# CRITICAL CONSIDERATIONS FOR THE MODERN DATACENTER INFRASTRUCTURE

AN IT PRACTITIONER'S GUIDE

## INTRODUCTION – EVOLVING TECHNOLOGIES, SAME IT CHALLENGES

The modern datacenter has evolved and the rate of innovation continues to quicken with unprecedented speed. Virtualization is a basic building block of infrastructure. Cloud is no longer just simply cloud—it's hybrid cloud and multi cloud. Cloud itself is ushering in the age of containers and microservices. Edge computing is the new "cloud". The advent of the Internet of Things (IoT) and the corresponding explosion of devices and data ushers in new technologies for data collection and analysis. On top of that, artificial intelligence and machine learning (AI/ML) continues to make its way into the datacenter.

All these technology trends and applications force IT organizations to view infrastructure differently. Some of the more emergent workloads such as data analytics and AI/ML have unique computational requirements that cannot be served by the traditional volume "rack" servers. Further, these emerging workloads can have a cascading effect on the traditional infrastructure as "specialized" server hardware may be required to meet the computational requirements.

Adding a layer of complexity to the above scenario is the fact that the IT-to-Business Unit (BU) service relationship is shifting to an automated "as-a-service" model enabled largely by software-defined constructs. This shift is necessary as IT evolves into a provider of intellectual property (IP) that drives digital transformation.

### INFRASTRUCTURE MATTERS MORE THAN EVER

This market dynamic is not unfamiliar to IT organizations. Disruptive technologies with unique needs emerge with regular cadence. IT is required to deliver to the BU to enable first mover status in an increasingly competitive market. At the end of the proverbial day, though, the needs of IT are a constant. Organizationally, IT must:

- Increase organizational agility

- Increase infrastructure agility by deploying a resilient server farm that can support the range of workloads and applications that populate the datacenter and their unique compute requirements for today and tomorrow
- Maximize resource utilization to drive the highest levels of infrastructure efficiency, including reducing costs

Moor Insights & Strategy (MI&S) believes infrastructure matters more than ever, especially in the software-defined world. The server choices IT organizations make can separate winners from losers. Terms like “time-to-value” and “time-to-market” are more than buzzwords. Performance matters and server performance starts − but doesn’t end − with the central processing unit (CPU).

Given this dynamic, questions emerge that require critical thought. How does the IT organization balance the needs of the digitally transformed business while maintaining the more traditional infrastructure that keeps an organization ticking? Does IT need to re-examine its infrastructure deployments and look for ways to drive greater levels of efficiency and flexibility? Is infrastructure simply a commoditized platform that enables real differentiation through software and services?

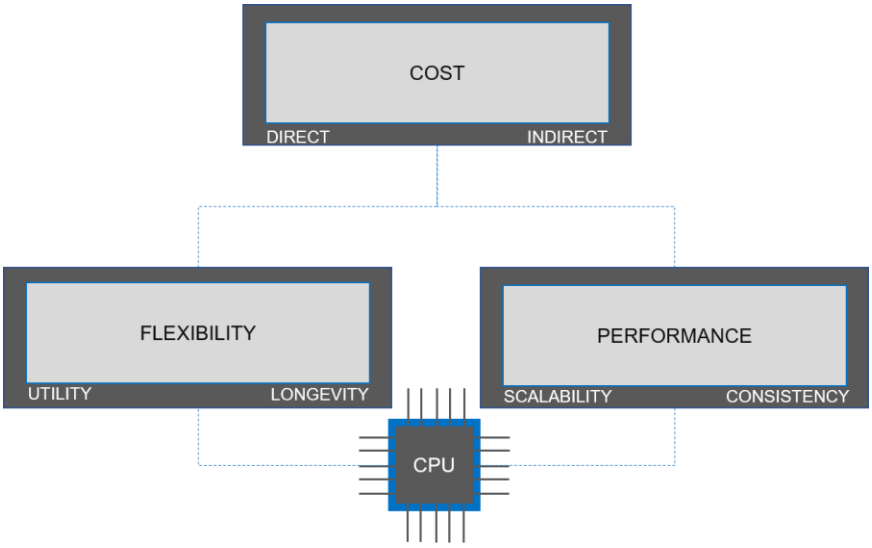
This paper will examine these questions and dig deeper into why MI&S believes infrastructure choices can have a big impact on the modern business. Further, this paper will discuss why companies like Intel should be considered a strategic cornerstone of the modern datacenter.

## CRITERIA TO CONSIDER WHEN EVALUATING DATACENTER INFRASTRUCTURE

Any IT professional knows there are a number of factors to consider when choosing the infrastructure that powers a datacenter. As each environment is unique, needs and priorities differ from organization to organization. However, MI&S recognizes three general categories that should be considered “table stakes” for any organization.

**Platform flexibility, performance,** and **cost** are all terms that can be interpreted differently. However, the ability to maximize the return on infrastructure investments while providing optimal performance across a variety of workloads is the goal of any IT organization.

## FIGURE 1: DATACENTER INFRASTRUCTURE CONSIDERATIONS




*Source: Moor Insights & Strategy*

When considering server platforms, one must start at what makes servers performant—the CPU. It's not an understatement to say the CPU is the cornerstone of the datacenter. This is not to say accelerators aren't important, they are, but applications and workloads are powered by the CPU. In fact, these very workloads and applications are often optimized to best utilize the CPU architectures.

When evaluating the fitness of the CPU, it's important to look beyond spec sheets. While we will dive into this more deeply later in the paper, it's worth noting that advantages on a CPU spec sheet sometimes − but don't always − translate into better performance. Per core performance is critical to powering enterprise IT workloads and core-to-cache architectural affinity is an often-overlooked consideration that can end up impacting performance when deployed.

## CALCULATE THE **REAL** COSTS

Total cost of ownership (TCO) is a term and number bandied about by hardware and CPU vendors to demonstrate a business value. The challenge with TCO is it is a broad-reaching concept that can easily be manipulated by marketers in an attempt to demonstrate competitive advantages. Additionally, IT departments place different premiums on different elements of TCO models, meaning the impressive numbers seen on websites and in product datasheets don't always translate into reality. MI&S believes there are real costs that any IT organization should consider.

### Overall cost of running workloads

The days of “rack ‘em and stack ‘em” are past, if they ever really existed. The enterprise always had tier-one applications with specific requirements around application acceleration, memory, or storage. The mainstreaming of data analytics and the emergence of workloads such as AI and ML shine a spotlight on these requirements.

The cost to deploy a system goes beyond initial capital expenditures (CapEx). IT organizations should consider the overall cost of deploying resource-hungry workloads. They should also consider the levels of integration individual components share to deliver best-in-class performance. For example, is the server platform able to take advantage of storage class memory (SCM) technologies such as Intel Optane? Or FAST (Fully Automated Storage Tiering) cache and storage technologies such as Samsung’s Z-SSD? Does a server platform require more investment in memory or a dual in-line memory module (DIMMs) to achieve best performance? Is there a level of resource investment required to optimize workloads? Does a CPU have the flexibility through instructions to boost performance (for example, Intel AVX512)? These are just a few important considerations when calculating the real cost of application support.

### Cost to maintain

MI&S sees a number of factors IT organizations should consider when calculating the cost of maintaining server infrastructure. One of the first considerations should be the cost of software licensing. As most IT organizations can attest, the move to core-based licensing for many commercial software providers such as Microsoft and Oracle can be financially difficult to manage. Furthermore, this licensing model seems to be trending toward broad market adoption, which should cause organizations to reconsider their approach of server deployment based solely on a CPU core count. Instead, looking for the best “per core” performance can be a better approach that allows IT organizations to potentially save considerable money on licensing fees.

Consider the total equation even for those software licenses that appear to favor “sockets” over core counts. Does the deployment of “fewer sockets” require additional server infrastructure to power a workload? Simply shifting the cost from software licenses to additional hardware is not an adequate solution, especially when considering the potential performance degradation associated with such an increase in infrastructure footprint.

Another factor in the cost to maintain is the cost of management. Servers from different vendors (or powered by different CPU vendors) may or may not require multiple

consoles to manage effectively. More challenging is the management of multiple virtualized environments, which is far more complex for the IT organization that delivers an "as-a-service" model wherein a catalogue of services is available via self-service portals.

**Cost to extend**

Virtualization extends the server lifecycle considerably. However, many servers in the latter years of that lifecycle tend to be relegated to supporting lightweight virtual machines due to somewhat limited CPU processing power and capabilities. Investing today in servers built on CPUs with better per core performance may protect against underutilized servers tomorrow. And investing in CPU technologies with advanced feature sets can enable higher levels of utility.

As previously mentioned, TCO is a vague term that often gets co-opted by technology solutions providers to promote the latest chip, server, or software. MI&S suggests IT organizations develop a criterion for TCO and ROI that reflects their business needs and use that to assess relative value to infrastructure deployment decisions. Be holistic and pragmatic. Above all, understand that TCO models are just that – models. They are based on assumptions that may or may not reflect the reality of your environment and workloads.

In theory, servers based on Intel Xeon Scalable should make the life of an IT professional tasked with measuring TCO and ROI easy. With Xeon's consistency of performance, capacity planning should be more predictable. Additionally, Xeon's wide adoption in the marketplace could make analogous modeling less challenging for those IT organizations that are new to cost analysis.

## PERFORMANCE MATTERS, BUT DON'T BE FOOLED

As with TCO models, CPU performance can be interpreted in many ways. What really matters to an IT organization is how a server supports its specific application needs in its unique environment. Infrastructure technology vendors promote performance in a few ways and those are sometimes not reflective of real-world performance.

- *Speed/feed comparison* – A listing of specifications is perhaps the crudest proxy of performance because it's not even a rough estimate of performance. Core counts, memory capacity, and the like are important, but they don't indicate how those cores and memory enable the workloads that run the business.

- *Benchmarks* – While better than a speed/feed comparison, benchmarks performed in a pristine environment may not be a good proxy for how workloads perform on a platform in one's environment. Benchmarks like Standard Performance Evaluation Corporation (SPEC) for CPU, TPC for database, and VMmark for virtualization are all good guide posts, but don't necessarily indicate real-world performance.

What really matters is how a server and its CPU perform in an environment running workloads and applications with an enterprise's data, whether that environment is fully virtualized or bare metal, running commercial software or homegrown line-of-business (LoB) applications.

***"In theory, servers based on Intel Xeon should make the life of an IT professional tasked with measuring TCO and ROI easier. With Xeon's consistency of performance, capacity planning should be more predictable. Additionally, Xeon's wide adoption in the marketplace should make analogous modeling less challenging for those IT organizations that are new to cost analysis."***

MI&S believes IT organizations should be critical in evaluating server platforms. Consider the underlying architecture. How many cores does a CPU support? How performant are the individual cores that make up a CPU (not all cores are created equal)? What is the core-to-memory relationship? Does a server require over-provisioning of memory to achieve best (and consistent) performance? Does the cache architecture enable or hinder performance? Does the CPU have the instructions and instruction extensions to enable your future workloads? Is the ecosystem of that CPU support the applications you run today or may want to run tomorrow?

The previous questions can be difficult to answer and not every IT organization has the resources or time to consider so many factors. As a pragmatic approach, MI&S suggests the following when evaluating servers and the underlying CPU architecture:

- Look for performance proxies that are indicative of the applications you want to run. Are there third-party tests that best simulate the performance of a workload "in the wild"? Are there benchmarks that simulate the workloads and applications in one's environment? A good example of this is STAC, an organization that provides objective technology assessments for the financial services industry (FSI). Through STAC, organizations can better understand the real-world performance of those applications critical for driving the industry.
- Make sure the software you want to run is supported.

- Look for customer case studies that reflect a similar environment to that of the organization.

## PERFORMANCE MATTERS – SCALABILITY AND CONSISTENCY ARE KEY

With the industry's historical focus on raw performance, the scalability of platforms and consistency of performance are often overlooked. However, their importance cannot be overstated. To be clear in definitions, scalability of performance is at the platform (server) level and consistency of performance is driven by the CPU.

Platform scalability is important because it impacts cost, performance, and manageability. It may be wise to set minimum internal server and CPU standards around the most demanding workloads that support the business, with other server platforms cascading from those biggest demands. For example, if an IT organization deploys SAP HANA on a server powered by eight processors, then infrastructure choices could cascade downward to perhaps four-way servers for large scale database and virtualized two-way servers for back office support.

Performance consistency is critical to any enterprise IT organization because its impact is both broad and deep. CPU, cache, and memory design all have implications on performance consistency. MI&S suggests thorough testing of database and virtualized workloads that will fully tax the CPU, cache, and memory to determine the fitness of architectures to support more demanding workloads.

## FLEXIBILITY – UTILITY MEETS LONGEVITY

Perhaps one of the most important considerations when deploying server infrastructure is the utility of servers and the longevity of that utility. Utility is the ability of IT organizations to take a "swiss army knife" approach with the servers deployed in the datacenter. In other words, which server platforms can deliver the most usefulness (support the widest variety of workloads) at the lowest cost? The logic of the argument is, servers with high utility support the digitally transformed business more effectively as they are provisioned and re-provisioned hour-to-hour and even minute-to-minute. For example, servers may be provisioned to support virtual desktop infrastructure (VDI) deployments by day and re-provisioned to support AI training after hours. A traditional two socket server is an ideal platform in such a usage model.

In addition to utility, IT organizations should look to invest in server technology that will stay technologically relevant through its life cycle, supporting the next big application

that will undoubtedly emerge. The rate of innovation results in new and emerging workloads with unique requirements on a seemingly daily basis. While there are no sure bets, a wise investment in IT infrastructure may prevent having to invest in new hardware for each new workload that powers the business.

The platforms powering these new and emerging workloads extend beyond the CPU and memory. When choosing silicon technology partners, look for those with a breadth of IP portfolio and a broad ecosystem of support. Do they demonstrate a vision that aligns to the workloads that will power your modern datacenter? Further, do they map products and technologies to that vision? These are important considerations, along with improved performance and price.

The workloads of the future will require an entire platform that extends beyond the CPU. Seek out solutions providers that make investments and/or create partnerships in complementary technologies to enable tightly integrated platforms and support their CPU roadmaps.

## INTEL XEON SCALABLE – DESIGNED FOR THE DATACENTRIC ORGANIZATION

The role of the modern IT organization has shifted from that of a break/fix service organization to a broker of services that enable the digitally transformed business. As such, IT organizations must be able to deploy infrastructure with confidence. The software-defined datacenter enables IT agility, but it is only as good as the underlying hardware that supports it and the levels of integration that exist between silicon, server platforms, firmware, operating systems, and software. Hardware matters more than ever.

The digitally transformed organization generates unprecedented levels of data. In many cases, the successful digitally transformed organization will be the business that can transform mounds of data into actionable intelligence faster than the competition. To achieve this, it takes more than a CPU. It takes a compute platform that integrates the latest technologies in compute, memory, storage, and networking.

MI&S believes Intel is competitively positioned in the IT infrastructure market for the long term for a few reasons:

***Excellence in silicon design and breadth of portfolio***
Simply put, Intel's Xeon portfolio is performant and scalable. The Intel portfolio scales

from the device to the edge to the cloud to the core datacenter and claims its core design delivers performance that powers the workloads across that continuum. While there are no known documented case studies to point to, in theory, the ability to standardize on a single architecture from end point to the core datacenter should drive cost and performance efficiency and streamlined management. For instance, a common architecture across the enterprise could enable IT to standardize on a single management console.

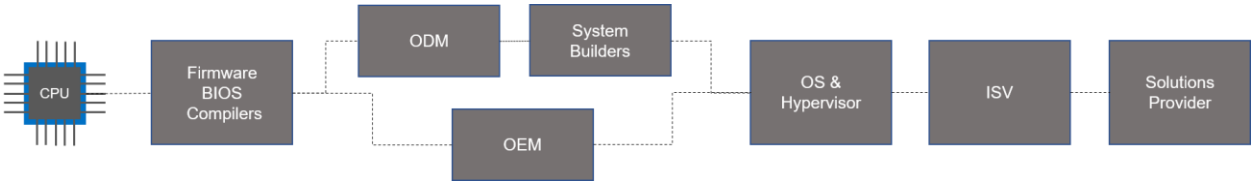
***Strategic investments in delivering a complete compute platform***

As mentioned, the resources required to drive optimal performance scale beyond cores and memory as workloads continue to evolve. Accelerators, storage, and network innovation is not simply nice to have – it is a necessity for the modern enterprise to take advantage of the explosion of data populating the datacenter. To be clear, CPU innovation is critical, but innovation cannot end with the CPU.

Intel is making investments through research, product development, partnerships, and acquisitions and claims to deliver a complete solution for enterprise IT from the CPU to cache to memory to storage to networking.

***Breadth and depth of ecosystem relationships***

CPU technology is a critical component to IT solutions being deployed, but it is just one component. This fact requires a company like Intel to approach the market differently. From a design perspective, features of the CPU must be visible and easily adopted by ecosystem partners "up the stack" that span silicon, BIOS, firmware, compilers, OS, applications, solutions providers, and more. Otherwise, the benefits of silicon innovation are never realized by the IT consumer.




The level of partnership required to make such a complex ecosystem "work" requires resources, investments and time. MI&S believes Intel has excelled in each one of these focus areas. The company claims to have the most comprehensive ecosystem in the industry.

***Experience and expertise in all verticals and market segments***

There is a reason Intel Xeon powers more than 90 percent of the servers in datacenters

around the world. The applications and workloads that drive the innovators across all verticals are, for the most part, built and validated on Xeon platforms. Obviously, developers cannot ignore the sheer weight of the installed base when bringing solutions to market, but this also show's Intel's close relationship with those communities.

***Experience that informs the future***

Intel claims its experience informs its investments in product roadmaps and strategic partnerships, and one can reasonably argue that these informed investments make Intel a formidable leader in the datacenter infrastructure market. A good example of this is Intel's development of Optane DC persistent memory. Over years, Intel has developed a new way of how it looks at memory and storage. While it is not the only SCM solution on the market, Optane DC is an example of Intel's work with the ecosystem that made Optane a viable product instead of a "science project" that never saw the light of day because of a lack of enablement. Now that Intel is fully committed to its own FPGAs, datacenter GPUS, and discrete ML accelerators, we believe Intel is better aligned to fulfill the modern datacenter's needs than it was before.

## CALL TO ACTION

Today, cloud computing, edge computing, digital transformation, and other factors drive the evolution of the datacenter and the transformation of IT operations. Still, some things never change. IT success factors include lowering costs, simplifying management, maximizing utility, and increasing the viability of the server platforms deployed.

Choices abound in an ever-competitive infrastructure market. The question is, "What is the right choice?"

MI&S recommends vetting the next investment in infrastructure as follows:

- Take a hard look at different architectures and determine what's best designed for the workloads and applications that reside in one's datacenter. Does the platform provide the performance for today and tomorrow's applications? Are cores performant enough? Are adjacent platform technologies easily enabled for your environment? Ask the tough questions of server OEMs and CPU partners. Understand the magnitude of interoperability challenges and how these challenges can limit the agility required for the modern IT organization.
- Test the server in its environment. This may seem obvious, but it's not always done. Look at performance holistically. How does that virtualized environment

perform? Is the performance consistent? Does a workload scale as more resources are applied? Are additional investments required to achieve acceptable performance levels?

- Analyze direct and indirect costs of deploying workloads on a given architecture: software licensing, management, over-provisioning, qualification, enablement, optimization and more.
- Look for breadth and depth of ecosystem. Is investment being made in a technology partner with mature and established partnerships relevant to your enterprise already in place? Are they the right partnerships representing where the industry is heading? Balance this against price and performance for the workloads you run today and may want to run in the future.
- Look for vision. Does the CPU vendor demonstrate an understanding of the technologies that will drive your datacenter of tomorrow? And does the company have a product roadmap that supports the computational needs of these new and emerging workloads?
- What populates your datacenter today? How easy will the path to integration be?

MI&S believes IT organizations should put Intel on the short list when undergoing infrastructure modernization exercises. The company's demonstrated and mostly consistent record of delivering CPU architectures and adjacent technologies spans the usage continuum from devices residing in the IoT to the largest supercomputers solving the world's most complex problems. Additionally, its investments in adjacent technologies demonstrates a vision centered on the needs of tomorrow's workloads.

## IMPORTANT INFORMATION ABOUT THIS PAPER

### *CONTRIBUTOR*
Matt Kimball, Senior Analyst at Moor Insights & Strategy

### *PUBLISHER*
Patrick Moorhead, Founder, President, & Principal Analyst at Moor Insights & Strategy

### *INQUIRIES*
Contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

### *CITATIONS*
This 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*


### *DISCLOSURES*
This paper was commissioned by Intel. 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.