WHITE PAPER

Data Center
Cloud Computing

Accelerate Business Growth With Multi-Cloud and Optimal Workload Placement Strategies

Strategies to help IT decision makers select cloud architecture and services based on their organization's business needs and technical requirements

Most organizations that complete a thorough workload placement analysis discover that it is the right mix of *both* public and private clouds that accomplish the business' goals.

Table of Contents

Executive Summary

In an age where the digital economy has compelled some enterprises to modernize and has disrupted others, cloud technologies have a large role to play. Cloud architecture and technologies have improved time to market and value for new applications, lowered operational and capital expenditures, and provided organizations with an ability to dynamically scale to meet evolving business needs. In an IDC survey, "50 percent of IT decision makers identified cloud-based infrastructure and applications as one of their highest investment priorities for meeting their business goals over the next five years."[1] Business transformation is dependent on the agility of cloud solutions, whether that be on-premises private cloud, public cloud, software as a service (SaaS), or a mix of all these. At Intel, we believe in the advocacy of cloud architecture as the most efficient and agile computing solution regardless of where it is deployed (public, private, hybrid, or multi-cloud).

Determining the right (or optimal) cloud infrastructure for each workload is a common challenge for both small and large organizations. The needs and goals of each organization and industry differ, making it impossible to adopt a one-size-fits-all cloud strategy—or even the same strategy for each workload within an organization. Organizations that conduct workload placement assessments often set out with a single goal: Should we migrate to the public cloud or keep the workloads on-premises? Based on our analysis, the answer to that question is, it depends. There are multiple considerations, not the least of which is the individual workload characteristics. After completing a thorough workload placement analysis, most organizations discover that the right mix of both public and private clouds accomplish their business goals.

This paper examines enterprise workload placement based on multiple evaluation vectors, and while there is no straightforward answer, the framework in this paper will help IT departments improve their decision making as they consider what cloud solutions and services best satisfy their business needs and technical requirements.

Intel Data Center Group Authors

William Giard, Principal Engineer

Danielle Hearn, Enterprise Marketing Specialist

Shawna Meyer-Ravelli, Cloud Service Provider Marketing Manager

Key Terms Overview

Cloud solutions have grown exponentially, vendor offerings have expanded, and the underlying technologies have matured immensely—making it critical to use a common set of terminology. For a full list of terms used in this paper, refer to Appendix A: Terminology and Definitions.

Characterizing Enterprise Workloads and Deciding on Placement

We recommend a multi-faceted approach to workload placement using these key decision factors (see Figure 1):

- **Business Requirements.** The top business problems that an organization is working to solve—or, put another way, the main use cases that decision makers want to enable or enhance. These can include time to market, agility, total cost of ownership (TCO) reduction, and legal and regulatory requirements.
- **Technical Requirements.** Technical attributes that must be addressed per workload under evaluation for right-sizing—workload performance, security, back-end integration, data size/storage, and elasticity.
- **Ecosystem Offerings.** Factors such as SaaS maturity, cloud service provider (CSP) offerings, or the market accessibility of cloud expertise.

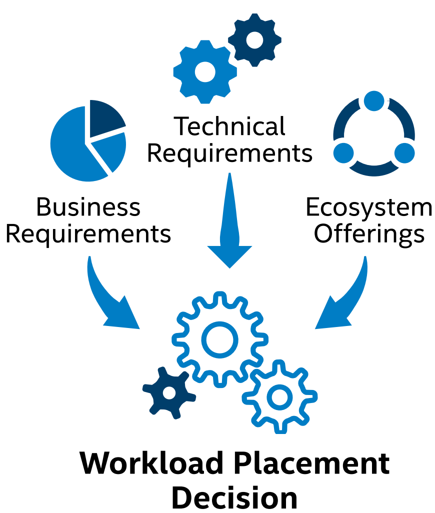

Figure 1. The decision factors for workload placement can be divided into three main categories: business, technical, and ecosystem.

Business Requirements

The highest priority in determining workload placement stems from the business needs, pain points, and growth opportunities. The issues that keep organizational leaders up at night—data growth, data control, cost, IT scale, rapid insights, security, and so on—will typically dictate IT strategies. Additionally, the future vision and business imperatives equally guide this decision process. Understanding and addressing some potentially determinant business requirements prior to further assessment will not only save time but will limit workload movement or relocation later.

Agility and Time to Market

In today's business environment, change happens quickly, forcing businesses to respond and alter their business model to meet customer demands. To keep pace in the digital economy, agility is critical—whether you are deploying new services or scaling to new markets, the degree of agility and scale required by an application can determine placement. In many cases, public cloud services offer a benefit to business in achieving faster time to market, scale, and provide the agility needed to investigate new applications and services.

Considering and calculating the durability of your application or project will have a significant impact on workload placement. Workloads that may be transient or needed only for a specific duration are especially suited for public cloud services. For example, in cases where a project is rapid development or conceptual in nature, public cloud services provide an advantage by lowering long-term impact of provisioning equipment that may be repurposed at the end of the project. Additionally, projects with longer durations may benefit from public cloud services until a point where the cost-benefit tips in favor of on-premises workload placement. An example of this type of project is online gaming applications—a business can leverage public cloud for initial development, launch, and maintenance of the game when success is uncertain, and transition to on-premises once customer adoption reaches a certain level and technical or business requirements dictate the migration.

Example: Research Company - Accelerates TTM for Results

A startup research company could not afford to wait months to prove its viability and deliver time-sensitive research and predictions in face of strong competition. By adopting public cloud to quickly scale compute, networking, and storage, the research lab was able to process PB-scale data sets in hours versus months and significantly accelerate the company's time to results and market for new findings.[2]

Mission Critical Intellectual Property (IP) or Services

Maintaining control over competitive advantages and critical products or services must be the first priority when considering placement. Because IP security and policy compliance is a top priority for every CIO, some businesses categorically require certain information to reside only on internal infrastructure. Investigate and understand if corporate policies dictate how IP or customer data is handled, doing this at the onset is important during migration efforts.

Enterprises with sensitivity to moving certain workloads or data outside the four walls of the data center should check first with the cloud provider to understand their offerings. Public cloud providers' confidential computing services and security offerings continue to mature and increase to support IP business requirements. Determine whether a migration to public cloud will result in a need to re-architect the workload to open communications between the private and public cloud. Factor in these integration efforts and costs to ensure ultimate agility, flexibility, and operational effectiveness for all workloads, including mission critical apps. Doing so will be increasingly important as an enterprise increases investments in the public cloud.

Legal and Regulatory Compliance

Fully comprehending the legal and regulatory environment of the organization, larger industry, and every country where business is conducted is a critical factor in determining optimal workload placement. Do organizational or local regulations (such as General Data Protection Regulation (GDPR)) require the institution to keep sensitive data on-site or within the borders of a particular country? Many countries now require certain data, like personal identifiable information (PII), remain in specific geographic or country boundaries, similar to classifications for export and control regulations for International Traffic in Arms Regulations (ITAR).

With global cyber-attacks at an all-time high, adhering to these regulations is critical for every business tasked with managing sensitive data. Despite the fact that public cloud environments have proven to be highly secure, regulated workloads often remain on-premises given the high level of control that is possible and the ability to more easily meet some audit requirements.

Example: Financial Institution – Sensitive Data Challenge

A Fortune 500 financial institution had an initiative to move applications and data to the public cloud. During the process of developing the strategy, it was discovered that their corporate policy prohibited placement of personally identifiable information (PII) and other sensitive data beyond their internal network and firewall. Although many security standards are supported by public cloud providers, because of the internal policy, the organization kept their data on-premises.[3]

Data Control and Security

When it comes to data control and security, off-premises cloud requires a certain level of trust, as it can lack the level of transparency some organizations are accustomed to with their data. Hosting data in the cloud necessitates that the organization hands over control of service availability and data protection to a third party organization. Enterprises must decide if they prefer a third party to secure their data, or if they would rather maintain the data in-house and under their own control. Start by asking these questions:

- What level of risk is the business willing to accept?
- What are the acceptable trade-offs?
- What security protections does public cloud enable vs. on-premises?

While the decision to migrate data off-premises requires trust, in exchange a business could gain benefits such as agility, ease of deployement, or improved security. The benefits of moving to the public cloud could far outweigh these realities, so a known trade-off is made. These are discussions to have with IT management and line of business leadership as their sentiments can tip the scale one way or the other.

Total Cost of Ownership

Understanding and evaluating placement decisions based on cost is important, and although spend does not always outweigh other business drivers, it should be well documented and understood. This is particularly important because public cloud costs are generally based

> "The transparency of public cloud pricing can create the impression that the public cloud is less expensive compared to a modern on-premises data center. However, when looking at TCO it's important to not only look at 'pay as you go' costs, but also associated migration costs. The majority of our surveyed customers reported unexpected costs when they migrated all their applications in a 'lift and shift' manner to the public cloud."
> *– Brian Kelly, founder and CEO of CloudGenera*

on operational expenses (OpEx), or "pay as you go," where usage may not be initially available or well understood. Completing a holistic cost analysis for both on-premises and off-premises resources will allow comparison and the ability to deploy the right workload in the right environment when cost is the number one requirement. When looking at TCO, it's important to look at both "pay as you go" costs and migration costs associated with lift and shift.

In the last year, migration trends back to on-premises workload placement (often considered repatriation) have been on the rise as businesses have internalized the true cost of operating workloads outside their own environment. Both traditional enterprises who moved workloads to the public cloud in a "lift and shift" manner, and "born on the web" companies who started in the public cloud are adjusting their strategies based on their learnings. Consequently, workloads determined to be better suited for on-premises are being migrated—80 percent of customers surveyed by IDC have reported "migrating applications or data that were primarily part of a public cloud environment to a private cloud or on-premises environment."[4]

Interestingly enough, some of the top CSPs now offer their customers private cloud deployment options (i.e., Microsoft Azure Stack*, AWS Outposts*, Cisco Hybrid Cloud Platform for Google Cloud*)—further supporting the premise that enterprises maturing their digital transformation journey will take advantage of public and private clouds to help deliver better long-term TCO. Today, the IT strategy of choice for the enterprise involves running multiple environments—81 percent of enterprises report having a multi-cloud strategy and on average use 4.8 clouds across both public and private.[5]

With so many companies reporting that trial and error helped achieve their optimal TCO strategy, how does an organization plan for success from the start and determine a cost structure that achieves the best business outcome? Engaging in a thorough evaluation of all options and collecting accurate data will guide an informed decision. Analyzing environment costs and TCO per workload is a complex task; therefore, organizations may consider consulting an assessment-tool provider with deep expertise and the ability to evaluate an individual situation based on real-world system and application data. This kind of investment will always produce better results.

Example: Domain and Web Hosting Company – Lowers TCO with Hybrid Cloud

A domain and web hosting company foresees expansion to over 90 markets in the near future. To prepare for this staggering growth path, the company must cost-effectively augment its existing global data centers and scale its private-cloud hosting environment to meet new customer demands. By adopting a hybrid cloud model, the business can expand into new markets faster and experiment at a lower price point, all while maintaining a consistent customer experience. Developing a hybrid strategy that rightly optimizes workloads to the right location with the right density lowers overall infrastructure TCO and allows the business to sell their products at a much lower price point.[6]

Regardless of whether the evaluation process is performed in-house or with a consultant, the steps shown in Figure 2 are helpful in outlining the best practices. (The steps in the figure are discussed in detail below.)

Using this framework should yield significant insights into which resources are no longer required within the enterprise and where modernization and migration opportunities could return positive results.

1. **Identify the most important business drivers.** These may include the following:
   - Cost and efficiency (reduce data center costs, improve data center efficiencies)
   - Agility (reduce turnaround time, faster TTM for new projects)
   - Control, regulatory, and security (retain control/ security of mission-critical services or sensitive data, comprehend legal and local requirements)
   - Resource capabilities (fill in-house resource gaps)
2. **Conduct a cost analysis.** If the most important business drivers are cost or efficiency, conduct a thorough cost analysis for the workloads under investigation. There

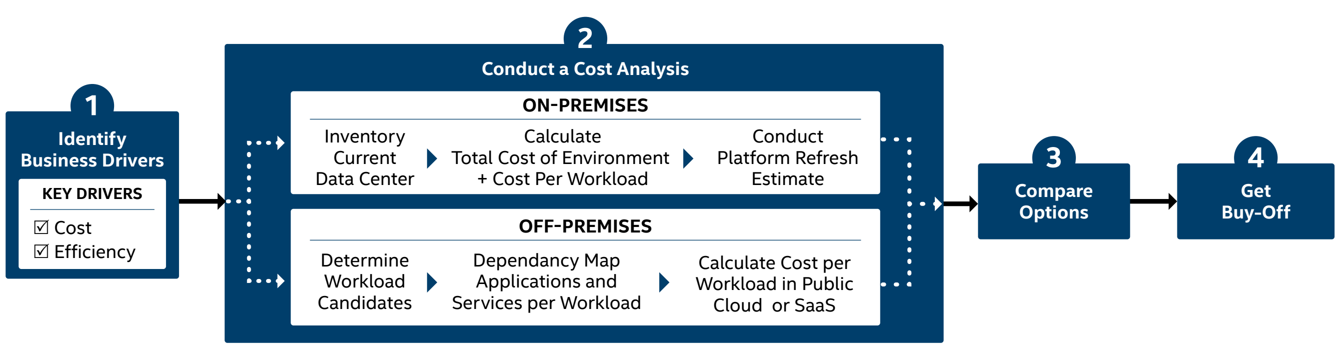

Figure 2. Following a formal assessment process can help determine a cost structure that achieves the best business outcome.

are two primary aspects of this analysis, described later: conducting an on-premises inventory and conducting an off-premises inventory.

3. **Compare the options.** How much is being spent today in total and per workload? Where are workloads better suited to a different environment? Can investments in modern infrastructure and hardware reduce overall spend on-premises? Can moving a group of applications to the cloud reduce spend on those workloads? The final analysis should provide a view of cost per group of applications and services across public cloud and on-premises options.
4. **Get buy-off.** The final step should be reviewed by the line-of-business owners, application owners, and solution architects to ensure the performance and connectivity to any potentially new architecture model meets the needs of the business.

ON-PREMISES INVENTORY

Complete an accurate inventory of the current on-premises data center. This assessment should include all compute, memory, storage and network hardware, management software and support services, facility costs (i.e., energy consumption), connectivity costs, data center staff resources and training, and software licenses. This assessment should include only resources that would not be relevant if they were moved. Costs such as operating systems may be included if they will vary in cost between solutions. Once the inventory is complete, it is possible to calculate the total cost of environment plus the cost per workload. Cost simulators and consultants offer faster and more accurate estimates of on-premises TCO to establish a baseline to compare against projected costs of other solutions. Ideal results will yield an approximate breakdown of the cost per workload running in the environment and show costs beyond a point-in-time analysis. Conducting analysis over time (past as well as extending into the future) is key to making strategic decisions, a one-time analysis has limited effectiveness.

ON-PREMISES MODERNIZATION

When evaluating the on-premises environment, it is critical to also consider if modernization efforts will yield future cost advantages. Continuing to upkeep outdated and monolithic infrastructure not only slows IT down, but adds significant maintenance and facilities cost. Look for energy and maintenance costs that rise as infrastructure ages, this indicates that a refresh to latest technology, such as the Intel® Xeon® Scalable platform, might be due. To determine if a platform refresh will lower TCO complete the following:

1. Begin by properly sizing the existing resources to a new high-density environment. Upgrading to new hardware with better performance and higher storage and memory capacity should increase utilization, reducing the total number of servers needed to run the existing workloads. Be sure to factor in any anticipated growth to ensure you have the resource capacity ready for the needs of both today and tomorrow.
2. Use this new host count and associated new hardware spend to calculate the total cost of a platform refresh plus the cost per workload. This should include estimates for the hypervisor licenses, storage connections, software licensing, maintenance support and services, energy consumption, and rack space for the new environment.
3. Compare the total cost of the modernized and consolidated environment against the cost of running the original environment. This new TCO can also be compared against the cost of migrating these virtual machines (VMs) to the public cloud after the off-premises inventory is completed.

By consolidating outdated servers into fewer, more performant systems, hypervisor licenses, per-CPU OS and application licenses, and IT support resources can be reduced. For example, an environment that can be downsized from 150 older-model servers, with low performance and low memory volumes, to 20 higher-density servers with significantly more memory and higher-performance storage and network, may introduce significant resource savings, and also capitalize on other system-level improvements from latest generation technology. The exact degree of cost savings depends on the organization; however, choosing to modernize with the latest Intel Xeon Scalable platform can deliver up to 60 percent TCO savings compared to a 5-year old system.[7]

OFF-PREMISES INVENTORY

Begin by identifying the migration candidates, or the workloads that have not previously been omitted for evaluation based on the business requirements. Next, complete a dependency mapping of these applications and services. Group all resources by their dependency and level of integration with other systems and processes. This will ensure dependent resources are kept together upon migration to a new environment. Lastly, calculate the cost per workload grouping in the public cloud as well as identify SaaS options and associated costs.

Identification of SaaS opportunities can dramatically reduce on-premises costs and potentially increase availability and scale for core general-purpose business services such as email and communications, search, social, and some IT systems management. These are often quick migration opportunities with immediate resource and cost recovery.

Ultimately, the economics of public cloud depend heavily on the size of the organization, the level of scale and resource utilization required, and the businesses' IT purchasing power. To avoid guessing, calculate the cost per workload using cost simulators, which offer projected costs of moving workloads

Example: Small Financial Provider – Cost Savings On-Premises

Businesses that closely monitor their workload utilization rates and efficiently manage IT resources can achieve significant cost savings. A small financial organization serves as an excellent real-world example. The organization operates 14 physical hosts with 157 virtual machines (VMs). These are consuming only 50 percent of available RAM and have a very low virtual CPU consolidation ratio of 1.5:1. In total, these use 11.5 percent of CPU potential. Right-sizing these VMs to increase the utilization rate results in a cost reduction of 47 percent. Sizing hardware for on-premises, the environment requires only four servers to accommodate their compute and availability requirements—a reduction of 10 hosts and an estimated cost of USD 65,300 per year in licenses, hardware, and software over four years. In the cloud, the instances cost an estimated USD 133,000 per year. Without the ability to share resources, the cloud model penalizes the inefficient usage of compute resources. In this case, upgrading to on-premises high-density computer hardware offers significant savings compared to the cloud.[8]

Example: Healthcare Insurance Provider – Cost Savings Off-Premises

When the cost of on-premises physical hardware is distributed across dozens of workloads, efficiencies are found in virtualization. However, these efficiencies decrease when workloads operate at a near 1:1 ratio. If these workloads are seasonal or are not required 100 percent of the time, resource allocation on-premises becomes very expensive. For example, consider a regional healthcare insurance payer that operates a cluster of eight hosts for 24 production virtual machines (VMs). The eight hosts offered 128 physical CPU cores. At its peak, this cluster runs 100 percent utilization for several days; all 24 VMs consume the full eight hosts. Running this environment incurs an annual cost of USD 186,194 per year on-premises. After evaluating three major cloud service providers, the firm found that all three could provide the same performance and availability for less—the savings ranged between USD 10,000 and USD 40,000 per year. The major cost-savings factor was that the cost of physical hardware drove up the cost locally and made the cloud savings more substantial.[8]

to a specific cloud provider so that organizations can compare multiple vendors.

Know the True Cost of Cloud Migration

When considering workload placement, the scope of the analysis should not be just one-way. To avoid vendor lock-in, analyze the costs of relocating—that is, bringing the workload back on-premises or moving it to another cloud platform. Agility and flexibility are critical factors to consider when choosing between different cloud instances, and it is important to understand the possible workload portability limitations. For example, if it seems best to move a workload to the public cloud, consider whether that workload can move back on-premises, and at what cost. Fully understand the CSP's egress charges to move data out from their environments, and how this affects future strategies and the ability to keep workloads agile and flexible.

Additionally, the actual migration of a workload can also involve costs. These can be one-time costs, but closer study often reveals ongoing operational costs. Some dimensions of relocation costs include:

- **Network requirements.** Are new networks required with new address spaces and security policies/mechanisms?
- **Uptime requirements.** Can the migration be done live? If not, what are the implications associated with scheduling downtime? For more information on how to plan your CPU resources to achieve ultimate portability via live migration, see this Principled Technologies study.
- **Data/volume format conversions.** Are data conversions or volume reformats required?
- **Disaster recovery/business continuity (DR/BC) compatibility.** What is the location of secondary storage and predicted growth?
- **Operational management.** Is a separate operational management stack required to monitor, provision, and troubleshoot the new cloud platform?
- **Capacity planning and cost-structure modeling and analysis.** Does the cloud platform require a new capacity planning model for a new hybrid structure to adjust for OpEx or time-based financial accounting versus capital expenditures (CapEx)?

By examining and considering the needs and drivers of the business, a company can ensure workload placement decisions support the overall direction and needs of the organization and leadership. Typically, the demands on the business determine the cloud strategy for particular workloads; therefore, it is always best practice to document these indicators before completing a thorough assessment on the technical requirements or ecosystem offerings.

Example: Financial Institution – Data Migration and Storage Challenge

The data growth of a Fortune 500 financial institution was expected to exceed the current IT budget within two years. They evaluated several options to consolidate storage, including using the public cloud and new hardware platforms. Due to the size and extent of their data, the cost of migration to and from the public cloud presented serious concerns. As a result, the organization invested in a modern, software-defined storage (SDS) platform for their internal data center. The system significantly reduced costs and can accommodate growth into the foreseeable future.[8]

Technical Considerations

After reviewing critical business considerations, assess the technical aspects of the workloads to select the solution that best supports both business and technical goals. In 2016, Intel conducted more than 125 customer and systems integrator (SI) focus groups worldwide across all major industries. This internal research identified the four most important technical characteristics to consider for cloud right-sizing:

- **Performance.** The primary use cases that apply to workloads with very high performance scores include the following:
  - Performance and latency in relation to an end user's location, such as engineering solutions that reside physically near engineering departments. No matter the network, three primary factors significantly contribute to latency: propagation delay, routing and switching, and queuing and buffering.
  - Performance for resource-intensive transactions (compute, memory, and I/O) with guaranteed quality of service and response agreements. This also includes requirements to support application-specific performance goals and the ability to meet strict Service Level Agreements (SLAs) and monitor performance factors to ensure acceptable Quality of Services (QoS) levels. See Appendix B: Additional Performance Requirements for details.
- **Security.** Some applications process and house data—intellectual property, PII, and personal health information (PHI)—that could harm an organization if affected by malicious or accidental actions. This rating also incorporates whether security solutions are broadly available for a particular workload—security solutions are fairly mature for email workloads, for example. See Appendix B: Additional Security Requirements for details.
- **Back-end integration.** Back-end connections to other databases, frameworks, applications, workflows, and endpoints present challenges to both traditional and cloud migrations. Applications that require integration with other systems may be limited by proximity, hardware, or security requirements. The complexity and quantity of integrations impact the workload placement because of the increased cost to integrate into multiple clouds. Each integration must be assessed, modified, and refactored to meet the Operational Level Agreement (OLA). See Appendix B: Additional Integration Requirements for details.
- **Data size/storage.** Two major factors should be considered: data size and location (where the data is created and managed). Distance, availability, latency, and cost will quickly influence data volume decisions. Locality of data to applications and data need are critical to placement because large datasets can be challenging to transfer across distances. See Appendix B: Additional Data Size/Storage Requirements for details.

The Intel Affinity Model for Public vs Private Cloud Deployments

In an effort to refresh the body of research and work from 2016, Intel conducted a series of additional interviews and focus groups with IT and cloud subject-matter experts to document new data on workload affinity to cloud deployments. Additionally, Intel consulted with IT analysis providers to identify common workload trends and successful cloud strategies based on their experience with over 11,000 IT environments. The findings provide an updated and representative perspective on workload placement.

By scoring the four technical requirements for a specific workload, then adding together the individual scores for performance, data volume, integration, and security, we calculated a total attribute score. In comparing the multiple attribute scores across workloads, their propensity to be deployed on-premises or off-premises becomes clearer. The Intel Affinity Model for Public versus Private Cloud Deployments (see Figure 3) serves as a good foundation and baseline for evaluating workload placement based on a workload's technical characteristics. We do not, however, propose that this model is representative of every enterprise environment; as business requirements and cloud service offerings evolve, the affinity score represented here will shift—movement is to be expected as technologies mature and ecosystem offerings increase. For example, SAP* has traditionally, and according to our affinity model, been deployed on-premises; however, over the past 12 months, new service offerings have emerged in the public cloud and a trend has begun towards off-premises.

This model indicates that workloads well suited for SaaS or public cloud are those that require a low to medium amount of performance, security, back-end integration, and data size/storage—but as these requirements increase, so does the propensity toward private cloud. For example, based on this model, workloads with a lower total affinity score, such as email, may be best served by the public cloud, and workloads with a higher score, such as government and defense, would benefit most from being in an on-premises environment. The large majority of the workloads land in the middle section of the model, signifying that the better part of enterprise workloads today have equal propensity toward both public and private, suggesting that other considerations (business and ecosystem) should aid in workload placement decisions.

Intel Affinity Model for Public Versus Private Workload Placement

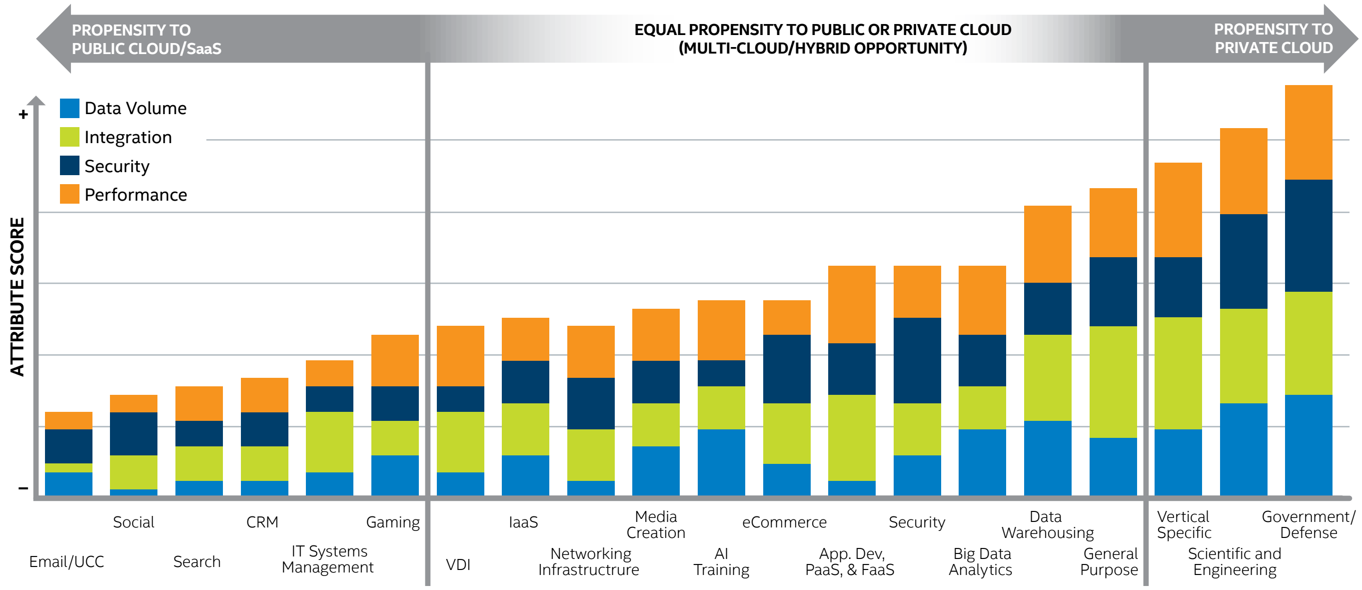

Figure 3. The technical workload characteristics determine whether an application is better suited to public or private clouds. Source: Intel internal analysis, January 2019. Note that different businesses will have different workload deployments, and these deployment differences may influence the attribute score.

This high-level view is helpful for understanding what Intel sees happening in the industry and with our customers, and it provides an informative baseline for understanding how performance, security, integration, and data size influence workload placement determinations. We fully encourage organizations to conduct their own workload analysis, using the findings in this paper as a starting point.

Ecosystem Offerings

Finally, it is important to consider ecosystem maturity in workload analysis. Ecosystem offerings to consider include factors such as the availability of mature SaaS and CSP services, and the accessibility of cloud expertise.

SaaS Offerings

SaaS offerings are increasingly popular because they provide organizations with fully managed and maintained expert solutions. Since SaaS solution providers manage all updates and upgrades, applications that require frequent upgrades and enhancements benefit considerably from a SaaS solution. SaaS vendor expertise provides organizations with industry-leading business solutions and the opportunity to offload these functions and free up precious IT skill and resources for more critical applications that require in-house expertise. Consider the maturity of the SaaS solution and the level of expertise the provider can deliver when evaluating vendors; common apps that have mature SaaS offerings include email, CRM, HR services, and collaboration and communication solutions.

As a prerequisite to undertaking SaaS, the organization should clearly understand the extent of business process re-engineering and technical integrations required to meet the business needs. If the organization's existing business processes are close to mainstream, or the technical evaluation team determines few major integrations are required, SaaS may be the right fit for the workload.

Example: Intel IT – Using SaaS for Commodity Apps

Intel's deployment of a SaaS-based human resources (HR) solution is an excellent example of how commodity apps such as payroll and expense platforms are well-suited for public cloud and SaaS. These "mobile-first, cloud-first" services contribute to Intel's digital transformation by boosting efficiency and productivity. The new cloud-based platform replaced Intel's 18-year-old legacy HR system. It provides mobile access to HR data and tools and empowers Intel's workforce by putting analytics and intelligence into the hands of company leaders, managers, and employees—at their desks or on the go.

The resulting platform handles all HR tasks, including payroll processing for nearly 106,000 employees in 49 countries and more than 1,800 business processes. Intel IT redesigned 224 integrations to external suppliers and internal downstream applications, and it simplified the overall HR system landscape by eliminating more than 250 legacy applications and interfaces.[9]

Specialized Services

Public CSPs are delivering capabilities to better enable cloud technologies, introducing code constructs and new specialized services at a rapid pace. Larger CSPs offer developer frameworks and rich integrated development environment (IDE) tool suites. These unique capabilities, services, and offerings can be a huge driver for many organizations. When choosing a CSP, evaluate their specialized service offerings, and match these to the needs of the business to drive faster innovations where required.

For example, many CSPs today provide expert analytics as a service offerings to their customers, including database as a service and data lake services. Because creating and maintaining complex artificial intelligence (AI) applications with the required service agility is difficult for some organizations, the public cloud model is appealing. As result, start-up organizations and some geographies (i.e., Asia Pacific and Japan) are rapidly embracing AI training and big data analytics in the public cloud because they lack the AI expertise in-house and cannot duplicate the CSP offerings on their own. The OpEx financial structure of public cloud is another driver for AI applications, as they often start in the experimentation phase and require time to prove long-term business value—CapEx investment in these cases is not ideal.

When considering specialized services from the public cloud, verify the extent to which the service will increase subscription costs and the difficulty and cost associated with migrating to other CSPs or back on-premises once data has been moved or application code rewritten. Be sure the offerings comply with local and legal regulations, and understand the size of data being migrated off-premises and the level of back-end integrations to existing systems.

For example, in the enterprise, AI is part of a larger analytics pipeline so understanding the level of integrations with existing data sets is critical. If deep integrations to private cloud exist, consider moving the AI development into the public cloud and deploying the final application on-premises. If building and maintaining AI and machine-learning (ML) models in the public cloud require the entire data set to also be migrated, consider the cost to do so. Very large data sets will likely be more cost effective to maintain in the original environment, so build out new AI models within that same environment. Alternatively, introduce public cloud AI solutions for *new* applications and databases, keeping these distinct from already existing data sets and processes.

Personnel Cloud Expertise

Cloud technologies are still relatively new compared to traditional models, and many organizations have only begun the multi-year journey of training their development and operational engineers to use cloud technologies to their best advantage. Developers, for example, need to understand how to develop cloud-optimized applications. For organizations with a deep technical bench, chances are some early adopters have already acquired cloud expertise. On the other hand, some industries and geographies have limited cloud expertise or narrow internal knowledge of cloud technologies. With increasing cloud demand, these organizations may prefer public, hybrid cloud, or SaaS models, which allow them to contract expertise from their cloud providers instead of acquiring this expertise (which is often expensive) for their own company.

Conclusion

There is no one-size-fits-all cloud strategy or a single cloud model that can satisfy every organization's needs. We live in a multi-cloud world that requires every enterprise to conduct analysis before defining a cloud strategy for their business. Successfully determining which cloud deployment models suit the organization's business needs and workload requirements is critical to serving customers and remaining relevant in this digital age. Evaluating workload placement and creating a modern IT strategy is complex and nuanced. However, the rewards are worthwhile.

We're confident that the workload placement approach outlined in this paper can help organizations achieve a positive outcome and avoid common pitfalls and mistakes. Start by identifying the most important business needs and drivers, including conducting a full TCO analysis to determine how different combinations (modernize on-premises, integrate public cloud, adopt SaaS, and so on) will deliver the most cost-effective mix of solutions. Secondly, consider where each workload is best suited based on its technical requirements —the degree to which performance, security, integration, and data size influences workload placement plays an important role. Finally, a full investigation of the available cloud ecosystem offerings will determine what services, solutions, and providers are most suitable, taking into consideration how much cloud expertise an organization can or cannot tap into within their own staff.

Find more information about how Intel technologies power key enterprise workloads at **intel.com/yourdata**.

About Intel

Intel (NASDAQ: INTC), a leader in the semiconductor industry, is shaping the data-centric future with computing and communications technology that is the foundation of the world's innovations. The company's engineering expertise is helping address the world's greatest challenges as well as helping secure, power, and connect billions of devices and the infrastructure of the smart, connected world—from the cloud to the network to the edge and everything in between.

Appendix A: Terminology and Definitions

Cloud Implementation Models

Modern nomenclature for workload placement is simplified into two locations: on-premises and off-premises. On-premises requires that the hardware, software, and services are supported in the local private data center. Off-premises are typically the inverse, and can include public data centers and/or co-located hosted data centers; however, the basis is that the solutions are supported by a third party.

Segmenting these two locations creates four cloud implementation models:

- **Private cloud.** Virtualized compute, storage, and networking services offered either over the internet or a private internal network for select users instead of the general public. They may reside on-premises, managed internally, or can be located off-premises, managed by a third party and connected through virtual private networks (VPNs). Managed solutions may be provisioned as single-tenant or multi-tenant configurations.
- **Public cloud.** Virtualized compute, storage, and networking services offered and managed by third-party providers over the public internet, making them available to anyone who wants to use or purchase them. They may be free or sold on-demand, allowing customers to pay only per usage for the CPU cycles, storage, or bandwidth they consume.
- **Multi-cloud.** The use of multiple cloud architectures, offerings and services in a heterogeneous solution. This also refers to the distribution of cloud assets, software, applications, and so on across multiple cloud-hosting environments. A typical multi-cloud architecture uses two or more public clouds as well as multiple private clouds to eliminate the reliance on any single cloud provider. It differs from hybrid cloud in that it refers to multiple cloud services rather than multiple deployment modes (public, private, and legacy).
- **Hybrid cloud.** A computing environment that combines a public cloud and a private cloud by allowing organizations to place their applications or data on either location. Hybrid cloud provides orchestration between platforms and assumes the clouds are homogeneous. Homogeneity means portability of services between the public and private cloud is provided by the vendors of each.

Cloud Service Types

There are three primary types of cloud services:

- **Software as a service (SaaS).** Network-based services offered through commercially available software that is running on private or public clouds, including applications such as customer relationship management (CRM), productivity and collaboration tools, online file storage, and backup services.
- **Platform as a service (PaaS).** A pre-provisioned environment, offered in both public and private clouds, that includes an operating system, abstracted middleware, and infrastructure that allows developers to rapidly deploy applications without having to procure or provision servers. PaaS environments generally include self-service on-demand tools, resources, automation, and a hosted platform runtime container.
- **Infrastructure as a service (IaaS).** Virtualized hardware (compute, storage, and network) delivered as a service. IaaS is used to give developers more control over the entire application stack or to support applications that require isolation from other applications within the stack. Infrastructure services are exposed as APIs, with a user interface that provides easy access to compute, network, and storage resources.

Two emerging service models, function as a service (FaaS) and container as a service (CaaS), are still evolving and they are not specifically discussed in this paper.

- **FaaS.** An event-driven computing model in which applications, bundled as one or more functions, are uploaded to a platform and then executed, scaled, and billed in response to the exact resource demand needed at the moment. Software developers can leverage FaaS to deploy an individual "function," action, or piece of business logic.
- **CaaS.** A cloud service that allows software developers to manage and deploy containers, applications, and clusters using container-based virtualization. Container management and orchestration are becoming key services as developers move quickly into the container space and organizations seek turnkey solutions.

Other Terms Defined

- **Workload.** An application or group of applications that deliver a specific business function. Workload tasks can also be defined as: an application and/or dataset as run, with reproducible measurements; an application under real load; or a capability that can conceivably run in different physical or virtual environments; something running at a given point in time.
- **Cloud service provider (CSP).** A company that offers cloud computing services such as infrastructure as a service (IaaS), platform as a service (PaaS) or software as a service (SaaS) for purchase through the internet.
- **Latency.** The amount of time it takes a packet to travel from source to destination. Together, latency and bandwidth define the speed and capacity of a network.
- **Service.** A self-contained, modular piece of software that is used for exchanging data among systems across a network.
- **Virtual machine (VM).** Software that can run an operating system and applications to imitate a physical computer. Multiple VMs can run on a physical computer, delivering an efficient computing environment.

Appendix B: Additional Technical Requirements

This appendix includes more performance, security, integration, and data size/storage aspects to consider when determining optimal workload placement.

Additional Performance Requirements

Beyond those performance considerations discussed in the main body of this paper, organizations should also consider latency, memory and I/O, service-level agreements (SLAs), and elasticity.

Latency-Sensitive Workloads

When considering hybrid, on-premises, cloud or bare metal, all latency values should be considered. No matter the network, three primary factors significantly contribute to latency: propagation delay, routing and switching, and queuing and buffering.

- **Propagation Delay.** Propagation delay is computed as a function of distance over wave propagation speed (d/s). Propagation speed also varies based on the physical medium of the link in use. When transmitting over copper cable, propagation speed can be as low as 60 percent[10] less than the speed of light.
- **Routing and Switching.** During its journey, data passes through various controllers, routers and switches that help it reach its destination. Each one of these gateway nodes is responsible for a different task in figuring out what to do with the data. With the advent of software defined wide area networking (SD-WAN), the routing of data can take a minimal amount of time. The main delay through routers and switches is the queuing delay.
- **Queuing and Buffering.** No two networks are exactly alike. One network may not experience a high volume of traffic while another may serve a multitude of users. If one link is heavily saturated with traffic, TCP/IP attempts to avoid packet loss and preserve data integrity by placing packets in a queue for later processing. This ensures packet loss is kept at a minimum within the network. The amount of time a packet sits in a queue is referred to as queuing delay and the number of packets waiting to be processed is referred to as a buffer.

CPU, memory, storage, and network all must work together to offer the best possible performance, because any one constrained resource may negatively impact the workload. Oversubscribed compute resources will result in perceived slowness in an application, resulting in secondary resources (such as storage and network) also suffering from not having access to CPU cycles.

Increased network latency is often a symptom of oversubscribed compute as well, due to network resources being oversubscribed by high bandwidth or highly transactional services. Databases and heavy read/write storage services suffer most from latency at the storage subsystem or storage connection layers. Streaming and near-real-time network services will also incur compute and bandwidth latency often associated with network distance and network hops. Streamlining network paths and reducing compute contention can dramatically improve network performance in on-premises environments. However, if on-premises resources cannot be scaled fast enough to meet the demands for latency sensitive workloads, consider public cloud services.

Memory or I/O Bottlenecks

Applications that perform specific computations or I/O-intensive tasks may be better suited for private or hybrid cloud environments where performance is more easily tuned. To address and reduce bottlenecks, on-premises resources can be isolated, and applications can be assigned affinity rules to enable maximum memory performance. These types of optimizations are often not available in off-premises environments because host resources are not manageable unless a bare metal or dedicated hardware solution is purchased. These dedicated resources often come at significant cost compared to on-premises designs that can be customized for workloads with multiple performance-critical points. When considering the public cloud, find cloud service providers (CSPs) that offer instance classes that can be fine-tuned to the unique peaks required by a high-performant or latency-sensitive workload.

Service-Level Agreements (SLAs)

For some enterprises, there is reassurance in knowing that a CSP can offer and commit to SLAs for each application and service. Maintaining the strict level of availability and redundancy of resources on-premises can be prohibitively expensive for smaller organizations.

Larger organizations that consistently require a specific level of performance should also assess private cloud or a managed services infrastructure for reliable, detailed, and visible performance monitoring and alerts. Meeting the requirements of an SLA requires additional hardware and software on-premises, and specialized or dedicated instances in the public cloud. Conduct a cost analysis between on-premises and the public equivalent to understand the cost of achieving and maintaining the SLA. In general, SLAs with very demanding requirements on performance and availability may involve close control of the underlying hardware and software stack supporting the application. And as CSPs are continuing to add and improve services and features with a goal of delivering increased performance and availability, it's important to regularly assess the available instances types and service offerings and compare them to private/on-premises solutions.

Elasticity

Elasticity is among the top reasons that organizations choose public clouds because it provides a way to adapt to a workload's changing needs—the degree to which demand grows, shrinks, pauses, is transient, or is unpredictable. Some workloads that take advantage of elasticity include prototyping, development, test, batch processing, seasonal bursting, month-end processing, and testing new hardware capabilities.

For workloads that require extensive computing power at irregular intervals, such as online retail holiday sales or internet-based voting, public clouds provide a distinct advantage as long as the application can scale across multiple instances and does not have latency dependencies. Smaller organizations with elastic performance needs can tap into a public CSP's large pools of on-demand hardware that can be consumed when needed—a significant benefit if this capability could be cost-prohibitive to provide on-premises.

There are also private cloud solutions that take advantage of scale-out architecture to manage spikes in demand and provide a good customer experience. For example, some providers now offer new forms of on-premises elastic computing with consumption-based models. Extra capacity is placed on-premises but there is no cost until the customer bursts onto that reserved capacity. It is also possible to consider the availability of spare capacity for scaling with on-premises infrastructure as a service (IaaS) solutions before it is actually needed. Although this costs additional capital up front, forward-thinking organizations can plan ahead and easily mimic the agility and scale of public cloud solutions.

Additional Security Requirements

In addition to those security considerations discussed earlier, organizations should also take into account organizational culture, data breach sensitivity, and security audits.

Organizational Culture and Sensitivities

While most public cloud platforms can comply with data privacy and security regulations, some organizations still prohibit storing data outside of the corporate domain. Many sophisticated IT organizations have policies and practices to reduce the number of attack points across configuration, network access points, management interfaces, administrative accounts, and shared multi-tenancy. Typically, they also require protection controls for data at rest, runtime, and in transit, while certifying the underlying infrastructure from malware and misconfiguration. IT organizations with a deep cultural history of data security practices may be more comfortable maintaining sensitive workloads on-premises, even though public clouds are very secure.

Workload Data Examinations

In an age where attack surfaces are ever-growing, data breaches are even more prevalent. It is critical to understand exactly the nature and content of what is being stored (for example, unique business value, mission-critical intellectual property, and customer data) in every workload under consideration for migration to ensure the data can move outside the enterprise domain. A thorough analysis of data storage is highly intrusive and typically requires deploying agents into the OS on which the application runs. These agents scan the file system or a database. One agentless approach that is not intrusive and much more lightweight is to identify the application and its associated metadata tags; the tags help indicate data privacy and data security classifications. Using that information, a workload's classification will determine whether it can be located outside of the domain according to company policies.

Security Audits

The ability to perform security audits is a critical element in meeting security and regulatory attestation requirements. To prove they are secure, most IT organizations require transparency and tracking in environment management, changes, and transaction execution. If an organization is subject to these expectations, security auditors may prefer more direct control over, and accessibility to data, and therefore prefer an on-premises environment. Many public cloud providers have also enhanced audit and compliance support; be sure security audit terms and SLAs meet your requirements.

Additional Integration Requirements

Legacy Systems

Legacy systems raise additional integration concerns because legacy systems integration can unexpectedly add weeks or even months to the implementation schedule. Very few enterprise workloads are stand-alone, and many have multiple components. They often require resources from other teams within the enterprise, and integrate with other applications to support critical organization functions, such as enterprise resource planning or product lifecycle management. It is also important to consider enterprise security and authentication systems such as a directory services application, single-sign-on, and identity and access management applications.

Some legacy or third-party applications may not have open APIs, making it challenging to integrate with new systems. It may be cost-prohibitive to replace, migrate, or modify these applications without addressing the dependencies, and re-integrating services from other systems with external solutions can add significant time to the migration schedule. Additionally, migrating an application to a different cloud can increase wide area network (WAN) traffic and costs if the dependent applications remain on a different environment. Most organizations find more value in connecting traditional and modern cloud applications through services instead of rewriting an entire application. This enables them to manage both traditional and modern cloud-aware applications on a common infrastructure, delivering new capabilities and automations using cloud techniques alongside mature solutions. Such an approach delivers both agility and the cost-effectiveness of cloud solutions.

Cloud Orchestration

For organizations with many types of cloud computing in use simultaneously across an organization, there is a need to simplify the complexity with cloud orchestration tools and strategies. Cloud orchestration refers to the arrangement and coordination of automated tasks resulting in a consolidated process or workflow. Many organizations implement automation in an ad hoc, opportunistic fashion that yields islands of automation with low operational agility and high costs. IT operations leaders should avoid these types of implementations, which introduce risk and result in expensive mistakes. Cloud orchestration offers a systematic approach that maximizes the automation and integration between clouds, providing a unified portal and cloud-inspired IT service model with full-stack automation and monitoring. Planning for a single cloud orchestration strategy results in accelerated delivery for new innovations, applications, and infrastructure, and ultimately delivers a better customer experience.

Additional Data Size/Storage Requirements

Data growth and location are important technical considerations. Large data volumes can be expensive to store, difficult to migrate, and may increase transaction time. Yet, storing data is critical to workload placement and the lifeblood of any organization. The reality is that as data grows, the cost for both on- and off-premises increases significantly.

Organizations can find opportunities to improve internal storage infrastructure, reduce cost, and improve operational efficiencies by taking advantage of both new storage innovations in their own data centers (such as software-defined storage), and integrating public cloud storage offerings (such as back-up and recovery). When working with large data stores, locating the data nearby the business applications using the information is an important design decision for both public and private cloud solutions. When looking at where data will live, consider where the primary consumers of this data reside and with what regularity the data is expected to grow, change, and move. Reads cost more than writes for most cloud solutions, so understanding the data access profile helps determine which solution most benefits the application.

1 Source: IDC's Cloud and AI Adoption Survey, Jan 2018, Doc #DR2018_GS4_MB; February 2018; twitter.com/IDC/status/1092531479278702594

2 Source: Taking on the World case study; January 2019; intel.ai/wp-content/uploads/sites/69/Descartes-Lab-A-GeoSpatial-Google-Cloud-Intel-Case_-002.pdf; and Descartes Labs Helps Customers Understand the Planet; Intel® Chip Chat episode 625; December 2018; soundcloud.com/intelchipchat/helping-customers-understand-the-planet

3 Intel customer engagement.

4 IDC's Cloud and AI Adoption Survey; January 2018; twitter.com/matteastwood/status/1095482380134072320

5 Right Scale, Cloud Computing Trends: 2018 State of the Cloud Survey; rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2018-state-cloud-survey

6 GoDaddy Reduces Infrastructure TCO with Hybrid Cloud; intel.in/content/www/in/en/cloud-computing/godaddy-reduces-tco.html; and GoDaddy Moves to Hybrid Cloud and Optimizes Infrastructure TCO; Intel® Chip Chat episode 620; December 2018; soundcloud.com/intelchipchat/godaddy-hybrid-cloud-tco

7 Performance results are based on testing as of June 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit intel.com/benchmarks. **Configuration details:** Per node 4X higher integer throughput performance: estimate based on SPECrate*2017_int_base on Intel internal platforms as of June 2018: 1x node, 2x Intel® Xeon® Processor E5-2690, 128GB total memory, 16 slots/8 GB/ 1600MT/s DDR3 RDIMM, Benchmark: SPEC CPU2017 V1.2, Compiler: Intel® Compiler IC17 update 2, Optimized libraries/versions: IC18.0_20170901, Other Software: MicroQuill SMART HEAP. uCode: 713, OS: Red Hat Enterprise Linux* 7.4, Kernel: 3.10.0-693.11.6.el7.x86_64 x86_64, Score 65.5 vs. 1x Node, 2x Intel® Xeon® Platinum 8180 Processor, 384GB total memory, 12 slots/32 GB/2666 MT/s DDR4, Benchmark software: SPEC CPU® 2017, Compiler: Intel® Compiler IC18 OEM, Optimized libraries: AVX512, ucode:0x043, Red Hat Enterprise Linux* 7.4, 3.10.0-693.11.6.el7.x86_64, Score: 281. Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction. **Configuration details:** Up to 60 percent TCO savings with Intel® Xeon® Scalable processor compared to 5-year old system. Example based on estimates as of June 2018 of equivalent rack performance over 4-year operation on integer throughput workload (estimate based on SPECrate*2017_int_base on Intel internal platforms) running VMware vSphere Enterprise Plus on Red Hat Enterprise Linux Server and comparing 20 installed 2-socket servers with Intel® Xeon® processor E5-2690 (formerly "Sandy Bridge-EP") at a total cost of $737,460 [Per server cost $36.8K: acquisition=12,5K, infrastructure and utility=4.5K, os & software=10.2K, maintenance=9.7K] vs. 5 new Intel® Xeon® Platinum 8180 (Skylake) at a total cost of $294,540 [Per server cost $58.9K: acquisition=12,5K, infrastructure and utility=10.1K, OS & software=10.1K, maintenance=9.7K]. Assumptions based on xeonprocessoradvisor.intel.com, assumptions as of June 6, 2018.

8 Source: CloudPhysics, Inc. customer data

9 2016-2017 Intel IT Annual Performance Report; intel.ly/2qXnCaz

10 Copper vs. Fiber – Which to Choose? Multicom, Inc.; multicominc.com/training/technical-resources/copper-vs-fiber-which-to-choose