THE PATH TO HYBRID CLOUD

INTELLIGENT BURSTING TO AWS AND AZURE

TURBONOMIC WHITE PAPER

turbonomic

EXECUTIVE SUMMARY

The hybrid cloud has been heralded as a promising IT operational model enabling enterprises to maintain security and control over the infrastructure on which their applications run. At the same time, it promises to maximize ROI from their local data center and leverage public cloud infrastructure for an occasional demand spike.

Public clouds are rapidly assuming a prominent position in the IT landscape, with more and more enterprises prioritizing their adoption and multiple vendors now offering solutions as well as improved on-ramps for workloads to ease the transition to a hybrid cloud model.

With these advances and the ability to choose between a local data center and multiple public cloud offerings, one fundamental question must still be answered: What, when and where to run workloads to assure performance while maximizing efficiency?

In this whitepaper, we explore some of the players in Infrastructure-as-a-Service (IaaS) and hybrid cloud, the challenges surrounding effective implementation, and how to identify and time the bursting of appropriate workloads.

One fundamental question must still be answered: What, when and where to run workloads to assure performance while maximizing efficiency?

A BRIEF HISTORY OF HYBRID CLOUD

The seeds of hybrid cloud were planted on March 8, 1999 in a cluttered San Francisco apartment shared by three developers and a bedroom closet-turned-server room. Salesforce.com sojourned into unchartered territories when it opened its humble doors. The then-longshot startup was looking to disrupt the market of Sales Force Automation and Customer Relationship Management with a new business model. In doing so it successfully delivered one of the first enterprise Software-as-a-Service (SaaS) business applications, accessible through an easy-to-use Web-based interface. It was one of the firms to prevail when the dot.com bubble burst.

What does Salesforce.com have to do with the hybrid cloud? In a word, everything.

Salesforce.com represented a complete paradigm shift from on-premises, private, hosting and delivery to off-premises, or public, hosting and delivery. When Salesforce.com capably proved that it could be done, the SaaS revolution ensued, driven forward by the likes of NetSuite, Constant Contact, Taleo, Google Docs and many more.

Gartner defines hybrid cloud computing as policy-based and coordinated service provisioning, use, and management across a mixture of internal and external cloud services.[1] Because of the way a hybrid cloud service crosses provider and isolation boundaries, it cannot be put into one category of service – hence its "hybrid" designation. Salesforce.com was the catalyst which created the permeability for business applications between the public and private domains.

Gartner defines a hybrid cloud service as a cloud computing service that is composed of some combination of private, public and community cloud services, from different service providers.

[1] http://www.gartner.com/it-glossary/hybrid-cloud-computing

It was not long before virtualization, and specifically, VMware®, forever changed the datacenter as we knew it. On November 10, 2003, the vMotion was born – and with it, live workload portability.

The year prior in 2002, Amazon Web Services was founded, which in 2006 launched Elastic Compute Cloud (EC2) and Simple Storage Service (S3), enabling companies to rent virtual computers for developing and deploying their own enterprise applications – all in the public cloud.

The SaaS movement, set in motion by Salesforce.com, paved the way for public Infrastructure-as-a-Service (IaaS). IaaS is a standardized, highly automated offering, where compute resources, complemented by storage and networking capabilities are owned and hosted by a service provider and offered to customers on-demand. Customers are able to self-provision this infrastructure using a web browser that serves as a management console and grants API access to the infrastructure.[2]

Ironically, virtualization enabled the economies of scale required for profitable IaaS, which conveniently also allows the portability of workloads between the private and public domains.

The public cloud, which we submit was born in a San Francisco apartment in 1999, has undergone widespread commercialization in the fifteen years since. In 2016, public cloud spending netted $96.5 billion, up from $47.4B in 2013, and is expected to grow at 20.4% compound annual growth rate (CAGR) through 2020 – or to $195 billion.[3]

For enterprises, the sheer number of options – across SaaS, IaaS and PaaS (Platform-as-a-Service) – is stymieing. Once effectively implemented, hybrid cloud can accelerate speed to deployment as new workloads get introduced. However, it is not as simple as one might desire. Performance, security and availability remain imperfect across all options.

[2] http://www.gartner.com/it-glossary/infrastructure-as-a-service-iaas

[3] http://www.idc.com/getdoc.jsp?containerId=prUS41669516

PUBLIC CLOUD PLAYERS

According to Gartner, "this phase of the [IaaS] market has already been won." In 2015, the market for IaaS consolidated dramatically around two vendors, Amazon Web Services and Microsoft Azure*, with AWS taking the lead and majority of market share.[4]

That being said, there are still thousands of service providers offering IaaS, and Gartner argues that the next phase of the IaaS market has not yet even emerged. Furthermore, the continued commoditization of x86 architecture, compounded by virtualization economies, has made it feasible for major vendors to also enter the IaaS space; in addition to the aforementioned players, Google, CenturyLink, IBM, VMware, Virtustream and Rackspace, among others, have devised their own IaaS offerings often aimed towards addressing specific barriers to adopting the public cloud.

The continued commoditization of x86 architecture, compounded by virtualization economies, has made it feasible for major vendors to also enter the IaaS space.

Furthermore, many niche vendors such as Vision Solutions, Hotlink, Stackdriver, OneCloud and CloudVertical are addressing capability gaps aimed at helping accelerate enterprise adoption.

*Microsoft® Azure refers to Azure IaaS, as opposed to PaaS, herein.

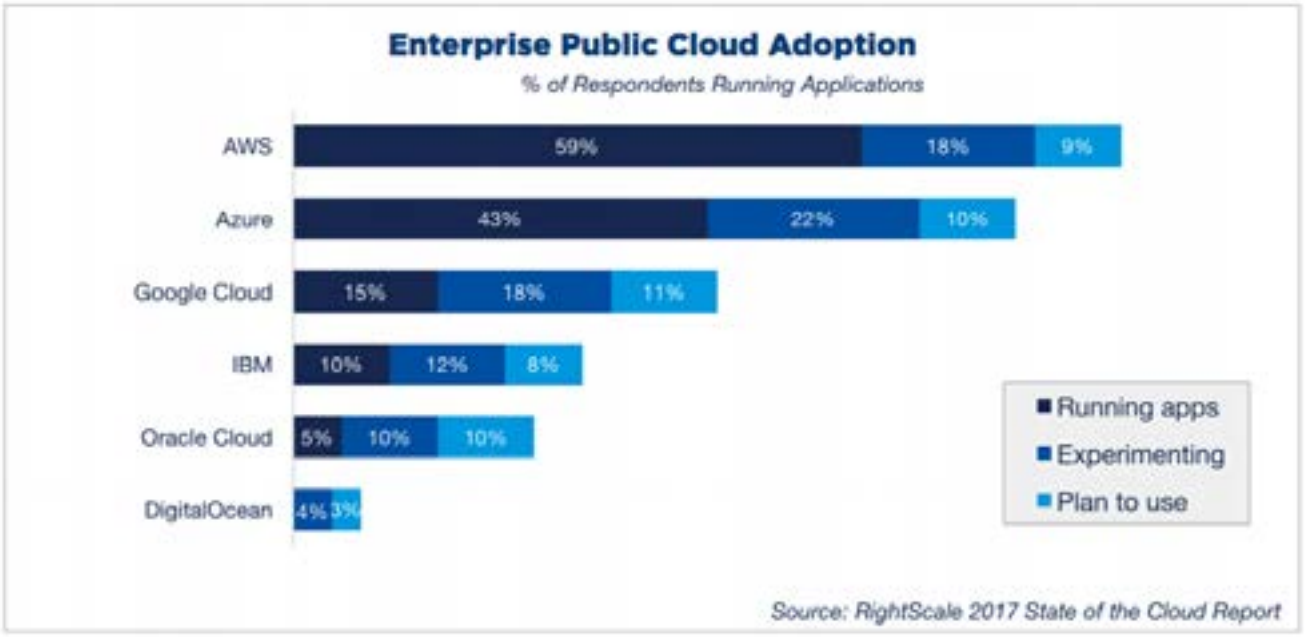
[4] https://www.gartner.com/doc/reprints?id=1-2G2O5FC&ct=150519

PUBLIC CLOUD ADOPTION

The RightScale 2017 State of the Cloud Report surveyed 1,002 technical professionals across a broad base of organizations and industries, yielding some interesting findings†.[5]

- 67% of organizations are leveraging a hybrid cloud deployment, compared with 22% using only public cloud and 5% using only private
- 50% of IT teams cite leveraging a hybrid cloud as their predominant strategy moving forward
- 95% of organizations are experimenting with IaaS
- Azure adoption among enterprises increased significantly, from 26% to 43%, while AWS maintained the lead, increasing from 56% to 59%

The chart below shows public cloud penetration among the top six vendors:



[5] https://www.rightscale.com/lp/state-of-the-cloud

† Margin of Error = ±3.0%

In addition to the above findings, the report also found that among organizations focused on cloud strategy, the three largest challenges were cost management, compliance, and lack of resources/ expertise.

Throughout the past several years, cloud adoption has gained significant momentum on both an organizational basis and a workload basis. In 2014, former VMware® EVP & GM of Hybrid Cloud, Bill Fathers, averred that just 6% of all virtualized workloads reside in a public cloud. Fast forward to today, Cisco contends that by 2018, there will be more workloads deployed in the public cloud than there are in private datacenters.[6]

Clearly, with this rapid acceleration comes significant challenges in paving the path to mainstream hybrid implementation and success.

CHALLENGES WITH ADOPTION

The significant benefits promised by hybrid cloud deployment are not without multiple challenges. There are technical, organizational and logistical obstacles that any organization must consider as they shift to a hybrid cloud model. Many of the public cloud vendors mentioned above as well as niche providers have been aiming to solve these challenges.

There are technical, organizational and logistical challenges that any organization must consider as they shift to a hybrid cloud model.

Security & Compliance

This has and continues to be one of the most salient concerns with the adoption of public cloud, and the same challenges are also present in a hybrid cloud scenario. As organizations consider which applications can run where, they need to consider compliance, identity management and data protection associated with those application workloads. Privacy regulation associated with data sovereignty may limit certain workloads from crossing geographical boundaries. In addition, regulatory requirements such as the Health Information Portability and Accountability Act (HIPAA), Payment Card Industry (PCI) or Sarbanes-Oxley Act (SOX) require that the infrastructure where data resides for a specific application has been deemed compliant.

[6] http://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf

On the identity management and credential side, Operations teams need to ensure user permissions and unique credentials propagate from a private to a public cloud environment. Lastly, assuring the public cloud provider has the basic data protection and cryptographic mechanisms in place and is diligent about updates and patches with minimal disruptions is paramount. In some cases, there is an expectation by the provider that its customers will handle securing and patching the servers across their private and public cloud environments.

To address these concerns, both AWS and Azure maintain that their infrastructure has been certified and complies with multiple standards including PCI, HIPAA and many more.

The impact of latency between the public cloud location(s) and the private infrastructure may not be something the end-user or the application tiers are willing to tolerate.

Network Configuration and Latency

A hybrid cloud requires a thoughtful network design and considerations for multi-tier applications. For starters, secure Layer 2 network connectivity to support VM migration needs to exist to support VM mobility between on-premises infrastructure and cloud providers. But that is just the starting point. The impact of latency between the public cloud location(s) and the private infrastructure may not be something the end-user or the application tiers are willing to tolerate. In particular, chatty application workloads may struggle to maintain a quality of service over a wide area network.

For example, consider a simple multi-tier analytics application in which the business logic in the application layer has been separated into two virtualized components each running on a Tomcat Application Server: one for defining subsets of the data and one for defining the queries on individual data sets which sit on a virtualized MySQL database.

If one component is burst to a public cloud while one remains local, or even if both are burst but end up on different hosts separated by a top of rack switch, performance will suffer.

And with the recent push to break up complex calculations across massive data sets into more and more components, the above scenario would include 10s or 100s of virtualized components that need to communicate to render a query, and their chattiness must be considered before bursting.

As organizations consider easily moving workloads between environments they need to consider the metadata and configuration associated with applications and virtual machines.

The network bandwidth must also be considered as a key cost driver or impediment to migrating workloads. Appropriate bandwidth is required for transferring large data sets and should be taken into account when deciding what is burst-worthy. In addition, most public cloud providers do not charge an organization for uploading data (unless they are leveraging a direct connection) but will charge for downloading the data back which is obviously required in a burst scenario.

In terms of configuration, the IP blocks that have been assigned for the network topology may need to be reconsidered in a hybrid scenario. In addition, the traffic network policies connecting the various tiers of a multi-tier application and the control mechanism for routing traffic between the tiers will also have to be evaluated. Lastly, network security policies (e.g. VPN setup), firewalls and encryption of the communication flow need to be extended from the local data center to the cloud.

Compatibility and Portability

There is a good chance that there will be material differences between the infrastructure and software stacks in an organization's private cloud and a public cloud provider. A dependency on a particular hypervisor or change management process tool will need to be evaluated.

As organizations consider easily moving workloads between environments they need to consider the metadata and configuration associated with applications and virtual machines. If the hybrid cloud is based on identical platforms on both ends, this challenge may be easier to overcome, but if it's not it may significantly hamper the ability to seamlessly burst.

Web-scale public cloud environments are inherently built differently than most any private cloud. A private cloud or application management approach that has a dependency on a particular hypervisor may face challenges when trying to burst to a public cloud that uses a different hypervisor or doesn't expose one at all. Most organizations will inevitably face these mismatch challenges and should be prepared to re-evaluate which applications can be burst given compatibility issues.

To address this challenge, offerings like VMware's vCloud Air provide a common platform for migrating workloads from a private cloud built on VMware vSphere & vCloud Suite to its IaaS offering. In addition, niche vendors such as Vision Solution, Hotlink and OneCloud offer capabilities to automate data replication and workload mobility between a local data center and a public cloud for production workloads or specific use cases such as disaster recovery.

Performance & Availability

After security, this may be the second most common challenge to adopting a hybrid cloud or considering running a mission-critical application on a public cloud. While organizations are now more trusting of public cloud providers when it comes to availability, even the behemoths experience outages from time to time, such as the February 2017 AWS S3 outage that affected a multitude of well-known websites across the internet.[7]

While this incident elicited a prompt response from Amazon, it reinforces the reality that it ultimately remains the responsibility of individual organizations to architect for and manage failure scenarios.

Considerations need also be made on how the public cloud platform in a hybrid scenario throttles inbound queries. Are the same patterns and tools to move data or process business events still relevant regardless of where the application resides? In addition, whether or not the application is designed to gracefully handle downtime of components that reside in different parts of the hybrid cloud is a major consideration.

Lastly, some built in auto-scaling mechanisms that are being exposed to the application architect in a public cloud may be designed to drive increased consumption of public cloud resources. But the logic for how those resources are being provisioned and the trade off between the performance penalties vs. the cost impact must be considered.

To address some of these challenges IaaS providers are exposing more and more capabilities to the enterprise. In addition, platforms such as Verizon's Intelligent Cloud Control enable organizations to determine which types of workloads should be run on which public cloud services to achieve the best performance.

Ultimately, it remains the responsibility of the organization to architect for and manage failure scenarios.

[7] https://aws.amazon.com/message/41926/

Cost Considerations

The economic considerations of private vs. public have been hotly debated. Given the above considerations, an apples-to-apples comparison is not a simple task. What is obvious is that despite continued price reduction by public cloud providers, bursting a workload is not free.

Most pragmatic and savvy CIOs have aligned to the realization that for the foreseeable future, the market will provide them with plenty of private and public cloud alternatives. In this market, maximizing the value from their private cloud infrastructure is paramount. And for many of those organizations, cost reduction is seen as the key business driver to pursue a hybrid cloud strategy. After all, why spend capital or operational dollars to set up infrastructure for an unpredictable period of time?

The key consideration for organizations deploying hybrid clouds is when to burst back as demand on a workload subsides or when demand on those workloads which have stayed on-premises has decreased – freeing-up room for the workloads to return.

To address this challenge, many cloud providers offer transparent pricing and a variety of models by which they charge. In addition, niche vendors such as CloudVertical aggregate data from multiple cloud providers to give their customers better visibility into how much they are spending and how to optimize this expense.

Cost reduction is a key business driver to pursuing a hybrid cloud strategy. Why spend capital or operational dollars on infrastructure for an unpredictable period of time?

WHAT, WHEN, WHERE

With all of these advances and multiple approaches to address the challenges, it appears that broad hybrid cloud adoption is imminent. We agree. However, a fundamental challenge remains: which workloads should be burst to the public cloud, when should we burst them, when should we bring them back and where should we place them? This is not a simple decision.

The decision requires taking the load on your current private data center into account, the seasonality of the load over a granular period of time and the trade off between provisioning additional capacity and the cost of moving. The lifecycle of the workload also requires consideration. Is this workload seasonal in nature, is it for development or testing and will be retired once the application is in production or is its resource requirement unpredictable?

Capacity planning also changes when dealing with elastic and potentially unlimited resource pools. A common view of virtual applications and resources across the data center and cloud service provider will enhance planning. Operations teams can tie their hybrid VM management tools into cloud orchestration platforms to increase automation of application deployment and network provisioning and create a unified operations and management view.

However, treating a set of public cloud resources as an infinite pool may lead organizations down the wrong path; nothing is truly infinite. Each environment has unique, natural constraints that have to be taken into account when assessing planned usage.

Furthermore, the ability to predict future demand on a local data center and the future potential capacity that will need to burst is a challenging task.

Operations teams can tie their hybrid VM management tools into cloud orchestration platforms to increase automation of application deployment and network provisioning and create a unified operations and management view.

The nature of capacity planning also changes when dealing with elastic and potentially unlimited resource pools. A common view of virtual applications and resources across the data center and cloud service provider will enhance planning. Operations teams can tie their hybrid VM management tools into cloud orchestration platforms to increase automation of application deployment and network provisioning and create a unified operations and management view.

However, treating a set of public cloud resources as an infinite pool may lead organizations down the wrong path; nothing is truly infinite. Each environment has unique, natural constraints that have to be taken into account when assessing planned usage. Furthermore, the ability to predict future demand on a local data center and the future potential capacity that will need to burst is a challenging task.

HOW TURBONOMIC CAN HELP

Turbonomic's autonomic performance platform for cloud and virtualized environments assures application workload performance while maximizing the efficiency of the underlying infrastructure. It does so by abstracting cloud environments as markets of buyers and sellers: every data center entity—applications, VMs, hosts, storage—is a buyer and a seller in a supply chain, every resource—CPU, memory, network, storage, IOPS—has a price. As a resource's utilization increases, so does price. This abstraction allows applications to self-manage, independently making decisions to get the best overall price for all the resources they need to perform.

A fundamental challenge remains: Which workloads should be burst to the public cloud, when should we burst them, when should we bring them back and where should we place them?

In a hybrid cloud environment, the public cloud is another market with resources. Workloads burst to the cloud when the price for public cloud resources is cheaper than on-premises resources—in other words, when resource utilization in the private data center has gotten so high that performance degradation is a risk. Likewise, when utilization (price) decreases on-premises, workloads will return.

Turbonomic's economic abstraction allows it to make performance-cost driven placement, sizing and auto-scaling decisions across clouds.

Lastly, Turbonomic is aware of the network flow (communication matrix) between workloads and factors those parameters into bursting decisions so that tiers of an application that frequently communicate will be kept together. In addition, any policies or constraints which have been placed on a workload for compliance (e.g. PCI or HIPAA compliant infrastructure or business unit that needs to keep data in a certain geography), or compatibility reasons (e.g. unique hypervisor or OS requirements) are taken into account before recommending which workloads to burst.

With Turbonomic you are enabled to extend your private cloud management approach to:

- Burst workload into the cloud when demand increases and cannot be met with local resources
- Maintain control on workload performance and resource utilization to decide when to move workloads back on-premises
- Allow applications to auto-scale and clone into the cloud
- Load balance across private and public clouds

With Turbonomic, public and private resources are continuously analyzed taking into account multiple dimensions of continuously fluctuating trade offs. Trade offs between QoS vs. budget and costs, between workload demand and infrastructure supply, between application performance and infrastructure utilization, between compute, storage and network latency, and so on and so forth. No human, regardless of how many reports, trends or alerts he or she examines can solve this problem. It must be solved by software, in an ongoing fashion, at runtime speed.

As workloads become more and more mobile across hybrid clouds the complex question of “What workload to run where and when?” becomes critical and must be continuously answered. Turbonomic’s autonomic performance platform is the only solution that continuously answers the question and enables transparent mobility of workloads across hybrid clouds.

ABOUT TURBONOMIC

Turbonomic delivers an autonomic platform where virtual and cloud environments self-manage in real-time to assure application performance. Turbonomic's patented decision engine dynamically analyzes application demand and allocates shared resources to maintain a continuous state of application health.

Launched in 2010, Turbonomic is one of the fastest growing technology companies in the virtualization and cloud space. Turbonomic's autonomic platform is trusted by thousands of enterprises to accelerate their adoption of virtual, cloud, and container deployments for all mission critical applications.