

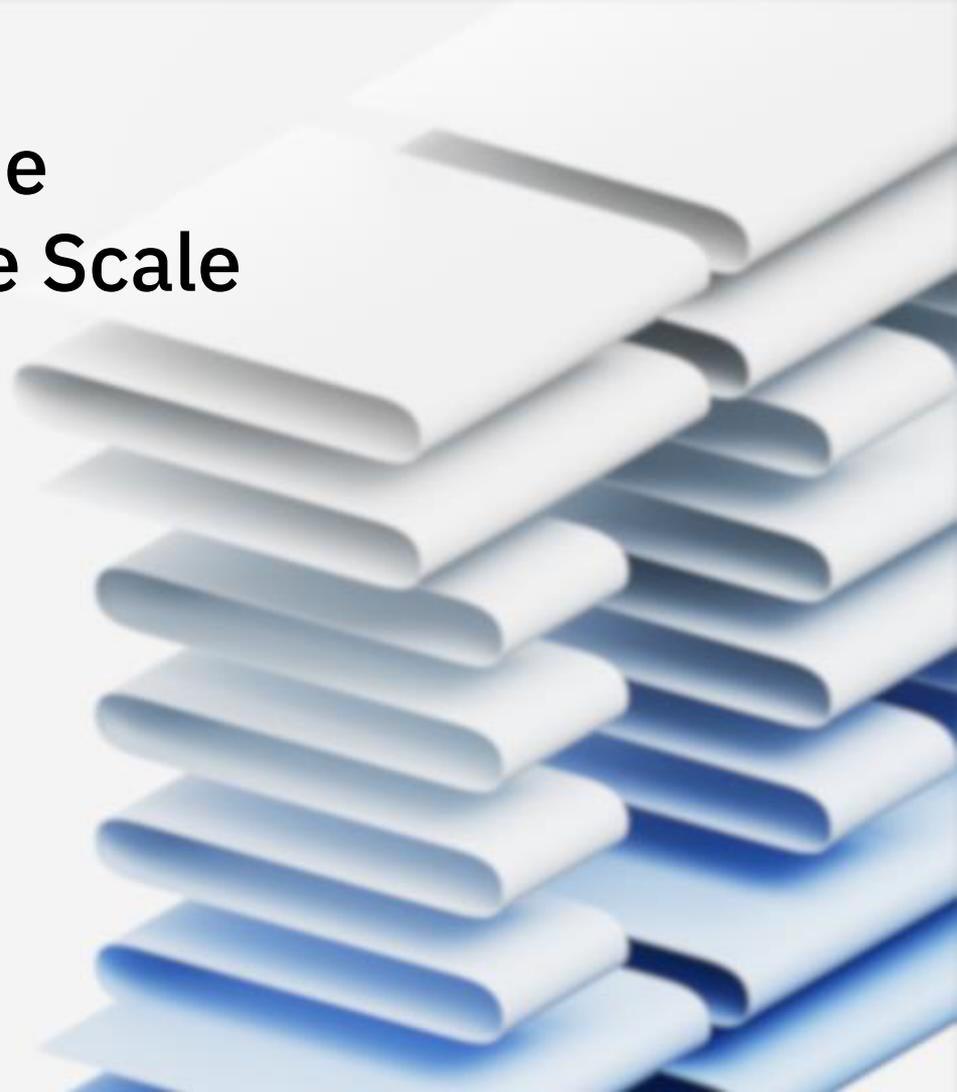
E09: Brückenschlag in die Cloud(s) mit IBM Storage Scale

IBM Storage Scale Days 2025 DE

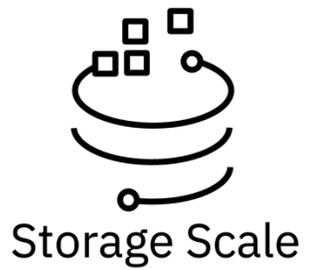
March 18th – 20th, 2025 | Heidelberg, Germany

Frank Kraemer (IBM)

<mailto:kraemerf@de.ibm.com>

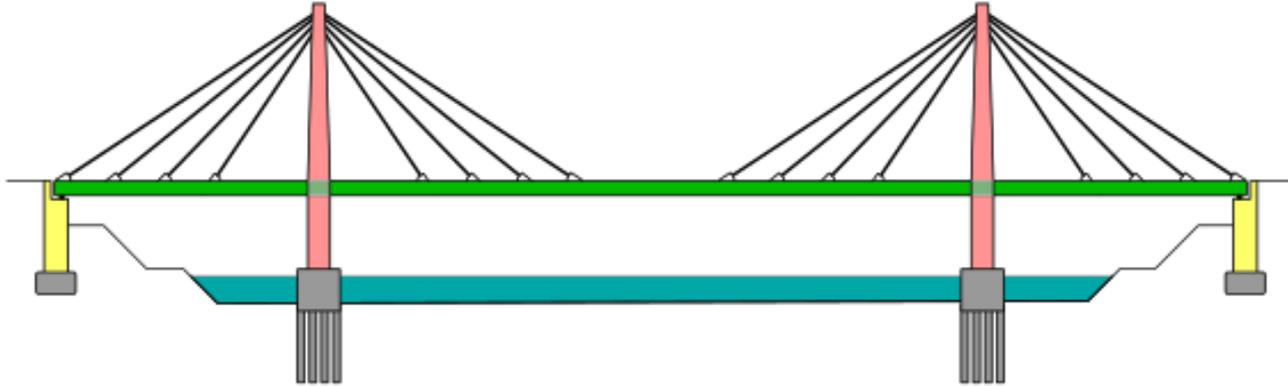


Disclaimer



- IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.
- IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Was ist ein Brückenschlag?



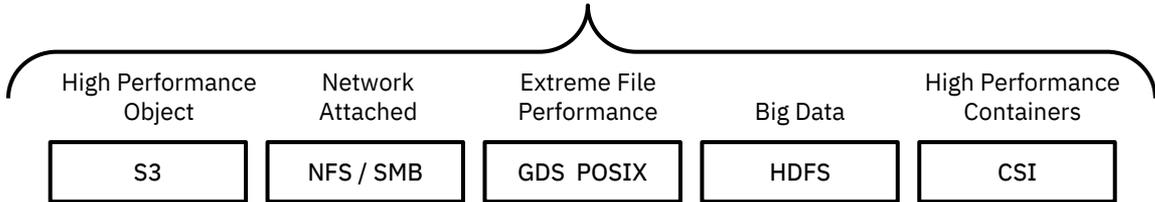
Brückenschlag bezeichnet den Bau einer Brücke. Eine Brücke ist ein Bauwerk, das Verkehrswege oder Versorgungseinrichtungen über natürliche Hindernisse oder andere Verkehrswege hinwegführt.
(Quelle: <https://de.wikipedia.org/wiki/Br%C3%BCckenschlag>)

A Global Data Platform for Unstructured Data

Unifying File and Object Storage to Provide Common Data Services



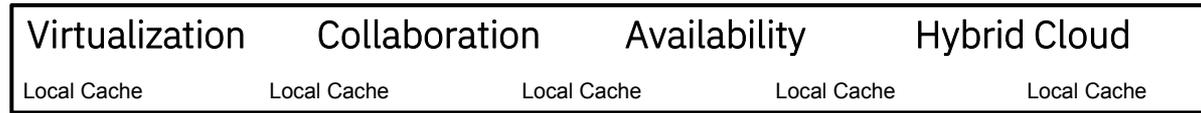
Data Access Services



Data Resiliency Services

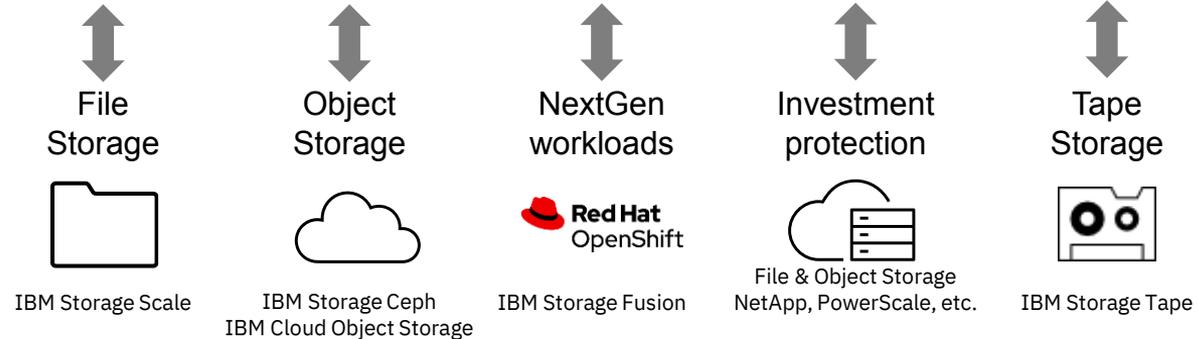
Identify

Data Caching & Core Services



Protect

Data Management Services



Detect

Respond

Recover

A Global Data Platform for Unstructured Data

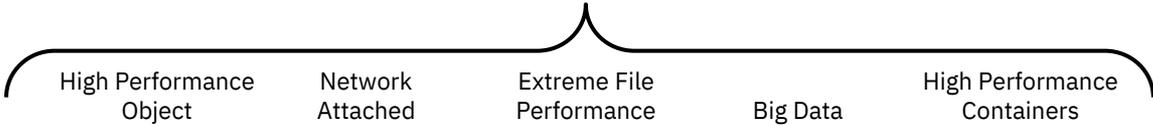
Unifying File and Object Storage to Provide Common Data Services



Data Access Services

Data Caching & Core Services

Data Management Services



Data Resiliency Services

Identify

Protect

Detect

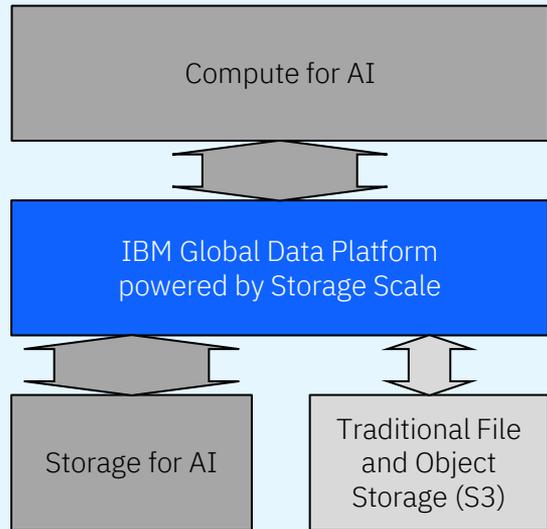
Respond

Recover



Source: Victorinox.com

Storage Scale for AI and the IBM Global Data Platform



1/ File and object storage

Most corporate unstructured data resides on traditional file and object storage.

Storage for AI is emerging as new file storage in addition to traditional file and object storage.

2/ Performance challenge

Traditional file and object storage does not provide the performance needed for AI.

Storage for AI provides the performance and scaling which is required for AI at scale.

3/ Data integration

How can I make my existing data available for AI training and AI inferencing?

The IBM Global Data Platform integrates traditional file and object storage and accelerates AI workloads by seamless caching of unstructured data.

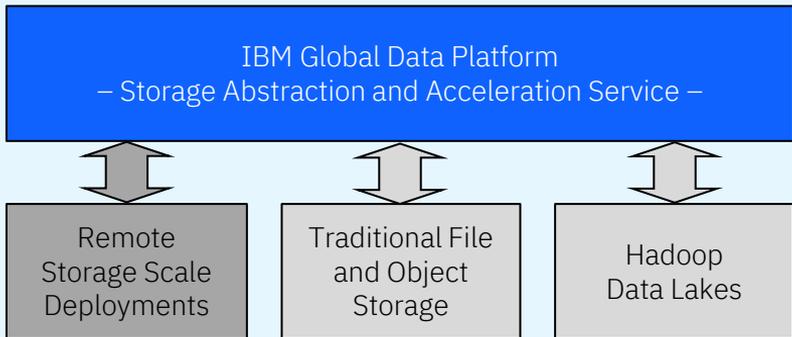
4/ Investment protection

How can I add GPUs and high-performance storage to my existing IT landscape?

The IBM Global Data Platform integrates Storage for AI and makes all unstructured data available to the GPUs with high throughput and low latency.

The IBM Global Data Platform

– Storage Acceleration and Abstraction Services –



1/ Storage Abstraction Services

Provides globally distributed storage for unstructured data.

Connects globally distributed file and object storage to single global namespace.

Integrates your existing file and object storage in your Storage for AI.

2/ Storage Acceleration Services

Provides seamless caching of remote file and object data.

Keeps data in cache in sync with connected file and object storage systems.

Serves data-hungry GPUs from data in cache.

3/ Bring your own storage

Flexibility for varying operational requirements.

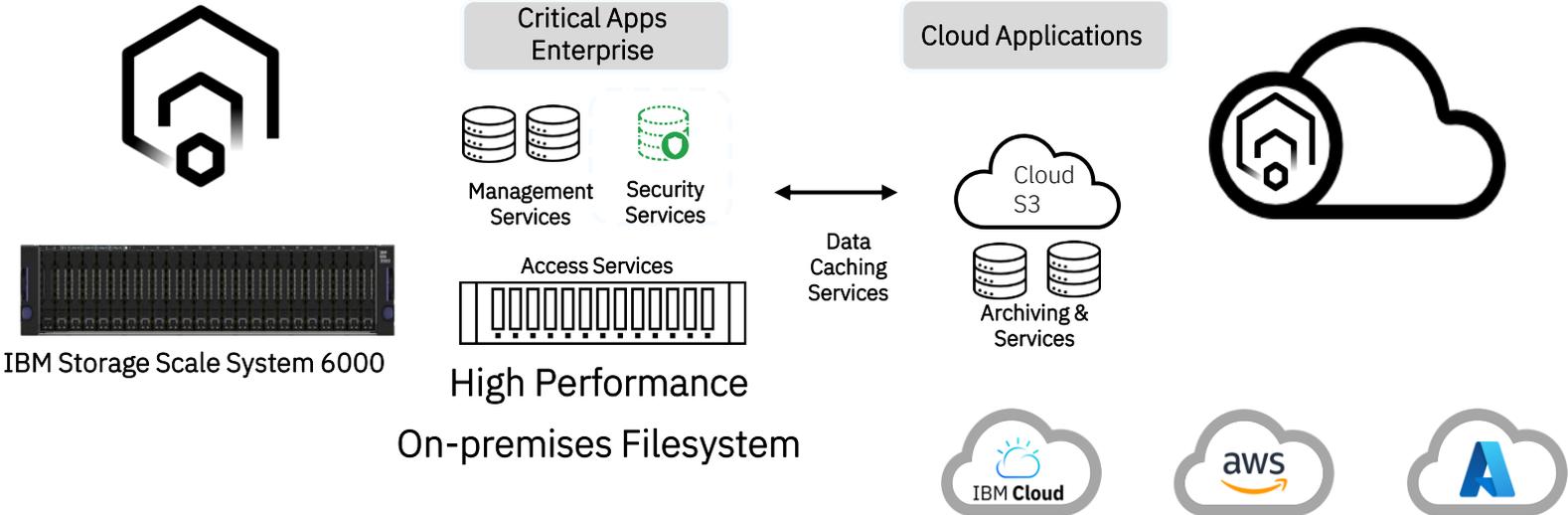
Continue to use your established file and object storage vendors for operational efficiency.

4/ Investment Protection

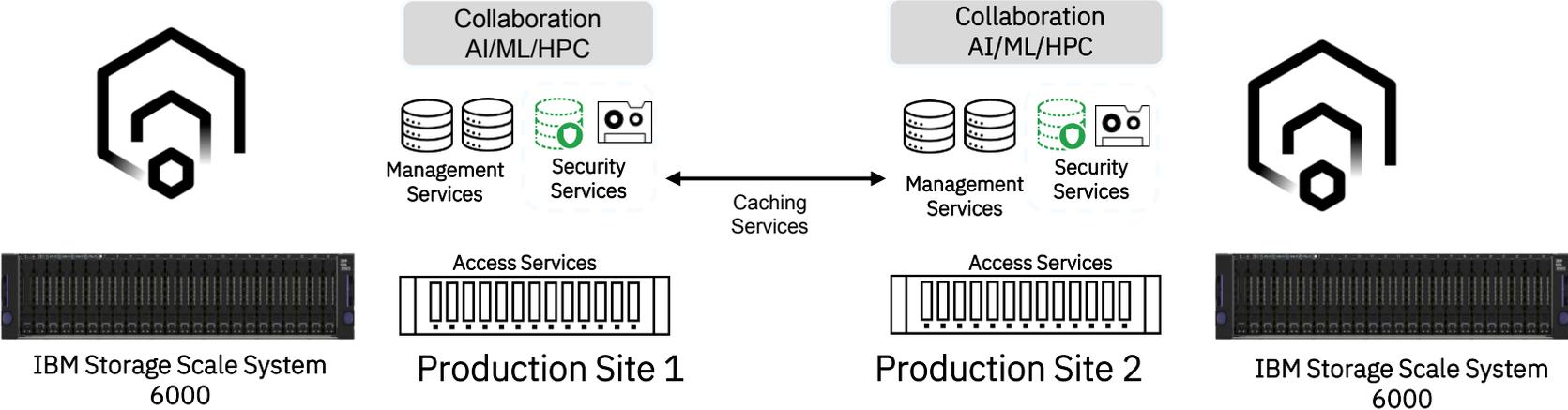
Integrate AI workloads in your existing IT landscape.

No need to replace your existing infrastructure. Add and integrate Storage for AI for performance hungry workloads.

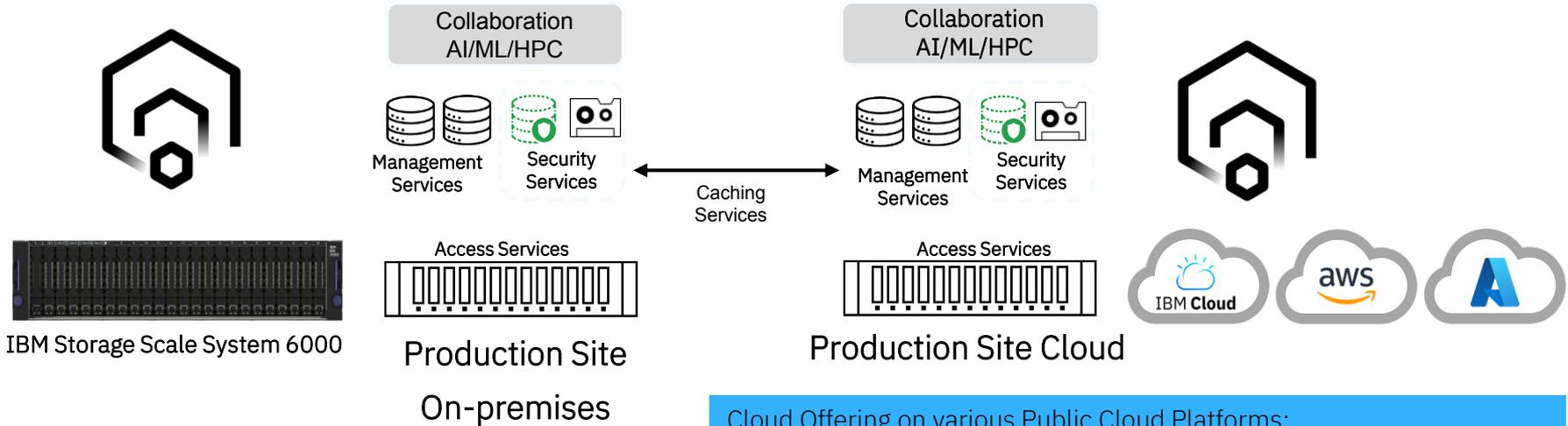
#1 Storage Scale to S3 (Cloud)



#2 Storage Scale to Storage Scale (Dual Site concept)



#3 Storage Scale to Storage Scale (Cloud Site concept)



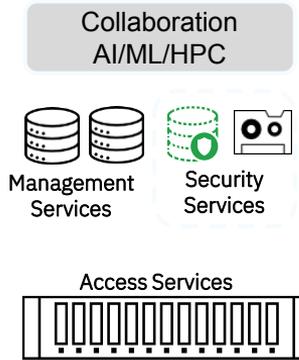
- Cloud Offering on various Public Cloud Platforms:
1. File Storage As a Service (Partner Managed)
 2. Marketplace Offering BYOL (Bring Your Own License)
 3. Self Managed via Storage Scale Cloudkit (Bring Your Own License)

#3.1 Example: Oracle Cloud Infrastructure (OCI)

<https://blogs.oracle.com/cloud-infrastructure/post/accelerate-ai-ml-workloads-oci-nvidia-ibm>

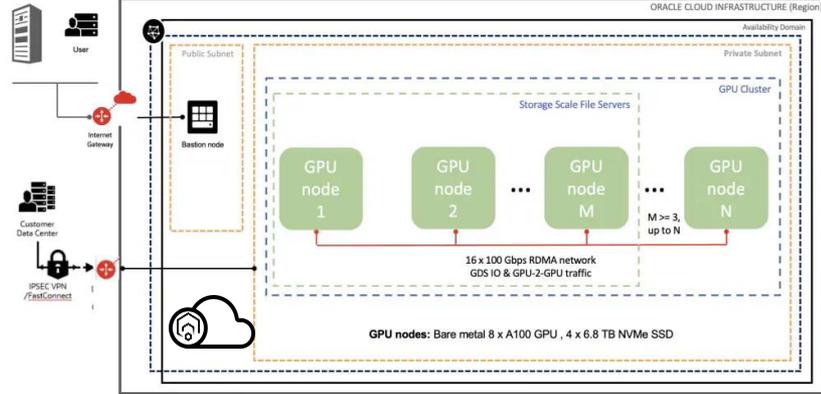


IBM Storage Scale System 6000



Production Site
On-premises

Caching Services



In collaboration with NVIDIA and IBM, Oracle Cloud Infrastructure (OCI) validated NVIDIA Magnum IO Storage on OCI using a hyperconverged cluster of bare metal NVIDIA A100 Tensor Core GPU compute nodes.

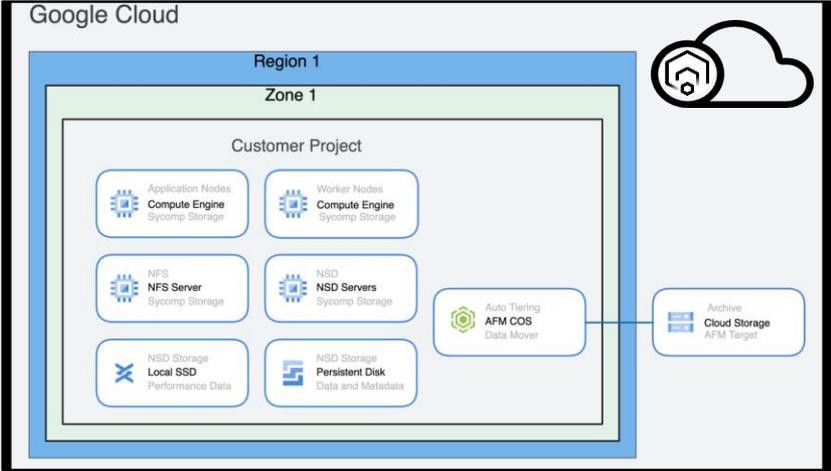
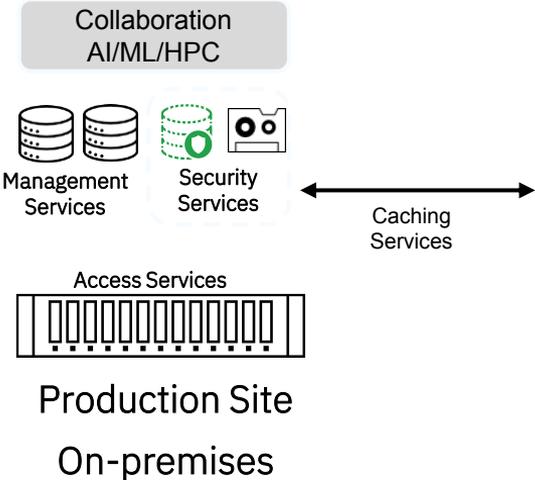
#3.2 Example: Google Marketplace Solution

(by IBM partner Sycomp)

<https://cloud.google.com/blog/products/storage-data-transfer/announcing-ibm-spectrum-scale-is-now-available-in-google-cloud/>

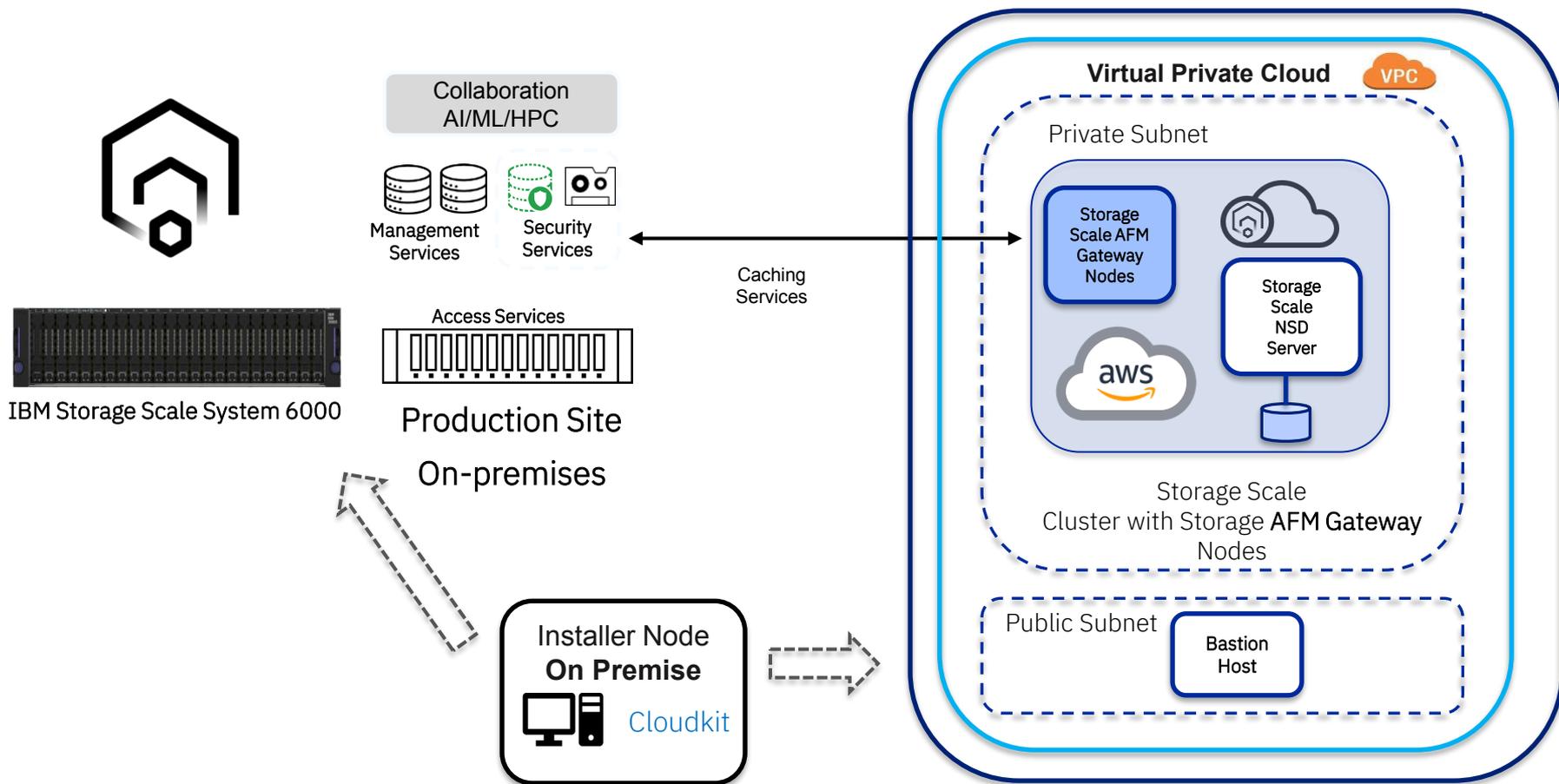


IBM Storage Scale System 6000



Deploy an instance of Sycomp Storage Fueled by IBM Spectrum Scale using Terraform provided by Sycomp through Google marketplace.

#3.3 Example: AWS by Scale Cloudkit (Self Managed)



IBM Storage Scale Cloudkit: Introduction



What is Storage Scale Cloudkit?

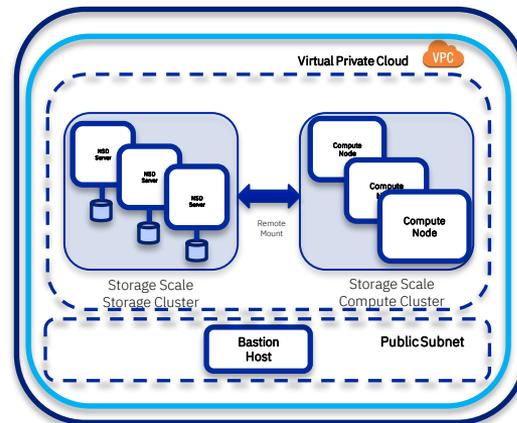
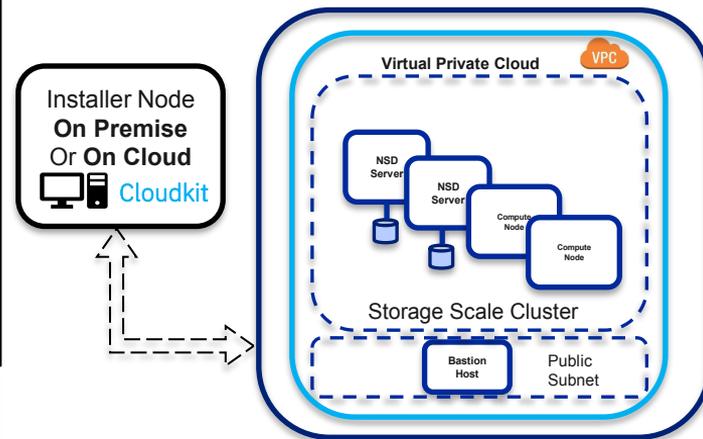
- Command Line Interface to create Storage Scale clusters on the cloud
- Provides end to end automation to deploy and configure Storage Scale cluster(s) on public clouds:
 - Automates infrastructure provisioning on the cloud
 - Automates the deployment of Spectrum Scale on the cloud
 - Applies Spectrum Scale best practises for deploying on the cloud

Advantages of Storage Scale Cloudkit

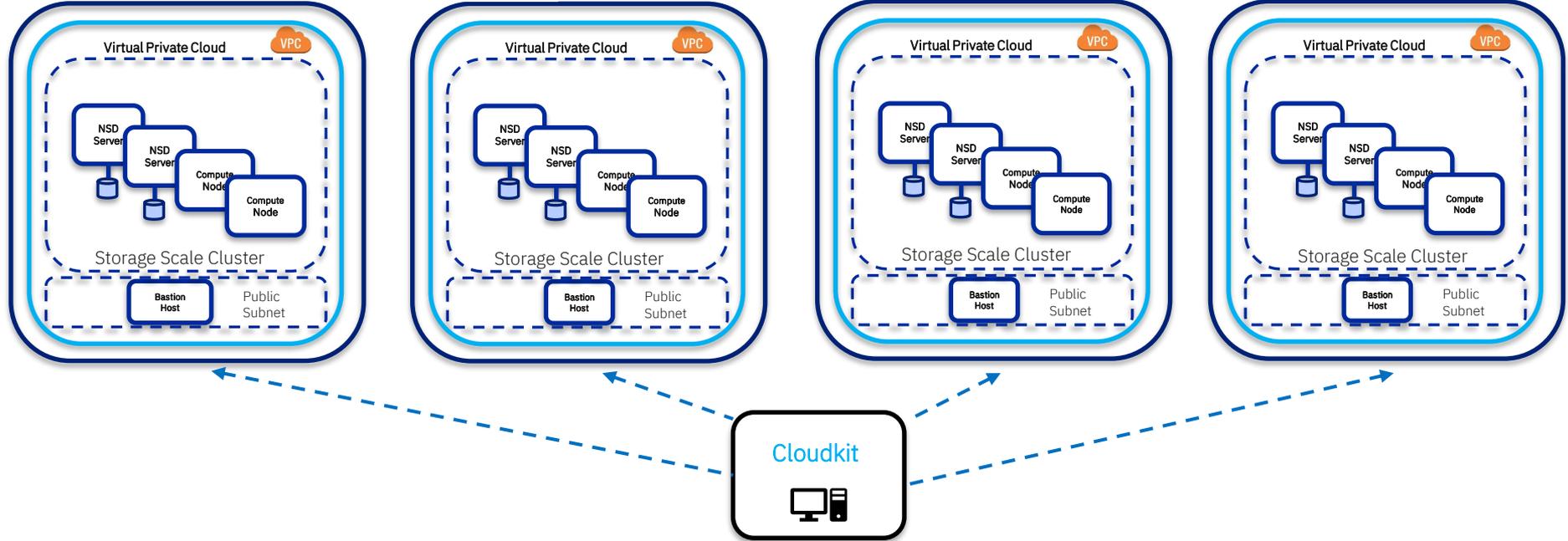
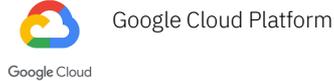
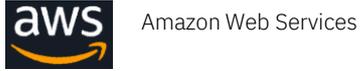
- Support for all major public clouds (more details see Roadmap)
- Rapid deployment and configuration of Storage Scale on the Cloud (in minutes)
- Out of the box “best practice” configuration
- Easy to use, guided interface
- Ability to provision multiple clusters from a single cloudkit instance
- Flexible deployment (supports multiple deployment models to suit customer needs)

- Deploy into existing customer environment (existing VPC)
- Create and deploy into new VPC
- Support Direct Access Network connection between On-prem and Cloud clusters

- Deploy separate Storage Scale Storage and Compute clusters
- Deploy a single Storage Scale cluster with Storage and Compute nodes



Cloudkit: Multi Cloud Support



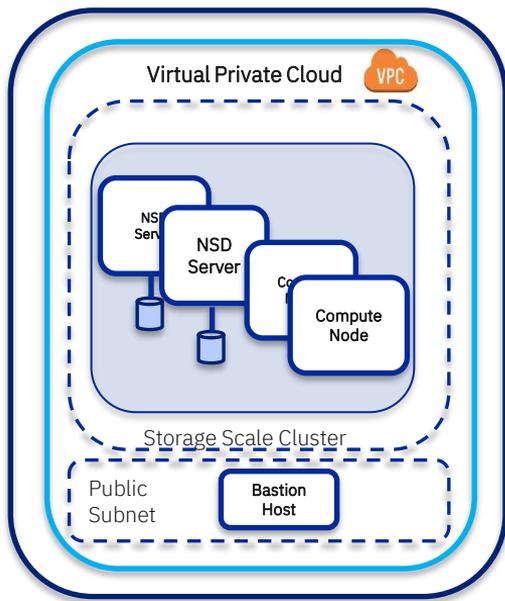
A single instance of Cloudkit can deploy Storage Scale clusters on different public clouds:

- Unified Deployment Experience
- No longer necessary to use different Cloud specific deployment technologies to provision on different clouds
- Cloud Agnostic Best Practise Deployment
- Enables Global Data Platform

Cloudkit: Flexible Deployment Architecture

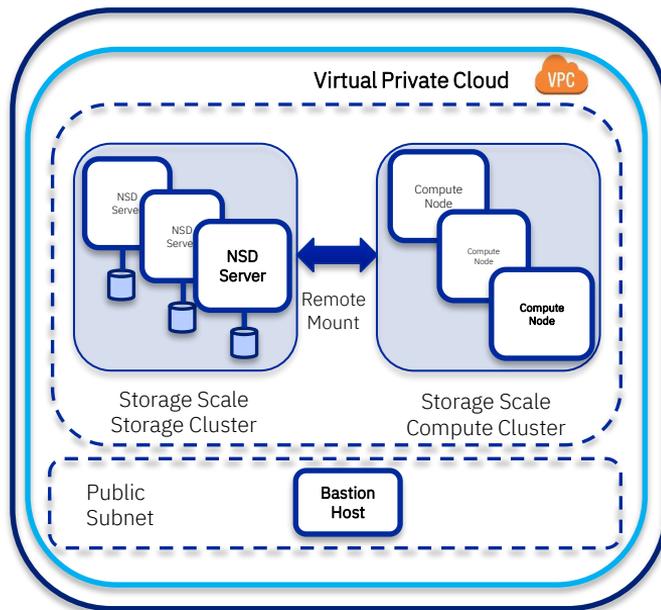


Single Unified Cluster



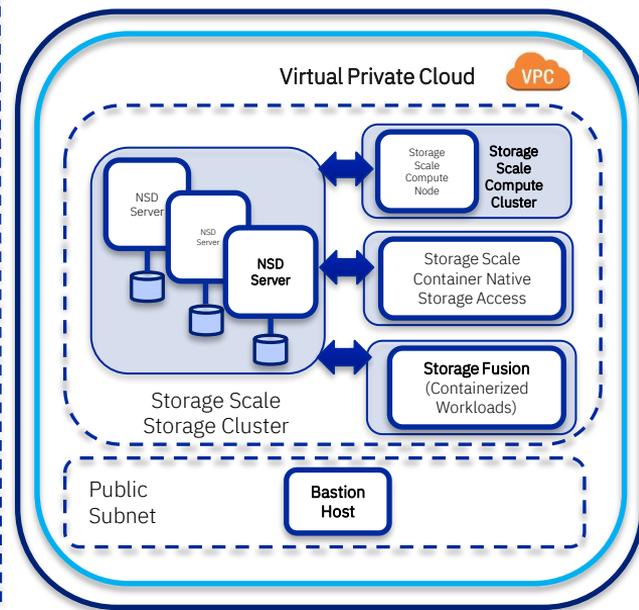
- Single administrative domain
- Storage and Compute nodes in the same cluster
- Ideal for small organizations

Separate Storage Cluster remote mounted Compute Cluster (s)



- Separate administrative domains
- Rapid Provision/Deprovision of Compute Clusters as required
- Storage Cluster persistent and isolated

Single Storage Cluster remote mounted to multiple Compute Clusters

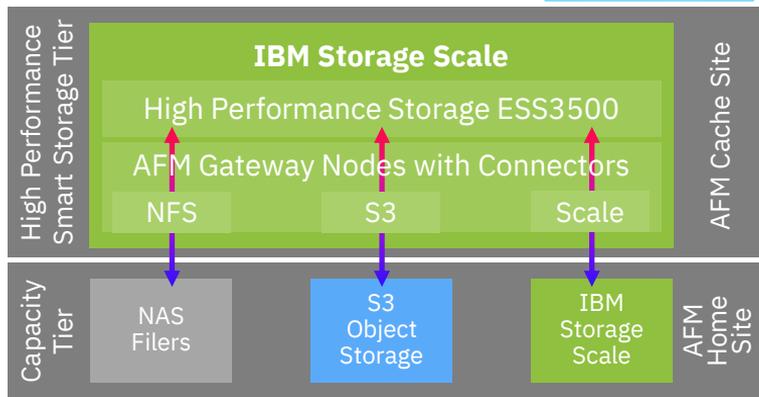


- Single Storage Data Lake for entire organization (Storage Cluster)
- Separate Compute for each Department administrated independently
- Provides Persistence, High Performance Storage for IBM Storage Fusion (Containerized Workloads)

Scale Data Caching Services (AFM) - Use case details #1

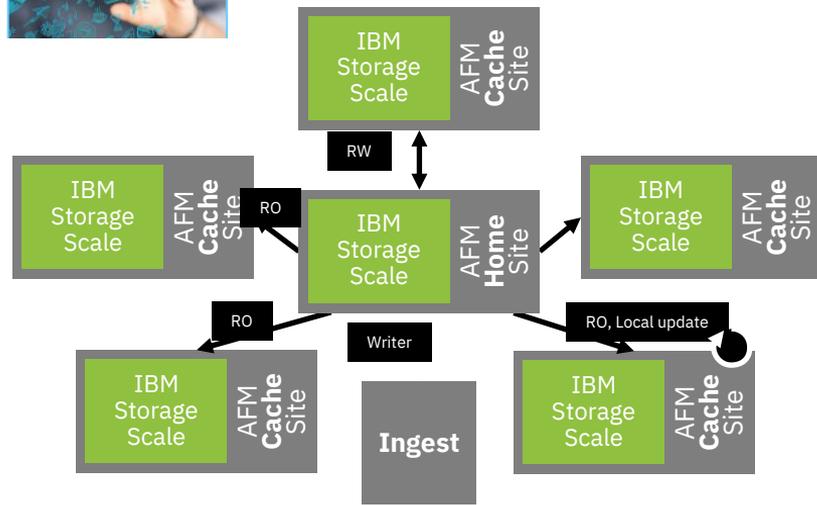


Data Virtualization



- Vertical caching
- Common namespace across isolated data silos in legacy 3rd party data stores
- Transparent access to all data regardless of silos
- Scale-out Posix performance
- Data export via NFS, SMB, HDFS, Object
- Can be used to seamlessly migrate data to new storage

Data Collaboration

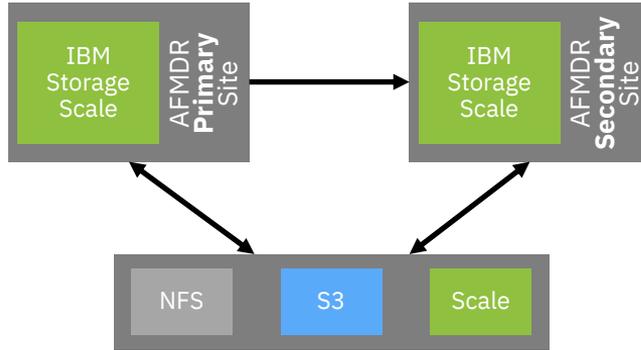


- Consistent cache provides a single source of truth with no stale data copies
- Horizontal caching
- Bi-direction traffic from Edge to Center
- Eventually Consistent data cache
- Transparent on-demand data access and transfer
- Policy driven data prefetch and eviction

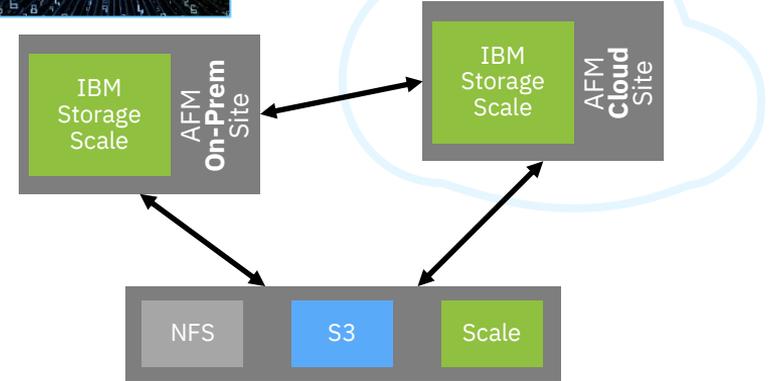
Scale Data Caching Services (AFM) - Use case details #2



Data Resilience



Hybrid cloud / Bursting



- Active-Passive DR over WAN or Cloud
- Designed for high latency and asynchronous DR
- Hot standby failover to DR site
- Automatic fallback data reconciliation
- Read-only access / analytics to all data at passive site

- Rapidly expand compute resources to cloud or data centers
- Common file system creates a single namespace across all locations
- Transparent access to data
- Cost effective way to increase compute on existing data
- Analytic results automatically pushed to home site

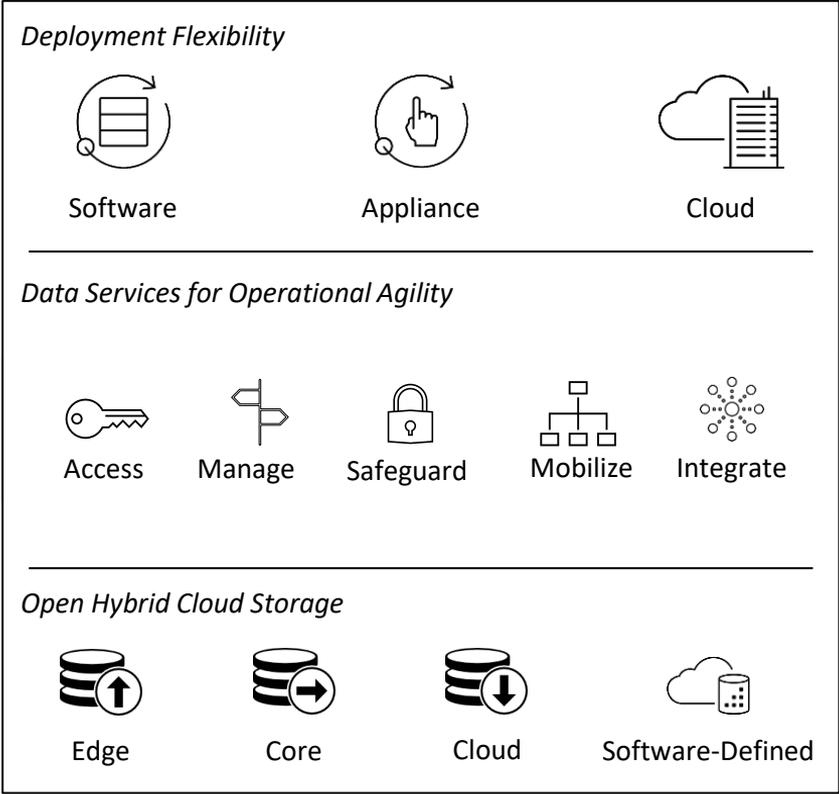
Summary: IBM's Global Data Platform



An Open Hybrid Cloud Experience

Data Sources

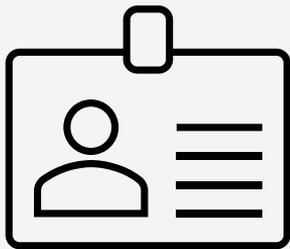
- Transactional
- Application
- Semi-Structured
- Mixed Media
- IoT



Outcomes

- Data Monetization
- Faster Time to Market
- Org Growth
- Digital Business

Thank you



Frank Kraemer (IBM)
<mailto:kraemerf@de.ibm.com>

