

The Results Are In!

23andMe Selects AWS & Weka for their HPC Research Platform



Executive Summary

23andMe, Inc. (23andMe) is a privately held biotechnology company based in Sunnyvale, California. It is best known for providing genetic testing to retail customers.

23andMe collects billions of data points. The medical breakthroughs this data can unlock are limitless. However, legacy storage systems hinder medical research and innovation as they were not designed for today's modern workflows, that require multiple and often simultaneous access to data. Given the company's current growth trajectory, 23andMe was challenged to find a solution that is not only able to address their challenges today, but also support them as they scale in the future.

Customer Challenge

23andMe had most of their non-research computing infrastructure on-premises in a high-performance computing (HPC) environment. Their Cloud Engineering team, SKY, had previously investigated migrating the HPC cluster, used for genomics research, to the cloud. However, cloud storage cost, home grown tooling, managing data locality and ownership led the company to continue investing in their on-premises cluster.

In July of 2020 the SKY team was tasked to build out a migration plan to the cloud for the HPC cluster. They had to develop an architectural design that would address the filesystem and data requirements as well as a design that made financial sense.

It was clear that moving the HPC cluster to AWS would address current challenges of physically maintaining hardware with a limited staff and would reduce the time needed to add compute and storage resources. However, addressing the economics of HPC storage in the cloud and the flexibility of a file system was a big concern.

The SKY team also needed to ensure they could address the research teams concerns of a unified interface for one time and batch job runs analysis and a shared file system for data.

23andMe had already ported some of their jobs from a POSIX filesystem to AWS S3 so a key requirement for the future filesystem would be support of POSIX, Object and Block storage. As part of the solution, 23andMe would also need the ability to rapidly provision file systems without the need to re-configure the environment each time. Additionally, they needed a filesystem that would allow them to preserve users home directories during this process.

Addressing the latency concerns was going to come down to performance testing, however addressing how the 23andMe researchers submit jobs, store and access data and improve the HPC research cluster that was built on old tooling and processes was going to require a mix of AWS native services and 3rd party solutions.

About 23andMe



23andMe wants to disrupt the healthcare experience by building a personalized health and wellness experience that caters uniquely to the individual by harnessing the power of their DNA. 23andMe pioneered direct access to genetic information as the only company with multiple FDA clearances for genetic health reports.

23andMe built the world's largest crowdsourced platform for genetic research, with 80 percent of our customers electing to participate. This platform allows us to accelerate research at an unprecedented scale, while providing information back to participants. Our Therapeutics team leverages this research platform to identify and develop drug targets rooted in human genetics across a spectrum of disease areas, including oncology, respiratory, and cardiovascular diseases.

Migrate to Modernize

Embracing the cloud is a journey most customers are embarking on today.

It presents a new approach to IT, complete with a new set of challenges, opportunities, tools and skills.

Most companies have very complex IT environments where the opportunities for cloud migration must be balanced with legacy technologies and data that must remain on premises.

Being successful on the cloud requires a sound strategy, the right expertise, and the perfect partners to put it together.

Converge's AWS cloud migration consulting and services delivery team leverages AWS programs and best practices to de-risk and accelerate the adoption of cloud computing, which allows organizations to achieve their innovation and digital transformation goals.

Solution

Converge leveraged the AWS Migration Acceleration Program (MAP) to build the business case for change and provide funding that would help accelerate and de-risk the migration project.

During the assess phase of the MAP process, issues around managing the home directories presented a file system challenge. After some initial review of native and 3rd-party options, Converge introduced Weka, an AWS Technology Partner solution available via the AWS Marketplace to the Sky team.

Storing and analyzing large data sets in the cloud, whether it is next-generation sequencing, imaging, or microscopy requires a modern approach for faster insights and better economics. Weka would provide 23andMe a unified file system, accelerate time to insights by eliminating the performance bottlenecks across the Life Sciences data pipeline, while significantly reducing the cost and complexity of managing data at scale.

After a successful POV (proof of value) 23andMe selected WekaFS data management platform to underpin the AWS solution stack.

Results

23andMe made a strategic decision to move their HPC production environment to AWS. The technology collaboration with Converge, Weka and AWS account teams allowed 23andMe to take full advantage of the flexibility and scalability that AWS has to offer, while driving down infrastructure cost and management overhead. Specifically, 23andMe realized the following benefits:

- Dynamically scale performance and capacity up and down based on real-time application requirements utilizing auto-scaling groups, while controlling AWS infrastructure spend on instances
- Support mixed workloads for a multi-petabyte dataset with zero tuning needs
- Seamless integration of flash and S3 storage into a single namespace that scales infinitely
- Access the same dataset across multiple protocols (POSIX, NFS, SMB and S3) at the same time
- Maintain flexibility, as Weka is completely software-defined enabling them to adopt new technologies as they emerge on-premises, hybrid, and in the cloud

With their HPC platform now in AWS, 23andMe can accelerate research without worrying about resource availability, performance, or data management issues.

"Weka was absolutely a slam dunk," said Arnold De Leon, Operations Manager at 23andMe. "Moving to the cloud comes with a new set of requirements. We (23andMe) needed a solution that was compatible with our existing applications and gave us the performance we required. We chose Weka because the protocols just ran, scale in and scale just worked, and we haven't found the limits of their performance."

About Converge Technology Solutions Corp.

Converge Technology Solutions is a global software-enabled IT & Cloud Solutions Provider focused on delivering industry-leading solutions and services. Our global solution approach delivers advanced analytics, application modernization, cloud, cybersecurity, digital infrastructure, and digital workplace offerings to clients across various industries. Converge is an AWS Advanced Consulting Partner, with the Master Specialist Competency for VMware Cloud on AWS.