

New data supports finding that 30 percent of servers are ‘Comatose’, indicating that nearly a third of capital in enterprise data centers is wasted

June 3, 2015

Jonathan Koomey* and Jon Taylor**

*** Research Fellow, Steyer-Taylor Center for Energy Policy and Finance, Stanford University
www.koomey.com**

**** Partner, Anthesis Group
www.anthesisgroup.com**

Contact: jgkoomey@stanford.edu or jon.taylor@anthesisgroup.com

Data centers are among the most sophisticated industrial facilities in modern society, but are rife with inefficiencies. While much has been written about the need for greater adoption of efficient technologies, a major part of data center energy use results from poor management of existing capital equipment. Fixing these management failures is a major opportunity for virtually all companies.

A case in point is the low utilization of information technology (IT) equipment inside data centers. According to McKinsey and Company, utilization of servers in business and enterprise data centers “rarely exceeds 6%” (i.e, they deliver no more than six percent of their maximum computing output on average over the course of the year) and up to 30% of servers are *comatose*— using electricity but delivering no useful information services.¹

Data for comatose servers

The Uptime Institute has also estimated that up to 30 percent of the servers in the US are comatose².

A new compilation of consolidated and anonymized data, taken from TSO Logic’s³ growing North American installed base across multiple industries, support this estimate. Available data from a pool of nearly four thousand servers indicate that 30 percent of the physical servers in the sample were comatose.⁴ These servers were running both virtualized and non-virtualized environments. Based on this

¹ Kaplan, James M., William Forrest, and Noah Kindler. 2008. Revolutionizing Data Center Efficiency McKinsey and Company. [http://www.mckinsey.com/client-service/bto/pointofview/pdf/revolutionizing_data_center_efficiency.pdf]

² The Uptime Institute estimate; <https://uptimeinstitute.com/research-publications/asset/comatose-server-savings-calculator>

³ TSO Logic is a leading provider of IT efficiency software for data centers, and was a 2013 Uptime Green Enterprise IT award winner. www.tsologic.com

⁴ We define comatose servers as those in the sample that have delivered no information or compute services in 6 months or more. Our sampling method excludes servers that are scheduled to be provisioned, focusing instead on abandoned servers and those that will never be used (because overcapacity is so large).

preliminary finding (supported by previous data released by Uptime and McKinsey and Company) there are now about 3.6 million comatose servers in the US, and 10 million worldwide.⁵

Excess server capacity

Data centers often have excess server capacity because identifying unused or over provisioned hardware with certainty is difficult with conventional tools. Managing resources simply by measuring CPU and memory usage alone isn't enough to ensure that a particular server is comatose.

Data center operators also need to look at upstream traffic or user access information per server from central IT management, virtualization and workload distribution systems. This approach identifies IT resources not doing any useful work so they can be decommissioned without adding risk to the business.

Insights into data center environment

Figure 1 gives an example of what state of the art software can reveal about comatose servers in most data center environments. Solutions like this one, which queries servers for operational data over the network, clearly show how incoming workload correlates with utilization, performance and power levels, and provides actionable reports that identify servers that have had no workload or traffic directed their way.

The participants in a spring 2012 Uptime Institute Server Roundup identified about 20,000 servers that were comatose. Shutting these servers off saved about 5 megawatts of IT load (and another 4 MW of associated cooling and infrastructure load). Applying these savings to the potential comatose servers worldwide could yield more than 4 gigawatts (GWs) of combined IT and infrastructure load reduction.

Alternatively, that freed up infrastructure could support additional IT loads of more than 2 GW, but those new loads would actually be creating value instead of mindlessly wasting electricity and financial resources.

In the twenty first century, every company is an IT company, but too many enterprises settle for vast inefficiencies in their IT infrastructure. The existence of so many comatose servers is a clear indication that the ways IT resources in enterprises are designed, built, provisioned, and operated need to change. The needed changes are not primarily technical, but revolve instead around management practices, information flows, and incentives.⁶

We will update these results periodically as the data set continues to grow, with special focus on virtualized vs non-virtualized environments.

To quantify the potential efficiency improvement in your data center from removing comatose servers, try using one of these online calculators:

www.uptimeinstitute.com/research-publications/asset/comatose-server-savings-calculator

www.tsologic.com/savings-calculator/

⁵ IDC 2012 world server count was 35 million; 10 million comatose servers assumes that 30% of servers worldwide are comatose. This estimate includes a mix of standalone servers as well as those acting as host servers (Hypervisors) in a virtualized environment.

⁶ To learn how to fix these problems, see the online class titled "Management essentials for transforming enterprise IT": <https://www.heatspring.com/courses/management-essentials-for-transforming-enterprise-it>

Figure 1: Sample dashboard from TSO Logic's Software

Notice the green line in the capacity graph – it indicates the number of servers that are idle and not doing any useful work.

