

The Economics of Infrastructure Scalability

Find the most cost-effective
and least risky way to scale



Executive Summary

Customers today have high standards when it comes to service delivery. If a website or application momentarily grinds to a halt, they won't stick around to wait for it to come to life again. Downtime can be extremely costly, which is why scalability is so important.

The question is, which scalability strategy will enable your business to deliver always-on services to customers without breaking the bank? What are the pros and cons of scaling on premises, into public cloud, or into hybrid cloud?

Discover how to weigh the risks and benefits of each option so you can choose the right scalability strategy for your business.

What Is Cloud Economics?

Cloud economics puts tried-and-true economic theories to work, helping IT managers make better, more cost-effective decisions about cloud adoption and usage.

A blend of traditional and behavioral economics, cloud economics suggests that you need to weigh the best choice from both cost and performance perspectives, while being aware of common biases and blind spots that affect cloud decisions.

The impact of cloud on the bottom line needs to be considered from multiple angles, not just the upfront cost. Rather, it should consider things such as:

- Refactoring and rework – How long might it take to refactor applications to run on public cloud?
- Talent reskilling – Are your team's current skills transferable to new platforms?
- Environment lock-in – Can your data center or applications be easily scaled to a new environment?
- Operational costs – Are there ongoing operational costs that you may not be aware of?

As well as potential costs, you also need to weigh the risks. Every IT project—cloud notwithstanding—will carry risk. Your job is to work out whether the risks, be they financial or performance, outweigh your risk tolerance. If they do, it might be time to explore a new option.

Find out how to measure your risk tolerance and exposure in our complete guide to cloud economics.

[Download now](#)

The Ability to Scale on Demand Is Non-Negotiable

Across industries, organizations face more pressure than ever before to deliver round-the-clock service and a flawless customer experience. As well as speed, simplicity and personalization, 24x7 availability is now a minimum expected requirement.

And if you don't deliver? Your customers may look elsewhere.

In retail, for example, one in four shoppers will abandon their shopping cart if a website crashes.¹ Retail events, such as Black Friday, have been known to bring websites to a standstill—and brands to their knees—with eager shoppers unable to grab a bargain because their page won't load.

In the finance industry, where daily transactions can stretch into the millions, the ability to scale securely is mandatory. Yet many struggle to strike the right balance between having the resources to meet spikes in demand while avoiding wasteful overprovisioning. As much as 35 percent of data center capacity goes unused across industries.²

To provide uninterrupted availability in a cost-effective manner, enterprises must be able to scale their data center usage to meet the ebbs and flows of market demand.

MOBILE RETAILERS

Mobile retailers have risked losing as much as \$300 million in sales on Black Friday due to app crashes.³

1. Shopify. "Why Online Retailers Are Losing 67.45% of Sales and What to Do About It." Michael Zipursky. August 6, 2013.
2. RightScale. "RightScale 2018 State of the Cloud Report." February 2018.
3. Crowd Research Partners. "2018 Cloud Security Report." March 2018.



3 Ways to Scale

Right-sizing your infrastructure capacity to meet current and future demands requires some planning. Do you invest in more hardware? Do you take things to the public cloud? Or is there another way?

Before you can answer these, you need to know what you're scaling.

Up or out?

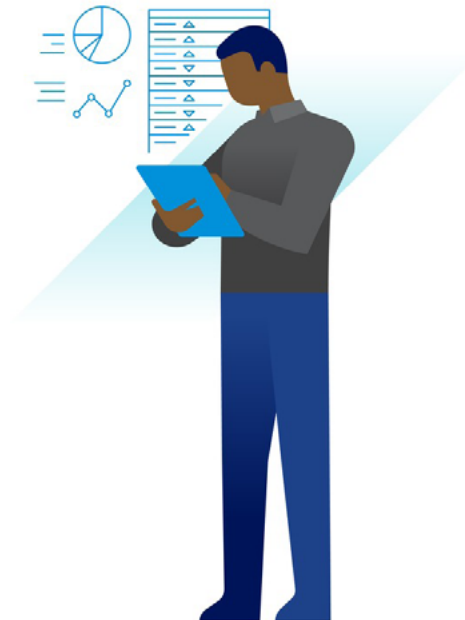
There are two approaches you can take when addressing scalability.

Scaling up involves adding additional resources (CPU, RAM and storage) to an existing virtual machine (VM) to make it more powerful. A key benefit of scaling up is that you don't need to re-engineer applications; the infrastructure remains the same. Front-end webpages are generally easy to scale up in a high-traffic period. Yet when customers start interacting with shopping carts or apps that connect to a database, scaling these services can be harder.

Scaling out involves adding new VMs to an application to increase the number of discrete processing units. By spreading the load, scaling out can minimize the risk of downtime and improve availability. Databases and other apps that use proprietary APIs and software, or have complex operational parameters and service-level agreements (SLAs), can be much harder to meet.

Whether you're planning to scale up or out, the good news is you've got options: on premises, public cloud, or hybrid cloud. But, as any prudent economist would advise, it pays to do your research before you pursue a particular scaling strategy.

Through the lens of cloud economics, let's take a look at the risks and benefits of the three ways to inject scalability into your IT environment.



Option 1: Buy more machines

Purchasing new infrastructure to boost the capacity of your on-premises environment is one way to scale. This can get costly, though. Not only is hardware expensive, but the resources that lie idle during quieter periods are effectively money down the drain.

Risks and challenges of on-premises scaling

- It demands capital expenditure – Building out your on-premises infrastructure to cope with scalability demands can be a massive line item on the balance sheet. Not only is there the infrastructure itself, there are the real estate and utility costs, too.
- The cost of standby capacity – It costs a lot to maintain on-premises data centers, so you want to make sure you make the most of your investment.
- It takes time – Unless you have the expansion resources powered up, configured, secured and accessible to the application (which is a cost in itself), you can't just flick a switch and have extra capacity ready to go. The procure, configure and deploy tasks all take time, and this time can prove costly.
- It may not be enough – A big risk with on-premises scalability is its finite ceiling. You may think you have enough capacity, but a big spike in workload could cause an app to crash or slow things down significantly.

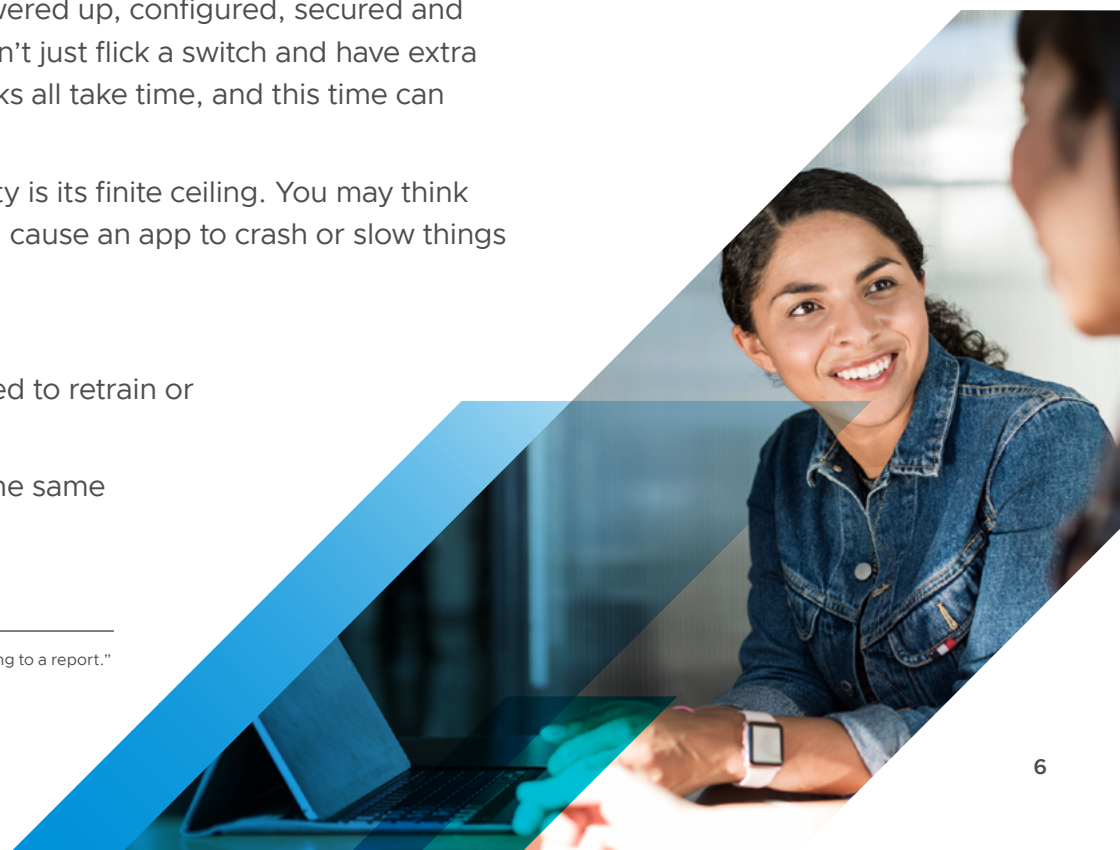
Benefits of on-premises scaling

- Familiar environment – There is less likelihood that you'll need to retrain or upskill employees.
- App compatibility – If you build out your data center using the same infrastructure, there's no need to refactor apps.

4. Business Insider. "Companies waste \$62 billion on the cloud by paying for capacity they don't need, according to a report." Becky Peterson. December 2, 2017.

CAPACITY WASTAGE

Companies waste \$62 billion on redundant capacity each year.⁴



Option 2: Scale into public cloud

It's possible to scale up or out in a public cloud environment. Scaling up involves paying more for extra grunt by increasing the size of an instance. Scaling out involves adding new instances, which also need new load balancers and schedulers.

The public cloud is certainly on the enterprise radar. Any workloads migrated to public cloud will be able to tap into the inherent scalability of cloud services.

But for workloads that remain on private clouds, scaling into public cloud can create more headaches than benefits.

Risks and challenges of public cloud scaling

- The cost of re-engineering apps – If you're scaling out in public cloud, applications may need to be rewritten to suit the new environment. This can cost hundreds of thousands of dollars per application.
- The cost of add-ons – Depending on the public cloud provider, you may need to develop and configure the VMs, load balancers and schedulers for the appropriate workloads. This can take time and money.
- Service interruptions – Disruption to an application's performance is common when scaling into public cloud (and again when contracting back to free up scaled resources). Consider the cost of these interruptions to the customer experience or to business performance, compared with the cost of providing mitigation to avoid disruption.
- Upskilling or hiring new staff – Moving into a new environment such as public cloud can require significant retraining of your IT team. Some organizations even hire new staff who bring the requisite skills to the job.
- Subscription costs – The cost of the public cloud subscription itself should also be taken into account.

5. LogicMonitor. "Cloud Vision 2020: The Future of the Cloud Study." January 2018.

ENTERPRISE WORKLOADS

It's predicted that 41 percent of enterprise workloads will be run on public cloud platforms by 2020.⁵



Benefits of public cloud scaling

Near-infinite expansion – Additional resources are always available in public cloud. You can either resize VMs to scale up, or add additional instances to scale out.

A key difference in scalability: public cloud compared to hybrid cloud

On the surface, it would seem that the cost of scaling into public cloud would be the same whether you adopt a hybrid approach or not. You still need to subscribe and add new hosts when needed, right?

A key difference lies in the automation of the scaling process. Hybrid cloud solutions, such as VMware Cloud™ on AWS, automatically scale and attach a new host node when resource consumption reaches a trigger threshold. It takes a matter of minutes.

In public cloud configurations, IT staff would need to purchase and configure the new host—including load balancers, schedulers and so on—to be ready for usage. While much of this can be automated, it still costs more and takes longer than the hybrid approach.

Option 3: Take a hybrid approach

Hybrid cloud is proving a popular choice for businesses seeking to solve the scalability puzzle. By adopting a hybrid cloud approach, organizations can maintain mission-critical data and applications on premises, yet can quickly expand into the powerful realm of cloud computing when workloads demand it.

Risks and challenges of hybrid cloud scaling

- Possible incompatibilities – Depending on your hybrid cloud infrastructure, there can be a lack of compatibility between environments. If so, then the cost of re-engineering applications to scale across these environments must be considered.
- Inefficient use of budget – Even if you only need a single VM, you may need to subscribe to an entire host when resource consumption reaches a trigger point (this is also true for public cloud scaling), which may mean some resources lie idle.



Benefits of hybrid cloud scaling

- Take advantage of elastic capacity – Scale cloud usage to meet customer demand, taking advantage of the elastic capacity of public cloud. When your hybrid cloud environment is set up correctly, you can scale capacity up or down without change or friction for any workload.
- Use cloud for disaster recovery and backup – Realize cost savings by leveraging the cloud for disaster recovery and backup, instead of maintaining a secondary data center.
- Grow into new regions faster – Scale workloads to specific regions and rapidly expand into new markets, while complying with rigorous regulation, such as the General Data Protection Regulation (GDPR).
- Protect existing investments – Protect, extend or consolidate existing on-premises data center investments.
- Do not need to retrain staff – Continue using your existing teams, skillsets, tools and processes if your hybrid cloud environment uses consistent infrastructure and operations.
- Gain uninterrupted service delivery – Keep your customers happy and your team productive by responding to demands for increased capacity in an instant.

Weighing the Risks and Benefits

Any scale-on-demand strategy—be it on-premises (traditional three-tier), public cloud, or hybrid cloud—comes with its challenges. These challenges can have a flow-on effect to your bottom line, so it’s important to weigh them all carefully.

Risks	On-premises (traditional 3-tier)	Public cloud	Hybrid cloud*
Capital expenditure	•		
Unused capacity	•		
Refactoring applications		•	
Add-ons (load balancers, schedulers etc.)	•	•	•
Subscription costs		•	•
Time to scale	•	•	
Limit on scalability	•		
Hiring or upskilling IT staff		•	

*Assumes the hybrid cloud environment supports seamless scalability with consistent infrastructure and consistent operations, as with VMware Cloud on AWS.



Benefits	On-premises (traditional 3-tier)	Public cloud	Hybrid cloud*
Easy to manage	•		•
Consistent environment	•		•
Application compatibility	•		•
Capacity		•	•
Speed			•
Protect existing investments			•
Seamless delivery of services			•

*Assumes the hybrid cloud environment supports seamless scalability with consistent infrastructure and consistent operations, as with VMware Cloud on AWS.



Achieving True Scalability with VMware Cloud on AWS

VMware Cloud on AWS is built for scalability. This best-in-class hybrid cloud service seamlessly supports workloads on premises and in the public cloud, and gives IT the flexibility to choose where the workloads run. Better still, it tackles the risks and challenges of on-premises or public cloud scaling head on.

Key scalability capabilities of VMware Cloud on AWS

- It brings the VMware enterprise-class software-defined data center (SDDC) to the AWS cloud, delivered as an on-demand service with access to AWS offerings, including storage, databases, analytics and more.
- You own the entire host, which means that all of its additional resources are immediately available to an application that's working harder than usual. Built-in VMware vSphere® functionality enables new resources to be added seamlessly when required.
- Additional VMs can be provisioned and deployed for applications using automation, and the SDDC stack can easily handle the load balancing and scheduling for requests hitting the application. If you hit the host performance threshold, additional VMware ESXi™ hosts are acquired, provisioned, configured and added to the cluster as needed before services are impacted.
- VMware Cloud Foundation™ delivers consistent infrastructure and operations, so you can leverage existing skills, tools and processes.
- It includes VMware HCX®, the application mobility platform designed to simplify application migration, rebalance workloads, and optimize disaster recovery across data centers and clouds.

Overall, VMware Cloud on AWS enables seamless scaling up or out at a fraction of the cost of other options.

“With VMware Cloud on AWS, we can provide agile, reliable service that—when combined with Trend Micro’s Deep Security—offers consistent management, visibility, and protection across our data center and cloud environments, saving precious IT resources.”

ALEX KUO
SENIOR DIRECTOR OF IT,
TREND MICRO INC.

The Economic Benefits of VMware Cloud on AWS

Many hybrid cloud solutions require significant investment of time, money and IT resources. VMware Cloud on AWS is delivered as a service on a single, consistent infrastructure across a seamless hybrid environment, effectively eliminating the IT challenges caused by incompatibility between on-premises and public cloud environments. This delivers a range of business benefits.

Increased agility (scale on demand)

To remain competitive and deliver on your customers' expectations, you often don't have the luxury of time to add extra capacity to your infrastructure. You need to respond at the speed of your business.

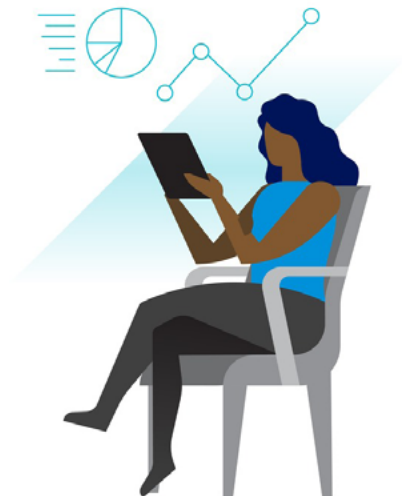
VMware Cloud on AWS gives your business new agility to meet the demands of customers for always-on service by enabling IT to move workloads with little or no downtime. This agility boosts customer engagement and, in turn, the bottom line.

As Table 1 shows, IT can deploy VMs on hybrid cloud much faster than previous environments. What this means for the business, as outlined in Table 2, is that more applications and features can be developed and released at a faster pace, which ultimately helps to meet and exceed customer expectations.



Deployment of new physical servers	Previous environment	With VMware Hybrid Cloud	Increased value with VMware Hybrid Cloud	Benefit with VMware Hybrid Cloud (%)
Time required per server (days)	39.5	1.4	25.9	66
Staff time required per server (hours)	38	17	20.9	55
Deployment of VMs				
Time required per VM (hours)	0.5	0.3	0.2	33
Staff time required per VM (minutes)	7.5	4.1	3.4	46

TABLE 1: Impact on IT and business agility.⁶



6. IDC White Paper, sponsored by VMware. "The Business Value of Running Applications on VMware Cloud on AWS in VMware Hybrid Cloud Environments." October 2020.

Number of new applications/features developed	Previous environment	With VMware Hybrid Cloud	Increased value with VMware Hybrid Cloud	Benefit with VMware Hybrid Cloud (%)
Number of new applications per year	9.8	15.1	5.3	5.4
Number of new features per year	14.5	21.2	6.7	46
Development lifecycle (weeks)				
New applications	15.8	15.4	0.4	2
New features	3.5	2.9	0.6	18
Productivity of impacted developers				
Developer productivity per organization (FTEs)	11.6	14.5	2.9	25

TABLE 2: Application development KPIs.⁷

7. IDC White Paper, sponsored by VMware. "The Business Value of Running Applications on VMware Cloud on AWS in VMware Hybrid Cloud Environments." October 2020.



Lower total cost of ownership (TCO)

VMware Cloud on AWS delivers clear cost advantages over traditional, multitier deployments. There is no need to re-engineer applications to suit native-cloud environments. There are no costly resources sitting unused and idle in non-peak periods. Your business can also shift from a CapEx to an OpEx model for building out capacity.

Economies of scale

Across the board, VMware Cloud on AWS delivers significant economies of scale. With consistent infrastructure and operations across your IT environment, you can use existing resources without having to upskill, retrain or hire new staff.

Further economies of scale are realized with VMware Cloud on AWS thanks to the ability to rapidly allocate and consume additional resources, without having to configure and purchase additional instances every time. It is a very efficient and cost-effective way for IT to meet application SLAs.

Management gets easier, too. Policies and procedures don't need to change—everything is standardized and familiar. From a business perspective, this consistency saves time, makes compliance easier to control, and enables you to focus on growth rather than keeping the lights on.





Conclusion: Reducing the Cost and Complexity of Scalability

Scalability is all about giving your business the ability to deliver a seamless user experience, so customers can always access your services, and employees can always get their jobs done. The cost of downtime—whether measured in loss of revenue, loss of customer faith, or a halt in production—is too great to ignore. Every business needs a scalability strategy.

Knowing this, it's then a matter of working out which strategy delivers economic benefit to your business. This involves weighing the risks, biases and blind spots, and the total cost of ownership of different solutions. Which solution gives you the best result with the least cost and complexity?

VMware Cloud on AWS offers the most efficient path to hybrid cloud, providing a robust and efficient foundation for scalability without the migration costs or risk that other options are saddled with.

Learn more about VMware Cloud on AWS

Take it for a test drive in the Hands-on Lab

Calculate total cost of ownership

