for the 21st Century. Carbon Disclosure Project Study 2011. Produced by Verdantix.
- 11 Miller, Rich. The Average PUE is 1.8. Uptime Institute. May 2011. <http://www.datacenterknowledge.com/archives/2011/05/10/uptime-institute-the-average-pue-is-1-8/>
- 12 Details of Facebook® data centers, which may operate at a PUE of 1.07, can be found at <http://opencompute.org/projects/mechanical/>
- 13 McKinsey. *Revolutionizing IT*. <http://www.mckinsey.com/client-service/bto/pointofview/Revolutionizing.asp>
- 14 US EPA. 2007. *Report to Congress on Server and Data Center Energy Efficiency*. However, the model uses more conservative assumptions for the utilization rates based upon the type of deployment (from on-premise to public cloud). These rates are provided in Table 2 and with references in Table 4.
- 15 The IT Energy Efficiency Imperative. Microsoft®. 2011. and <http://www.triplepundit.com/2011/07/overbuilding-server-utilization-and-energy-productivity/>
- 16 Ref Forrester Report and VMware website.
- 17 To adequately represent the impacts of changing hardware utilization and virtualization and the benefits of going to a private cloud, the model adjusts the number of "effective" servers, because changing the utilization of the same number of servers will just result in a linear increase in energy consumption as the utilization increases. The model works on the premise that virtualization reduces the effective number of servers to do the same job, so as the server uses more energy at a higher utilization rate, the physical server energy consumption is reduced proportionally. This theory was tested and developed in partnership with Accenture for the white paper: *Salesforce.com and the Environment: Reducing Carbon Emissions in the Cloud*. WSP Environment & Energy, LLC. 2011.
- 18 Liu, Huan. *A Measurement Study of Server Utilization in Public Clouds*. Accenture Technology Labs. *IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*. 2011.
- 19 *Data Center Trends Report*. Information Week Analytics. 2010.
- 20 Because there is very little published data on server utilization with respect to specific deployment types, the analysis takes a conservative approach by using the available data that is publicly available. Therefore, the figures used herein should be treated as indicative of trends and averages rather than measured data.

About WSP Environment & Energy, LLC

WSP Environment & Energy, LLC is one of the world's leading global consultancies, delivering solutions to environmental, energy, and sustainability issues and offers a full-service consulting practice to a range of commercial and industrial clients, many of whom are Fortune 500 and FTSE 250 companies.

We help clients increase business performance through process improvement, risk mitigation, and design and implementation of sustainable business practices. WSP Environment & Energy, LLC is a part of WSP which, through its combination with GENIVAR, is one of the world's leading professional services firms. The firm has 14,500 employees, based in more than 300 offices, across 35 countries, on every continent. Visit on the web at www.wspenvironmental.com/sustain.

About Natural Resources Defense Council

The Natural Resources Defense Council (NRDC) is an international nonprofit environmental organization with more than 1.3 million members and online activists. Since 1970, our lawyers, scientists, and other environmental specialists have worked to protect the world's natural resources, public health, and the environment. NRDC has offices in New York City, Washington, D.C., Los Angeles, San Francisco, Chicago, Livingston, Montana, and Beijing. Visit us at www.nrdc.org and follow us on Twitter @NRDC.

The research performed to prepare this report was funded by a grant to NRDC from the U.S. Environmental Protection Agency. Research and analysis were conducted by Josh Whitney and Jessica Kennedy of WSP Environment & Energy, LLC, and the project was managed by Pierre Delforge of NRDC.

The views and findings expressed herein are solely those of the authors and do not necessarily reflect those of the EPA.

Acknowledgements

The authors would like to thank:
Rich Garrison,
Senior Principal, ATCE

Jonathan Koomey,
Consulting Professor,
Stanford University

Drew Kramer,
Vice President of
Technical Operations, VMware

Mark Monroe,
Executive Director,
The Green Grid

© 2012 WSP Environment & Energy, LLC.

UNITED
BY OUR
DIFFERENCE

