



Extending data-intensive DNA screening workflows to the public cloud

Counsyl partners with healthcare providers to provide genetic screening and supporting services for women and their families. Whether it's starting a family or evaluating risk for cancer, Counsyl's genetic screens provide actionable information to guide critical and timely health decisions. Since 2009, Counsyl has worked with tens of thousands of healthcare providers to deliver over one million patient results.

Counsyl operates a high-complexity clinical laboratory that processes and analyzes patient samples to generate reports on risk for genetic conditions. They have requirements on their workflow that include:

- Fast-turnaround – patients / healthcare providers are awaiting answers to make treatment decisions
- Storage of large volumes of data produced through genome sequencing pipeline and movement through their data processing pipeline
- Data must be retained for 7 years and encrypted at rest

Additionally, there are regulations on data retention requirements for human samples that are used in diagnostic testing. They needed a storage platform that could accommodate these requirements and rapid growth.

Cloud-native, On-Premises Solution

Counsyl uses SwiftStack in their data sequencing pipeline. Data is streamed into SwiftStack directly

from the genome sequencers and then processed to transform millions of raw "reads" into a usable single-file representing the genome. The genome is analyzed to identify deleterious variants and then the data can be archived.

The on-premises SwiftStack footprint is accommodating the high-throughput requirements for data ingestion and computation throughout the workflow. With SwiftStack being cloud-native, it can interoperate with any ingestion pipelines that can source from Amazon S3 or Google Cloud. And with the core being based on open standards, it allows Counsyl to use technology that drives some of the largest storage clouds, while leveraging a vibrant community.

Extending Infrastructure to the Public Cloud

Counsyl extends their DNA screening workflow into the public cloud for disaster recovery, archiving, and compute. Leveraging SwiftStack's multi-cloud capabilities, production data is stored in an additional cold archive (such as AWS Glacier) to meet Clinical Laboratory Improvement Amendments (CLIA) requirements for long-term data storage.

With the ability to place data in the public cloud, Counsyl has access to elastic compute resources to support the needs of new product development and data processing techniques. This process data can then be retrieved back on premises to local workstations to use other analysis tools.

USES

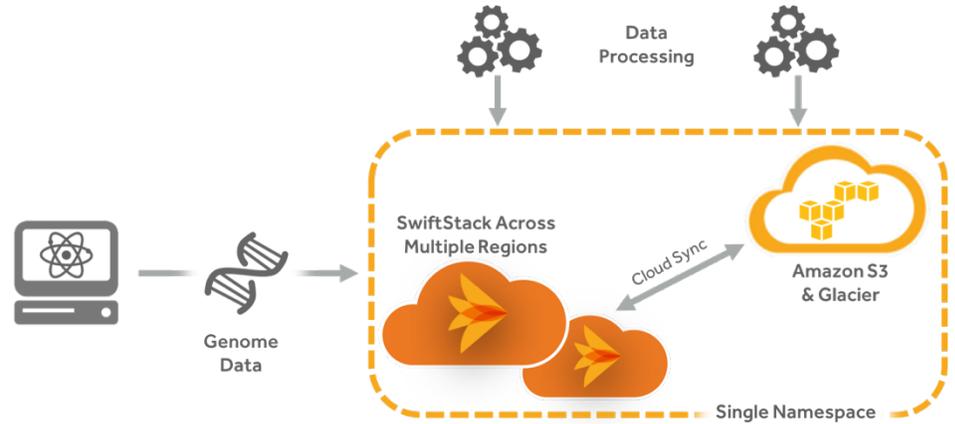
- Petabyte-scale storage namespace across on-premises datacenters and Amazon Web Services (AWS)
- Receives patient data from genome sequencers on-premises
- Extended to AWS for disaster recovery, archiving, and compute
- Manages data based on operator-defined policies

RESULTS

- Common API to access data in a single namespace across on-premises and public cloud resources
- Meets Clinical Laboratory Improvement Amendments (CLIA) requirements for long-term storage
- Enables the use of elastic compute resources in the cloud

"SwiftStack is a "single source of truth" for our data. Our infrastructure team can manage data placement based on policies to optimize workflows without application developers or researchers even knowing changes have occurred."

- Jeffrey Tratner, Technical Lead, Scientific Computing at Counsyl



Optimizing the Workflow

Using Cloud Sync in SwiftStack, Counsyl set a policy so all sequencer and analysis data is automatically copied to an Amazon S3 bucket. For archiving, additional policies were set to automatically move data into Amazon Glacier.

According to Jeffrey Tratner, Technical Lead, Scientific Computing at Counsyl, "Because SwiftStack presents a single namespace across private and public infrastructure, it greatly simplifies access to data. A single API endpoint can be used by application developers and researchers when they access their data. No applications need to change as the infrastructure team creates policies that will move data from on-premises to the public cloud."

Counsyl has an entirely automated workflow to drive all SwiftStack Cloud Sync policies. Moving forward, they will continue to implement more sophisticated data management policies and allow end-users to explicitly apply Cloud Sync policies to a data set.

Cloud Bursting

The SwiftStack approach to multi-cloud data management preserves native data formats in the public cloud. This is an important attribute of the archive solution as data is not stored in a proprietary format like other solutions do. It also means that the data is accessible for cloud bursting, since applications running in the public cloud can directly interact with the data.

"Being able to provide access to the data increases the agility of the research and development team to leverage the elastic compute resources in the public cloud. Counsyl R&D can select which datasets should be made available for data processing from the archive and can either access the data on-premises with their workstations or access the data directly in the public cloud. This simplifies access to data, provides necessary access controls, and R&D does not need to change their applications to access the data." said Tratner.

This access to data increases the speed of development, allowing the team to build and test additional products or increase its quality of service.

To try SwiftStack for free, go to swiftstack.com/test-drive.

For additional assistance or to learn more, always feel free to contact us. We're here to help.

Phone – (415) 625-0293 | **Email** – contact@swiftstack.com | **Chat** – Go to swiftstack.com and click chat