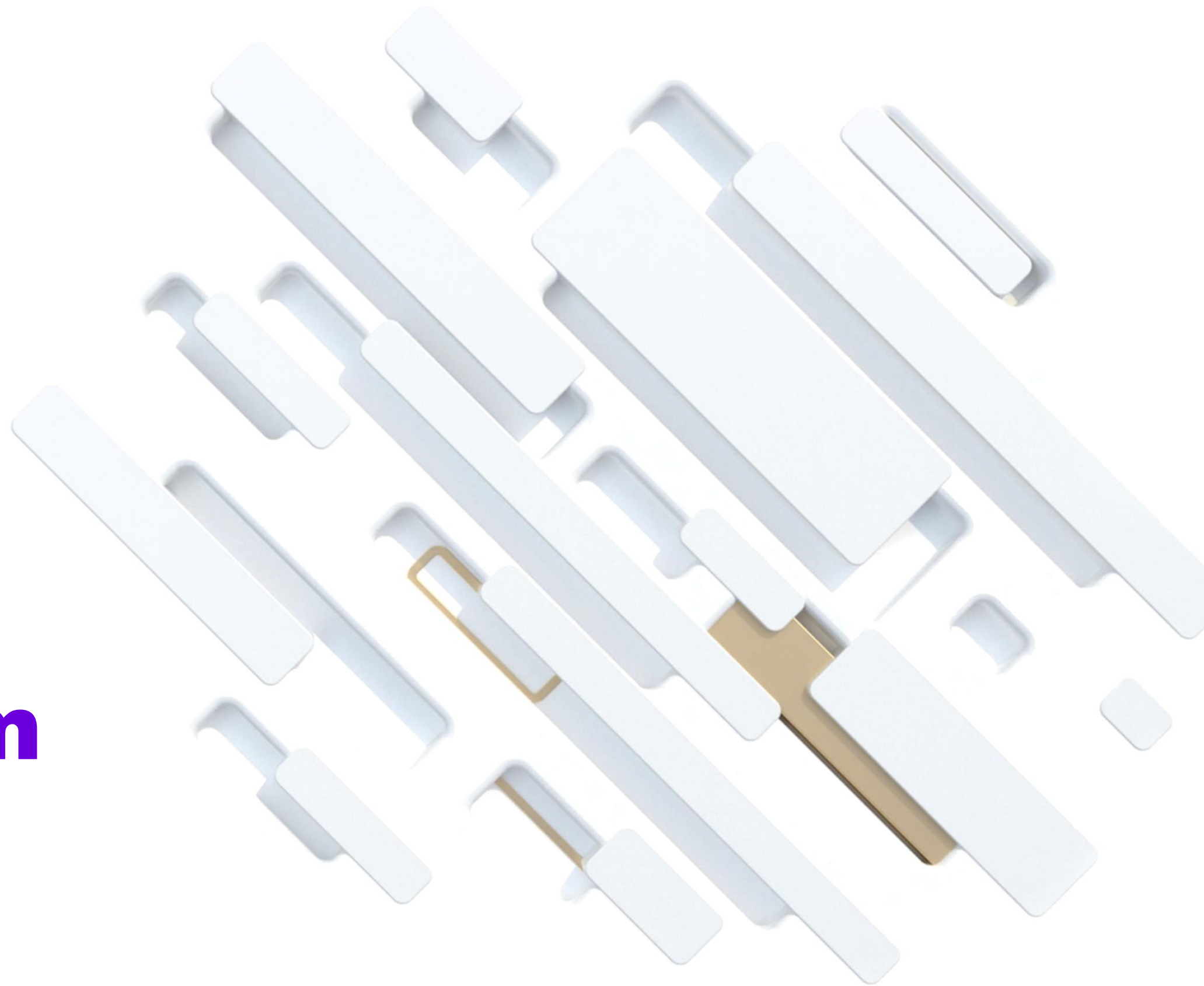




WEKA Data Platform

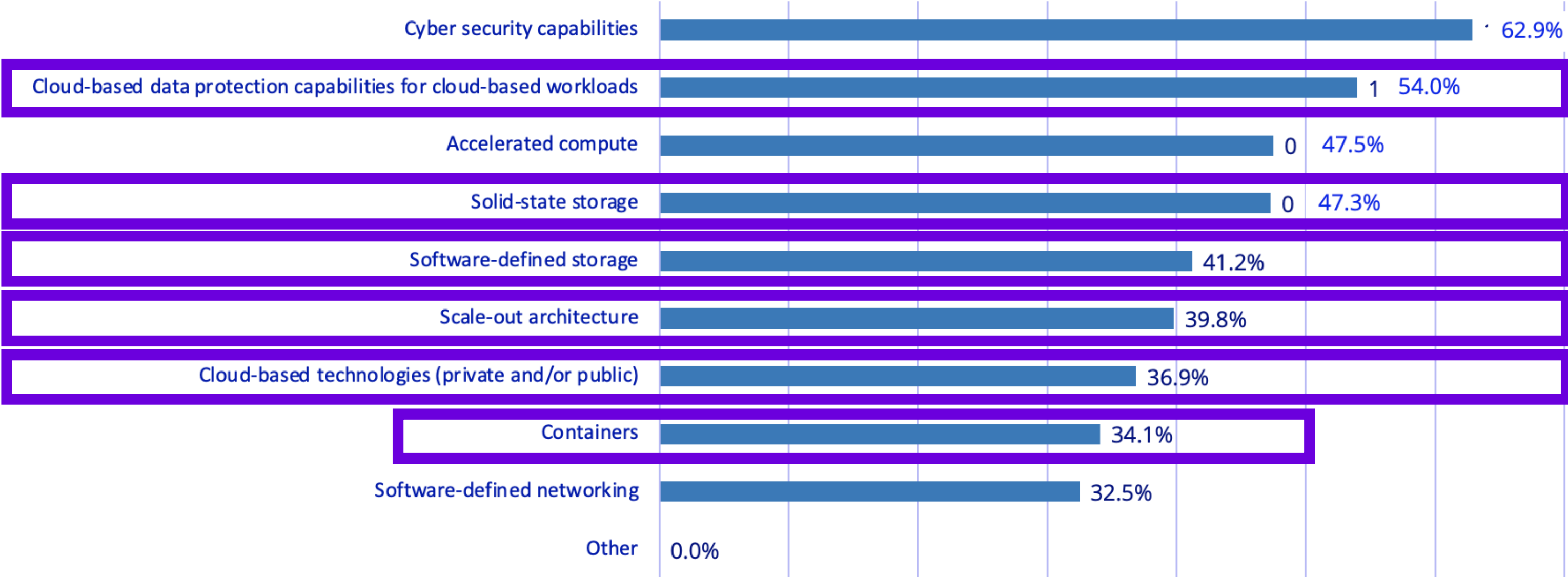
Shimon Ben David

CTO



IDC - Most Desired Technologies for Infrastructure Modernization

What key technologies do you want to leverage for your infrastructure modernization initiative?



DISRUPTIVE TRENDS ENABLE NEW USE CASES

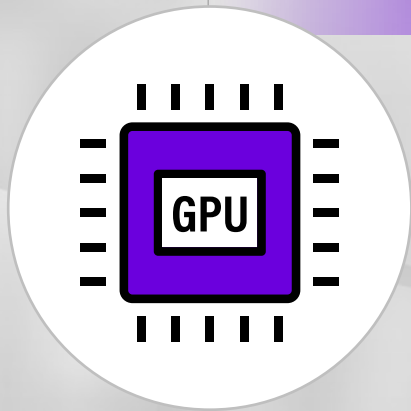
CLOUD



NVMe



GPU COMPUTE



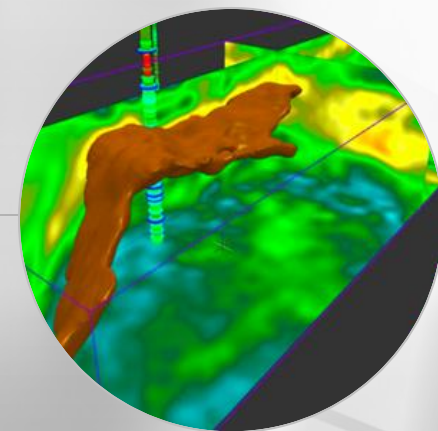
Life Sciences



Automotive



Financial Services



Oil & Gas



Media & Entertainment

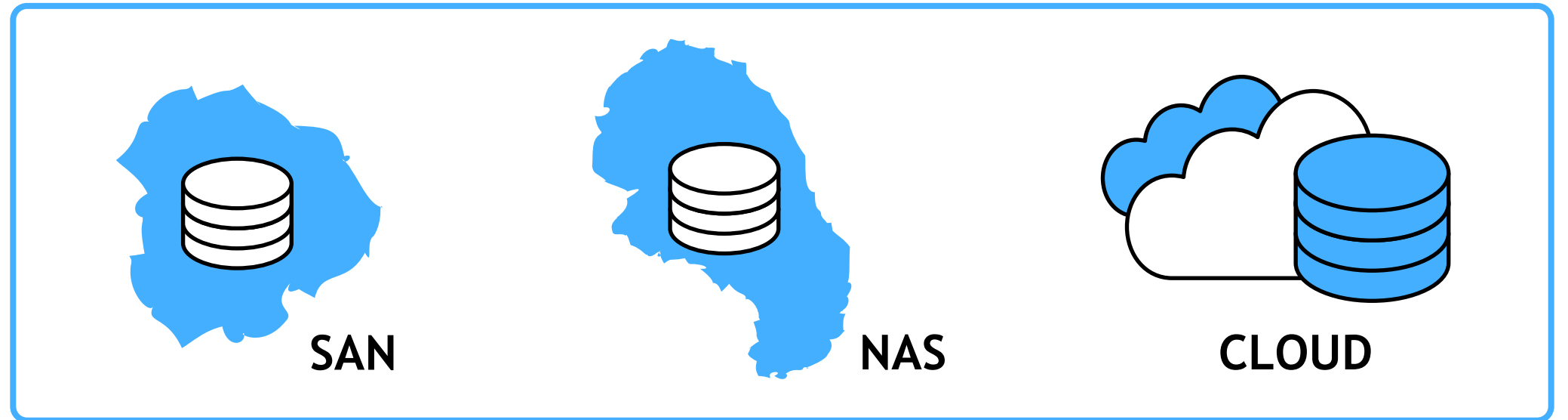


Research

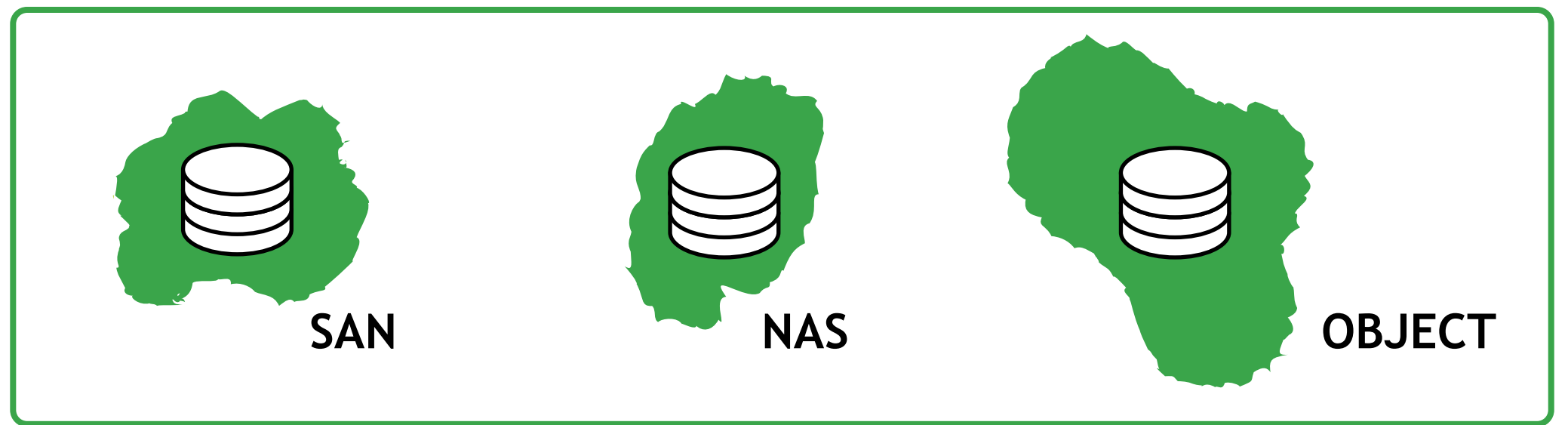
LEGACY DATA MANAGEMENT IS INEFFECTIVE

- Multiple architectures
- Incompatible tools and processes
- Complex management
- Multiple data copies are inefficient and ineffective
- Cloud adds new layers of complexity

PERFORMANCE



CAPACITY



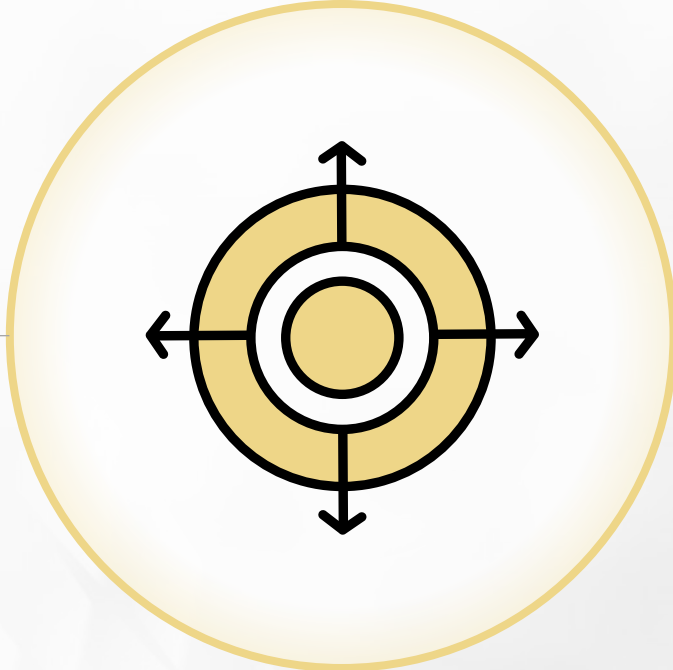
LEGACY APPROACHES NEED DIFFERENT PLATFORMS



SIMPLICITY

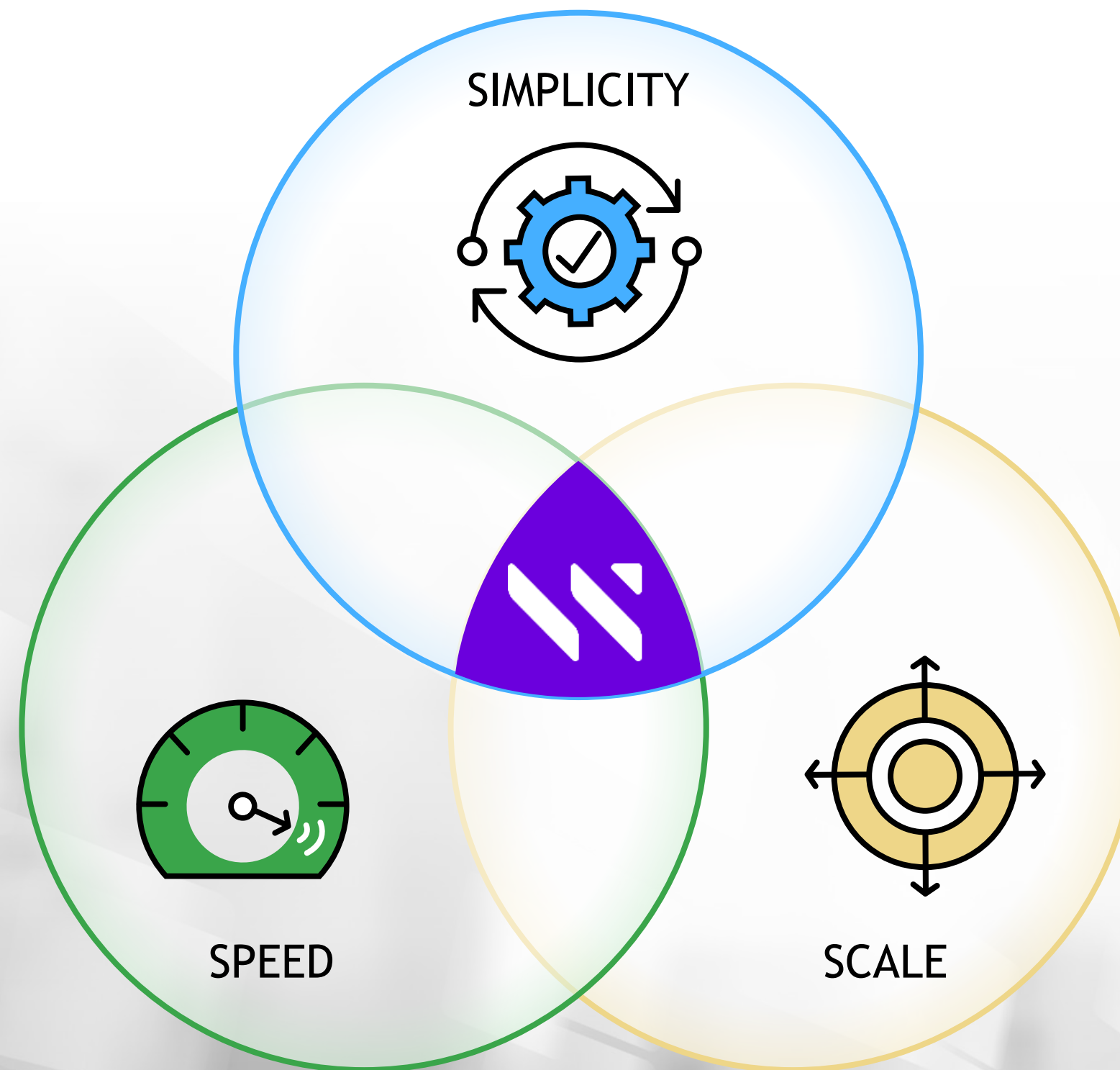


SPEED



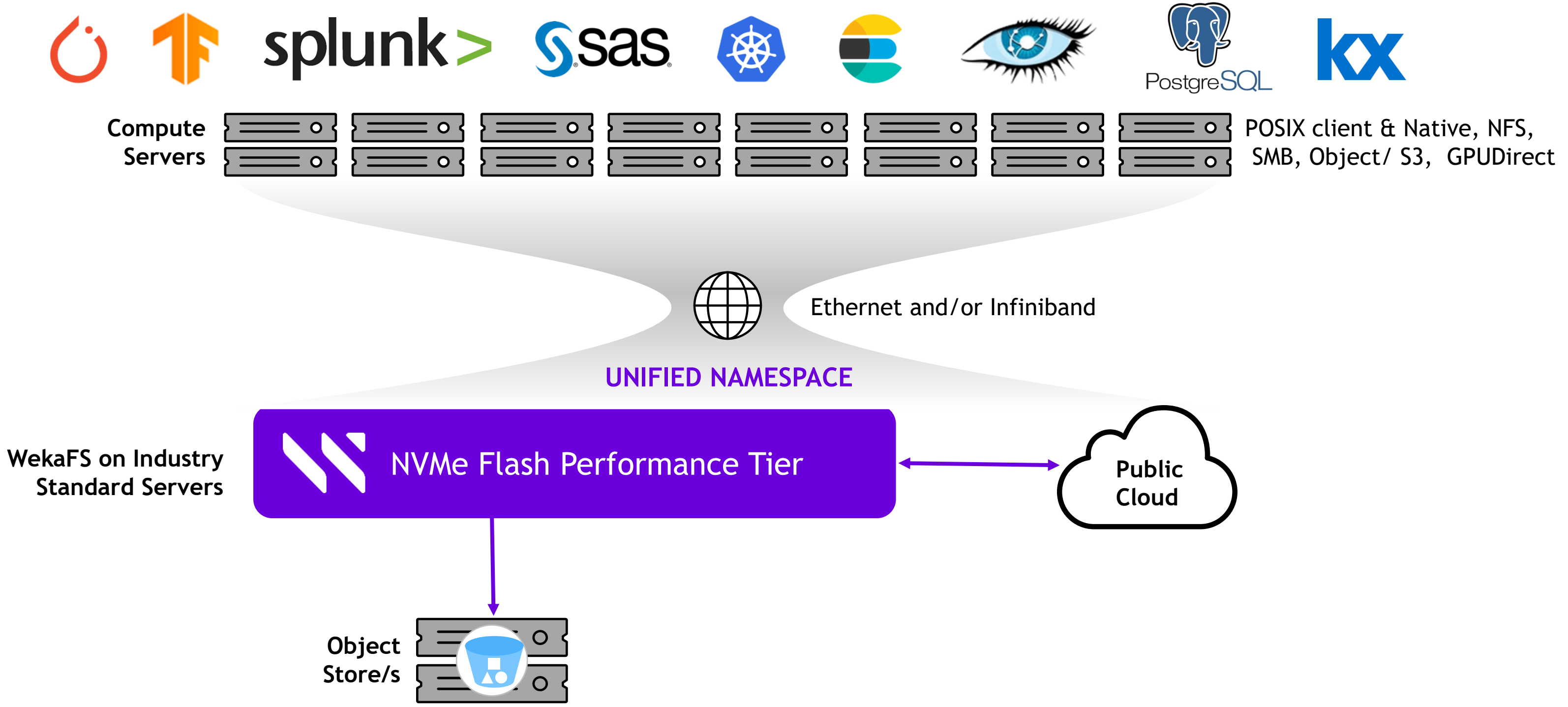
SCALE

LIMITLESS DATA PLATFORM

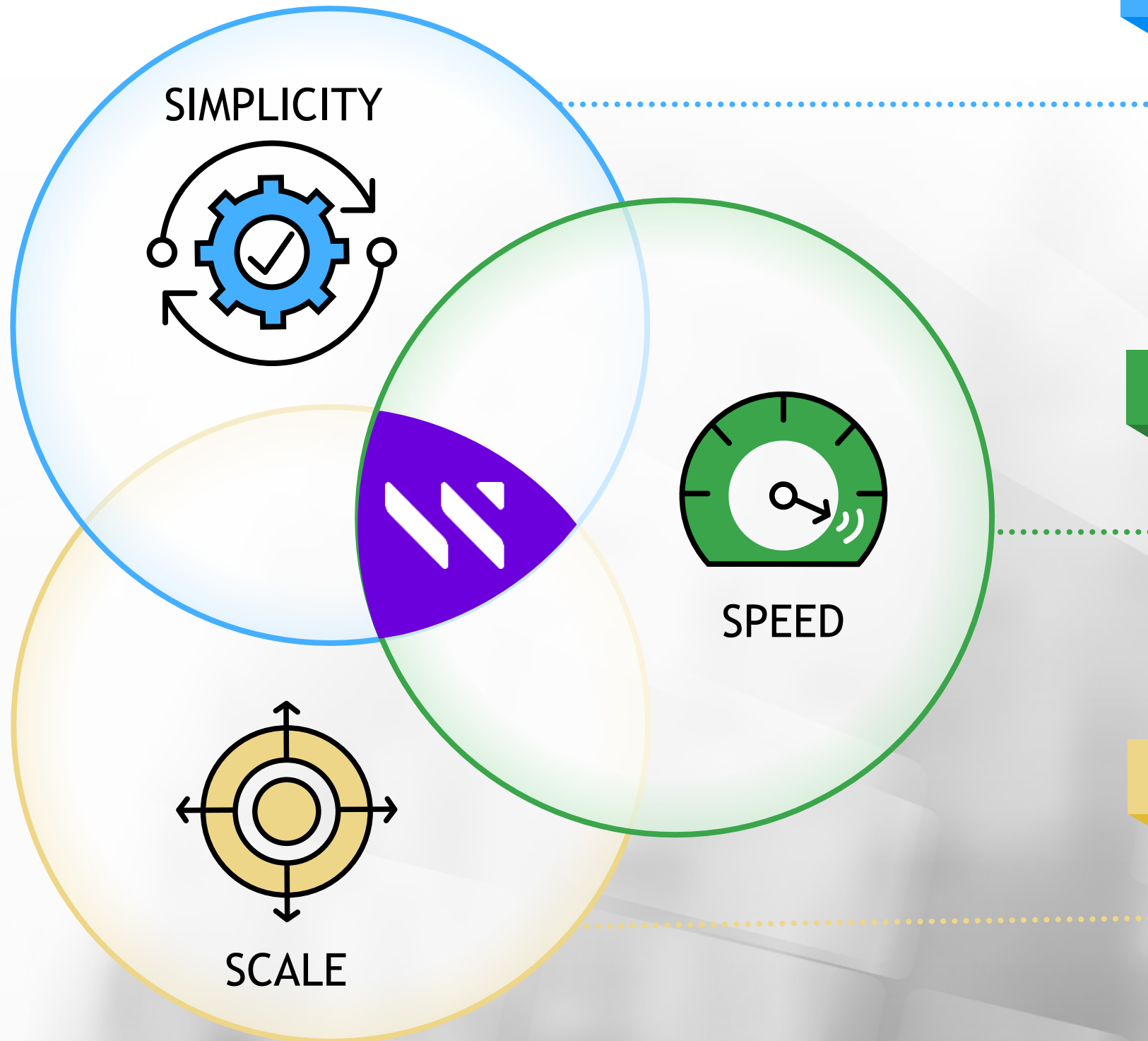


FOR MODERN ENTERPRISE APPLICATIONS

THE LIMITLESS DATA PLATFORM



LIMITLESS DATA PLATFORM



SIMPLICITY

Unifies all your data

Single pane of management

Any platform-cloud or on-premises

SPEED

Industry Leading Performance

Resiliency without performance impact

Amazing latency across diverse workloads

SCALE

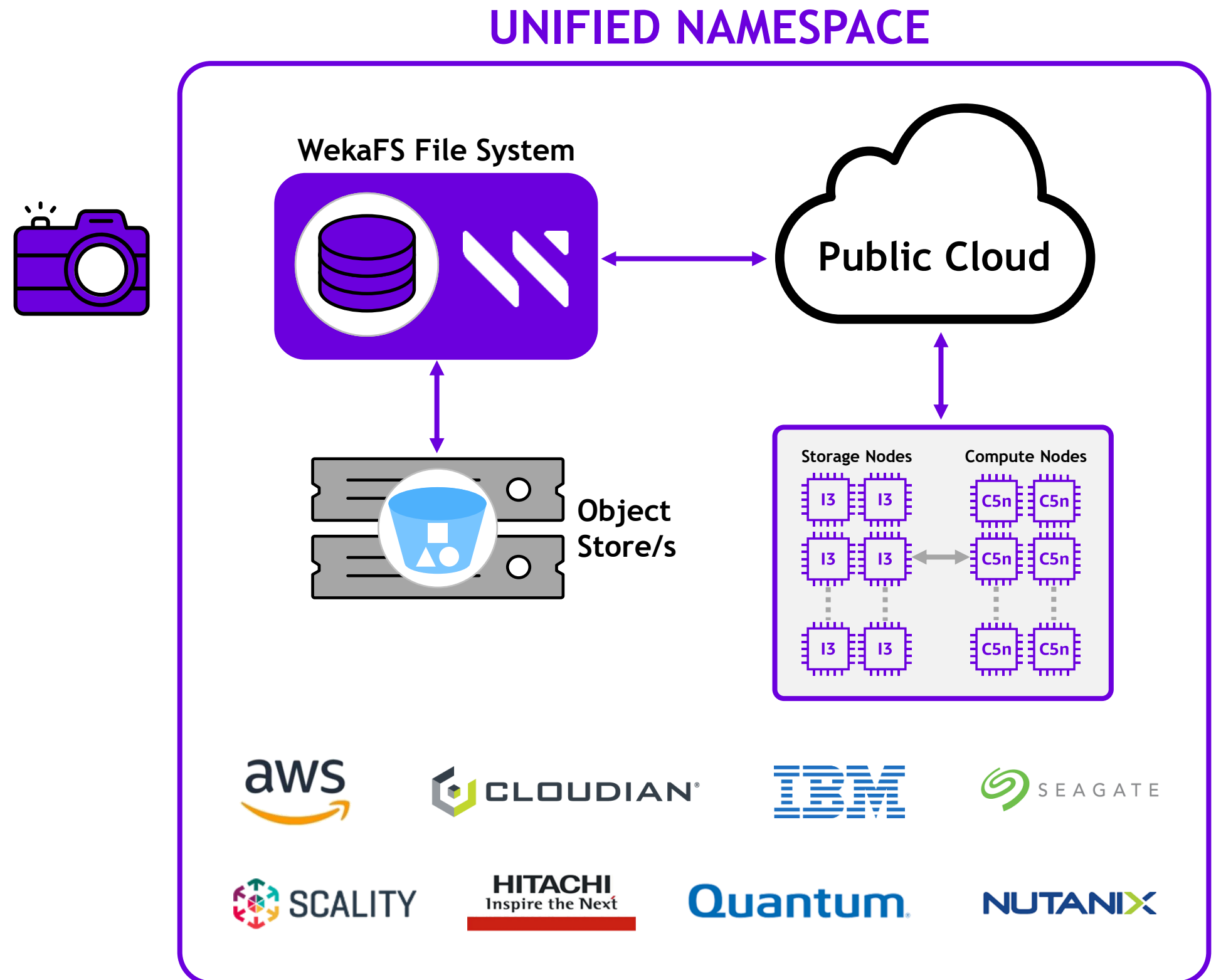
From Terabytes to Exabytes

Data mobility to cloud

Flash & disk for best economics

SIMPLIFIED DATA PROTECTION

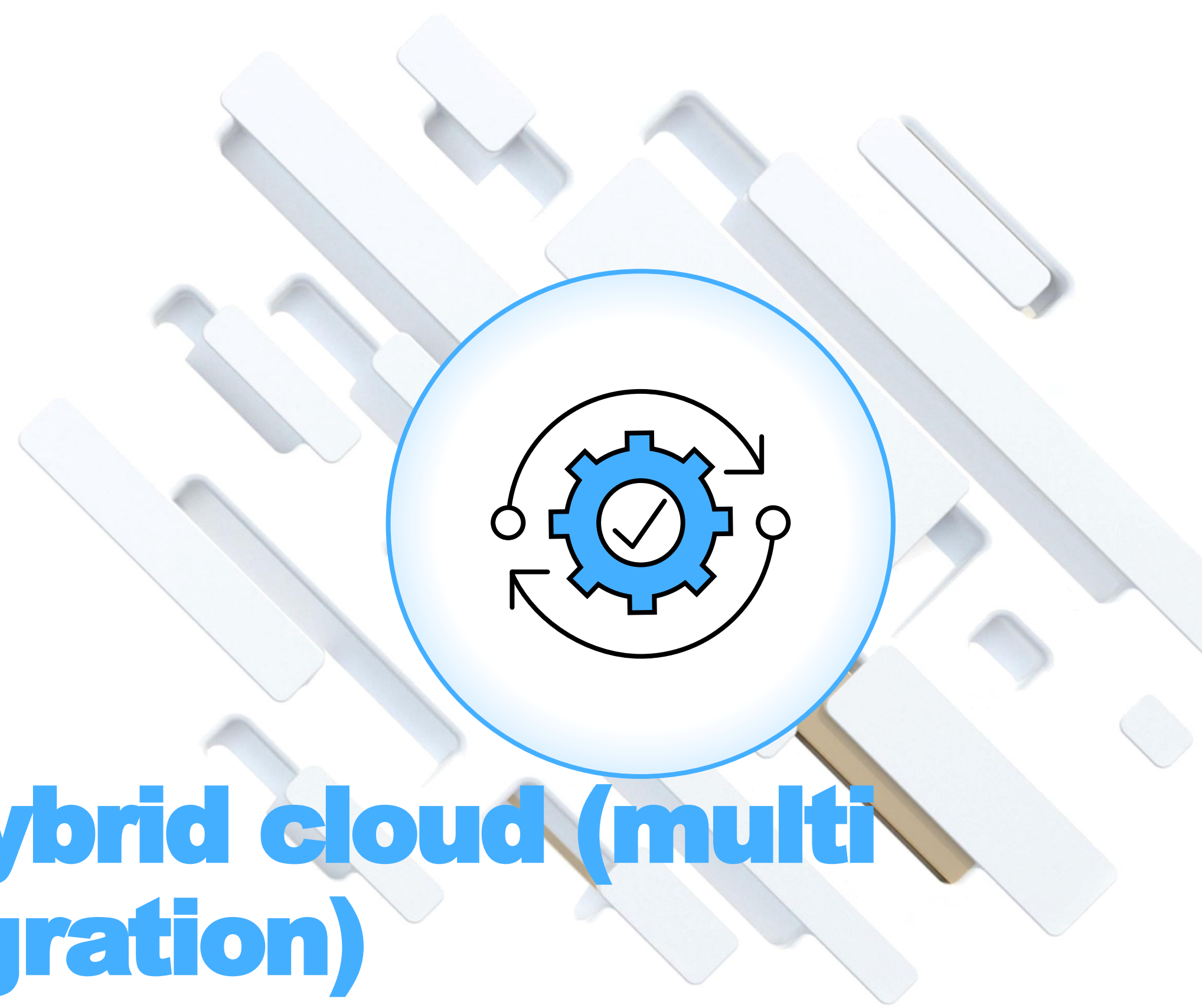
- Transparent backup, archive, data migration
- Snap-to-object allows burst to cloud
- Disaster recovery and business continuity baked into solution
- Tiering to more than one object storage solution increases durability



Storage Infrastructure Requirements in the New Era

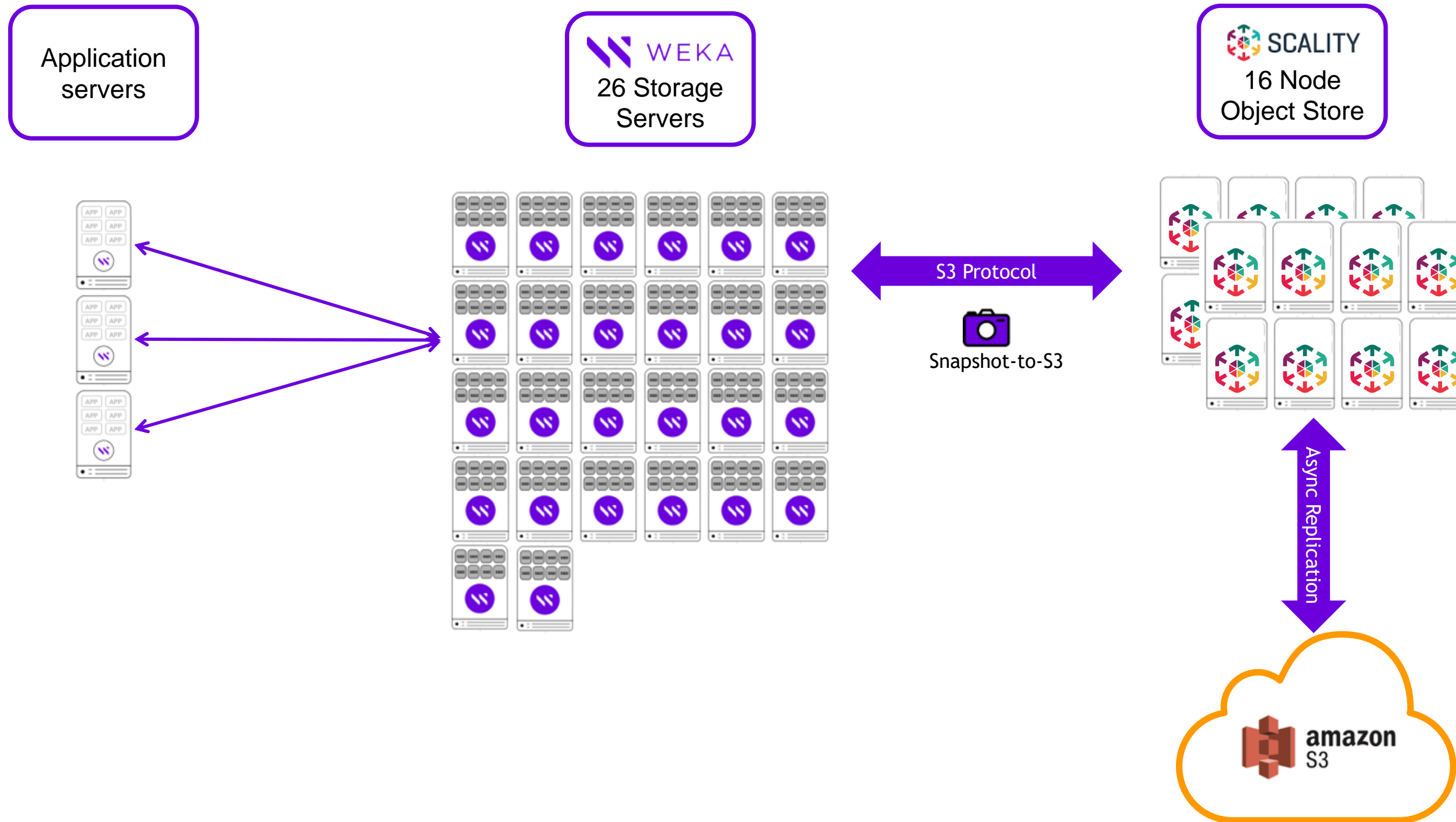


- Architected for hybrid cloud (multi public cloud integration)
- Multiple facets of performance (low latency, high throughput and bandwidth) with architectures optimized for solid-state storage
- Support for accelerated compute technologies (e.g. GPUs)
- Mission-critical high availability supported by flexible data protection and rapid recovery
- Extreme scalability - from tens of terabytes to multi-petabytes
- Ease of management, expansion and multi generational technology upgrades
- Infrastructure density for better efficiency



Architected for hybrid cloud (multi public cloud integration)

Pharmaceutical HPC Solution



One solution for many applications

Scientists never wait for technology

Integrated backup

AWS is now DR and BC

Drastic improvement in availability and performance

Technology is now an enabler

Weka Deployment at Genomics England

Single shared namespace to users and applications

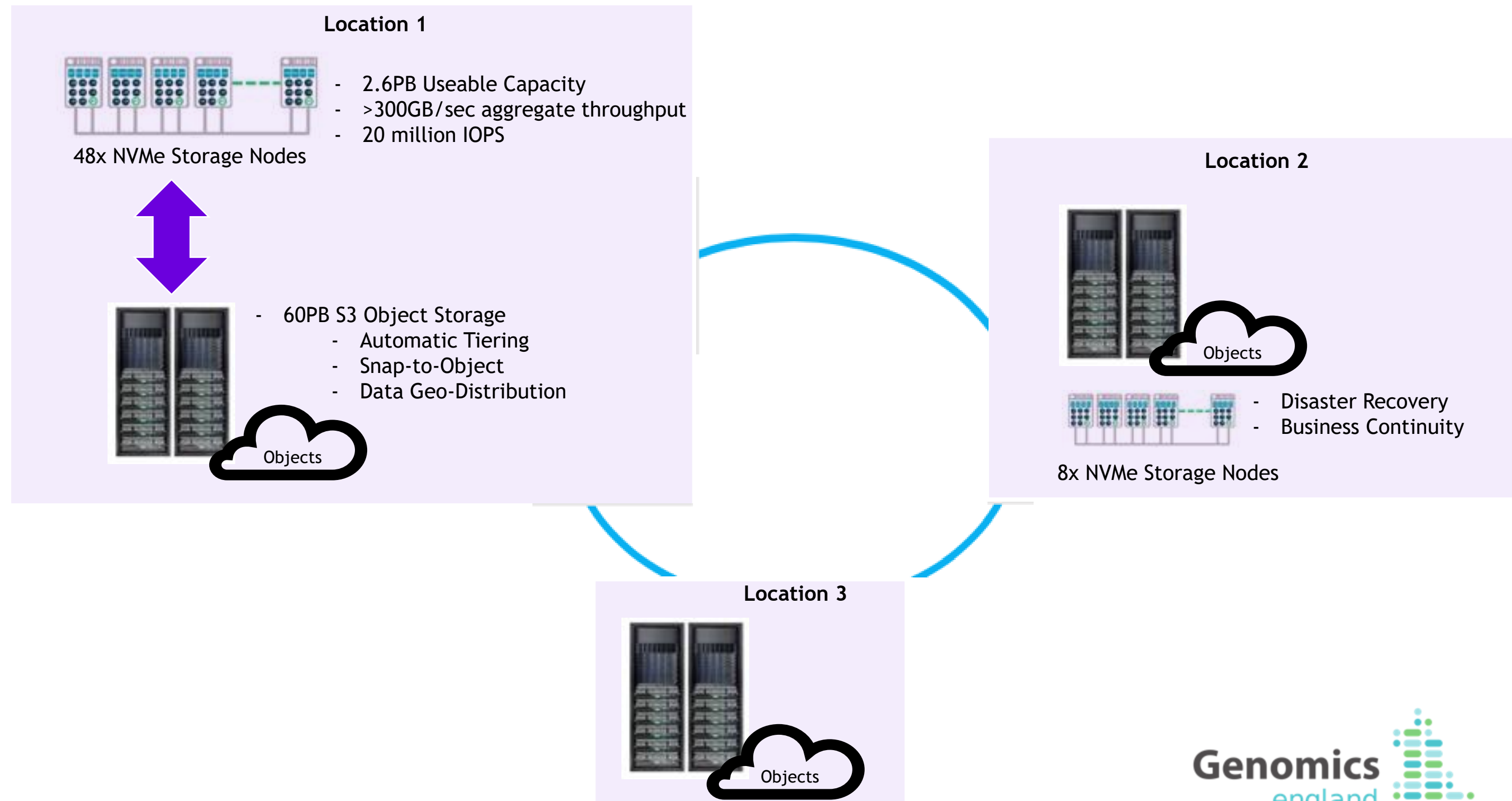
All data accessed via NVMe

Authenticated mounts, LDAP, active directory, extended ACLs, encryption in flight

Weka cluster at site 2 for DR and business continuity

NVMe tier will scale to 5PB & Object tier to 160PB within project lifetime

Over 2,500 Researchers access data for drug discovery & COVID research





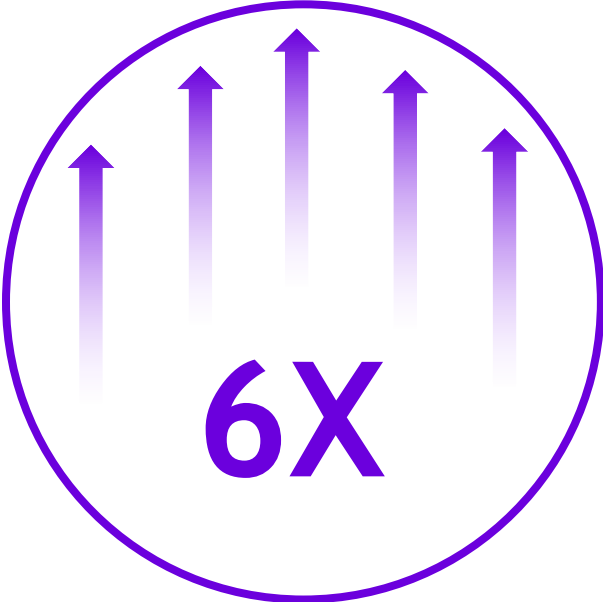
Multiple facets of performance

TGEN – BCL2FASTQ Conversion – Latency sensitive workload

6X Improvement in Pipeline Productivity

Before
WekaIO

Isilon



After
WekaIO

WekaFS - BCL2FASTQ Runs/Hour

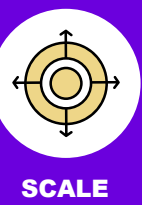
Weka WOW Case Study – Genomics DB



SIMPLICITY

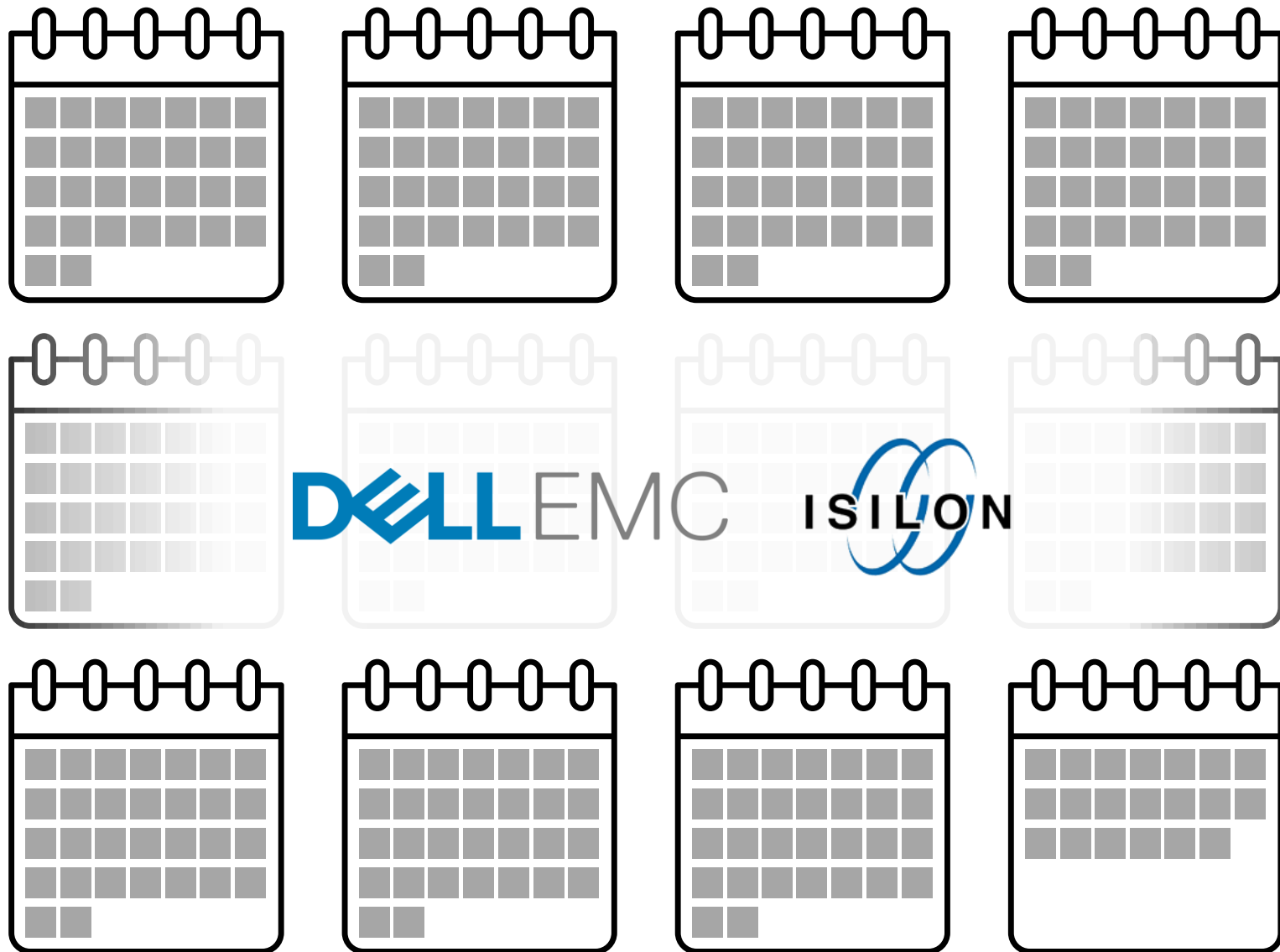


SPEED



SCALE

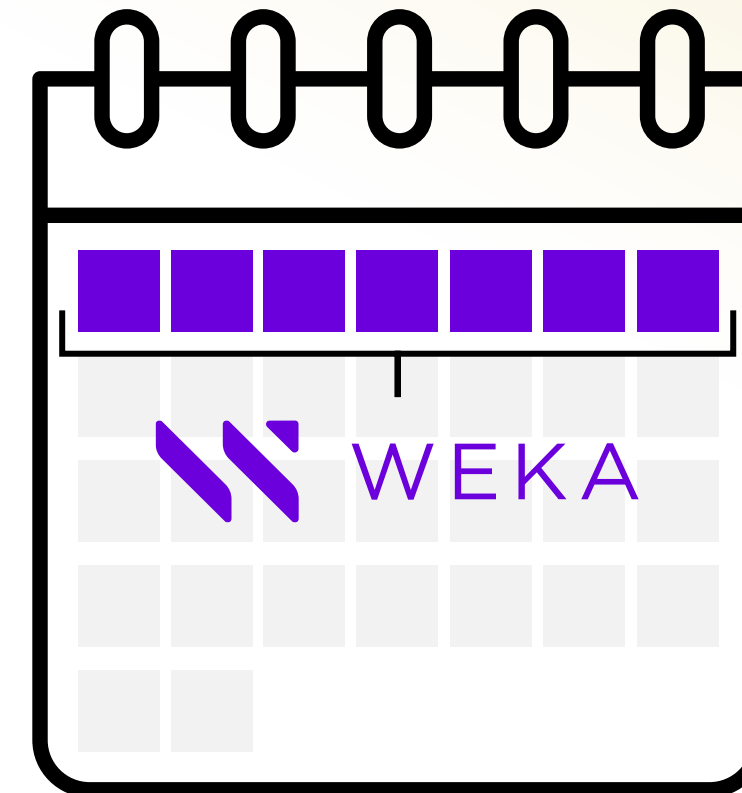
Before Weka



Compute workloads can take

70 DAYS +

After Weka



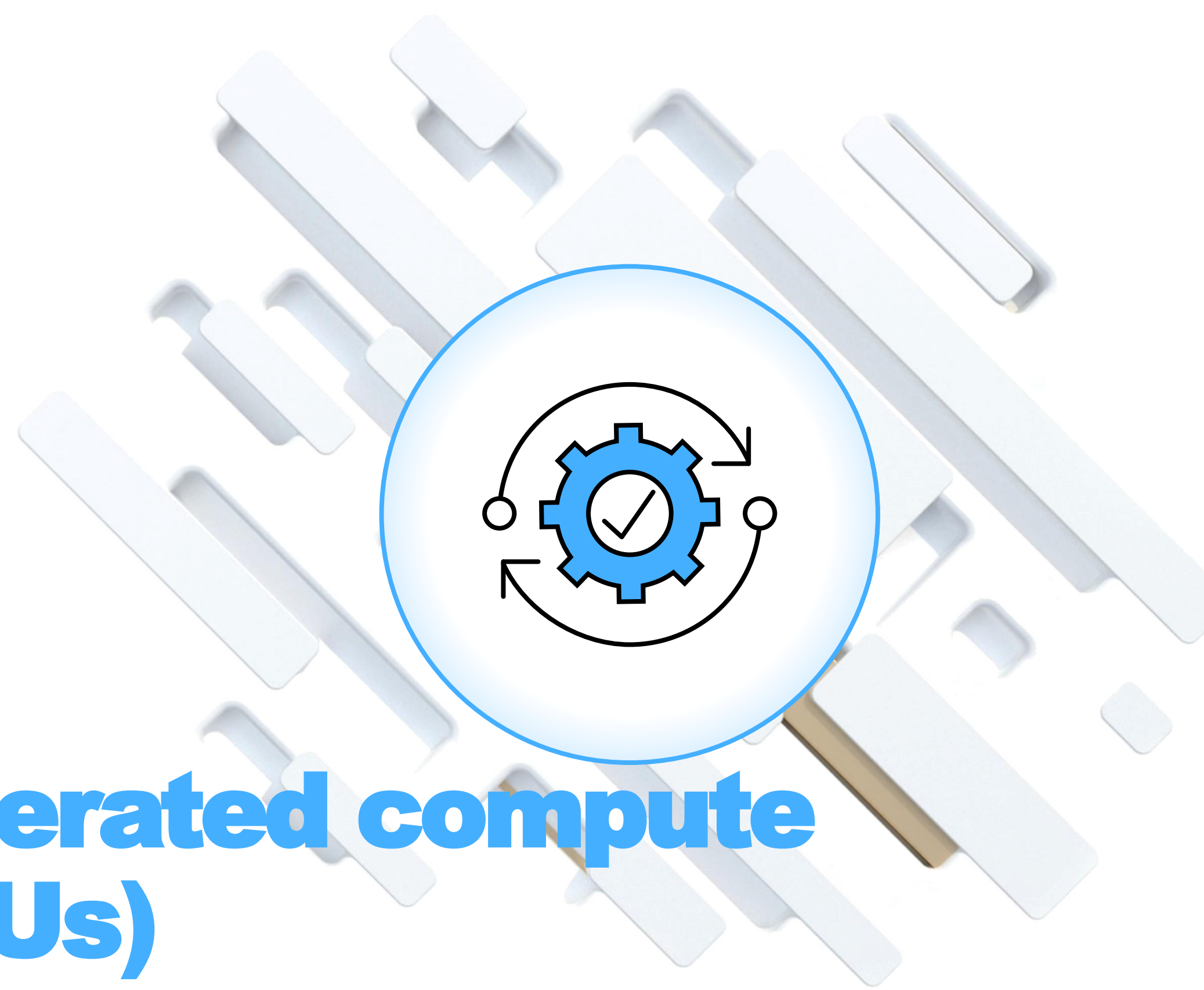
With Weka, compute workloads can take

ONLY 7 DAYS

50x improvement!

The jobs would run 70 days on Isilon and fail. Now they can aggregate thousands of genomes more efficiently and faster.

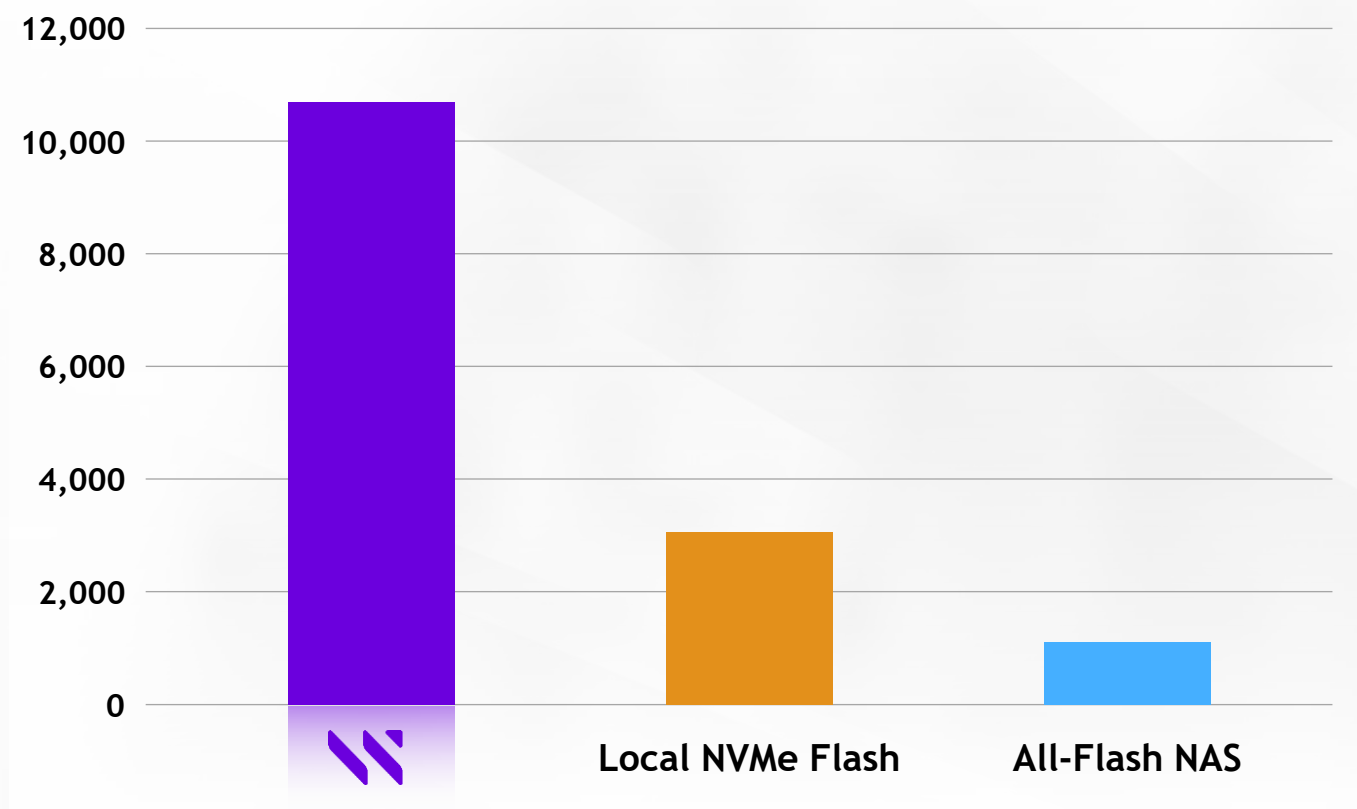




Support for accelerated compute technologies (GPUs)

INDUSTRY-LEADING PERFORMANCE

HIGH-SPEED NETWORK SATURATION



**FULLY SATURATE 100Gbe
NETWORK LINK**

MASSIVE SINGLE-CLIENT PERFORMANCE



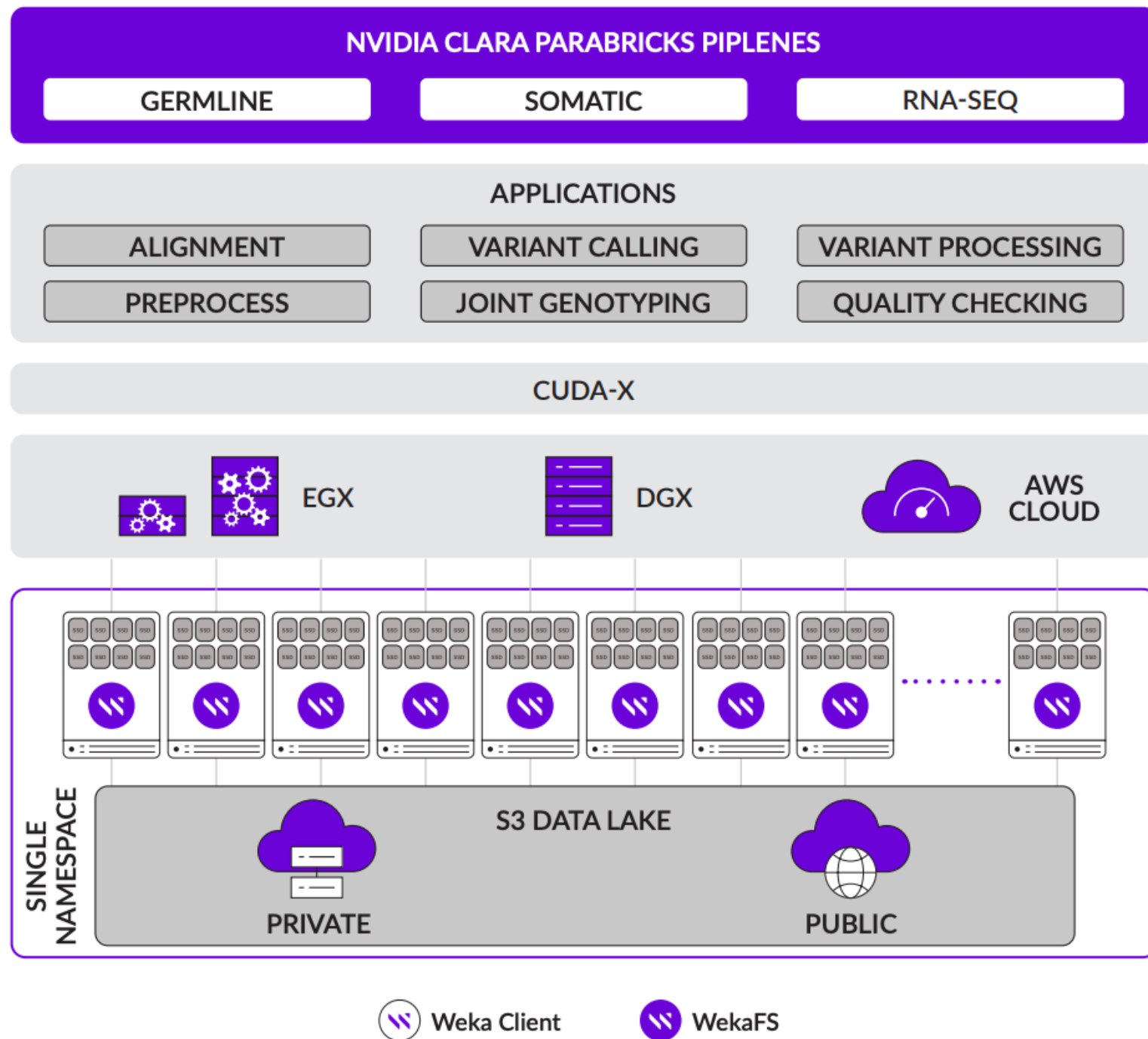
162 GB/S

1M IOPS

On NVIDIA DGX-A100

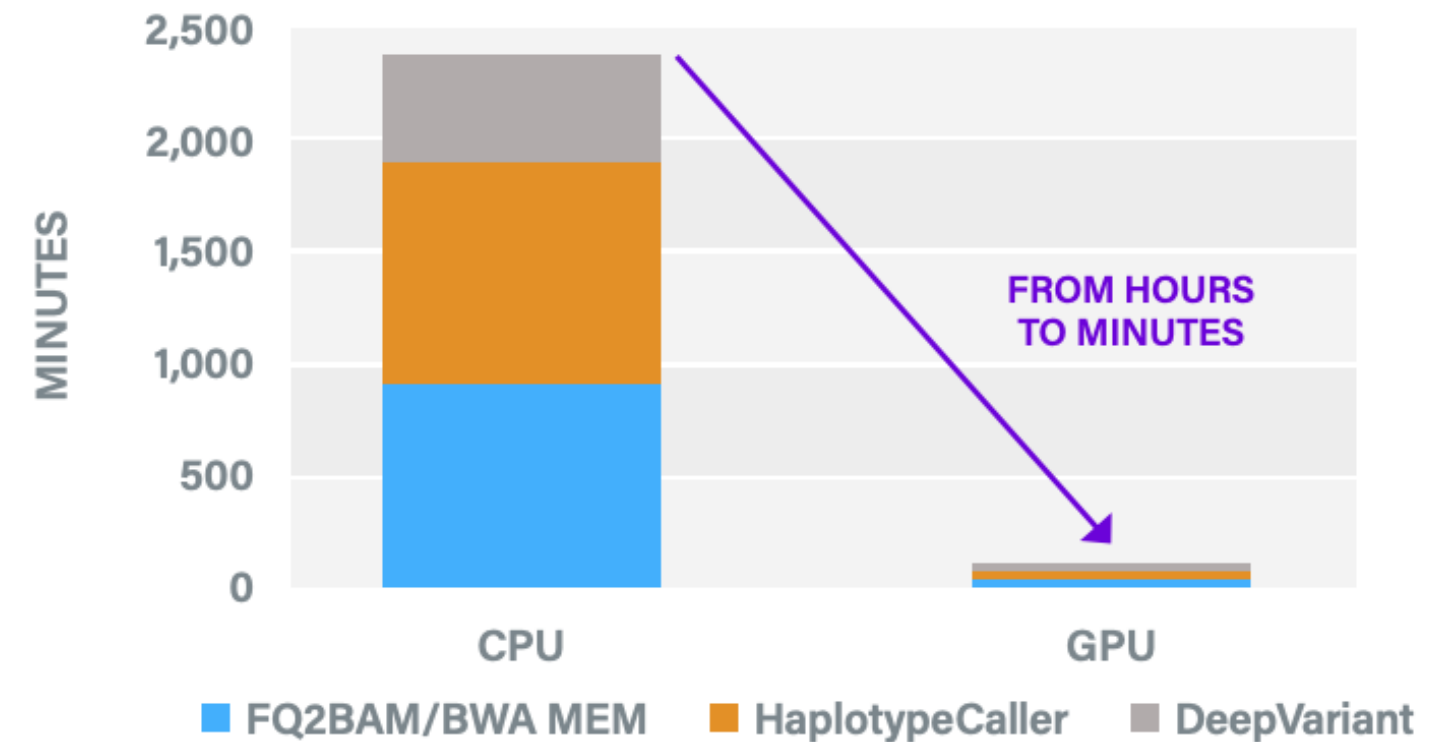
<https://www.weka.io/promo/nvidia-ai-reference-architecture/>

Genomics on GPUs – NVIDIA Clara and Parabricks



72x Speed Up in Genome Pipeline compared to CPU

WekaFS AVERAGE COMPLETION TIME CPU VS. GPU COMPUTE CLUSTER

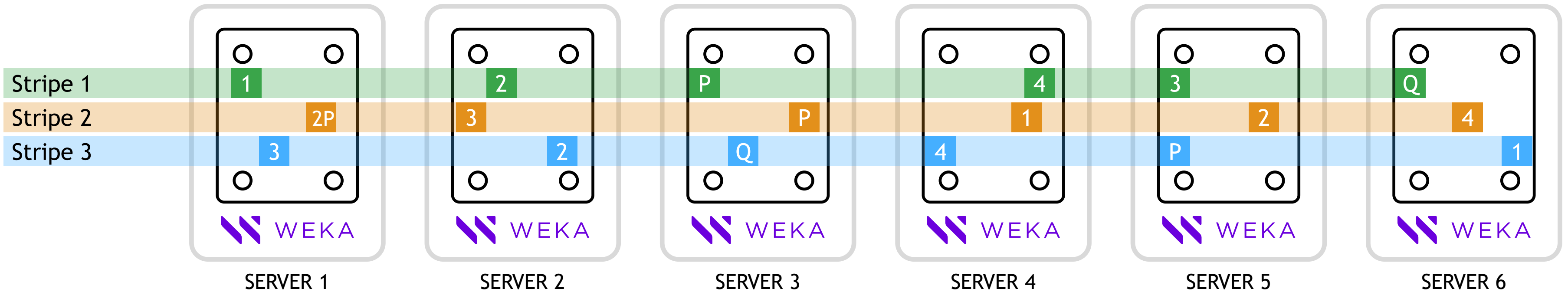


<https://www.weka.io/wp-content/uploads/files/2020/11/WekaFS-in-AWS-with-NVIDIA-Clara-Parabricks-for-Genomics-SB-W24SB20201143.pdf>



Mission-critical high availability

RESILIENCY WITHOUT PERFORMANCE IMPACT

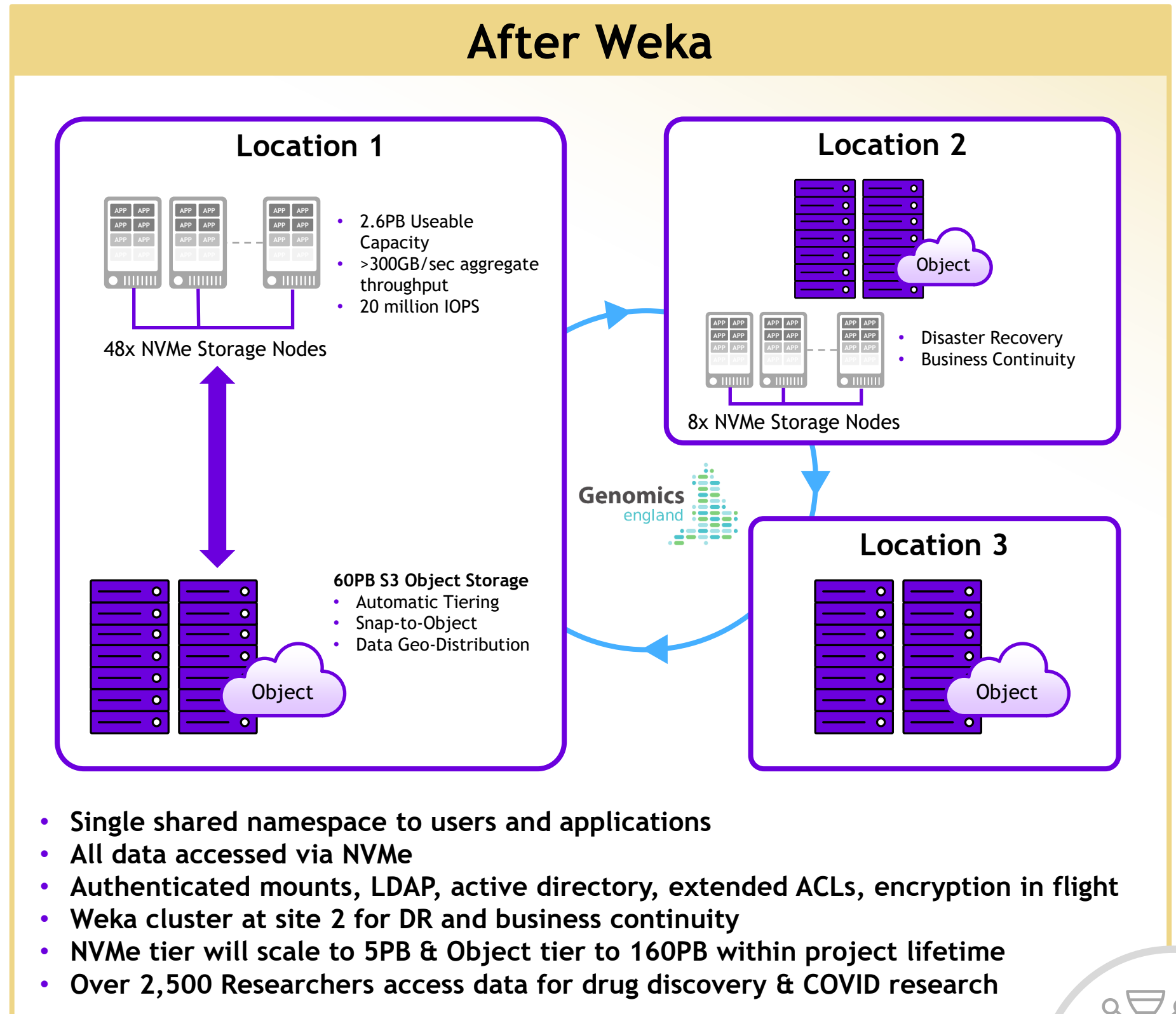
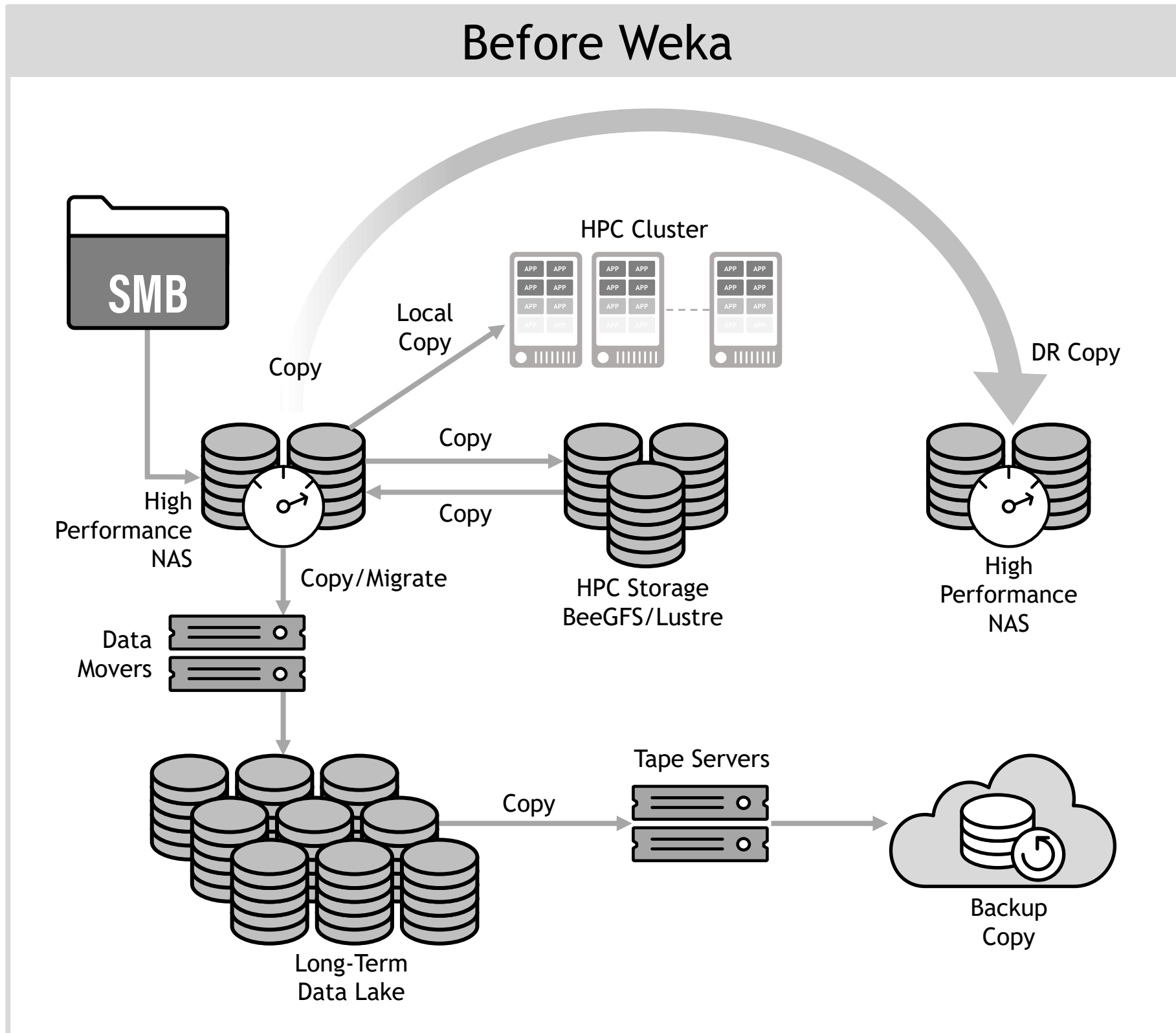


- Data protected across storage nodes
- No Performance impact during rebuilds
- Data protected at file level, so only need to rebuild small part of failed SSD/Server
- Smart rebuilds – dramatically faster than traditional RAID rebuilds
- Bigger the cluster, faster the rebuild

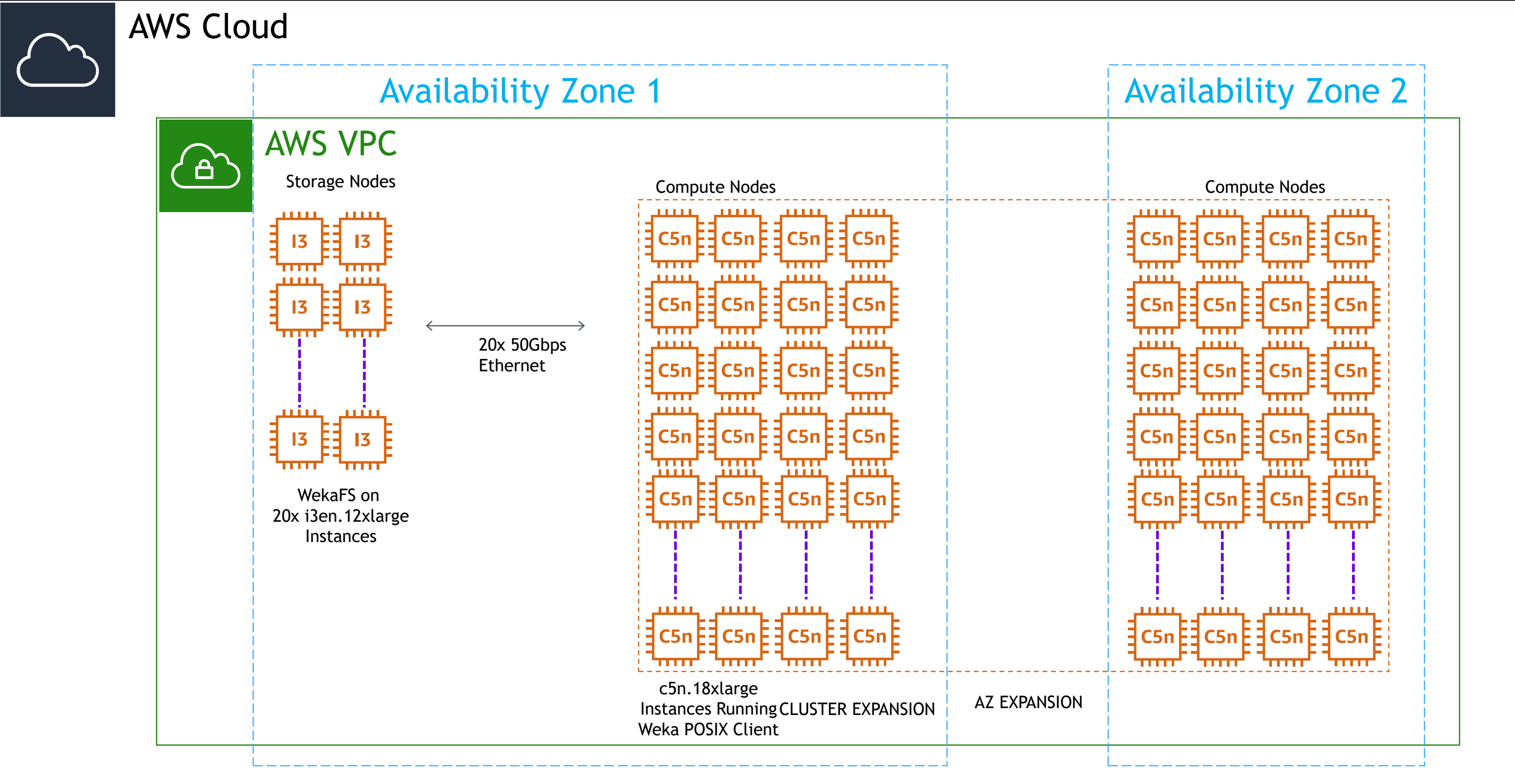


Extreme scalability - Petabytes

Genomics Environment with Legacy Solutions vs. Weka Deployment at Genomics England



Pharmaceutical in AWS - Elasticity





Ease of management – MGMT, Upgrades

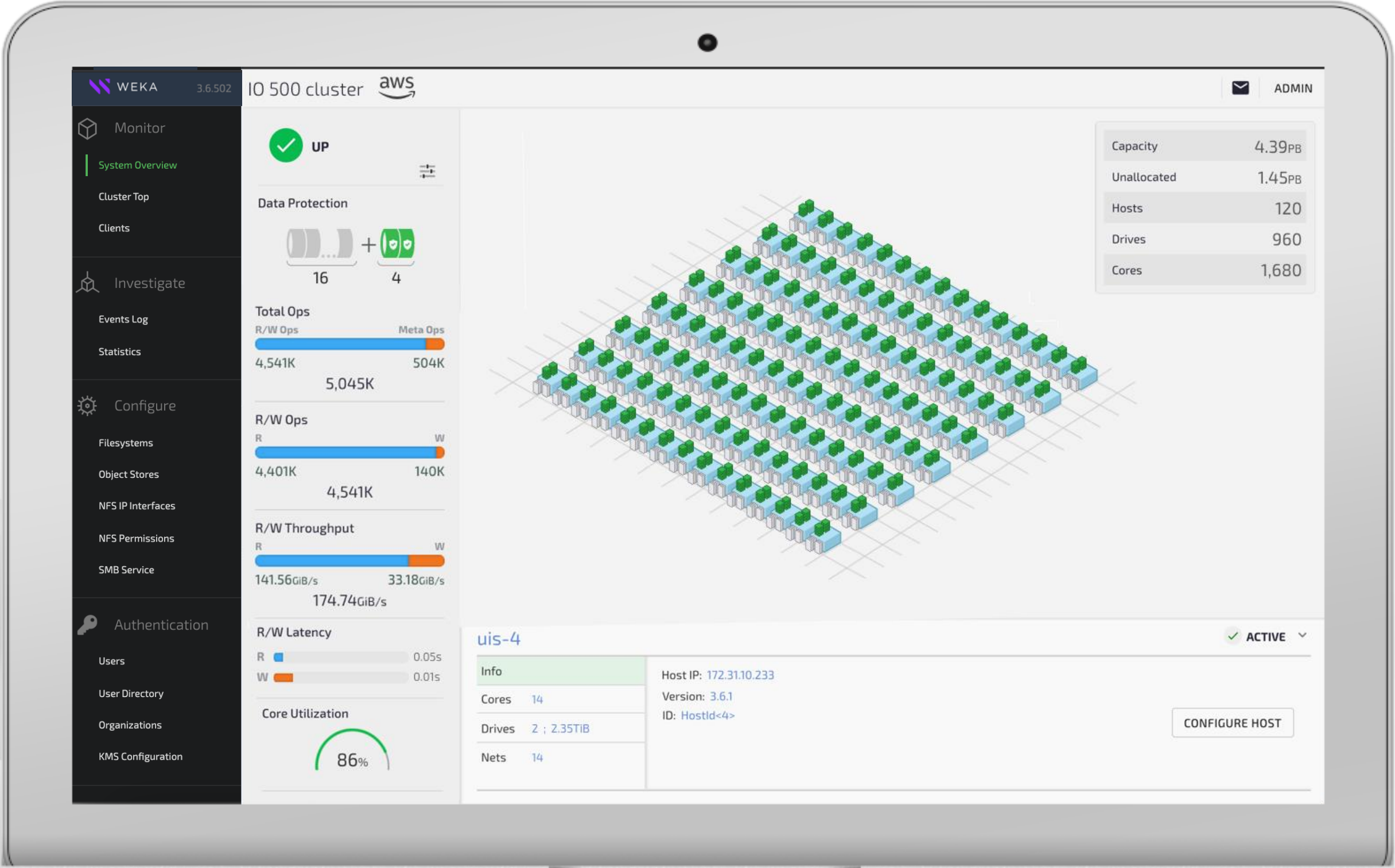


Weka can dramatically simplify the ease of use and management of our demanding customer's environments plus our recent testing reached over 200GB/sec. The ability to have persistent storage for Kubernetes and allow any protocol access expands the scale of our use cases that makes the Weka Data Platform a winner.

MEADOWGATE TECHNOLOGIES



INTUITIVE USER INTERFACE

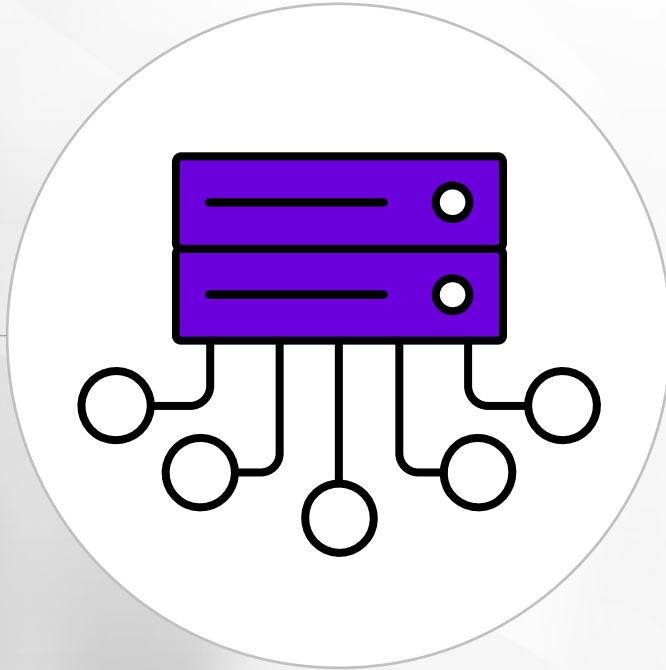


- 
- **Infrastructure density for better efficiency**

FLEXIBLE DEPLOYMENT MODELS



DEDICATED
STORAGE SERVERS



CONVERGED
COMPUTE/STORAGE



CLOUD

DEPLOY ON YOUR CHOICE OF HARDWARE



Hitachi Vantara



Lenovo



WEKA