



CASE STUDY

The Hormel Institute, University of MN, Deploys Quantum ATFS to Accelerate Cancer Research

The Hormel Institute, a leading cancer research department at the University of Minnesota, chose Quantum ATFS to help address its operational and large data set needs. ATFS's unique architecture enables the Hormel Institute to accelerate research and manage data according to grant or project requirements—all while realizing an increase in ROI.

 THE HORMEL INSTITUTE
UNIVERSITY OF MINNESOTA

FEATURED PRODUCTS

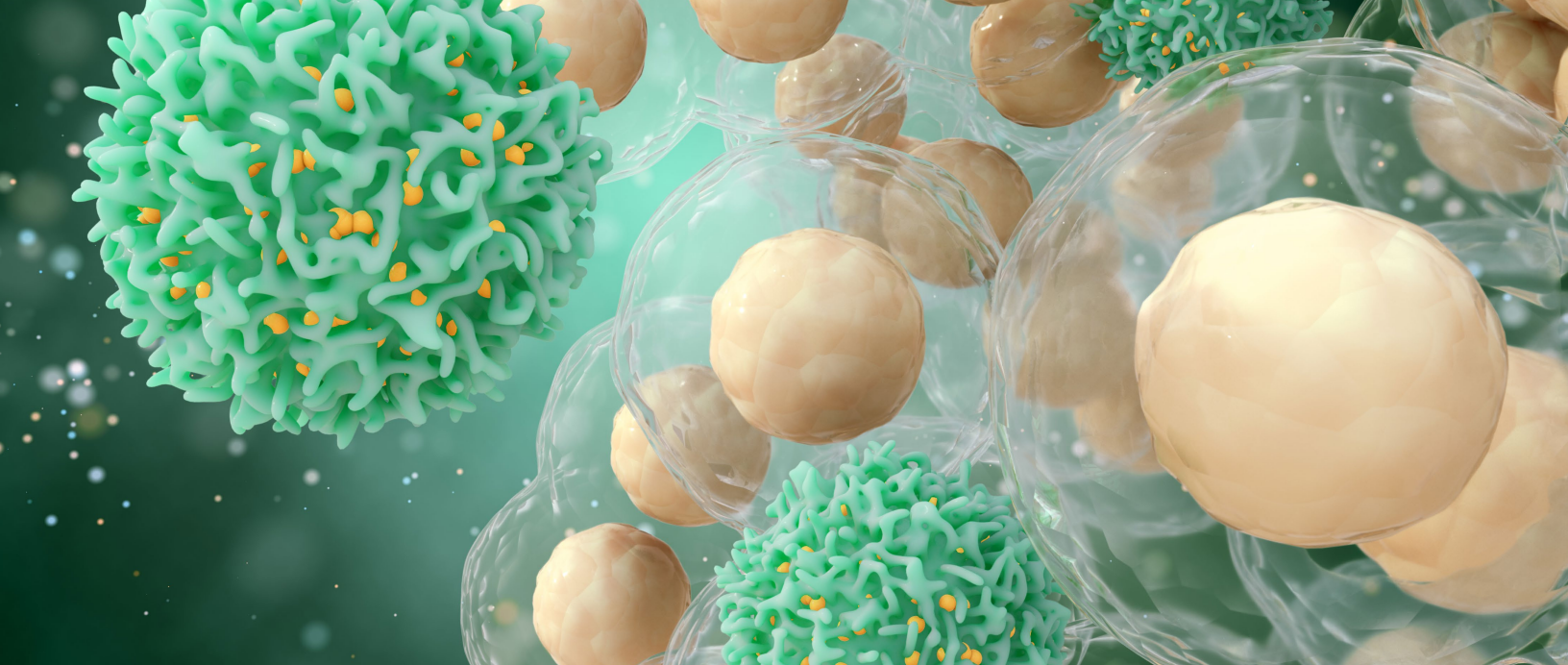
Quantum ATFS



“The engineering teams took a fresh look at the challenges teams face with growing storage needs and designed an innovative and unchained solution to those needs with cutting edge technology, data lifecycle and provenance tools and analytics to make any IT administrator exuberant about storage.”

Jeffrey McDonald

PhD, Director of Infrastructure at the Hormel Institute, University of Minnesota



I can view all of my data instantly; this used to take me days to get a usage report.

Jeffrey McDonald - PhD, Director of Infrastructure at the Hormel Institute, University of Minnesota



SOLUTION OVERVIEW

- Quantum ATFS Platform
 - 51.2 TB of NVMe flash
 - 2 PB of bulk storage using nearline disk

KEY BENEFITS

- Increased data mobility
- Greater data retention, collaboration and compliance
- Real-time reporting, status updates and issue resolution
- Consistent performance

Cancer, as a disease, dates back to some of the oldest civilizations. There is mention of the disease as far back as ancient Egypt—though it was Hippocrates in ancient Greece who first used the term cancer to describe the disease. Since ancient times, people have been searching for answers to prevent and treat cancer.

The Hormel Institute, a leading cancer research department at the University of Minnesota since 1942, is at the forefront of this research—with a focus on identifying the molecular mechanisms of cancer development and best ways to prevent, diagnose, and treat the disease. Currently the Institute includes more than 130 leading cancer scientists working across dozens of cancer research projects—each one contributing multiple terabytes of data to the Institute’s data storage system each week.

THE CHALLENGE

One aspect of the Institute’s research efforts involves the use of microscopes to study molecules. These powerful cryoelectromagnetic microscopes continuously capture images of

molecules over a period of a few days to a week. These images serve as the foundation for analysis performed by the Institute’s researchers.

Even just one microscope produces considerable data for researchers to parse through. A single cryoelectromagnetic microscope can produce 2 TB – 10 TB of data per run—and each run can take as little as a few hours or as much as a week to complete. The data created by the microscope are high resolution images of molecules, which need to be kept indefinitely for future research. Once data is captured, the images are processed, creating 3D images of the molecule. These images are used to identify effective ways for cancer drugs to achieve more desirable results.

After the microscope captures the images, it writes data into a shared storage repository where it can be accessed by researchers. The Institute hosts a 40 node HPC cluster whose schedule is managed by SLURM. Each researcher conducts investigations independent of others and prefers to keep their findings private until publication. Once processing has been

completed, data needs to be archived for either the life of the researcher or according to the terms of the grant.

Prior to the introduction of ATFS into the data environment, the Institute struggled to deliver data provenance, long-term retention, isolation of data by owner, real-time status updates, and workflow automation. There are more than 130 researchers at the Hormel Institute—each one funded primarily by grants—and each grant may dictate different data retention requirements. Without an accurate way to identify data, its owners, and its value, the Institute could not guarantee compliance with the grants' requirements.

In addition to microscope images, the Institute also conducts bioinformatics research. As part of this research, large data sets from DNA sequencers are analyzed, looking for patterns and commonalities that could help identify a cause or help with the treatment of a disease. The data used in this research is received based on a User Agreement from NIH and other research organizations. The User Agreements require the Institute to follow rules around data retention and deletion. An example of such rule: data must be deleted no later than six months after the completion of the analysis. Tracking data per the agreement and complying with the rules is a challenge—and noncompliance exposes the Institute to potential costly penalties.

THE SOLUTION

The Hormel Institute chose to implement ATFS after an extensive proof of concept, proving that ATFS's unique architecture and feature-set could help address its operational and organizational challenges. ATFS's unique architecture enables the Institute to accelerate research and manage data according to grant or project requirements—all while realizing an increase in ROI.

The Institute's new ATFS storage solution includes three main components:

The Zero-touch Classification Engine

This engine organizes data as it enters the system, capturing file metadata and applying user-defined tags based on business variables such as projects, data owners, or funding sources. This information can then be used to automate tasks, provide real-time status updates, determine data placement across media types and geographies—all while ensuring compliance with retention requirements. Additionally, the software is agile and can adjust to the use of different media, network, and other system components to maintain an optimal ratio of cost-value.

The Policy Orchestration Engine

Users can manage the data automatically according to its lifecycle requirements and simplify risk and compliance management by placing data according to policy. The engine can also deliver flexibility and customization according to user-defined variables—so one faculty member can keep data in flash for two weeks before archiving to bulk disk, while another faculty member places all their data in the cloud to enable public review and easy remote team access.

Flash Performance at Archive Pricing

The ATFS system has been designed to achieve optimal results in performance, cost, and capacity. The combination of NVMe, nearline disk, and cloud (public or private) delivers performance at a fraction of the cost compared to more traditional architectures. The hardware configurations can be modified to reflect changes in requirements for performance or capacity.

Hassle-Free Implementation

The system the Institute implemented consisted of 51.2 TB of NVMe flash and 2 PB of bulk storage using nearline disk. As the system ingests data, categorization tags are applied via the zero-touch classification engine and metadata is captured. Tags applied to data represent a user-defined category. The data orchestrator drives data placement, lifecycle, retention, and workflows.

ATFS's unique architecture enables the Institute to accelerate research and manage data according to grant or project requirements—all while realizing an increase in ROI.

ABOUT THE HORMEL INSTITUTE

The Hormel Institute, University of Minnesota, is a leading cancer research department of UMN and part of the Masonic Cancer Center, an NCI Designated Comprehensive Cancer Center. Their scientists are focused on accelerating answers to cancer by determining the basic molecular mechanisms of cancer development and finding better ways to prevent, detect, and treat cancer.



RESULTS

By adopting the ATFS storage system, the Hormel Institute was able to implement data retention and deletion policies to assure data provenance and silo research findings. The new solution also helped the Institute to realize an improved ROI while increasing performance and data mobility. Moving forward, the Institute believes these important updates will expedite research that can contribute to the eradication of cancer.

Data Mobility

The ATFS system moves and manages data according to its tag policies. Those policies often include movement of data to the appropriate media to match performance and/or retention needs. Data can be placed on media types, such as flash and nearline disk, on premises or in the cloud, or just copied to a different system as a backup. Data may also be moved per size, origin, file type, or access times. Using ATFS, the Institute can optimize data placement to reflect the needs of researchers.

Data Retention, Collaboration, and Compliance

Data can be classified to assure retention or deletion of data per each grant's specific requirements—including the isolation of data for compliance reporting purposes, or the isolation of data from others within the Institute or external collaborators. The ability to isolate data will also help protect research methodologies and results from disclosure before publication and in determining data provenance. Tags enable the presentation of data independent of its logical or physical location.

Data with long-term retention requirements may be archived to public or private cloud via S3. The Institute is considering using tape for long-term retention; when ATFS moves the data to an external system, file metadata and tags remain for continuous management of the data across its lifecycles.

Real-time Reporting, Status Updates and Issue Resolution

Everything that happens in the system is captured as it happens; this eliminates the

need to walk the file system using scripts or initiate data collections on performance or capacity consumption. The availability of stats as things happen allows for real-time status updates, searches, and problem identification. Alerts are sent as problems occur; once resolution has been achieved, another alert is sent out to update the status. With just-in-time management of its system, the Institute can be confident it has accurate information related to resource availability, hot-spotting, anomaly occurrences, and issues.

Consistent Performance

Using prefetch and read-ahead functionality, ATFS can accelerate access to data and streamline retrieval. Tags may also be used to identify data that requires higher performance storage at a predefined period of time. Moving data into flash when needed reduces the amount of flash required overall, which helps reduce storage costs and improves system performance across each of the Institute's research projects.

Quantum

Quantum technology and services help customers capture, create, and share digital content—and preserve and protect it for decades at the lowest cost. Quantum's platforms provide the fastest performance for high-resolution video, images, and industrial IoT, with solutions built for every stage of the data lifecycle, from high-performance ingest to real-time collaboration and analysis and low-cost archiving. Every day the world's leading entertainment companies, sports franchises, research scientists, government agencies, enterprises, and cloud providers are making the world happier, safer, and smarter on Quantum. See how at www.quantum.com.

©2020 Quantum Corporation. All rights reserved. Quantum and the Quantum logo are registered trademarks of Quantum Corporation and its affiliates in the United States and/or other countries. All other trademarks are the property of their respective owners.

www.quantum.com

CS00498A-v01