AMPERE®

8x8 | Case Study

Ampere® Improves Cost-Performance and Enables Seamless Migration for 8x8's Video Services Platform

SNAPSHOT

Organization: 8x8, Inc. is a leading Software as a Service provider of 8x8 XCaaS™ (Experience Communications as a Service™), an integrated contact center, voice communications, video, chat, and SMS solution built on one global cloud communications platform. 8x8 uniquely eliminates the silos between Unified Communications as a Service (UCaaS) and Contact Center as a Service (CCaaS) to power the communications requirements of all employees globally as they work together to deliver differentiated customer experiences.

Challenge: In 2020, during the pandemic, video services from 8x8 accelerated from around 200K users per month to more than 20M active users per month worldwide, resulting in a peak of almost 1.5 PB a day of video data being delivered to users. This is equivalent to downloading 10 billion photos per day from the most used social network Facebook. This level of engagement pushed 8x8 to seek a cloud service provider that could meet their demands of scale.

Important factors to consider in selecting a partner were high performance and low latency in compute and network services as well as significant cost reduction to accelerate the onboarding of millions of monthly active users.

Solution: Flexible shapes of Ampere processors, with predictable low latency performance meets the demanding requirements of uncompromising compute and network transfer for high QoS. The Ampere partnership with Oracle Cloud Infrastructure (OCI) helps 8x8 expand its global footprint for 8x8's resource-intensive applications like video conferencing and streaming content. In addition, running the communication platform on Ampere processors on OCI provides both cost and energy savings for 8x8 services.

Results: Initially, 8x8 moved its video meetings services to Oracle, experiencing more than a 25% increase in performance per node on Oracle Cloud Infrastructure, global reach, and savings of more than 80% in network outbound costs. Later, migrating the video services on Ampere A1 instances resulted in further cost-performance efficiencies. With Ampere, 8x8 was able to sustain 80% CPU utilization vs. the normally practiced 60% with x86 processors, while achieving a linear, predictable performance and latency. These factors, along with low cost-performance and the best QoS, were the key reasons 8x8 moved to Ampere processors. By migrating its video services platform to Ampere A1 instances on OCI, 8x8 was able to achieve all three of its goals: exceptional price-performance, low costs and sustainable energy.

INTRODUCTION

8x8 Inc. is a leader in cloud contact center and communication services for small- to mid-sized enterprise customers globally. The product portfolio includes integrated cloud contact center, business phone, video meetings, team chat, and SMS on a single platform called 8x8 XCaaS (Experience Communications as a Service).

8x8 has over 3 million business users across 186 countries that are utilizing its services and enjoying cloud native contact center and unified communication solutions. 8x8 solutions span various industry segments including transportation, education, retail, manufacturing, healthcare, financial, professional services, and government.

8x8 is an innovation and R&D led company which is evident in the fact that they are delivering one of the most difficult applications over the public internet- real-time conversational voice and video to customers, which requires high performance of compute, network, and storage to maintain a standard of high QoS.

WHAT MOTIVATED 8X8 TO WORK WITH AMPERE COMPUTING

8x8 wanted a solution that was reliable, performant and cost effective in collaboration with a cloud services partner that could support the expanding demand of video/voice services. To find the right partner that would allow for flexibility and 8x8's rapid expansion of the business, 8x8 selected OCI (Oracle Cloud infrastructure), for its global coverage, cost, performance, and security.

8x8 chose Ampere processors on OCI for their cost effectiveness, reliability and consistent linear performance - factors critical for real-time voice and video applications. Additionally, Oracle had recently moved their own managed services to Ampere processors which further encouraged 8x8 to try out A1 Ampere instances on OCI.

While 8x8 initially came in with a conservative approach towards Ampere A1 instances, they discovered that the solution delivered three key requirements – low cost, no degradation in performance and low energy consumption. In addition, the experience of migrating

to Arm Ampere architecture was much simpler than they had anticipated. Mehdi Salour, SVP DevOps at 8x8, expressed delight at the results: *"A1 processors have been performing really well for us. The fact there is no HyperThreading they are very consistent linear performance, which is very important for applications such as real-time voice and video, so the performance has been really good. Cost-performance has also been excellent. We are saving money, and we are doing something good for the environment, so we are hitting all three goals".*

AN EASY PORT TO AMPERE

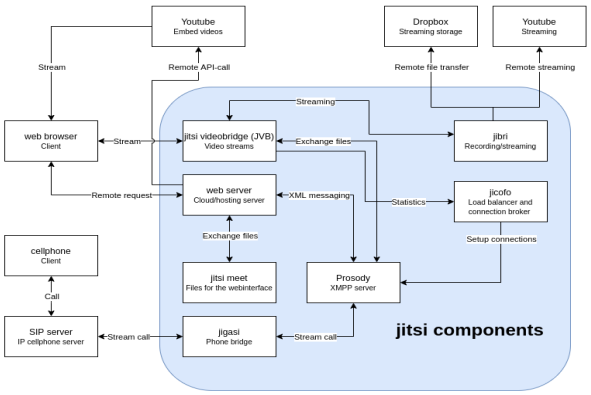
8x8 is the main contributor to the open-source secure video meetings technology, and the standalone and integrated versions of 8x8 Video meetings are powered by Jitsi. The Jitsi.org code has been hardened with over 30 million downloads and is widely used in applications like banking video conferencing, etc globally.

8x8 Video Meetings utilizes the WebRTC standard, which enables attendees to instantly join meetings without any downloads or plugins required.

Due to the global nature of 8x8 services, all tools need to support the Arm 64 architecture, Arm images and builds. After initiating the migration process, the 8x8 team discovered that there was only one module of native code that had to be recompiled, while the rest of the efforts were not a heavy lift. Most of the work focused primarily around making tools and images aware of Arm 64 arch, or finding the right binaries, making right builds, orchestration, and tooling.

For the fastest migration to A1 and to get to production quickly, the video team (with video bridges component) was chosen because of two reasons: first, the video team has always been keen on Arm architecture development and was familiar with HashiCorp distribution of Arm binaries as well as the security tools on Arm 64. Second, the video team is agile, innovative and has a deep understanding of Arm64 arch.

The overall architecture for 8x8's video stack was based on VMs before migration to OCI. After moving to A1 VMs with OCI, 8x8 witnessed reduced noisy neighbor issues and an overall improved cost/performance for their video environment. 8x8 uses OKE in OCI and their proprietary Cloud8 K8s in their data centers as a service to the engineering teams. While 8x8 tends to be cloud and infrastructure-agnostic, they optimize for geo redundancy, overload protection, autoscaling etc. as part of the services provided to their customers.



Reference: *https://jitsi.github.io/handbook/docs/architecture/#testing*

In terms of ease of Arm Ampere migration, Java workloads and the CI infrastructure did not require any changes and 8x8 was able to use the existing selenium grid infrastructure to perform end-to-end load testing for the video stack. This included simulating the real environment of starting chrome hitting the site, joining the video bridges, playing video, etc, as an end-to-end workflow. Since performance testing was an important part to ensure processors can perform well in production, 8x8 ensured there was no roll back, least usage regions were tested first, and the video bridges were loaded with A1 before deploying on a global scale.

PERFORMANCE METRICS

The complete end-to-end workflow was simulated to ensure great QoS- starting from chrome browser, joining the video bridges, playing video games etc. Latency between ingress and egress of packets is a key metric used to measure real performance for workloads in production.

According to Salour –
"Before Ampere A1, max CPU utilization was observed at 60% after which performance starts to degrade, but Ampere A1 processors could be loaded up to 80% while maintaining great performance at low power."

While compute and network are both critical for video services, an important metric to measure performance for 8x8 is latency since any jitter introduced due to latency can cause degradation in audio or video quality. It is important to note that 8x8 is known for its high QoS for enterprises and signature audio quality compensation in less-than-ideal network conditions. 8x8's global reach technology provides the highest level of service with end point correction to their customers.

With all Arm binaries being readily available, 8x8 team found the migration to be a fairly simple process with little effort of selecting the right images for Arm64-Ampere. The complete simulation of the production environment was quick, with no performance degradation observed in the production environment. This allowed 8x8 to roll out their enterprise grade services within three weeks of starting the process. The solution went live in three weeks to full production and achieved more than 1M users on the platform- a big milestone achievement for 8x8.

PARTNERSHIP WITH ORACLE

8x8 currently runs services such as core voice infra services including session border controls, core call flows, media servers, as well as video and many of its over 300 Kubernetes microservices utilizing Oracle Container Engine for Kubernetes (OKE) on OCI globally. 8x8 found the OCI team to be highly collaborative, providing supportive partnership at all levels. The working culture between product teams has been based on constructive discussions and realistic prioritizations.

MEETING THE SUSTAINABILITY GOALS

On 8x8's sustainability goals, Salour mentioned that 8x8 has successfully reduced their GHG output in the UK by one-third from 2020 to 2022. In his words, "our current solution portfolio was born in the cloud and 8x8 is fully committed to helping others reduce hardware footprint for contact center applications, while driving unified communications. Utilizing technologies like Ampere is an easy way to help the environment and improve the C footprint as part of our ESG strategy while saving money without sacrificing performance, so absolutely that is an important factor for us."

CONCLUSION

In this era of soaring cloud communications, 8x8 saw its business grow its user count exponentially over the past few years. In collaboration with Ampere and OCI, 8x8 has received the benefits of global coverage, low power, price-performance, and significant cost reductions across every service essential to 8x8—compute, network transfer and storage. The company was up and running quickly on Ampere A1 instances using OCI, with production workloads running within three weeks of starting the migration process. Lower cost/core for A1 relative to x86 shapes on OCI, great performance and significantly less power consumption provided a unique value proposition, enabling 8x8 to continue serving its millions of users. On future partnerships, 8x8 looks forward to A2 introduction to power its expanding 8x8 business.

About Ampere

Built for sustainable cloud computing, Ampere Computing's Cloud Native Processors feature a single-threaded, multiple core design that's scalable, powerful, and efficient.

Learn more

See our solutions for a variety of demanding workloads:
https://amperecomputing.com/solutions

Visit our Developer Center:
https://amperecomputing.com/developers

Disclaimer

All data and information contained in or disclosed by this document are for informational purposes only and are subject to change.

This document is not to be used, copied, or reproduced in its entirety, or presented to others without the express written permission of Ampere®.