



TidalScale

**7 Trends driving the Industry to
Software-Defined Servers**

The Death of Moore's Law. The Birth of Software-Defined Servers

It has been over 50 years since Gordon Moore saw that transistor density doubles every two years.

Over the decades, the interpretation of “Moore’s Law” has evolved to represent that the performance of microprocessors, and computers in general, is doubling every 18 months.

Amazing innovations in semiconductor technology, as well as microprocessor architecture including multi-core, hyperthreading, direct-attach memory and glueless symmetric multiprocessing (SMP), have played an important part in maintaining that Moore’s Law is an enduring given, a truth not open to debate.

For years, we saw performance increases in computers outpacing the needs of applications and databases. The resulting underutilization of compute resources opened the door for hypervisor-driven virtualization technology allowing a single server to be split into many, often dozens, smaller virtual machines.

We are now in a “perfect storm.” Data growth rates are exploding due to social media, mobile devices, and the Internet of Things (IoT) - a world where everyone and everything is connected. It’s a data tsunami. And, if that wasn’t enough, the compute performance improvements we are seeing have slowed to a relative trickle and are increasingly dependent the complex task of finding and exploiting parallelism in applications. Moore’s Law is dead.

The birth and rise of the Software-Defined Servers

Before you lament the passing of Moore’s Law, consider that something very good has emerged from the demise of this half century-old rule. That something is the Software-Defined Server.

TidalScale is the pioneer of the Software-Defined Server. Simply put, our technology allows one or more physical servers to be combined into a virtual server of essentially any size. The technology aggregates the CPUs, memory and I/O of all the connected systems - all with no changes to the operating system or applications and built on commodity hardware. The effective result is that we can deliver, today, what computers on their own might deliver a decade from now.

This is a breakthrough which counteracts the death of Moore’s Law, and then some. To learn more about the technology, view our video explainer or read my earlier blog on why 2018 is the year of the Software-Defined Server.

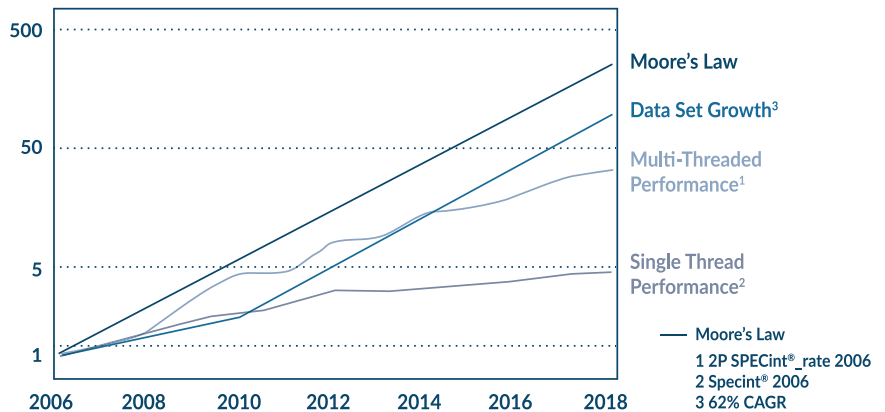
In an upcoming paper from TidalScale featuring research from Gartner, we’ll address how Software-Defined Servers are the missing piece in the Software-Defined Data Center.

7 reasons why Software-Defined Servers are necessary

For those who can't wait, we're offering this excerpt that lists seven technology realities that have given rise to the Software-Defined Server; note how the first three are directly related to the slowing of processor performance gains—the breaking of Moore's Law.

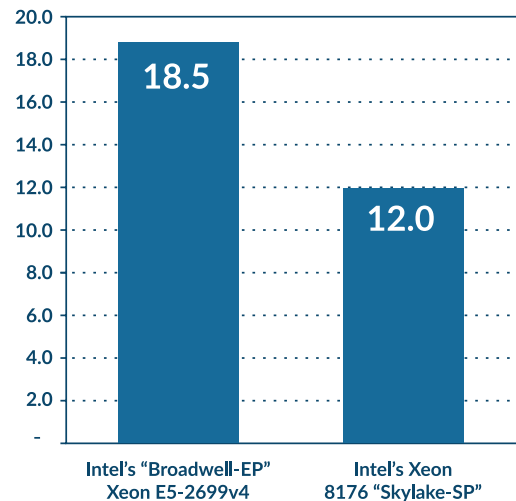
1. CPU performance increases are slowing overall.

The industry response to physical limitations on CPU frequency has been incremental improvements in microarchitectural improvements including hyperthreading, larger caches, and increasingly large (often impractical) core counts. The net result is a strong rate of improvement in integer and floating point performance per socket, but a much more muted improvement in single thread performance. If all applications were embarrassingly parallel, this might be fine. They are not. One law that is not dead is Amdahl's law which points us to single thread performance being a fundamental limiter.



2. Memory bandwidth per core is decreasing.

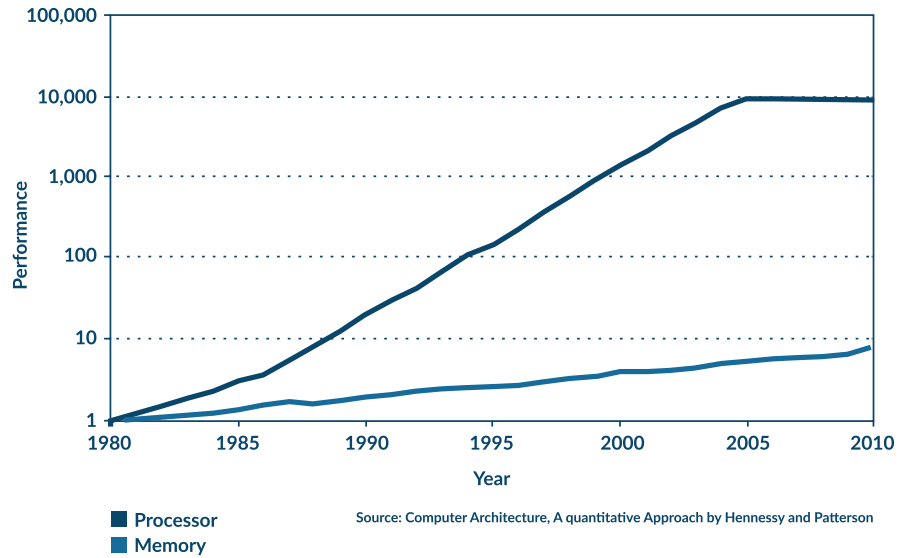
While core counts are increasing rapidly, physics is working against us in scaling memory bandwidth to keep pace. The shared parallel DRAM interface, overall limitations in processor pin counts, and motherboard layers have severely limited aggregate bandwidth per socket. Overall, memory bandwidth per core in recent years is increasing at just 23% per year. Bandwidth per core decreases in the Skylake processor generation. Intel's Xeon 8176 "Skylake-SP" shows only 12MB/sec bandwidth compared to the prior generation's Intel "Broadwell-EP" Xeon E5-2699v4, which sported 18.5MB/sec bandwidth.



3. Memory latency is getting worse.

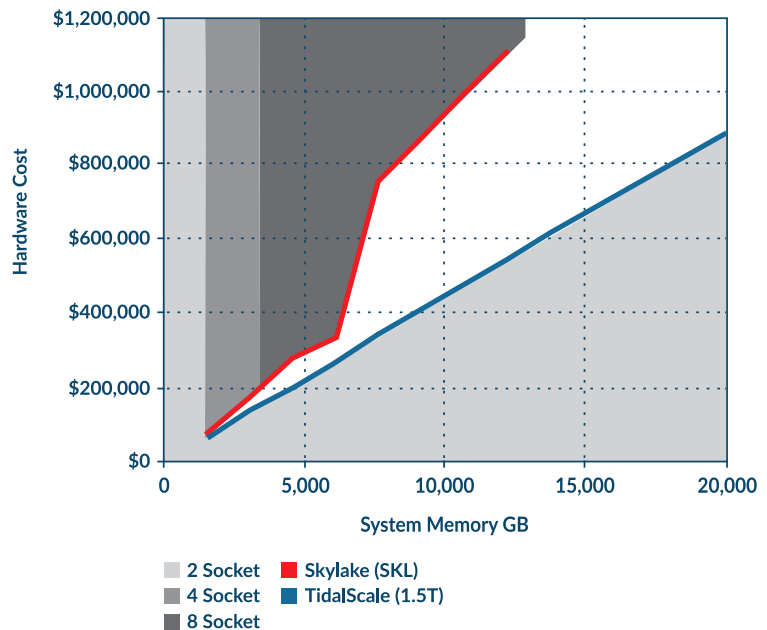
Flat to decreasing processor clock speeds with rapidly increasing core counts leads to sluggish increases in memory latency, despite heroic increases in hardware complexity. The result is a three-orders-of-magnitude increase in the performance gap between processor performance improvements and memory performance improvements that is caused by three factors:

- Slow rate of memory pin speed improvements
- Slow rate of increase of memory interface width
- Virtually unchanged DRAM cell cycle time



4. Sweet spot servers are compellingly cost effective.

The volume sweet spot price points for standard 2-socket servers are significantly more cost-effective compared to 4- or 8-socket servers. On a 3-year TCO analysis of a popular database application on a Software-Defined Server, standard 2 socket servers with cost-effective processors saved 50% of hardware costs and 61% of software licensing costs for a total of 56% in cost savings overall. A software-defined layer that can utilize standard two socket servers as building blocks becomes progressively more cost effective as system size increases. The advantages of using 2-socket servers as the basis for Software-Defined Servers actually increase in the SkyLake hardware generation where Intel starts charging a larger premium for big memory server configurations – it takes twice as many sockets to get to the same memory as the Intel Xeon® v4 systems. With SkyLake there are literally no processor versions that support 2 sockets that are in the High Performance Per Core category, as if customers who had 2-socket servers did not also want high performance.



5. New server designs increase system complexity.

In pursuit of performance improvements, hardware designers are introducing a confusing array of new interconnect technologies: CCIX, 3D Xpoint, GENZ, HCM, HBM, NVMe. The only thing that is clear about these technologies is that system administrators will be forced to move up a layer in abstraction: these increasingly complex node memory hierarchies strongly suggest the need for the existence of a simplification layer that can enable applications to adapt to whatever technologies are underneath them and extract the best performance automatically. In other words, these hierarchies call for software definition of servers.

6. Composable infrastructure responds to a desire for on-demand resources.

Many organizations are deploying composable infrastructure platforms because they recognize the business and IT advantages of on-demand resources. Compute resource disaggregation in next-generation rack scale hardware designs is a key part of this effort because it empowers business efficiency and agility. However, most of these platforms lack an easy way to aggregate the most critical resources in the rack, compute and memory, in a logical layer. The need for this missing logical layer is aggravated by shifting user expectations for auto-scaling infrastructures implied behind the push to serverless architectures.

7. Machine learning enables software-defined resources to dynamically adapt.

Machine learning technologies enable the possibility of software-defined systems tuning themselves to adapt to application needs automatically and dynamically. Affinities between resources like memory, CPU, network, and I/O, can be automatically and efficiently identified so that their virtual placement on physical hardware can deliver optimal performance, and these adjustments can be made orders of magnitude faster than anything currently available. This self-tuning capability is unique to TidalScale's .

These technology factors affect in-memory performance solutions and they come at a time when there is a rapid shortening of the time-to-decision window for business users.

The low latency requirements for analytics, ad placement, relevance assessment, fraud detection, medical diagnosis, supply chain management, IoT analysis, identification of persons of interest, and more are pushing the requirement for in-memory performance to an extreme.

It's not long before that journey leads to Software-Defined Servers.