

OMRF Accomplishes More Big Science Research Faster with the Weka Limitless Data Platform

CHALLENGES

- Mixed workload
- Poor storage performance
- Increase in workload size, complexity, and throughput
- Faster, easier access to object storage data archive
- Budget constraints

BENEFITS

- Simple architecture
- Speedier jobs with a 10x reduction in run time
- Limitless capacity scaling by tiering to object storage archive
- More research with multiple concurrent projects in a shorter time

Weka's Limitless Data Platform™, the fastest-growing data platform for modern enterprise workloads from WekaIO™ (Weka), facilitates more and bigger science research faster than any other solution for the Oklahoma Medical Research Foundation (OMRF). Located in Oklahoma City, Oklahoma, OMRF is an independent, nonprofit biomedical research institute with more than 450 staff and over 50 labs studying cancer, heart disease, autoimmune disorders, and diseases of aging.

Established in 1946, OMRF is dedicated to understanding and developing more effective treatments for human disease. Scientists at OMRF have done more than explore the mysteries of human disease—they have helped develop medications that fight deadly illnesses. Discoveries at OMRF led to the first targeted therapy approved in the U.S. for sickle cell disease and the first approved treatment for neuromyelitis optica spectrum disorder, a rare autoimmune disease. OMRF's critical research is helping people live longer, healthier lives, one discovery at a time.

THE CHALLENGE: ENABLE MORE COMPUTE, MORE STORAGE, AND FASTER STORAGE FOR BIGGER DATA SETS AND MIXED WORKLOADS

The research efforts at OMRF change with the changing problems in human health. And for every question answered, more questions are asked. With over 300 scientific staff members, including some of the world's foremost immunologists and cardiovascular biologists, OMRF has become one of the nation's leading independent medical research institutes.

The key focus of the research computing team at OMRF is supporting high performance computing (HPC) for general bioinformatics work. One very common workflow is Next-Gen Sequencing (NGS) analysis using the GATK pipeline for sequence alignment and variant calling. However, the cluster supports numerous research jobs running simultaneously with unique toolsets. The challenge for the OMRF research computing team was to architect a system for scientists that have growing informatics needs for their research: more compute, more storage, faster storage, and "bigger" data.

The data the team manages comes from high throughput, high dimensional lab instruments, such as NGS and Microscopy. And they support mixed workloads that can vary greatly ranging from jobs with a few very large files (100s GB), to jobs with many very large files, to jobs with thousands of tiny (< 1MB) files, and jobs doing lots of meta operations. The decade old Isilon system used to meet the needs of capacity and was never a significant bottleneck to productivity when software stacks were not well developed. But with the increase in workload size, complexity, and throughput and the maturing of analysis workflows (highly-tuned CPUs and emergence of GPUs) the legacy storage system was no longer keeping pace with the growing performance demands. For the "hot" active tier, OMRF needed faster storage that could handle mixed workloads and a cost-efficient object storage back-end solution to manage larger capacity. To summarize, OMRF needed a storage solution that would:

- provide better throughput and remove the storage I/O bottleneck
- enable concurrent research jobs
- tier seamlessly to object storage archive
- be cost-efficient
- speed data access to the applications
- future-ready the data center that has begun exploring GPUs to accelerate compute



“Our scientists can more easily complete their analysis and no longer have to deal with the storage. Our old workflow was complex and time-consuming, involving staging data to a compute nodes local SSD for better application performance.”

With Weka we have so much performance and expandable capacity that we don't have to think about it.

Stuart Glenn, Assistant Director of Infrastructure & Research Computing, OMRF

THE SOLUTION: WEKA LIMITLESS DATA PLATFORM ON SUPERMICRO SERVERS

Providing a fast and cost-efficient storage system was the primary driver for the OMRF data center team to choose Weka. The team considered other products from their legacy storage vendor, even considering all-flash solutions. They also considered BeeGFS-based systems, both as a vendor supplied appliance and as a build your own model. But the alternative solutions didn't provide the simplicity, speed, or scale that OMRF required. Ultimately, OMRF deployed the Weka Limitless Data Platform on-premises and on Supermicro® servers and achieved all three as benefits.

OMRF manages 950TB of data—100TB active on Weka, 250TB on a slower scale out NAS with standard file shares to satisfy a medium-term data retention strategy, and 600TB in an object storage archive. The cluster consists of standard x86 CPUs from Supermicro connected via 100Gb Ethernet switches. The system was architected to stream data from the lab to the Weka storage via SMB. The team, who has begun experimenting with GPUs, now has peace-of-mind knowing that Weka provides the best I/O throughput to GPU-enabled servers rendering their updated cluster architecture future-ready. Weka's Limitless Data Platform is built on the Weka File System (WekaFS™), a software-defined storage architecture that delivers the industry's best performance and cost-efficiencies by leveraging the latest technologies in storage such as NVMe, networking technologies like NVMe-oF, NVIDIA Mellanox InfiniBand, 100Gb Ethernet, and advances in computing technologies like GPU acceleration.

THE WEKA INNOVATION NETWORK™ (WIN) DELIVERS COMPLETE SOLUTIONS

Silicon Mechanics, a WIN Leader Partner, is one of the world's largest private providers of high-performance computing (HPC), artificial intelligence (AI), and enterprise storage solutions. The team has been working with OMRF for several years and was well-versed with their unique workloads and organizational requirements. When OMRF began this new project, Silicon Mechanics designed a new, custom hardware infrastructure that would support its specific computing goals. The in-house manufacturing team quickly built and tested the solution, then sent it to Oklahoma. Since 2001, Silicon Mechanics' clients like OMRF have relied on its custom-tailored open-source systems and professional services expertise to overcome the world's most complex computing challenges. With thousands of customers across the aerospace and defense, education and research, financial services, government, life sciences, healthcare, energy, and oil and gas sectors, Silicon Mechanics solutions always come with "Expert Included"™. Learn more at <https://www.siliconmechanics.com/>

The Supermicro Storage Appliances, featuring WekaFS, are preconfigured and optimized for maximum acceleration and reduced training times, delivering unmatched performance at scale. This family of appliances from Supermicro, a WIN Innovator Partner, are based on the Supermicro Ultra SuperServer® and BigTwin® server platforms you can take advantage of a plug-and-play engineered solution that helps extract greater value from data. This solution is available in a set of appliances that allows you to easily select a choice that best suits your performance, capacity, and footprint requirements. For more information on Supermicro Storage Appliances Featuring WekaFS, go to <https://bit.ly/3tLYdlh>

BENEFITS AND RETURN ON INVESTMENT

By implementing the Weka Limitless Data Platform, the OMRF team was able to achieve better throughput and run more research jobs concurrently without negatively impacting other jobs or workloads. In addition, the turn-around time is better, because the jobs finish faster the results get to the scientist quicker which accelerates the next stage of their research. The research workflows are greatly simplified, because using Weka eliminated the complexity of staging-in and -out data into a compute node's local SSD. Research outcomes are no longer limited by how much data can be stored on a compute node's local SSD, with WekaFS acting as a front-end they have faster and easier access to their object archive tier ensuring the applications have access to all the data. Ultimately, OMRF now has so much performance and expandable capacity available to all nodes that nobody has to think about storage any longer—instead focusing on saving lives.

- **Faster Time to Answer:** OMRF research jobs were reduced 10X, one job was reduced from 70 days to 7 days; another common analysis workflow was reduced from 12 hours to 2 hours.
- **Supports Multiple Concurrent Projects:** OMRF can support multiple and simultaneous new research initiatives; faster turn-around time and jobs finishing faster gets results to researchers faster.
- **Cost-Efficiency:** The object tiering feature of Weka doubled the available capacity of the scratch space, for a lower overall storage cost. OMRF shopped alternative solutions and found all-flash combined with object storage for additional capacity to be 1.9-2.4X the price of Weka per usable TB. And while the SATA-based hybrid solutions may have been slightly less \$/TB, the 30% performance improvement delivered by Weka offset the slight difference in acquisition cost. By comparing \$/IOP or \$/RW throughput, the Weka solution came out 8-10X ahead of both the all-flash and hybrid solutions.

For more information or to locate a partner in the Weka Innovation Network, go to: <https://www.weka.io/partners>

For more information on OMRF, go to: <https://omrf.org>



910 E Hamilton Avenue, Suite 430, Campbell, CA 95008 T: 408.335.0085 E: info@weka.io www.weka.io

©2018-2021 All rights reserved. WekaIO, WekaFS, WIN, Weka Innovation Network, Weka Within, the Weka brand mark, the Weka logo, Limitless Data Platform, and Radically Simple Storage are trademarks of WekaIO, Inc. and its affiliates in the United States and/or other countries. Other trademarks are the property of their respective companies. References in this publication to WekaIO's products, programs, or services do not imply that WekaIO intends to make these available in all countries in which it operates. Product specifications provided are sample specifications and do not constitute a warranty. Information is true as of the date of publication and is subject to change. Actual specifications for unique part numbers may vary.