CREATING A HYPER-EFFICIENT HYPER-SCALE DATA CENTER

By Jimmy Pike
Ty Schmitt
Frank Frankovsky
Todd Brannon

A growing number of organizations today utilize computing "clouds" to deliver many applications familiar to Internet users as well as flexible access to powerful compute and storage resources. Integrated solutions designed to minimize acquisition and operating costs, maximize energy efficiency, and enable rapid scalability can help operators of these and other hyper-scale data centers speed deployment, lower total cost of ownership, and reduce environmental impact.

Related Categories:

Cloud computing

Visit DELL.COM/PowerSolutions for the complete category index.

As computing architectures continue to evolve, server-centric "cloud" and large cluster environments are increasingly being utilized to deliver common applications as well as flexible access to powerful compute and storage resources. A cloud computing environment is one that is designed to cluster and/or load balance computation across thousands, or even tens of thousands, of compute nodes. Users typically connect over the Internet to tap into these resources "in the cloud." Internet search, free e-mail, online media, and multiplayer gaming are representative of this emerging computing paradigm. Cloud computing platforms are also increasingly being deployed by organizations that offer computing resources as well as software and storage hosting services over the Internet. In parallel with this trend, traditional clusters employed across a wide range of industries to perform compute-intensive tasks such as financial risk analysis, geophysical modeling, electronic design analysis, graphics rendering, and medical research continue to grow ever larger.

Cloud and large cluster computing environments have distinctly different design and management requirements from traditional general-purpose environments. For example, in the massively scaled-out architecture of a hyper-scale data center, any inefficiency—such as excess power consumption or poor space utilization—can significantly drive up the cost of computation. As a result, maximizing the efficiency of a compute cloud at the node, rack, and facility level is a key source of business value and an essential design consideration.

Cloud and large cluster computing environments also have different hardware requirements from traditional computing environments. For example, in a hyper-scale data center, it is usually the *application* that provides availability and resiliency, not redundant hardware within the individual nodes. As a result, server-level availability is not required—which significantly changes the configuration and characteristics of individual compute nodes.

UNDERSTANDING HYPER-SCALE DATA CENTER REQUIREMENTS

In a hyper-scale computing environment, efficiency is a key driver of enterprise success. But unlike traditional computing environments, hyper-scale data centers generally have thousands, or even tens of thousands, of compute nodes. In this type of massively scaled-out architecture, even very small inefficiencies can significantly drive up the cost of computation. Also, because the workload requirements of individual nodes in a hyper-scale environment are generally less stringent than those of nodes in a general-purpose environment, over-provisioning can occur and significantly drive up the cost of computation.

For example, most general-purpose servers and storage come equipped with standard features such as redundant components for availability and reliability that are not needed in a cloud environment. As a result, organizations often pay a "feature tax" on unneeded capabilities. In addition, in most general-purpose servers, the power supply and cooling components are designed to accommodate the power and cooling needs of the maximum configuration of components available on that server. As a result, general-purpose servers often consume more electricity and generate more heat than necessary to support the workload and configuration required of a node in a specific cloud or cluster environment.

Furthermore, because a hyper-scale environment utilizes thousands of servers, the excess heat generated by unnecessary features and over-provisioned power and cooling components typically found in general-purpose servers requires additional cooling that can significantly increase facility-level energy costs. Excess heat and bulk also limit node density, which can increase the amount of space required to deploy a server farm. In the multi-thousand-node environment of a fast-growing compute platform, such space requirements can significantly add to facility costs and slow project schedules.

Data center efficiency considerations are not just limited to individual nodes, however. The design and layout of a data center at the *facility* level is equally important. For example, the layout of racks, power, networking, and heating, ventilation, and air-conditioning (HVAC) systems within a facility is a critical factor contributing to the cost of computation. A facility that is not carefully designed to minimize heat and optimize air flow can generate tremendous excess energy and space costs. It is essential that great care be taken when designing a hyper-scale data center to minimize inefficiency at the node, rack, and facility level.

CO-DEVELOPING DESIGNED-TO-ORDER COMPUTING SOLUTIONS

To help organizations maximize the efficiency and effectiveness of their hyper-scale data centers, Dell recently created the Data Center Solutions (DCS) Division. The DCS team comprises solution consulting, engineering, supply chain, and project management experts with in-depth knowledge regarding the needs of hyper-scale data centers and cloud computing environments. The DCS Division offers a wide range of services, including data center design consulting, designed-to-order hardware, and custom service offerings for rapid deployment and system maintenance.

By employing a close co-development process with organizations, the DCS team can help improve the efficiency of an existing computing platform or assist in designing a new hyper-scale data center from the ground up. The DCS team is organized and scoped to support the needs of these specific computing environments, as shown in Figure 1. By building integrated solutions designed to minimize acquisition and operating costs, maximize energy efficiency, and enable rapid scalability, DCS can help organizations speed deployment, lower total cost of ownership (TCO), and reduce the environmental impact of their compute clouds.

To help organizations maximize the computing power of a hyper-scale data center while minimizing overall costs, the Dell DCS Division offers data center design consulting. In data center design consulting, a DCS team works closely with an organization to help understand its current and future computing requirements, analyze and assess the efficiency and effectiveness of the data center currently in place, and co-design a comprehensive data center solution.

For example, in the assessment phase, the DCS team evaluates the features and specifications of the compute nodes currently in place to help identify areas of inefficiency, such as unneeded features or excess power consumption. The team then evaluates the layout of the facility—including the placement of nodes, racks, networks, and power and cooling systems—to help identify inefficiency in data center layout. The team also applies advanced computational fluid dynamics (CFD) analysis to assess the effectiveness and efficiency of the power and cooling systems.

Following the assessment phase, the DCS team develops a comprehensive data center solution designed to optimize the

	Solutions for typical data center environments	Solutions optimized for cloud computing environments
Deployment size	Any	Typically 1,500 nodes or more
Application	Any	Homogeneous cloud applications, not requiring hardware high availability at the node level
Validation	Broad enterprise interoperability and deep product testing	Point solution only
Support	Comprehensive enterprise hardware, OS, applications, and solutions	Engineer-to-engineer and point solutions only
Service	Numerous packaged options	Site-specific service plan, with parts on-site
Deployment span	Multiple locations and/or staggered over a long time	Limited number of deployment locations with large-scale builds
Design	General-purpose or optimized for widely used solutions	Specific to each organization and optimized for clouds

Figure 1. *Cloud computing platforms are optimized for tightly scoped requirements*

efficiency and effectiveness of the computing environment at all levels of scale. At the node level, for example, the DCS team designs custom-built hardware tailored to meet an organization's specific computing needs, to avoid a "feature tax" for unneeded capabilities. DCS designers also use best thermal practices and state-of-the-art technology to help ensure that custom-designed hardware is as energy-efficient as possible. In particular, DCS designers take advantage of advanced technologies such as the following:

- **High-performance, energy-efficient components:** To help optimize computing power and efficiency in processors, memory, and hard disk drives
- **Right-sized power supplies and fans:** To help meet the specific requirements of each compute node's feature configuration
- **Low-flow fan algorithms:** To help optimize both fan efficiency at the server level and return air temperature to computer room air-conditioning units
- **Increased memory density:** To reduce dual in-line memory module (DIMM) count using technologies such as quad-rank DIMMs to help reduce excess heat and overall power consumption
- **Custom-designed chassis:** To help optimize density and airflow in each node, taking advantage of the absence of components such as redundant power supplies

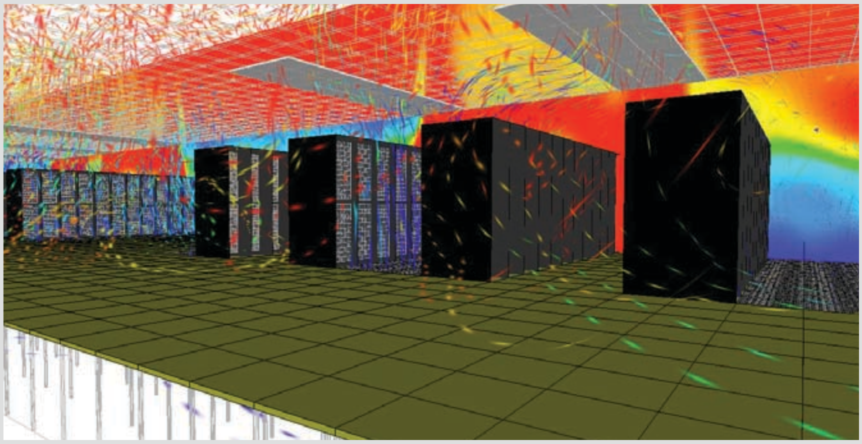
Furthermore, the DCS team applies best thermal practices to design the layout of the entire facility. For example, using state-of-the-art CFD modeling, the DCS team can design the optimal structure and layout of racks, networks, and power and cooling systems, as shown in Figure 2. In particular, the team can suggest layouts at the component, server, rack, and row level that help optimize power flow, air flow, and heat removal. The team can also size HVAC and power distribution configurations based on a careful assessment of specific power consumption and cooling requirements.

When working with organizations to co-develop custom solutions, DCS system architects leverage industry-leading Dell technology and manufacturing expertise to deliver hardware that can meet specific hyper-scale computing needs with outstanding energy efficiency and functionality. Moreover, because DCS has full access to the outstanding Dell supply chain, manufacturing infrastructure, and global sales and service organization, DCS custom-built hardware can be manufactured and delivered quickly, and offer the same world-class warranty service as general-purpose Dell™ servers and storage.

CUSTOMIZING SERVICES FOR RAPID DEPLOYMENT AND HYPER-SCALE SUPPORT

DCS can also help organizations enhance efficiency by helping reduce deployment time and simplify maintenance. To do this, DCS offers a comprehensive range of services tailored to meet the service needs of hyper-scale data centers. Key DCS service offerings include the following:

- **Single point of contact:** Each organization is assigned a dedicated services account manager to help ensure that issues are properly routed and rapidly resolved.
- **Fully integrated racks:** By installing fully integrated racks, DCS can minimize the time required to deploy a hyper-scale data center. Deployment services include integrating the hardware solution into a rack and setting up cabling off-site, delivering it to the data center, unpacking it, connecting it to network uplinks and a power source, inspecting it, powering it up, conducting "burn-in" testing, and removing trash. DCS can deploy the solution into an existing rack in the data center as well.
- **On-site parts:** Dell-owned parts can be inventoried on location to enable rapid systems repair. Service technicians can easily scan parts in and out of inventory, and parts can then be replenished automatically. Alternatively, parts can placed off-site near the organization.



- Optimize layout and airflow paths and minimize air consumption, bypass, and recirculation
- Utilize techniques such as cold or hot air containment based on environment
- Make cable management, raised floor and drop ceiling plenum, and return path to the HVAC system as unrestrictive and efficient as possible
- Use closed loop-feedback to drive HVAC and power distribution units
- Size power distribution based on an accurate understanding of power consumption requirements
- Balance all parameters critical to the operational efficiency of a data center

Figure 2. *A data center facility and nodes are an intrinsically linked ecosystem*

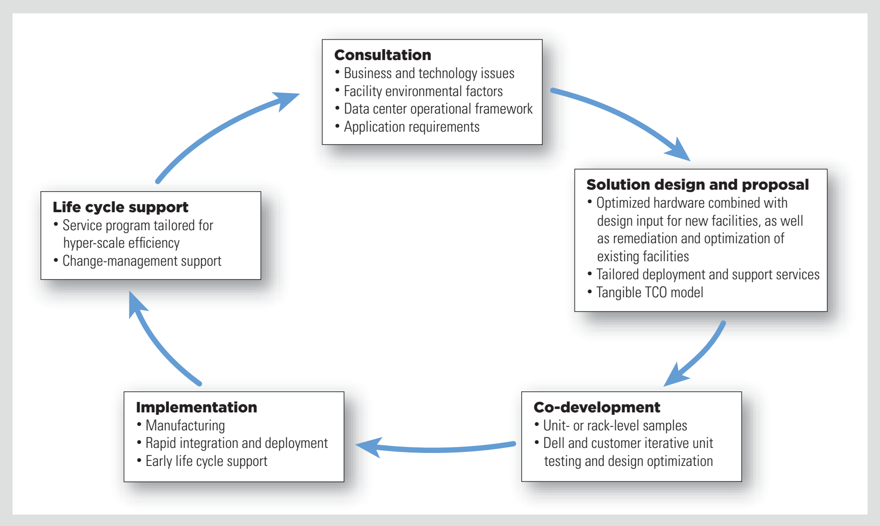


Figure 3. *DCS offers optimized solutions throughout the life cycle of a hyper-scale data center*

- **Node-level exchanges:** Fully tested and burned-in whole units can be located on-site, which allows an impaired system to be replaced immediately while it is repaired outside the rack, helping increase system uptime. Alternatively, whole units may be located off-site and delivered on a scheduled basis at the same time that impaired units are removed for repair at the off-site location.
- **On-site hardware troubleshooting and repair:** To help organizations rapidly troubleshoot and repair problems, DCS can deploy highly skilled personnel on-site either periodically or full-time.
- **Custom maintenance and support services:** To streamline ongoing support, DCS also provides several maintenance and support service options that can be customized to meet an organization's specific needs.

ENHANCING EFFICIENCY AND PERFORMANCE IN AN ENVIRONMENTALLY FRIENDLY WAY

In a hyper-scale computing environment, efficiency is a key source of business value. By offering optimized solutions throughout the life cycle of the data center, DCS can help organizations dramatically improve efficiency and lower the TCO of their hyper-scale data centers, as shown in Figure 3. In particular, by offering data center design consulting, designed-to-order hardware, and customized service offerings, DCS can help organizations rapidly deploy and cost-effectively maintain a hyper-scale data center that is designed for lower acquisition costs, lower power consumption, higher HVAC efficiency, lower network infrastructure costs, and lower deployment and management overhead than a traditional data center.

Moreover, by optimizing for energy efficiency at all levels of scale, Dell also helps ensure that DCS-designed data centers are environmentally friendly. Plus, DCS offers an optional carbon offset program in partnership with Carbonfund.org, which plants trees to offset the estimated "carbon footprint" of a participating organization's data center.

Jimmy Pike is director of system architecture and distinguished engineer in the Dell DCS Division. He is responsible for the strategic system architecture of the enterprise product line, and has more than 25 years of experience in the server industry. He was previously employed by Intel as part of its Enterprise Product Group, where he served as the director of telecom server development, director of server architecture, and director of engineering. He has received numerous patents through extensive work in the area of symmetric multiprocessing for both large and small systems.

Ty Schmitt serves as principal thermal and mechanical architect in the Dell DCS Division. He previously served seven years as the senior manager for the enterprise thermal, acoustical, and mechanical architecture division at Dell, and seven years as an enterprise mechanical architect and lead mechanical engineer responsible for the mechanical development of 10 enterprise and storage platforms. Ty has a bachelor's degree in mechanical engineering from Texas A&M University and holds 37 U.S. patents.

Frank Frankovsky leads the Solution Architect team in the Dell DCS Division and was a key leader in the inception and launch of DCS. Since joining Dell in 1991, he has held leadership roles in Dell enterprise marketing and technical sales organizations. Frank has a B.B.A. in Marketing from Stephen F. Austin State University and completed the Executive Management Program at the Jesse H. Jones Graduate School of Management at Rice University.

Todd Brannon is market development manager for the Dell DCS Division. Todd joined Dell in 1995 as a founding member of the System Performance Analysis lab and has since held a variety of enterprise marketing roles, including leading the launch of the Dell server and storage businesses in Japan. Todd has a B.S. in Electronic Media from Syracuse University.

MORE ONLINE
DELL.COM/PowerSolutions

QUICK LINK

Dell Cloud Computing™ Solutions:
DELL.COM/CloudComputing