

QF2 for Genomic Sequencing

Qumulo File Fabric (QF2) is an ideal solution for storing, managing and accessing genomic sequencing data. It handles small files efficiently, and support of SMB, NFS, FTP and REST means that all phases of the genomic analysis pipeline can use the same QF2 cluster. QF2 is a modern file storage system that can scale to billions of files and that runs in the data center and the public cloud.



Performance

QF2 handles small files, such as TIFF and BCL, as efficiently as large ones. With QF2, researchers can perform their analyses in real-time, which translates into cost efficiencies and faster time to market.



Efficiency

QF2 makes 100% of user-provisioned capacity available for file storage, in contrast to legacy scale-up and scale-out NAS that can only support 70% to 80% of usable capacity. QF2's efficient use of disk space decreases the data footprint and saves not just on the cost of the storage system but on infrastructure costs such as power and cooling.



Real-time visibility and control

QF2's real-time visibility and control provides information about what's happening in the storage system, down to the file level, no matter how many files are in the system. System administrators can apply quotas in real time. The capacity explorer and capacity trends tools give IT the information it needs to plan sensibly for the future and not waste money because of overprovisioning.



Cloud and on-prem

Organizations that want to move some of their genomic analysis workloads to the cloud can take advantage of QF2 for AWS. QF2 has the highest performance of any cloud offering and is the only file storage system in the cloud with a full set of enterprise features, such as multi-protocol support and real-time visibility.

QF2 uses continuous replication to move data where it's needed, when it's needed. Continuous replication creates a copy of the data in a directory on your primary cluster and transfers it to a directory on a second, target cluster. Continuous replication is always running (unless you configure it not to). QF2 takes your latest changes and replicates them without you needing to worry about it.

Continuous replication means you can easily transfer data from your on-prem QF2 cluster to your QF2 cluster in AWS, perform your computations, and then transfer the results back to the on-premises storage.

Universal-scale file storage

Scales to billions of files

Use any mix of large and small files and store as many files as you need. There is no practical limit with QF2's advanced file-system technology.

Highest rated support

Get help fast from our team of storage experts with your own Slack channel.

Cloud-based monitoring

QF2 proactively detects potential problems, such as disk failures. You can also access historical trend data about how your system is being used.

Out-of-the-box simplicity

Skip the pain. From the moment QF2 is unboxed to when it can start serving data is a matter of hours, not days. QF2 for AWS can be set up instantaneously.

Complete REST API

Use the QF2 REST API to build and manage a modern application stack. It's the future of infrastructure, available today.

Our research organization falls between the cracks for most storage vendors, with giant imaging sets and millions of tiny genetic sequencing scraps. Finding a system that reasonably handled all our complex workflows was difficult, and in the end only QF2 was the right fit.

Bill Kupiec

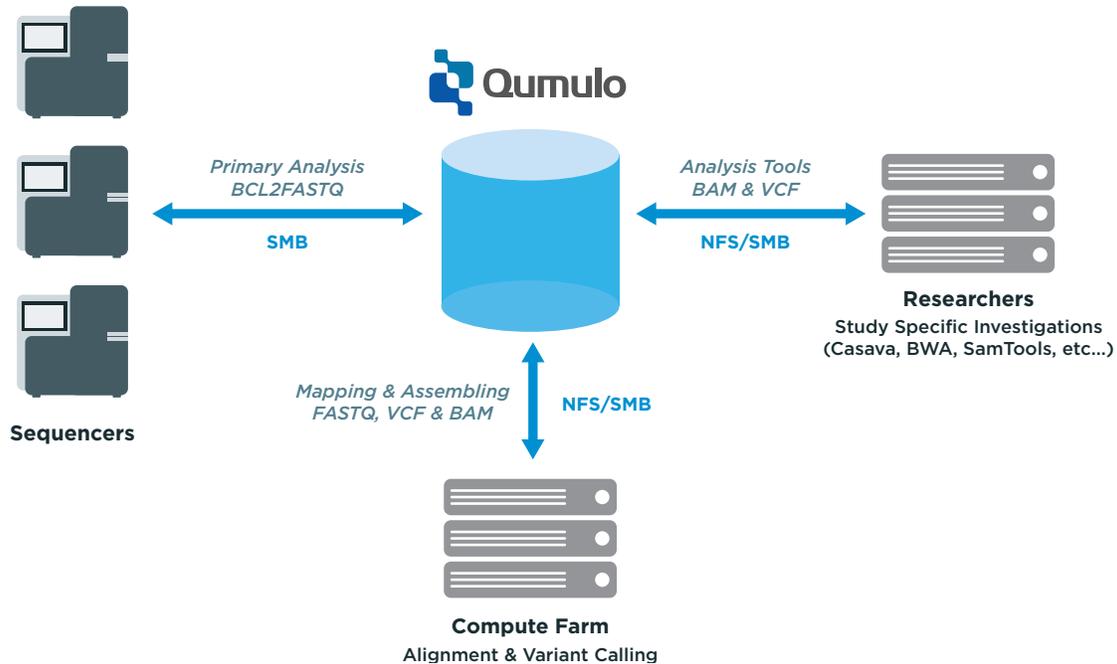
IT Manager
Department of Embryology
Carnegie Institution for Science



CARNEGIE
SCIENCE

NGS workflow on premises

Here is an example of an on-premises NGS workflow. (You can also run your compute farm in the cloud with QF2 for AWS.)



In this example, the DNA sequencers are generating many small BCL files or base calls, which are unordered DNA sequence fragments. A process of demultiplexing assembles BCL files into a FASTQ file, which is a text file that stores the combined output results of the BCL files along with corresponding quality scores.

The compute farm performs alignment and variant calling. In alignment, sequence fragments are quality checked, preprocessed and aligned to a reference genome. A BAM file is a binary file that stores this alignment data. Variant calling looks for differences between the data and the reference genome. Results are stored in a VCF file.

Once these data stores are ready, they can be used for application-specific analysis, which is done by researchers for their own projects. For example, a researcher might be working on a targeted therapy for patients whose tumor has a specific gene mutation. Researchers may use all the data that is generated contained in the BAM and VCF files.

QF2 provides a central file storage system that is suited for all types of genomic data. QF2 has industry-leading small file efficiency and has the throughput to handle all phases of the workflow.

About Qumulo

Qumulo is the leader in universal-scale file storage. Qumulo File Fabric (QF2) gives data-intensive businesses the freedom to store, manage and access file-based data in the data center and on the cloud, at petabyte and global scale. Founded in 2012 by the inventors of scale-out NAS, Qumulo serves the modern file storage and management needs of Global 2000 customers. For more information, visit <http://qumulo.com>.

