BUILDING VALUE: YOUR CRITICAL WORKLOADS
The benefits of running enterprise workloads in the cloud are clear. 87 percent of enterprises say they are able to drive business acceleration using the cloud, while 41 percent attribute business growth to their cloud deployments\(^1\). In addition, 30 percent use the cloud to launch new products, speed time to market, and expand into new markets\(^2\).

However, deciding to run or migrate critical workloads to the cloud is just the first step towards getting more performance for every dollar of your investment. It’s critical to determine the right – or optimal – cloud infrastructure for the size and purpose of your organization. Even within your enterprise, you may have different workloads that require very different strategies. In short, a one-size-fits-all approach won’t deliver.

Intel and Amazon Web Services\(^*\) (AWS\(^*\)) are collaborating to help companies place their critical workloads and achieve the cloud performance they need at the value their CFO demands. Together, Intel and AWS have jointly serviced hundreds of real-world workloads and have optimized infrastructure hardware and software for these diverse workloads based on their particular characteristics. These optimizations make our customers’ lives much easier and their cloud deployments much faster\(^3\).

Whether you are looking to buy, build or expand your enterprise cloud workloads, this guide is designed to help you explore the key considerations you should be thinking about.
ACHIEVE A LOW TCO ACROSS CRITICAL WORKLOADS

Today, the IT strategy of choice for many enterprises involves running multiple cloud environments. 81 percent of enterprises report having a multi-cloud strategy, and on average use 4.8 clouds across public and private⁴. However, many companies have also reported that achieving an optimal total cost of ownership (TCO) required a process of trial and error⁵. For example, 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 learnings.

Developing an optimal TCO strategy requires a holistic approach, considering the full context of your enterprise’s business priorities and operating environment. This framework has been designed to help you think through some of the core considerations you should take into account.

Of enterprises report having a multi-cloud strategy
FRAMEWORK FOR EVALUATING TCO ACROSS CRITICAL WORKLOADS

Identify your top business objectives

Depending on your pain points, use cases and growth opportunities these could include:

- Reducing data center costs or improving efficiencies.
- Reducing turnaround time, or faster time to market for new products/services.
- Retaining or improving control of mission-critical services.
- Helping to ensure adequate protection of sensitive data to comply with local or global regulatory requirements.
- Filling in-house resource gaps or boosting in-house skills.
- Meeting the requirements of a Service Level Agreement (SLA).

Conduct a cost analysis

Consider the expenditures and opportunities associated with your workloads. Factors to consider include:

- The upfront cost of migrating workloads to the cloud, and/or ongoing operational expenditures – including network and uptime requirements.
- Required connections to other databases, frameworks and applications.
  The complexity, quantity and location of multiple data sources can impact the workload deployment because of the cost of integrating into multiple clouds.
- The total cost of your operating environment, as well as the cost per workload.

Compare options for each critical workload

Any overarching TCO analysis should also include an assessment of the characteristics of individual workloads. It’s important to consider:

- The level of performance required – including any resource-intensive requirements, and whether (and for how long) you can afford downtime.
- If any workloads will be processing intellectual property or sensitive personal information that might require additional layers of security.
- The size and location of data managed and processed by each workload. If you’re just getting started, you may wish to begin with a less demanding workload before scaling.

Get buy-in from key stakeholders across the organization

Compared to traditional models, cloud technologies are still relatively new. So, unless your enterprise was an early adopter, levels of familiarity with what’s required to run enterprise workloads in the cloud may be uneven. Considerations for securing buy-in naturally vary from team to team:

- IT/Engineering: Consider the degree of in-house support your team might need, depending on their levels of cloud expertise.
- Finance: Try to define the up-front and ongoing cost implications of running critical workloads in the cloud, and the potential cost-benefit of doing so.
- C-Suite: Explore how cloud deployments can support competitive advantages, and help your enterprise deliver on its wider strategic business objectives.
A comprehensive analysis of all the factors related to TCO across your critical workloads may lead you to conclude that – for now at least - not all your applications should operate in the cloud.

That's why it's also important to consider specific enterprise-critical workloads in their own terms to better understand if and how it is possible to deliver the performance you need for the value you want.

This section explores specific critical workload instances, and how enterprises can ensure they are set-up to run at a low TCO.

**A. DATABASE WORKLOADS**

Hosting your database in the cloud can give you greater reliability, scalability and flexibility. Considering database workloads are one of the most likely to suffer from latency at the storage subsystem or connection layers this is a key value proposition, especially when it comes to scaling. But, closely monitoring budgets remains important.

To get more value for money, choose an Amazon EC2 memory-optimized instance featuring the Intel® Xeon® Scalable processor. For HammerDB® and PostgreSQL® workloads, it delivers up to 1.85x the performance per dollar, and for MongoDB®, it can achieve up to 2.84x the performance per dollar.

There are further benefits from the savings too: if you can do more with fewer vCPUs, you may be able to make additional savings on your software licensing costs. Moving from a C4 instance featuring the Intel® Xeon® processor E5-2666 v3 to a C5 instance featuring the Intel Xeon Scalable processor cuts the core count by 40%. If a commercial database is licensed at a cost of $1,800 per core annually, that adds up to $72,000.

**B. HIGH PERFORMANCE COMPUTING (HPC) AND ARTIFICIAL INTELLIGENCE (AI)**

Many organizations turn to the public cloud for AI training and big data analytics because it can otherwise be challenging to achieve the required service agility. Similarly, most HPC processing is still conducted on-premise, where there is often competition for limited capacity, risking bottlenecks.

Choose an Amazon EC2 instance running on the Intel Xeon Scalable processor to get the low TCO you want and the high performance you need. According to test results from two popular benchmarks (LAMMPS® and High-performance Linpack®), Amazon EC2 instances featuring the Intel Xeon Scalable processor offer 4.15x higher performance per dollar.

If you’re using a compute-optimized EC2 C instance, which AWS recommends for science and engineering applications, you can cut the core count by 40% if you switch from an older c4.8xlarge instance to c5.4xlarge. If you’re licensing software per core, there can be huge cost savings.
C. MEMORY AND COMPUTE-INTENSIVE

Demanding workloads can put a lot of pressure on both memory and compute. Oversubscribed compute resources can result in perceived slowness in an application, which can have resulting impacts on resources such as storage and network, which will also struggle without access to CPU cycles. Equally, insufficient memory can create bottlenecks.

Public cloud instances can provide the scale necessary to meet the demands of such latency sensitive workloads. By using Amazon EC2 instances featuring Intel Xeon Scalable processors, you can get up to 2.25x higher performance per dollar\(^v\), as estimated using the STREAM\_OMP Triad benchmark. For floating point maths, M5 instances featuring the Intel Xeon Scalable processor deliver up to 1.85x the performance per dollar\(^vi\) of M5a instances, according to tests based on the SPEC\_rate2017\_fp\_base benchmark.

D. WEB-BASED

For web-based workloads, speed is critical to delivering a competitive customer experience. The longer a user has to wait for a page to download, the more likely it is they will abandon the website altogether. Despite recent reductions, the average loading time for a mobile webpage is 15 seconds\(^v\). However, 53 percent of mobile site visitors will leave a page that takes longer than three seconds to load. In other scenarios, a web-based workload might be integrated into a business process, in which case other processes are waiting for it to complete. Balancing these high wire demands while keeping costs at a manageable level is challenging.

Choose the Intel Xeon Scalable processor to unlock savings. A 96 vCPU Intel Xeon Scalable processor instance delivers up to 1.74x higher performance per dollar\(^vii\). That’s true for three common web workloads: server side Java\(^v\), Wordpress\(^v\), and PHP/HHVM\(^v\).
WHY CHOOSE AMAZON EC2 INSTANCES FEATURING INTEL® TECHNOLOGY?

Clouds built on Intel® architecture deliver the most predictable, industry-leading per core performance and workload acceleration. With Intel’s rich experience working with enterprises on mission-critical business infrastructure and meeting the stringent service level agreements (SLA) of business applications, Intel Xeon processors boast industry-leading server uptime and reliability.

Intel has a long history of working with AWS, including deep investment with the software ecosystem to optimize code for our architecture, which has led to broad workload performance.

Workloads running on Intel architecture-based AWS instances can take advantage of these optimizations to increase performance and security, but those software-optimization gains may be lost if the virtual machine (VM) is transferred to a non-Intel instance.

This is especially important if your business is global, and you need to consider instance availability around the world. To minimize TCO at worldwide scale, one possible strategy is to replicate identical VMs in identical instances wherever a local presence is required to meet your own SLAs or comply with applicable laws. The cost of creating and maintaining both Intel and other VM versions of the same service in non-uniform instances may offset any cost savings associated with instance price.
How much data do you currently store and what are your data growth predictions?

If you’re running cloud workloads already, where does your sensitive data live currently?

How important are business continuity and disaster recovery (BC/DR) practices to your workloads?

What security risks do you need to plan for, and how might these change over time?

How accessible is cloud expertise in your organization?

Would your workloads benefit from burst capacity to cover irregular demand?

How much elasticity do your workloads require, now and in the future?

How do your cloud instances integrate with your wider IT infrastructure?

PERFORMANCE

Eligibility is a key driver of public cloud adoption. It provides a way to adapt to a workload's changing needs, which can often be unpredictable. This is particularly helpful for prototyping, development or testing new hardware capabilities, for example.

Would your workloads benefit from burst capacity to cover irregular demand?

Workloads such as online retail which may be prone to bursts during promotions or holidays can benefit from the scalability across multiple instances that public cloud offers.

Don’t assume that on-premise is the best option for workloads that include sensitive data. 52 percent of companies say they experience better security in the cloud. Companies using the cloud with Cloud Access Security Brokers (CASB) to protect their data are also able to accelerate faster.

Your CIO will expect full compliance with security and regulatory requirements – this is even more critical if your workloads require you to perform regular security audits. Be sure to understand your company’s handling policies for sensitive information before migrating or scaling critical workloads.

Your CIO will expect full compliance with security and regulatory requirements – this is even more critical if your workloads require you to perform regular security audits. Be sure to understand your company’s handling policies for sensitive information before migrating or scaling critical workloads.

It’s important to understand the strengths and limitations of your internal workforce, not just in IT, but other business-critical areas you will need to successfully migrate critical workloads to the cloud. You may, for example, need to engage an external consultancy to help fully evaluate TCO.

DATA MANAGEMENT

If you’re working with large data volumes, locating the data close to the business applications that use the information is an important design decision, both for public and private cloud solutions. Large data volumes can otherwise be expensive to store and difficult to migrate, so it's important to plan ahead if you expect data volumes to increase.

BC/DR is one of the most common public cloud workload usages because it enables data to be stored in multiple locations worldwide without investing in data center facilities. For workloads designed to take advantage of cloud-aware architectures, business continuity can be significantly improved.

As you continue to build your enterprise cloud strategy – whether to deploy for the first time or to scale – this checklist provides you with some key, actionable questions to consider before you invest.
CONCLUSION

Every enterprise – and workload – is different. Whether you’re just getting started or preparing to scale, it’s important to identify the next steps best suited to your critical workloads. As this guide has shown, these will depend on your specific TCO, performance, security and talent requirements. Ultimately by underpinning your critical workloads with an Amazon EC2 instance powered by the latest Intel Xeon Scalable processors, you can ensure you deliver the performance you need at the value your CFO demands.


⁹ https://www.thinkwithgoogle.com/marketing-resources/data-measurement/mobile-page-speed-new-industry-benchmarks/


³ https://itpeernetwork.intel.com/intel-aws/#gs.molao8


¹ Results calculated by Intel P2CA using AWS pricing ($/hour, standard 1-year term, no up-front) as of 12th January, 2019. Performance per dollar testing done on Amazon EC2® R5 and R5a instances (https://aws.amazon.com/ec2/instance-types/), comparing 96 vCPU Intel® Xeon® Scalable processor performance per dollar to Competitor processor performance per dollar.

Workload: HammerDB® PostgreSQL

Results: Competitor processor performance per dollar = baseline of 1; Intel® Xeon® Scalable processor performance per dollar = 1.85X (higher is better)

Database: HammerDB® – PostgreSQL [higher is better]

Amazon EC2 r5.24xlarge (Intel) Instance, HammerDB 3.0 PostgreSQL 10.2, Memory: 768GB, Hypervisor: KVM; Storage Type: EBS(io1), Disk Volume 200GB, Total Storage 200GB, Docker version: 18.06.1-ce, RedHat® Enterprise Linux 7.6, 3.10.0-957.el7.x86_64, 64, 6400MB shared_buffer, 256 warehouses, 96 users. Score "NOPM" 439931, measured by Intel on 12/11/18-12/14/18.

Amazon EC2 r5a.24xlarge (Competitor processor) Instance, HammerDB 3.0 PostgreSQL 10.2, Memory: 768GB, Hypervisor: KVM; Storage Type: EBS(io1), Disk Volume 200GB, Total Storage 200GB, Docker version: 18.06.1-ce, RedHat® Enterprise Linux 7.6, 3.10.0-957.el7.x86_64, 64, 6400MB shared_buffer, 256 warehouses, 96 users. Score "NOPM" 212903, measured by Intel on 12/10/18.

Results: Competitor processor performance per dollar = baseline of 1; Intel® Xeon® Scalable processor performance per dollar = 4.15X (higher is better)

Workload: Server Side Java

Server Side Java (higher is better):

Amazon EC2 m5.24xlarge (Intel) Instance, LAMMPS version: 2018-08-22 (Code: https://github.com/lanl/lammps#download.html), Load: 512X Particles, Intel ICC 18.0.3.20180410, Intel MPI Library for Linux® OS, Version 2018 Update 3 Build 20180411, 48 MPI Ranks, RedHat® Enterprise Linux 7.5, Kernel 3.10.0-682.67_x86_64, OMP_NUM_THREADS = 2, Score 137.5 timesteps/sec, measured by Intel on 11/16/18.}

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 changes 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 www.intel.com/benchmarks.