Keystone STaaS services

Keystone

NetApp
February 10, 2026

Table of Contents

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

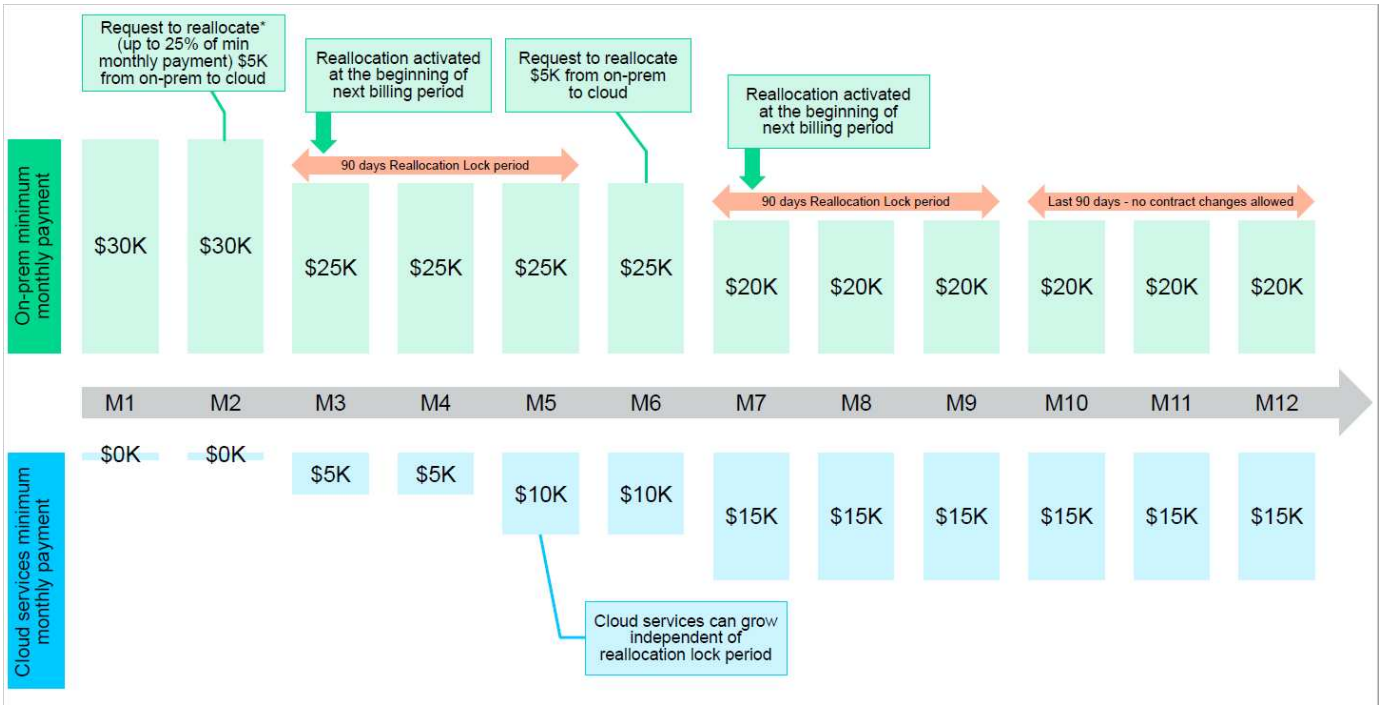
Capacity adjustments

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.



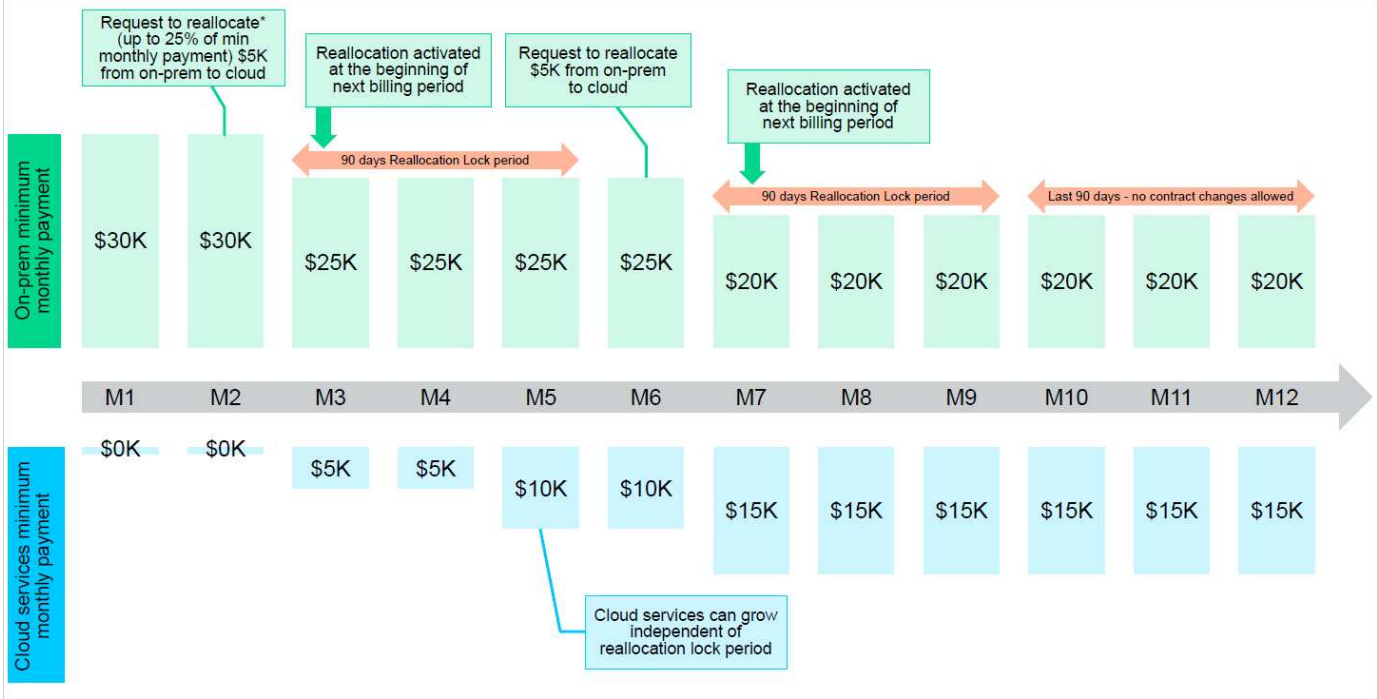
Quarterly spend reallocation

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.

Copyright information

Trademark information