MOOR INSIGHTS & STRATEGY

Rack Scale Server Segmentation 2014

The question is," said Alice, "whether you can make words mean so many different things?"* [Through the Looking-Glass, and What Alice Found There](#)*, Lewis Carroll, 1871

Executive Summary

Moor Insights & Strategy has published commentary and detailed analysis describing the changing scale-out datacenter market for almost two years. In that time, datacenter server hardware segments and descriptive terminology have become fairly complex.

Service oriented datacenters must balance their mix of compute capability and power consumption within an increasing density per cubic meter of datacenter space. Optimal efficiencies with increasing density can only be achieved through advances in network architecture and specialized computing. There is no way to meet the future datacenter needs without changing our approach away from individual server chassis and toward rack and datacenter level architectures.

We describe emerging opportunities by introducing a simple top-level taxonomy:

- Small Scale
- Virtualized
- Hyperscale
- Hyper Virtualized

Our segmentation is based on relative integer performance of full racks of server nodes and with the effective north-south (N-S) network bandwidth delivered to those server nodes from a top of rack switch (TOR). See the “Methodology” section below for detail. Future versions will attempt to incorporate east-west (E-W) network fabric impact to rack level performance and eventually rack level power consumption metrics. But we note that, because Hyperscale systems focus on reducing N-S network traffic, they are clearly differentiated from Virtual and Hyper Virtual without explicitly measuring the impact of E-W fabrics.

If ARM server licensees wish to be competitive, they must raise themselves out of the Virtual competitive pack and solidly into the Hyperscale range. Lower power consumption is a tough sell, at best. The performance impact of E-W fabrics will also be hard to sell in the short-term. Better density means better performance per rack for a set of high value workloads, period.

We expect more from OpenPOWER. POWER-based racks must land in Hyperscale as a baseline, but Hyper Virtual is the real target to differentiate from ARM-based servers.

Please see the end of this research note for references to our previously published opinions on future datacenter architecture.

Traditional Datacenter Server Segments Fail At Scale

The question when one assigns meaning to words is, quite simply: *which will be the master?*

Our datacenter infrastructure industry has built up a taxonomy of chassis types over the last few decades, ever since servers shrank smaller than individual racks (remember minicomputers?). These taxonomies are aimed at footprint planning; they are designed to tell buyers how much rack space a server (until recently running a specific workload or a single application) requires. But they don't adequately address system level performance in a modern datacenter, because each server is assumed to be independently configured for its own application.

Traditional quarterly reports for server shipments record information such as:

- **How many processor sockets or chips share coherent memory?** This is recorded as a 'P' (processor) or recently 'S' (socket) count. Until recently it also reflected shared system resources, but blades bent that definition and new architectures break it. Reports detail the number of processor cores, their speeds, *etc*. but do not give the same treatment to network bandwidth and topologies, or storage subsystem architecture and performance. Compute, networking, and storage are no longer cleanly divisible for many modern workloads.
- **What is the chassis type?** Does a server fit in a rack, and if it does, does it contain multiple server boards or nodes, and do they share resources? Until recently this meant separating blade chassis from everything else. But new architectures span a range of shared power and cooling infrastructure to large scale I/O and coherent memory architectures.
- **If a server fits in a rack, how large is the enclosure in terms of 'U' height?** (Each U is 1.75 inches.) This is an important question in markets where customers buy individual servers, one at a time, and where individual servers are replaced immediately if they fail. But the unit of installation for many datacenters is now at least one entire rack, and "fail-in-place" (FIP) maintenance cycles are becoming more common.

This taxonomy is aimed at traditional commercial IT buying practices: small, medium, large, and humongous organizations buying servers for internal consumption to reduce operating friction and increase agility. CIOs and enterprise IT organizations read the reports. Practically speaking, this means that server manufacturers design within existing feature categories if they want to be reported within the existing categories.

At rack scale, all of the current means of categorizing server hardware sales fail. Service oriented customers who deploy datacenter infrastructure to directly drive external business buy differently than traditional IT customers who serve internal needs as an OPEX line item (*i.e.*, "overhead").

We also considered terminology already in use for describing energy efficient servers:

- Gartner uses "extreme low-energy" to characterize many new architectures, but everyone is trying to reduce their energy consumption. At this point compute power consumption is already outweighed by network and storage power consumption. It's not a significantly differentiating characteristic for describing different kinds of compute architecture.
- IDC uses "density-optimized" to describe servers sold into heterogeneous, cloud, hyperscale, and similar large scale datacenter customers. But they rely on server manufacturers to brand servers as density optimized, which leaves the definition open to vendor interpretation. Similarly Forrester uses "dense" or "ultra-dense" terminology.
- Intel and several OEMs use "microservers" as a characterization of using smaller or less capable cores or SoC designs. It is also used to describe small business towers and non-datacenter form factors. We believe that thread performance can be better described than by using today's processor core design attributes to describe tomorrow's SoC features and performance attributes.

Almost universally we hear "virtualized" and "hyperscale" referred to as markets, and the concept of "hyper virtualized" is starting to surface as a market descriptor (and not as a reference to Microsoft's Hyper-V product).

Table 1 describes the mapping we created from readily obtained system physical and performance characteristics to these market segments, based on how the server is designed and provisioned.

Table 1: Moor Insights & Strategy Rack Scale Server Segmentation

Small Scale	Local provisioning for a single application with limited dataset and users or a small office / branch office. Includes tower chassis, tower chassis designed for optional rack mounting, and other stand-alone form factors.
Virtual	Virtualized enterprise IT provisioned as chassis but now often deployed at rack scale.
Hyperscale	Scale-out, service oriented IT provisioned in units of racks at row to multiple site scale.
Hyper Virtual	Virtualized enterprise IT provisioned as chassis but at higher densities and with much more network bandwidth. Intended for multiple site private and hybrid cloud usage models. We include workloads like analytics and High Performance Computing (HPC) in this segment, and some service oriented datacenters buy this gear in addition to Hyperscale.

Segmentation Considerations

We drive our market model with the relative performance of server nodes and with the effective N-S network bandwidth delivered to those server nodes. It is a simple model for understanding both the compute density of a rack and the ability to move data to and from server nodes.

- The number of processors or sockets per server node is only one dimension in determining the coherently connected thread capacity of a socket. We chose a flat model of node performance. It is not perfect, but it is quantifiable using publicly available sources and has a reasonable level of relative accuracy for most of the systems we included.
- We chose to segment based on integer processor performance only. We believe that HPC and analytic workloads will diverge from other workloads in Hyper Virtual when we do account for floating point performance. But for the vast majority of non-analytic workloads, floating point performance is not a purchase consideration. We also do not account for specialized accelerators, such as graphics acceleration for Virtual Desktop Interfaces (VDI).
- The location of storage (DAS, distributed, NAS/SAN) and first level network switches can be external or internal to the chassis or rack. Their placement is not architectural on a broad scale, it is an implementation decision. We measured the bandwidth from each node to TOR based on publicly available vendor specifications. To normalize our comparisons for volume buying, we solely used configurations with 1Gbps and 10Gbps Ethernet, and we ignored configuration options for InfiniBand, Fibre Channel, and other specialty interconnects. These interconnects will make a difference when we look at Hyper Virtual workloads such as HPC.

For this initial cut at our segmentation, we did not include three important elements of emerging datacenter infrastructure:

- Rack level E-W fabric architectures have yet to quantify their performance benefits to specific workloads and applications. We believe these fabric architectures will have an impact, but we are in the early market phase, and it is a challenge to vendors and customers alike to quantify E-W fabric benefits.
- Reduced power consumption (and by extension cooling costs) are near the top of everyone's list of critical features. We ignore rack level power consumption in this document but intend to figure out a way to include it in the future. Our challenge is that, over the last few years, processors have become much more efficient at consuming power—and that makes memory, network, and storage components much more of an issue at a rack level. This is a general challenge for the entire datacenter supply chain. We are not confident that simply adding lower power processors results in significantly lower power consumption per unit of rack performance.
- Pricing is an important component of service oriented datacenter purchasing and a daunting challenge. We do not address pricing here and probably won't in the future. Large datacenters are buying components and semi-custom server, storage, and networking gear at a scale that makes volume pricing difficult to assess and therefore difficult to compare with any accuracy. We do note that

vendors brand and price some systems as “dense” specifically to sell them into Hyperscale markets, but our charts show that they do not provide a measureable rack level performance per bandwidth advantage over competing systems.

We show Small Scale in our charts but exclude it for the rest of this paper, because these form factors are, for the most part, no longer sold into datacenters, and especially not into scale-out datacenters. They are sold to small and medium business (SMB) customers for use in home and small offices, branch offices, *etc*. They tend to be desktop derivatives, and the market is small and declining. Likewise, we do not address tower chassis that can be slotted sideways into a rack. This is a tremendously inefficient form factor that is also small and declining. Both are adequately accounted for in existing market research.

We also do not directly address high-performance computing (HPC) in this taxonomy. We include it in our Hyper Virtual segment, because HPC and HyperVirtual share a lot of “overprovisioning” characteristics when compared to the other segments. HPC is a market defined by analytic usage models, and we do plan on addressing HPC in the future—as it is colliding with Big Data analytics, and the result will be a profound impact to both markets. HPC is tracked as a workload, and so it is included in the form factor shipments tracked by secondary market research. We believe that HPC will also undergo significant transformation by the end of this decade. In the process, HPC deployments will most likely end up looking a lot like a hyperscale datacenters, or they will be subsumed by them.

The Data and Market Dynamics

Figures 1 and 2 show our high level segmentation summary. We note that we were surprised by two things:

1. There are very few server chassis that enable differentiated rack level Hyperscale performance (given our inability to quantify the impact of E-W fabrics). This may explain some of the challenges vendors are having in differentiating their “dense” chassis designs at customers who can do their own math for whole rack configurations.
2. Hyper Virtual separates from the other segments much more cleanly than we thought we would see. It has well-defined boundaries and value propositions.

Figure 1: Segment Map of Server Models

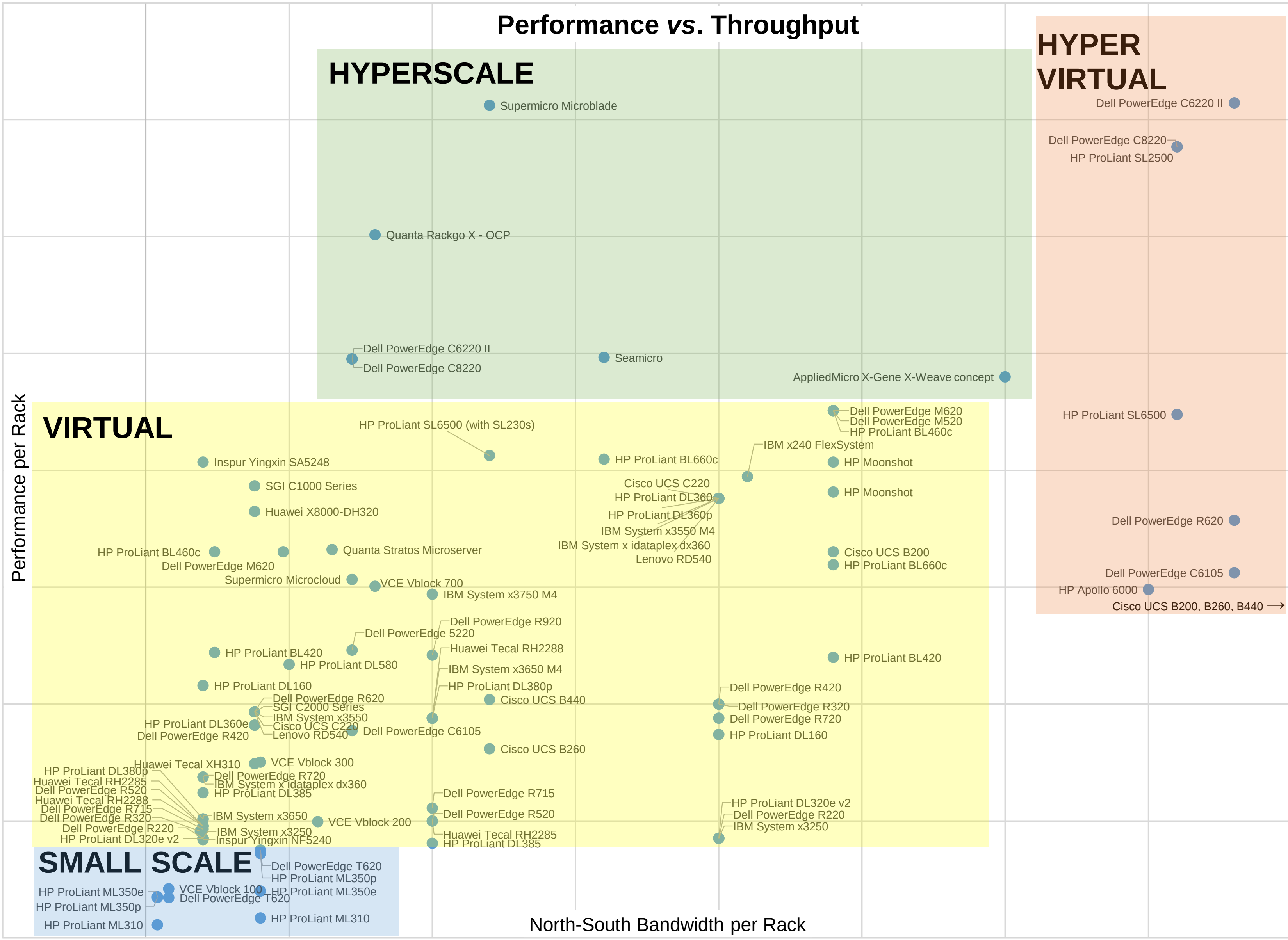
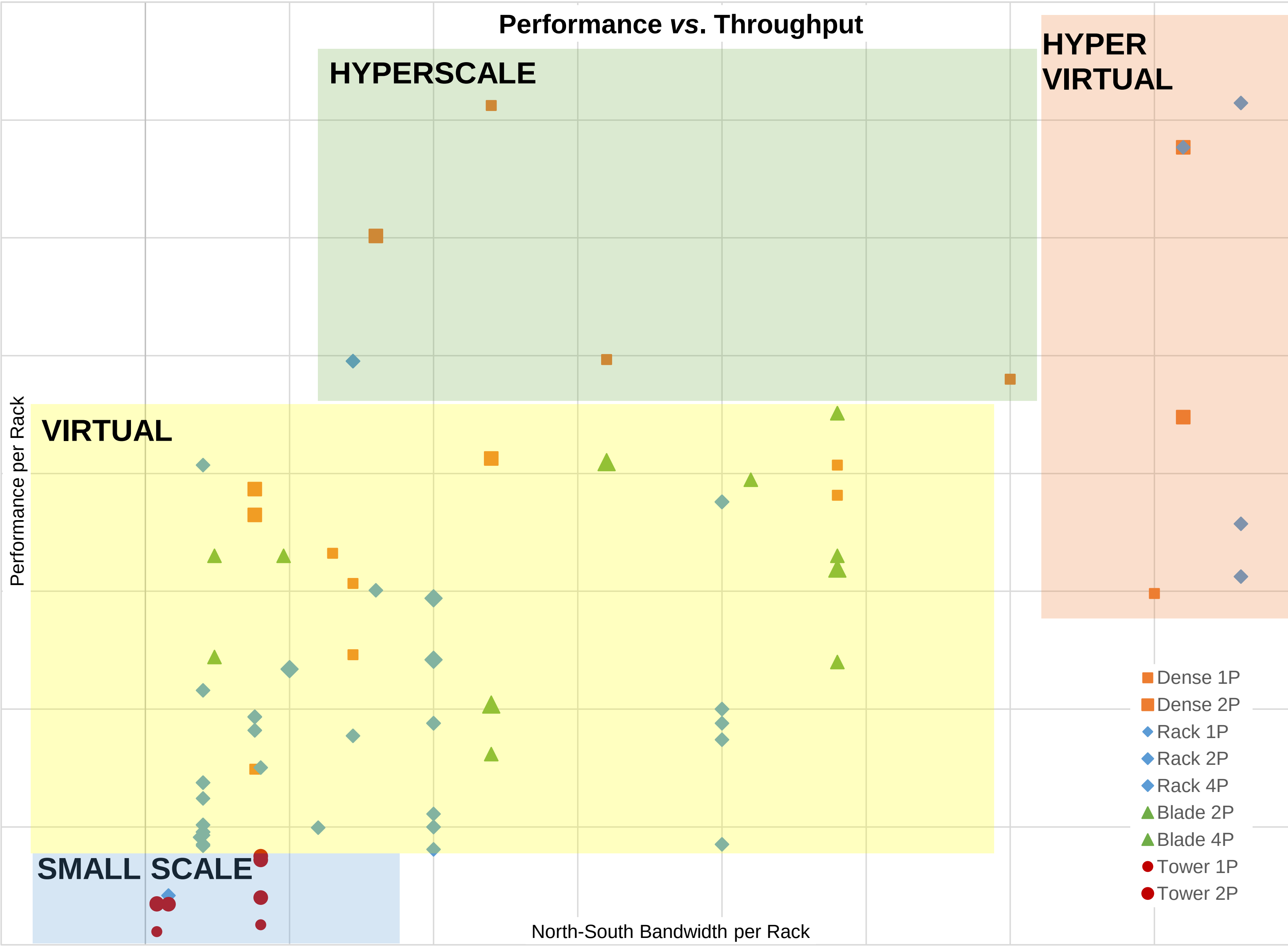


Figure 2: Segment Map of Form Factor and Sockets per Node



Note that many models are shown twice; we included both a volume configuration and a highly-provisioned configuration to highlight design constraints.

We attempted to include a dense HPC rack level design, HP's new Apollo 8000, but it broke both scales by extending the graph far up and to the right. We plan to address HPC in the future, as mentioned above.

We noted a few interesting market dynamics that are highlighted in these charts.

- First is the scarcity of true Hyperscale architecture. Our charts highlight the tension in balancing compute density with TOR network traffic. Based on pricing dynamics outside of the scope of this paper, many Hyperscale buyers are buying "dense" chassis that balance compute and bandwidth within the Virtual segment. We believe they will move to truly dense Hyperscale architectures when those Hyperscale architectures bump up their rack level performance at a reasonable cost and move buyers away from looking only at the chassis. This will clearly differentiate Hyperscale systems from the Virtual melee. We believe that aiming for differentiation in the Hyperscale segment is the real opportunity for both ARM's server licensees as well as the OpenPOWER consortium.
- A cluster of three blade systems at the upper-right corner of Virtual (two Dell chassis and one from HP) could probably be included in Hyperscale if they were not shared infrastructure blades. They are close, but service oriented datacenters tend to avoid blade chassis and their unique management subsystems.
- We excluded aggressive configurations of Cisco UCS blade models B200, B260, and B440. They extended our graph too far to the right and compressed the rest of the graph, but they are clearly in the Hyper Virtual segment. This might explain some of Cisco's success in the blade market.
- Similarly, Dell's PowerEdge C6220 and C8220 sport configurations that are clearly in the Hyperscale segment. They clearly contribute to Dell being the leading OEM in service oriented datacenters.
- HP's Moonshot chassis using Intel Atom and AMD Opteron processors (both on the low-end of x86 server socket level performance) are solidly in the Virtual range. This is where HP has them positioned: Atom for dedicated hosting (but capable of virtualization) and Opteron for hosted desktops. Future, more aggressive cartridges should move Moonshot into Hyperscale territory.
- HP's new Apollo 6000 System separates from HP's Moonshot quest for Hyperscale and effectively targets Hyper Virtual.
- AppliedMicro's X-Weave enabled rack architecture, using their X-Gene 2 server SoC, starts to edge over toward Hyper Virtual. But, similar to the three blade systems above, the Hyper Virtual market is an x86 market except for a few high end RISC "engineered systems" models. It is a novel Hyperscale performance point for the ARM instruction set.

From our charts it is clear that some of the traditional form factors are catching up to hyperscale as they implement robust networking with four-port 10Gbps Ethernet per 2P server node. Conversely, hyperscale laggard systems are showing up in Virtual (Quanta Stratos, Supermicro), because they implement only 1Gbps per node. 10Gbps per node has emerged as the current state-of-the-art for mainstream star topology network connectivity. For example, Quanta Stratos uses 10Gbps Ethernet as its standard network interface, but the network port to processor socket count is low, and LAN on Motherboard (LOM) options are not available, which impacts that model's per server node bandwidth.

Closing Notes

Segments are useful in understanding design points. Do not focus on segment names. Focus instead on how segments are defined and differentiated. We believe that the datacenter infrastructure market needs a rack level segmentation model that is independent of prior compute, storage, and networking product silos. And it needs to be independently measurable, regardless of brand or market positioning is applied to the system.

Our segmentation does not represent the end of server evolution. We expect that after a few more years we will have to sub-segment Hyperscale as it becomes a more dominant force in the market. We believe that service oriented datacenter operators will buy specialized architectures for different large scale workloads. And we believe that branding and technology differentiators will emerge for infrastructure vendors to address these new market opportunities.

For many potential Hyperscale buyers, computing, storage and networking are no longer separate markets. Lines between product categories will blur within enterprise IT, as public cloud architecture influences private cloud architecture and provisioning.

Moor Insights & Strategy References

- Blog: Tech Giants: Move to Specialized Computing or Die
- Blog: SC13: Two Oxen vs. Two Hundred Chickens
- Blog and Research Note: Decoding Intel's Silicon Photonics and Rack Scale Architecture
- White Paper: AMD's Disaggregated Servers
- White Paper: HP Moonshot: An Accelerator for Hyperscale Workloads
- White Paper: NextIO Enables Multivendor Converged Datacenters
- Research Note: Calxeda: Rack Trumps the Chip
- And a host of blogs and articles covering the potential for ARM in hyperscale datacenters

Methodology

Our charts are useful for highlighting gross architectural and high level design differentiation; this is their only intended use. We tried to make our charts as accessible and as replicable as possible, and so we selected proxy data that we could normalize across a wide range of server models for a balanced architectural comparison without boiling a proverbial ocean.

To be explicit, our charts should not be used for competitive positioning - they do not contain model-specific performance benchmarks and we do not claim a level of accuracy suitable for making buying decisions for one model over others. Please do not send us competitive data "proving" that a model should be somewhere else within a segment. If you believe a system should be in a different segment on our chart via upgrading architecture, processor or networking options, then please see the section "Want to Add a System to Our Charts?" below.

We created our charts using publicly available data for integer processor performance and for chassis network bandwidth to top of rack. We made simple assumptions for how much rack height would be consumed by infrastructure like mid-rack and TOR switching, external power supplies, etc., and otherwise modeled a packed hypothetical rack full of each server SKU. Anyone can create similar charts, and we believe that everyone should...but please be rigorous and inclusive when you do.

We believe that our charts cover at least half of 2013 worldwide unit shipments in every existing server chassis category except for "dense" form factors. There will be variation by region as we used English-language websites as our source. There are many models and configurations sold into public cloud customers that we did not have enough information about to adequately model. If we missed an important model or if you have a rack configuration that moves your model to a different segment, please let us know (see the section **"Want to Add a System to Our Charts?"** near the end of this paper).

The important thought behind these charts is that the datacenter infrastructure industry is moving away from deploying individual chassis within racks and toward deploying entire racks at a time. Our charts illustrate that some new server models are clearly separating from established design norms to attack this new market reality head-on.

Want to Add a System to Our Charts?

Or provide an update to a system already in our charts? Please send the following information to serverconfig@moorinsightsstrategy.com, and when we next publish these charts we will include the new data point. Please note that if the sent data puts the system it describes into a crowded section of the Virtual segment then it will not be readable. We assume that all information sent is not NDA nor is it embargoed; we assume that it is intended to be printed in public forums immediately.

It will help us if you send two versions: a midrange configuration and a fully loaded configuration. It will also help to send a complete rack architecture, if you have a specific architecture in mind. Here is the minimum information we'd like to see, at a chassis level:

- **Manufacturer**: You
- **Brand**: product line
- **Model**: unique identifying sequence of alphanumeric characters
- **Processor**:
    - Manufacturer
    - Model
    - Speed
- **Form Factor:**
    - Enclosure U height
    - Number of processors per enclosure
    - Node SMP count, such as 1P, 2P, etc.
- **Network Uplinks:**
    - Number of ports per enclosure
    - Network controllers
        - Number of controllers
        - Number of ports
        - Type of controller
        - Adapter card options: make and model
- **Product Web Page/Site**: URL (Used to verify the above information, without it we will not publish a data point.)

Important Information About This Paper

Author
Paul Teich, CTO & Senior Analyst at Moor Insights & Strategy
Gina Longoria, Analyst at Moor Insights & Strategy

Review & Publish
Patrick Moorhead, President & Principal Analyst at Moor Insights & Strategy

Editors
Scott McCutcheon, Director of Research at Moor Insights & Strategy

Inquiries
Please contact us here if you would like to discuss this report and Moor Insights & Strategy will promptly respond.

Citations
This note or paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

Licensing
This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

Disclosures
Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness or adequacy of such information and shall have no liability for errors, omissions or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.