

Data Should be as **Elastic as Cloud**

SOLUTIONS BRIEF

Virtual Data Lakes, On-Demand

THE CHALLENGE

It is attractive to take advantage of elastic compute resources to run analytics jobs in the cloud, but the number one reason analytics jobs fail is that they took too long to run, and often the most time-consuming stage is finding data and staging it for the job. Data operations teams struggle to gain better visibility and access to unstructured data sprawled across on-premises storage systems and cloud services by pooling data into cloud data lakes, creating yet another silo. Too often, the exercise of filling a data lake does not take data locality into consideration, resulting in a data swamp, filled with data of questionable quality.

THE SOLUTION

Hammerspace's data-centric approach allows you to serve and manage data independently from the infrastructure. Built for the hybrid multi-cloud, Hammerspace serves data at high-performance from any storage across the infrastructure. Using standard protocols, Hammerspace abstracts data at file-level granularity from the infrastructure, making it easy to burst-to-cloud unstructured file data and instantly create virtual data lakes using enriched user-defined metadata, ensuring that the data lake is filled with quality data. The global cost of data is tightly managed using global data reduction technologies to reduce capacity and leveraging data virtualization avoiding unnecessary copies of data.

To serve data across the hybrid multi-cloud, Hammerspace separates the control plane (metadata) from the data plane (data) reading, writing, and moving data across sites through a Universal Global Namespace at file level granularity. Hammerspace metadata servers (Anvil) are present at each site, replicating metadata so that every site has a complete view of all data, with the assistance of machine learning-driven automation to direct resource optimization. When non-local data is accessed, Hammerspace data services (DSX) moves data live to where it needs to be, even while actively being read/written. DSX data services are architected to scale-out on-demand so that performance is parallelized to meet application SLAs. In Hammerspace, data is protected against the loss of infrastructure by services like snapshots, undelete, and replication.

AGILITY, CONTROL & EFFICIENCY

Instantly create virtual data lakes, anywhere

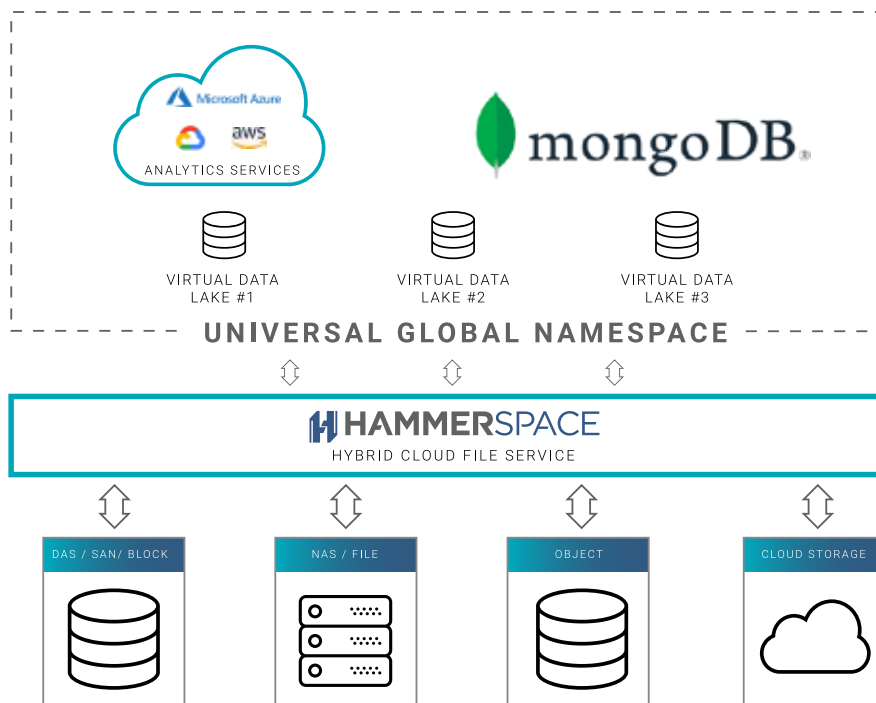
- Present data from multiple on-prem sources into the cloud as a unified view
- Discover and present data collection via user-defined metadata
- Skip manual copy processes with live data mobility

Secure data transport, on-prem to cloud

- Periodic checkpoints protect data with global snapshots
- Data air-gap connects to cloud without exposing on-prem infrastructure
- Data is secure and encrypted with KMS
- Custom metadata identifies data origin

Serve data from any storage infrastructure

- Reduce unnecessary copies of data
- Consolidate storage resources
- Optimize capacity with global dedupe and WAN optimization
- Data is served at high-performance at any location



AGILITY: INSTANTLY CREATE VIRTUAL DATA LAKES, ANYWHERE

With Hammerspace, data can be burst-to-cloud from multiple data centers to create virtual data lakes, on-demand. Data management is fully automated and served through a unified view into the Universal Global Namespace, simplifying access across multiple data stores. Discover and present data with searchable global collections that can be filtered via the metadata in order to keep the data lake clean. Non-disruptive live data mobility makes it easy to consume unstructured file data without waiting for a lengthy manual copy process. Data is visible at all locations, but only copied on-demand or pro-actively by policy so that analytics jobs can start sooner.

CONTROL: SECURE DATA TRANSPORT, ON-PREMISES TO CLOUD

Global snapshots automate data protection with periodic checkpoints of data lakes, making it easy to self-service roll-backs. Also, users can create a data air-gap to connect sensitive data to the cloud without exposing on-premises infrastructure, and identify data origin with custom extensible metadata.

EFFICIENCY: REDUCE THE COST OF DATA LAKES WITHOUT SACRIFICING PERFORMANCE

Reduce unnecessary copies of data by consolidating storage resources while leveraging cloud native storage services to expand the availability of on-demand data lakes. With Hammerspace, data is presented virtually so data is only copied as necessary file-by-file. Global dedupe takes back data reduction from cloud providers and WAN optimization keeps transfer costs low, while scale-out parallel data services deliver high-performance.

ABOUT HAMMERSPACE

Hammerspace, the Data-as-a-Service company, manages and protects data on the hybrid multi-cloud overcoming the challenges of making unstructured data cloud-native, independent of the infrastructure. With non-disruptive, autonomic data management Hammerspace reduces the complexity of embracing hybrid, multi-site, or Kubernetes workflows. To learn more, visit us at www.hammerspace.com