Red Hat
OpenShift
Data Foundation

Multicloud object gateway object and namespace buckets

Overview:

- Deploy Red Hat OpenShift Data Foundation and multicloud object gateway for data access that is fast, efficient, and secure.
- Simplify data management between multiple on-premise or cloud-based data resources.
- Flexibly span multiple back-end data resources with granular control over issues like data protection and redundancy.
- Provide a consistent experience for developers and data administrators.
- Deploy multicloud object gateway with Red Hat OpenShift Data Foundation or as a standalone configuration.

Adapting to growing object storage needs

Object storage provides massive scalability and enhanced metadata search capabilities essential for managing large amounts of data, typical in cloud-native environments. Modern cloud-native applications often feature an object storage back end. Object storage is also increasingly popular for edge and core datacenter platforms, providing an economically viable solution to address expanding data growth.[1]

Managing multiple cloud-based data resources can quickly become complicated. Organizations now routinely use numerous data resources across multiple public cloud providers and employ a mixture of on-premise and cloud storage. Just keeping track of data resources and credentials can be complex and a source of avoidable human error. With cloud-native applications built on microservices, individual workloads often need to access multiple data sources that may reside in more than one cloud. Replication of large data sets between sites is not always possible or desirable.

Red Hat® OpenShift® Data Foundation and its integral multicloud object gateway allow storage administrators to address these issues while dramatically reducing complexity. As a lightweight object storage service for Red Hat OpenShift, multicloud object gateway delivers simplified data management, flexibility across multiple storage back ends, and a consistent experience for both data consumption and management.

Red Hat OpenShift Data Foundation

Running as a Kubernetes service, OpenShift Data Foundation is persistent software-defined storage integrated with and optimized for Red Hat OpenShift. Deployed and managed through the Red Hat OpenShift administrator console, the platform includes Ceph® petabyte-scale persistent cloud storage, the Rook Kubernetes storage operator, and Noobaa multicloud object gateway technology. OpenShift Data Foundation runs anywhere Red Hat OpenShift does—on-premise or in hybrid cloud environments.

Red Hat OpenShift and OpenShift Data Foundation are available on all major public cloud platforms. For a complete on-premise solution, Red Hat OpenShift Platform Plus provides all the Red Hat OpenShift components organizations need at a lower cost than purchasing individual application subscriptions. Red Hat OpenShift Platform Plus includes subscriptions for both Red Hat OpenShift Container Platform and OpenShift Data Foundation, as well as:

- Red Hat Advanced Cluster Management for Kubernetes
- Red Hat Advanced Cluster Security for Kubernetes

1 Red Hat video. Data explosion for high-performance computing, accessed 4 January, 2022.

- Red Hat OpenShift Data Foundation Essentials
- Red Hat Quay

Multicloud object gateway

Multicloud object gateway is a high-level data service that enables object data access and data flow to and from different "buckets" across hybrid and multicloud environments. It uses policies to define data services behavior, yielding:

- Flexible access across multiple back ends (both on-premise and in a public cloud setting).
- A consistent experience everywhere.
- Simplified management of diverse data resources.

Mulicloud object gateway is installed natively with OpenShift Data Foundation. Beginning with OpenShift Data Foundation 4.9, administrators can also elect to deploy multicloud object gateway in a standalone configuration. This ability offers deployment flexibility and can help reduce resource consumption.

As shown in Figure 1, multicloud object gateway exposes a consistent Amazon Simple Storage Service (S3) endpoint to developers and applications. On the back end, multicloud object gateway provides different kinds of policy-driven data buckets:

- **Object buckets** contain object data and offer data safety with available mirroring, spreading, and tiering approaches.
- **Namespace buckets** provide data federation capabilities, letting administrators and data managers organize, configure, and manage diverse data resources.
- **Object Bucket Claims** dramatically simplify data creation and management for developers, allowing the dynamic creation of object buckets for workflows.

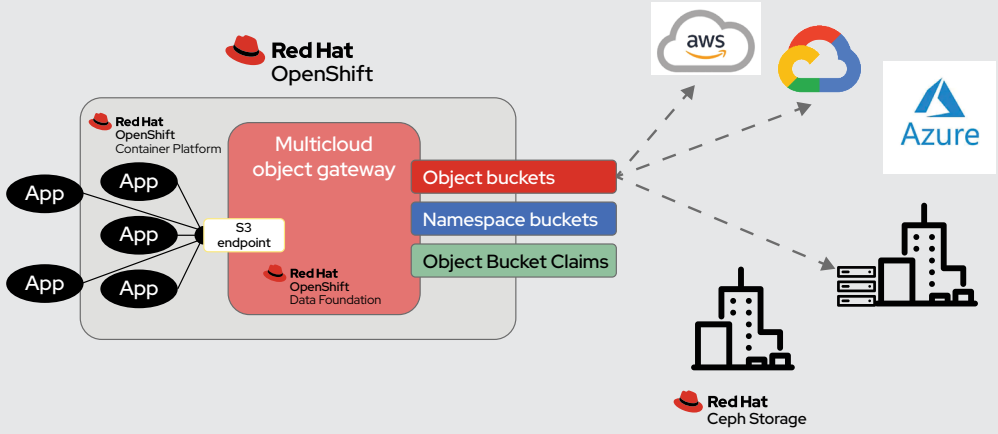

Figure 1. Multicloud object gateway supports multiple policy-defined data bucket types across on-premise or cloud-based data resources

Data buckets

Data buckets in multicloud object gateway provide flexible data safety for data resources across multiple backing stores. When an application writes data, it ends up in a data bucket with advanced data safety features as well as optional replication across multiple infrastructures. This backing store can exist within a public cloud, on-premise infrastructure, or both environments (Figure 2).

Data buckets provide extensive data storage functionality, offering:

- Deduplication.
- Compression.
- Encryption.
- Fragmentation into smaller chunks.
- Replication across different backing stores.

Administrators can dynamically choose the back-end storage location where data is written. Data can be replicated across multiple back-end infrastructure locations and read from a single S3 endpoint location. Data can still be read from other back-end locations when a replicated back-end location is no longer available.

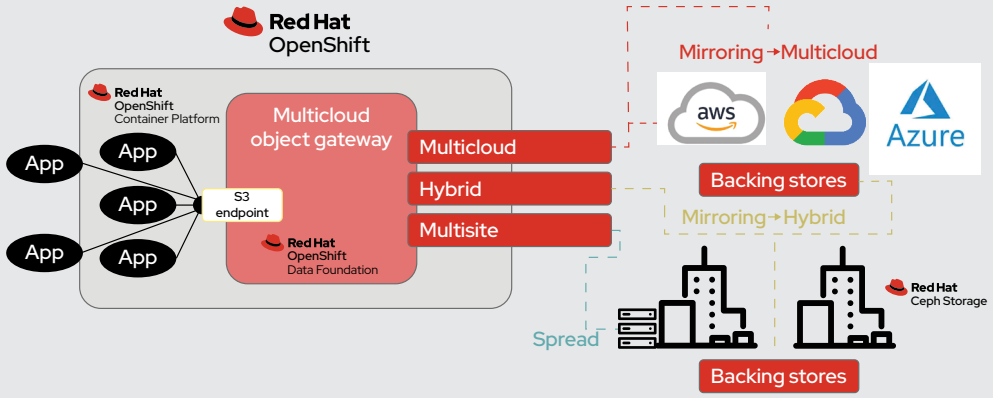

Figure 2. Object buckets provide data safety across hybrid, multicloud, and multisite data resources

The multicloud object gateway actively manages data buckets with metadata determining where data resides. All data in data buckets are chunked into small data particles and cannot be accessed without multicloud object gateway intervention, providing an extra layer of data security. For instance, when data is replicated across multiple back-end locations, metadata retained by multicloud object gateway is aware of the actual storage location. This capability is essential as the smaller data chunks are stored encrypted, preventing third parties from accessing the data without a multicloud object gateway. Data managed through multicloud object gateway cannot be accessed independently through the public cloud provider.

Object Bucket Claims

Similar in concept to persistent volumes (PVs), multicloud object gateway supports Object Bucket Claims that create object buckets along with necessary credentials for access (Figure 3). This feature allows developers to easily create storage when they need it without having to consult storage administrators. Object Bucket Claims automatically delete themselves when they are no longer required.

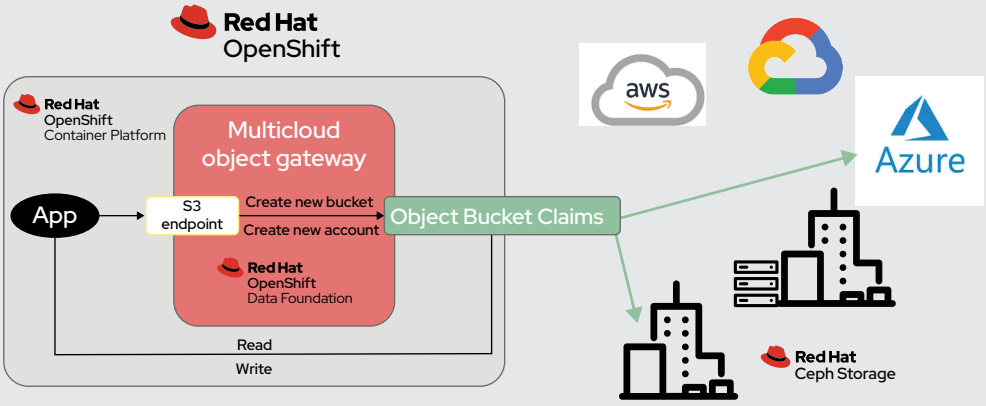

Figure 3. Object Bucket Claims dynamically create object buckets and necessary credentials, easing storage management for developers

Object Bucket Claims use an application's YAML file to refer to the object storage class. They then create a new object bucket and application account with permissions (access key, secret key). The object bucket information is passed dynamically to the application using environment variables, allowing the application to immediately write to and read from the bucket.

Namespace buckets

Namespace buckets provide an aggregated view of multiple buckets and potentially diverse data resources (Figure 4). Namespace buckets do not contain data or metadata. There is no data movement, no replication, no mirroring. Instead, namespace buckets only point to one or more datasets, allowing storage administrators to decide how they wish to view and organize their data resources. Namespace buckets point to data buckets, which could either be new empty data buckets or existing buckets that already contain data.

Namespace buckets allow an easier way to engage with data across multiple locations without needing to copy back-end data. Data can stay in place and be made available in alternative locations without time-consuming copying of data. A namespace bucket can hold multiple datasets across different infrastructure locations (e.g., Red Hat Ceph Storage RADOS Gateway, Amazon AWS, Microsoft Azure, IBM Cloud, and on any other S3 compatible bucket)—making all that data accessible from within a single access location. Multiple namespace buckets can exist within a single OpenShift Data Foundation cluster or across multiple clusters. Each namespace bucket can point to the same or different target buckets.

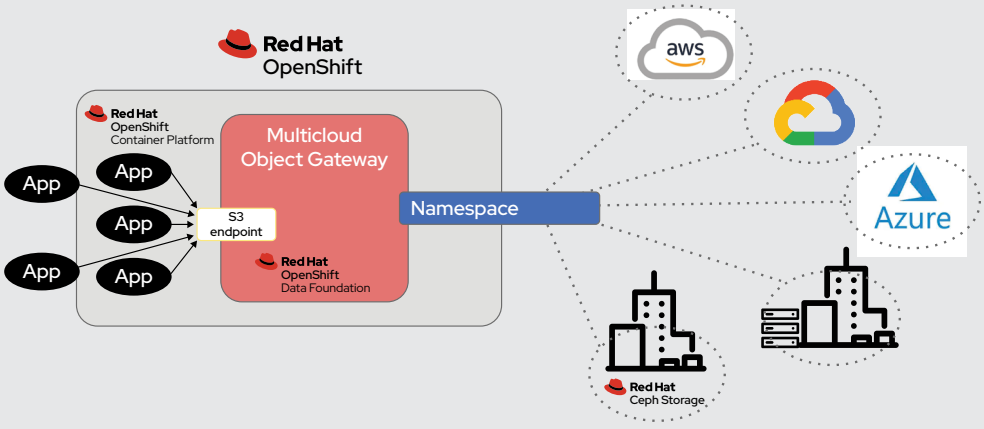

Figure 4. Namespace buckets provide an aggregated view of object buckets across diverse infrastructure

Flexible policies are available through namespace buckets. For example, a namespace bucket policy can determine which bucket(s) data will be read from and which bucket will be the default location for write operations. There will always be one location for write operations, but applications can access multiple reader buckets simultaneously.

Namespace buckets provide additional functionality that adds essential features for administrators and storage managers.

- **Multicloud object gateway caching.** Multicloud object gateway supports object data caching (Figure 5), functionality that enriches the namespace bucket concept with additional intelligence. Staging multiple datasets across different infrastructure locations can be problematic, particularly when large datasets are involved. Accessing a dataset multiple times typically means the same remote bucket data would have to be accessed repeatedly. This excess network traffic can translate into egress costs when using public cloud resources.

  To mitigate these costs, multicloud object gateway offers a data cache mechanism. Organizations can create multiple projects with simple namespaces pointing to their relevant datasets—without having to worry about data being copied repeatedly, with each project accessing it individually. When an application within a project or cluster accesses a namespace bucket, the multicloud object gateway will cache the remotely retrieved data. When that same data is requested again by the application or another, it will then be served locally from the multicloud object gateway's cache, preventing repetitive fetching of remote bucket data.

  Remote data remains where it resides and does not need to be copied for an application to access the data. After the data processing has finished, data does not have to be moved back to its original location, saving on bandwidth, egress costs, and other specific data administration.

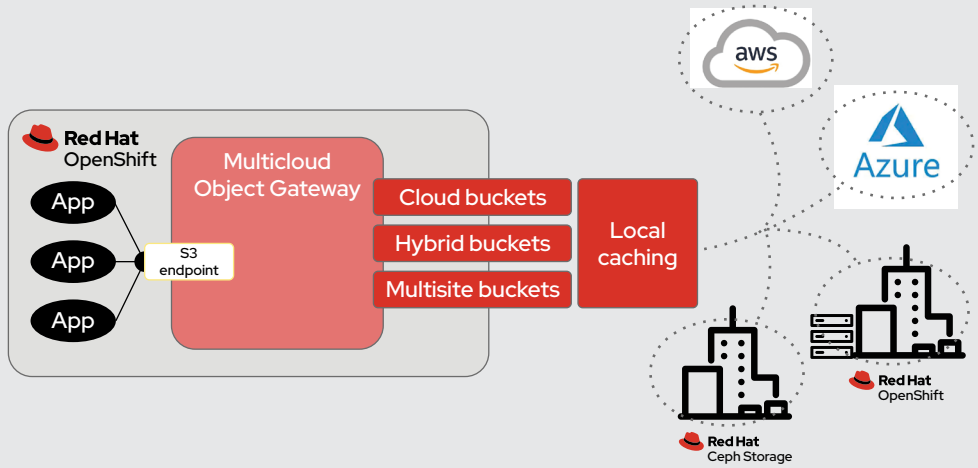

Figure 5. Multicloud object gateway caching avoids copying data and saves on public cloud egress costs

- Data infrastructure location. Local caching does not imply it is only applicable for on-premise use. The functionality is relevant regardless of where organizations want to place their data. The caching functionality can also run within public cloud environments while data stays on-premise. This ability can be beneficial for temporary workloads where required compute power is not immediately available on-premise. In this case, the cluster can run in a public cloud environment and can be removed after finishing the temporary workload without needing to copy data over the public network.

- Namespace bucket replication. Namespace bucket replication (Figure 6) is another functional layer that lets organizations stage data across different locations for multiple purposes, including disaster recovery. Unidirectional or bidirectional replications are possible options. With namespace bucket replication, data can be accessible across multiple public cloud service providers. For example, namespace bucket replication might be helpful if an organization needed to run specific data analytics processes within a public cloud infrastructure environment while their data remained on-premise.

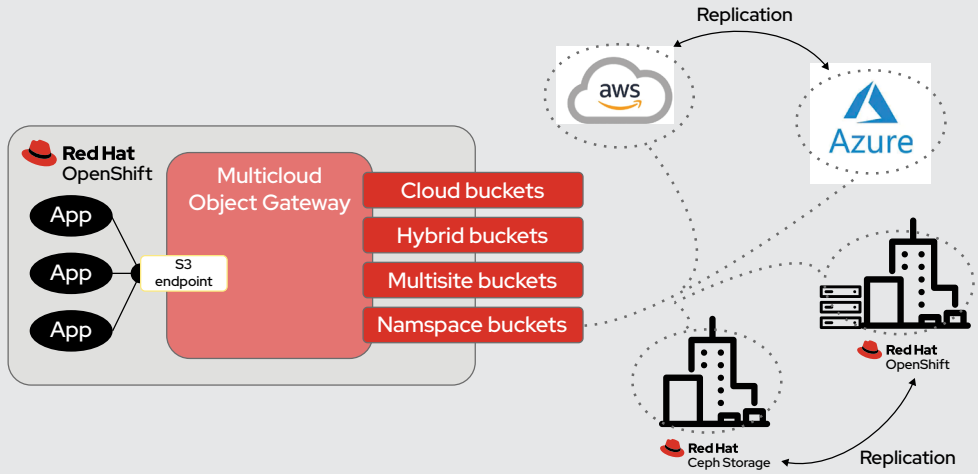

Figure 6. Namespace bucket replication provides for disaster recovery and other management convenience

Use case examples

Organizations need flexibility for running their applications, but that need for flexibility also extends to their data services.

Use case 1: Data collaboration and resiliency with data bucket replication

In this use case, a company maintains multiple infrastructure locations across public cloud environments and on-premise locations. Having specific data accessible at just one location is not practical and involves business challenges and continuity risks. Presenting relevant and valuable business data at multiple locations would be helpful, particularly if copying data back and forth could be avoided.

Figure 7 shows how data bucket replication could work across two independent sites. Each site produces and consumes its own data. The data is replicated to ensure availability in case of disconnection. The entire dataset is available on both sites.

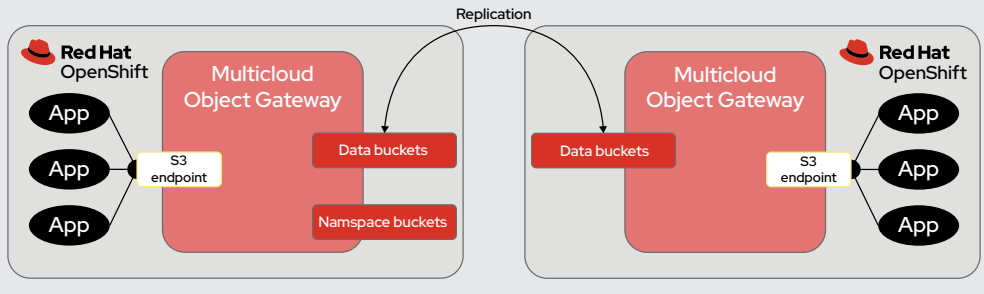

Figure 7. Data bucket replication helps ensure access and availability

Use case 2: An hybrid cloud solution using namespace buckets

In this example, an organization might suddenly need data analytics capabilities without sufficient available on-premise compute resources. The company could be a retail organization that maintains customer data inside its on-premise datacenter but wants to run artificial intelligence (AI) analytics on the data. In this scenario, obtaining public cloud resources is the logical next step. However, simply transferring the required compute processes does not resolve the entire issue. The data to be processed must also be available on the public cloud infrastructure.

As shown in Figure 8, namespace buckets could be used in conjunction with a data bucket that contains the customer data. The dataset inside the data bucket is made available to the AI analytics software in the public cloud environment through a namespace bucket. Using this approach eliminates the need to copy over the entire dataset first before beginning the analytics process. In addition, the original dataset remains untouched.

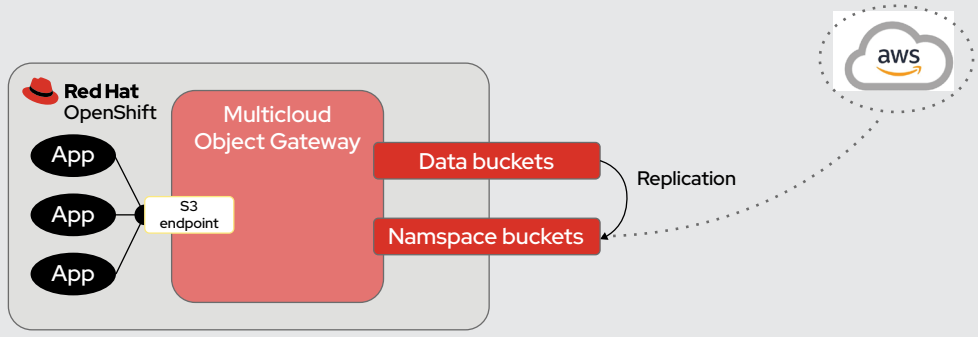

Figure 8. Using namespace bucket replication to enable hybrid processing while data remains on-premise

Learn more

Read more about Red Hat OpenShift Data Foundation and watch Understanding multicloud object gateway (video), which offers a detailed explanation of the multicloud object gateway included with OpenShift Data Foundation.

Contact us if you would prefer to talk to a Red Hat representative, and we will get in touch.

About Red Hat

Red Hat is the world's leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers develop cloud-native applications, integrate existing and new IT applications, and automate and manage complex environments. A trusted adviser to the Fortune 500, Red Hat provides award-winning support, training, and consulting services that bring the benefits of open innovation to any industry. Red Hat is a connective hub in a global network of enterprises, partners, and communities, helping organizations grow, transform, and prepare for the digital future.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

redhat.com
#F30739_0622

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com