IJCTA, 9(23), 2016, pp. 113-118
© International Science Press

Environmental Benefits of Cloud Computing

Kapil More*

Abstract : As computing needs increase worldwide, the electricity consumption and recycling efforts for a data center will keep on increasing. This will increase the environmental impact as it has upward impact on carbon footprint of a data center. Most of the data center resources remain grossly underutilized for various reasons. In this paper we will see how cloud computing helps in increasing resource utilization and thus helping reduce the carbon footprint of a traditional data center.

Keywords : Cloud computing, Environment, Green, Carbon, Virtualization, Utilization, Energy.

1. INTRODUCTION

With increasing number of computing worldwide, the environmental impact of these servers is becoming a concern. Traditional data centers are scattered across various companies, organizations which maintain their servers, storage, and other equipment's in them. There are two primary environmental load of a data center as below.

- Electricity consumption
- Electronic waste generation

Electricity is required to run servers and various network equipment's like routers, switches etc. Also a massive amount is required for cooling of the facility.

Electronic waste is generated when these equipment's are decommissioned.

According to a US Govt study, US data centers consumed about 70 billion kilowatt-hours of electricity in 2014, representing 2 percent of the country's total energy consumption. That's equivalent to the amount consumed by about 6.4 million average American homes that year

2. SCOPE FOR LOAD REDUCTION

In order to reduce electricity consumption and electronic waste, the number of devices has to be reduced. To do that utilization of existing devices should be increased. Currently if we study the device utilization we find that there is lot wastage in device utilization in the form of idle CPU time, free space in hard drives. Below are some of the statics of underutilization of resources.

A Study by Anthesis group and Stanford university researcher Jonathan Koomey & another by Uptime Institute reveal 30% of the servers in a data center are comatose *i.e.* zombie or not used even in once in past 6 months. There are approximately 75million servers in the world. Considering that 30% servers are comatose:

Energy consumed by US data centers (A)	70 billion kilowatt-hours
Idle servers ie zombies or comatose (B)	30%
Energy saving from shutting idle servers (C)	= 30% of A = 21billion kilowatt-hours

Here we are only talking of comatose servers *i.e.* once which are idle for more than 6 months.

* Persistent University Persistent Systems Ltd, Pune, India Email-Kapil_more@persistent.com

1. Underutilization of Servers

If we consider the idle CPU cycles the numbers are huge. Below are some of the reports done by various studies on CPU utilization.

1. A McKinsey study in 2008 pegging data-center utilization at roughly 6 %.
2. A Gartner report from 2012 putting industry wide utilization rate at 12 %.
3. An Accenture paper sampling a small number on Amazon EC2 machines finding 7% utilization over the course of a week

Even if we consider average 10% utilization, that means a huge 90% CPU remains unutilized. If we can reduce idle CPU cycles by even half in US, we can save 0.45 * 70 = 31 billion kilowatt-hours of electricity.

Another report by Stanford[11] points that data centers can slash their greenhouse emissions by 88% *"Of the potential 88 percent reduction in greenhouse gas emissions, though, IT device efficiency accounts for about 80 percent and facility energy management for only about 8 percent. Once those two areas are maximized, sourcing electricity from renewables like wind and solar power, plus green handling of retired equipment, can get a typical data center's emissions down 98 percent."* With this if we redo our statistics

Total energy used by US data centers (A)	70 billion kilowatt-hours
Idle servers that can be shut down (B)	30%
Savings because of shutting down idle servers (C)	= B % of A = 30% of 70 billion kwh = 21billlion kwh
Energy consumption after shutting idle servers (D = A-C)	= (70-21) billion kwh = 49 billion kwh
Unused server capacity (E)	90%
Considering 50% reduction in idle CPU capacity (F = 50% of E)	= 50% of 90 = 45%
Energy saving if idle CPU reduced by 50% (G = 45% of D)	= 45% of 49 billion kwh = 22 billion kwh
Total energy reduction = 100% idle servers shutdown +50% reduction in CPU idle time (H = C + G)	= 21billion kwh + 22 billion kwh = **43 billion kilowatt hours**
Total energy reduction in terms of H % of A	= H/A * 100 **= 60.42 % net reduction in energy consumption**

2. Disk Space Underutilization

Various reports peg the disk space unitization at 25-50%. Considering this 70% of disk space is underutilized. If this is even reduced by half we can save 35% disk. This will reduce the recycling effort and power consumption of the disks by 35%.

"*Data utilization rates were abysmal with data centers on average using just 25% to 30% of their hard disk capacity***"** says a computer world report[3].

3. REASON FOR UNDERUTILIZATION

Most resources are provisioned to handle peak load. Hence during non-peak time they remain underutilized. For *e.g.* all end-users accessed application servers have very low utilization during 10 hours of night. Typical rush hours for most applications are 3-4 hours a day. So the resources provisioned to handle peak load of rush hours

remain idle or underutilized during non-peak period. Below the internet traffic history of Walmart.com as given by Alexa.com and published in cloud provider Rackspace's white paper10 on moving infrastructure to cloud. It shows big difference in peak and non-peak load and also the fact that peak is reached only for a short period of time.

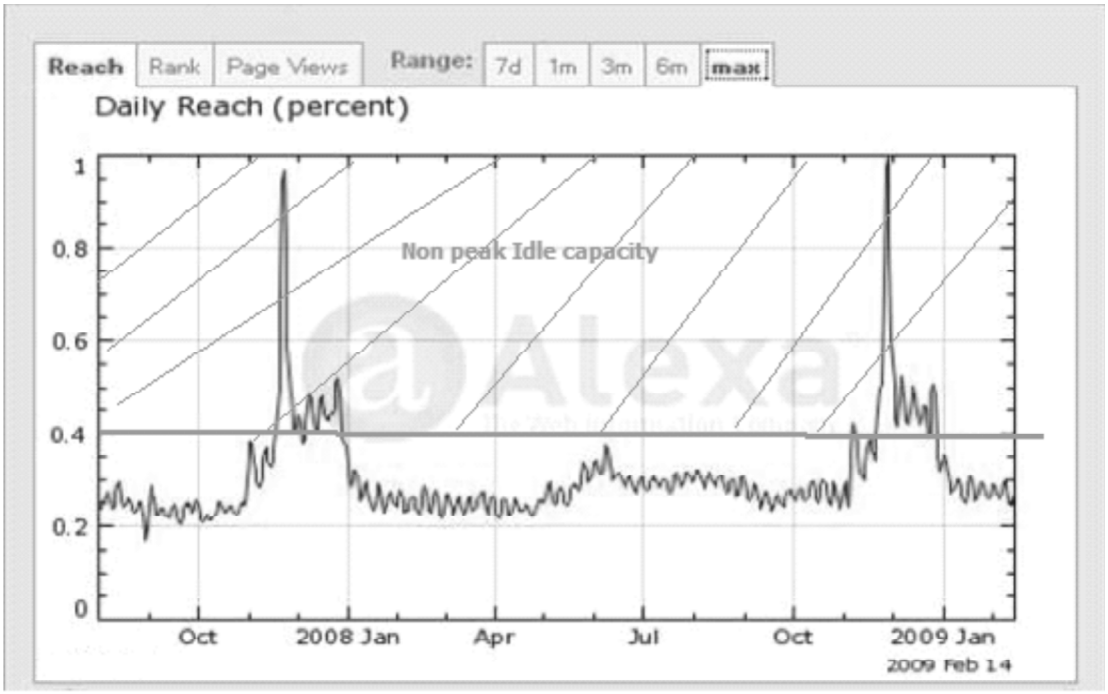

Fig. 1. Walmart Traffic history from web analytics site Alexa.com.

Similarly storage space is provisioned for peak load, **which doesn't happen during normal operations. And henever** peak storage is reached, immediately a new storage is provisioned which then has more unused capacity.

4. PROBLEMS FACED IN INCREASING UTILIZATION

Most companies find it difficult to reduce this wastage due to following:

- Lack of easy, efficient & cost affordable tools, software.
- Lack of awareness and interest in micromanagement,
- Effort vs benefit mismatch in micro managing the wastage.

For small enterprises it become costly to micro-manage this wastage. For example regularly monitoring the CPU resources and shutting down non used CPU's and again when required powering them on, software for this is complex and not many companies has the expertise to build it. To have human labor do this makes it very costly. Similarly there weren't good software solution which data centers can use to consolidate unused disk space. This was until the advent of cloud.

5. BENEFITS OF CLOUD COMPUTING

With advent of cloud computing it has become possible to better manage data center utilization. Primarily cloud has made possible

- Better monitoring and reporting of IT infrastructure
- Pooling of unused CPU resources
- Pooling of unused disk space
- Sharing of unused CPU and disk resources
- Making all above Application agnostic

1. Virtualization

Virtualization has made it possible to use a single server for multiple purposes. It's possible to now have multiple web servers , email servers or multiple versions of database servers all configured on a single server in a application agnostic way. It helps in letting a multiple applications or multiple instance of a single application to run a single server thus helping consolidation of server computing power. This has reduced need of application dedicated servers and has helped reduce idle server times. The rapid growths of virtualization software's like Dockers and VMware have proven how cost effective they are. Clouds provide such containers services which work on virtualization techniques.

2. Server Less Architecture

The granularity of requesting computing resources is not only limited to requesting servers or virtual machine. In server less architecture CPU cycles , or processing time can be the smallest granularity of resource that can be accessed. For eg for a given minute if application is processing 10000 requests and in next minute the requests are 5000, the cpu power of 5000 request for that minute can by shared with other customer.

Server-less architecture allows applications to be written in terms of functions which are mapped to requests and whenever there is request, computing power for that request is made available .Once function ends, computing power is freed. This takes resource consolidation and micromanagement to as low as functions or CPU cycles level. Applications which are not used in night hours or which have server utilization as low as 10% , thus can free entire non utilized capacity to be shared with other applications. Most clouds now provide the framework for server less applications for eg Amazon Lambda, Google Cloud functions, and IBM Swift runtime.

3. Hard Disk Consolidation

Various disk virtualization tools , frameworks are not provided by cloud providers. With this applications a get a virtually unlimited storage space and there is no need to do micromanagement of disk. Amazons S3 , Google File systems, various cloud dB's are such storage space virtualization software's which do the management and consolidation of storage space by provided a abstraction over actual storage devices.

4. Easy Scaling

This ability to scale as per load, allows only to provision resources which are required as per current load. When load is less the extra resources can be shut down or shared with other applications.

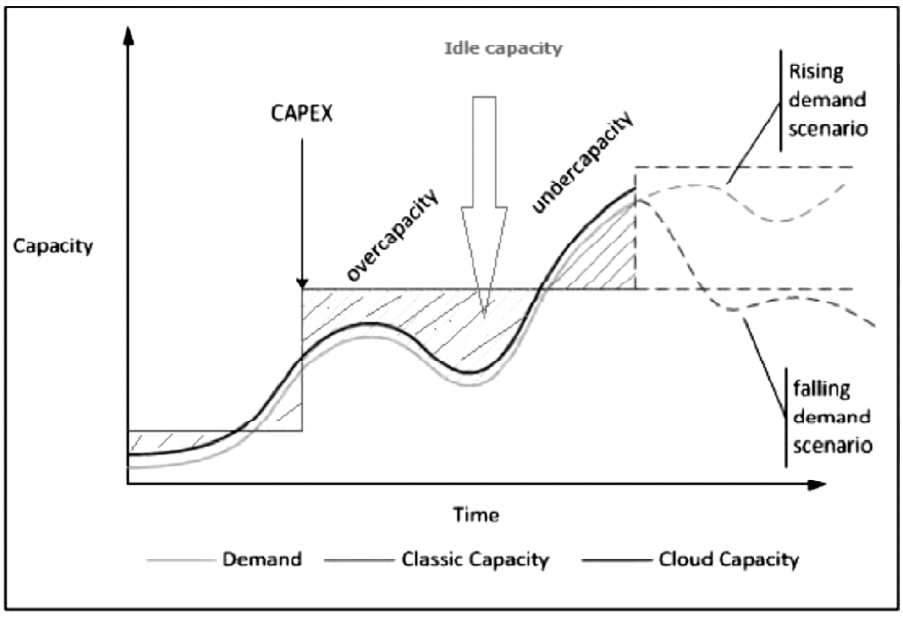

Fig. 2. Capacity Vs Utilization graph.

For *e.g.* application for public utility like banks will have more load in day hours, less in evenings and least at night. They need not have all the resources they need at peak capacity to be on for all 24hrs of the day. During non-peak time, these resources can be shut down or made available to other customers.

Below figure shows that in absence of easy scaling, the capacity is provisioned for peak load and during non-peak time, most of the capacity remains underutilized. However with auto scaling, provisioning follows the load closely and thus improves utilization.

Cloud provided for application agnostic ways of scaling the application on the fly. Auto scaling a feature provide by most clouds. This scaling is for computing resources like CPU cycles, servers as well as for storage space. The days of manually scheduling resources considering the peak need are gone. Everything is automated and can also be programmed through API's.

5. Monitoring and Reporting

Cloud systems provide better monitoring and reporting services. Now its very easy to monitor the utilization of every resource, device a, application and apply rules to take necessary actions. For e.g. if a application reaches 90% of CPU utilization a new server will be auto provisioned and load shared . When then load gets reduced the extra server then can be auto shutdown. Thus combining monitoring and auto scaling we have can easily micro manage resource utilization.

6. Location Consolidation

Instead of having small data centers spread across various companies, with big clouds likes the once by Amazon, Google, Microsoft, IBM we have data center consolidation. This consolidation also helps in recycling of electronic waste as the collection of electronic waste will be from places far less dispersed.

Collection and recycling of electronic waste becomes less costly. For E.g. Google claims "*We recycle 100% of the electronic equipment that leaves our data centers. Since 2007, we've repurposed enough outdated servers to avoid buying over 300,000 new replacement machines* ".

7. Green Energy

Using green energy like wind, solar requires high initial investment in addition to space and a suitable location. Because of this a small company having a data center might not find it affordable to power the data center by using green energy. With data center consolidation, big clouds can easily afford to have their energy needs fulfilled by green energy like wind and solar. For *e.g.*, Apple powers all its data centers using 100% renewable energy. This is possible because of large scale.

6. CONCLUSION

Current utilization percentage of data center resources like computing power and storage is very low. Mostly 90% of CPU cycles and 70% or storage remains underutilized. This has a huge environmental impact in terms of increased electricity consumption and recycling requirement. If this utilization rate is increased there will be big reduction in servers and storage required. Cloud computing technologies like auto scaling, virtualization, server less computing, storage abstraction etc. help in consolidation & better management of these resources thus increasing their utilization. With increased utilization, the environmental load of electricity and recycling effort can be reduced, thus help in reduction on carbon footprint of a data center.

7. REFERENCES

1. Mark Golden, Precourt Institute for Energy at Stanford University https://energy.stanford.edu/news/data-centers-can-slash-co2-emissions-88-or-more
2. Alex Benik, Battery Ventures, https://gigaom.com/2013/11/30/the-sorry-state-of-server-utilization-and-the-impending-post-hypervisor-era/

3. Lucas Mearian, http://www.computerworld.com/article/2519639/data-center/a-waste-of-space—bulk-of-drive-capacity-still-underutilized.html
4. Jonathan Koomey and Jon Taylor Anthesis group http://anthesisgroup.com/30-of-servers-are-sitting-comatose/ [2015]
5. 1 Kaplan, James M., William Forrest, and Noah Kindler. 2008. Revolutionizing Data Center Efficiency McKinsey and Company. [http://www.mckinsey.com/clientservice/bto/pointofview/pdf revolutionizing_data_center_efficiency.pdf]
6. J Miyuru Dayarathna, Yonggang Wen, Senior Member, IEEE, and Rui Fan Data Center Energy Consumption Modeling: A Survey http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7279063
7. DennisBouley"Estimatingdatacenterelectricalcarbonfootprint" https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf
8. Unused servers survey results analysis Mark Blackburn, IE & Andy Hawkins, http://www. thegreengrid.org /~/media/WhitePapers/Unused% 20Server% 20Study_ WP_101910_v1.ashx?lang = en
9. YEVGENIY SVERDLIK "Here's How Much Energy All US Data Centers Consume" http://www.datacenterknowledge.com/archives/2016/06/27/heres-how-much-energy-all-us-data-centers-consume/?utm_source = dck&utm_medium = widget&utm_campaign = trending_articles
10. Rackspace.com https://support.rackspace.com/white-paper/moving-your-infrastructure-to-the-cloud-how-to-maximize-benefits-and-avoid-pitfalls/
11. Mark Golden, Precourt Institute for Energy at Stanford University https://energy.stanford.edu/news/data-centers-can-slash-co2-emissions-88-or-more