08 Common Advanced Storage Technologies

www.huawei.com

HUAWEI

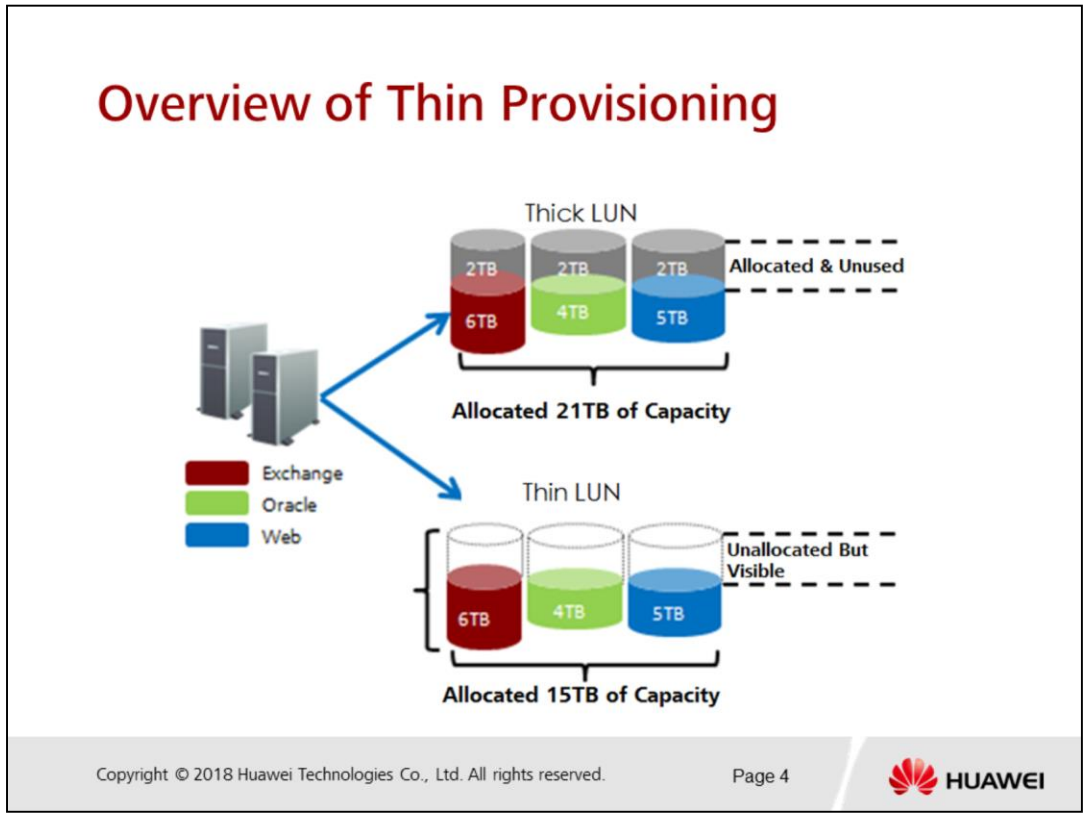
Foreword

- This module introduces the features of the common advanced storage technologies such as:
  - SmartThin
  - SmartTier
  - Smart QoS
  - SmartPartition
  - Snapshot
  - SmartQuota

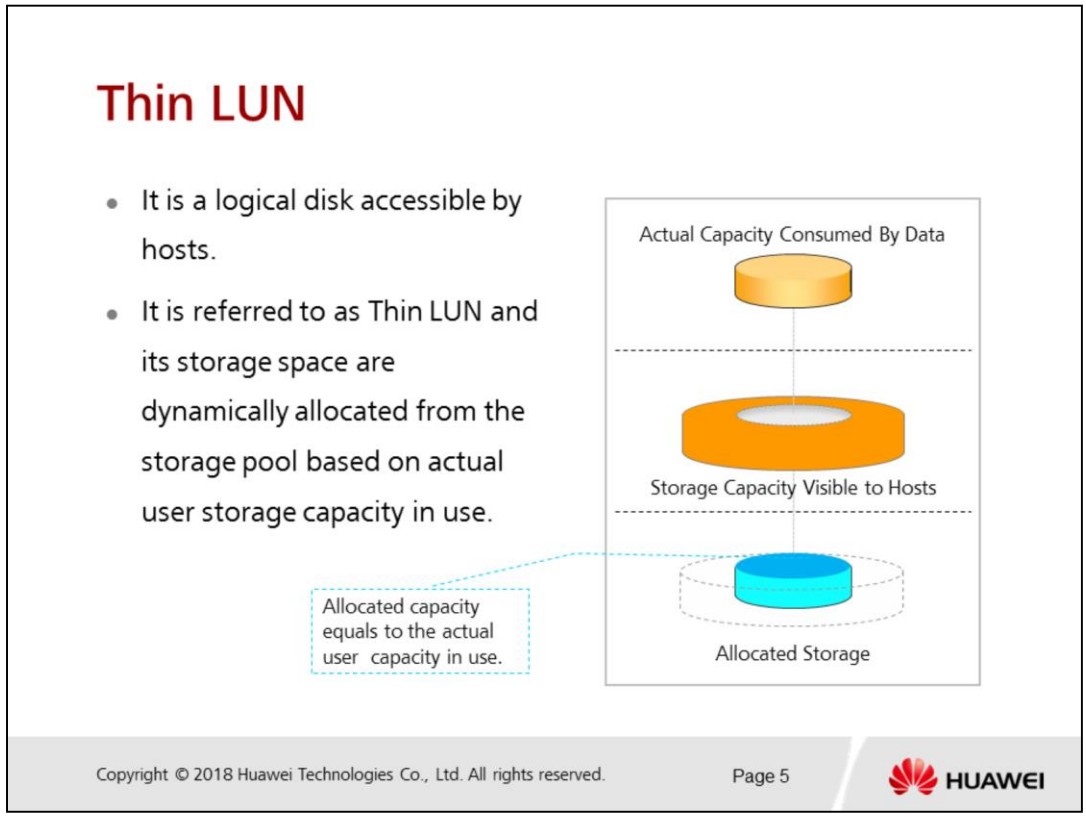
Objectives

- Upon completion of this module, you will be able to:
  - Understand the principles, configuration process and application scenarios of SmartThin technology.
  - Understand the principles, configuration process and application scenarios of SmartTier technology.
  - Understand the principles, configuration process and application scenarios of SmartQoS technology.
  - Understand the principles, configuration process and application scenarios of SmartPartition technology.
  - Understand the principles, configuration process and application scenarios of SmartQuota technology.

Contents

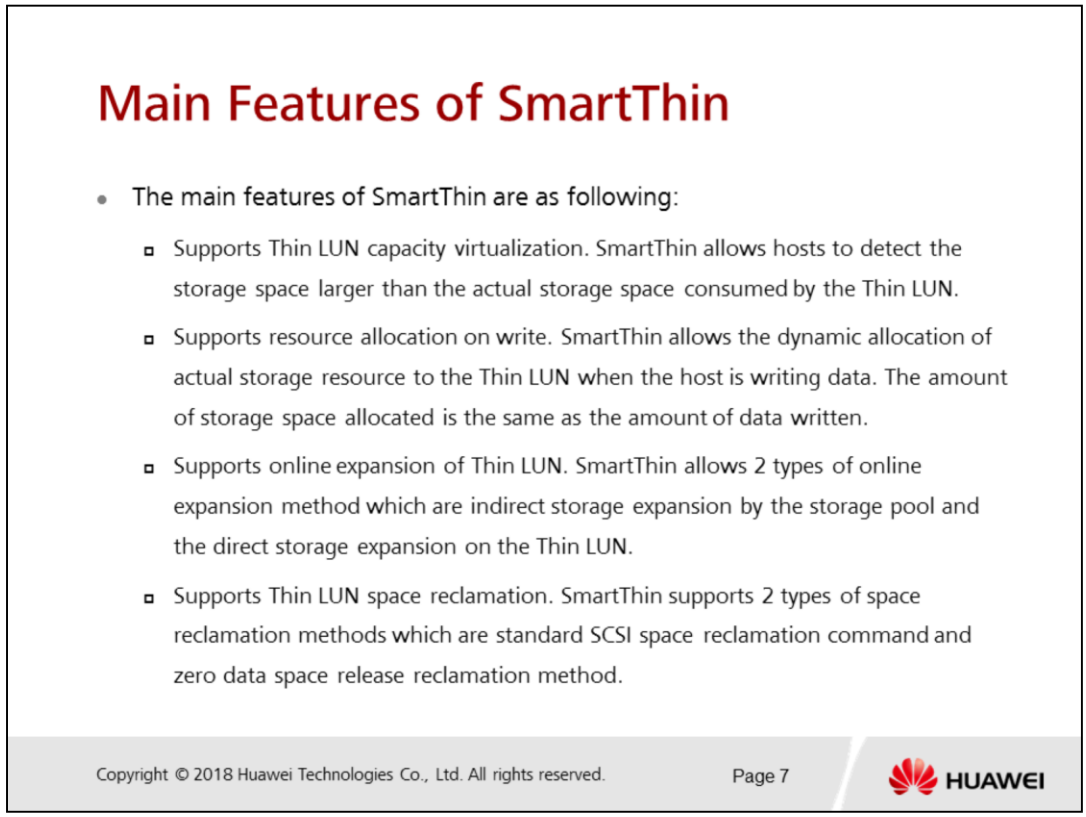


- The diagram above shows the differences in storage space allocation between the Thin LUN provided by SmartThin and the Traditional LUN.
- SmartThin technology has improvements as the following on top of traditional LUN provisioning:
  - Improvement 1: The storage space of LUNs that uses SmartThin technology is not allocated upon creation, but only allocated when it actually needs that storage space.
  - Improvement 2: Based on the fundamentals of improvement 1, the storage space of LUNs that uses SmartThin technology can be dynamically adjusted in size.
- When data capacity has exceeded the initial prediction, we can dynamically adjust the storage space of that LUN. The unused space can be used as public storage space that can be allocated to any LUNs that require the storage space. Thus, there won't be any private unutilized space in LUNs, which increases the utilization rate and efficiency ratio. At the same time, dynamic storage space adjustment provides the capability of online configuration of storage space size, which allows online expansion of storage that doesn't affect the business in operation.

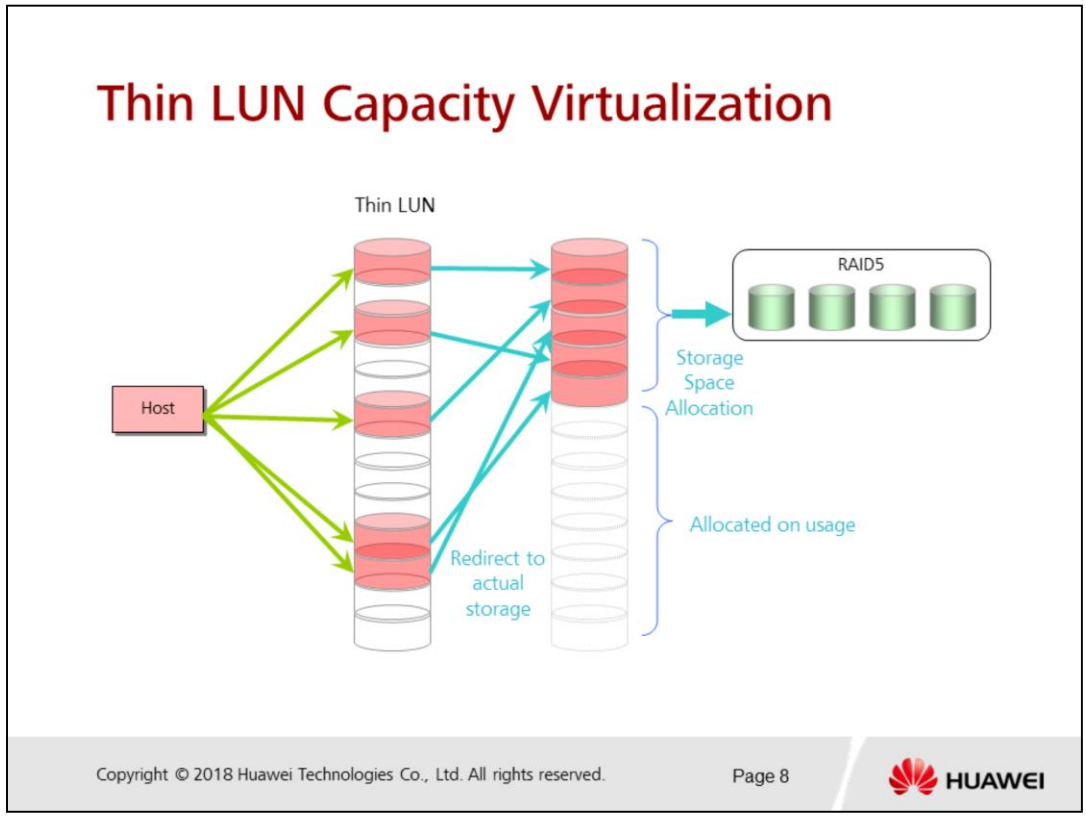


- Data Integration: It is a LUN that can be mapped to host from the perspective of storage systems.
- Fully Usable: Normal Read/Write on the Thin LUNs.
- Dynamic Allocation: Storage resource are allocated on write.
- Two different types of LUNs may be created on Storage Pools – Thick LUNs and Thin LUNs. While both allocate space on demand, there are significant differences between them in terms of both operation and performance.
- When a Thick LUN is created, the entire space that will be used for the LUN is reserved. If there is insufficient space in the Storage Pool, the Thick LUN will not be created. For example, if a Thick LUN of 100GB is created, 100GB worth of storage space from the storage pool is allocated upon creation to the Thick LUN.

- However when a Thin LUN is created, the actual storage space will be allocated when needed. For example when a Thin LUN of 100GB is created, and users stores 10GB of data within the Thin LUN. This means that the users and hosts will be able to view the full storage capacity of 100GB but the actual storage capacity used from the storage pool is just 10GB.
- Oversubscription is allowed, so the total size of the Thin LUNs in a Storage Pool can exceed the size of the available physical data space. Thus, monitoring is required to ensure that out of space conditions do not occur. There is appreciably more overhead associated with Thin LUNs than with Thick LUNs and Traditional LUNs, and performance is substantially reduced as a result compared relatively to the other types of LUN.



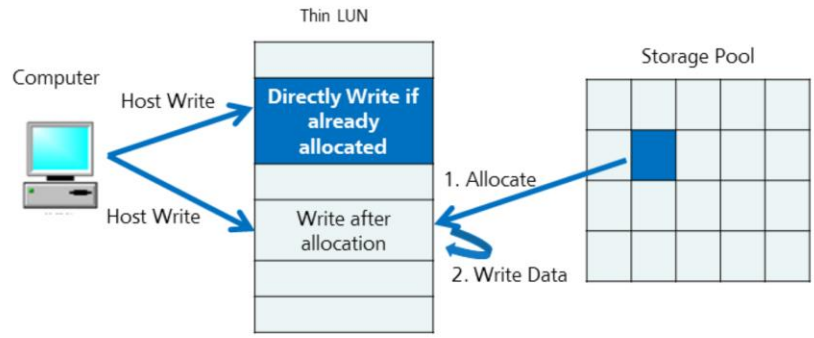
- SmartThin provides a storage management mode that supports on-demand allocation of storage space.
- SmartThin applies to the following scenarios:
  - Core system services that have high requirements for service continuity use SmartThin for online expansion, ensuring ongoing services. For example, SmartThin applies to financial systems.
  - Services whose data growth is hard to predict use SmartThin to allocate physical storage space based on requirements, avoiding wasting storage space. For example, SmartThin applies to email and web disk services.
  - Mixed services that have diverse storage requirements use SmartThin to contend for physical storage space and achieve optimal configuration of physical storage space. For example, SmartThin applies to carrier's services.



- SmartThin creates Thin LUN based on RAID2.0+ virtualized storage pools, which means that the Thin LUN and the traditional LUN coexists in the same storage pool.
- Thin LUN is a logical unit that can be created in the storage pool and mapped to host for direct access.
- The capacity of Thin LUN is not the actual physical storage space, but it is a virtualized value. The physical storage space is only applied through on write allocation policy from the storage pool when there is actual I/O process on the Thin LUN.
- In the RAID2.0+ environment, storage system will divide the storage capacity of the storage pool into individual small data blocks called CHUNK. These CHUNKs are used to form the CHUNK Groups (CKG) for RAID. This allows the data to be evenly distributed across all the hard disks in the storage pool, and resource management is based on CHUNK as the unit. SmartThin uses CKG and divides it further into smaller Extent (the smallest storage unit that can be allocated) as the unit for storage space organization.
- Thus, Thin LUN and Thick LUN exists within the same storage pool, and can utilize the physical storage capacity within the storage pool making storage planning much easier and flexible, and avoid the issues of providing separate storage pool space for Thin LUN and Thick LUN.

Capacity - on - write

- Write I/O request to the Thin LUN will trigger the storage capacity allocation.
- When the currently allocated physical storage is lower that the threshold, it will apply for new storage space from the storage pool.

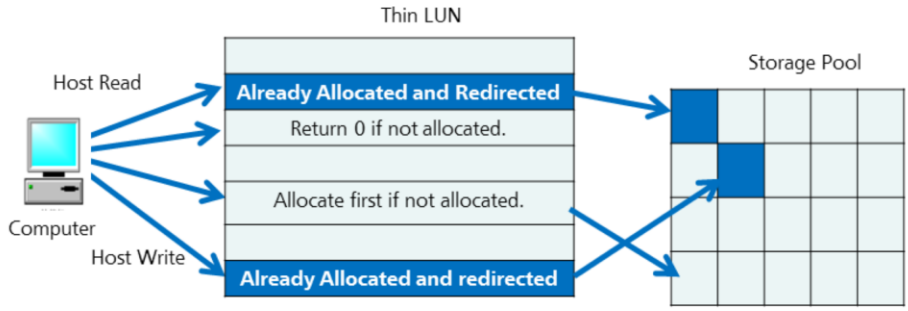


HUAWEI

- Capacity-on-write: When Thin LUN receives a data write request from the host, it will firstly use the direct-on-time technology to determine whether the logical storage area for the write request has already been allocated with physical storage space. If it is not allocated, then it will trigger the space allocation, and the smallest unit for the space allocation for Thin LUN is grain, where the size of the Grain is 64Kb. Once the space is allocated, the data will be written to the newly allocated actual storage area.

Direct - on - time

- Due to the adoption of the capacity-on-write technology, the actual storage area for the data is not fixed, thus when there is an read/write I/O request for the Thin LUN, it need to be directed on time to the actual storage area.



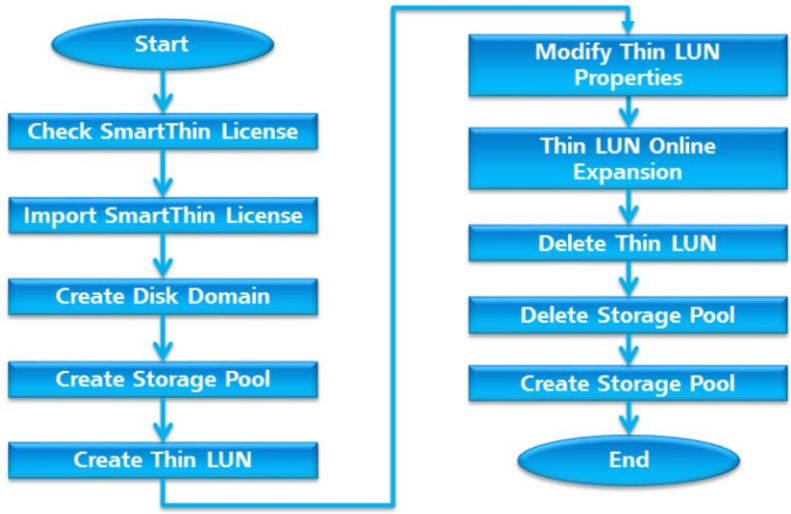
- Due to the adoption of capacity-on-write technology, the relationship between actual storage area of data and the logical storage area of the data is not fixed and not constant, thus can no longer be calculated based on a exact formula, but calculated during write time based on the principle of direct-on-time. In simpler words, since the storage space is allocated on write, it requires a calculation during the write process to map the logical storage area on the LUN to the actual storage area in the storage systems. Similarly, when there is a read request, a check is required to determine the mapped actual storage area to access the data. Both these calculations are run on-time when the request is initiated and thus are referred to direct-on-time.
- Thus when there is a read/write request to the Thin LUN, it needs to be redirected based on the relationship of the logical and actual storage area, and these mappings relies on the mapping table.
- The mapping table's main function is to record the mapping relationship between the actual storage area and the logical storage area. During the write process the mapping table is dynamically updated, and the same mapping table is checked during the read process. Hence, the Direct-on-time operation can also be divided into Read Redirection and Write Redirection.

- Read Redirection: When Thin LUN receives the data read request from the host, it will first check the mapping table, if the logical and actual storage area is already mapped then it redirects the read request to the actual storage area, then after retrieving the data then returns it to the host. In contrast, if the data read request is not mapped in the table, or the space is not allocated, then it will return all 0 as the data from the logical storage area to the host.
- Write Redirection: When the Thin LUN receives the write request from the host, it will first check the mapping table, and if the data to be written to the logical storage area has already been allocated with the physical storage, then it will redirect the date write from the logical storage area to the actual storage area and writes it, then return a write success signal back to the host. In contrast, if the data write operation finds that the intended logical storage area has not been allocated with the physical storage space, then the capacity-on-write technology is triggered.

Application Scenario

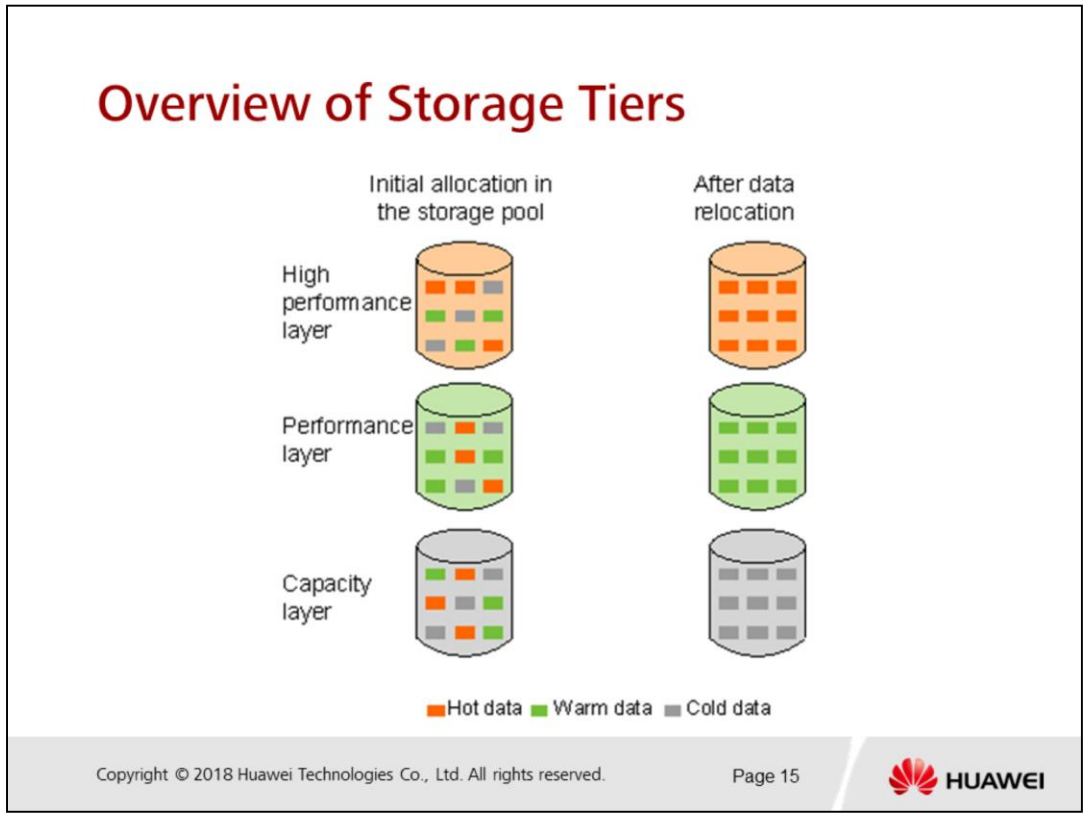
- For core business services that has high requirements for business continuity, using SmartThin configuration allows online expansion of the system without interruption to business service. For example: bank transaction systems.
- For businesses that has high application data growth speed that is unable to be accurately evaluated, using SmartThin configuration allows on demand allocation of physical storage space and avoid wastage in utilization rate. For example: Email services and Web Disk services.
- For business with complex systems and has different requirements for storage for different services and applications, using SmartThin configuration allows different business services to compete for the physical storage space, allowing optimum configuration for physical storage space. For example: Carrier Service Providers.

Configuration Process



HUAWEI

Contents

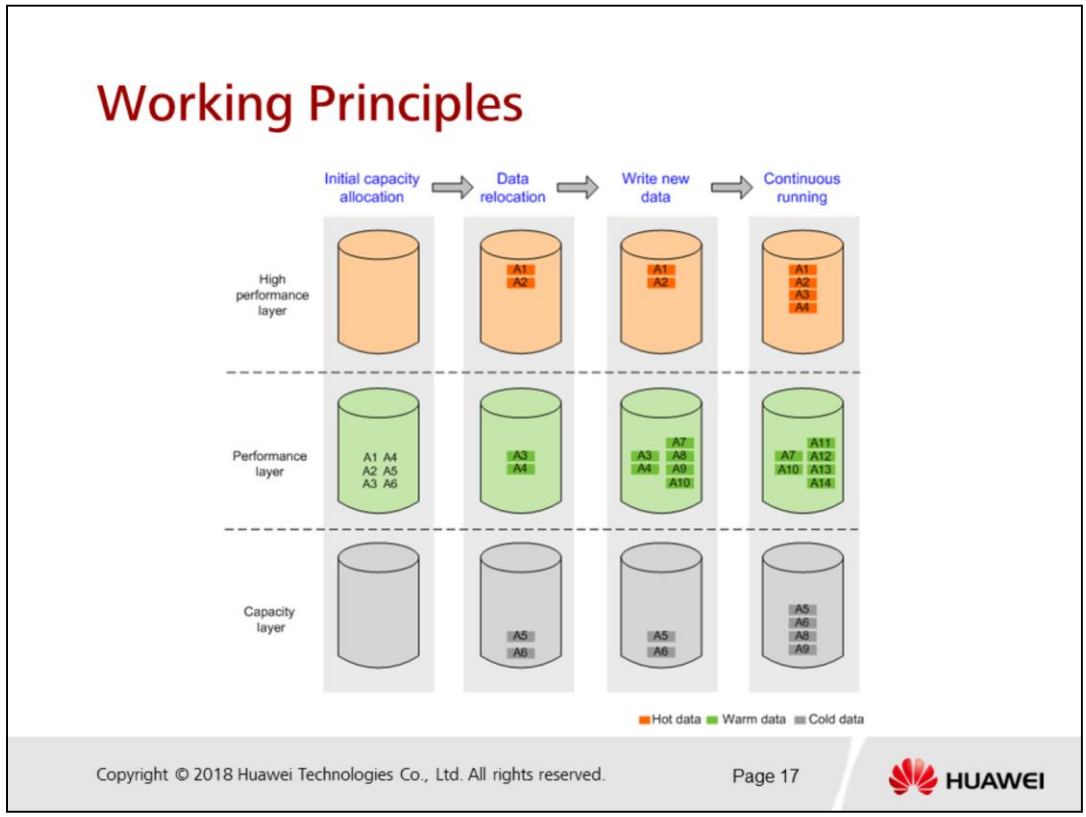


- SmartTier is a proprietary storage tier software developed by Huawei based on the RAID2.0+ technology. It automatically matches the data of different activity level with the different characteristics of the storage media, in order to increase the storage system performance and lower the user costs.
- Tier storage technology completes tiered storage of data through data migration. Currently, tier storage technology is divided into manual migration and automatic migration.
- Manual migration is performed by system maintenance personnel based on the current state of operation of the storage system, and the pressure on each business systems. They manually migrate the data from one LUN to another LUN.
- Automatic migration is based on the access frequency of the files or blocks of data. The access frequency helps to differentiate between the hot data which are frequently accessed and cold data that are less accessed. Different types of storage media is matched by their characteristics to the hot or cold data. Frequently accessed hot data is automatically migrated to the high performance storage media, and cold data are automatically migrated to the high capacity and low cost storage media.

Levels of Storage Tier

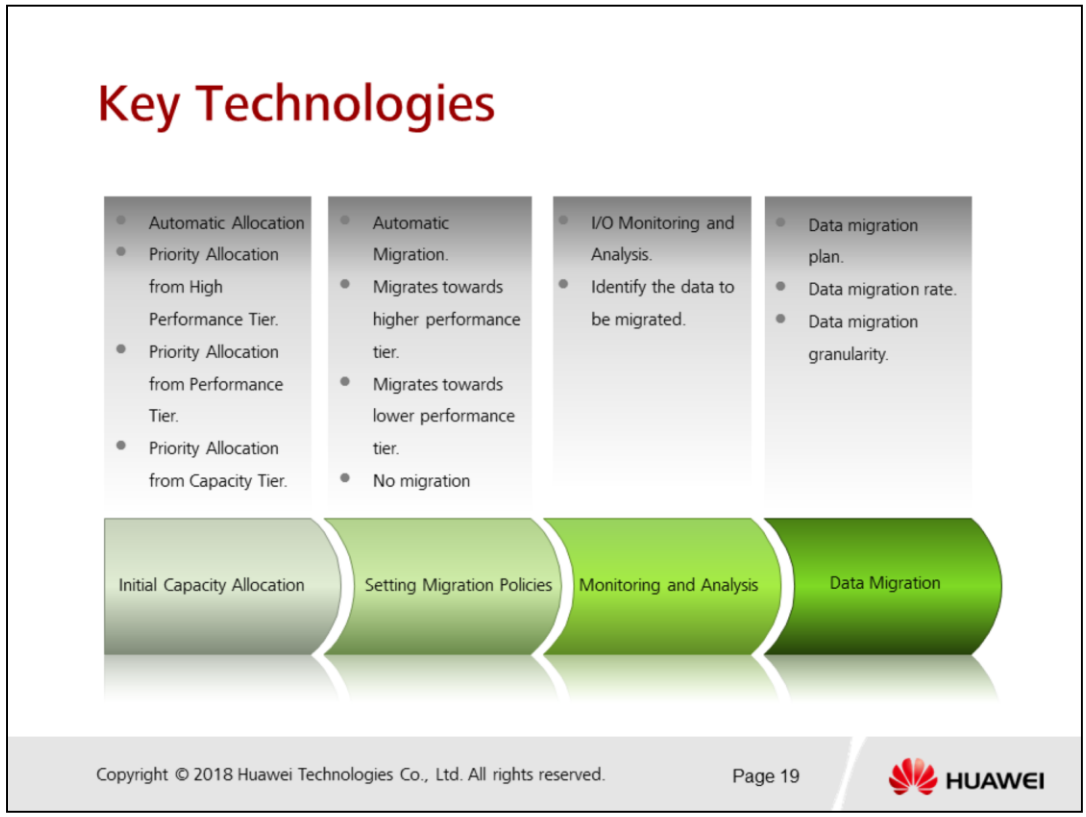
Storage Tier	Disk Type	Disk Characteristics	Data Characteristics
High Performance	SSD	SSD has very short response time, each unit of storage capacity has high costs.	High Performance Tier is suitable for data with high frequency of access.
Performance	SAS	SAS disks has short response time and each unit of storage capacity has medium costs.	Performance Tier is suitable for data with medium frequency of access.
Capacity	NL-SAS	NL-SAS disks has relatively longer response time, and each unit of storage capacity has low cost.	Capacity Tier is suitable for data with low frequency of access.

- SmartTier divides the different types of storage media into 3 levels of storage tiers based on the performance. SSD drives with the highest performance is categorized into the High Performance Tier, SAS drives forms the Performance Tier and NL-SAS drives forms the Capacity Tier.
- Each storage tier independently uses the same disk type and RAID policies.

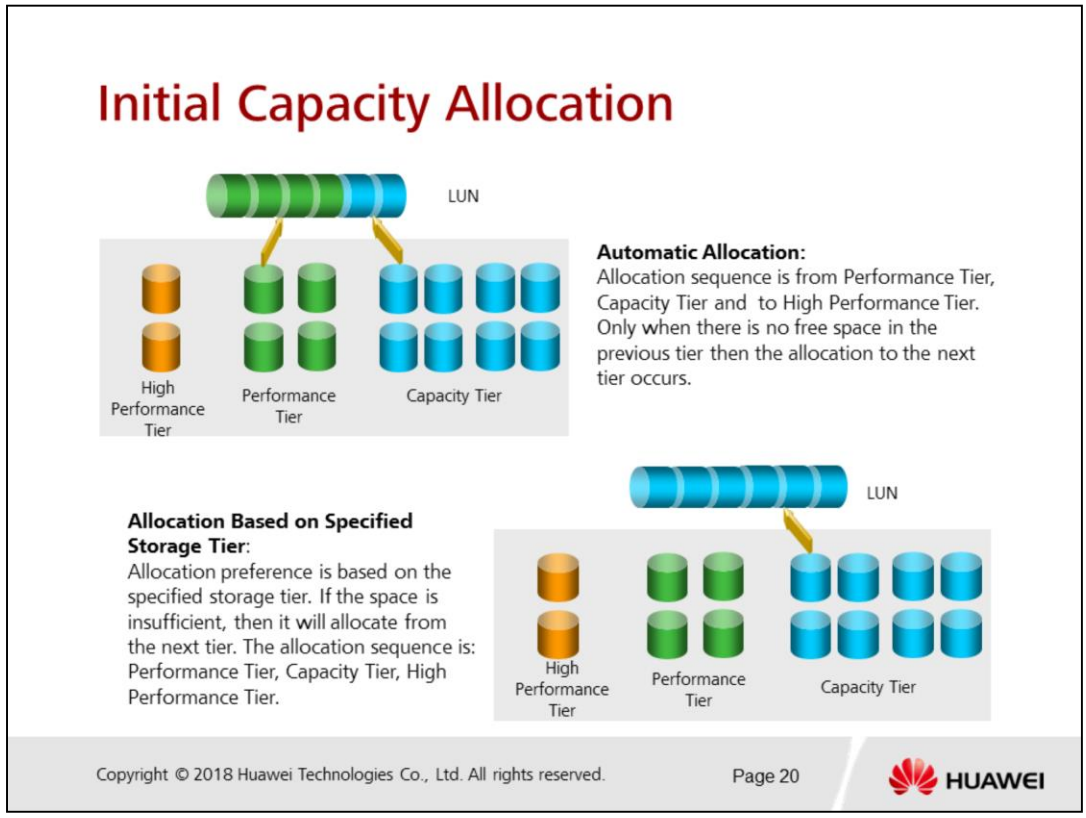


- Storage Pool is the logical combination of one or more storage tiers, and can support up to 3 types of storage tiers. The disks contained within the storage pool determines what kind of storage tiers can be created in the storage pool. A storage pool that consist of a single disk type cannot create different storage tiers, and thus cannot use SmartTier for intelligent data storage management.
- LUNs are created in the storage pool, it distributes the data in the LUN to the different storage tiers within the storage pool before SmartTier functions are applied on those data.
- The SmartTier partitions LUN data by certain granularity. Such granularity is called Data Migration granularity or data block. The data migration granularity is defined when creating a storage pool and the value cannot be changed after being set.
- If a storage pool contains more than one type of disks, the SmartTier can be used to fully utilize the performance of each storage tier. Data migration is an action that promotes data blocks with higher activity and demotes data blocks with lower activity. In this process, storage pool uses blocks as the unit to differentiate the data activity level and then migrates the whole block of data to the other storage tier.

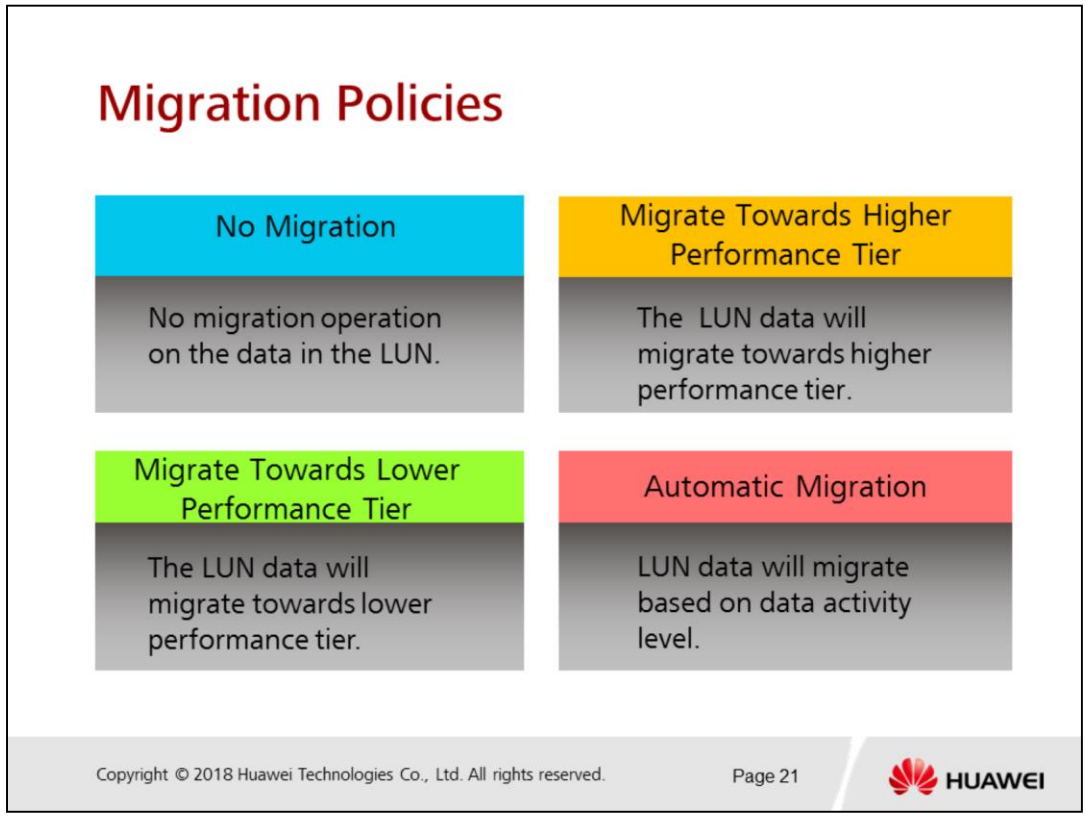
- The SmartTier experiences three stages in managing storage system data at the data block level which are I/O monitoring, data placement analysis, and data migration.
- The storage system is also configured with an initial capacity allocation policy to determine the location of a new data block within available tiers.
- When new data is written into LUN, the storage system will write the new data into the matching storage tier based on the initial capacity allocation policy.
- As the data lifecycle progresses, the data activity level changes. Then the SmartTier relocates the data to another storage tier to meet the change and improve storage system performance. The whole data relocation process does not affect the writes of new data.



- The initial capacity allocation policy determines the tier to which a new data is written. For example, if the allocate from the performance layer first policy is chosen, new data blocks written into the storage system are placed into the performance tier. There are four initial capacity allocation policies available: automatic allocation, allocate from the high performance layer first, allocate from the performance layer first, and allocate from the capacity layer first. Automatic allocation means that the storage system distributes new data to the performance tier, the capacity tier, and then the high performance tier in sequence.

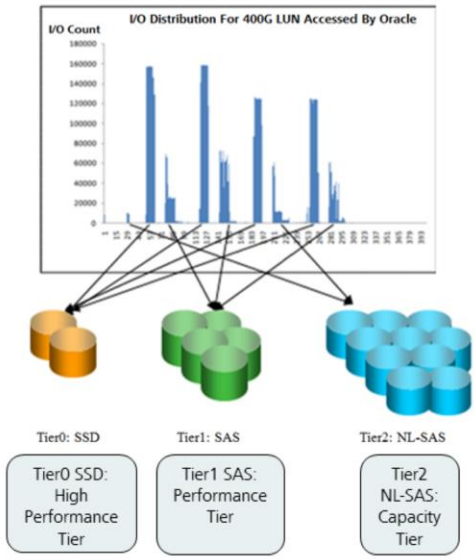


- There are 4 types of initial capacity allocation policy which are: Automatic allocation, Allocate from the high performance layer first, Allocate from the performance layer first, and Allocate from the capacity layer first.



- Automatic Migration: When the LUN is configured with Automatic Migration policies, the LUN will do data migration based on data access frequency. The data with higher access frequency migrates towards higher performance tier, and data with lower access frequency migrates towards lower performance tier. If there is no special requirements, automatic migration is recommended for better overall performance.
- Migration Towards Higher Performance Tier: When the LUN is configured with this setting, no matter if the data block access frequency is high or low, the data blocks in the LUN will migrate towards higher performance tier, and consume more high performance tier storage resources. Configure this setting only if you have special requirements for the LUN.
- Migration Towards Lower Performance Tier: When the LUN is configured with this setting, no matter if the data block access frequency is high or low, the data blocks in the LUN will migrate towards lower performance tier. When the LUN is used for archive services, or business services with low performance requirements, this setting can be configured for that LUN.

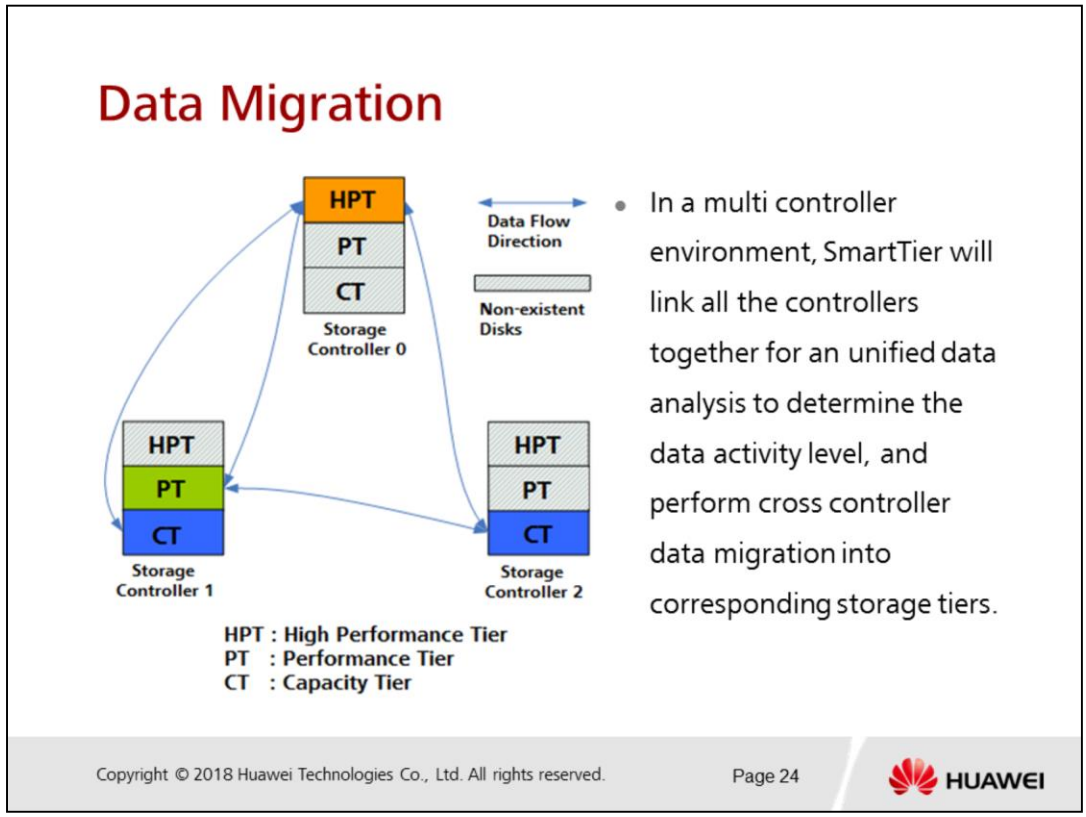
I/O Monitoring and Analysis



- Based on the I/O statistics information during the monitoring period of the monitored storage pool, such as Read/Write I/O, Average I/O size, the I/O distribution in the storage pool can be determined periodically. Thus, it can be used to determine the data activity level (hot/cold) and then figure out the data blocks to be migrated through analysis.

- The is no absolute value for differentiating between hot and cold data. The high performance storage tier has the capacity to store A number of data blocks, while the performance tier has the capacity to store B number of data blocks. During migration, all the LUNs within the storage pool that has automatic migration policies configured will move A number of data blocks with the highest access frequency to the high performance tier and move B number of data blocks to the performance tier. Other data blocks that has lower access frequency will be moved to the capacity tier.
- The main principle is to allows data with the highest access frequency to use the storage media with the best performance level.

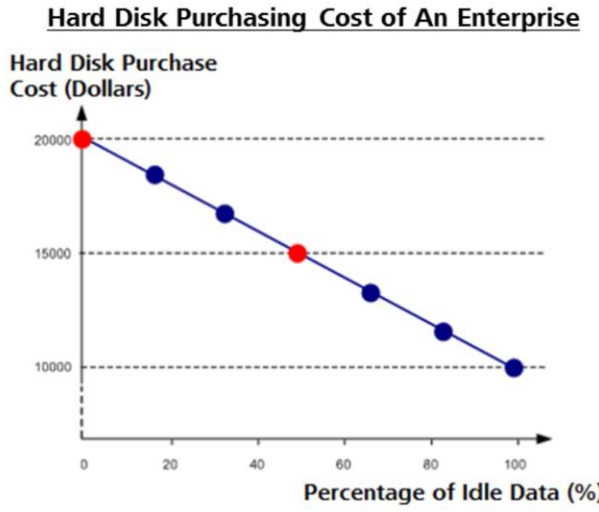
- The diagram above shows the experiment done by simulating an Oracle relational database in a 400G LUN.
- In order to more intuitively reflect the database IO characteristics, in the experimental environment, the database relational table is stored in the LUN of 400G, and the test tool is used to simulate the database application, and the access data of the host when accessing the database is counted in units of G(Gigabyte).
- As shown in the figure above, the amount of frequently accessed data only accounts for a small part of the entire database space, and most of the database space has little or no access. In a customer's real environment, only 10% of the data in the database is accessed frequently. Compared with the database IO characteristics simulated in the laboratory, the data in the customer's real environment has more obvious hot spots.
- For relational databases, index data is the most frequently accessed, so when the database is large, the speed of the index determines the delay in accessing the database. Hence, by using the SmartTier features, it allows data that is frequently accessed to be placed at high performance SSD, which increases the overall storage performance and at the same time increases the performance of the database in responding to application database requests.



- The data migration speed can be configured to High, Medium and Low which correspond to the actual data migration speed of a single storage controller which are 100M/s, 20M/s and 10M/s. The default migration speed is set to Low.
- To lower the impact of SmartTier data migration towards the performance of host business services, the configured data migration speed is the upper limit of the migration speed allowed in the storage system. Data migration always gives priority to the host business services, and uses the idle time of the storage devices to perform data migration, and it adjusts the migration speed based on the real-time pressure on the host business services. When the host business services has lower pressure, the migration speed is adjusted higher whereas when the pressure on the business systems is high, it will lower the data migration speed. No matter how the speed of the migration is adjusted automatically, the actual data migration speed will not exceed the configured value of data migration speed.
    - Low: Suitable for data migration period where there is high pressure on the business systems.
    - Medium: Suitable for data migration period where there is medium pressure on the business systems.
    - High: Suitable for data migration period where there is no business services are running, or when the host and business services are not sensitive towards the performance requirements.

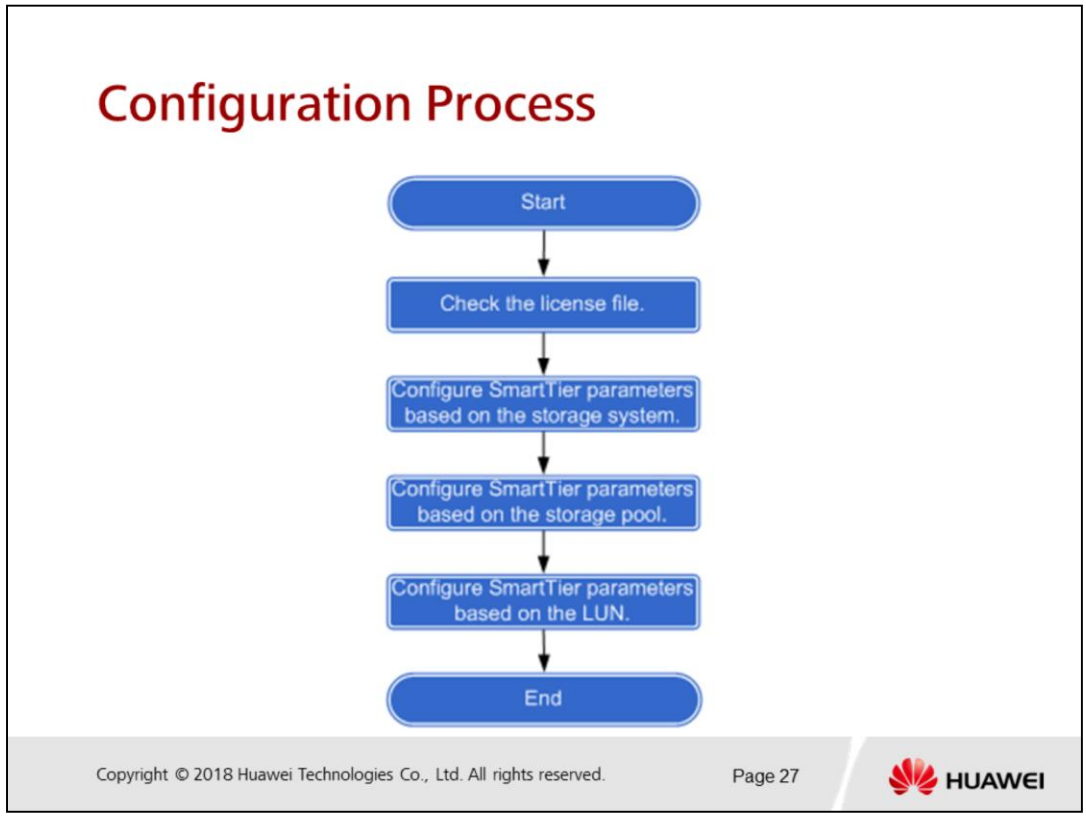
- Data Migration Granularity:
    - 512KB~64MB (Based on business service configuration, such as video monitoring service which are suitable for bigger granularity).
- Data Migration Plan:
    - Manual: Can perform the data migration at any given time.
    - Scheduled: Can only perform the data migration in the specific time period configured.
- Data Migration Rate:
    - Provides 3 types of data migration rate: High, Medium and Low.
    - Based on the current business service load, it can dynamically adjust the migration rate to ensure that there is no visible impact towards the current business services.

Application Scenario - Lowers TCO



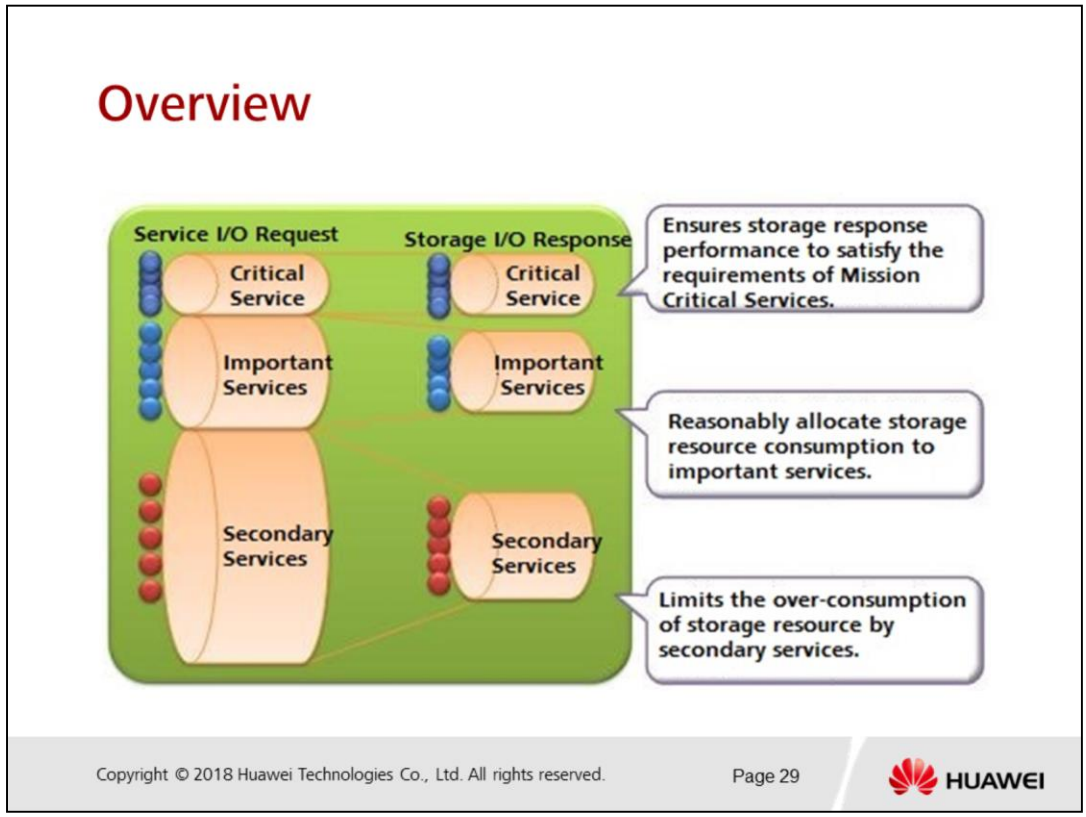
HUAWEI

- Due to the free data (Not highly accessed) are stored in the NL-SAS drives, high performance SSD drives could release their storage space and allows more of the highly accessed data to be stored in the SSD drives. SSD drives provides shorter response time and higher IOPS for highly accessed data which in turn increases the performance of the storage system.
- Better utilization rate of the hard disks allows savings on the amount of hard disks purchased, which effectively lowers the TCO(Total Cost of Ownership) of the storage system.
- SmartTier allows better usage of hard disks of different performance levels, which not only increases the overall performance of the storage system but also lowers the cost required for hard disk purchase as it allows better utilization of hard disks with different performance levels with data of different activity levels.

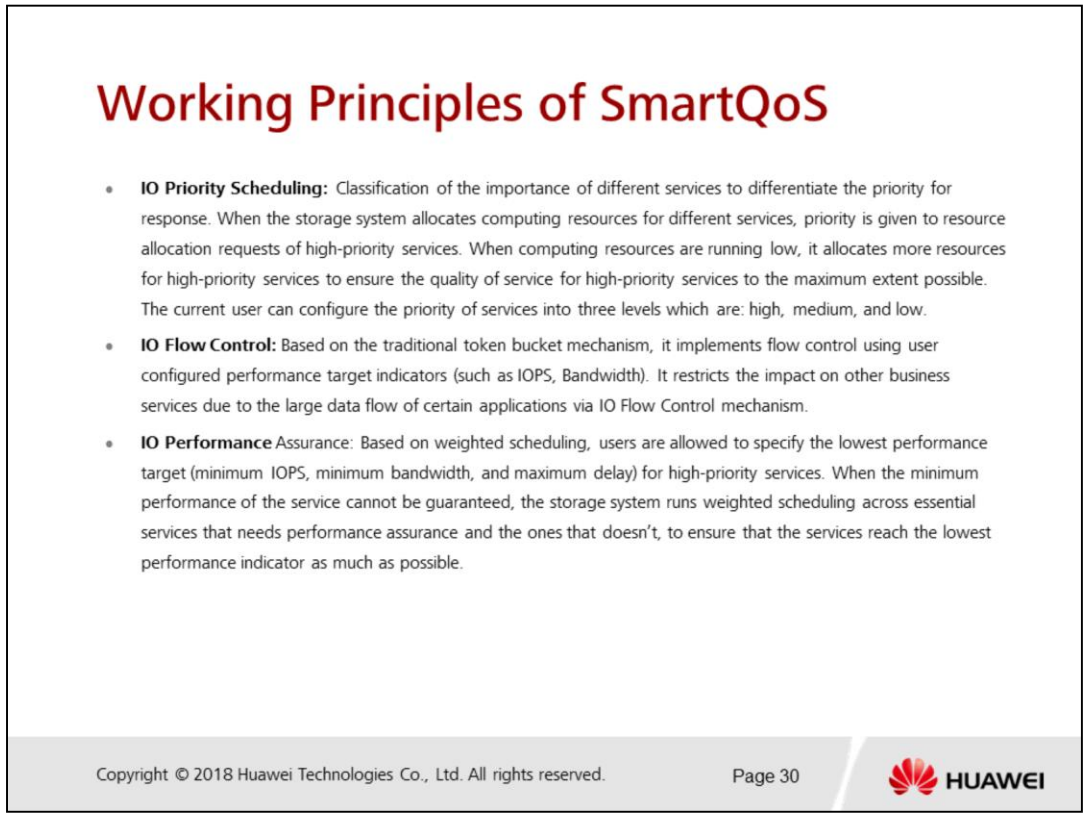


- The flowchart above shows the process of configuring the SmartTier feature in Huawei OceanStor V3 storage systems.
- The process for configuring SmartTier within a storage system consists of checking the license file and configuring SmartTier parameters based on storage system levels, storage pool levels, and LUN levels.

Contents

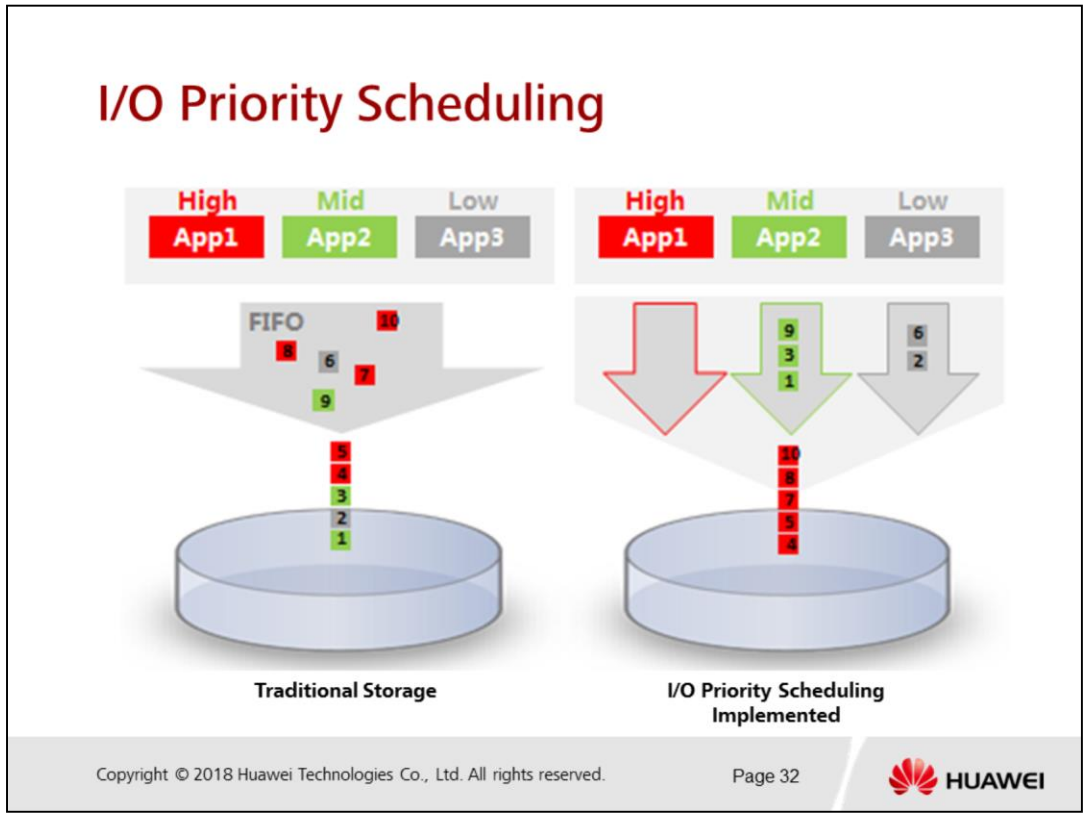


- SmartQoS is a storage Quality of Service(QoS) feature provided in the OceanStor V3 converged storage systems.
- SmartQoS is an intelligent QoS control feature developed by Huawei that dynamically allocates a storage system's resources to meet specific performance goals of certain applications.
- As shows on the diagram above, it can dynamically allocate and adjust the resources within the storage system, and has fine granular control from one end to another end along the data I/O path of the storage system, to ensure and fulfill different QoS requirements in different applications of various importance on the same storage device.

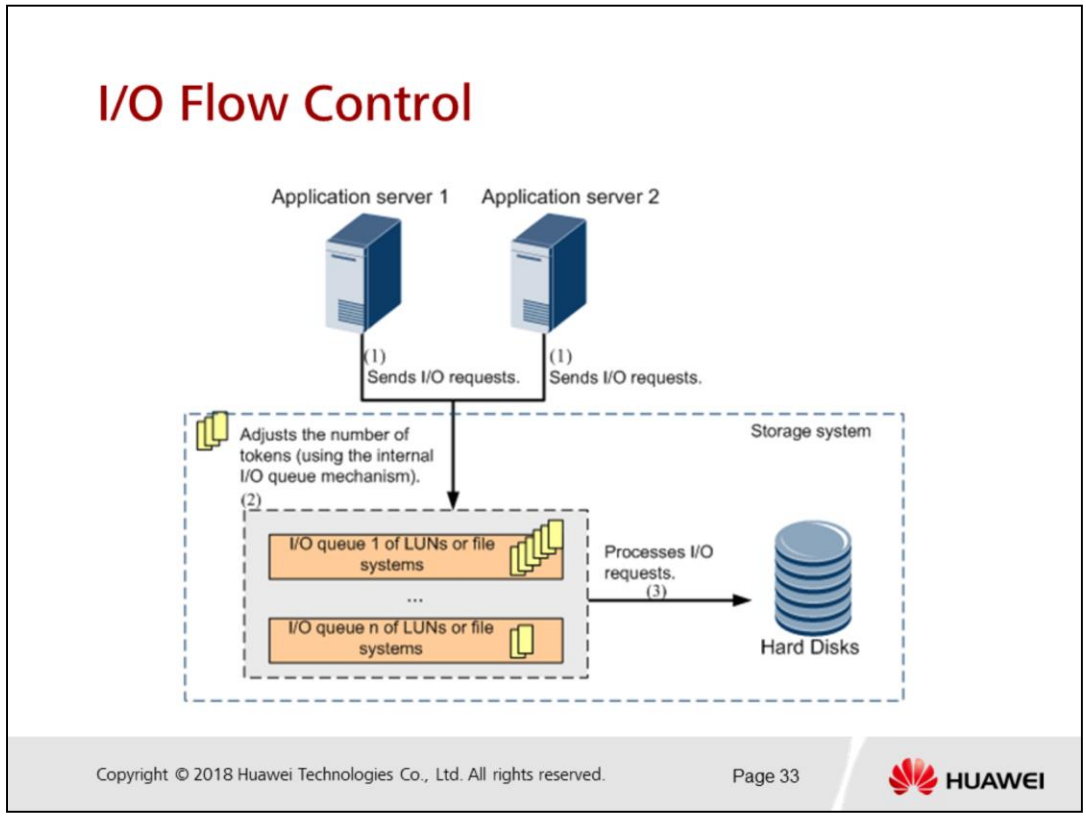


- SmartQoS Features:
  - Full IO Path Priority Scheduling:
    - SmartQoS provides priority scheduling on the whole IO path through computing resource scheduling and allocation control. Compared to the other vendor's priority scheduling policies, the control granularity is much finer.
  - Different QoS policies for different application scenarios:
    - SmartQoS has finer classification of application scenarios, and provides different QoS policies for different application scenarios. When the user has specific performance indicator target requirement, it can use flow control or lowest performance assurance policies to satisfy user requirements. When the user doesn't have specific performance indicator target requirement, it can use the priority mechanism for applications to ensure best performance for high priority applications.

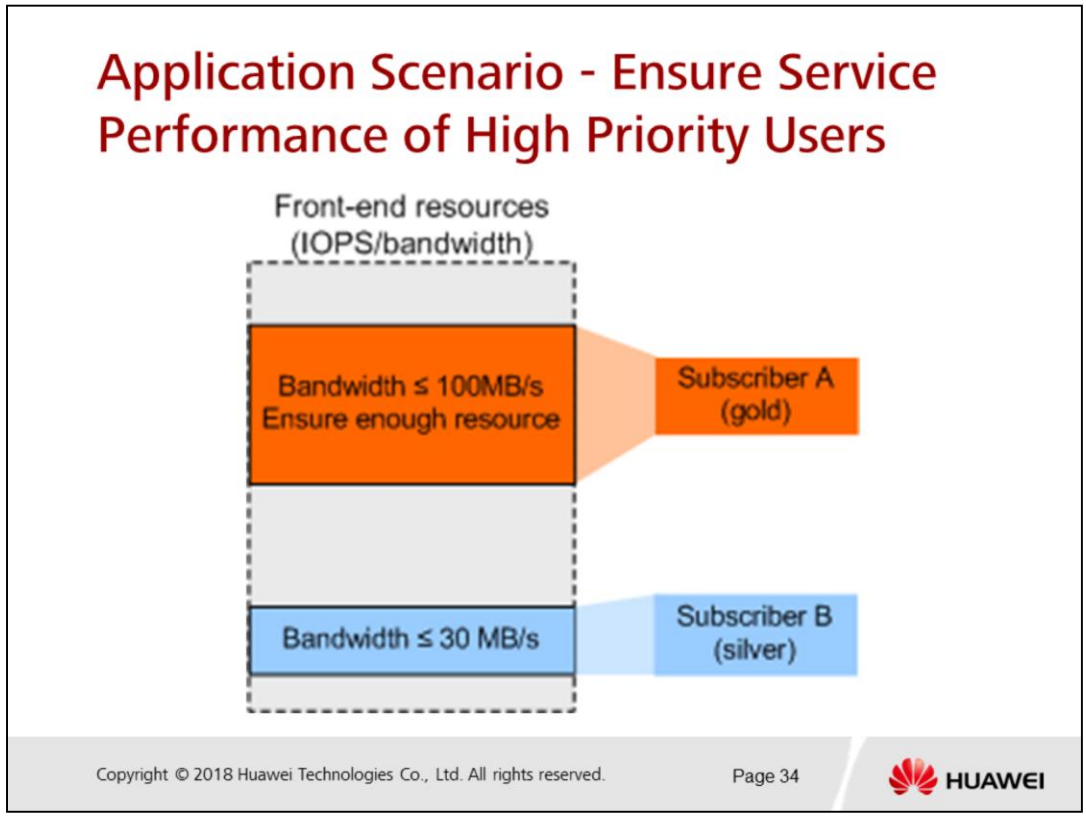
- IO Performance Assurance:
    - OceanStor V3 converged storage system SmartQoS feature has a performance assurance technology by weighted scheduling to ensure the lowest performance requirement of critical services are met at all times. In simpler words, users/applications can be configured into different QoS levels with different priority with a lowest performance threshold set. Thus, the SmartQoS feature will ensure that these user/applications has the sufficient amount of resources for operation based on the threshold set.
    - Users can set the lowest performance indicator target value(IOPS, Bandwidth) for critical services. When these target are unable to be met, the system will perform weighted scheduling based on priority level to adjust and allocate resources to ensure that the critical application meet their lowest performance targets. For example, core application A has higher priority level than normal application B. Thus, in the event where resources are insufficient to maintain the lowest performance target of the core application A, SmartQoS will allocate more resources to the core application A to ensure its performance level.



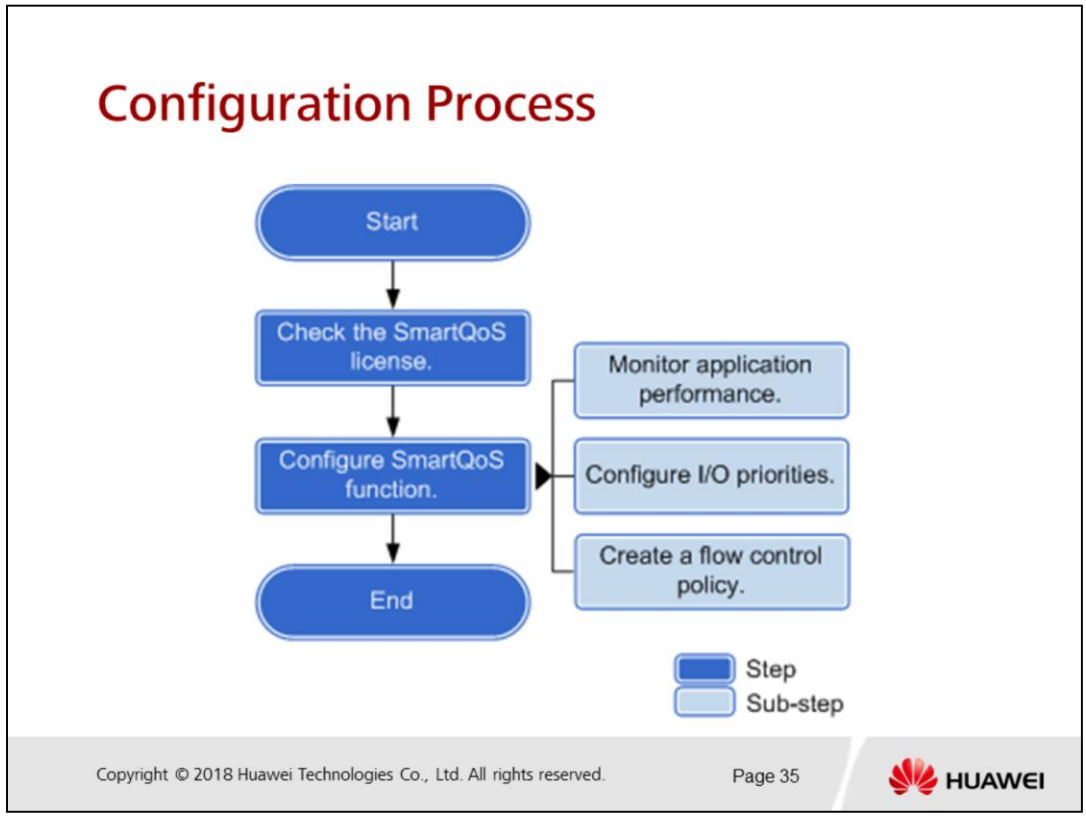
- SmartQoS feature's priority scheduling technology is based on LUN and File System priority levels.
- Hence, each LUN and File System has a priority level properties, and these properties are user configured and saved in the database. When an I/O request is sent from host to storage array, this I/O will obtain its priority level property information from the LUN or File System it belongs to and carries this priority information along the whole I/O path until it reach its destination.
- Users need to configure the priority level when creating the LUN or File System. If it is not configured, the LUN or File System created by have the lowest priority level by default.
- After the LUN or File System is created, its priority level can be manually changed based on user requirements.



- SmartQoS I/O Flow Control's performance indicator target are achieved through a mechanism of token distribution and control. When users configured the performance upper limit for a certain flow control group, this upper limit will be converted into corresponding tokens. In the storage system, if the users wants to restrict the flow type of IOPS, then each IO corresponds to a token. If the performance indicator is set to Bandwidth, then each sector corresponds to a token.
- Each flow queue has a token bucket, and SmartQoS will periodically put in a certain amount of tokens within the bucket of each flow queue. The amount of tokens depends on the user configured performance upper limit of this flow queue. Thus, if the user configure the performance indicator upper limit as the IOPS=10000, then the token distribution algorithm will set the amount of tokens in the token bucket of the flow queue to be 10000 tokens per second.
- When the flow queue is managed, it will check whether there is sufficient token within the bucket of the flow queue. If there is tokens in the bucket, then it will retrieve an I/O to be processed, and consume the tokens. If there is insufficient tokens in the bucket, then the flow queue must wait till the bucket has enough tokens to be used.

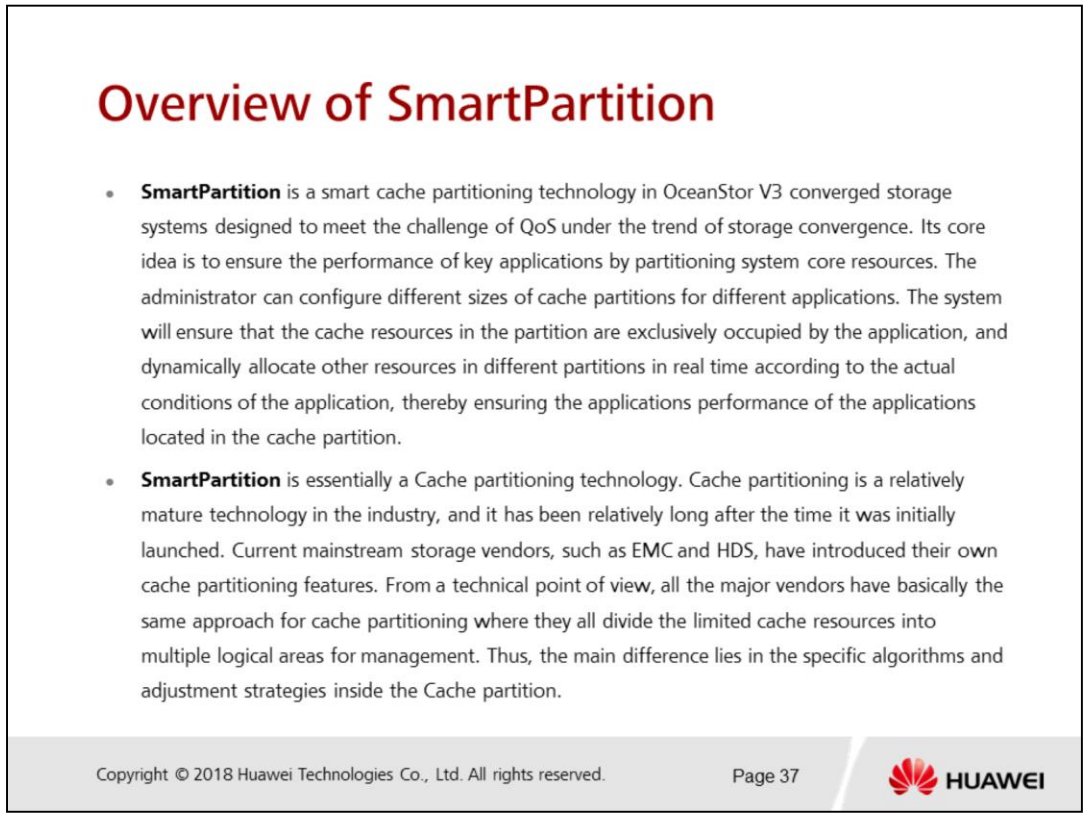


- To satisfy the high priority users, we can configure the I/O priority level on the LUN to ensure the operations of the service.
  - Configure user A with higher priority, to ensure that the services for user A will operate as normal.
  - Configure user B with lower priority, to ensure that services for user B will not impact the normal operations of user A's services.

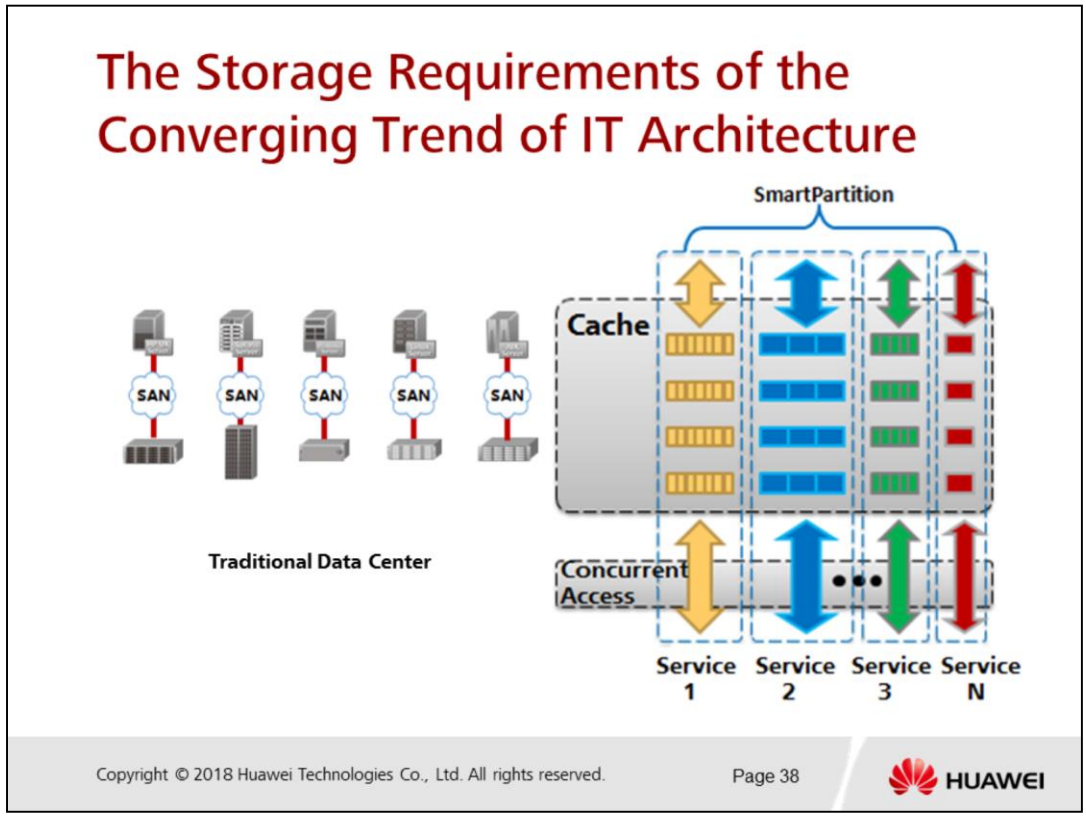


- The flowchart above shows the process of configuring the SmartQoS feature in Huawei OceanStor V3 storage systems.
- The SmartQoS configuration process includes checking the SmartQoS license and configuring the SmartQoS function.
- If performance is a major consideration for the services in a storage system, make sure you understand the true performance requirement of each application. The storage system provides a service monitoring function to help you find the performance bottlenecks of each application.

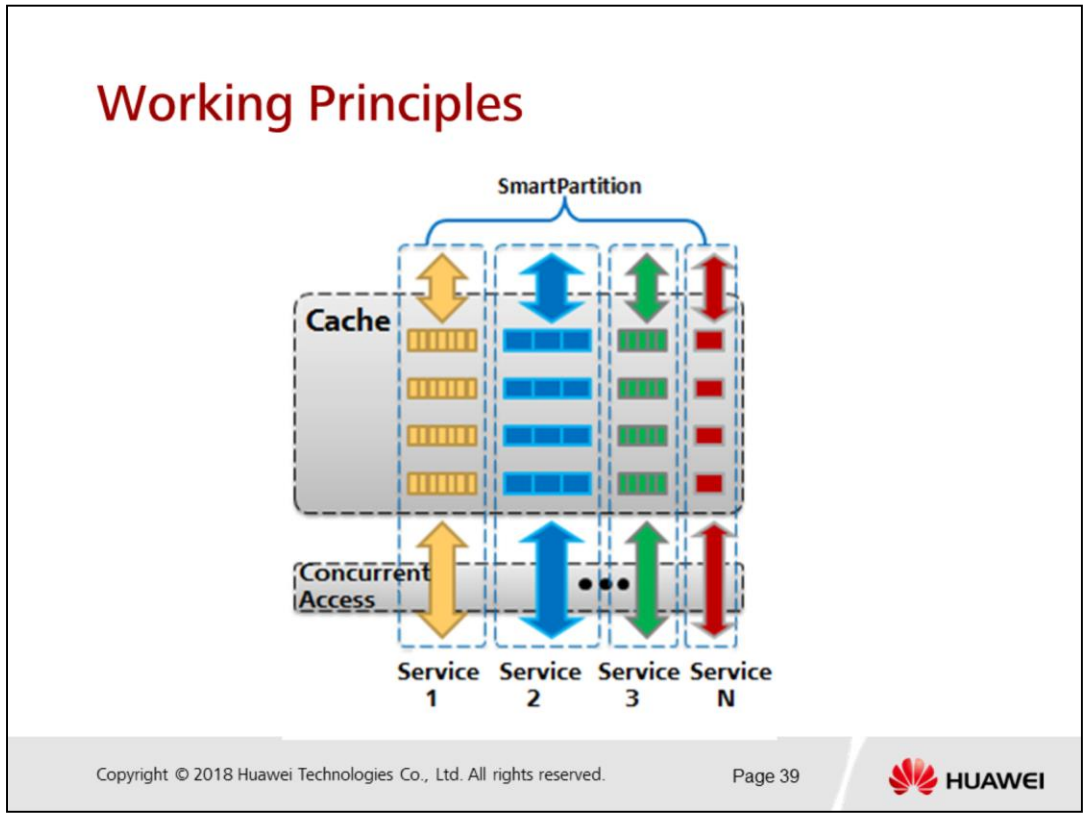
Contents



- SmartPartition can even cooperate with the other QoS technologies such as SmartQoS within the OceanStor V3 converged storage systems, in order to achieve better service quality assurance effect.

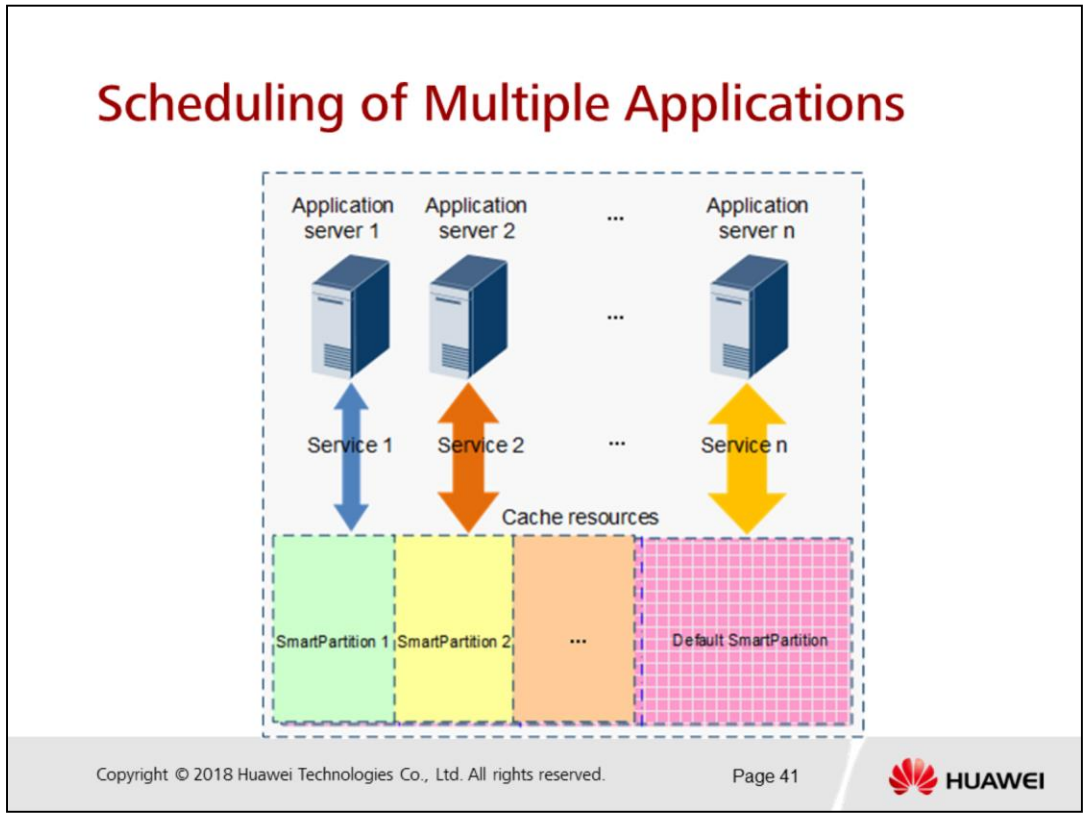


- Typical IT architecture consist of the combination of Compute, Network and Storage. In the traditional "Silo" architecture, different application systems are isolated, and a single storage only needs to face a low amount of application (commonly 5 or below) that consume the storage resources.
- The diagram on the right above represents different kinds of services using a different color. We can see that the I/O mode of these applications are different and has different requirements for cache and concurrent access.
- Different kinds if services are mixed together in the storage system and they compete for concurrent access resources and cache resources, causing unguaranteed quality of service. For example, concurrent access of service N doesn't match the concurrency of its host, causing the data for service N to become more and more in the cache and impact the performance of other services. If service 3 is a key business service, then it means that the business service performance cannot be guaranteed.

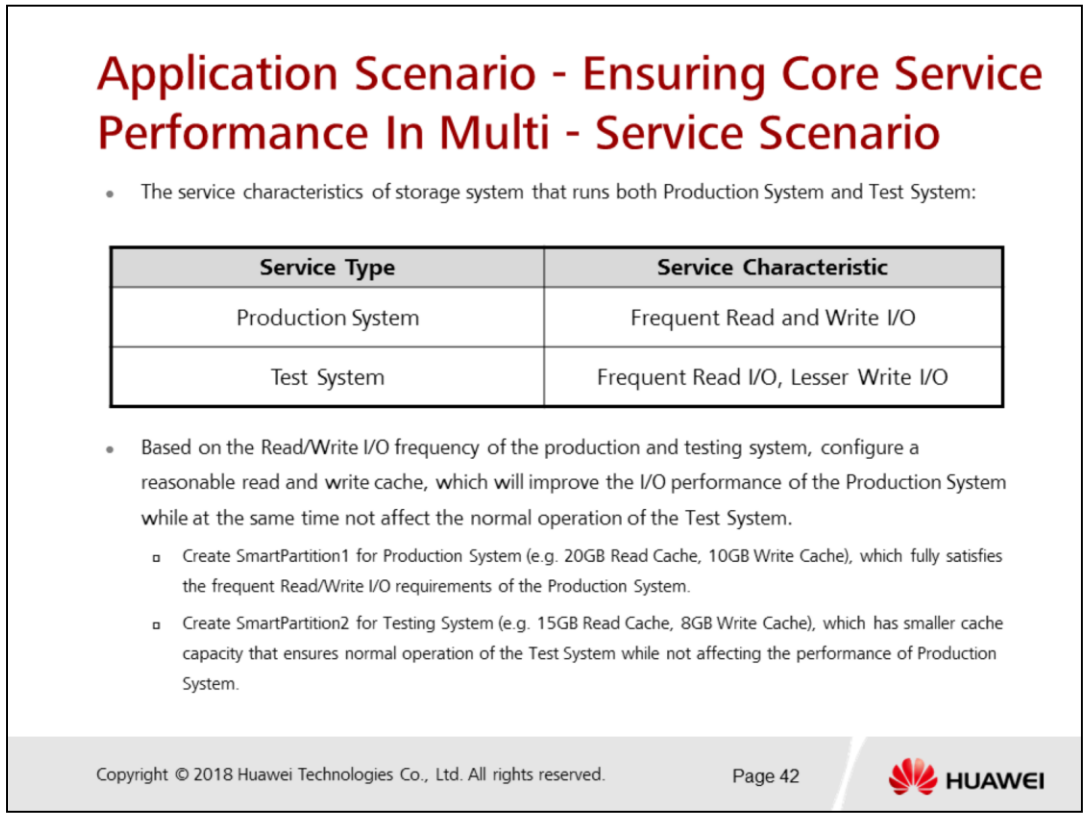


- Cache partitioning technologies ensure the service quality of key services through isolating and partitioning the required cache resources for the different services.
- In storage systems, the amount of cache that can be used by one service is the main factor of whether it can impact the service quality of other services.
- The size of the cache that a service can occupy on the storage system is the most important factor that affects the performance of the storage system. A large cache capacity translates into:
  - Higher write combination rate, higher write hit ratio, and better disk access sequence for an application that delivers write I/Os.
  - Higher read hit ratio for an application that delivers read I/Os.
  - For an application that delivers sequential I/Os, its cache capacity does not need to be large but must meet I/O requirements.
  - Higher access rate and better performance for an application that delivers random I/Os.

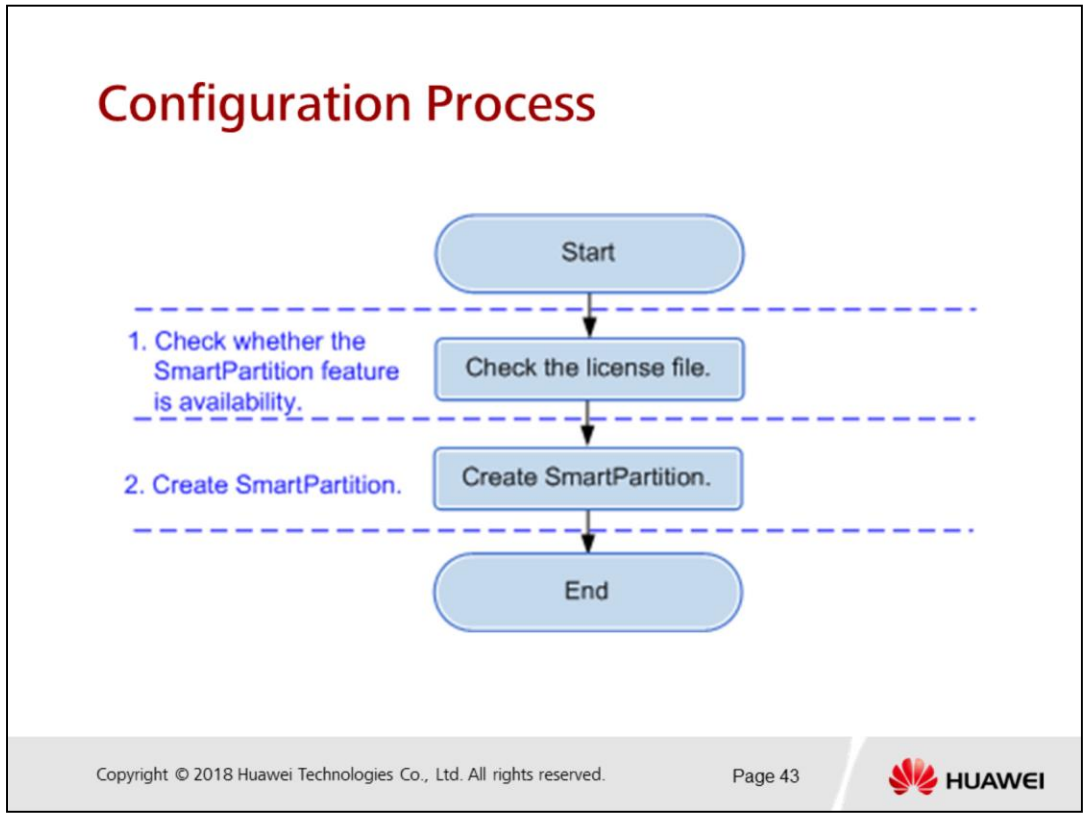
- Cache resources are divided into read cache and write cache.
    - Read cache improves the read hit ratio using the read prefetching mechanism.
    - Write cache accelerates host's access to disks using methods like combination, hit, and sorting.
- SmartPartition can allocate different sized partitioned cache resources to different services (actual control objects are LUN and File System), to ensure service quality for mission critical applications.



- SmartPartition partitions are created based on service LUNs or file systems. Each SmartPartition partition is exclusively accessed and does not interfere with other partitions.
- After you manually set cache capacity for a partition, SmartPartition will periodically analyze the number of I/Os processed by each partition to achieve an optimal configuration and ensure the QoS of mission-critical services.



- As the performance and capacity of a storage system keep growing, multiple applications are typically deployed in the same storage system to simplify the storage architecture and reduce configuration and management costs. However, those applications contend for storage resources, seriously affecting the performance of each service.
- Based on service characteristics, SmartPartition can allocate different sub partitions to different services, so that cache resources in a sub partition are exclusive to the corresponding service, thereby meeting the performance needs of different services and ensuring that mission-critical services run smoothly.
- At the same time, according to service read and write pressure, the corresponding read partition and write partition size can be set separately. If there is a large amount of read I/O traffic, you can allocate more read partition capacity; if more traffic on write I/O services, you can configure a smaller read partition capacity.

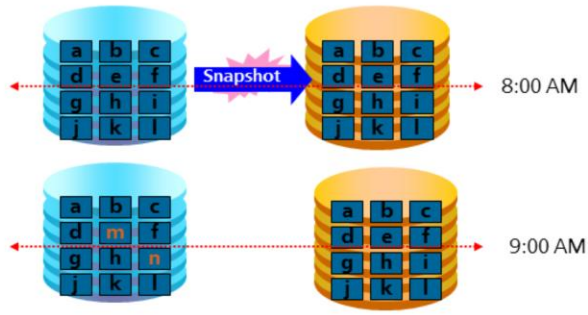


- The flowchart above shows the process of configuring the SmartPartition feature in Huawei OceanStor V3 storage systems.
- The SmartPartition configuration process includes checking the SmartPartition license and configuring SmartPartition functions.

Contents

LUN Snapshot Overview

- Definition: Snapshot refers to consistent data copies from the source data at a certain point of time. After the snapshot is generated, it can be accessed by the host or used as a backup of data at a certain point of time.
- The main features of Snapshot includes:
  - Instant Generation: Storage system can generate a snapshot in seconds, and obtain a consistent copy of the source data.
  - Takes up less storage space: The generated snapshot is not a full complete copy of the physical data, it does not takes up much of the storage space. Thus, even if the source data is huge, snapshots only takes up very less amount of storage space.

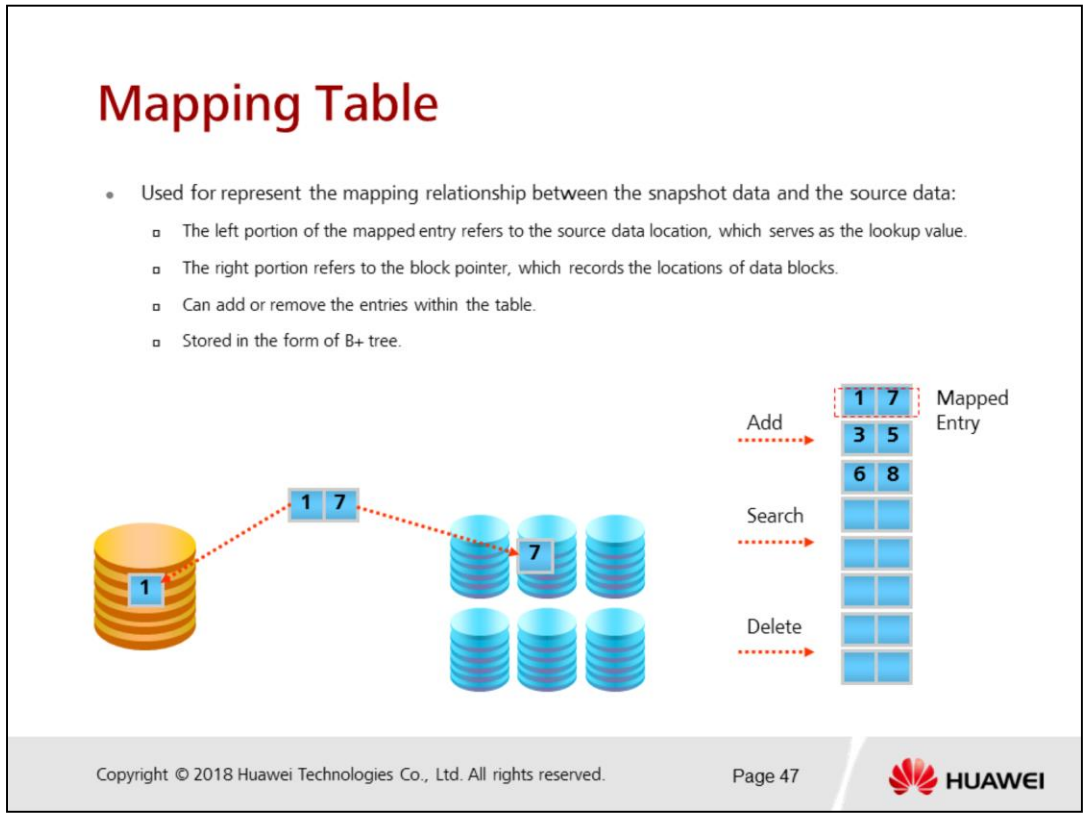


- Snapshot not only can quickly generate a consistent copy of the source volume at a certain point of time, it also provide the mechanism to restore the data in the source volume.
- When the data within the source LUN was intentionally deleted, destroyed or infiltrated by virus, we can quickly restore the source LUN using snapshot rollback to restore the data in the source LUN to the time when the snapshot was activated, in order to reduce the loss of data.

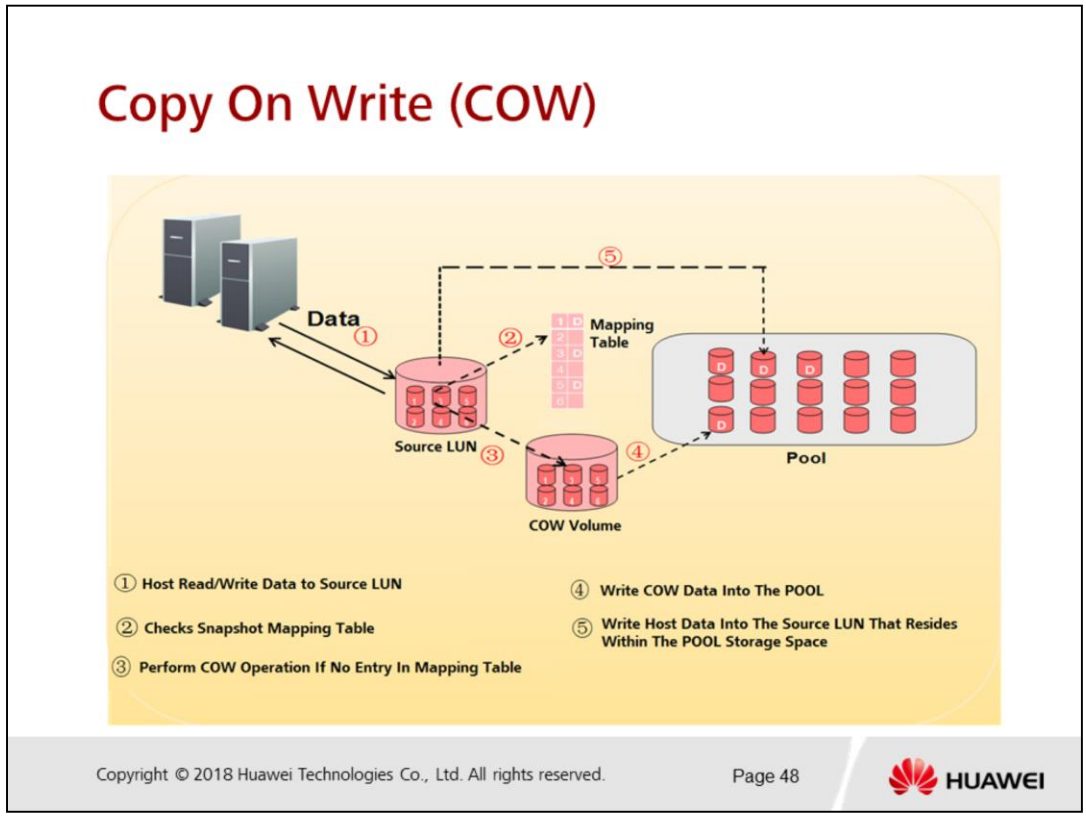
Related Concepts

Terms	Definition/Description
source volume	The volume where the source data resides for snapshot operation, it is represented to the host as LUN. Source Volume includes Meta Volume and Data Volume. -Meta Volume: Records the location of source data. -Data Volume: Records the data stored in the source volume.
COW data space	COW data space is dynamic storage allocated from the storage pool where the source LUN resides. This storage is used to store COW data following snapshot generation and activation. All snapshot LUNs of a source LUN share a single COW data space.
snapshot volume	It is the logical data copy generated after the snapshot is created from the source volume. It is represented as a LUN to the users.
snapshot rollback	Replicate the data from the snapshot LUN to the source LUN, which restores the source LUN data to the point where the snapshot was made.
mapping table	It is used to record the changes of data in the source and snapshot LUN at a certain point of time and the changed storage location of data. It is divided into shared and private mapping tables.
inactive	It is a status of the snapshot. In this status, snapshot are unable to be used, and requires activation operation before snapshots can be used as normal.

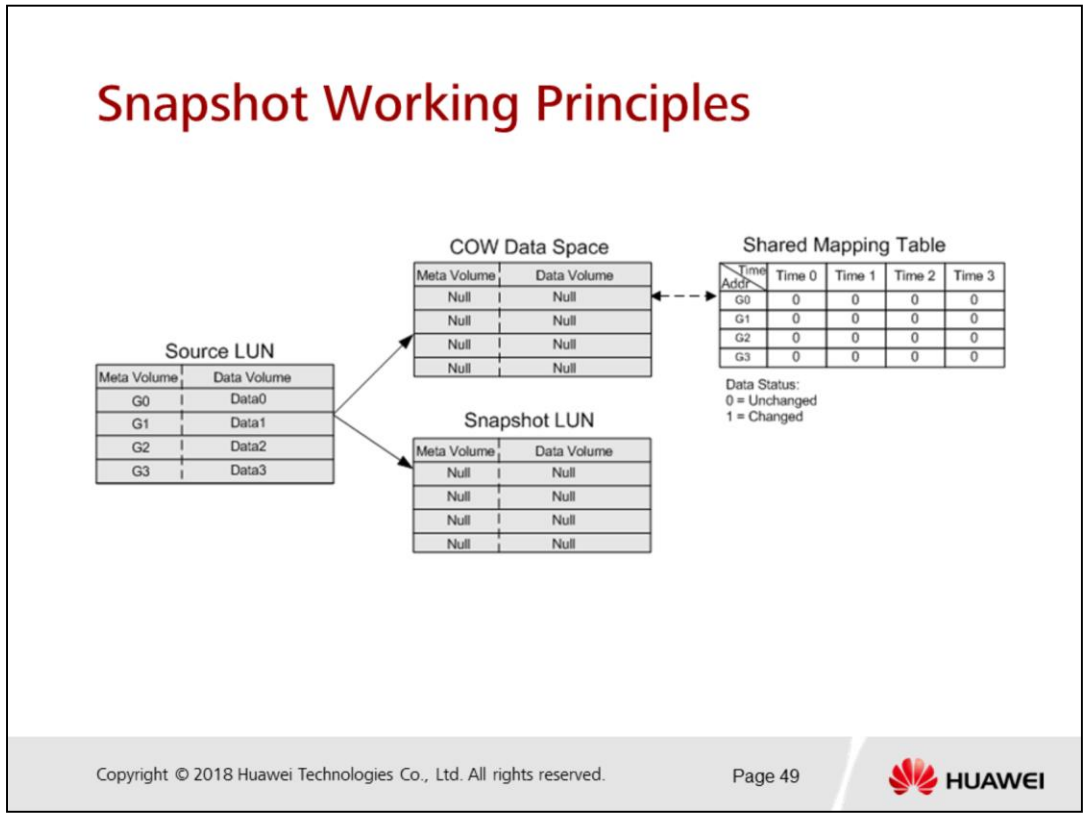
- The technical terms explained on the table above are quite commonly used to explain the different steps and status of Snapshot feature.



- Mapping table is used to represent the relationship between the snapshot data and actual data.
- Mapping table is divided into 2 types: Private mapping table and Shared mapping table. Both have the same working principles, but the main difference lies where the private mapping table records the changes of data in the snapshot LUN while the shared mapping table records the changes of data in the source LUN.

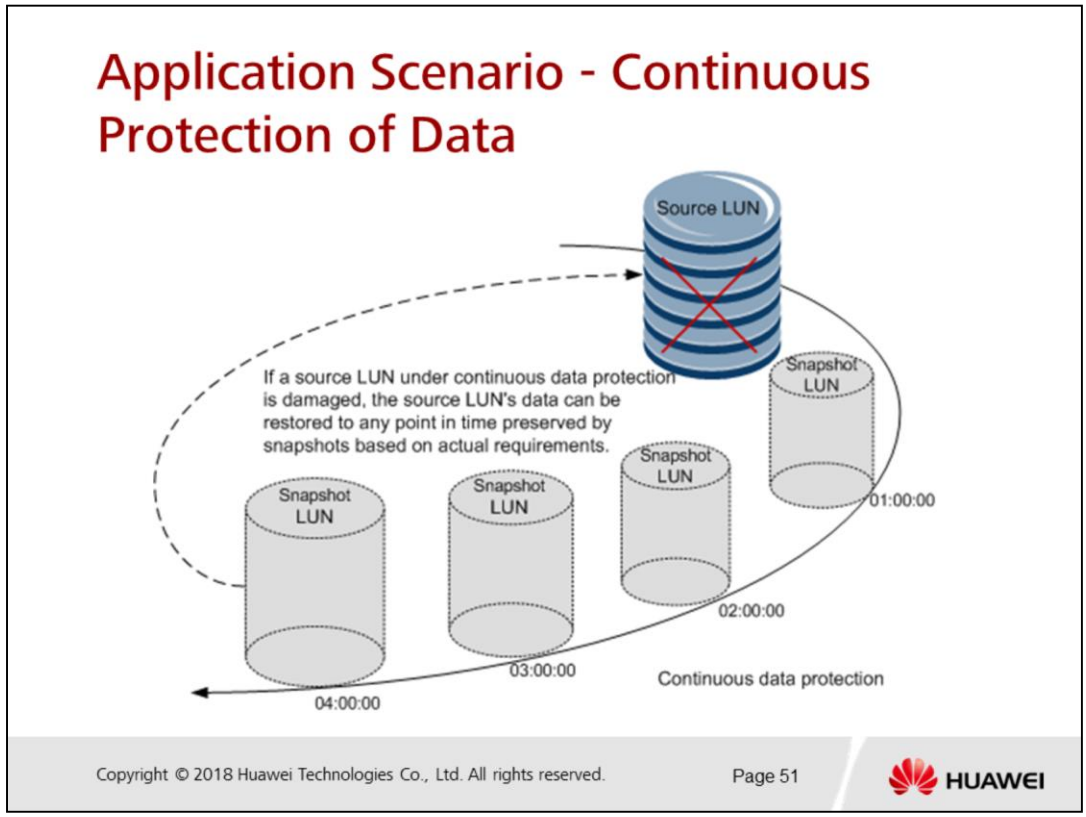


- When the snapshot is activated (snapshot point-in-time), the read/write operation of the hosts are as following:
  - After snapshot is activated, data is written to the source LUN.
  - It will first check the mapping table, if there is no mapping entry for the corresponding location, then it will perform copy-on-write(COW). Afterwards, it will record a backup copy of the source LUN data in the mapping table. If the mapping entry already exists, then it will directly overwrite the location on the source LUN.
  - COW reads the data in the corresponding location and copies it to the COW data space.
  - COW data space and the source LUN data space are distributed in the same POOL, which means that writing to the COW data space is the same as writing to the POOL data space.
  - After COW is completed, write the data from the host to the POOL space where the source LUN resides.
- Copy on write (COW) is a technology that saves changed data to a source LUN. If snapshots are active, when an application server writes data to the source LUN, the storage system performs the following steps:
  - Copies the to-be-replaced original data (COW data) to a COW data space.
  - Changes the mapping relationship of the COW data.
  - Records the new location of the COW data in the COW data space.
  - Writes the new data to the source LUN.

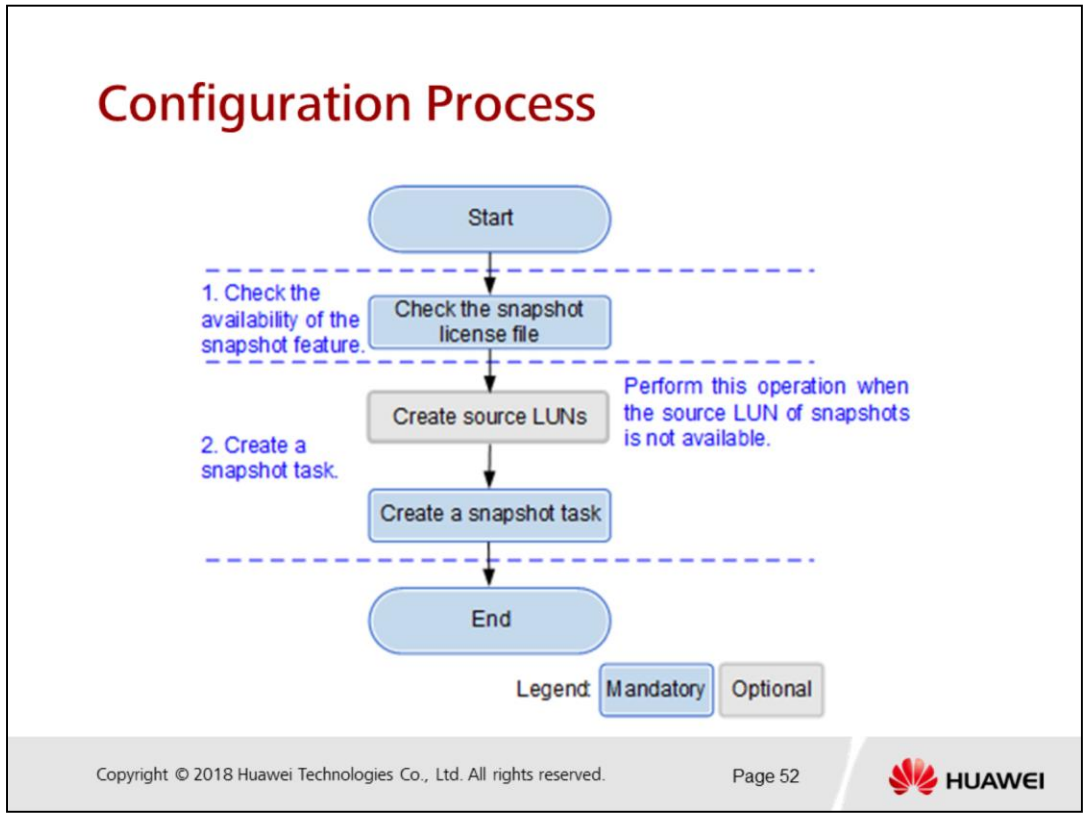


- Following the snapshot creation and activation, a data duplicate of the source LUN is generated. The storage system dynamically allocates COW data space from the storage pool where the source LUN resides, and generates a snapshot LUN.
- Due to the fact that no data is written to snapshot or source LUNs, thus no data is recorded to the COW meta and data areas or to the snapshot meta and data volumes. Diagram above shows the initial state of a snapshot after creation and generation.
- If an application server attempts to write data to the source LUN of an activated snapshot, the storage system performs the following before processing the write request:
  - Uses the COW mechanism to copy COW data to the COW data space.
  - Changes the mapping relationships in the mapping table.
  - Writes new data to the source LUN
- In a snapshot period, COW is performed only once for data of each location. Following that, newly written data overwrites existing data. For example, if the corresponding value of DataX is recorded as 1 in the mapping table, COW has been performed for DataX. For data write requests made in the future, the storage system will write data directly to the source LUN without copying DataX to COW data space.

- Following snapshot activation, application servers are able to send read and write requests to the snapshot LUN. Write requests are directly processed by the snapshot LUN, and the private mapping table records the data location in the snapshot LUN.
- Following snapshot activation, when an application server attempts to read snapshot data, the storage system processes the read request as follows:
    - The storage system searches the private mapping table for the data location in the storage volume.
    - If the data is in the snapshot LUN, the data is returned to the application server.
    - If the snapshot LUN does not contain the data, the storage system searches the shared mapping table.
    - The storage system searches the shared mapping table for the data location in the COW data space and source LUN.
    - If the search result is 0, data is read from the source LUN.
    - If the search result is 1, data is read from the COW data space.
- If the application server has written data into the snapshot LUN, the application server can read snapshot data directly.
- If the application server has written data to the source LUN but not the snapshot LUN, the application server reads snapshot data from the source LUN or COW data space.
- Consider a copy-on-write system, which copies any blocks before they are overwritten with new information (i.e. it copies on writes). In other words, if a block in a protected entity(source LUN) is to be modified, the system will copy that block to a separate snapshot area before it is overwritten with the new information. This approach requires three I/O operations for each write: one read and two writes. Prior to overwriting a block, its previous value must be read and then written to a different location, followed by the write of the new information. If a process attempts to read the snapshot at some point in the future, it accesses it through the snapshot system that knows which blocks changed since the snapshot was taken. If a block has not been modified, the snapshot system will read that block from the original protected entity. If it has been modified, the snapshot system knows where the previous version of that block is stored and will read it from there. This decision process for each block also comes with some computational overhead.

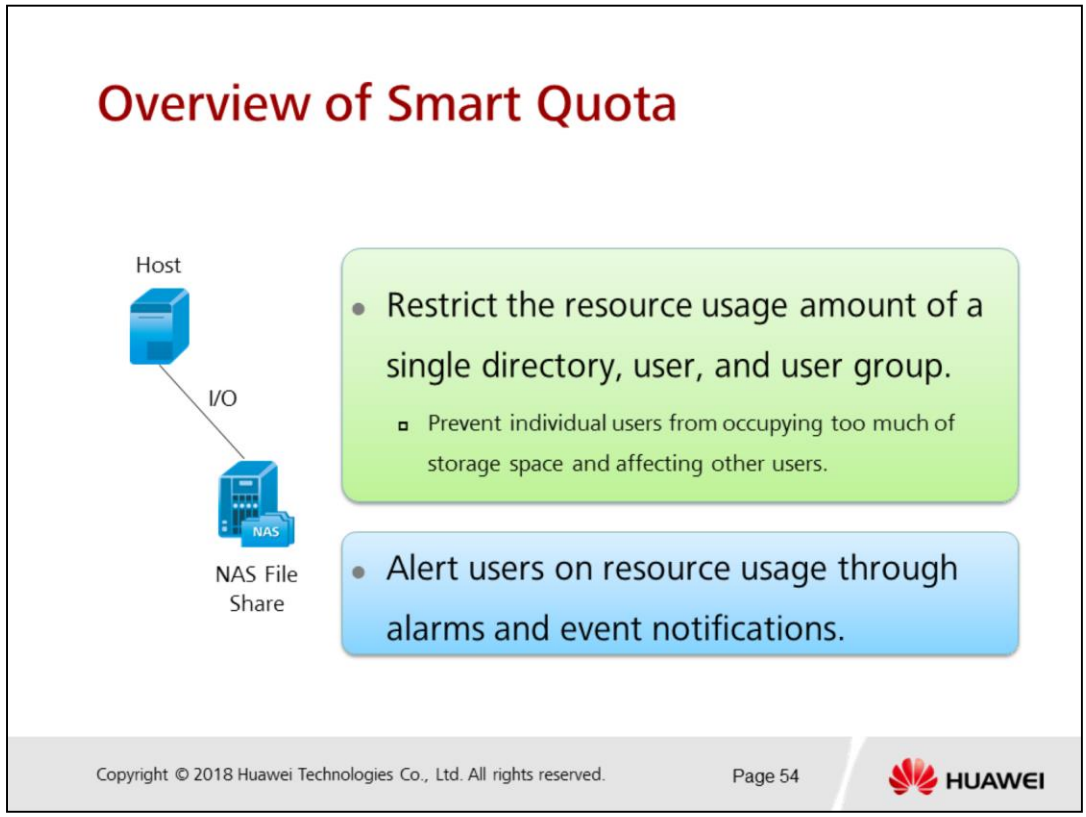


- Snapshots can be used directly as point in time backup copies of data. The usage of snapshot allows quick recovery of data in the following scenarios:
  - Virus infection.
  - Human error in operation.
  - Malicious tampering.
  - Data corruption due to system downtime.
  - Data corruption due to BUG in applications.
  - Data corruption due to BUG in storage systems.
  - Damaged storage media (Only snapshots based on split mirror technology can recover data in this case).

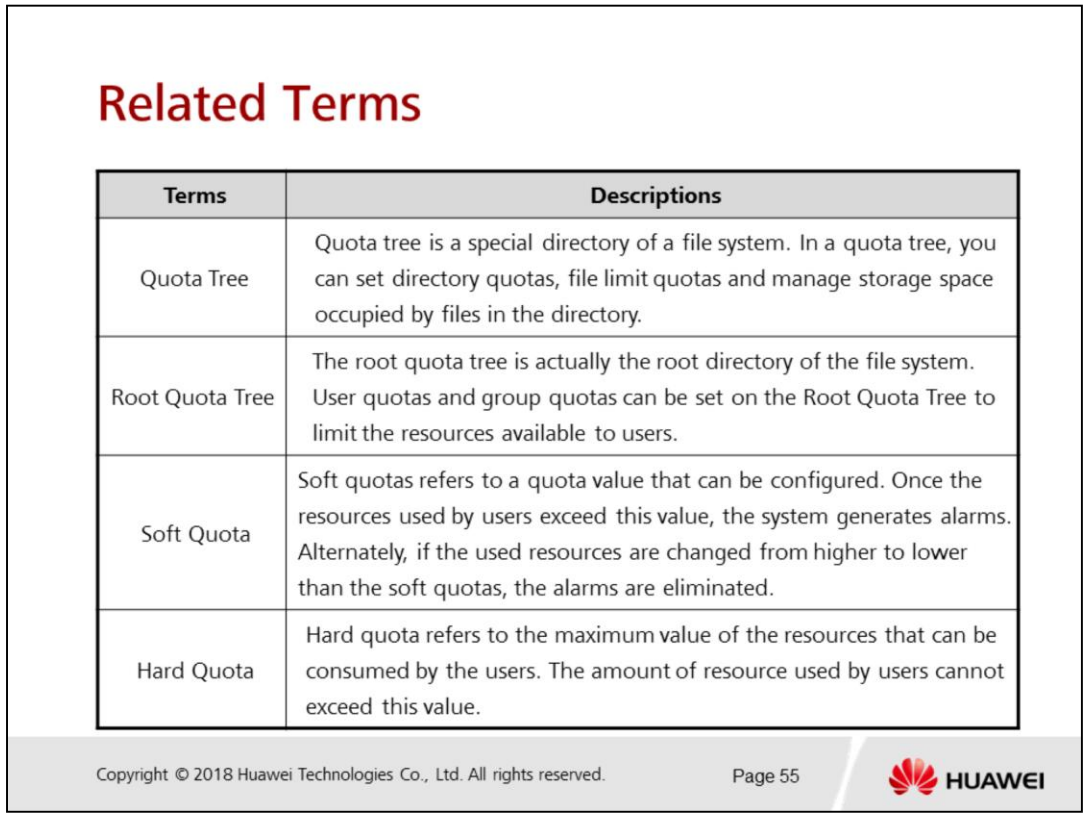


- The flowchart above shows the configuration process for Snapshot feature in Huawei OceanStor V3 storage systems
- Configuring a snapshot task includes checking the license file for the snapshot feature, creating LUNs, and creating snapshots.
- Steps of configuring a snapshot task:
  - Check the availability of the snapshot feature by checking the snapshot license file.
  - Create a snapshot task by either optionally create a source LUN for snapshot activation or create a snapshot task following the wizard instruction on a existing source LUN.

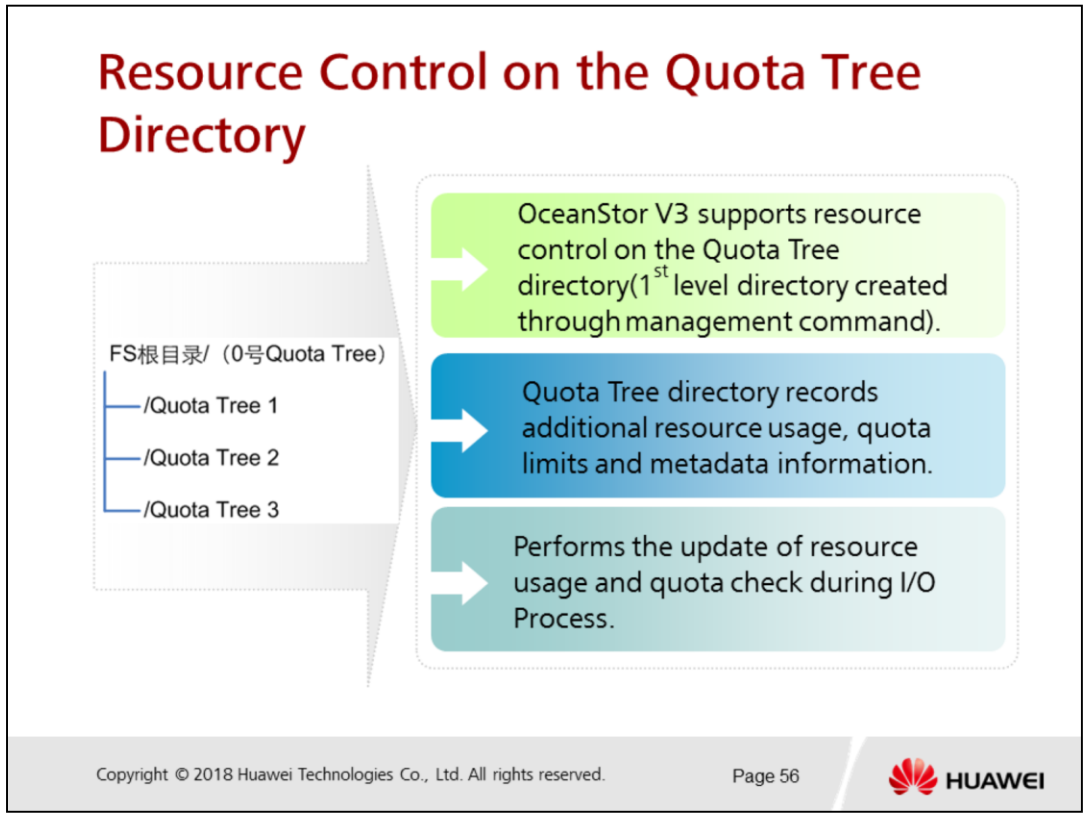
Contents



- With the development of virtualization and cloud computing technologies, IT systems faces the challenges of resource utilization efficiency and how to effectively manage them. In a typical IT storage system, as long as there are usable storage resource (disk space), then it will be utilized by the users until it has depleted. From an economical point of view, we need a method to control the usage of storage resources and its growth.
- In the NAS file service environment, resources are usually provided to departments, organizations or users in the form of shared directories, and each of them has unique storage requirements and limitations. Thus, systems need to allocate resources and restrictions to each user based on shared directory or local conditions. SmartQuota is the technology developed to solve this requirement. It can be used to perform resource allocation and restrictions based on directories, users, and user groups.
- The main role of SmartQuota is to make things easier for system administrators to control the usage of the storage resources by the users (including directories, users and user groups) and to limit the disk space usage of specific users to avoid the problem of over-consumption of storage resources by some users.

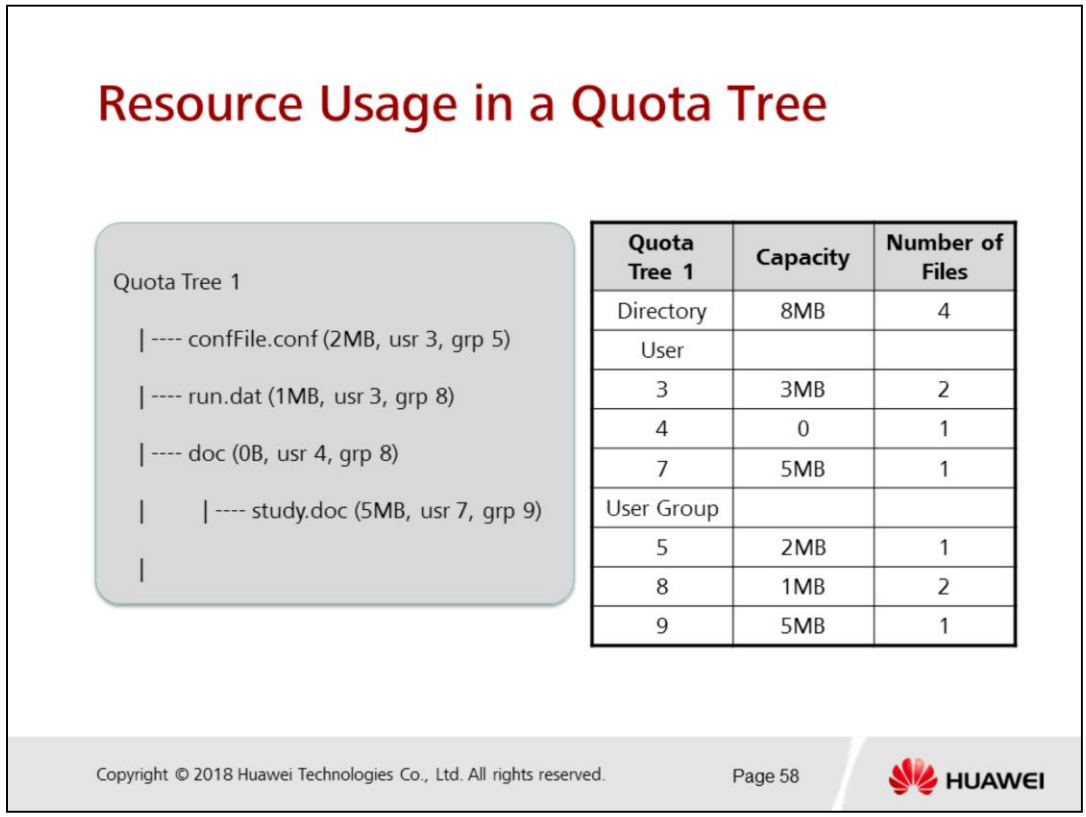


- The table above explains the key terms involved in the SmartQuota feature. It is important to differentiate the key differences between the soft and hard quota when configuring the resource limitation for users using the SmartQuota feature.

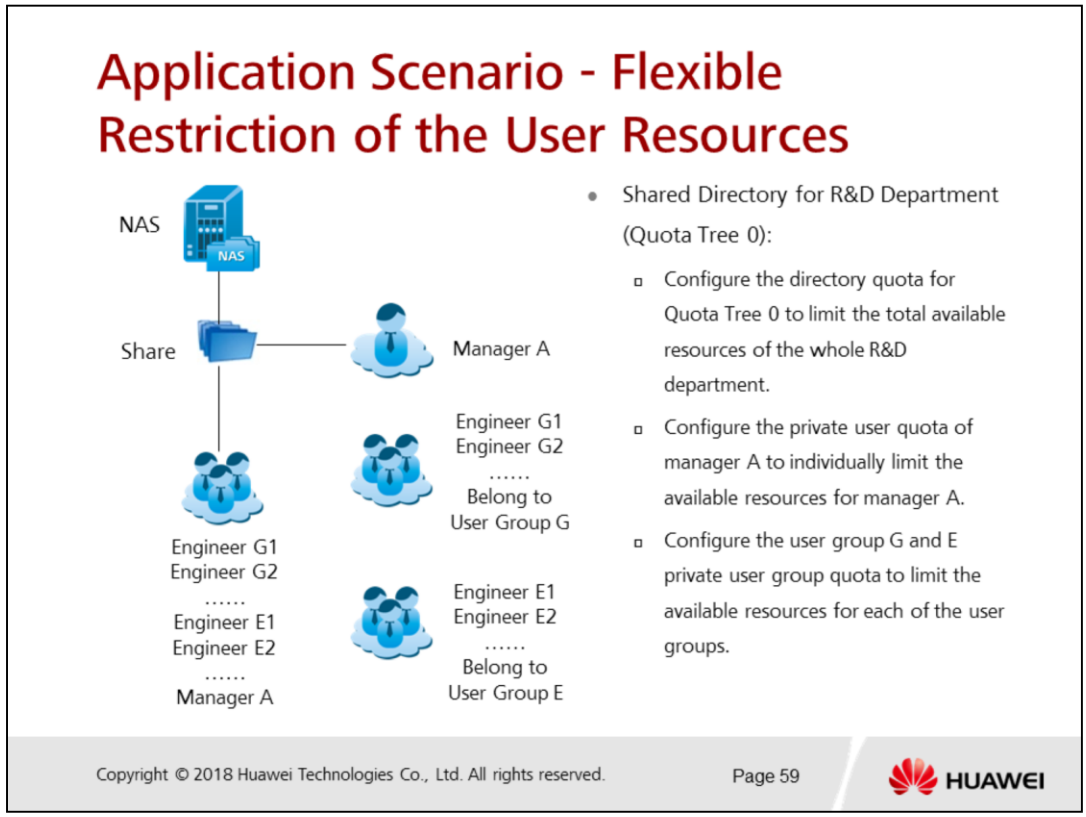


- In the SmartQuota feature, the Quota Tree is an important concept. In simpler words, Quota Tree is the first level directory of the file system. From the management viewpoint, Quota Tree is not just a directory, but also a configuration entity. This means that, Quota Tree can only be created, deleted or modified through the management terminal (Command Line or GUI Management Interface), and it cannot be modified through the client terminal or hosts. Additionally, it serves as the entity to configure the directory quota, user quota, and user group quota, and these quotas can only be configured on the Quota Tree itself. In the following, lets summarize the differences between the Quota Tree and normal directories:
- Quota tree can only be created, deleted, renamed through command line or GUI management interface by the system administrator. System administrators can only delete empty quota trees.
- Quota tree can be shared using protocols, but no rename or deletion is allowed when the quota tree is shared.

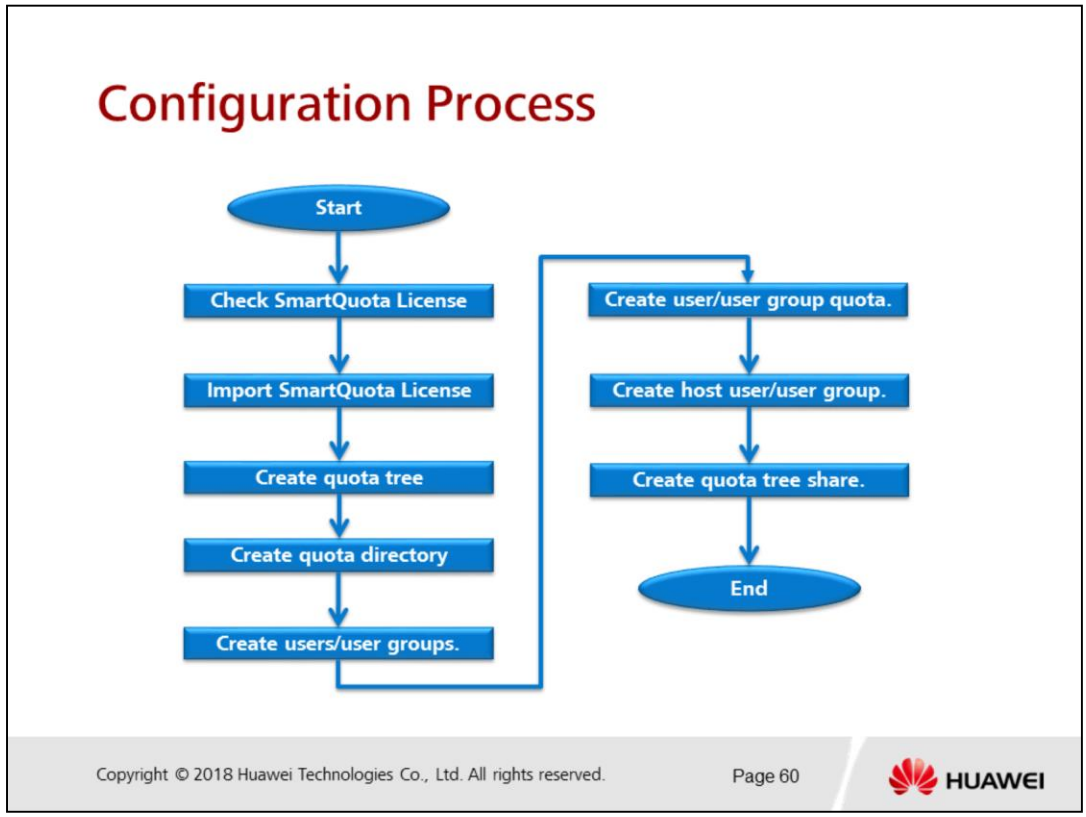
- File moving operations (under NFS protocol) or File Copy & Pasting operations (under CIFS protocol) are not allowed across Quota Tree. This means that the files cannot be moved or copy/pasted between 2 different Quota Tree directories under CIFS/NFS protocols.
- Hard links between Quota Trees are not allowed, which means that hard link operation between 2 different Quota Tree are not permitted.
- In computing, a hard link is a directory entry that associates a name with a file on a file system. All directory-based file systems must have at least one hard link giving the original name for each file. The term “hard link” is usually only used in file systems that allow more than one hard link for the same file.



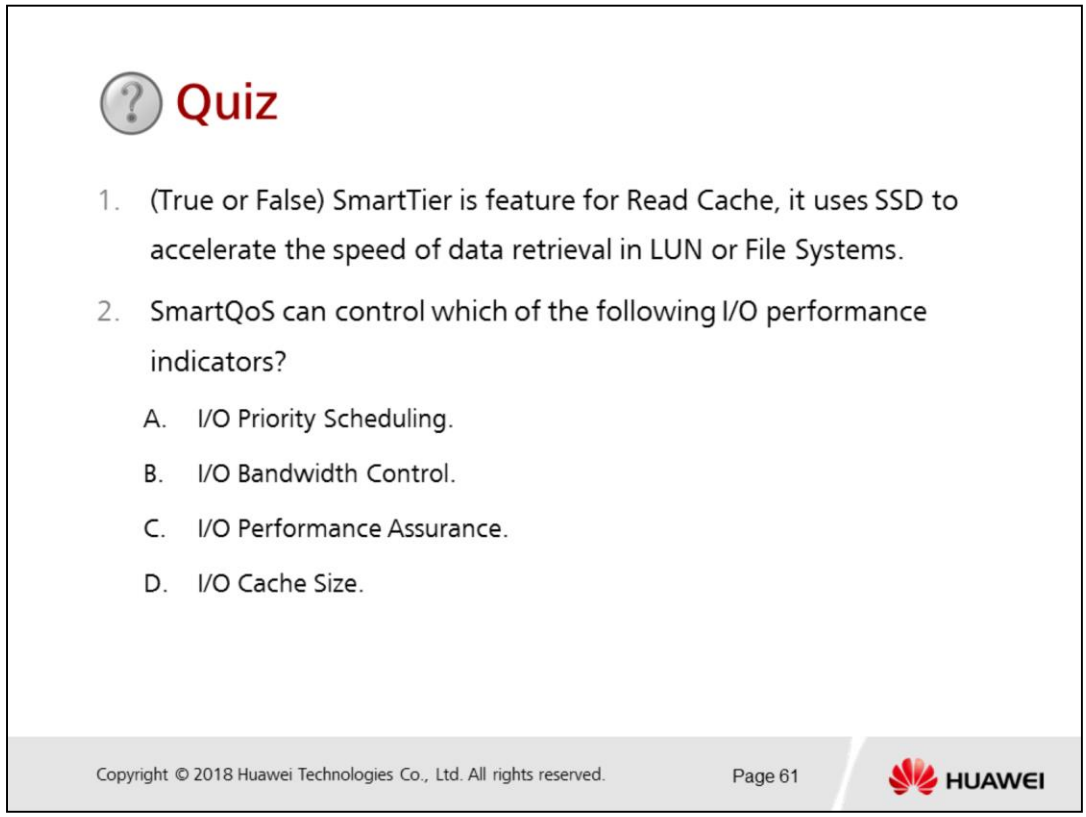
- Resource usage of directories (Statistics of directory quotas)
  - The total amount of the storage capacity and number of files for all the files within the directories of Quota Tree.
- Resource usage of user/user groups (Statistics of user/user group quotas)
  - The total amount of the storage capacity and number of files for all the files in the directory created for the specific user/user group within the directories of Quota Tree.



- SmartQuota feature allows flexible restriction or limitation on the user resources within department by implementing 3 different types of quota which are :
  - Directory Quota.
  - User Quota.
  - User Group Quota.
- Within a single Quota Tree, the system administrators can configure the directory quota to limit the total amount of resources available for the corresponding department, then configure the user quota and user group quota for each user and user groups in the department to flexibly limit the amount of resource usage of each users and groups.



- The flowchart above shows the configuration process of the SmartQuota feature in Huawei OceanStor V3 storage system.
- The configuration process involves checking the license for the feature availability, if it is not available then the license must be imported to activate this feature.
- Subsequently, create the quota tree and the quota directories and configure the users/user groups and their quotas respectively.
- Finally, create the hosts user/user groups and link them together with the quota directories through sharing protocols.



- Answers:
  - False.
  - ABC.

Summary

- This module introduced the common advanced features used in storage systems which are:
  - SmartThin
  - SmartTier
  - SmartQoS
  - SmartPartition
  - Snapshot
  - SmartQuota

Thank You

www.huawei.com