Uptime Institute®
INTELLIGENCE

2020

Beyond PUE: Tackling IT's wasted terawatts

This report is an excerpt of a comprehensive report available to Uptime Institute Network members. This document includes the highlights of the material contained in the larger original. For more information about membership, see: www.UptimeInstitute.com/UI-Network

Dr. Rabih Bashroush

Andy Lawrence

Uptime Institute

January 2020

Abstract

Analysis of data center energy efficiency trends over the past decade, supported by detailed power usage data from over 300 data centers, shows that significant improvements have been made. However, Uptime Institute's analysis also shows that very substantial energy reduction opportunities still remain untapped. While gains in mechanical and electrical efficiency have stalled over the past few years, it remains that over 65% of the power used by IT in data centers is used to process just 7% of the work, due to aging equipment inefficiencies. But if it is time for operators to pay more attention to IT energy efficiency, they need to do their analysis carefully: the slowdown in Moore's law is creating new complications.

Key findings

- Efforts to improve the energy efficiency of the mechanical and electrical infrastructure of the data center are now producing only marginal improvements. The focus needs to move to IT.
- The most-implemented practices for energy reduction are those that do not require substantial investment but do require process, discipline or relatively minor and incremental investments.
- Initiatives that span IT, that involve most cultural and multi-disciplinary changes and that require major strategic operational changes are the least-implemented energy efficiency practices.
- Major energy efficiency opportunities involving IT remain untapped – partly due to a misplaced management focus on infrastructure.
- Energy-saving opportunities on the IT side are so great that if fully addressed, they would significantly reduce data center energy use and carbon footprint, would slash energy bills and would likely lead to reduced demand for cooling and critical power equipment.
- In a study of 300 data centers, aging IT kit (older than five years) accounted for 66% of IT energy use but contributed just 7% of the compute capacity.
- All these issues are well-known and can only be resolved by senior management, which is empowered to make decisions that cross the IT/facilities boundary or drive behavior among suppliers and clients. An understanding of the sheer scale of the energy savings should encourage executives to address the issues more directly.
- Over the past few years, while processor lithography has stagnated at 14 nanometers, the increase in performance per watt has been accompanied by a steady increase in idle power consumption (perhaps due to the increase in core count to achieve performance gains). This is one reason why the case for hardware refresh for more recent kit has become weaker: Servers in real-life deployments tend to spend a substantial part of their time in idle. As such, the increase in idle power may overall offset energy gains from performance.
- If a server spends a disproportionate amount of time in active idle mode, the focus should be on active idle efficiency (i.e., choosing servers with lower core count) rather than just on higher server performance efficiency, while satisfying overall compute capacity requirements.

Introduction

It is widely known that the aggregated energy consumption of the infrastructure of global IT (data centers, servers, networks, devices) has been rising steadily for many years – even if the scale of overall energy consumption is a matter of debate and requires further research). It seems likely that the annual consumption of energy by data centers is somewhere between 400 terawatt-hours (TWh) and 500 TWh, depending on what is counted as a data center. To put things in perspective in terms of demand, research by Uptime Institute Intelligence shows that every time an image is posted on Instagram by the Portuguese soccer star Cristiano Ronaldo (who at the time of writing has the highest number of followers on the platform), his more than 195 million followers consume nearly 30 megawatt-hours of energy to view it.

Some forecasters (most notably the International Energy Administration) have predicted a flattening or even downturn in overall energy use by data centers – the result of improving efficiency. Certainly, the explosion in IT demand in the past decade (2010-2020) did not translate directly into the same rate of growth for infrastructure energy consumption. However,

Uptime Institute expects that demand for IT services and data centers will substantially outpace the gains from efficiency practices over the next five years, resulting in steadily increasing energy use.

Increasing energy demand by IT – especially if that energy is used inefficiently -- has both direct and indirect consequences for data center operators and major customers. First, power is a major cost component of all IT services and inefficiency, therefore, increases costs and, very likely, reduces overall margins. Second, a shortage of power in some geographies increases power prices; and third, most power sources produce carbon dioxide, and all polluters will come under increasing political and economic pressure to reduce this in the coming years.

Summary, conclusions and opportunities

Data shows that the past decade has seen substantial gains in IT and data center energy efficiency. This can largely be put to three factors: A big focus on mechanical and electrical infrastructure, improved designs and equipment, and optimized management and processes. This is highlighted and emphasized by the way power usage effectiveness (PUE) data has evolved over time. However, for IT, efficiency gains were largely driven by steadily increased equipment throughput relative to power consumption, thanks to technological advancement driven by the likes of Moore's law. There has been little active adoption of energy conservation best practices by operators.

With PUE gains plateauing, the major energy-saving opportunities in the future will have to come from shifting the focus to IT optimization best practices. The five main opportunities that we believe will present the best savings opportunities are:

1. Optimize the server refresh lifecycle. With 40% of deployed servers in the sector older than five years doing just 7% of the work and consuming more than 66% of energy, the optimization opportunity is enormous.

2. Increase server utilization. Uptime Institute's analysis shows that increasing server utilization can yield much more savings than reducing PUE, for much less upfront investment. With average sector utilization levels at around 25%, there is plenty of low-hanging fruit to be had, at least until we've reached the 40% average utilization mark.

3. Right-size redundancy based on workload requirements. While it seems easier to design for the highest common denominator in terms of workload resiliency requirements, best practices show that right-sizing redundancy levels (and tracking the appropriate key performance indicators to reflect it) can yield energy savings of up to 90%.

4. Consolidate infrastructure to benefit from economies of scale. The industry has come a long way in terms of consolidation, but 80% of data centers still contain fewer than 25 racks. This represents a significant consolidation opportunity that could be achieved through internal organizational infrastructure rationalization or a move to colo or the cloud.

5. Address energy consumption across traditional design boundaries. Work over the past decade has shown that the next order of magnitude in energy efficiency will only be achieved by working across traditional design boundaries and organizational silos, engaging business, IT and infrastructure teams.

About the Authors

Dr. Rabih Bashroush is Uptime Institute's Research Director. An international authority in the field of information and communication technology digital infrastructure and consultant for various European governments (including the European Commission), Dr. Bashroush has led numerous research programs designed to provide a better understanding of digital infrastructure, its impact, and how it can operate more effectively. Contact: rbashroush@uptimeinstitute.com

Andy Lawrence is Uptime Institute's Executive Director of Research. Mr. Lawrence has built his career focusing on innovative new solutions, emerging technologies, and opportunities found at the intersection of IT and infrastructure. Contact: alawrence@uptimeinstitute.com

Want more detail?

Become an Uptime Institute Network member! It's easy and provides a wealth of information needed to own and/or operate your infrastructure most effectively, reduce downtime and create a predictable hybrid computing infrastructure.

Uptime Institute is the world's global authority on the design and operations of data centers and has more than 25 years of experience helping customers understand what it takes to keep their data centers running smoothly and efficiently. Uptime Institute brings together thousands of data center professionals who wish to continue to refine their organization's IT service delivery proficiency as well as their own personal skills. Membership in Uptime Institute's community program profiles the essential tools, knowledge and experience to assure business continuity, disaster preparedness and operational efficiency. Members of Uptime Institute experience fewer disruptions, lower costs of operations, reduced unplanned downtime and can defend their IT choices.

Full versions of this and many other reports, roundtable discussions, face to face meetings with like-minded peers, data center tours and access to online incident research tools are available to members of Uptime Institute. See: https://uptimeinstitute.com/ui-network for more information about joining the industry's most respect community of data center professionals and practitioners.