

Case Study: Clinical Genomics Gothenburg

Genomics research facility reduces storage footprint to one-seventh of prior level



UNIVERSITY OF
GOTHENBURG



The Background

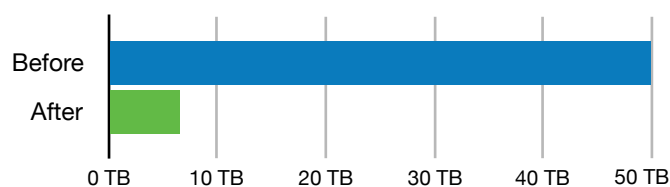
The Clinical Genomics Gothenburg facility at the Sahlgrenska Academy, University of Gothenburg performs clinical research in the areas of acquired disease, hereditary disease and microbiology.

The facility also works with the Sahlgrenska University Hospital to expedite translational research. They use 10X, Illumina and IonTorrent sequencing platforms.

The Challenge

The Clinical Genomics Gothenburg team stores whole exome, whole genome and deep panel clinical data as BAM files on local object storage. These data need to be accessible for ad-hoc remote visualization tasks by researchers and clinicians using Integrative Genomics Viewer (IGV) over an Apache webserver.

The amount of data already in storage (50 TB), combined with the anticipated increase in need associated with the wholesale shift from exome to whole genome sequencing during 2019, was about to place the existing infrastructure under considerable strain. Per Sikora, Head of Facility at Clinical Genomics Gothenburg was tasked with resolving the issue. He set out to find a solution that would create as much capacity as possible in the current infrastructure but not require modifications to existing tools and pipelines.



Clinical Genomics Gothenburg's file sizes before and after compression with PetaSuite

The Solution

Clinical Genomics Gothenburg now compresses its data using PetaGene's PetaSuite compression software and stores it on the same local object storage as previously.

When IGV is used with an Apache server, Apache accesses the data in small chunks which can cause workability challenges for other tools. PetaGene's on-the-fly decompression library had to interact effectively with this Apache access pattern. The University of Gothenburg/Clinical Genomics team was pleased with PetaGene's responsiveness in hardening the PetaSuite product for deployment with Apache. The client found that PetaGene was open, technically adept and effective.

The Results

By using PetaGene's PetaSuite compression software, Clinical Genomics Gothenburg now achieves an average compression ratio of 87%. This means that the university's storage requirements are now less than one-seventh what they would be without compression. The clinicians use IGV to access the data in exactly the same way as they did before and there is no discernible impact on visualization performance. Even during the compression process there was no downtime, and uninterrupted access to the data was maintained throughout – users saw and accessed either an uncompressed file or the equivalent PetaGene virtual file with the original filename in the original location, depending on whether the file had been compressed yet.

The university chose to use the optional BayesCal feature, as their main objective was to compress their data to the maximum extent. However, an additional welcome benefit is an increase in genotyping accuracy across the entire receiver operating characteristic (ROC) curve by using BayesCal.



Per Sikora, Head of Facility commented

“By using PetaSuite compression software for our data we have achieved our primary aim of dramatically increasing our storage capacity. This means that we do not need to spend precious resources on replacing or adding to it. The PetaGene team were responsive to our needs, including managing the demands of using IGV to efficiently access the compressed data via Apache server without decompressing the data first.”