

# Tame the Hybrid Cloud

## Increase Organization-Wide Flexibility and Scalability

### Challenges

- Inconsistencies between data center and cloud storage leads to creation of new data silos adding costs and complexity..
- Cloud storage performance limitations force some workloads to remain entirely on-premises.
- Lack of tools to orchestrate operations across data silos slows hybrid cloud workflows.

### Solution

- WEKA enables hybrid cloud workflows for demanding applications including video production, genomics processing, drug discovery, and autonomous vehicle development.

### Benefits

- Eliminate data silos between cloud and data centers.
- Obtain consistent performance and experience across on-premises and cloud deployments.
- Deliver a data environment that scales up and down independently for performance and capacity.
- Improve organizational agility and reduce time to results with seamless collaboration across hybrid cloud workflows.
- Enable 7x performance improvement versus traditional file storage offerings.
- Support the cloud of your choice: AWS, OCI, GCP, and Azure infrastructures.

### From Binary Decisions to Hybrid Possibilities

Fifteen years ago, many organizations believed the cloud would eventually become the sole solution for all their data and applications. However, this proved not to be the case, as security, performance, and cost concerns led organizations to determine that keeping some parts of their data in on-premises data centers was best.

Today, things have changed. Most workloads can be run anywhere, so organizations have the opportunity to choose where to run each application based on which location delivers the best performance, scalability, agility, security, and cost outcomes. Organizations have to decide, therefore, which workloads to put in the cloud and which to keep in on-premises data centers. The challenge today is not choosing between the cloud or on-premises but finding the optimal place to run each workload.

### Why Hybrid Cloud Might Not Deliver the Best of Both Worlds

The early faith that the cloud should be a cure-all solution has been replaced with a more nuanced approach in which organizations would try to optimize a mix of cloud and on-premises data centers. This has forced them to make significant trade-offs when it comes to data storage in a hybrid cloud environment.

In general, data centers rely on monolithic applications running on hardware that couples compute and storage resources in a fast, low-latency network. Thus, high performance is achievable in a data center. But this approach can be expensive and may limit scalability. In cloud architectures, applications are distributed across a pool of virtual compute and storage resources connected via shared physical networks. This approach offers massive scalability at a lower cost than a data center but often does not meet the performance requirements of some applications.

**KEY INSIGHT**

By eliminating the barriers between cloud and on-premises resources, organizations gain the ability to easily move data across platforms and achieve new levels of creativity and success.

The data storage solutions available to organizations typically support one architecture or the other. In the cloud, there are simple, scalable services like Amazon S3. In data centers, expensive options such as NetApp or Pure Storage provide greater speed capabilities. These options rely on different—and incompatible—storage formats, forcing organizations to make binary choices between cloud or data center. This scenario is a primary driver of data silo creation.

The spillover caused by data silos can cascade in ways that significantly impact organizations. The disconnect and isolation among data and applications create inefficiencies that hinder organizations from extracting insights from their data. The lack of data integration across platforms increases costs, decreases agility, and limits scalability. Organizations find themselves hindered from using techniques such as cloud bursting, leading to slow processing times, decreased productivity, and missed opportunities.

Inadequate data management can further restrict data-driven decision-making and an organization's ability to adapt its IT environment to business needs. With few tools available to integrate and scale operations between on-premises and cloud workloads, matching the storage environment to changing business requirements is difficult. Gaps between on-premises and cloud storage silos create an environment that is difficult to manage. IT teams are stuck employing manual processes and awkward custom tools that inflict a huge burden on their resources and skill sets. Achieving consistent data protection and maintaining resilience is challenging, at best.

It doesn't have to be this way. By breaking down silos and eliminating the barriers between the cloud and on-premises resources, as well as between different cloud platforms, organizations can access new levels of creativity and success by having the ability to move data across platforms easily. This hybrid cloud approach promises organizations the performance and cost-effectiveness they've long desired, along with the scalability and agility of the cloud.

## Overcome On-premises Limitations with Cloud Bursting

Cloud bursting is a compelling solution to the problem of limited on-premises computing power, offering organizations the promise of leveraging cloud-based resources to augment their on-premises data processing and analysis infrastructure.

Cloud bursting is a configuration method that uses cloud computing resources whenever on-premises infrastructure reaches peak capacity. When organizations run out of computing resources in their internal data center, they burst the extra workload to external third-party cloud services. For example, the organization may collect data in one location, but that facility doesn't have the computing power to process or analyze it. With cloud bursting, the organization can use the cloud to access the additional resources needed quickly.

However, cloud bursting can also create new challenges. One difficulty of cloud bursting is that the storage used in the data center is usually incompatible with that used in the cloud. So to do cloud bursting effectively, organizations must copy data across and maintain consistency between the two locations. Organizations often have to deploy expensive and complex data replication tools to do this. The second big challenge is that cloud storage typically delivers lower performance than data center storage, so jobs take longer or can't run. In an attempt to solve this, organizations over-provision storage and networking resources to get more performance, but this also adds costs and complexity.

## Support Distributed Teams with Efficient Hybrid Cloud Workflows

Hybrid cloud workflows aim to support distributed teams working on data-intensive projects. For example, in drug discovery, scientists may collect data in the lab while others analyze it in the cloud. The drug discovery pipeline is a multi-step process typically involving six to seven stages, each with unique performance requirements. Data ingestion demands massive concurrency and storage throughput, while exposure processing necessitates massive read throughput for search and indexing. 2D/3D classification is data throughput-intensive, 3D image refinement requires many concurrent streams, and the creation of the final structure requires high write performance.

Unfortunately, using traditional storage platforms to meet these diverse requirements has not been feasible. Researchers have attempted to address this challenge in two ways. First, they create separate copies of the data for each data pipeline stage, which significantly increases costs. Alternatively, researchers can optimize a single storage system for only one step in the pipeline, but this approach can delay the data analysis phase by weeks or months.

A hybrid cloud approach can provide a solution by offering high-performance storage on-premises and massively scalable storage in the cloud to support different stages in the data pipeline. However, the fundamental problem of data silos remains. Therefore, organizations must consider a hybrid cloud strategy that addresses data silos while supporting the unique performance requirements of each pipeline phase.

**KEY INSIGHT**

WEKA enables organizations to avoid the trade-off between cloud economics and performance by breaking down data silos to unlock new levels of innovation and flexibility in a hybrid cloud environment.

## **Create an Accessible Solution to Long-Term Data Retention**

The goal of long-term data retention is to store large amounts of data cost-effectively without worrying about the data becoming inaccessible in the event of a disaster. However, traditional solutions like Iron Mountain can be expensive and create disconnected data silos that add complexity to the data pipeline. By moving to the cloud, organizations can store large quantities of data, but these, too, can become separate data silos. This situation creates the need for multiple tools and steps to access this data, leaving organizations with no way of understanding what data is stored in these silos.

## **Deploy Effortless On-premises to Cloud Backup and Recovery**

Data protection aims to provide an organization with a reliable backup and disaster recovery (DR) solution. Initially, organizations turned to object storage offerings such as Amazon S3 or Azure Blob Storage due to their low cost, highly accessible nature, and ability to work independently of the customer's data center infrastructure. However, this created another data silo and additional, inconsistent copies of the data used elsewhere in the organization, requiring different special tools to access them on premises.

Organizations are now seeking a solution that provides backup and recovery and allows them to extract more value from their DR strategy without compromising resilience. Forward-looking organizations realize they can use the data in their DR site for other value-add purposes such as cloud bursting, test and development, and analytics. For example, once data is copied to a secondary cloud target, it should be easily mounted into a usable format that cloud-based applications could access for non-production purposes. Organizations could use this site to analyze data for customer behavior patterns or feature use. Log analytics is a significant first early use case to improve overall application performance. Another possibility is to use the DR copy as part of a comprehensive migration strategy.

## Maximize Cost Savings, Scalability, and Flexibility with WEKA's Data Platform for Hybrid Cloud

The WEKA Data Platform is a cloud-native, software-defined storage solution for next-generation workloads like artificial intelligence and machine learning that supports large-scale data collection and processing on high-performance computing infrastructures. The platform provides a unified experience, allowing organizations to configure storage to meet their specific requirements without forcing a choice between cloud and on-premises solutions. It offers performance, agility, and scalability, no matter where the workload runs, enabling organizations to deliver applications and data where they provide the best experience for their customers. With WEKA, organizations avoid the trade-off between cloud economics and performance by breaking down data silos to unlock new levels of innovation and flexibility in a hybrid cloud environment.

### Streamline Workflows and Enhance Data Management to Work at Exabyte Scale

WEKA enables the creation of a single namespace between on-premises and cloud storage tiers so organizations can leverage both on-premises and cloud storage resources as needed without compromising their workflows. Using WEKA Snap-To-Object makes it possible to create a fully usable and consistent copy of an organization's on-premises data in any cloud. Because the on-premises WEKA cluster and cloud-based object storage are part of a single storage namespace, storage operations for data retention, back-up, and disaster recovery needs are much simpler. With WEKA, organizations can store just a single copy of data in the location that best fits the requirements of each workload. They can also extend their existing workflows without switching between different systems or platforms.

### Effortless Cloud Bursting and Optimized Cost Control

WEKA enables organizations to mount incremental file system snapshots to any cloud of their choice, providing ready access to the resources needed to run applications or perform data analysis in the cloud. Advanced features like cloud and data center application autoscaling, multi-protocol support, and a distributed architecture make it possible for organizations to rapidly scale their environment to meet peak demand. Once the cloud bursting operation is complete, the data can be archived or deleted, after which it is possible to scale down the cloud environment to optimize costs. This feature of WEKA allows organizations to take advantage of the scalability and flexibility of the cloud while maintaining control over costs and resources.

## Unified Data Management, Effortless Data Portability, and Streamlined Backup and DR

WEKA provides a unified data platform built on a single code base that operates the same no matter where your data is located. This eliminates the need for multiple management tools, providing a simple and consistent data management solution for every application in the cloud and the data center. With WEKA's Snap-To-Object capability, organizations can quickly create a usable copy of their file system to any object stored in any cloud or on-premises, ensuring backup and disaster recovery are simple and seamless. The same capability also enables fast and easy data portability between locations.

WEKA enables organizations to back up their on-premises data to the cloud easily, ensuring it is protected and secure. With WEKA, businesses can recover data in the cloud or back up to their on-premises infrastructure, providing them with a flexible and reliable disaster recovery solution. The platform also allows organizations to run applications in the cloud using their DR snapshots for recovery, cloud bursting, or cloud migration. The backup and DR capabilities offered by WEKA are available at no added cost, making it a cost-effective solution for organizations looking to protect their data.

WEKA enables long-term data retention by allowing organizations to take incremental snapshots of their entire file systems and archive them to any object store, on premises, or in the cloud. This cost-effective solution eliminates expenses associated with Capex and allows for cold storage of large quantities of data without having to differentiate what is important and what isn't. By linking the stored data to the rest of the data pipeline, WEKA also eliminates disconnected data silos, simplifying the process of data retention while reducing costs.

## About WEKA

WEKA addresses the complex and diverse requirements of modern hybrid cloud environments. With WEKA, organizations can optimize data storage costs, get cloud scalability for every workload, enjoy consistent and straightforward data management for every application, and improve data resilience and portability.

### For More Information or to Arrange a Free Trial

Visit us at <https://www.weka.io/get-started> or email us at [info@weka.io](mailto:info@weka.io).



[weka.io](https://www.weka.io)

844.392.0665

