



CASE STUDY:

HPC and Storage Strategic Development Plan to Maximize Scientific Innovation



Challenge

Staying at the forefront of scientific innovation is difficult for any organization, but it's especially hard for medical research groups. One such institution, upon realizing they needed a change, decided to ease their transition by working with BioTeam. They knew that Life Sciences research had changed significantly in the past two decades and that the required skills and infrastructure needs had evolved. Finally, decision makers realized they needed some help understanding where they could improve.

The institute's scientists already had massive amounts of data and worked with complex, next-generation workflows that required enormous amounts of computational power. They also knew that the amount of research data would only continue to increase exponentially, as would the complexity and computational needs of the workflows.

Therefore, the medical research institute needed experts to provide strategic recommendations. Specifically, they needed HPC and storage capabilities that could handle the current and future needs of their Computational Biology and Computational Sciences research teams. Additionally, they required technology and staffing resources that focused on the highest return on investment.

Approach

The institute hired BioTeam to develop an HPC and Storage Strategic Development Plan that (1) provided guidelines for the institute to improve its infrastructure and staff, and (2) invested in the best solutions to support its scientific research teams' current and future needs.

BioTeam's experts believe that a successful strategic plan must be founded on a holistic perspective that goes beyond technology. For this reason, BioTeam evaluated the impact of altering the institute's scientific research needs regarding data science, computing, and technology infrastructure. At the same time, BioTeam needed to evaluate organizational, cultural, process, policy, and operational issues that could impact successful deployment of a new infrastructure.

To gain insights, BioTeam first performed an assessment which included infrastructure surveys and interviews with scientists, technologists, and HPC staff covering a broad spectrum of scientific and service needs. Insights gleaned from the interviews identified significant improvement opportunities in staffing, HPC provisioning and queuing, storage, data management, data sharing, and Identity and Access Management (IAM).

In the interviews, we discovered there was no process in place that allowed scientists to make decisions about data management. For example, identifying what data was of high value or part of an active analysis and what data was irrelevant or easily recreated. This resulted in a large amount of budget being spent on fast, expensive storage for all data. BioTeam recommended a data management solution that could seamlessly tier data from one system to another and evaluate data value using metadata. The result introduced a tiering solution that enabled scientists to identify what data should be considered more active or less active and allowed the organization to optimize the storage spend to include larger slower storage for less active data without impacting the active data that required faster storage.

Outcome

After discussing critical scientific and technical challenges, BioTeam outlined an HPC and Storage Strategic Development Plan that the institute's computing staff could act on in the short and long term. Recommendations and/or detailed implementation plans included the following:

- Increasing the HPC group staffing and their scope of responsibilities.
- Simplification of HPC storage and replacement of cumbersome parallel file systems.
- Network reorganization to create separate research zones while enabling data sharing.

- Improvement of data management and collaboration between HPC staff and researchers.
- Software, cloud, data management, visualization, and workflow management services.
- Simplification and optimization of the HPC environment, using tools and resources already available at the institute to add capabilities. This included detailed recommendations regarding ways to:
 - Develop a consistent provisioning process
 - Develop a data management strategy
 - Consider all-flash storage for HPC
 - Consider moving from LSF to Slurm
 - Integrate HPC into the AD environment

BioTeam's recommendations simplified what appeared to be a sprawling, disjointed HPC environment. The result simplified the scientists' experience and decreased the management burden on the HPC team. BioTeam's experts were thrilled to help this institute with streamlining their processes. Efficiency is a vital component in medical research, and the improvements implemented here will help increase the institute's performance and productivity.

Find more case studies at bioteam.net/tag/case-studies/