February 6, 2012

Server And Data Center Predictions For 2012

by Richard Fichera
for Infrastructure & Operations Professionals

FORRESTER® Making Leaders Successful Every Day

February 6, 2012

Server And Data Center Predictions For 2012

Continuation Of Old Trends And A Few Disruptions

by Richard Fichera
with Christopher Voce and Lindsey Kempton

EXECUTIVE SUMMARY

This report outlines Forrester's solution for infrastructure and operations executives working on data center infrastructure. This report is designed to help you build a road map for 2012. Looking at the remainder of 2012, the landscape is a combination of the familiar and the new. The drumbeat of new servers and new x86 CPU offerings will continue, with announcements from all of the major players. Forrester also expects to see the first commercialization ARM CPUs, which will offer additional options for energy-efficient processing. On the data center front, expect that the pressures on power-efficient operation will continue unabated, driving adoption of emerging data center infrastructure management (DCIM) software. Modular data center options will proliferate as entry prices decline and additional options appear. This report focuses on eight key predictions that I&O leaders should consider when planning for their server and data center infrastructure moves.

TABLE OF CONTENTS

NOTES & RESOURCES

This research leverages the many conversations Forrester analysts have with I&O professionals, software vendors, and enterprise thought leaders.

Related Research Documents

"Updated Q3 2011: Power And Cooling Heat Up The Data Center"
September 21, 2011

"Updated Q2 2011: Are Converged Infrastructures Good For IT?"
June 22, 2011

"Put ARM-Based Servers On Your Server Planning Horizon"
March 15, 2011

2011 PRESSURES CONTINUE UNCHECKED INTO 2012

Infrastructure and operations professionals (I&O pros) can expect the underlying pressures they saw in 2011 to continue into 2012 unabated — with some newer emerging issues to contend with. These include:

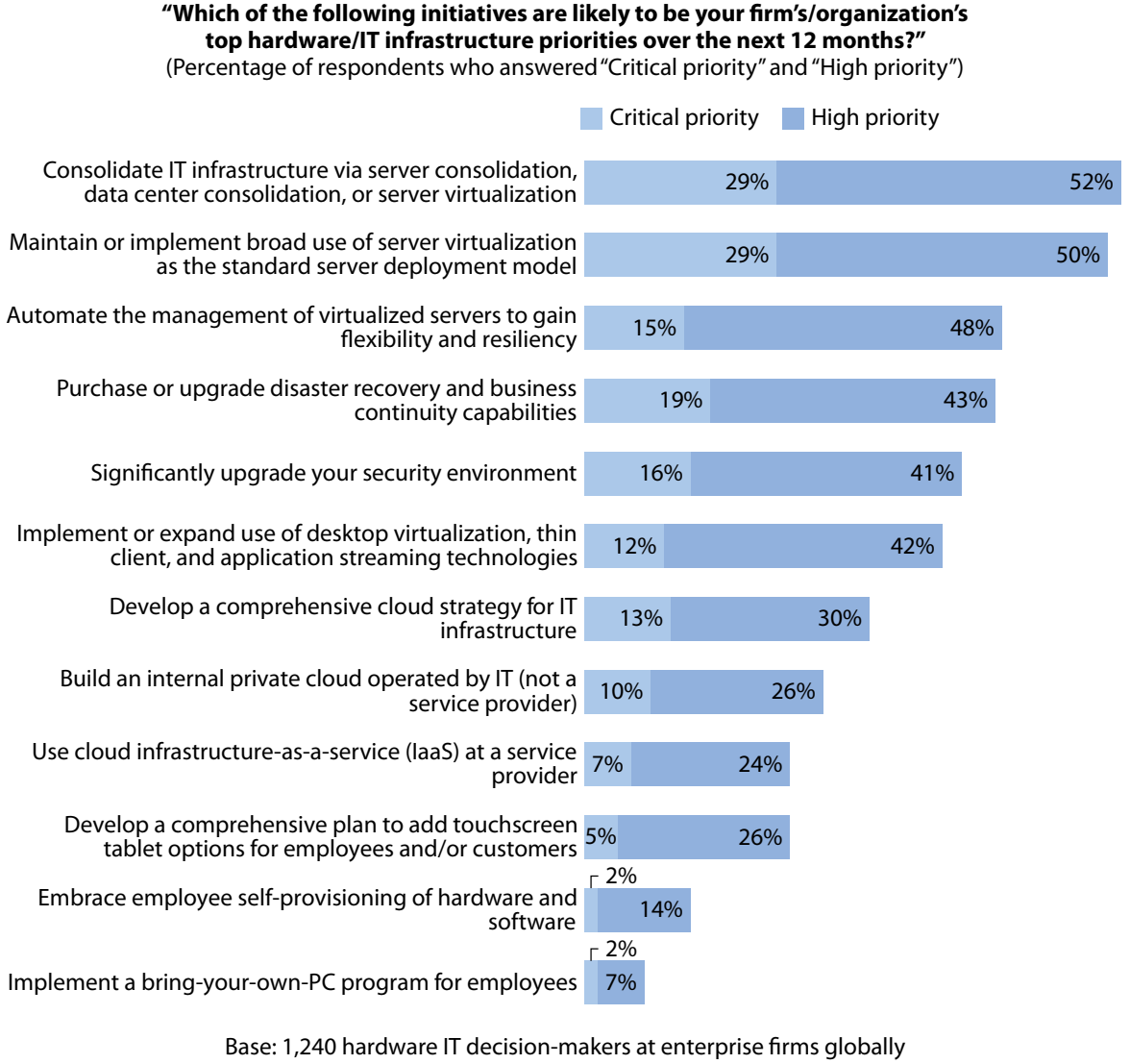
- **Continual expansion of existing and new workloads . . .** While some workloads such as cloud computing and analytics report higher growth than others, Forrester has not seen any indication of flat or negative growth in underlying processing requirements. The pressures on some application sectors are immense and show no signs of letting up. For instance, analytics and big data; biomedical computing; and online web services such as search, social media, digital media, gaming, and a host of others are driving requirements for core computing as well as storage and networking, at an accelerating pace.

- **. . . but an ongoing need to maintain operational efficiency.** At the same time, IT does not get a blank check from executive management. Financial pressures remain extreme, with continual pressure to expand processing resources at rates well in excess of the flat to single-digit-growth budgets allocated to IT. In addition, I&O pros still must contend with regulatory pressures. While still coping with the ramifications of Sarbanes-Oxley, I&O groups are bracing themselves for the implications of the new Dodd-Frank Wall Street Reform and Consumer Protection Act, which is expected to trigger new requirements with its more than 200 new rules regulating financial services firms.

- **Continued execution of infrastructure transformation initiatives.** The pace of infrastructure transformation as part of the process of providing new capabilities continues (see Figure 1). I&O pros are carrying on with consolidation projects including server as well as network, storage, and entire data centers. This effort continues alongside and as part of continued efforts to virtualize and standardize the data center environment.

- **The need for long-term capacity visibility.** As processing requirements escalate, many organizations are discovering that they need to pay particular attention to long-term capacity planning and make better use of their existing physical facilities. Rather than just cutting the electric bill, many IT organizations will deem it more important to optimize data center capacity and resource consumption to avoid new data center construction or a major outsourcing contract.

Figure 1 IT Infrastructure Priorities



Base: 1,240 hardware IT decision-makers at enterprise firms globally

Source: Forrsights Hardware Survey, Q3 2011

61442 Source: Forrester Research, Inc.

- **What I&O needs to know.** ARM servers are no longer science fiction, and in 2012 you will be able to buy one from one or more of the largest server vendors. They will not be suitable for all, or even the majority, of workloads. But for selected ones they can potentially deliver five times or more throughput per watt.[2]

- **What I&O needs to do about it.** Initially, the burden is on the application development function to evaluate the suitability of these new systems, but as these investigations mature, I&O will have to get involved to understand the implications.

3. Competition Between Intel And AMD Will Remain Healthy

Despite the dominance of Intel in the server space, competitive pressure from archrival Advanced Micro Devices (AMD) will continue. AMD's recent Magny-Cours CPUs, with up to 16 cores, have put substantial price-performance pressure on Intel's current line. While the Intel 22 nm products will probably allow Intel to reclaim an overall performance advantage again, AMD will probably follow up in late 2012 with its anticipated "Piledriver" core, which is the successor to the "Bulldozer" core of the current Magny-Cours. Regardless of the details of the faceoff between the two vendors, AMD's existence and focus on emerging cloud computing platforms will be one of the major factors in keeping Intel aggressive and competitive and preventing them from abusing their near monopoly position in server CPUs:

- **What I&O needs to know.** AMD will survive as the No. 2 server CPU vendor. Aside from performance differences driven by the differing number of cores and performance of the cores, AMD products will remain completely software-compatible with Intel products.

- **What I&O needs to do about it.** Reassess your Intel versus AMD server investments in the light of the new announcements, and for large procurements, don't be afraid to play one vendor off against the other. In the final analysis, there are few workloads that can't be handled adequately by either vendor.

4. Hybrid Scalar/GPU Computing Will Become Mainstream

In 2010, we saw the introduction of general purpose graphics processing unit (GPGPU) technology as an adjunct to conventional computing architectures. GPUs are highly effective for some computing tasks, once the high cost of programming them is overcome. By the end of 2011, all the necessary elements were in place for an accelerated acceptance into mainstream computing: ubiquitous hardware support by all the major systems vendors, mature software development environments (despite some vendor-driven splits between OpenCL and Cuda), and at least two major suppliers in Nvidia and AMD.[3] The penultimate step in the mainstreaming of GPUs was the successful introduction of AMD's Fusion desktop processors, which combine an x86 scalar core with an integrated GPU. This served as a catalyst for further GPU development tool activity and raised the awareness of GPUs as a potential solution across a wide range of developers:

- **What I&O needs to know.** All major vendors offer servers with integrated GPUs, in both rack and blade or dense rack configurations. There are also better development tools and skilled programmers to take advantage of GPUs. As with ARM servers, the initial burden is on the internal app dev groups and the independent software vendor (ISV) community, but for any application that has inherent parallelism, GPUs offer major potential performance gains.

- **What I&O needs to do about it.** I&O should work with the major application stakeholders to determine which applications can be rewritten to take advantage of GPU capabilities. Many ISVs, including Matlab, Mathematica, MSC Nastran, Ansys, Abaqus, and Autodesk, currently support GPUs, and I&O groups should work with major application stakeholders to identify potential candidates.

5. Converged Infrastructure Offerings Will Continue To Evolve And Differentiate

Over the past two years we've seen a considerable churn in converged infrastructure (CI) technology, and Forrester expects this pace to continue in 2012.[4] Vendors will focus on improving the integration of storage and network virtualization and increasing integration of the systems-level virtualization capabilities with higher-level management software. The major system vendors with blade server offerings such as Cisco Systems, HP, and IBM will lead the charge. Forrester expects HP and IBM to refresh their blade platforms this year, as HP's c-Class blade servers were introduced in 2006 and IBM's BladeCenter's architecture is even older. The updates should include the upcoming Intel 22 nm Intel CPU technology and with an eye toward the future Intel "Ivy Bridge" as the main deployment platform. While we can't predict exact feature sets, we would expect to see some combination of:

- **More granular abstraction of physical components.** Cisco one-upped all incumbent vendors with its UCS architecture, which offered a richer presentation of physical components as software objects. The others are likely to follow suit.

- **Increased use of integrated fabrics for system scaling.** We can see clues about future server implementation by looking at the HP "Redstone" server, unveiled as part of their ARM-based server announcement. Expect the passive backplane architecture of current blades to give way to a more scalable high-speed fabric, with more flexible topologies for aggregation and management.

- **Even more focus on power and cooling.** Power and cooing efficiency was one of the primary values offered by the early generations of blade servers, and Forrester expects that this focus will continue. Blade enclosures offer more opportunities to optimize power and cooling than a conventional rack server because of their integrated design. The next generation of blade servers will also likely offer better federated management across multiple enclosures.

- **Better systems management.** Expect vendors to improve the current blade server management tools to accommodate the more complex software environments of future systems. While all the vendor implementations will differ, Forrester expects that one common thread running through them will be federation of multiple system enclosures into a seamless management presentation. Some vendors have begun this process, but the next generation of blade servers will probably contain the necessary software and hardware hooks to make this even more seamless and useful.

Above and beyond the underlying hardware and system management tools, the vendors will compete to extend the CI concept to include composite application services, presented as abstractions including servers, network, storage, and security policies. HP, with its CloudSystem Matrix, is a prime example of this trend, and will continue to enhance this already strong offering. Meanwhile, other vendors will be following closely on HP's heels, and the proliferation of offerings in 2012 will be at best bewildering and at worst overwhelming for most IT groups:

- **What I&O needs to know.** The major challenge for I&O groups will be understanding the complexities of the available options and coordinating with all the affected stakeholders in the organization. The complexity of CI, along with its inherent lack of standardization above the basic VM layer, tends to lead to vendor lock-in. This isn't necessarily bad as long as the tradeoffs for lack of vendor choice are mitigated by tangible benefits from the solution.

- **What I&O needs to do about it.** Aggressively pursue your current vendor for updates and visibility to futures — the sheer volume of new product technology in the CI space will be difficult to track. Where there is the opportunity for a major architectural change to infrastructure as well as applications, there is also an opportunity to introduce a new cycle of competition into your environment.

6. DCIM Solutions Will Proliferate, Ushering In An Age Of Confusion

In 2011 we saw the emergence of the first generation of DCIM solutions. These offer data center operators the ability to take a holistic look at their entire physical data center environment and to not only understand what is happening in real time but also make intelligent decisions about future changes and investments. Forrester defines a rough taxonomy of functions for DCIM software as follows:

- **Inventory and discovery.** The solution collects real-time physical infrastructure data, including physical rack and equipment location, circuit capacities, CRAC/cooling data, and thermal data. To perform these functions, the software vendor needs to collect data from the equipment of multiple, and in many cases competing, vendors, along with data from other sources including other CMDBs. Prospective purchasers need to understand in detail how the initial facility discovery and ongoing updates will be handled.

- **Continuous data collection.** Once the DCIM software is installed, it needs to collect data from the environment on a continuous basis. The granularity of the data collection, the type of database in which it's stored, and the extent to which it can be used for other analyses are characteristics that differentiate the vendors.

- **Consolidated display.** DCIM solutions offer a display and reporting of real-time data, trends, and exceptions. One of the earliest functions of DCIM software was as a unified dashboard for displaying a single-source view of power, physical, and thermal information. Bridging the worlds of facilities and IT operations was one of the initial selling propositions of the first-generation DCIM applications and continues to be a focus of many DCIM products.[5]

- **Trend analysis.** The next major step in functionality is the ability to display trends for power and cooling. This will offer data center operators the ability to better predict and diagnose conditions that have exceeded preset limits.

- **Model solutions for implementation.** This is still a work in progress for most vendors, but the ability to model potential solutions to power and workload problems is one of the areas where the system vendors and infrastructure equipment suppliers intersect. To effectively propose an optimal set of allocations, the software needs to understand the behavior of the system at a granular level, the potential workloads, and the details of the power and cooling environment. High-end IT systems management ISVs, such as HP, CA, and BMC Software, have traditionally offered this, but it's appearing in many of the DCIM offerings as well. The degree to which the potential solution simulations are aware of the underlying workloads and the accuracy and granularity of the simulations are all differentiators.

The universe of DCIM vendors is evolving and expanding — more than 60 vendors currently identify themselves as supplying DCIM solutions, with an average of two or three new ones cropping up each month. If DCIM follows the same trajectory as other emerging solution spaces, we can expect considerable dropout of smaller players combined with the M&A activity that is already occurring. The integration of DCIM with conventional management tools is a work in progress, and the competitive battle among DCIM vendors will likely be defined by the degree to which this integration can be made seamless and useful for I&O professionals. The challenge for I&O professionals will be separating myth from reality in a rapidly changing vendor landscape:

- **What I&O needs to know.** The need for DCIM is not new — problems with managing the instantaneous operational aspects of power and cooling and planning for future changes have been around for years. What is new is the availability of single solutions that now begin to merge the world of facilities, I&O, and even application service delivery. At a minimum, I&O groups need to forge new working relationships with facilities to begin to realize the power of these solutions.

- **What I&O needs to do about it.** I&O groups should be prepared to invest time in evaluating and selecting a DCIM offering. Forrester believes that the benefits are real — such efforts will be rewarded with incremental power and cooling efficiencies ranging from 5% to 30%, depending on the initial state of the environment.

7. Modular Data Center Options Will Continue To Evolve

Even after upgrading systems and optimizing power and cooling, eventually there comes a time when new data center space is needed. Either as a result of major corporate changes such as M&A, growth of current business requirements, or new lines of business, many enterprises can be expected to outgrow a current facility in the next three to five years. The story has been the same for the past several decades. What has changed are the options available for new data center capacity. To the old menu of hosting providers and colocation we now have two new major options — cloud and modular data centers.

Leaving aside cloud for the moment, which in many ways implies a more fundamental architectural change and is not mutually exclusive with new data center requirements, modular data centers represent a significant shift in provisioning new data center capacity. Forrester expects modular data centers to continue their rapid growth in 2012, bringing continued segmentation of offerings, with more choice in "room within room" offerings, and smaller modular entry points for self-contained pod offerings, along with enhanced management software as differentiators:

- **What I&O needs to know.** Building new data center capacity is a major undertaking for most enterprises and involves almost the entire chain of corporate stakeholders, including business units, finance, facilities, I&O, and executive management. For I&O, the most commonly missing element is accurate visibility into future requirements.

- **What I&O needs to do about it.** I&O needs to build long-range capacity planning processes into IT and encourage business units to participate. Modular data centers' two stellar benefits will remain intact in 2012: They are somewhat cheaper than conventional data centers and much, much faster to acquire. A new pod can be ordered and delivered from all of the major suppliers in anywhere from six to 12 weeks with infrastructure installed, compared with 18 to 24 months for conventional data center space.

8. Data Center Industry Initiatives Will Be Interesting, Influential, But Not Definitive

Forrester sees organizations such as the Open Data Center Alliance, the Green Data Center Alliance, and the Facebook-led Open Compute Foundation as having a legitimate role in promoting awareness of problems and solutions, disseminating best practices, and educating the vendor community about requirements. But they are not going to directly define products, despite vendor participation. Vendors will participate for incremental advantages, such as seeing early trends and getting hints about new products, but it's unlikely that vendors will deliver more standardized or interoperable products as a result of these organizations. There is simply not enough mass in these organizations to dictate products, despite the attractions of some of the proposals:[6]

- **What I&O needs to know.** I&O professionals should be aware of the activities of these organizations. They offer many insights into the future of infrastructure technology, and many of them offer excellent lists of best practices and workshops.

- **What I&O needs to do about it.** I&O professionals should take notice of what these organizations are up to, but should not be making product decisions based on these groups. Products are still going to be delivered through the conventional vendor ecosystem that they are used to dealing with. Any direct impact on the working world of the I&O professional will be very indirect.

ENDNOTES

[1] ARM Holding has a different business model than most companies in the CPU business. It does not manufacture directly, but rather licenses its IP at different levels, ranging from a completely predefined CPU design to what they call an architecture license, which allows the holder to actually change the instruction set of the system should they desire. There are a lot of ARM licensees because ARM accounts for as much as 90% of the 32-bit embedded CPU population, including products from such high-profile licensees as Apple (ARM is the basis of Apple's A4 and A5 processors used in the iPhone, iPod, and iPad), Broadcom, Qualcom, Texas Instruments, Marvell, Alcatel-Lucent, and a host of others.

[2] For more information, see the March 15, 2011, "Put ARM-Based Servers On Your Server Planning Horizon" report.

[3] Note that not all apps/segments can benefit from GPUs, but the ones that can, such as image processing for entertainment, medical and security, geophysics, CAD/CAE, and biomedical computing, can show 10 to 50 times performance improvement in core algorithms and easily five to 10 times improvement in total problem runtimes.

[4] For more information, see the June 22, 2011, "Updated Q2 2011: Are Converged Infrastructures Good For IT?" report.

[5] For more information, see the September 21, 2011, "Updated Q3 2011: Power And Cooling Heat Up The Data Center" report and see the December 9, 2009, "Put DCIM Into Your Automation Plans" report.

[6] Facebook, for example, makes a compelling case with its Open Compute Foundation for its unique form factors and its high-voltage AC power supplies. It's likely that elements of this architecture, such as high-efficiency HV AC power supplies, will show up in future general purpose products, but Forrester does not believe that the total package of form factor, fans, and sheet metal will ever appear in any form except in very high-volume customer procurements. On the other hand, Facebook has suddenly enabled more sources for the next time they do go to market for another large procurement.

FORRESTER®

Making Leaders Successful Every Day

Headquarters

Forrester Research, Inc.
60 Acorn Park Drive
Cambridge, MA 02140 USA
Tel: +1 617.613.6000
Fax: +1 617.613.5000
Email: forrester@forrester.com
Nasdaq symbol: FORR
www.forrester.com

Research and Sales Offices

Forrester has research centers and sales offices in more than 27 cities internationally, including Amsterdam, Netherlands; Beijing, China; Cambridge, Mass.; Dallas, Texas; Dubai, United Arab Emirates; Frankfurt, Germany; London, UK; New Delhi, India; San Francisco, Calif.; Sydney, Australia; Tel Aviv, Israel; and Toronto, Canada.

For the location of the Forrester office nearest you, please visit: www.forrester.com/locations.

For information on hard-copy or electronic reprints, please contact Client Support at +1 866.367.7378, +1 617.613.5730, or clientsupport@forrester.com.
We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Research, Inc. (Nasdaq: FORR) is an independent research company that provides pragmatic and forward-thinking advice to global leaders in business and technology. Forrester works with professionals in 19 key roles at major companies providing proprietary research, customer insight, consulting, events, and peer-to-peer executive programs. For more than 28 years, Forrester has been making IT, marketing, and technology industry leaders successful every day. For more information, visit www.forrester.com.

FORRESTER®