More to Store with a Global Data Platform

Highlights

- Eliminate duplicate data
- Improve application agility with Kubernetes integration
- Speed AI results and accuracy
- Consolidate global data on a single storage system
- Developers can choose the interface to access any data
- Optimized data resources for performance and costs
- Connect multiple data sources into a single data lake
- Create an open ecosystem and avoid vendor lock-in
- Quickly expand storage as needed
- High performance options from edge to core to cloud

Our story begins with a race. The race to become agile and for organizations to adjust the constantly changing business requirements that they face. CIOs are constantly looking for ways to improve IT and digital assets to help achieve business goals. There are many reasons that CIOs pursue infrastructure modernization. The top reasons are agility and helping the organization speed it's time to market with new applications or by pursing new opportunities. Agile organizations can quickly redirect their resources and priorities toward value-creating opportunities. Quality of Service or reliability and cost efficiencies follow closely as the 2nd and 3rd reasons for pursuing infrastructure modernization.

Organizations Need to Unlock Value From Data...
Wherever it Resides

"We fundamentally believe that core to the competitiveness of every company going forward will be their ability to use AI to unlock real-time value from their data wherever the data resides.

Organizations need to unlock the value from their data. As Arvind Krishna, CEO of IBM has so well stated on the Q2 2021 Earnings call "We fundamentally believe that core to the competitiveness of every company going forward will be their ability to use AI to unlock real- time value from their data wherever the data resides."

The Challenges Organizations Face

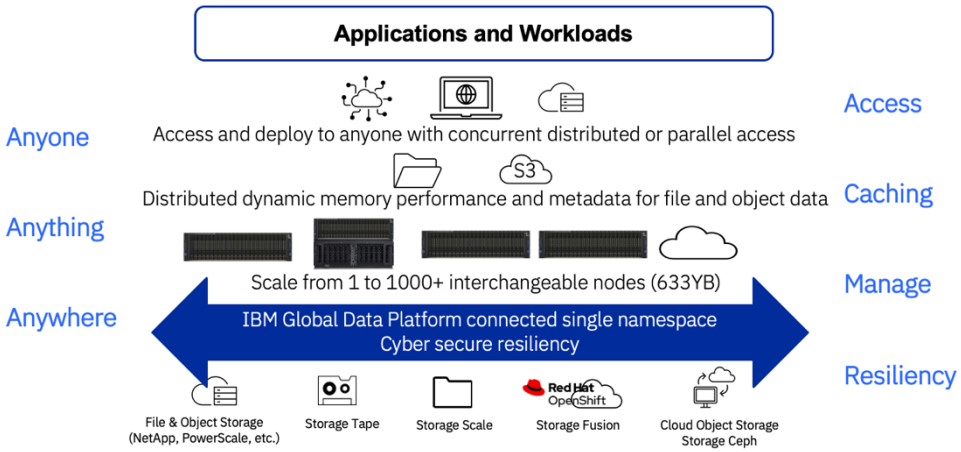
As customers strive to create smarter ways to work and drive digital transformation throughout the enterprise there are multiple challenges that are limiting the value that IT can deliver to organizations. Enterprises are restricted by data silos as data is everywhere making data access very difficult throughout the organization. Many digital modernization efforts are failing to meet objectives due to cost overruns and performance limitations. As many businesses have experienced or will experience a cyber-attack the focus on data security, availability and data governance has increased. Finally, as organizations are faced with managing their growing capacity of file and object data there is a need to automate and enhance efficiencies so that managing all their data does not hinder the business results that are desired.

Current infrastructure is blocked from unlocking the all the value that is hidden away in silos of data.

- Data everywhere but not connected so its difficult to access
- Costs and performance are not leveraged across the organization
- Security, availability, and governance are managed by each application loosing economies of scale for the organization
- Manageability is hard and limited by each workload

Organizations now require an infrastructure for their data that breaks through the challenges they are facing and enables more business benefits with more innovation, business resilience and efficiency for data intensive and AI applications. What is needed is more of a global view of data and what we call a global data platform. This global data platform needs to break down the barriers and silos so the organization can Speed innovation with flexibility and agility so the business can move quickly to adapt to changing requirements and access data that is connected and not soloed. The Global Data Platform needs to strengthen business resilience so the business can continue safely and maintain stability with its infrastructure. Finally the Global Data Platform needs to increase efficiency to enable new workloads to maintain high performance and the growing amounts of data does not overwhelm budgets and storage becomes affordable.

A Global Data Platform that provides More in Store

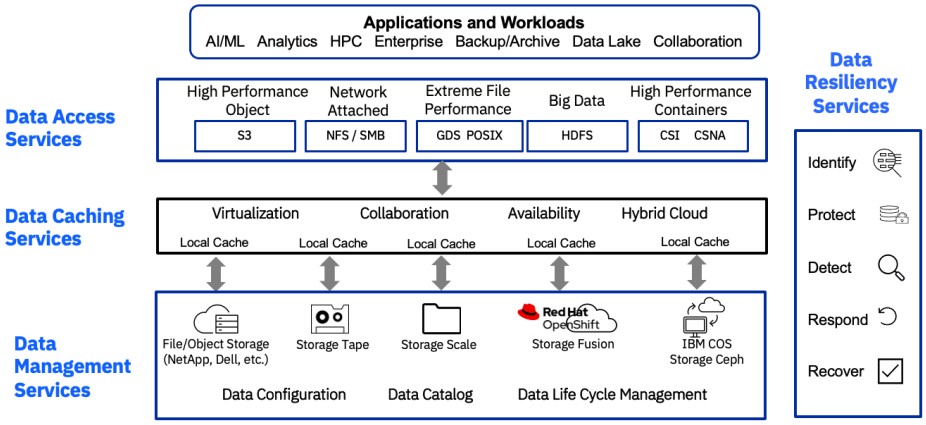


We now introduce IBM's Global Data Platform. engineered for anyone to access anything anywhere. Customers can access and deploy to anyone, no matter where they are located on the edge, in the data center or in the cloud. They can also access anything anywhere with concurrent distributed or parallel access to data. Data is distributed with dynamic memory performance and metadata for file and object data. Finally the IBM Global Data Platform creates a connected single namespace maintaining cyber secure resiliency so critical data can be protected when the unexpected happens. IBM uniquely organizes our platform into four catagoes of data services: Access Services, Caching or Core Services, Management Services, and Resiliency Services.

IBM Global Data Platform delivers More in Store

- More performance
  - IBM Storage Scale provides Ultra performance with TB/s and Millions of IOPs with dynamic optimization and direct connect parallel paths that support multi-protocol access to a single source of truth.
- More affordability
  - IBM Storage Scale provides energy and cost savings with computational storage and up to 380% ROI with data that turns off and on as needed.
- More resiliency
  - Multi-site protection and cyber security automation with faster recovery and air gap using Safeguarded Copy and CyberVault services
- More agility with connectivity
  - Break down silos creating a global single source of truth by connecting data across sites and in the cloud
- More efficiency
  - Extend benefits by connecting storage systems and reusing resources without rearchitecting infrastructure

A look inside IBM's Global Data Platform



IBM's Global data platform creates a true distributed and global parallel file system. IBM's data and AI solution creates an information supply chain for your global data that not only connects IBM storage but also connections non-IBM storage as a single cluster of data. IBM Storage Scale is the controller that connects all the storage together and accelerates data with parallel access and highest performance to the data.

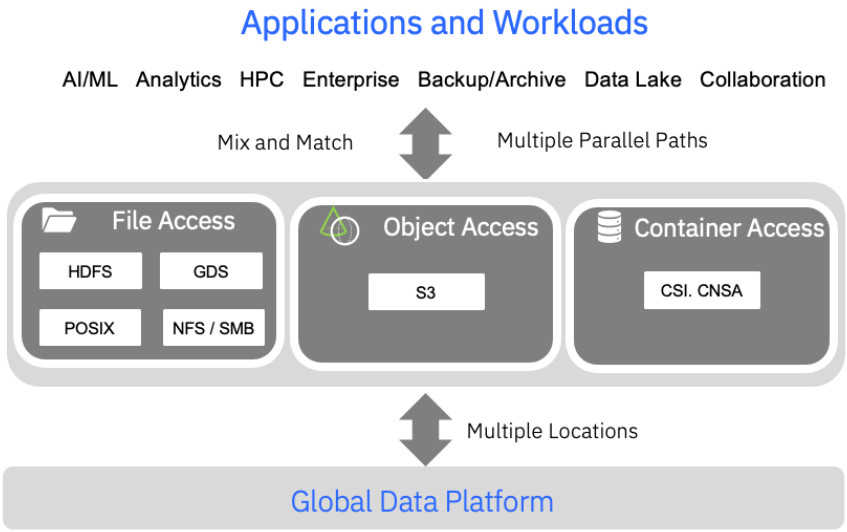
Starting at the top of the chart, a global data platform must "speak the language" of the applications and use cases. Data access services must be "multi-lingual" meaning some applications will create and access data with a certain protocol, and others may require access to the same data with a different protocol. It's also critical that the data access services provide shared access to file and object data. For example, if data is ingested with file protocols, it must allow for other use cases to access that same data with other protocols – such as S3/object.

Looking at the middle of this chart, a global data platform must provide data access independent from where the data resides - without creating copies of the data. For example, when cloud-native applications require high performance access to data stored in an S3 bucket, a global data platform transparently executes "vertical data caching services" and automatically fetches the data from its home location. Likewise, for cloud-bursting or data collaboration use cases, the global data platform transparently executes "horizontal data caching services", making data available to public cloud infrastructure or other locations. Again – without copying the data and acting as a local cache.

Moving to the bottom of the chart, a global data platform must provide visibility, control and automation facilitating data orchestration and policy-driven data life cycle management. With the various IT infrastructure options, mentioned earlier, finding the correct data for business analytics, machine learning and AI use cases can become extremely challenging. IBM's solution provides not just the visibility to the single source of truth, but also has the capability to ensure data is automatically moved to the most effective storage tier based on an organization's policies such as cost, performance, retention, or data access.

Finally, the right side of the chart highlights data resiliency services that provide data protection and effective detection and prevention of cyber security attacks and near instant data access to critical data in the event of a successful attack.

Data Access Services

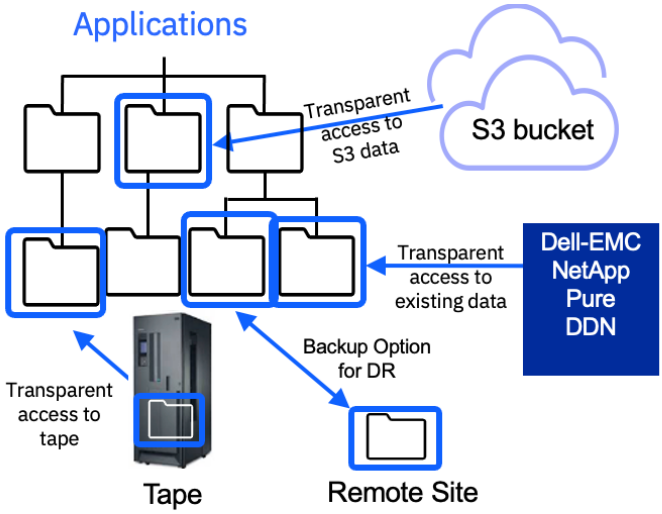


- Drive user productivity and application performance
- Eliminate silos and lower costs

Organizations have a variety of use cases, workloads, and applications requiring access to both file and object data. With its global data platform, IBM lets the application define the necessary access protocol and provides a unified storage solution to meet these data access requirements. Organizations can run traditional workloads like backup and archive and Enterprise IT applications, or perhaps they need to run data analytics, or high-performance simulation workloads. Maybe they want to take advantage of modern and containerized applications. IBM's global data platform supports all of these workloads, and each application can dictate whichever protocol it requires to access the data. For example, an Enterprise IT application might require NFS access and a machine learning application requiring high-performance needs POSIX or S3 access. With point solutions for file and object storage, they cannot easily access the same file or object data sets. IBM's global data platform solves this by allowing applications to write data with file or object protocols and other applications can access the same data using either file or object protocols. It provides applications access to the same data, at the same time, no matter which protocol the application speaks. These services is what connects the application to the data and provides the performance and storage resources that can be optimized, secured and connected with the other data services. The data access services are unique to many other file and object solutions in the way that one can create multiple paths to the same data and scale the capacity to 633 Yotta Bytes and scale performance to TB/s throughput and 100s

of millions of IOPs with low latency access. IBM Storage Scale supports multiple data types and multiple API interfaces including POSIX, S3, HDFS, NVIDIA GPU Direct Storage (GDS), NFS/BMB, Container Native and Container CSI interfaces. Multiple interfaces provides the flexibility and the consolidation advantages, but the application performance is achieved with the parallel data access to each storage node.

Data Caching Services



- Lower costs with multi-vendor and multi-cloud support
- Quickly scale your data with resources you choose and performance you require

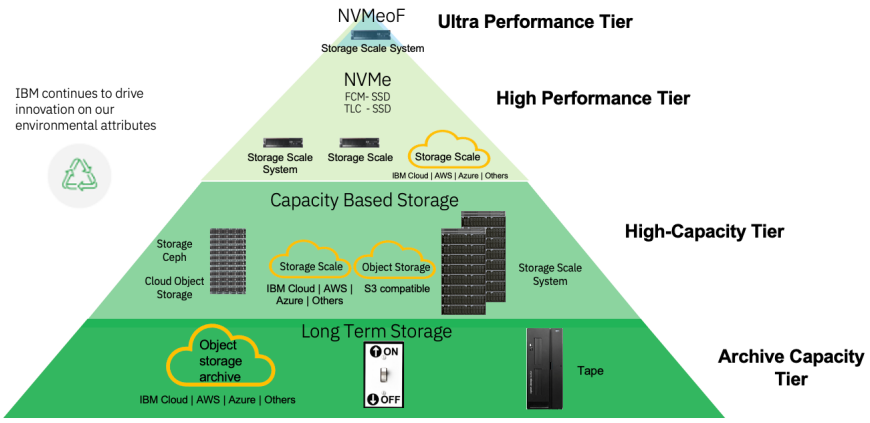
Having multiple interfaces to the same data gets even more powerful when that data can be accessed from multiple data sources that each may have a unique interface. We maximize resources and connect to what your applications need… data.

- Turns remote file and object data into active capacity (open ecosystem)
- Masks wide-area network latencies and outages by transparently caching data locally
- Individual files in the file set can be compressed

Multiple sources provide **investment protection** with an open ecosystem of storage options leveraging multi-vendor and multi-cloud resources. We also provide **increase application agility** from edge to core to cloud by bringing more data to applications wherever they are deployed. Multiple data sources allow you to **quickly scale your data** from resources you choose with performance you require

- Investment protection with an open ecosystem of storage options leveraging multi-vendor and multi-cloud resources
- Increase application agility accessing data from edge to core to cloud by bringing more data to applications wherever they are deployed
- Quickly scale your data from resources you choose with performance you require
- Faster access to remote data by transparently caching remote data locally when needed

Data Management Services



Moving on to data management services, the global data platform provides integrated tools for visibility, control, and automation, facilitating data orchestration and policy-driven data life cycle management. IBM's solution provides not just the visibility to the single source of truth, but also has the capability to ensure data is automatically moved to the most effective storage tier based on an organization's policies.

Data Cataloging and Orchestration capabilities allow users to benefit from having a single copy of the data for all their workloads. Customers are now able to optimize performance and lower storage costs through the ability to pre-fetch or fetch data on-demand as needed into the global data platform. Additional features, such as metadata augmentation and automated data enrichment for contextualizing data with semantics and knowledge help customers optimize time to results for multiple workloads.

Information life cycle management services include a highly flexible policy engine that allows customer to define rules for optimizing the storage of file and object data. These services transparently move data to the appropriate tier of storage, optimizing both cost and performance based on an organization's policies. These policies can also be utilized to support the organization's retention and archiving requirements.

Infrastructure management is a critical aspect of petabyte scale file and object storage solutions. Simplified methods for deployment and management in the public cloud, at edge or remote locations and in traditional data centers are included with IBM Storage Scale software and IBM Storage Scale System.

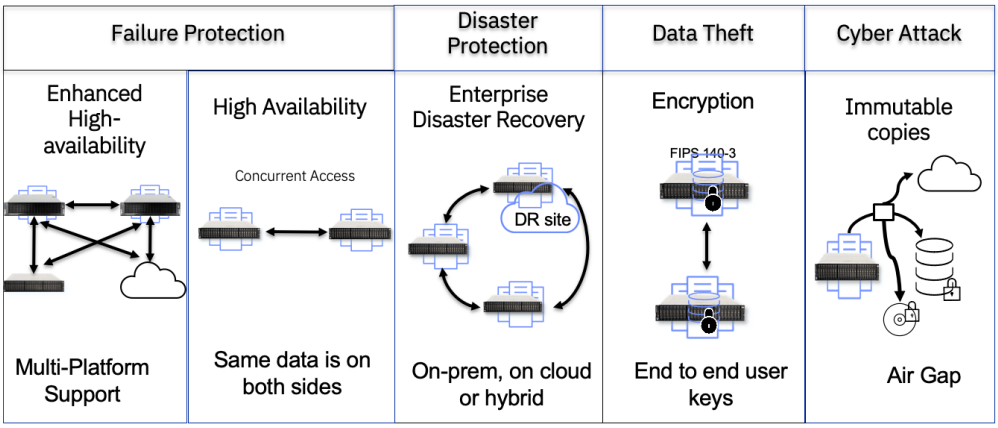
There are multiple benefits of the management services including:

- Admins can manage more data with easy interface and globally connected data
- Integrated meta data for policy-based optimization to help maintain SLAs
- Integrated backup and recovery to maintain always on access to data
- Lower cost of storing data with transparent tiering and policy-based archiving

The IBM Storage Scale management GUI provides an easy way to configure and manage various features that are available with the IBM Storge Scale and Storage Scale System. Since the system supports global data and the GUI can manage all the data one truly has a global interface and view to all the data being access by IBM Storage Scale.

- You can perform the following important tasks through the IBM Spectrum Scale management GUI: Monitoring the performance of the system based on various aspects
- Monitoring system health
- Creating and managing file systems
- Creating filesets and snapshots
- Managing Object and NFS and SMB data exports
- Creating administrative users and defining roles for the users
- Creating object users and defining roles for them
- Defining default, user, group, and fileset quotas
- Creating and manage node classes
- Monitoring the capacity details at various levels such as file system, pools, filesets, users, and user groups
- Monitoring and managing various services that are available in the system
- Monitoring remote clusters
- Configuring call home
- Configuring and monitoring thresholds
- Creating and managing user-defined node classes
- Configuring user authentication for NFS and SMB users.

Data Resiliency Services



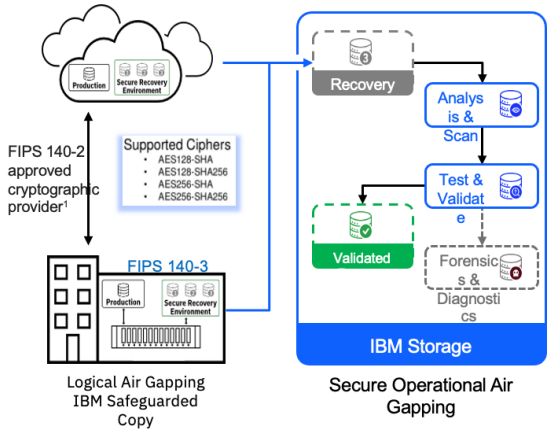
Ensuring your data stays available to your applications is the primary function of storage. In a hybrid cloud or distributed environment, organizations are working with multiple locations comprised of on-premises and public clouds. With a strategic storage software foundation like IBM Storage Scale deployed on-premises in storage systems and on public clouds, you have great flexibility in how you leverage your infrastructure locations for operational resiliency. The following are common configurations and it's certainly not an exhaustive list.

Resiliency Services address cyber secure challenges capabilities:

- Safeguarded Copy and air gap cyber security
- CyberVault cyber security services
- Encryption disk support (data at rest)
- End to end encryption (data in transit)
- Key management support
- SEC 17a-4 compliant locks and AES512 encryption
- Alert management (event log) and data catalog

Benefits

- Recovery in minutes from cyber attack
- Meet government compliance recruitments
- Alert applications such as Splunk or QRadar for continuous monitoring
- Enables a comprehensive security strategy
- Meet availability SLAs and provide multiple secure access points for data



Safeguarded Copy

- Space efficient snapshots (scheduled, manual, triggered) stored in an immutable secure recovery location
- Simple GUI Interface with single screen policies
- Up to 256 recovery points per fileset
- Logical air gapping solutions with security controls using separation of duties for checks and balances

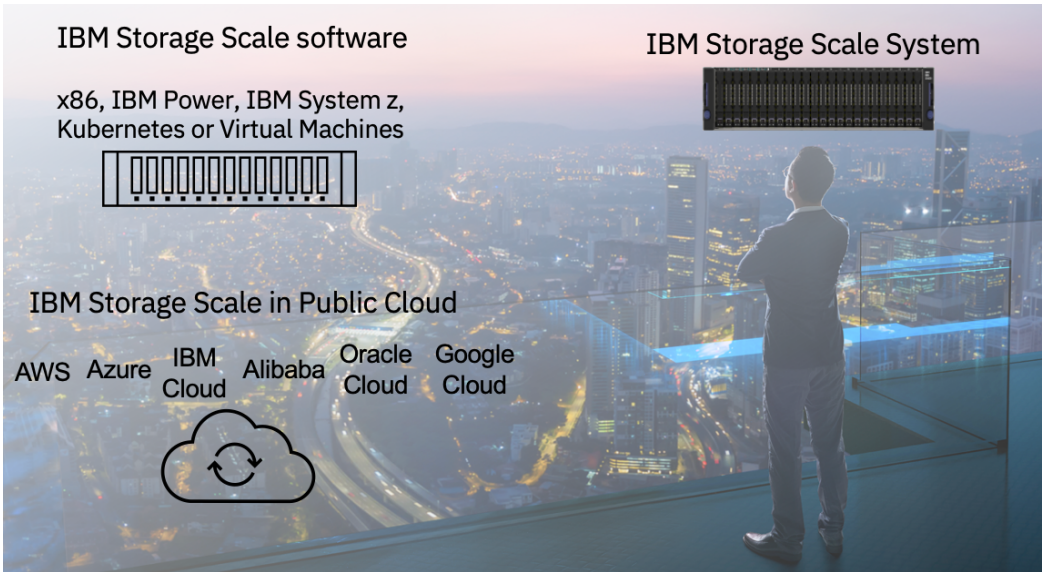
Benefits

- Reduce recovery from days and/or hours to minutes
- Backups cannot be accessed or modified by unauthorized applications/staff
- Enables a comprehensive security strategy with Security Information and Event Management (SIEM) tools for response automation

IBM's air-gapping approach begins first on-premises (or in the cloud) with IBM Safeguarded Copy, which creates a logical air-gap relationship between the source volume and the secure recovery environment. These restore points can be used to quickly bring data back online following a cyberattack if a crash-consistent restore operation is required We refer to this as a secure operational air-gap as the data set being analyzed for restore operations is separated form the source array, but often in the same datacenter.
IBM Safeguarded Copy for cloud scale file and object storage is a new protection mechanism for data on IBM Storage Scale and Storage Scale System.
Safeguarded Copy sessions secure data to prevent it from being compromised, either accidentally or deliberately.
Safeguarded Copies can be used to take many frequent copies of a production environment (for example, hourly copies maintained for several days), while Storage Scale snapshots continues to be used to take a small number of less frequent copies (such as weekly copies maintained for 1-2 weeks).

The Safeguarded Copy function provides backup copies to recover data in case of logical corruption or destruction of primary data.
Safeguarded Copy uses a backup capacity, filesystem, and filesets can be created for any production volume. The size of the backup capacity depends on the frequency of the backups, as well as the duration that backups need to be retained.
The Safeguarded Copy session creates a consistency group across the source files to create a safeguarded backup, which stores the required data in the backup capacity.
Production fileset(s) are the source files for a Safeguarded Copy relationship.
Recovery fileset(s) are used to restore a backup copy for host access while production continues to run on the production fileset(s). The recovery fileset is the target volume for a Safeguarded Copyrecovery, which enables a previous backup copy to be accessed by a host that is accessing this fileset. The recovery fileset is typically thin-provisioned, but it does not have to be.

Multiple ways to deploy a Global Data Platform



There are multiple ways to deploy the IBM Storage Scale System and they all can be interchanged because they are all leveraging the same Storage Scale software. IBM innovation provides a plug and play high performance flash storage solution that consolidate all file and object data on a global data platform . Clients can start with a single storage node with as little as 48TB and can grow to 1000s of nodes or to YB capacity. Customers can also leverage their own hardware or even a virtual machine. You can use almost any x86 hardware or use a certified partner offering like Seagate, Lenovo, Pixit Media or HPE. IBM Storage Scale can also be deployed in the public cloud as a service you deployed on cloud resources. IBM has several offerings available through IBM or partners on clouds such as AWS, Azure, IBM Cloud, Alibaba or Oracle and recently introduced Google Cloud Platform.

Why IBM?

Data matters. When planning high performance infrastructure for new or existing applications it's easy to focus on compute resources and applications without proper planning for the data that will drive the results for the applications. Our products are all about solving hard problems faster with data. IBM helps customers achieve business value with a clear data strategy. Our strategy is simple, unlock data to speed innovation, de risk data to bring business resilience and help customers adopt green data to bring cost and energy efficiencies. Value needs to be delivered by connecting the multiple organizational data sources with business drivers to create business value that mean something to the organization. Many organizations focus on a single driver with a storage solution, but the best solution is driven by an infrastructure strategy than can accomplish most if not all the drivers for maximum benefits. Our story is not just about another storage product but is about innovation and a storage portfolio that is powered by our global data platform.

For Further Information

For further information on IBM Storage file and object products please visit
https://www.ibm.com/ai-storage

Next steps

Contact your IBM Representivie or your IBM Business Partner
https://www.ibm.com/partnerworld/bpdirectory/