FILE SERVICES FOR DNA SEQUENCING

Morro Data Enables Efficient Storage and Data Transfer

SeqMatic (https://www.seqmatic.com/) is a next generation DNA sequencing service provider. SeqMatic was founded by world leading scientists who were part of the revolutionary R&D team at Illumina (a company whose products enable researchers to explore DNA at an entirely new scale). SeqMatic'sclients include some of the largest genomics and transcriptomics research facilities like Stanford and Harvard University, the USDA and FDA, and the Allen Institute for Brain Science.

©2010, Illumina Inc. All rights reserved.

Compared to the initial Human Genome Project which took over 10 years and cost nearly $3 billion, Next-Gen DNA sequencing (NGS) today is incredibly fast and inexpensive, making large-scale whole-genome sequencing accessible and practical for the average researcher. The Illumina HiSeq X Ten System, for example, can churn out 18,000 genome studies a year. This is enabling researchers to use DNA sequencing across many fields and purposes. As Next Generation Sequencing Machines have high upfront costs, researchers turn to DNA sequencing service providers like SeqMatic. Once SeqMatic receives the samples, they run the DNA sequence and send the results files back to the customer for further analysis.

However, a single DNA sequencing run can churn out up to 1 TB of raw data, which then needs to be stored, protected, and made available to the scientists. All of this has to be done cost effectively, too. This is not just at budget-constrained universities, even large pharmaceutical firms need their IT departments to operate with utmost efficiency. File handling issues include:
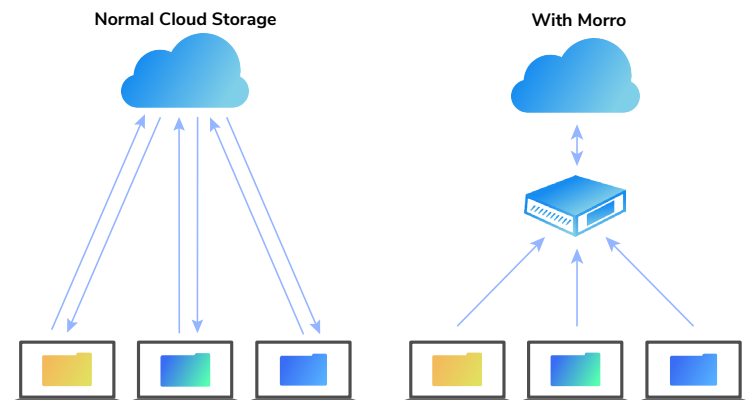
1. **Data Scale:** DNA sequencers churn out a tremendous amount of unstructured data. Relying on traditional file transfer methods is error prone and time consuming.

2. **Data Growth:** New Sequencer Machines operate at much higher resolution and speeds, generating up to 4x – 16x more data then prior generation machines per run, and each run can now be more extensive.

3. **Protection:** This data is tremendously valuable – especially at pharmaceutical companies. The file transfer must be highly reliable and secure, and ideally should be retained once transferred, causing rapid growth in storage requirements.

4. **Server Maintenance:** Storage needs can rapidly grow, requiring more file servers and increasing maintenance and IT overhead.

A simple, secure method to easily transfer and store the results file for ongoing analysis and long term access is required. Morro Data has the solution.

Transferring Large Files to Clients

Every day, SeqMatic has to send hundreds and thousands of large files all at once to their clients. Currently, they have been transferring these files through FTP which is time consuming as there are six steps to sending and downloading a file through FTP:

1. Generate an MD5 checksum for each file that needs to be uploaded.
2. Transfer the files up (user babysits the upload).
3. Verify the checksums of the uploaded files.
4. Contact the target user that the files are ready for download.
5. Target user transfers the files down (user babysits the download).
6. Target user verifies the checksums of the downloaded files.

Normal Cloud Storage   **With Morro**

Morro Allows for Easy Transfer of Large Files

In a matter of seconds, through Morro Connect (on a PC/Mac), SeqMatic's employees can simply drag and drop the files to the CacheDrive using SMB protocol on their LAN and let the CacheDrive do the rest as it syncs with the cloud on its own. Morro Connect sends clients a notification that their files are ready for download and clients are able to easily download the files through either their own CacheDrive or a share link.