



# Scaling Large Genomic Data Transfers

## AT A GLANCE

### Industry

Life Sciences

### Products

Aspera On Demand

### Challenge

Moving large genomics data from clinical providers to GenoSpace's cloud platform.

### Solution

Aspera On Demand moves genomic data to the cloud at high speed, maintaining patient privacy and security with encryption in transit.

### Results

- Automatic retry and resume of incomplete transfers ensures that temporary network problems do not result in failed transfers.
- Cloud-ready solution saves the cost and implementation effort of infrastructure build-outs.
- Aspera will scale to accommodate the large and ever-growing genomics data that GenoSpace receives.

While the cost of genomic sequencing has decreased drastically in recent years, BioIT companies are still faced with the challenge of collecting, managing, analyzing and distributing large sequencing data sets efficiently.

GenoSpace offers a cloud-based software-as-a-services (SaaS) platform for the storage and analysis of large sets of complex genomic data, providing the tools to unite researchers, clinical labs, doctors, and patients for the development and delivery of personalized medicine. Their research gateways intersect clinical data with genomic parameters, allowing clinical providers to gain utility from data that would otherwise be difficult to query and interpret.

GenoSpace selected Aspera On Demand to transfer genomics data from the clinical provider to the GenoSpace database in the cloud at high-speed, benefiting from enterprise-grade security and reliability.

## CHALLENGE

Physicians, clinical labs and researchers conducting clinical studies typically collect a large amount of clinical information from the subjects, such as age, weight, prior treatments, blood type, and lab test results. While this data is fairly easy to store in a

database and interpret, the challenge arises when moving into molecular medicine, where genomic parameters are collected alongside clinical data, and genome sequences can be extremely large and difficult to run regular relational queries on.

GenoSpace was founded to create a new type of secure data portal that could analyze clinical laboratory results and patient responses to therapy, using genomic information with high-level statistics to identify patterns and causes in patient responses to further segment the patient population. Once a clinical provider has conducted a test, collected

samples, and run the samples through their instruments, the resulting data is then sent to the GenoSpace cloud platform for analysis and report generation. Genomics data can vastly exceed a gigabyte per patient with full genome sequencing, and often

times clinical trials involve thousands of patients. GenoSpace set out to find a file transfer solution that could scale to accommodate large and rapidly growing genomic data files while also maintaining security, reliability, and remaining cost effective for smaller data sets, allowing them to focus on their core competency of data analysis and correlation.

*“Aspera has achieved everything that we hoped it would achieve, plus it gives us the ability to scale further in the future.”*

 **Niall O'Connor**  
Director of Engineer, GenoSpace

## BENEFITS

### Maximum speed

Enables transfers of large genome sequences over any network at maximum speed, from clinical provider to GenoSpace cloud, regardless of network conditions or distance.

### Reliable transfers

Reliable, with automatic retry and resume for partial transfers and maximum bandwidth utilization, Aspera provides a fast and worry-free transfer experience.

### Ease of integration

GenoSpace was able to easily integrate Aspera with their existing software running on AWS, plus Aspera easily integrates with third parties, benefitting both GenoSpace and their potential partners.

### Cloud-ready solution

Aspera On Demand moves large genomics data to the cloud and allows GenoSpace to take full advantage of the massive scalability offered by the cloud without requiring expensive infrastructure build-outs.

### Strong security

With built-in SSH authentication, encryption in transit and at rest, and data integrity verification for each transmitted block, Aspera protects patient privacy and assures GenoSpace's data remains safe throughout the transfer process.

## SOLUTION

Finding Aspera's solution was in line with their goals, GenoSpace implemented Aspera On Demand and the Aspera Embedded Client, easily integrating the solution with their software running on Amazon Web Services (AWS).

To deliver content to the GenoSpace database, the clinical provider simply logs into the GenoSpace portal, opens an order request, encrypts the data, and then connects to the Aspera On Demand server to complete the high-speed transfer. Once the data is uploaded to AWS, administrators at GenoSpace pull files from the Aspera server using the Aspera Embedded Client, decrypt the contents, and place the files into their secure database for analysis.

## RESULTS

Selecting a cloud-based solution coupled with Aspera's FASP®-powered high-speed transfers to the cloud proved to be a cost-effective option for GenoSpace's use case. While avoiding heavy investments in infrastructure build-out, GenoSpace still reaps all of the benefits of the endlessly scalable cloud.

For GenoSpace, reliability is a key concern. GenoSpace wanted to avoid the unnecessary troubles that arise when a transfer fails or data pieces unexpectedly go missing mid-transaction, which with other technologies typically required the transfer to be restarted from the beginning. Aspera's FASP is extremely robust, and with automatic resume and retry for partial transfers, Aspera offers the reliability and recoverability GenoSpace can depend on.

Plus, Aspera delivers the security, and scalability that GenoSpace desired to provide the best possible solution to their clients.

"The security of the transport mechanism is an essential requirement. If it can't compete with SSH, I wouldn't even consider using it," said Niall O'Connor, Director of Engineering at GenoSpace. "Aspera's over-the-wire encryption ensures patient privacy while data is in transit."

*"Aspera's over-the-wire encryption ensures patient privacy while data is in transit."*

 **Niall O'Connor**  
Director of Engineer, GenoSpace

In the future, GenoSpace plans to use Aspera Connect Server with the Connect browser plug-in to disseminate information back to researchers, physicians, clinical labs, and patients, who will be able to easily pull and view project files using a standard Web browser.

"Aspera has achieved everything that we hoped it would achieve, plus it gives us the ability to scale further in the future," said O'Connor.



## ABOUT EMBL

At Genospace, our mission is to deliver the platform that makes biomedical data useful and usable by everyone. Genospace has built a comprehensive platform for precision medicine to enable interpretation, analysis, reporting and collaboration on high-dimensional genomic and other biomedical data. With specific applications supporting research, development, pathology and clinical care, many of the most advanced precision medicine organizations are powered by Genospace.

Learn more at [www.genospace.com](http://www.genospace.com)

## About Aspera

Aspera, an IBM Company, is the creator of next-generation transport technologies that move the world's data at maximum speed regardless of file size, transfer distance and network conditions. Based on its patented, Emmy® award-winning FASP® protocol, Aspera software fully utilizes existing infrastructures to deliver the fastest, most predictable file-transfer experience. Aspera's core technology delivers unprecedented control over bandwidth, complete security and uncompromising reliability. Organizations across a variety of industries on six continents rely on Aspera software for the business-critical transport of their digital assets.

Learn more at [www.asperasoft.com](http://www.asperasoft.com) and follow us on Twitter @asperasoft for more information.