Deploying HPC Infrastructure In Your Data Center

White Paper

Table of Contents

EXECUTIVE SUMMARY

The development, customer adoption, and deployment of HPC (high performance computing) infrastructure is a significant trend in IT. The use of parallel processing enables advanced applications and programs to be run reliably and efficiently, allowing engineers and scientists to solve complex computing and data-intensive problems. CPU, GPU, and FPGA servers are the ever-growing and more powerful workhorses being put to the task.

However, there are important implications for your data center when you deploy HPC in your existing environments. In this whitepaper we will describe a number of important capabilities your data center will require in order to make your HPC deployments successful.

ADOPTION OF HPC PLATFORMS

The world of IT continues to evolve at rapid pace in response to data; there are more and more endpoints collecting data at ever-increasing velocities, requiring real-time storage, modelling, and analysis to become insightful, actionable information points for companies to benefit from. Whether it's from the Internet of Things, websites, commercial transactions or other locations, the amount of data pouring into servers in data centers is growing at an exponential rate.

But companies need to make that data useful. Turning Big Data into meaningful insights has invited many participants and solutions into the HPC space, leading to impressive developments in artificial intelligence, deep learning, self-driving cars, digital manufacturing design and testing, and fraud detection among many others. Capitalizing on this data requires sophisticated, powerful computing platforms paired with Big Data platforms and data analytics tools; dense, power-hungry servers are ultimately required to make it all happen.

HPC server adoption continues to grow rapidly with IDC estimating the market at $23B in 2016 and growing to $31B in 2019 as HPC and Big Data platforms continue to provide business value with their insights. Companies of all shapes and sizes are all seeking to benefit from data, and HPC server and component manufacturers are enjoying this strong demand while making their infrastructure smaller, faster, and more powerful.

HPC PLATFORMS AND POWER CONSUMPTION

Developments in server processing are the keys to unlocking HPC performance and value. General purpose CPUs with only a few cores have ceded way to GPUs with thousands of cores and FPGAs hold further promise for processing efficiency. Ultimately performance per watt consumed is the critical benchmark, with manufacturers seeking to maximize computational capabilities in the smallest footprint requiring the least amount of energy consumption and expense.

Independent of improved server performance per watt consumed, HPC platforms draw significantly more power per server rack unit *in aggregate* than traditional servers because of the power requirements of the underlying processing components. Single GPU cards can draw 200-300 watts of power alone; in fact, the Nvidia GTX Titan X has been tested drawing power of 374 watts at max consumption. The latest Nvidia GPU workhorse, the Tesla P100, claims a max power consumption of 250 watts and in a single node can replace half a rack of commodity CPU nodes in terms of computational performance. More than 20 rack units have been consolidated into a 2 or 3U body!

The power draw of an HPC server is ultimately a function of many variables, but illustratively SuperMicro makes a 1U server that can house 4 Nvidia Tesla K80 GPUs, each of which can theoretically draw 300 watts. So, at maximum capacity, a single HPC rack unit could draw 1.2 kW of power. The Nvidia DGX-1, self-described as the "world's first AI supercomputer in a box," is a 3 rack unit system delivering performance equal to 250 conventional servers… and similarly specifies a maximum power draw of 3.2 kW in that 3U chassis, or just more than 1 kW per rack unit.

In summary, a good rule of thumb is that an HPC server will draw 600 watts per server rack unit, and as much as 1 kW. A traditional server might draw 100-200 watts in comparison. This fact has enormous implications for your data center, which we will discuss next.

Data Center Rack Power Requirements

	Legacy Server	Modern Blade	HPC Server
Avg. Power Draw Per Rack Unit (watts)	125	250	600
Data Center Cabinet Capacity (rack units)	45	45	45
Total Power Required Per Rack (kW)	**5.6**	**11.3**	**27.0**

IMPLICATIONS FOR YOUR DATA CENTER

HPC platforms promise the ability to harness and utilize massive processing power, but with regards to your data center, there's a rub. Simply put, most data center environments are not built or engineered to handle such powerful servers! As a result, if you are going to deploy HPC infrastructure successfully in any data center you must critically evaluate the facility's power, cooling, and scalability. Insufficient power, cooling and scalability can lead to poor performance, frustration, and potential I.T. failure.

We recommend a thorough evaluation of the following:

1. <u>Can your data center supply enough power per cabinet that HPC platforms draw?</u>

Power density is a measure of how much power can be supported by an individual cabinet. The standard capacity of a legacy data center is 3-5 kW of power per rack ("critical load"), which is a function of how much power and cooling capacity have been engineered per square foot when the data center was constructed. This power limitation is very important because of the power requirements of dense HPC servers.

As provided earlier, a typical HPC server rack unit will draw 500-600 watts and it is not uncommon for dense, GPU-heavy servers to draw 1 kW per rack unit. If a data center cabinet has 45 rack units, depending on the configuration of your rack (i.e. in terms of the mixture of HPC, storage, and networking units), it is very easy to provision an HPC rack requiring **20-30 kW** of aggregate power. Even just a couple of HPC 3U platforms can draw the entire power capacity of a legacy data center's rack! It is conceivable that 6 rack units out of 45 total, or *only 13% space utilization of available rack units*, could max out your legacy data center cabinet. This is terribly inefficient and wasteful.

5 kW in a rack is no longer sufficient, and if that is the limitation on power in your data center, then it means upon an HPC deployment you will have to spread your infrastructure across multiple racks, leaving tons of empty, wasted space (ie, racks only 10-20% full of gear). This leads to greater operational complexity (more racks, cabling and top of rack switching to manage) and costs (rack space, power, and cross-connects). In addition, if you are using Infiniband for cabling, distance limitations are critical in terms of performance so being forced to spread your servers across large areas and many cabinets may not be feasible to begin with.

- **In summary, you must understand how much power per data center rack your HPC deployment will require. Your data center should ideally be able to support at least 20 kW per rack to serve as a scalable home for your HPC hardware, and across the *entire* data center floor, not just a few "high-density" racks.**

2. Can your data center cool the resulting increase in heat generated per square foot?

Directly related to the power considerations above, your data center's cooling capacities will need to be evaluated; more power is now being drawn per rack unit, which means more heat is being emitted per square foot, and this can be challenging in legacy data center environments.

Data centers are provisioned to cool a certain expected amount of power drawn and heat generated per rack, more commonly referred to as watts per square foot. This cooling capacity is generally fixed and typical legacy environments cool 125-150 watts per square foot. This translates to the 3-5 kW of power per rack mentioned previously.

Understanding cooling limitations are critical because often times data center providers can allocate or bring more power to certain racks, but the question now becomes: can it be cooled? Because data centers are designed and engineered from the ground up to cool a certain amount of watts per square foot on average through the facility, they cannot magically increase cooling capacity to support a significant increase in heat per square foot.

Based on prior examples we know a full-populated HPC infrastructure rack can easily draw 20 kW or more. What assurances can be provided that the significant increase in heat generated in that rack will be adequately cooled? If the data center allocates cooling capacity for a higher-power rack, it is directly robbing that cooling capacity from another section on the data center floor. This is a zero-sum game; cooling capacities are generally fixed. So at some point a data center operator cannot cool higher and higher power draws per rack. This can create a major bottleneck for your HPC deployment and scalability, especially if your data center's floor space is highly occupied.

In particular, air-based cooling systems may not be robust enough to support HPC server heat loads. For a variety of reasons, air-based cooling systems tend to reach maximum cooling capacities per rack around 8-10 kW of load when measured on a scaled basis (meaning when there are more than just one or two racks drawing 10 kW of power). Hyperscale data centers being run by web giants and a few colocation providers are now utilizing liquid-based cooling to support these dense power requirements, given liquid is thousands more times efficient than air in its cooling capabilities.

- **We recommend that you evaluate and validate the designed and engineered watts per square foot capacity in your data center. If the facility is not designed to cool 750 watts per square foot and higher throughout the entire data center, you may find significant limitations on your ability to deploy HPC infrastructure on any scaled basis. In addition, liquid-based cooling systems are preferable to air-based systems given liquid's inherent advantages as a cooling medium versus air.**

3. Are your power circuits adequate to provide robust, consistent power?
   Often overlooked in data center operations is the importance of having adequate power circuits to provide reliable connectivity. In short, as we summarized in an earlier White Paper entitled "Challenges of Multi-Circuit Power Distribution in Colocation," managing many smaller power circuits versus fewer larger circuits inside the data center rack invites risk and potential operating disruptions. Utilizing many smaller circuits has a number of disadvantages, including reducing usable space inside the cabinet, causing delays when more circuits need to be installed if more power is desired at the rack, increasing risk by having multiple sources of failure, having little overhead to support "spikes" in power draws due to workload processes, and driving increased costs (many providers bill per power circuit deployed).

   With the robust power requirements per HPC rack, you must take into account the type of power circuits you are utilizing. Do they supply enough power to handle the increased critical load? Is there enough overhead in the power circuit to support a 50% increase in power draw when an HPC server workload spikes? Tripping a power circuit is a real possibility in legacy environments with the underlying nature of HPC infrastructure and processing cycles.

   And don't forget in this evaluation to understand the true max capacity of the power circuit, which needs to be analyzed for power capacity delivered at the 80% derated value, not gross 100% value. This is critical for planning and operating assumptions. For example, a 208 volt/ 20 amp/ 3-phase circuit at the 80% value actually only provides 5.8 kW of usable capacity.

- **We recommend that you evaluate the power circuits currently resident in your data center cabinets. If they are traditional single-phased circuits with less than 60 amps of capacity (ie, 30 amp/ 120 volt, 20 amp / 208 volt, or 30 amp / 208 volt), they will not be adequate to support your HPC infrastructure unless you utilize many of these circuits, which has major disadvantages as noted above. We recommend utilizing 3-phase circuits with at least 60 amps and 415 volts of capacity, which can deliver up to 34.5 usable kW of power draw per rack.**

CONCLUSION

The proliferation of data and our ability to capture it, store it, and manipulate it to serve important business and technical objectives is one of the compelling I.T. stories unfolding these days. Unlocking the value of that data and utilizing it on an instantaneous basis will have critical and transformative effects on our economy and lives.

But please be advised that the underlying computing infrastructures supporting high performance computing and applications requires magnitudes of order more power and cooling capacity in order to function properly. Very few data centers are engineered to support these requirements, so you must be very thoughtful and critical as you consider where and how best to deploy and scale your HPC gear.

Peter Harrison, CTO and Co-Founder
May 2017