

Introduction

The Italian Institute for Genomic Research (IIGM) is a not-for-profit entity established in 2007 by Compagnia di San Paolo, the University of Turin, and Turin Polytechnic. Their aim is to support research and postgraduate training in the field of human genetics, genomics, and proteomics, with a focus on the bioethical aspect.

The IIGM research team focuses on genomic research and precision medicine, analyzing next generation sequencing data to try and understand how cancer evolves and whether they can assign therapies including immune therapies. Together, the researchers generate around 1800 samples of data per week and use different platforms to achieve their goals. The main platform used for sequencing is the Illumina machine, but they also use long-range sequencing technologies provided by Oxford nanopore technologies.

Challenges

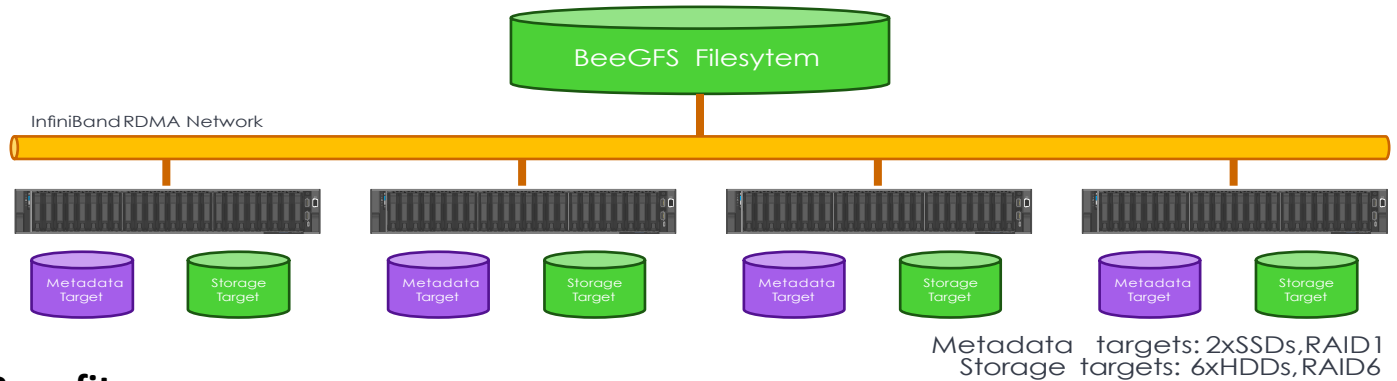
With a focus on translational research and precision medicine, and aiming to become a reference center in Italy, IIGM's team quickly realized that their previous infrastructure was not able to support the daily workload required for the team of scientists in order to continue to develop research projects with potential in terms of both scientific excellence and IP impact in the field of precision medicine.

The previous infrastructure consisted of a single outdated machine with 200 cores, and when the institute relocated to the cancer research center, they knew they needed a more powerful and cost-effective solution that could handle their data bandwidth and provide faster access to their data that would deliver more accurate results.

"In terms of time savings, we have noticed a significant difference in the speed of our projects since implementing these technologies. Not only are analyses completing faster, but the delivery of results is also quicker. This has allowed us to provide faster support to our collaborators and to make progress more efficiently" Professor Matteo Cereda, IIGM

Solution

To address these challenges, IIGM contacted ThinkParQ's partner Do IT Systems for a solution that would meet their budget and performance requirements on a platform that could scale in several dimensions. Do IT Systems designed a small but powerful turnkey system powered by the award-winning and leading parallel file system, BeeGFS. BeeGFS is used as the scratch and global file system for different workloads, providing very high-performance access from InfiniBand, and was installed on a Dell HPC system made up of 16 nodes, 8 of which have 28 cores and 8 have 48 cores, plus one node with two GPU's. The BeeGFS filesystem is delivered to the compute nodes by two additional Dell PowerEdge R740 equipped with fast SSDs for metadata and HDDs for storage. In total, 20 terabytes of data are run through BeeGFS.



Benefits

The implementation of the BeeGFS system has provided IIGM with a faster and more standardized infrastructure that offers ease of use, scalability, and improved performance. With over 30 people accessing the system across a broad range of roles, and many of them not having any HPC systems experience, setting up the environment and accessing the system was easy for everyone. There have been no user limitations or issues, and the users immediately found the benefits of using a HPC infrastructure that provides powerful access to their data, improved usability, and a strong line of support provided by Do IT Systems and ThinkParQ.

A key aspect of BeeGFS that has been highly appreciated by the system administrators is its ease of use and high performance speed, which makes scaling up and out completely transparent. For the users, BeeGFS allows them to run their projects within view hours while it runs behind the scenes, improving their workflow and delivering quicker results.

“The solution we have implemented is ready to scale up and scale out, allowing us to add additional servers as needed to increase performance and service capacity. Overall, the use of genomic technologies at our research institute has allowed us to save time and improve the efficiency of our projects, ultimately benefiting those we serve” Professor Matteo Cereda, IIGM

Future Plans

The future benefits that IIGM foresee from their use of BeeGFS include the opportunity to scale the system and upgrade it to connect with other institutes worldwide in all dimensions and is ready to move to the next level by integrating artificial intelligence to improve bioinformatic services for their patients. PowerEdge R740 equipped with fast SSDs for metadata and HDDs for storage. In total, 20 terabytes of data are run through BeeGFS.



Do IT Systems is the leading HPC service provider in Italy, offering solutions for industries operating in all the main HPC verticals (manufacturing, life sciences, energy, financial services). Do IT Systems services cover every stage of an HPC infrastructure life cycle (consultancy, installation, optimization, support). Do IT Systems operates in EMEA under the Do IT Now Alliance umbrella.”

About ThinkParQ:

ThinkParQ GmbH is a technology company that specializes in developing software solutions for high-performance computing (HPC). Main product is BeeGFS, a parallel file system that is designed to support high-bandwidth and low-latency I/O for HPC environments.