AMD Disaggregates the Server, Defines New Hyperscale Building Block

Fabric Based Architecture Enables Next Generation Data Center Optimization

Executive Summary
AMD SeaMicro’s disaggregated server enables large and small data center operators to optimize their hardware performance profile for specific applications. Today’s modular x86 servers are compute-centric, designed as a least common denominator to support a wide range of IT workloads. Those generic, virtualized IT workloads have much different resource optimization requirements than hyperscale and cloud applications. They have resulted in a “one size fits all” enterprise IT architecture that is not optimized for a specific set of IT workloads, and especially not emerging hyperscale workloads, such as web applications, big data, and object storage.

The server ecosystem is fragmenting as hyperscale workloads become the high volume market value drivers. Cloud and hyperscale infrastructure vendors have been shifting their hardware designs away from modular x86 form factors for the past few years as their buying scale has enabled them to investigate new rack, row, and data center levels of workload optimization. Facebook is leading the Open Compute Project to further optimize rack-level hardware designs for their needs, but their disaggregated rack architecture concept will only be efficient at a large (and multiple) data center scale – hyperscale data centers will have to be designed around it to take full advantage.

AMD shifted the focus from previous compute-centric IT architectures with an innovative disaggregated architecture that scales from small to hyperscale data centers. This approach also brings disaggregation to IT buyers who don’t yet host massive social media sites. Businesses and web giants (even aspiring not-so-giants) can more cost effectively and power efficiently tune applications to balance compute, storage, and network demands within the SeaMicro server’s 10U rack increments.

Datacenter Market Dynamics
The familiar modular x86 server drove a transformative wave of IT consolidation by commoditizing server CPU cycles – it became the fundamental building block of modern IT data centers by creating server design standards and a small set of “dominant designs” that enabled vendors to lower costs dramatically. Implicit in this transformation is the assumption that IT must continue to perform the same functions they had been performing, but with lower infrastructure costs and tighter budgetary constraints. The result is difficult to differentiate, converged IT server infrastructure designed to run a wide range of workloads with a high degree of compatibility, assurance, and management tools – a “one size fits all” approach to workloads.
Three mega-trends have recently emerged that counter the IT status quo and increases in hyperscale workloads: IT as a business partner, “bring your own device” (BYOD), and Business Intelligence (BI) coupled with Big Data. Foremost is that IT has become a business partner in the C-suite. Non-business-differentiating workloads are top targets for offloading to private and hybrid clouds and for potential outsourcing to public cloud services. At the same time, the BYOD trend is forcing IT to make trade-offs for different classes of data security, and together they are accelerating the trend toward offloading and outsourcing hosted applications. As “traditional” IT workloads migrate to hosted services, a new generation of BI is emerging. Companies are instrumenting their business process to generate real-time insight into their operations – a combination of machine-to-machine and Big Data workloads. That instrumentation can take the form of new software metrics, new physical sensors, and combinations of software and sensors. The result is an explosion in both stored data and analytics techniques. These trends are driving a new generation of data center economics and workload requirements.

**Hyperscale Challenge**

In a modern scale-out data center, efficiency is an overriding operational imperative. While virtualized, interchangeable modular x86 server chassis enabled IT shops to consolidate workloads, when they are deployed at scale they are not optimized for any specific workload. Each workload class has a general profile for a mix of compute, storage and networking resources. As importantly, each application is coded to a unique combination of design choices, coding environments, and runtime deployment choices that also impact its mix of resource demands. Server applications today balance compute performance with memory and storage sizing plus network interconnect latency and bandwidth.

After general workload attributes, specific application software design and deployment choices, there are two major influencers for optimizing an application’s mix of resource requirements – scale and hardware modularity.

Scale matters because running a workload across a set of resources places different demands on those resources as demand scales. Sub-rack, rack, multiple racks, and rows of racks have different balance points to run the same application, as no application can perfectly scale its resource demands. Right-sizing resources for expected demand impacts both capital and operational expenses.

Modularity refers to the ability of different hardware architectures to reduce operational expenses as resources need to be replaced or upgraded. Data center operators want to replace and upgrade resources independently for three reasons: to isolate the impacts to their operational environment, to reduce the cost of those replacement parts, and to upgrade performance of individual components as those performance upgrades become available. For example, today compute and network resources are designed onto the same modular x86 server motherboard, so upgrading one means paying to
upgrade both. These modularity considerations substantially impact image stability, thermal management, and a host of other operational parameters.

**Open Compute Disaggregated Rack**

Facebook’s disaggregated data center proposal to the [Open Compute Project](http://www.opencompute.org) (OCP) separates compute, storage, and network resources via fiber-optic cabling and high-speed local switching. In their eventual end-state, a data center might implement a tray full of processor cards and a separate tray containing storage – the design supports populating a rack full of only one type of tray, for instance a rack full of compute trays. Network I/O will be virtualized through local switches, and behind them will be a simpler network interface to data center core switches. Eventually compute resources will be disaggregated from their memory, so that the rack full of processor trays has no memory chips or modules on the processor cards, they will be located in a separate tray full of memory.

This radical disaggregation is a break from current data center architectures and comes at a cost – the introduction of network protocol overhead, switching latencies, and bandwidth constraints between devices that were previously directly connected. The network that enables disaggregation directly adds cost and performance penalties, and will impact every aspect of data center operations – provisioning, management, maintenance, and a host of other procedures.

Over the next few years, the move to disaggregated racks will depend on Ethernet switching silicon. That silicon supports switching at Ethernet bit rates, so a high performance switch that supports 64 x 10GbE or 16 x 40GbE ports would be limited to 640Gbps non-blocking throughput.

Data centers will have to be massive to see capital and operational returns from disaggregated infrastructure at this scale. The large web-based companies are already there. However, this architecture does not address operating at anything less than massive scale. It does not fit into many other business models, data retention strategies, or data center architectures.

**AMD SeaMicro Disaggregated Fabric-Based Server**

AMD SeaMicro’s disaggregated server architecture is designed to address smaller scale to hyperscale data centers and focuses on balancing compute and storage for specific workloads. SeaMicro Freedom Fabric is the core technology in the SeaMicro SM15000 server and the literal center of that 10U system. The Freedom Fabric is a single network mesh with a 3D torus topology and an aggregate throughput of 1.28Tbps, connecting a modular mix of compute, storage, and network resources. It is self-healing and provides multiple paths for redundancy. AMD says that it is at least twice as efficient compute and storage density than standard server architectures – that is due to their fabric architecture, which allows them to share many of the common components in traditional servers, such as power, fans and network I/O. The system
consists of a set of modular sub-systems which “bolt on” to that fabric. Those modular systems are its disaggregated compute, network, and storage resources:

**Up to 64 “C-Cards” (‘C’ for compute).**
A node is a processor or System-on-Chip (SoC). Each C-Card today contains up to four server nodes and up to 64GB of DRAM per server node, and has the equivalent of a full duplex 10GbE connection to the fabric, with fabric headroom for twice that in the future. A range of low power processors from both AMD and Intel are already supported (AMD Opteron 4300 Series, Intel’s Xeon E3-1260L and E3-1265Lv2, and Intel Atom N570), with more planned in the future. Processor cards are standardized so that they are completely interchangeable – physical, electrical, and thermal – and are accessible through the side of the chassis.

**AMD Opteron C-Card**  
**Intel Xeon C-Card**  
**Intel Atom C-Card**

**Up to 8 network interface modules.**
Each has a 32Gbps connection to the fabric. Each network interface module can be configured for either 8 x 1GbE (64 x 1GbE system total) or 2 x 10GbE (16 x 10GbE system total) external ports, plus each has 2 x 6Gbps SAS fabric extenders that support SeaMicro’s Freedom Fabric storage products, enabling a single system to connect with up to 5.44PB of external storage capacity. These network interface modules are accessible through the back of the chassis and can be easily replaced as new interfaces become available.

**Up to 64 standard SATA II 2.5-inch hard disk drives in the chassis.**
The SeaMicro SM15000 server’s drive bays are located on the front of the chassis. Each drive has a 6Gbps connection into a storage card, 8 to a card, and like the C-Card each of the storage cards has the equivalent of a full duplex 10GbE connection to the fabric, with available future fabric headroom. With up to 8 storage cards per chassis, the maximum internal storage capacity is currently 64TB (1TB per drive), which can be extended externally with the additional 5.44PB mentioned above.

Drives, processor cards and network interface modules are hot-swappable, drives from the front of the chassis on the cold aisle and network interface modules from the rear. To add or replace a processor card, slide the chassis forward from the front of the rack, remove two screws on the side panel on both sides of the chassis to gain access to the
cards. A complete upgrade of processor cards for all 64 servers can be completed in about 15 minutes. Unlike the OCP architecture, infrastructure upgrades can be done without building a new data center, or even having to upgrade an entire server or rack.

SeaMicro SM15000 server

AMD’s SeaMicro server customers have the option of bypassing top-of-rack switches to network directly to end of row or core switches. Their solution simplifies rack architecture, eliminating hundreds of cables and other components that are redundant in its architectural context, for a more efficient and simple operational environment.

AMD SeaMicro’s disaggregated chassis enables any data center to balance their applications at a smaller scale and much lower performance and cost overheads than the disaggregated rack approach. As important, SeaMicro’s products are available now and easy to experiment with to discover optimal resource mixes for individual applications, whereas the OCP proposed disaggregated server is not yet on the market.

SeaMicro’s architecture also addresses repurposing – when an owner wants to run a completely different application on SeaMicro infrastructure than it was originally deployed to address. Optimizing a SeaMicro chassis for a new application requires removing easily accessible resource components where they are not needed and adding upgraded resources to easily accessible slots where they are needed. Tuning hardware infrastructure does not get much simpler than that.

Conclusion
Several years ago, SeaMicro (acquired by AMD in 2012) reimagined the server from a performance and hyperscale total cost of ownership optimization perspective. They
have delivered a forward-looking architecture that overcomes the “one size fits all” constraint and enables buyers to right-size compute, storage, and networking resources for specific applications. AMD says that the end product reduces power consumption and doubles compute and storage density.

We are still in the early years of hyperscale and cloud build-out. Therefore, it is not possible to standardize on one architecture that serves all applications with the cost and power efficiencies that buyers are asking for, whether we’re looking backward at virtualized and consolidated IT or forward to a disaggregated rack architecture.

AMD’s SeaMicro server disaggregates rack-level server, storage, and networking building blocks into a highly configurable 10U chassis that is in the market and shipping today to hyperscale data center customers. That is important when comparing it to OCP’s disaggregated rack and other evolving solutions. Additionally, SeaMicro leverages best of breed compute, storage, and networking technologies and is not dependent on future rack-level network architectures.
Author
Paul Teich, Analyst at Moor Insights & Strategy

Editor
Patrick Moorhead, President & Principal Analyst at Moor Insights & Strategy

Inquiries
Please contact us at the email address above if you would like to discuss this report and Moor Insights & Strategy will promptly respond.

Licensing
Creative Commons Attribution: Licensees may copy, distribute, display and perform the work and make derivative works based on this paper only if Paul Teich and Moor Insights & Strategy are credited with a link to the original work.

Disclosures
Moor Insights & Strategy provides research to and has a consulting relationship with AMD. This paper was commissioned by and is sponsored by AMD. No employees at the firm hold any equity positions with AMD.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors.

©2013 Moor Insights & Strategy.
Company and product names are used for informational purposes only and may be trademarks of their respective owners.