

Right Data, Right Place, Right Time: Managing by Objectives with DataSphere

Today's IT professionals have more choice than ever before when selecting storage to meet application needs. With different products offering a wide range of capabilities across performance, protection and price, IT can strategically choose the option that best matches expected data demands when adding new storage. However, IT is forced to take a bottoms up approach to meeting data requirements, purchasing and allocating storage based on expected peak performance and protection needs. DataSphere closes the gap between business and IT, enabling IT to define target objectives directly to data with fine granularity. DataSphere then automatically and non-disruptively moves data to the right storage to meet these objectives, ensuring desired service levels are always met. Let's examine how DataSphere ensures the right data is in the right place at the right time.

Modern Storage Offers Unique Capabilities

Thanks to varying levels of performance, protection, price and capacity, today's storage options including flash, shared and cloud storage each provide different benefits to the business. For example, server-side flash storage is very fast, providing low latency and high IOPS, but flash is considered less reliable and is more expensive than shared networked storage. Shared storage, such as NAS filers and SAN arrays, has lower performance and cost than server-side flash, but provides higher levels of data reliability through sophisticated RAID protection operations, error correction schemes, and disaster recovery redundancy. Cloud and object storage resources feature lower costs, and high capacity density, but today only deliver performance suitable for colder data when accessed remotely.

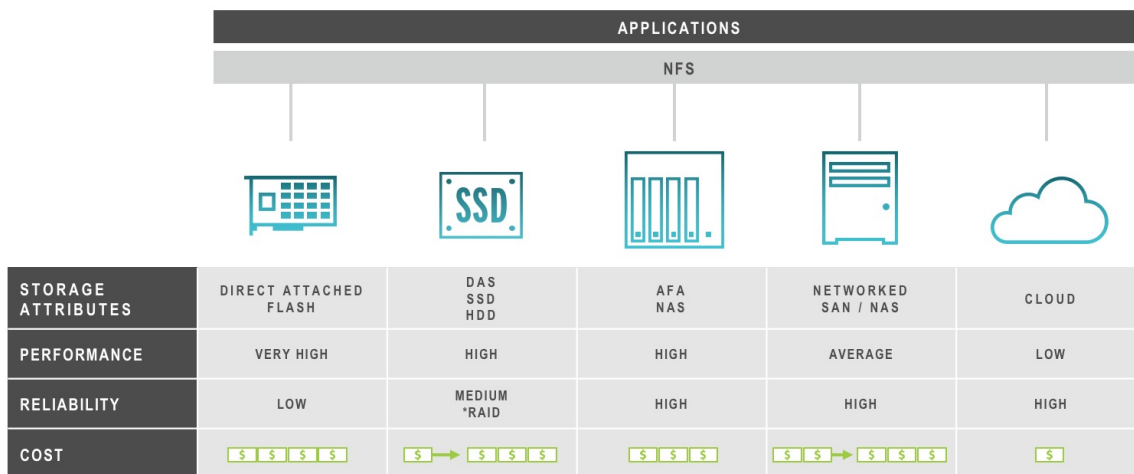


Figure 1. Different storage resources offer varying attributes across performance, reliability and cost.

Connect Storage Resources in a Global Namespace

DataSphere pools multiple storage resources and presents a logical global namespace to clients. This approach virtualizes the view of the client's data by using open standards-based storage I/O protocols already installed in most clients and the DataSphere metadata engine. When storage is mounted to the DataSphere, the system discovers the capabilities of the resource and adds it to a pool of storage with similar capabilities, or creates a new tier of service before adding it to the global dataspace.

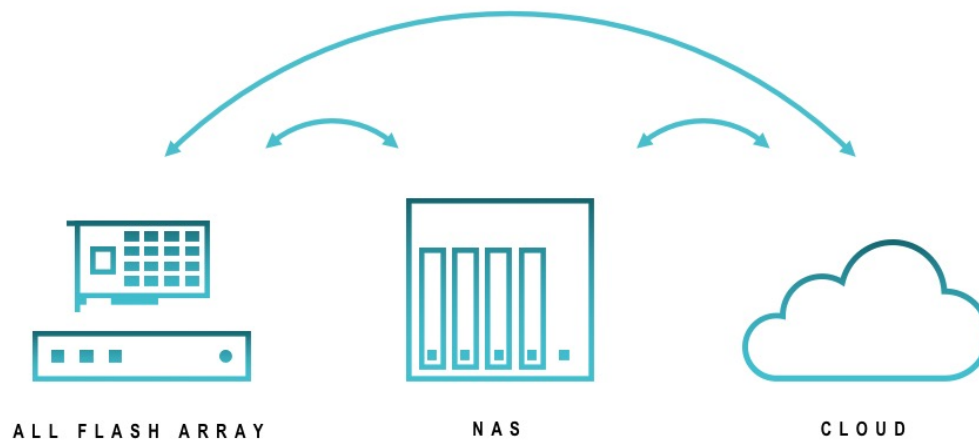


Figure 2. Connecting storage in a global namespace enables data to move to the right storage at the right time.

Automatically Provision Data Across Storage to Align the Right Resource for the Job

DataSphere enables IT or business administrators to manage data according to target objectives for the following attributes:

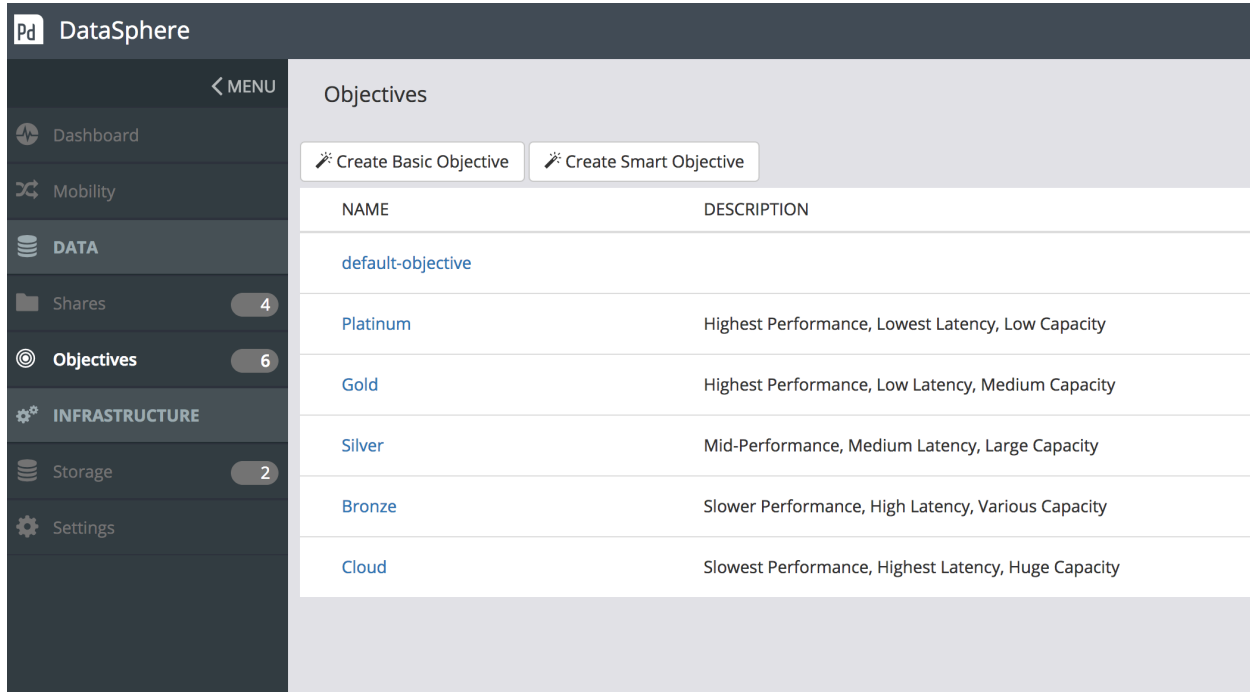
- **Performance:** Admins can set objectives for IOPS, latency, and bandwidth to ensure ideal application performance.
- **Protection:** Admins can set objectives for availability, reliability, durability, and security to meet application protection requirements.
- **Time:** Admins can set objectives based on file activity or inactivity. This objective is often combined with other objectives. For example, admins can apply objectives to a share that ensures that all active files in a share that have been accessed in the last day are placed on storage that can deliver 10,000 IOPS, 100MB/bandwidth, and 0.5ms latency and all files that have not been accessed in the last 30 days are moved to storage designed for preservation of data (with high reliability, durability, and low cost), such as object or cloud storage.
- **Pattern Matching:** Admins can set objectives based on regular expression pattern matching. For example, admins could set objectives that files matching “.tmp” be stored on the local storage tier, while all other files are on shared storage tier.

DataSphere features a powerful graphical user interface (UI) to create, define and manage objectives. The image below illustrates how an objective can be defined so that data can automatically tier across different types of storage as it ages:

The screenshot displays the 'Create Smart Objective' interface in DataSphere. The top navigation bar includes the 'Pd Primary Data' logo and user profile icons. The main content area is titled 'Create Smart Objective' and contains a 'Conditions' section with the instruction 'Select files and define objectives for those files based on their activity.' Below this, there are input fields for 'Objective Name' (containing 'Hot to cold') and 'Description'. The 'Select Files' section includes a search bar with an asterisk, 'Update' and 'Cancel' buttons, and a 'Define activity ranges' section. This section features a horizontal bar with 'Inactivity' and 'Activity' arrows, and three rows of activity ranges with corresponding tier selections: 'Tier 1', 'Tier 2', and 'Cloud Tier'.

Figure 3. Objectives ensure the right data is served by the right resource, automatically.

Extending the capabilities of objectives even further, storage administrators can also create a service catalog that they can make available to application administrators for different levels of service. Different levels can be applied down to file level granularity, so performance sensitive individual files can be placed on performant tiers, while less critical data can be located on lower cost sensitive tiers or all the way to a cold tier such as the cloud. Below, DataSphere's UI screen shows a five-level service catalog:



NAME	DESCRIPTION
default-objective	
Platinum	Highest Performance, Lowest Latency, Low Capacity
Gold	Highest Performance, Low Latency, Medium Capacity
Silver	Mid-Performance, Medium Latency, Large Capacity
Bronze	Slower Performance, High Latency, Various Capacity
Cloud	Slowest Performance, Highest Latency, Huge Capacity

Figure 4. Objectives can be used to create a menu of service level catalogs for data.

Integrate with the Cloud as Just Another Storage Tier

When it comes to cloud archival, the challenge is determining what data can be safely archived, and how to move that data once it is identified. Managing data by objectives allows IT to automatically identify data that meets the enterprise's criteria for cloud archival and move it between the cloud and on-premises storage, as needed.

Many archiving solutions move data using simple rules based on attributes like file creation date. These solutions are error prone, can impact productivity, and require IT intervention to fix. DataSphere makes decisions based on actual client, application, and user access, and can retrieve files automatically if they are needed again.

In addition to archival, companies are beginning to use the cloud as a store for their backup data, but restores can be costly due to the bandwidth charges associated with retrieving data from the cloud provider. Objective-based management can retrieve data granularly, at the file level, making it possible to restore and rehydrate just the file that is needed, without IT intervention, minimizing cloud bandwidth charges.

DataSphere includes many features that enable enterprises to increase the value of the cloud. WAN optimization is built-in, as all data that is sent to the cloud is automatically de-duplicated and

compressed. During transfer a secure link is used to ensure that data remains safe. In addition, data movement is not limited to one cloud. Mobility between clouds or back to on-premises is as simple as a single click or client data access to trigger data transfer, and rehydration to the target store. Data mobility between the cloud and on-premises storage is automatic and transparent to the user. Files that require performance when opened are automatically transferred back ensure application needs are met to maintain compliance with your objectives.

Automatically Adapt to Evolving Business Demands

Enterprises have long struggled to control budgets while scaling with data growth, but more complex uses of data are now making it critical to contain costs by managing data in real-time, using objectives to leverage storage resources more efficiently and effectively.

For example, more companies are looking to gain insight through business intelligence and data warehousing (BI and DW) applications that analyze highly transient data and mine old data. Since companies have difficulty determining whether data might become valuable in the future, they rarely delete it. Some organizations are even using their Big Data platform as a Backup/DR/archive repository, which compounds the amount of data stored in an enterprise.

As the amount of mostly unused data companies are storing continues to increase, objectives enable IT to leverage cloud or object storage to contain storage sprawl created by big data. DataSphere uniquely allows data to remain accessible by applications and can seamlessly promote data to other storage tiers that suit an application's data demands. Importantly, DataSphere will retrieve only the files that the application is accessing, minimizing cloud bandwidth charges. DataSphere delivers these capabilities without companies having to modify applications to use cloud storage.

Overcome the Cost and Waste of Overprovisioning

Many companies have data that goes cold quickly as it ages. For example, cell phone billing for telecommunication companies typically go cold after a 30 to 45-day billing cycle and is rarely accessed again. In addition, many applications have cyclic demands. For example, payroll might need higher performance once a month, but because IT can't easily move data from a capacity tier to one with higher performance, in both situations, data is typically left on storage that meets the application's peak needs, no matter how short that window of time may be.

Global enterprises are also challenged by availability requirements that dictate that many apps must be available 24/7/365 for both internal and external users. This means that IT is challenged to work within very short windows of time to archive data without affecting business. To work around the interconnected challenges of performance and uptime, many companies simply buy far more than enough capacity so data won't have to be moved over its lifetime.

This kind of waste, called overprovisioning, can be incredibly costly, particularly when it comes to the overbuying of high performance storage. Eventually, even this capacity runs out, requiring a massive project to archive data, upgrade storage, or both. Managing by objectives with DataSphere solves the need to overprovision by ensuring the right data is on the right resource to meet cyclical business cycles. In addition, DataSphere enables admins to view individual and aggregate resources and deploy more performance or capacity, as needed, in minutes. Once added, DataSphere automatically rebalances data across the new resources, according to the data's objectives, and thereby increasing utilization.

While overprovisioning waste creates a surplus of unused capacity, storing data on the wrong resource for its needs can also result in a great deal of overspending on storage, particularly when idle data gets stored on high performance storage. According to IBM, “approximately 75 percent of the data stored is typically inactive, rarely accessed by any user, process or application. An estimated 90 percent of all data access requests are serviced by new data—usually data that is less than a year old.”ⁱ

Managing by objectives with DataSphere solves this problem by automatically moving cool data to more cost-effective, higher capacity resources, enabling enterprises to reclaim up to 75% of their expensive performance storage capacity. DataSphere can also ensure performance sensitive, highly active data is stored on premium storage (such as flash based arrays) to increase application throughput (transactions per second or number of simultaneous users serviced). This decreases the need for additional application licenses and hardware and gives IT more effective compute capacity per application.

Automate Agility and Response to Changing Business Needs

DataSphere gives petabyte-scale enterprises the ability to automate the movement of data according to business and IT objectives, from creation to archival, including the integration of public clouds as an active archive. DataSphere also automates core management tasks, making it easy for companies to maximize storage efficiency and cost savings, while protecting performance and protection to meet required service levels.

ⁱ [The fundamentals of data lifecycle management in the era of big data](#), 2013.