hazelcast®

Accelerate Your Cloud Application Performance

High throughput, low latency, and scale for your cloud deployments

White Paper

Accelerate Your Cloud Application Performance

Introduction

When Amazon Web Services launched in 2006, it was little more than hosted servers where users could install, build, and test software and only pay for the processing power, storage, and bandwidth that they actually used. It was an incredibly novel concept, and has expanded dramatically over the years to what we currently see as leading cloud infrastructure.

Google Cloud Platform launched in 2008 as Google App Engine and provided additional tooling to make it easier for developers to build and host applications on Google infrastructure. By the time Microsoft launched Azure in 2010, it was clear that the cloud was here to stay. At that time, analysts estimated Amazon Web Services was already netting half a billion dollars or more and projecting strong future growth.[1]

Since their beginnings, the public clouds have touted their strengths against on-premises deployment, including the shift in IT spend from capital expenses to operational expenses on the balance sheet. While each provider offered unique technical differentiators, they all touted these generic benefits associated with the public cloud.

Enterprises that spent tens of millions of dollars on IT infrastructure every year perked up and took note. CIOs and CTOs began to strategize how they could operate deployment models that leveraged the cloud. Even in the early days, many IT professionals were touting that, "everyone is going to the cloud."

The Evolving Value of the Cloud

One of the key selling points in the early days of the cloud was cost. Instead of buying huge, expensive servers for your data center that also incurred costs around space, power, cooling, bandwidth, hardware repair, and administrative staff, you outsourced your computing infrastructure. Since you only paid for what you used, you wouldn't have to over-provision on hardware to accommodate unexpected load spikes.

Soon, businesses realized that the cost was not necessarily the key advantage, as the cloud was more about ease of deployment. This included fast provisioning, which enabled elasticity to scale up or down as needed. You no longer were required to spend months building out racks of servers before you could install new software for testing or a proof-of-concept. Instead, you could simply swipe your credit card, and all the resources you could ever need would be waiting at your disposal within minutes. If you needed even more computing power, that again was just minutes away.

However, moving forward with this infrastructure was not a simple decision. Data and IT leaders weren't always prepared for the big shift that required them to put their precious data into someone else's hardware. There were concerns about security, reliability, and even vendor lock-in. While cloud vendors continue to create innovations that try to allay these concerns, a variety of cloud deployment models emerged.

1 https://www.crn.com/news/applications-os/226500204/amazon-cloud-revenue-could-exceed-500-million-in-2010-report.htm

An Overview of Cloud Deployment Models

There are a number of cloud deployment strategies that businesses can implement. Let's look at the similarities and differences in more detail.

Hybrid Cloud

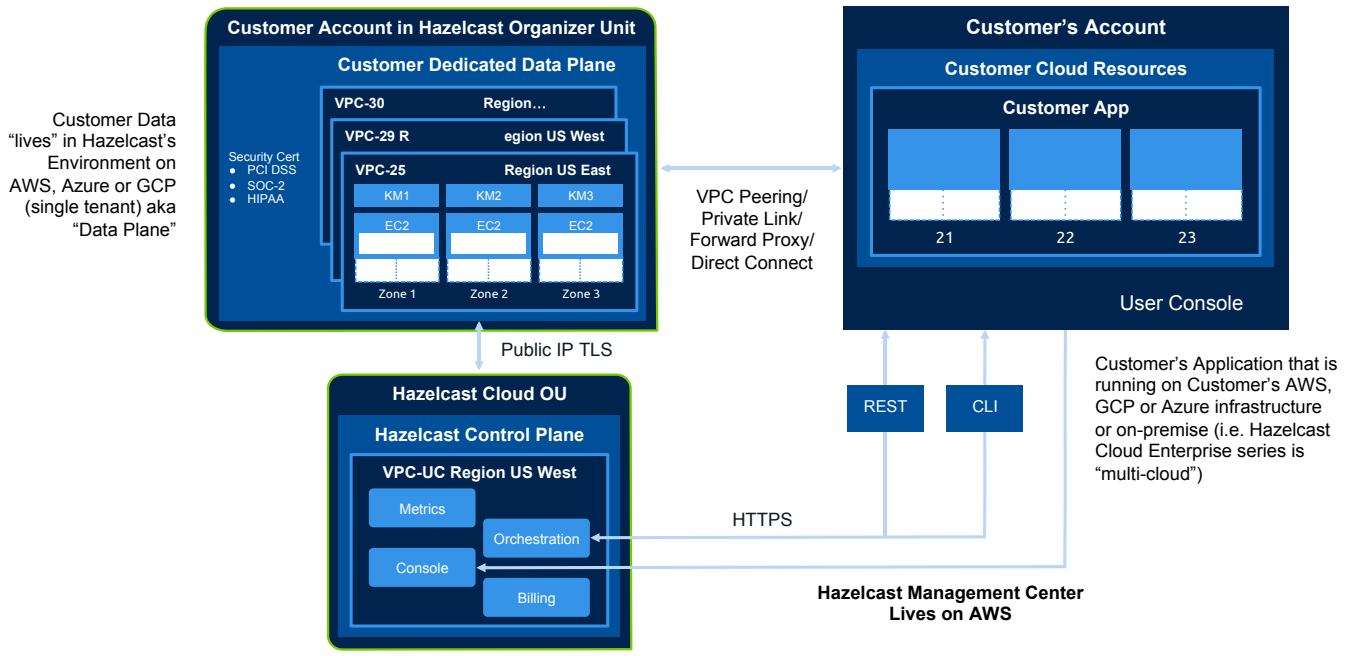
A hybrid cloud is an environment that incorporates both private infrastructure and cloud resources. It could be the case that on-premises applications send data to other applications in the cloud, or vice-versa. Or it may be the case that some on-premises applications "burst" to the cloud for additional computing and storage resources during peak loads. Or, the overall deployment could take advantage of the decoupling of compute and storage, so that the processing power resides on-premises, while the data is stored economically in the cloud (especially in cloud object storage). This hybrid deployment model requires a secure, reliable network between the enterprise and the cloud.

Multi-Cloud

A multi-cloud deployment can be used for similar use cases as a hybrid-cloud deployment. However, rather than leveraging private on-premises resources, applications run in and across different public clouds. For example, an application built on GCP may feed results to another application built on AWS or Azure. Enterprises see the multi-cloud as reducing the risk of relying too heavily on any one provider. A multi-cloud approach also enables organizations to leverage features that may exist with one provider but not others.

Shared Services

In a shared services environment, several enterprises share resources in a single cluster of servers, which can be ideal for enterprises looking to minimize costs. In some cases, the cloud vendor will inquire upfront about workloads to understand the likely peaks and valleys in resource utilization, then bundle customers with complementary usage patterns into the same environment. This way, their customers get the benefit of additional resources during heavy workloads by "borrowing" from other customers in their environment who are running lighter workloads at that time.



Dedicated Services

Some enterprises aren't keen to have their data in such close proximity to others, though, or they want to know that they can spin up additional resources solely for themselves as needed. For these enterprises, dedicated cloud services are the desirable option (and an easy transition on the backend from a shared service) to enable more predictability of the overall load on the system since there is no sharing with other customers. A dedicated service enables you to tailor your environment to your specific needs. This is especially useful when you have outgrown the resource allocation available within shared services.

Focus on the Business, Not Infrastructure

A key value proposition of any cloud deployment is that it frees you from managing the underlying infrastructure. This in turn frees up resources to focus on the bigger business goals. Unless your company is an IT services provider, your market differentiation is probably not around infrastructure technology management.

By leveraging a managed service in the cloud, you remove your team from the operational overhead of managing the underlying infrastructure. Instead, you can focus on the ways your business can drive more value from your data. For example, you may want to identify areas where you could re-architect and modernize existing applications or build new ones to take advantage of real-time microservices that create new revenue opportunities.

With a focus on business logic, you can digitally transform your business, ensure that it is built for speed, and outpace your competitors by delivering new, real-time experiences to your customers.

The public cloud has been good about taking advantage of technological innovations to further remove infrastructure concerns from customers, and that trend looks to continue. As solid-state drives (SSD), graphics processing units (GPU), and high speed interconnects emerged as required hardware technologies, cloud vendors have included them as options. The same will likely happen with innovations like Intel® Optane™ DC Persistent Memory, which can be used in the cloud in two different modes. The first is volatile storage mode as an alternative to RAM, with comparable speeds but at half the cost. This gives cloud customers a more cost-effective option at taking advantage of in-memory speeds to increase overall performance of cloud applications. The second mode for Optane is as a faster non-volatile storage mechanism, which can increase access speed to data that needs to be persisted.

Simplified Scalability

Perhaps one of the most important value propositions that the cloud provides for mission-critical systems is the ability to scale and ramp according to workload to maintain consistent responsiveness.

As an example, let's say you are deploying a large-scale analytics environment in the cloud. Your deployment may have relatively stable performance requirements, but the amount of processing horsepower, or memory, that you need may fluctuate significantly depending on time of day, season, new marketing promotions, or other factors.

Many cloud-native technologies have decoupled storage from compute processing. Therefore, in the example above, you could maintain your storage capacity but instantly increase or decrease the amount of CPU you need to meet workload requirements.

This ability to effortlessly scale is important not only for spiking workloads. It also adds value as other teams in your organization see the benefits of your deployment. This leads to greater adoption, and as usage grows, simplified scalability in the cloud will greatly reduce your operational burden.

What about Performance?

Any organization that has migrated an on-premises deployment to the cloud knows that trying to match performance characteristics as part of the migration planning is not a simple task. Even if you know what underlying hardware is used by the cloud vendor, the virtualized nature of the cloud instances adds a performance hit that is not always easily quantifiable. Certainly you can simply use more powerful instances to get to the performance levels you want, and that flexibility is another advantage of deploying in the cloud.

A better way to improve performance for your cloud environment is to leverage in-memory technologies. With a scalable, in-memory platform, you can grow your RAM usage independently from your compute power, depending on your data and processing loads. These technologies have worked well over the years for on-premises deployments that needed to overcome the latency incurred by disk-based systems, and can help in the cloud as well. You can use in-memory cloud platforms as databases, operational data stores, and caches to accelerate data accesses by your applications.

One historical concern of the use of RAM is the cost. Since RAM is significantly more costly than non-volatile storage media, businesses have made compromises on performance to reduce hardware expenditure. But with new innovations like the aforementioned Intel Optane technology, high speed in-memory processing is becoming an economically feasible option for more and more businesses today.

Introducing Hazelcast Cloud Enterprise

Recently, technologies have emerged to enable you to take advantage of the benefits of the cloud described in the previous section for high performance, low latency, in-memory computing.

Hazelcast Cloud Enterprise is the cloud-native version of Hazelcast IMDG, an in-memory computing technology. Hazelcast Cloud Enterprise enables enterprises to leverage three types of deployment models: public cloud, hybrid cloud, and multi-cloud. Highlights include:

- **Elastic scalability.** Scale up or down based on the demands of your workload so you never pay for more than what you need.
- **Right-size.** Hazelcast solution architects will work with you to right-size your environment based on workload demands.
- **Ease of management.** Hazelcast technical experts manage the cloud environment for you. No heavy lifting or complicated configurations to manage. You get high performance, low latency, and no headaches.
- **Portability.** Hazelcast Cloud Enterprise is identical across AWS, GCP, and Azure, so you're never locked into a public cloud vendor.
- **Engage data where it is created.** Most modern data sources such as mobile applications and websites are likely hosted outside your corporate firewall.

Enterprise-Grade Capabilities

Hazelcast Cloud Enterprise makes it simple to leverage the cloud for in-memory computing jobs without sacrificing any of the performance you would expect from mission-critical on-premises applications.

Hazelcast Cloud Enterprise leverages large allocations of memory for maximum performance. In the cloud, you can easily scale Hazelcast up or down to meet peak demands without over- or under-provisioning. Hazelcast Cloud Enterprise will soon support auto-scaling so it can monitor its own cluster metrics to automatically scale up or down when it crosses specified min-max thresholds.

In addition, Hazelcast Cloud Enterprise enables you to persist data to recover from cluster-wide failures, or to quickly restart after a cluster shutdown for maintenance. Hazelcast IMDG is currently certified to use Intel Optane as a fast, persistent store to enable fast recovery, and as cloud vendors adopt Intel Optane in their clouds, Hazelcast Cloud Enterprise will take advantage of that.

Deploy Anywhere

Hazelcast Cloud Enterprise can be deployed as part of any of the common cloud deployment models. It can be deployed standalone in the public cloud, in a hybrid cloud, or across multiple public clouds.

Hazelcast Cloud Enterprise will be available across all major public clouds - Amazon Web Services (now available), Google Cloud Platform, and Microsoft Azure. It is the exact same application across all three clouds. With such portability, customers can easily avoid vendor lock-in. You can even leverage WAN replication across the public clouds or availability zones.

As part of a hybrid cloud that incorporates both private infrastructure and cloud resources, on-prem applications can send data to Hazelcast Cloud Enterprise, or vice-versa. Or your on-premises Hazelcast IMDG could “burst” to the cloud for additional resources during peak loads. This deployment model requires a secure, reliable network between the enterprise and the cloud.

In a multi-cloud deployment, Hazelcast Cloud Enterprise can run across different public clouds. For example, an application built on GCP can leverage Hazelcast Cloud Enterprise there to feed results to another application built on Hazelcast Cloud Enterprise on AWS or Azure.

Hazelcast Cloud Enterprise is a dedicated service (as opposed to a shared service), fully managed by the Hazelcast DevOps team. When you need additional resources for peak loads, those resources are yours alone to use. You don’t have to worry about competing workloads in a shared environment.

Comprehensive Security with a Dedicated VPC

Security is a critical design criterion for Hazelcast IMDG, which will be included in Cloud Enterprise as well. You can encrypt data both in transit and at rest. This is especially important as you transport data from on-premises applications to the cloud in a hybrid-cloud deployment.

In addition, the entire cluster is secure by default as it runs in its own virtual private cloud (VPC on AWS and GCP, and Virtual Network on Azure). With VPC Peering or PrivateLink, you can securely connect between your VPC and the Hazelcast VPC. Hazelcast Cloud Enterprise will be packaged with the Hazelcast Enterprise security suite to protect your data. Hazelcast is in the process of certifying Hazelcast Cloud Enterprise for the most common security standards including SOC 1, SOC 2, HIPAA, and PCI-DSS.

Availability and Stability

Hazelcast Cloud Enterprise is based on the proven, commercial version of Hazelcast in-memory technologies. It offers high availability through several approaches:

- Within a single cloud provider and region, you can deploy a cluster across multiple availability zones, using partition groups to ensure that each partition's backup is in a different availability zone than its primary. This is identical to how organizations use partition groups in on-premises deployments of Hazelcast IMDG.
- With a single cloud provider and multiple regions, you can use WAN replication to keep the clusters in different regions in sync with either an active-active or an active-passive design.
- Also, you can also use WAN replication to synchronize clusters across different cloud providers. This multi-cloud topology can protect you from widespread outages affecting a provider's entire infrastructure, or enable you to deploy in a multi-cloud fashion to meet corporate requirements to avoid lock-in with a single cloud provider.
- Finally, you can use WAN replication between a cloud and an on-premises cluster.

Clients

Hazelcast provides clients for popular languages, including Java, .NET, Python, Node.js, and Go, that are managed and certified by Hazelcast for use with Hazelcast Cloud Enterprise. You can download a pre-configured Hazelcast client directly from the Hazelcast Cloud Console. Hundreds of Hazelcast customers use Hazelcast software on-premises, and the same application code can be used in the cloud. The common capabilities and clients make it easy for them to deploy Hazelcast Cloud Enterprise.

Conclusion

Hopefully this short overview has made it clear that the cloud is a capable channel for high performance, low latency applications, made possible by the availability of in-memory cloud systems. You can expect at minimum the same level of performance as you have seen on-premises from precisely architected cloud-native software such as Hazelcast IMDG. With Hazelcast Cloud Enterprise, you get a platform that is a fully managed service, cloud-native, scalable, cost-efficient, cloud neutral, consistent, and portable, with a rich feature set features and comprehensive security.

Hazelcast Cloud Enterprise is extremely flexible, mitigating any risk of migrating to the cloud. It is based on the proven, stable Hazelcast IMDG, which is used by hundreds of companies in production with the most demanding requirements. It enables you to work in a variety of deployment models across any of the public clouds. It scales with your business needs to handle growing data sets and growing processing requirements.

One thing to look forward to is the emerging use of new hardware innovations like the Intel Optane DC Persistent Memory class of technologies in the cloud. This technology can be used in two distinct modes to accelerate applications. In the volatile storage mode, Optane acts like RAM with similar performance at half the cost, letting more businesses use in-memory speeds while still staying within budget constraints. In non-volatile storage mode, Optane acts like a faster SSD to let technologies like Hazelcast IMDG recover fast from cluster-wide outages or from shutdowns for maintenance, thus improving overall uptime which is especially critical for 24x7 payment systems.

If you are investigating high performance, low latency, in-memory computing platforms and whether the cloud makes sense, consider Hazelcast Cloud Dedicated and Intel® Optane™ on any major public cloud provider.

For more information on Hazelcast IMDG and Hazelcast Jet, please visit https://hazelcast.com.

For more information on Intel Optane, visit this page: https://www.intel.com/content/www/us/en/architecture-and-technology/intel-optane-technology.html

Also check out:

- **Hazelcast's blog on Optane:** *In-Memory Computing for All Ft. Intel Optane* https://hazelcast.com/blog/in-memory-computing-and-intel-optane/
- **Our joint whitepaper:** *Advancements in High Speed, In-Memory Systems* https://hazelcast.com/resources/advancements-in-high-speed-in-memory-systems/

2 West 5th Ave., San Mateo CA 94402 USA
Email: sales@hazelcast.com Phone: +1 (650) 521-5453
Visit us at www.hazelcast.com

@hazelcast
https://www.linkedin.com/company/hazelcast
https://www.facebook.com/hazelcast