Concepts

Keystone

NetApp
February 10, 2026

Table of Contents

Concepts

Keystone STaaS services

Metrics and definitions used in Keystone

The NetApp Keystone STaaS service uses several terms to measure metrics. You might want to learn more about these terms as you use Keystone.

The following terms and definitions are used within the Keystone STaaS service to measure metrics:

- Capacity: Measured in GiB, TiB, and PiB.
- IOPS: Number of input/output operations processed per second.
- Service availability
- Durability in accurate data access
- Latency and speed

Metrics measurement

- **Capacity measurement in GiB, TiB, and PiB**: Measurements of data storage capacity using base of 1024 (1 GiB = 1024^3 bytes, 1 TiB = 1024^4 bytes, and 1 PiB = 1024^5 bytes).
- **Operations counter chart in IOPS**: The protocol operations per second, requested by the application.
- **Availability**: Measured as a percentage of the number of I/O requests successfully responded to by the service, divided by total number of I/O requests made to the service. This is measured at the service demarcation in a month and does not include the scheduled service downtime or unavailability of the facilities, network, or other services provided by the customer.
- **Durability**: Percentage of data accessed without loss of fidelity, excluding customer-caused deletion or corruption.
- **Latency**: Time to service an I/O request received from a client, measured at the service demarcation (storage controller I/O port).

Performance metrics

The following performance metrics are applicable to unified and block-optimized services:

Unified services:

- **IOPS**: For ONTAP 9.16.1 with NFS, each performance level instance supports random access with a 70% read and 30% write ratio, an 8 KB block size, and a latency of 1 ms (4 ms for Standard).
- **Throughput**: For ONTAP 9.16.1 with NFS, each performance level instance supports sequential access with 100% read and a 32 KB block size.

Block optimized services:

- **IOPS**: For ONTAP 9.16.1 with FCP, each performance level instance supports random access with a 70% read and 30% write ratio, an 8 KB block size, and a latency of 1 ms.
- **Throughput**: For ONTAP 9.16.1 with FCP, each performance level instance supports sequential access with 100% read and a 64 KB block size.

Supported storage in Keystone

Keystone STaaS service supports unified, block-optimized, and object storage of NetApp, and Cloud Volumes ONTAP.

The supported storage options are:

- **Unified storage**: Includes both file, block, and S3 object storage, available on NetApp ONTAP AFF as well as FAS systems.
- **Block-optimized storage**: Includes block storage available on NetApp ONTAP ASA systems.
- **Object storage**: Includes object storage available on NetApp StorageGRID systems.

Keystone STaaS provides standard and optional services for your storage.

Keystone STaaS standard services: Standard services are included within the base subscription and are not charged separately.

Keystone STaaS add-on services: These are optional, chargeable services that provide additional utilities and benefits on top of standard Keystone STaaS subscription services.

Keystone STaaS services can be used at the same time. For example, a cloud storage subscription can have the same term as unified, block-optimized, and object storage subscriptions. A cloud service can be included at any point during the service term of an existing storage subscription. However, if you do not plan to renew an existing unified, block-optimized, or object storage subscription, a cloud storage subscription cannot be added during the last 90 days of the subscription.

Services for unified, block-optimized, and object storage

Keystone STaaS services for unified, block-optimized, and object storage, support multiple features and protocols, and described in the following table:

Storage	Platform	Protocols	Supported features
Unified storage	ONTAP	NFS and CIFS	Supports all ONTAP One features
Block optimized storage	ONTAP	FC and iSCSI	Supports all ONTAP One features
Object storage	StorageGRID	S3	Supports all ONTAP One features

To learn more about ONTAP One, refer to ONTAP licensing overview and ONTAP One: The full power of ONTAP, now all in one.

Services for cloud storage

Keystone STaaS provides cloud storage services. Keystone STaaS supports Cloud Volumes ONTAP data management capabilities on Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform.

Hyperscalar-based compute, storage, and network services required by Cloud Volumes ONTAP are not provided by NetApp as a part of Keystone STaaS subscriptions; these subscriptions need to be procured directly from hyperscalar cloud service providers.

Supported storage capacities in Keystone

The NetApp Keystone STaaS service supports several types of storage capacities. Understanding these different capacity terms can help as you use Keystone.

Logical capacity

This is the amount of storage capacity required to store user data before any data efficiencies provided by the storage array are applied.

Committed capacity

The minimum logical capacity billed each month during the subscription:

- Capacity is committed to each performance service level.
- Committed capacity and additional performance service levels can be added during the term.

Changes to committed capacity

During the tenure of a subscription, you can change the committed capacities. However, there are certain preconditions:

- The committed capacity can be decreased based on certain conditions. For information, see Capacity reduction.
- The committed capacity cannot be increased 90 days prior to the expiry of your subscription, unless the subscription is to be renewed for an additional 12-month term.
- You can request changes to committed capacity through the Console or from your Keystone Success Manager (KSM).
  For information about requesting changes, see NetApp Keystone support.

Consumed capacity

Consumed capacity refers to the capacity (in TiB of storage) currently being consumed on the service. It is calculated differently based on the storage type:

- **Unified or block-optimized storage**: Consumed capacity is calculated based on the type of capacity (either logical or physical) selected during the ordering process. The calculation is performed per performance service level instance.
  a. **Logical capacity**: It is the sum of:
     - Metered logical capacity, before storage array data efficiencies, to store all instances and types of customer data, such as copies, mirrored copies, versions, and clones.
     - Physical capacity used to store metadata and differential data of snapshots and certain clones.
     - Any thick-provisioned physical capacity.
  b. **Physical capacity**: It is the sum of:
     - Metered physical capacity, after storage array data efficiencies, to store all instances and types of

customer data, such as copies, mirrored copies, versions, clones.

  - Physical capacity used to store metadata and differential data of snapshots.
  - Any thick-provisioned physical capacity.

- **Object storage**: Consumed capacity is calculated as the amount of metered physical capacity used to store all instances and types of customer data across all nodes. This calculation is based on the information lifecycle management (ILM) policies configured.
- **Cloud Volumes ONTAP**: Consumed capacity is calculated as the amount of metered provisioned capacity of all Cloud Volumes ONTAP volumes.

Burst capacity

The NetApp Keystone STaaS service enables you to use additional capacity on top of the committed capacity for a performance service level. This is referred to as the burst capacity usage.

Note these points:

- Burst capacity is agreed upon in the Keystone agreement. It is usually set up to 20% above the committed capacity, per performance service level instance, with additional options available to select burst capacity limits of 40% or 60% of committed capacity.
- Burst capacity consumption is invoiced at the same rate as the committed capacity corresponding to the selected performance service level.
- Keystone STaaS services provide a burst waiver period of 60 days from the start date.

Billed capacity

Monthly bill = (committed capacity [TiB] * committed rate [$/TiB]) + (daily average provisioned burst capacity [TiB] * burst rate [$/TiB]). The monthly bill contains a minimum charge based on the committed capacity.

The monthly bill varies beyond the minimum charge based on daily average burst capacity consumption.

Performance service levels in Keystone

Keystone STaaS offers data storage capacity at pre-defined performance service levels. Each volume managed by Keystone services is associated with a performance service level.

A subscription can have multiple rate plans and each rate plan corresponds to a performance service level. Each rate plan has a committed capacity per performance service level.

A performance service level can have multiple instances, with each instance representing a separate storage array assigned to that performance service level in the customer's environment. Each performance service level is defined by input/output operations per second (IOPS), throughput (GBps), and latency (ms), with these metrics measured and applied per performance service level instance.

You select performance service levels based on your storage environment, and storage and consumption needs. The base performance service levels are available for you by default. Specific performance service levels are additionally available, if you have opted for add-on services.

A detailed service description for NetApp Keystone STaaS performance service levels is available here.

The base performance service levels for the supported storage types, unified, block-optimized, object, and cloud services are described in the following sections:

Performance service levels for unified storage

Supported protocols: FC, iSCSI, NFS, NFSv4/RDMA, NVMe/FC, NVMe/TCP, SMB, S3

Performance service level (all specifications per performance service level instance)	**Extreme**	**Premium**	**Standard**	**Value**
Sample workload types	AI/ML, HPC, InMem DB	Analytics, EDA, OLTP	OLAP, IoT, Containers	Backup, Archive
Maximum IOPS[1]	1M	550K	500K	NA
Maximum GBps	40	20	20	NA
Target 90th percentile latency	<=1 ms	<=1 ms	<=4 ms	>4 ms
Minimum committed capacity	50 TiB	50 TiB	100 TiB	100 TiB
Incremental committed capacity increase	25TiB			
Committed and metered capacity type	Logical or physical			

Performance service levels for block-optimized storage

Supported protocols: NVMe/TCP, NVMe/FC, FC, iSCSI

Performance service level (all specifications per performance service level instance)	**Extreme**	**Premium**
Sample workload types	SAP HANA, Oracle, MS SQL Server, EPIC	
Maximum IOPS[1]	850K	450K
Maximum GBps	65	25
Target 90th percentile latency	<=1 ms	<=1 ms
Minimum committed capacity	50 TiB	50 TiB
Incremental committed capacity increase	25TiB	
Committed and metered capacity type	Logical or physical	

[1]Mutually exclusive targets. Actual performance may differ based on various factors, including the operating system version, hardware, workload type, and number of concurrent operations.

More on performance service levels for unified and block-optimized storage

The base performance service level metrics depend on the following conditions:

- The performance service levels support ONTAP 9.8 and later.
- For unified storage,
  - **IOPS**: For ONTAP 9.16.1 with NFS, each performance level instance supports random access with a 70% read and 30% write ratio, an 8 KB block size, and a latency of 1 ms (4 ms for Standard).
  - **Throughput**: For ONTAP 9.16.1 with NFS, each performance level instance supports sequential access with 100% read and a 32 KB block size.
- For block-optimized storage,
  - **IOPS**: For ONTAP 9.16.1 with FCP, each performance level instance supports random access with a 70% read and 30% write ratio, an 8 KB block size, and a latency of 1 ms.
  - **Throughput**: For ONTAP 9.16.1 with FCP, each performance level instance supports sequential access with 100% read and a 64 KB block size.
- Latency does not include the following:
  - application or host latency
  - customer network latency to or from the controller ports
  - overheads associated with the data transfer to the object store in case of FabricPool
- Latency values are not applicable to MetroCluster write operations. These write operations are dependent on the distance of remote systems.
- *Expected IOPS* is targeted for FabricPool only if the tiering policy is set to "none" and no blocks are in the cloud. *Expected IOPS* is targeted for volumes that are not in a SnapMirror synchronous relationship.

Performance service levels for object storage

Supported protocol: S3

Performance service level	Standard	Value
Minimum committed capacity per order	200 TiB	500 TiB
Incremental committed capacity increase	25 TiB	100 TiB
Committed and metered capacity type	Physical	

Cloud storage

Supported protocols: NFS, CIFS, iSCSI, and S3 (AWS and Azure only)

Performance service level	Cloud Volumes ONTAP
Minimum committed capacity per order	4 TiB

Incremental committed capacity increase	1 TiB
Committed and metered capacity type	Logical

- Cloud native services, such as compute, storage, networking, are invoiced by cloud providers.
- These services are dependent on cloud storage and compute characteristics.

Related information

- Supported storage capacities
- Metrics and definitions used in Keystone Services
- Keystone pricing

Capacity requirements for Keystone performance service levels

The capacity requirements for Keystone STaaS performance service levels differ between the unified, block-optimized, object, or cloud storage offerings supported by the Keystone STaaS subscription.

Minimum capacity requirements for unified and block-optimized storage

You can see the minimum capacity and incremental capacity allowed per subscription for unified and block-optimized storage in the following tables:

Unified storage

Capacity	Extreme	Premium	Standard	Value
Minimum capacity [in TiB]	50		100	
Incremental capacity (and in multiples) allowed at start of subscription [in TiB]	25			
Incremental capacity (and in multiples) allowed as add-on during subscription [in TiB]	25			

Block optimized storage

Capacity	Extreme	Premium
Minimum capacity [in TiB]	50	

Incremental capacity (and in multiples) allowed at start of subscription [in TiB]	25
Incremental capacity (and in multiples) allowed as add-on during subscription [in TiB]	25

The minimum capacity for each performance service level is the same across all Keystone sales.

Minimum capacity requirements for object storage

You can see the minimum capacity requirements for object storage in the following table:

Capacity	Standard	Value
Minimum capacity [in TiB] per order	200	500
Incremental capacity (and in multiples) allowed at start of subscription [in TiB]	25	100
Incremental capacity (and in multiples) allowed as add-on during subscription [in TiB]	25	100

Minimum capacity requirements for cloud services

You can see the minimum capacity requirements for cloud services in the following table:

Capacity	Cloud Volumes ONTAP
Minimum capacity [in TiB] per order	4
Incremental capacity (and in multiples) allowed at start of subscription [in TiB]	1
Incremental capacity (and in multiples) allowed as add-on during subscription [in TiB]	1

Capacity adjustments

Learn more about capacity adjustments:

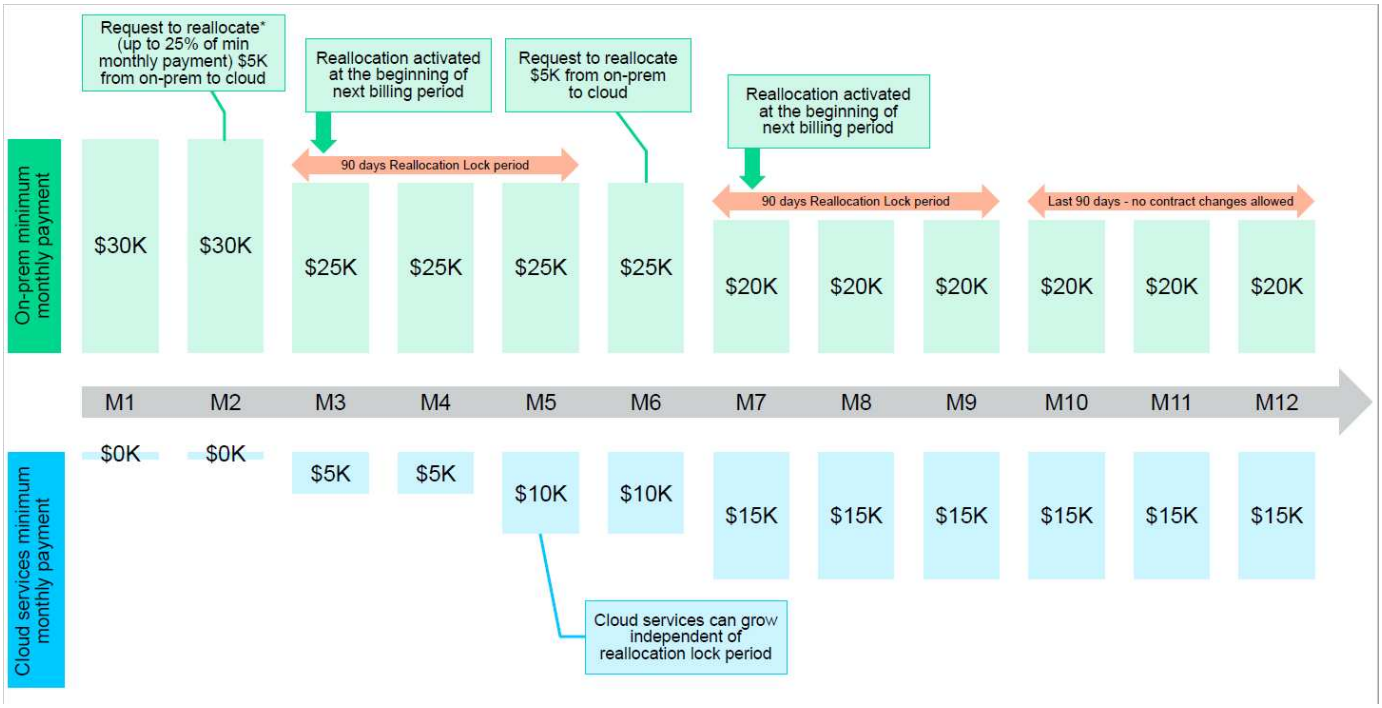
- Capacity can be added anytime during the term, except for the last 90 days of the contract term, in the increments per performance service level as described in the tables in the previous section. Addition of capacity or services is allowed within the last 90 days of the contract term as long as there is a consent of service renewal. Any addition in capacity, new service on-prem or cloud can co-term with the existing term. The invoice sent to you following the activation of the new services reflects the revised billing. Committed

capacity of cloud services cannot be reduced at any point during the subscription term. Meanwhile, committed capacity and committed spend on the on-premises services during the term of the contract can be reduced based on certain criteria as defined in the following section *Capacity reduction*.

- A burst capacity is available at each site, based on the Keystone agreement. Usually, it is set up to 20% above the committed capacity for a performance service level. Any burst usage is billed only for that billing period. If you have additional burst requirement greater than the capacity you agreed upon, contact support.
- Committed capacity can be altered during a contract term, only under certain conditions, as described in the following section *Capacity reduction*.
- Increasing capacity or changing to higher performance service level during a subscription term is allowed. However, moving from a higher performance service level to a lower performance service level is not permitted.
- Any change request in the last 90 days of the service term requires a renewal of the service for a minimum of one year.

Capacity reduction

Capacity reduction (annual) is applicable to the *Annual in Advance* payment model and on-premises only deployments. It is not available for cloud services or hybrid cloud services. It provides provision for on-premises capacity, which can be reduced by up to 25% per service level per subscription. This reduction is allowed once every year to be made effective at the beginning of the next annual billing period. On-premises service-based annual payments should be >=$200K anytime during the term in order to take advantage of capacity reduction. Because it is supported only for on-premises deployments, this billing model does not provide reallocation in spending from on-premises to cloud services. An example of annual capacity reduction is illustrated in the following image.



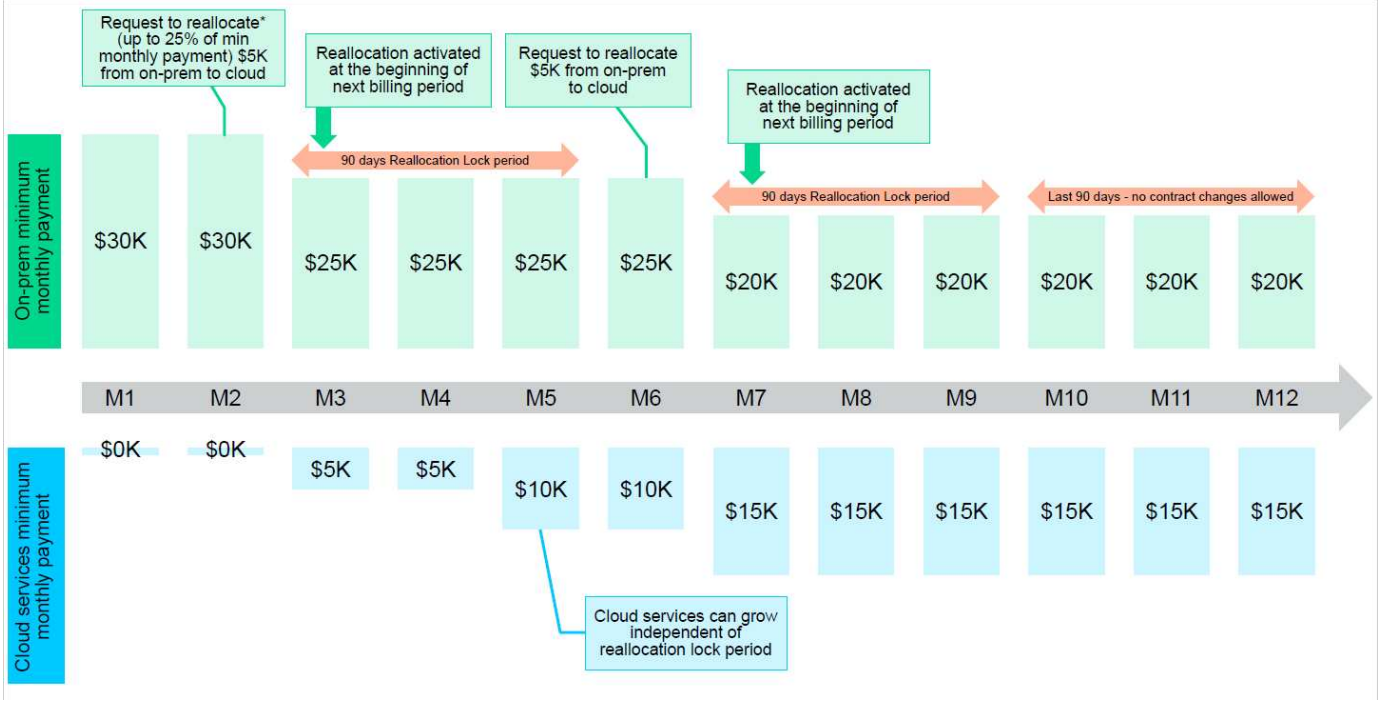
Quarterly spend reallocation

Keystone STaaS offers you the option to reallocate on-premises service spend to Cloud Volumes ONTAP spend.

Requirements and conditions at a subscription level:

- Applies only to monthly billing in arrear model.
- Applies only to subscriptions with 1, 2, or 3-year term commitments.
- Capacity for Cloud Volumes ONTAP and Cloud Backup service should be purchased through Keystone.
- Up to 25% of the existing on-premises, service-based monthly payments can be used for reallocation to cloud services.
- Reallocation requests are made effective only after 90 days from the previous activation date of the reallocation.
- Reallocation cannot be done from cloud services back to on-premises services.
- A request to reallocate should be formally submitted by the customer or partner to Keystone Success Manager (KSM) at least one week before the next billing cycle.
- New requests go into effect only from the consecutive billing cycle.

You can allocate a portion of your expenses towards your subscribed file, block, or object storage performance service levels to hybrid cloud storage services. Up to 25% of the Annual Contract Value (ACV) can be reallocated on a quarterly basis to Cloud Volumes ONTAP Primary and Cloud Volumes ONTAP Secondary services:



This table provides a set of sample values to demonstrate how the reallocation of expenses works. In this example, `$5000` from the monthly spend is reallocated to hybrid cloud storage service.

Before allocation	**Capacity (TiB)**	**Monthly designated expense**
Extreme	125	37,376
After reallocation	**Capacity (TiB)**	**Monthly designated expense**
Extreme	108	37,376
Cloud Volumes ONTAP	47	5,000

		37,376

The reduction is of (125-108) = 17 TiB of the capacity allocated for the Extreme performance service level. On spend reallocation, the allotted hybrid cloud storage is not of 17 TiB but an equivalent capacity that $5000 can purchase. In this example, for $5000, you can get 17 TiB on-prem storage capacity for the Extreme performance service level and 47 TiB hybrid cloud capacity of Cloud Volumes ONTAP performance service level. Therefore, the reallocation is with respect to the spend, not capacity.

Contact your Keystone Success Manager (KSM) if you want to reallocate expenses from your on-premises services to cloud services.

Add-on services

Advanced data protection add-on for Keystone

You can subscribe to the advanced data protection (ADP) add-on service with your Keystone STaaS subscription. While standard Keystone services include default data protection using SnapMirror, SnapVault, and Snapshot, this add-on service uses NetApp MetroCluster technology to ensure efficient data protection of your mission-critical workloads with a recovery point objective (RPO) of 0.

Keystone advanced data protection service can synchronously mirror data to a secondary site. In case of a disaster at the primary site, the secondary site can take over, without any loss of data. This feature uses the MetroCluster configuration between two sites to enable data protection. You can use the advanced data protection add-on services for only your file and block storage services. As a part of this add-on service, the `Advanced Data-Protect` performance service level is assigned to your subscription.

You can monitor consumption and health metrics for your MetroCluster configuration. For more information, see View consumption and health of your MetroCluster subscriptions.

Understand pricing

The advanced data protection add-on service is priced based on the committed capacity at each site. This helps determine the actual cost of advanced data protection service in $/TiB. The add-on charges apply to all capacities in your subscription, including source data, mirrored copies, and unmirrored data.

For the MetroCluster configuration, each site requires its own subscription, and you are charged for the committed capacity at each site independently.

Note the following:

- The service uses 100% of the committed capacity on associated storage as committed capacity.
- Charges apply to both primary and secondary sites, covering both source and destination clusters.
- Charges apply only to your file and block storage services.

Supported MetroCluster configurations

Keystone supports the following MetroCluster deployment scenarios:

Active/Passive configuration

In this configuration, primary data at one site is mirrored to a secondary site. For example, if 100 TiB of logical data is consumed at site A, it is replicated to site B. Both sites require identical subscriptions:

- **Subscription 1 (site A):** 100 TiB storage service + 100 TiB ADP
- **Subscription 2 (site B):** 100 TiB storage service + 100 TiB ADP

Active/Active configuration with full mirroring

In this configuration, both sites host primary data that is mirrored bidirectionally. For example, 100 TiB of primary data at site A is mirrored to site B, and 100 TiB of primary data at site B is mirrored to site A. Both sites require subscriptions for the combined capacity:

- **Subscription 1 (site A):** 200 TiB storage service + 200 TiB ADP
- **Subscription 2 (site B):** 200 TiB storage service + 200 TiB ADP

Active/Active configuration with partial mirroring

In this configuration, one site hosts mirrored data while the other site hosts both mirrored and unmirrored data. For example, 100 TiB of primary data at site A is mirrored to site B, while site B also hosts an additional 100 TiB of unmirrored data. The subscriptions differ based on the capacity at each site:

- **Subscription 1 (site A):** 100 TiB storage service + 100 TiB ADP
- **Subscription 2 (site B):** 200 TiB storage service + 200 TiB ADP

The following diagram shows these supported MetroCluster configurations:

Burst capacity add-on for Keystone subscriptions

You can opt for the burst capacity add-on service as part of your NetApp Keystone subscription. This service allows you to increase your burst capacity limits to 40% or 60% above your committed capacity, providing the flexibility to handle unexpected surges in workload demand.

Burst capacity refers to the additional storage capacity that can be utilized beyond the committed capacity of your subscription. It is measured and billed per performance service level. By default, your burst limit is set at 20% above the committed capacity. However, with this add-on service, you can increase the limit to 40% or 60%.

To change your burst limit to 40% or 60%, contact the NetApp Keystone support team.

To learn more about how burst capacity is billed, refer to Billing based on burst consumption.

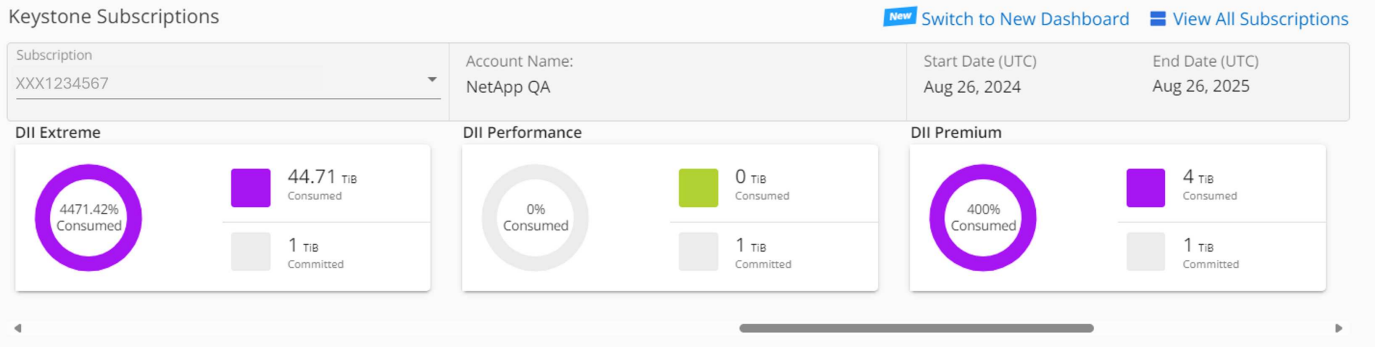
Data Infrastructure Insights add-on for Keystone

NetApp Data Infrastructure Insights (DII, formerly known as Cloud Insights) is an add-on offering for Keystone STaaS. The integration of this service with Keystone STaaS enhances the monitoring, troubleshooting, and optimization capabilities of Keystone-provided storage resources across public clouds and private data centers.

To learn more about Data Infrastructure Insights, refer to Data Infrastructure Insights documentation.

DII is available for both new and existing subscriptions. It can be integrated into a Keystone subscription as an add-on for the committed capacity. When DII is integrated into a Keystone subscription, it will have a corresponding performance service level for each base performance service level in the subscription. For example, Extreme maps to DII Extreme, Premium maps to DII Premium, and Performance maps to DII Performance. These mappings ensure that the DII performance service level aligns with the base performance service level of your Keystone subscription.

A view of DII performance service levels within a **Keystone Subscriptions** widget on the Digital Advisor dashboard:



Deployment of DII for Keystone

Customers can integrate DII for Keystone in two ways: either as part of an existing instance that monitors other non-Keystone environments, or as part of a new instance. It is the customer's responsibility to set up DII. If help is needed for setting up DII in a complex environment, the account team can engage NetApp Professional Services.

To set up DII, refer to Data Infrastructure Insights onboarding.

Note the following:

- If the customer is starting a new DII instance, it is recommended to begin with a DII free trial. To learn about this feature and the required startup checklist, refer to Feature Tutorials.
- For each site, an Acquisition Unit is required. To install an Acquisition Unit, refer to Install an Acquisition Unit. If the customer already has a DII instance and Acquisition Unit set up, they can proceed with configuring the data collector.
- For each storage hardware deployed, the customer must configure a data collector on the Acquisition Unit. To configure data collectors, refer to Configure Data Collectors. The required data collectors for Keystone storage, based on the underlying hardware, are as follows:

Storage hardware	Data collector
ONTAP Systems	NetApp ONTAP Data Management Software
StorageGRID	NetApp StorageGRID
Cloud Volumes ONTAP	NetApp Cloud Volumes ONTAP

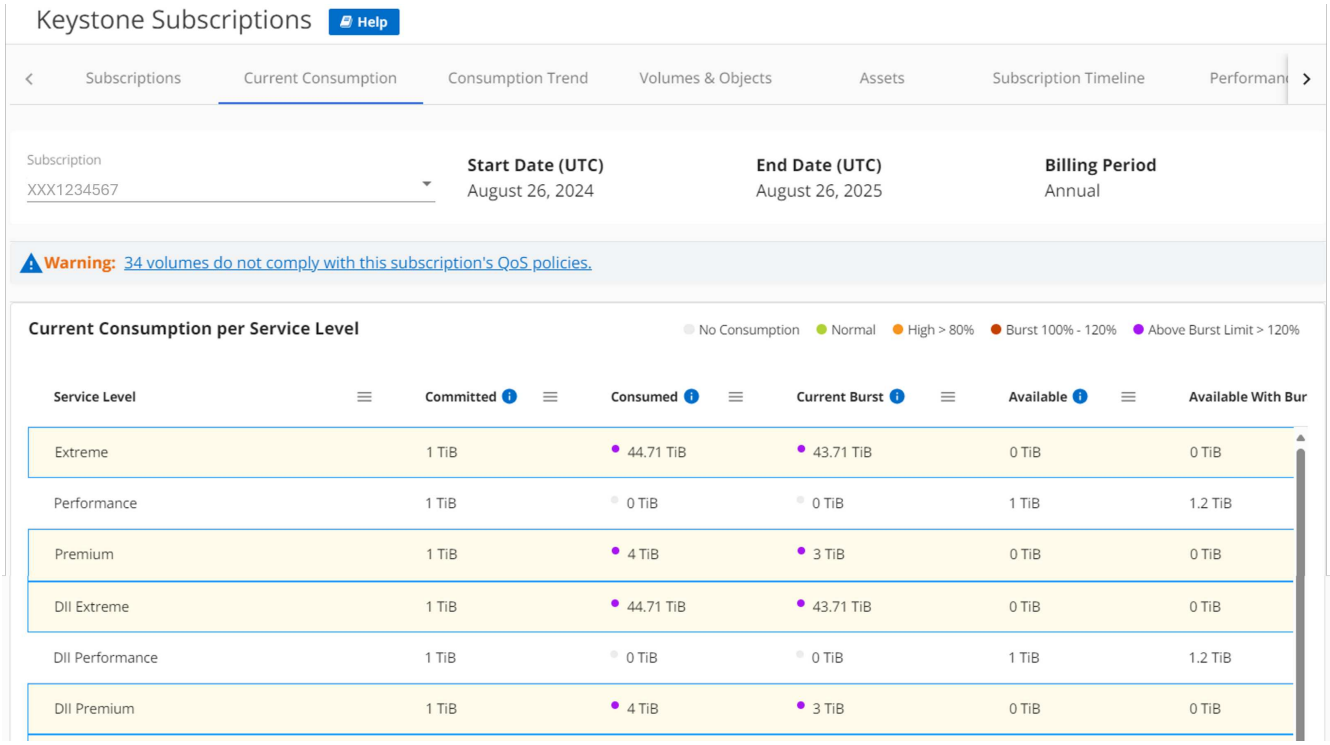
Once configured, the DII instance will begin monitoring the NetApp storage resources deployed as part of Keystone.

While DII offers extensive monitoring capabilities for the deployed hardware, it does not provide insights into your Keystone subscription, such as subscription usage or details on performance service levels. For subscription insights, see Keystone dashboard and reporting.

Billing and metering

The following outlines the billing and metering details of the DII add-on service:

- This service is measured in the same way as your main subscription. For example, if your main subscription includes 100 TiB of Extreme service and 100 TiB of Premium service, both measured on a logical usage basis, the add-on services, DII Extreme 100 TiB and DII Premium 100 TiB, are also measured on a logical usage basis. If your main subscription is measured on a provisioned basis, the add-on service is measured the same way for the same capacity. The add-on service will follow the same measurement method as your main subscription.

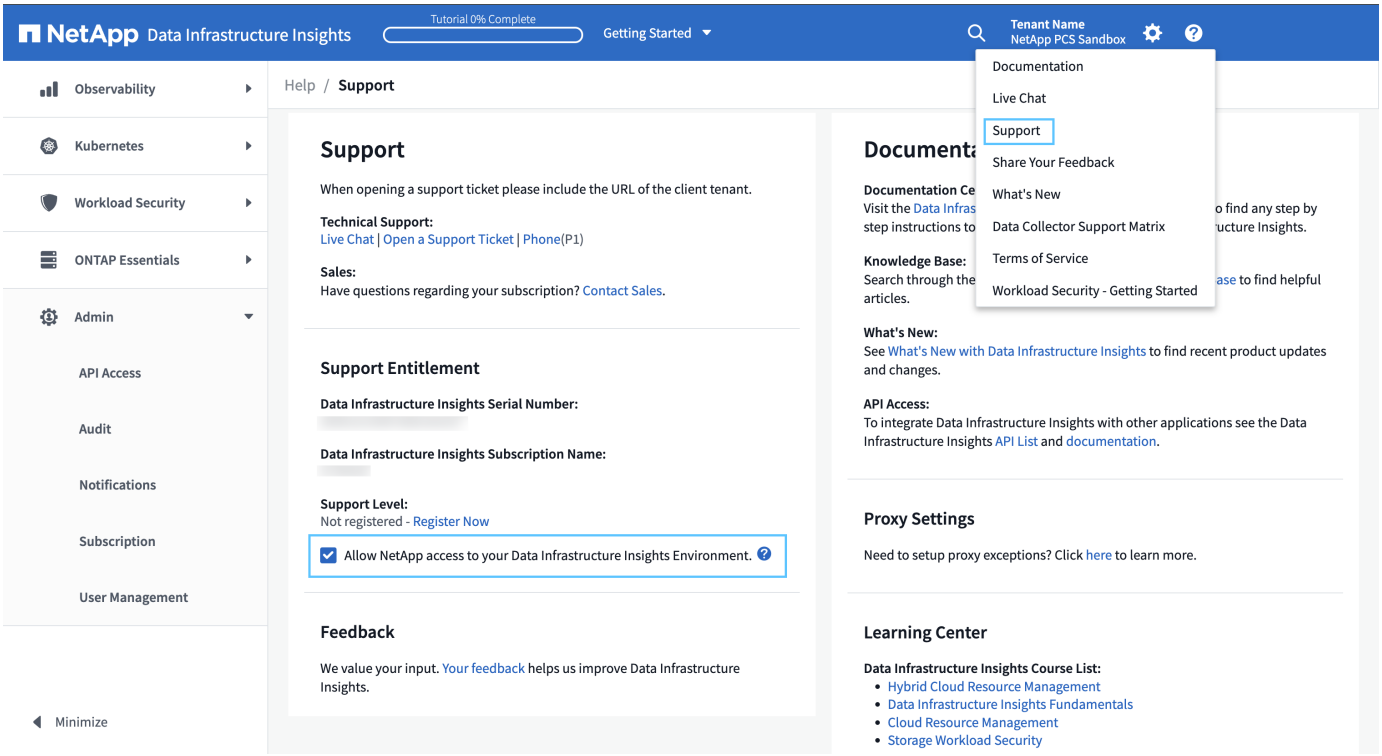


- This service is metered and invoiced along with your Keystone subscription on the same invoice. If you configure DII for Keystone before activating your Keystone subscription, billing still begins from the Keystone subscription activation date, or the modification date for existing subscriptions.
- This service is metered and invoiced based on both committed capacity and burst usage for DII performance service levels, in addition to the standard Keystone invoice. The metering of this add-on service follows the same methodology as the underlying Keystone subscription performance service level, which could be logical, provisioned, or physical.
- This service ends with the Keystone subscription to which it is linked. At renewal, you can choose whether to renew the add-on service. If you don't renew the Keystone subscription, the monitored hardware is decommissioned, and the add-on service automatically terminates.

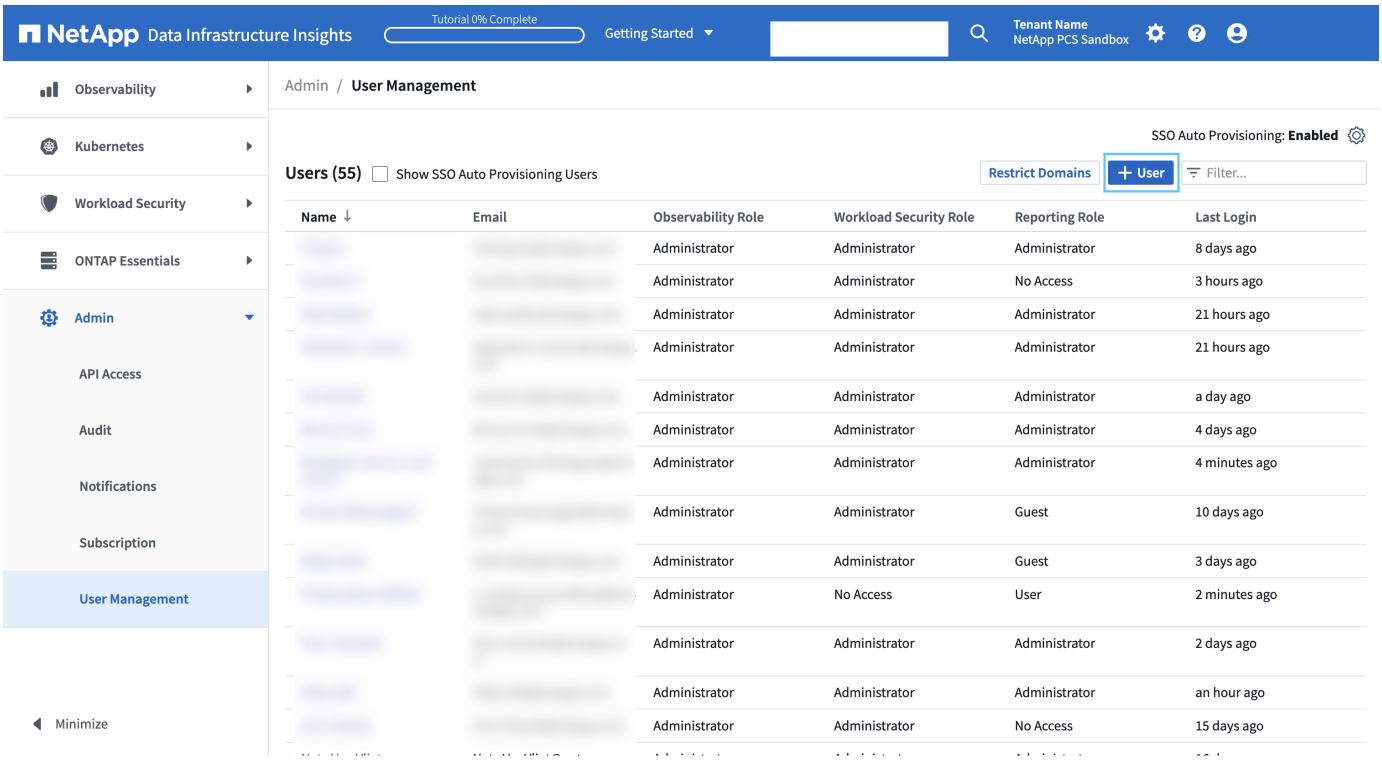
Support and user access

Approved NetApp Support team members can access the customer's DII instance if the customer enables the **Allow NetApp Access to your Data Infrastructure Insights Environment** option. To do so, go to **Help >**

Support, and enable the option.



Customers can provide access to internal or external users from the **User Management** screen using the + User option.



Data tiering add-on service for Keystone

Keystone STaaS standard services for file and block storage include tiering capabilities

that identify less-frequently used data, and tiers it to Keystone STaaS-supported NetApp cold storage. You can use data tiering as an add-on service if you want to tier your cold data to any Keystone STaaS-supported, non-NetApp storage.

For information about standard and add-on services, see Keystone STaaS services.
For information about performance service Levels, see Performance service levels in Keystone.

> The tiering add-on service is required only when data is tiered to any non-NetApp storage such as Amazon Web Services (AWS) S3, Azure Blob, Google Cloud Platform (GCP), and other Keystone STaaS-supported, S3-compatible, third party object storage.

The tiering capability leverages the NetApp FabricPool technology that enables automated tiering of infrequently accessed data to object storage tiers on and off premises.

The add-on data tiering service enables tiering from Extreme, Premium, Performance, Standard, and Value tier to an object storage target. The ratio of hot to cold data to be tiered is not fixed, and each tier is metered and invoiced separately.

For example, if the target for cold storage tier is:

- Keystone STaaS Value tier, Keystone STaaS StorageGRID Object Tier, or existing StorageGRID Webscale (SGWS) grid (customer owned) - There is no additional charge; it is part of the standard service.
- Public cloud (AWS, Azure, Google) or Keystone STaaS-supported, third party object storage - There is an additional charge for data capacity that is tiered to cold storage target.

The charges for add-on tiering services apply through the entire subscription term.

> Hyperscaler-based compute, storage, and network services required by Cloud Volumes ONTAP are not provided by NetApp as a part of Keystone STaaS subscriptions; these services need to be procured directly from hyperscaler cloud service providers.

Related information

How to approximate Keystone consumption with data tiering (FabricPool) using the ONTAP CLI

Non-returnable, non-volatile components, and SnapLock compliance add-on service for Keystone

As a part of your NetApp Keystone subscription, NetApp extends the non-returnable, non-volatile components (NRNVC) offering for your file, block, and object services.

NetApp does not recover the physical storage media used during the entire service tenure or at service termination when NetApp otherwise recovers all of its physical assets used in the delivery of the service.

You can subscribe to this add-on service as a part of your Keystone subscription. If you have purchased this service, note the following:

- You do not need to return any drives and nonvolatile memory at end of the service term or if they failed or were found defective during the service term.
- However, you need to produce a certificate of destruction for the drives and/or nonvolatile memory and cannot be used for any other purpose.

- The additional cost associated with the NRNVC is charged as a percentage of the total subscription services (includes standard service, Advanced Data Protection, and data tiering) monthly bill.
- This service is applicable only to file, block, and object services.

For information about the standard and cloud services, see Keystone STaaS services.

For information about performance service Levels, see Performance service Levels in Keystone.

SnapLock compliance

The SnapLock technology enables the NRNVC feature by making the drive unusable after the expiry date set in the volume. For using the SnapLock technology on your volumes, you need to subscribe to NRNVC. This is applicable only to file and block services.

For information about SnapLock technology, see What SnapLock is.

USPS add-on for Keystone

United States Protected Support (USPS) is an add-on offering for NetApp Keystone Subscriptions. It entitles you to receive delivery and support of ongoing Keystone services from U.S. citizens on U.S. soil.

Read the following sections to understand which elements of your subscriptions are bound by this add-on service and are provided under the terms of NetApp Keystone Agreement. [1]

NetApp USPS monitoring

NetApp USPS Keystone support team monitors the health of your products and subscribed services, provides remote support, and collaborates with your Keystone Success Manager. All personnel monitoring the products associated with the relevant Keystone subscription orders are U.S citizens operating on U.S. soil.

Keystone Success Manager

The Keystone Success Manager (KSM) is a U.S. citizen operating on U.S. soil. Their responsibilities are specified in your NetApp Keystone Agreement.

Deployment activities

Where available, onsite and remote deployment and installation activities are conducted by U.S. citizens on U.S. soil. [2]

Support

Where available, the necessary onsite troubleshooting and support activities are conducted by U.S. citizens on U.S. soil. [2]

Learn about Keystone STaaS SLO

Availability SLO for Keystone

Availability SLO targets an uptime of 99.999% during a billing period for all NetApp ONTAP flash storage arrays deployed to deliver the Keystone order.

Metrics

- **Monthly uptime percentage** = [(number of eligible seconds in a month - average of number of seconds of downtimes for all AFF storage arrays deployed to deliver the Keystone order in that month) / number of eligible seconds in a month] x 100%
- **Downtime**: The period of time when both controllers in a pair within a storage array are not available, as determined by NetApp.
- **Eligible number of seconds**: These are seconds in a month that count towards the uptime calculation. It does not include the time period when the STaaS services are not available because of planned maintenance, upgrades, support activities agreed upon with NetApp, or due to circumstances that are beyond control or responsibility of NetApp or Keystone services.

Performance service levels

All performance service levels that ONTAP flash storage arrays support are eligible for Availability SLO. To learn more, refer to Performance service levels in Keystone.

Service credits

SLAs and guarantees are available on a nomination basis.

If the availability of ONTAP flash storage arrays for eligible subscriptions falls below the 99.999% monthly uptime target within a billing period, then NetApp issues service credits as follows:

Monthly uptime (less than)	Service credit
99.999%	5%
99.99%	10%
99.9%	25%
99.0%	50%

Service credit calculation

Service credits are determined using the following formula:

Service credits = (impacted capacity / total committed capacity) X capacity fees X credit percentage

Where:

- **impacted capacity**: The amount of stored capacity affected.
- **total committed capacity**: The committed capacity for the performance service level for the Keystone order.
- **capacity fees**: The fees for the affected performance service level for the month.
- **credit percentage**: The predetermined percentage for service credit.

Example

The following example shows the method of calculation for service credits:

1. Calculate monthly uptime to determine the service credit percentage :
   - Eligible seconds in a 30-day month: 30 (days) X 24 (hours/day) X 60 (minutes/hour) X 60 (seconds/minute) = 2,592,000 seconds
   - Downtime in seconds: 95 seconds

     Using the formula:
     Monthly uptime percentage = [(2,592,000 - 95)/(2,592,000)] X 100

     Based on calculation, the monthly uptime will be 99.996%, and the service credit percentage will be 5%.
2. Calculate service credits:

Service level	Impacted capacity	Total committed capacity	Capacity fees	Credit percentage
Extreme	10 Tib for 95 seconds	100 Tib	$1000	5%

   Using the formula:
   Service credits = (10 / 100) X 1000 x 0.05

   Based on calculation, the service credits will be $5.

Service credit request

If a breach of the SLA is detected, open a priority 3 (P3) support ticket with NetApp Keystone support.

- The following details are required:
  a. Keystone subscription number
  b. Volumes and storage controller details
  c. Site, time, date, and description of the issue
  d. Calculated time duration of latency detection
  e. Measurement tools and methods
  f. Any other applicable document
- Provide the details in the excel sheet as shown below for a P3 ticket opened with NetApp Keystone support.

	A	B	C	D	E
1	**Subscription_No**	**Service_level**	**Volume_uuid**	**Date**	**Is_SLA_Breached**
2	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx5	2024-01-01	Yes
3	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx6	2024-01-02	Yes
4	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx7	2024-01-03	Yes
5	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx8	2024-01-06	Yes
6	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx9	2024-01-17	Yes

- A service credit request should be initiated within six weeks after NetApp Keystone support has validated a breach. All service credits should be acknowledged and approved by NetApp.
- Service credits may be applied to a future invoice. Service credits do not apply to expired Keystone subscriptions. To learn more, refer to NetApp Keystone support.

Performance SLO for Keystone

NetApp Keystone offers latency-based SLO per performance service level, as described in the Keystone order for consumed capacity up to the burst limit, according to the following listed terms and conditions.

Metrics

- **Degraded performance**: The amount of time, in minutes, per incident, during which the 90th percentile latency target is not met.
- The **90th percentile latency** is measured per volume, per performance level, for all volumes within a Keystone Order. Latency is sampled every five minutes, and the 90th percentile value calculated over a 24-hour period is used as the daily measure, considering the following points:
  - The volumes that record at least five IOPS at the time of metrics collection are considered for a sample.
  - Volumes with greater than 30% write operations at the time of metrics collection are excluded from the sample.
  - Latency added by AQoS for requested IOPS/TiB that is greater than target IOPS/TiB are excluded from the sample.
  - Latency added by AQoS to maintain minimum IOPS per volume are excluded from the sample.
  - For volumes that have FabricPool enabled, the latency incurred due to the transfer of data to and from the target (cold) storage is not counted.
  - Latency caused by the application, host, or customer network outside of the ONTAP cluster is not counted.
  - During a 24-hour period, at least ten valid metrics should be available. If not, the metrics will be discarded.
  - If one or more volumes on a storage array do not have a valid AQoS policy applied, then number of IOPS available to other volumes may be affected, and NetApp will not be responsible for targeting or meeting performance levels on that storage array.

- In FabricPool configurations, performance levels are applicable when all requested data blocks are on FabricPool source (hot) storage and the source storage is not in a SnapMirror Synchronous relationship.

Performance service levels

All performance service levels that ONTAP flash storage arrays support are eligible for Performance SLO and guarantee meeting the following target latency:

Service level	Extreme	Premium	Performance	Standard
Target 90th percentile latency	<1ms	<2ms	<4ms	<4ms

To learn more about the latency requirements of the performance service levels, refer to Perfromance service Levels in Keystone.

Service credits

SLAs and guarantees are available on a nomination basis.

NetApp issues service credits for the degraded performance:

Performance threshold	**Service credit**
90th percentile latency > target latency	3% for each calendar day of occurrence

Service credit calculation

Service credits are determined using the following formula:

Service credits = (impacted capacity / total committed capacity) X capacity fees X affected days X credit percentage

Where:

- **impacted capacity**: The amount of stored capacity affected.
- **total committed capacity**: The committed capacity for the performance service level for the Keystone order.
- **capacity fees**: The fees for the affected performance level as per the Keystone order.
- **affected days**: The number of calendar days impacted.
- **credit percentage**: The predetermined percentage for service credit.

Example

The following example shows the method of calculation for service credits:

Service level	**Impacted capacity**	**Total committed capacity**	**Capacity fees**	**Affected calender days**	**Credit percentage**

Extreme	10 Tib	50 Tib	$1000	2	3%

Using the formula:
Service credits = (10 / 50) X 1000 x 2 x 0.03

Based on calculation, the service credits will be $12.

Service credit request

If a breach of the SLA is detected, open a priority 3 (P3) support ticket with NetApp Keystone support.

- The following details are required:
  a. Keystone subscription number
  b. Volumes and storage controller details
  c. Site, time, date, and description of the issue
  d. Calculated time duration of latency detection
  e. Measurement tools and methods
  f. Any other applicable document
- Provide the details in the excel sheet as shown below for a P3 ticket opened with NetApp Keystone support.

	A	B	C	D	E
1	**Subscription_No**	**Service_level**	**Volume_uuid**	**Date**	**Is_SLA_Breached**
2	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx5	2024-01-01	Yes
3	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx6	2024-01-02	Yes
4	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx7	2024-01-03	Yes
5	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx8	2024-01-06	Yes
6	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx9	2024-01-17	Yes

- A service credit request should be initiated within six weeks after NetApp Keystone support has validated a breach. All service credits should be acknowledged and approved by NetApp.
- Service credits may be applied to a future invoice. Service credits do not apply to expired Keystone subscriptions. To learn more, refer to NetApp Keystone support.

Sustainability SLO for Keystone

NetApp Keystone delivers a guaranteed measurement of maximum number of actual watts per tebibyte (W/TiB) for storage services based on ONTAP flash storage arrays with Sustainability SLO. Sustainability SLO defines the maximum consumption of W/TiB for each eligible performance service level, helping organizations meet their sustainability goals.

Metrics

- **Watts**: The power consumption reported from daily AutoSupport, including the usage by the controller and attached disk shelves.
- **Tebibyte**: The maximum of:
  - the committed capacity + allocated burst capacity for the performance service level, or
  - the effective deployed capacity, assuming a storage efficiency factor of `2:1`.

  To learn more about storage efficiency ratio, refer to Analyze capacity and storage efficiency savings.

Performance service levels

Sustainability SLO is based on the following consumption criteria:

Service level	SLO criteria	Minimum committed capacity	Platform
Extreme	<= 8 W/TiB	200 TiB	AFF A800 and AFF A900
Premium	<= 4 W/TiB	300 TiB	AFF A800 and AFF A900
Performance	<= 4 W/TiB	300 TiB	AFF A800 and AFF A900

To learn more, refer to Performance service levels in Keystone.

Service credits

SLAs and guarantees are available on a nomination basis.

If W/TiB consumption during a billing period fails to meet the SLA criteria, then NetApp issues service credits as follows:

Days SLA missed in billing period	Service credit
1 to 2	3%
3 to 7	15%
14	50%

Service credit request

If a breach of the SLA is detected, open a priority 3 (P3) support ticket with NetApp Keystone support, and provide the details as requested in the excel sheet as shown below:

	A	B	C	D	E
1	**Subscription_No**	**Service_level**	**Volume_uuid**	**Date**	**Is_SLA_Breached**
2	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx5	2024-01-01	Yes
3	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx6	2024-01-02	Yes
4	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx7	2024-01-03	Yes
5	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx8	2024-01-06	Yes
6	192037XXX	premium	fxxxxxb1-fxxb-xxed-axxx-dxxxexxxxxx9	2024-01-17	Yes

- A service credit request should be initiated within six weeks after NetApp Keystone support has validated a breach. All service credits should be acknowledged and approved by NetApp.
- Service credits may be applied to a future invoice. Service credits do not apply to expired Keystone subscriptions. To learn more, refer to NetApp Keystone support.

Ransomware Recovery Guarantee for Keystone

NetApp guarantees the recovery of Snapshot data from SnapLock Compliance volumes in the event of a ransomware attack with the Ransomware Recovery Guarantee program. NetApp Ransomware Recovery Assurance Service is required to support the Ransomware Recovery Guarantee program and should be purchased separately from the associated Keystone order.

Service levels

Ransomware Recovery Assurance Service is required for all hardware supporting the Keystone subscription for the duration of the applicable subscription term.

Service credits

SLAs and guarantees are available on a nomination basis.

If SnapLock Compliance is deployed as per best practices, and NetApp professional services either configure it or validate it upon the purchase of Ransomware Recovery Assurance Service, then NetApp issues the service credits if the data protected by SnapLock is not recoverable. The criteria for these credits are as follows:

- Service credits can be applied to future invoices. The credits are capped at 10% of the Committed Contract Value (CCV) and are paid out on a per-subscription basis.
- Credits are provided during the active subscription term of the relevant Keystone order.
- For subscriptions with monthly billing, the credits will be divided over the next 12 months and can be used for any future Keystone invoices until the end of the subscription term. If the subscription ends in less than 12 months, it can be renewed to continue using the credits, or the credits can be applied to other NetApp invoices.
- For annual subscriptions, the credits will be applied to the next Keystone invoice, if available. If there are no future Keystone invoices, the credits can be applied to other NetApp invoices.

Understand billing

Learn about Keystone pricing

The NetApp Keystone STaaS pay-as-you-go subscription service offers flexible and scalable consumption with predictable and upfront pricing for your storage requirements.

Keystone provides you with the following billing facilities:

- You can pay based on IOPS and latency committed capacity to meet various workload needs. The different performance service tiers - Extreme, Premium, Standard, Value, Object, and Cloud Volumes ONTAP enable you to manage your storage based on your purchased performance service level.
- It presents predictable billing for the committed capacity and pay-per-use for variable (burst) capacity usage.
- You can select a bundle price for hardware, core OS, and support for one $/TiB price. You have a single invoice for each storage type, unified, block, optimized, object, or cloud storage services.
- You can select a flexible term for the services and payment options, such as monthly, quarterly, semi-annual, or annual.

Keystone billing is based on committed capacity and variable burst consumption.

For information about different capacities supported in Keystone, see Supported storage capacities in Keystone.

Related information

- Billing based on committed capacity
- Metering based on consumed capacity
- Billing based on burst consumption
- Billing based on miscellaneous volume types
- Billing schedules

Understand Keystone committed capacity billing

Committed capacity is the capacity committed for a particular performance service level while purchasing the subscription.

Committed capacity can be the total capacity for various performance service levels in a single subscription, as accepted by you and NetApp/partner. This capacity is stated on each Keystone order and is billed, regardless of the actual capacity consumption.

For information about different capacities supported in Keystone, see Supported storage capacities in Keystone.

Understand Keystone consumed capacity metering

Keystone STaaS has metering based on the capacity consumed by you during your service usage. Consumed capacity is the capacity that your workloads actually use.

As a part of the Keystone service deployment, NetApp continuously monitors and measures the consumption of the service. At least once in every five minutes, a consumption record is generated by the system, detailing the current consumed capacity for your subscription. These records are aggregated over the billing period to generate invoices and usage reports.

For information about different capacities supported in Keystone, see Supported storage capacities in Keystone.

Understand Keystone burst consumption billing

Keystone STaaS billing is based on *burst capacity*, which is the capacity consumed by you, on top of the committed capacity of your subscription.

Your burst limit is determined and specified in your Keystone agreement. By default, it is set at 20% above the committed capacity. You also have the option to choose burst capacity limits of 40% or 60% of the committed capacity. To learn more, refer to Burst capacity increase options.

Committed capacity is the capacity committed to you while purchasing the subscription. The committed capacity and burst capacity are measured per performance service level. Consumed capacity is the capacity that your workloads actually use.

When the consumed capacity is greater than the committed capacity for a performance service level, burst consumption is recorded and charged accordingly. The usage above the burst capacity is indicated as "Above Burst Limit".

This process occurs for each consumption record generated. Burst consumption, therefore, is a reflection of both the amount and tenure of your over-consumed capacities on top of your committed capacities. To learn more, refer to View consumption trends of your Keystone subscriptions.

For information about different capacities supported in Keystone, see Supported storage capacities in Keystone.

Learn about Keystone billing for specific volume configurations

Understanding Keystone billing for specific configurations can help you optimize service usage and manage costs. The configurations include cloned volumes, temporary volumes, SnapMirror destinations, LUNs, and system/root volumes.

Billing for cloned volumes

If volumes are cloned in ONTAP and you use them for backing up and restoring your data, you can continue using the clones without any additional payments. However, cloned volumes used for any other purpose in your business for an extensive duration are charged.

Note the following:

- Clone volumes are free from charging as long as their size is less than 10% of the parent volume (the physical capacity used in the clone volume compared to the physical capacity used in the parent volume).
- There is no 24-hour grace period for cloned volumes. Only the size of the clone is considered.
- Once the clone volume exceeds 10% of the physical size of the parent, the clone is billed as a standard volume (logical used capacity).

Billing for temporary volumes

Occasionally, temporary (TMP) volumes are created by ONTAP when moving volumes. These temporary volumes are short-lived, and the consumption on these volumes is not measured for billing.

Billing for SnapMirror destinations

The pricing for SnapMirror destination volumes, whether used for disaster recovery or long-term retention, is based on the performance service level assigned to the destination. There is no extra fee for data protection.

Billing for LUNs

For LUNs, the billing is based on the performance service levels of the volume.

System and root volumes

System and root volumes are monitored as a part of the overall monitoring of the Keystone service but are not counted or billed. The consumption on these volumes is exempted for billing.

Learn about Keystone billing schedules

Keystone STaaS subscriptions are billed on a monthly, quarterly, semi-annually, or annually basis.

Monthly billing

Invoices are sent monthly. For the month in which the services are availed, an invoice is sent in the next month. For example, the invoice for the services you have used in January is delivered at the beginning of February. This invoice includes the charges for the committed capacity and if applicable, any burst usage.

Quarterly, semi-annually, and annually billing

For quarterly, semi-annually, and annually billing, the process is similar with slight variations in timing:

- **Quarterly billing**: An invoice is generated at the beginning of each subscription quarter for the minimum payment of the committed capacity. Another invoice is sent at the end of the quarter for any burst usage accrued.
- **Semi-annually billing**: An invoice is generated at the beginning of every six months for the minimum payment of the committed capacity. Another invoice is sent at the end of each quarter for any burst usage accrued.
- **Annually billing**: An invoice is generated at the beginning of each subscription year for the minimum payment of the committed capacity. Another invoice is sent at the end of each quarter for any burst usage accrued.

For quarterly, semi-annually, and annually billing, if the committed capacity is changed during a subscription, an invoice is sent on the same day for the prorated minimum payments for the rest of that subscription year. Billing is calculated from the day the change in committed capacity becomes effective.

Copyright information

Trademark information

[1] The services and offerings described here are subject to, and limited and governed by a fully-executed Keystone Agreement.

[2] Availability of appropriate personnel for onsite activities is dependent on the geographical location at which the Keystone systems are deployed.