

WEKA for Life Sciences

Accelerate Research with The AI-Native WEKA™ Data Platform—the World’s Fastest Data Platform for Life Sciences

CHALLENGES

- Reduce research costs
- Accelerate research insights
- Optimize shared compute
- Scale up and down
- Manage mixed IO
- Improve time to market

WHY WEKA?

- Accelerate Drug Discovery with dynamic performance and capacity scaling
- Increase Collaboration with integrated cloud connectivity
- Protect Critical Data without sacrificing performance
- Speed Time to build models
- Simplify the Environment including backup & DR
- Lower the Cost to process and store data
- Scale to Meet the Demands of modern pipelines
- No Need to Copy datasets between locations

\$9 billion

expected value of AI in life sciences by 2032

90%

the amount researchers can accelerate project times with the WEKA Data Platform

up to 7x

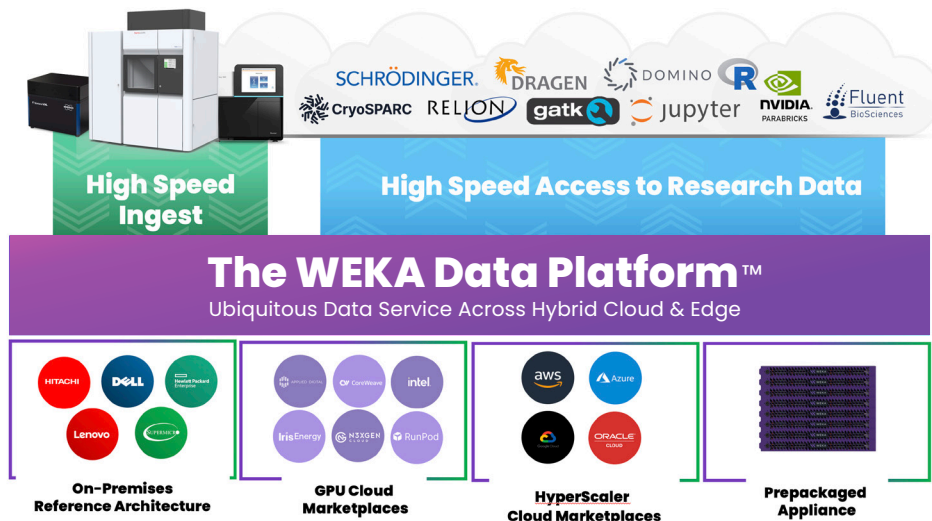
improved application performance with WEKA

Today’s scientific discovery requires conducting more experiments in less time with more data. Researchers need flexible infrastructure to affordably handle trillions of files and directories on-premises and in the cloud without slowing down their research. The WEKA Data Platform lets you tame unruly data management and be ready to take on any scientific application, at any time, and in any location.

WEKA eliminates the need to over-provision storage resources to meet performance-intensive life science application needs. The bi-directional scaling enables you to rapidly scale up your research workflows to meet your timelines and then scale back down so you never pay for resources you don’t use.

Keeping a GPU compute cluster fed with data is critical to overall pipeline performance and efficiency. WEKA powers GPUs up to 20x faster, which means your GPUs and modeling software spend less time idle waiting for data. Our zero-tuning capability also assures that every application in your analytics pipeline gets the performance it needs.

WEKA’s zero-copy architecture eliminates the need for multiple data copies and complex data operations, reducing time to insights for your researchers and scientists. Our single namespace for flash and object storage with smart tiering ensures hot data is always ready for analysis, improving application performance by as much as 7x. And because we know flexibility is key, WEKA is designed to be deployed on-premises, supports hybrid and cloud deployments, and integrates seamlessly with AWS.



20x

lower management overhead

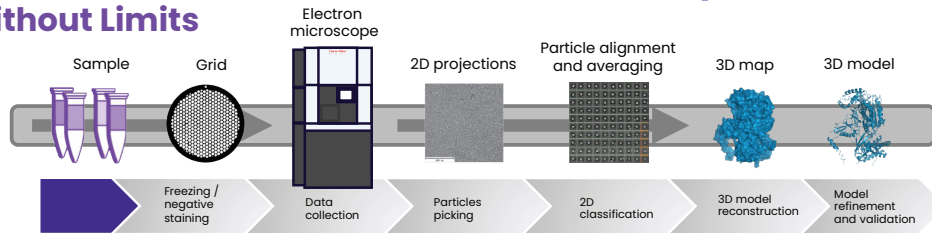
up to 75%

lower infrastructure costs

up to 30%

less in wall-clock time

How The WEKA Data Platform Accelerates CryoEM Research without Limits



- Run on-prem, public cloud, or hybrid
- Zero tunables out of the box
- Half the latency of NFS storage
- 12X faster model training
- Scale from 10's of TB to 100's of PB's
- Archive in low-cost "Glacier" storage

The WEKA Data Platform™
Zero Copy, Zero Tuning Architecture

WEKA Accelerates Many Research Areas & High-Performance Life Science Workloads

- Drug Discovery
- Drug Design
- Drug Repurposing
- Drug Development
- Genomics
- Genetic Engineering
- Synthetic Gene Design
- Molecular Dynamics
- Generating Novel Molecules
- Predictive Peptide Design
- Protein Engineering
- Biomarker Identification
- Med Device Design
- Medical Imaging Analysis
- Personalized Medicine
- Precision Medicine
- Designing Ligands de Novo
- Predicting Protein Stability
- Predicting Protein Expression
- Automating Data Collection
- X-ray Crystallography
- Cryo-EM
- Natural Language Models
- Mathematical Modeling
- Disease Diagnosis
- Research Collaborations
- Disease Prediction
- Disease Surveillance
- Supply Chain Efficiency



Problem: Needed to improve AI epoch times and implement a less costly solution that eliminated copying to multiple EBS instances

Solution: Migrated from native services to reduce epoch time from 3 months to 1 week with 100%+ savings

92%

reduced epoch time in AWS

40x

faster creating and copying 1GB files

30M

files being used to train the model



Problem: Needed a highly-performant solution that allowed for scaling a shared data set that had scaled beyond NAS capability but OBJ was too slow

Solution: Scaled capacity by 5x, increased performance by 10x and reduced costs by 75% vs Isilon

72x

faster genome pipeline compared to CPU

140PB

of data in storage

75%

less storage cost per genome

10x

your genomic workflows

WEKA genomics customers have experienced compute workloads being **reduced from 70 days to as low as 7 days**

Customers Who Trust WEKA with Their Life Science Research



SAMSUNG



Swiss Institute of Bioinformatics



weka.io

844.392.0665

