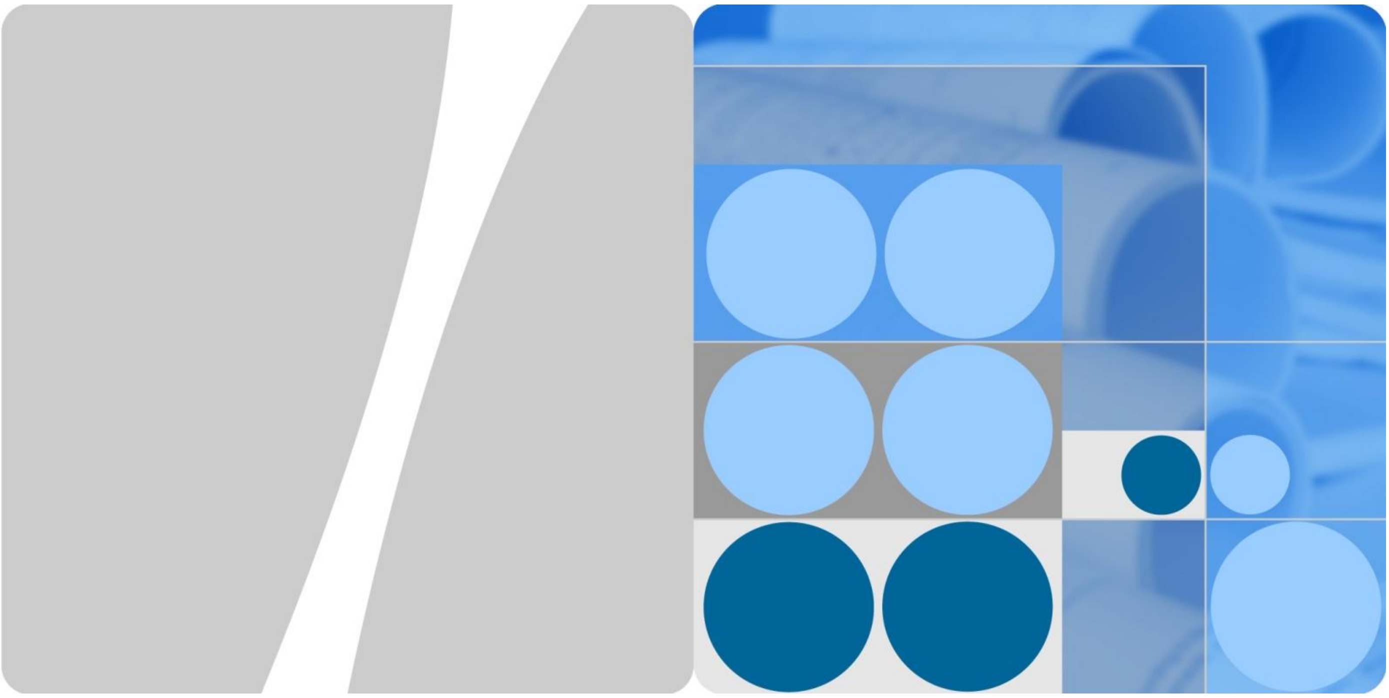
Doc. code



OceanStor VTL6000 Technical White Paper

Issue V2.2

Date 2012-05-05

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions

Notice

The purchased products, services and features are stipulated by the commercial contract made between Huawei and the customer. All or partial products, services and features described in this document may not be within the purchased scope or the usage scope. Unless otherwise agreed by the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute the warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://www.huawei.com

Email: support@huawei.com

Contents

Revision Records

Date	Version	Description	Author
2010-2-23	1.2	First release. Versions 1.0 and 1.1 are unavailable.	Wang Ye
2010-3-8	1.3	The description of customer benefits is modified in section 4.1 according to review comments. High performance and large capacity features are described separately.	Wang Ye
2010-4-13	V2.1	The description of SIR engines is modified in section 3.2.2	Wang Ye
2012-5-5	V2.2	Update some parameters	Zhu Feng

1 Executive Summary

This technical white paper analyzes the key technical features, application scenarios, solutions and customer benefits of the VTL6000, and shows that the VTL6000 is an ideal virtual tape library (VTL) solution for medium- and high-end users.

2 Introduction

As data increases explosively, traditional tape backup systems with low backup and recovery rate fail to satisfy the backup requirements of users. Meanwhile, the rapid development of disk technologies helps to increase the disk storage capacity greatly when reducing the disk storage cost substantially. Backup systems based on SATA disks are widely deployed in users' IT systems thanks to their excellent backup and recovery performance. VTL is one of the representative products for those solutions. VTL products feature high performance, easy maintenance, and advanced media management capacity of tapes. VTL products, which are well recognized once launched, enjoy promising market prospect.

For medium- and high-end users, a full backup may involve more than 10 TB of data. To back up such amount of data in eight hours, the backup system needs to process at least 1,000 MB of data per second. In addition, the backup devices need to support a storage capacity of several hundred TB if a user requires daily incremental backup, weekly full backup, and backup data storage for three months. It is costly to construct, operate and manage such a disk backup system. In addition, medium- and high-end users also expect high availability of the backup systems. Single-engine VTL products, however, are unable to meet the high availability requirements obviously.

In a word, the problems that medium- and high-end users have to solve are poor performance, insufficient storage capacity, high energy consumption, and low availability when choosing VTL products.

3 Solution

The VTL6000 is a VTL product specialized for medium- and high-end users. The VTL6000 integrates deduplication, HA cluster, and hibernation technologies to provide an ideal VTL solution for medium- and high-end users.

This chapter describes the deduplication, HA cluster, and energy-saving features of the VTL6000. For information on tape caching, remote replication, and other advanced VTL features, see the *Technical White Paper for the VTL3500*.

3.1 Deduplication

3.1.1 Introduction to Deduplication

Deduplication is a technology for deleting duplicate data from the stored data by software or hardware. Deduplication helps to save storage space.

In a backup system, data is transmitted from the backup client (the source) to the backup device (the target) under the control of the backup server. By location where the deduplication process takes place, deduplication is classified into two types: deduplication at the source and deduplication at the target. In the case of deduplication at the source, the backup client deletes the duplicate data before sending the data to the backup device. In the case of deduplication at the target, the backup client sends the data to the backup device without deduplication. It is the backup device that deletes the duplicate data from the received data.

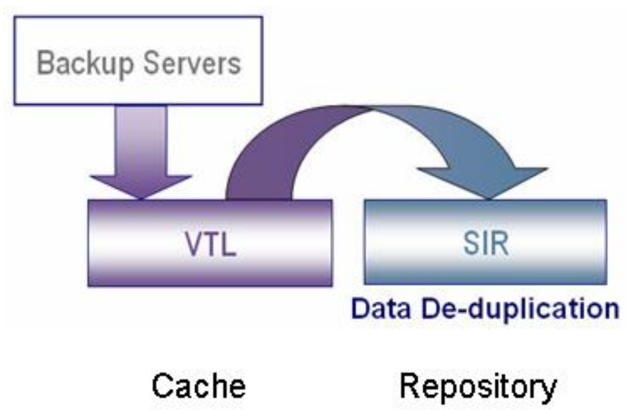
By time when the deduplication process takes place, deduplication at the target is also classified into the following two types: inline deduplication and post processing deduplication (also called offline deduplication). In the case of inline deduplication, the backup device deletes the duplicate data while receiving backup data from the backup client. Deduplication is completed when the backup device receives all backup data. In the case of post-processing deduplication, the backup device has to receive all backup data before deleting the duplicate data.

The ultimate purpose of deduplication is to eliminate duplicate data, regardless of the particular deduplication methods. Therefore, all deduplication methods involve comparison between the new data and the existing, thus determining whether the new data is duplicate. Duplicate data can be identified in either of the following ways: index-based comparison and content-based comparison. If index-based comparison is used, the duplicate data is identified through data index comparison. If content-based comparison is used, the duplicate data is identified through data comparion. In data indexing, data is divided into data blocks of fixed of unfixed length. Each data block is processed based on an algorithm, thus obtaining a unique value of the data block. The unique value is the index of the data block. Data indexes occupy much less storage space than data blocks do. Therefore, index-based comparison can occur in the memory and achieve higher efficiency than content-based comparison. Currently, index-based comparison is widely used in various deduplication technologies.

3.1.2 Deduplication Concept of the VTL6000

The VTL6000 adopts the offline Single Instance Repository (SIR) technology and index-based comparison for deduplication.

Figure 3-1 Concept of the SIR technology

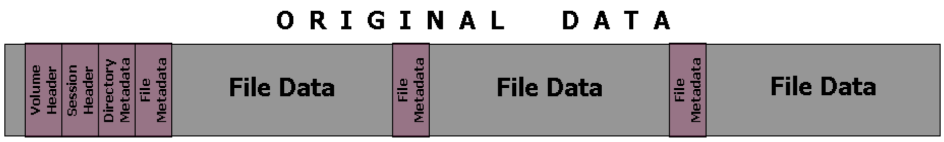


The VTL software of VTL6000 contains two modules: a VTL module and an SIR module. Correspondingly, the storage space of the VTL6000 is logically divided into the VTL storage space (also called the cache) and the SIR storage space (also called the repository) for the VTL module and the SIR module respectively. The VTL6000 receives backup data and store the data in the VTL storage space. The SIR module then reads the stored data at a certain time (for example, at a specific time, upon the completion of the backup, or when the storage threshold is reached), and compares them with the existing data blocks stored in the SIR storage space In the following way:

The SIR module divides the original backup data into data blocks of several KB, works out a unique hash number (also called an index) for each data block based on the SHA-1 algorithm, and then compares the hash numbers of the new data blocks with those of the existing data blocks to determine whether the new data blocks are duplicate with the existing ones. The SIR module discards the duplicate data blocks except for their data block pointers and store the new data blocks in the SIR storage space, as shown in Figure 3-1.

In the VTL6000, the VTL module and the physical device where the VTL module runs constitute the VTL engine. The SIR module and the physical device where the SIR module runs constitute the SIR engine. The VTL engine is a mandatory component that provides the virtual tape library service, while the SIR engine is an optional component that provides the deduplication function. The VTL engine and the SIR engine can be integrated into the same engine or configured as two separate engines. The VTL6000 supports up to three (2+1) SIR clusters.

Figure 3-2 Original data on the virtual tapes in the VTL6000



Before the SIR engine performs deduplication, the backup data in the VTL storage space is stored on the virtual tapes, as shown in Figure 3-2.After deduplication, the data on the virtual tape are replaced with pointers (here the virtual tapes are called virtual index tapes (VITs)). Each pointer refers to a unique data block in the SIR storage space. The released VTL storage space is available for new backup data, as shown in Figure 3-3.

Figure 3-3 Data distribution in the VTL6000 after deduplication

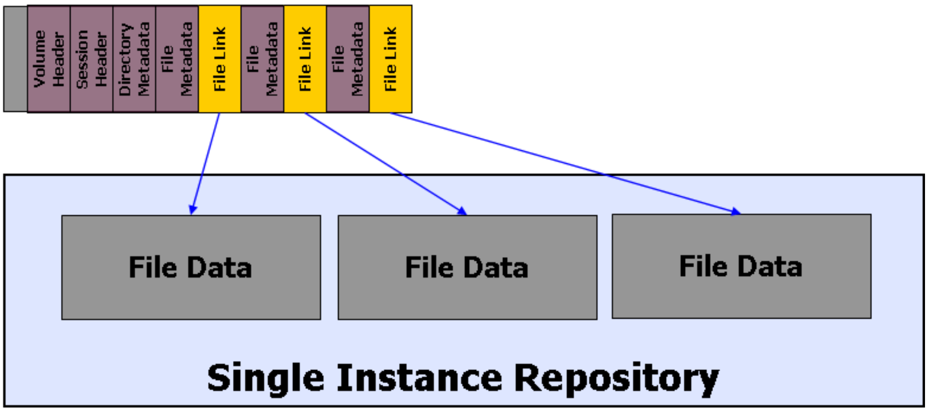
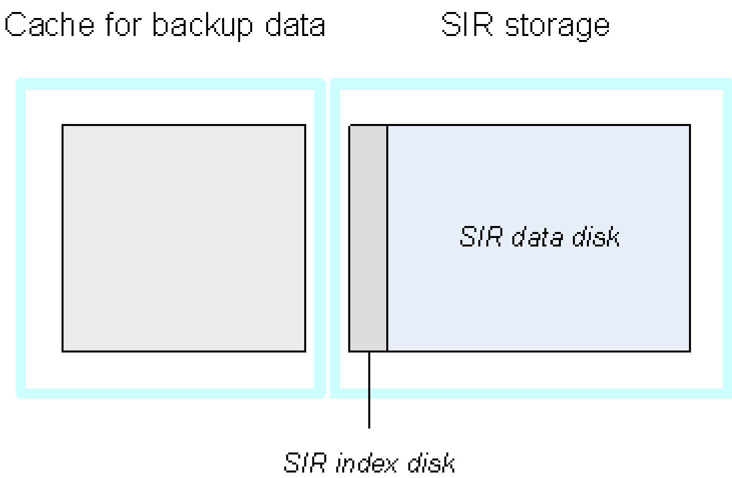


Figure 3-4 Storage space allocation in the VTL6000



As mentioned previously, the storage space of the VTL6000 is logically divided into the VTL storage space (the cache) and the SIR storage space (the repository). The SIR storage space is further divided into the SIR data disk and the SIR index disk, as shown in Figure 3-4. The SIR data disk stores the data blocks after deduplication. The new backup data in the cache are compared with the data blocks in the SIR data disk to identify duplicate data. The SIR index disk stores the indexes (that is, the hash numbers computed based on the SHA-1 algorithm) of all data blocks stored in the SIR data disk. Therefore, the more data the SIR data disk stores, the more indexes the SIR index disk has. In the operation of the SIR module, all indexes (the index table) in the SIR index disk are read into the memory of the SIR engine for quick search for duplicate data from the index table. Therefore, the more indexes the SIR index disk has, the larger memory the SIR engine needs.

For more details on deduplication technologies, see the *Technical White Paper for Deduplication*.

3.2 HA Cluster

3.2.1 Introduction to HA Clusters

If a node in a high availability (HA) cluster fails, another node in the cluster takes over the tasks of the failed node to guarantee high availability.

Nodes in an HA cluster are categorized into the active nodes and the standby nodes. The active nodes are nodes that perform the tasks, and the standby nodes are the backups for the active nodes. Once an active node fails, the standby node takes over the tasks of the failed node.

HA clusters work ased on resource failover. Resources refer to all things that the standby node needs to take over from a failed node in the HA cluster, such as the IP address, host name, disk volume, and application context. After taking over these resources from the failed node, the standby node can continue to provide the services originally provided by the failed node, thus minimizing the influence on the client.

Resource monitoring and failover rely on the HA software. Most operating systems are integrated with HA software, and various manufacturers have developed their own HA software. All HA software is used to enable the standby node to monitor the status of the active node; and thus the standby node can take over the resources from the active node and continue to provide the relevant service once a failure of the active node is detected.

NOTE

This section results from references to the cluster section in the book *About Storage*.

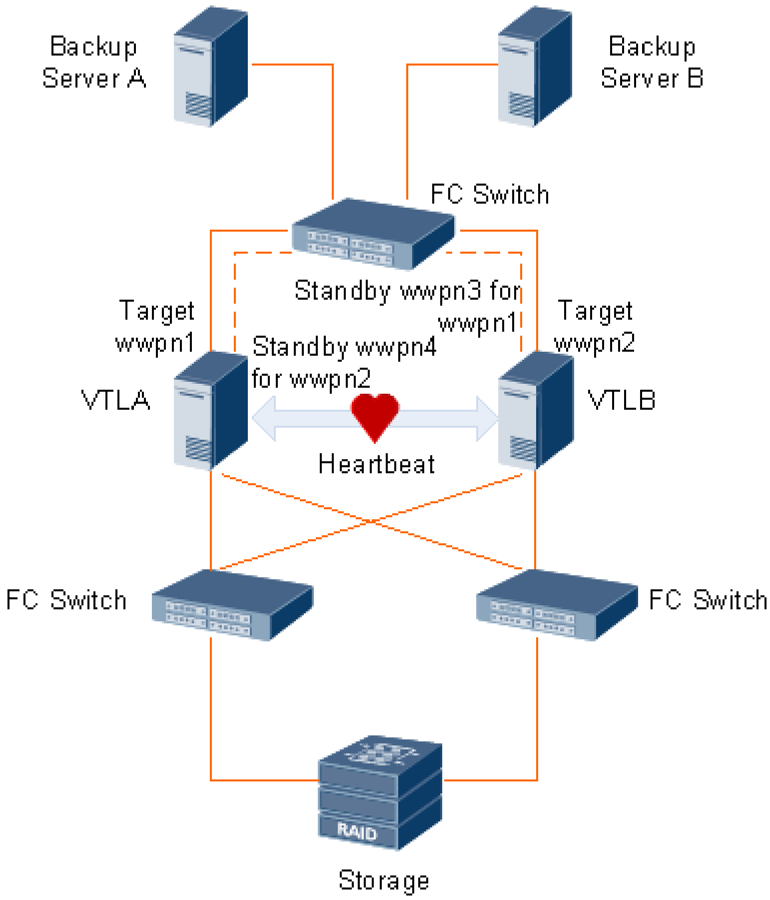
3.2.2 High Availability of the VTL6000

Both the VTL module and the SIR module of the VTL6000 support high availability. That is to say, high availability of the VTL6000 results from the HA configurations of the VTL engine and those of the SIR engine.

HA Configurations of the VTL Engine

The VTL6000 supports HA configurations on two VTL engines. The two VTL engines can be fixed as active or standby nodes (for unidirectional failover), or be configured as mutually converted active and standby nodes (for bidirectional failover).

Figure 3-5 HA configurations of VTL engines



As shown in Figure 3-5, both VTLA and VTLB provide the VTL service when they work normally. Assume that VTLA is the active node and VTLB is the standby node in the case of unidirectional failover. VTLB can monitor VTLA and take over the the VTL service of VTLA when VTLA fails. VTLA, however, cannot take over the VTL service of VTLB when VTLB fails. If bidirectional failover is used, VTLA and VTLB can monitor each other. If either engine fails, the other can take over the resources from the failed one and continue to provide the VTL service.

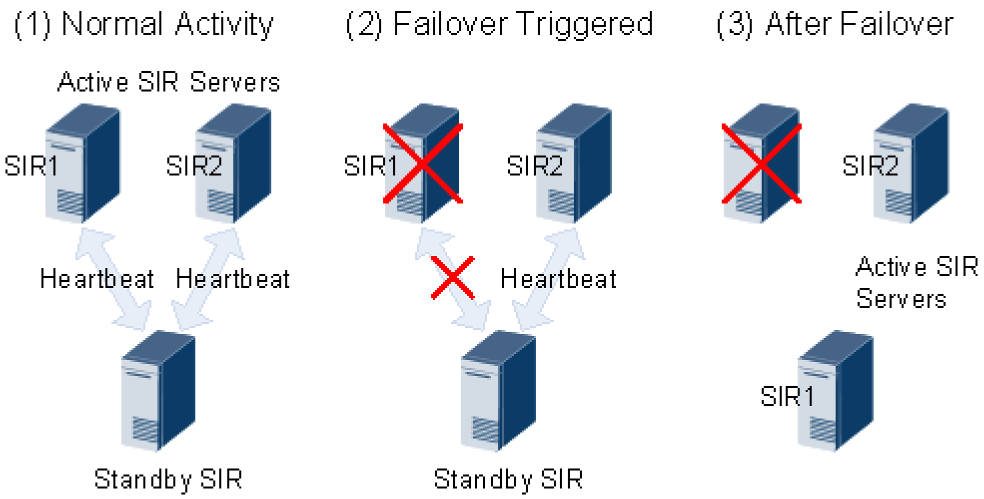
As shown in Figure 3-5, VTLA refers to the VTLA software application and the relevant physical device, and VTLB refers to the VTLB software application and the relevant physical device. Failover means that if VTLB takes over VTLA, the VTLA software application originally running on the physical device of VTLA is running on the physical device of VTLB. How does the VLTA software application running on the physical device of VTLB access their original VTL storage space (VTL storage) and continue to provide the VTL service for their original backup server? The answers are shared storage and standby backup host interfaces. Shared storage means that the physical device of VTLA and that of VTLB are connected to the storage unit of the VTL6000 through physical channels. The VTLA software application that runs on the physical device of VTLB can still access the original storage space through the physical channel that connects VTLB to the storage unit. Take fiber channels (FCs) as an example to explain standby backup host interfaces. Assume that VTLA provides the VTL service through the FC interface wwpn1. Correspondingly, VTLB provides an FC interface wwpn3 that serves as the backup interface of wwpn1. The interface wwpn3 is a standby backup host interface. The two FC interfaces are connected to the backup server of VTLA through FC channels. Therefore, the VTLA software application that runs on the physical device of VTLB can continue to provide the VTL service to the original backup server through the standby interface wwpn3.

If bidirectional failover is used, VTLA and VTLB monitor the status of each other through a heartbeat network. Once either engine detects a failure in the other engine's software, hardware, or channel connecting the storage unit (which affects the availability of normal VTL service), failover is triggered automatically. The monitoring node then provide the VTL service instead of the failed node. The failover process takes about four minutes. When the failed node recovers from the failure, the failback process is triggered automatically (or manually). The node that took over the resources from the failed node will give control back to the restored node automatically (or manually).

HA Configurations of the SIR Engine

The VTL6000 supports HA configurations on three SIR engines. Two of the three SIR engines are active nodes, and the other one acts as a backup node.

Figure 3-6 HA configurations of SIR engines



As shown in Figure 3-6, the HA configurations of the SIR engines differ slightly from those of the VTL engines. The SIR engines has a dedicated standby SIR engine. The SIR software application on the standby SIR engine can monitor the status of the active SIR engines. When with the standby SIR engine detects a failure in an active SIR engine's (SIR1 or SIR2) software, or hardware or physical channel connecting to the SIR storage space (which affects the availability of the normal deduplication service), failover is triggered automatically. The standby SIR engine then sends an IPMI instruction to power off SIR1, and takes over the work of SIR1. Failover here means that the SIR1 software application runs on the physical device of the standby SIR. When the failed SIR engine is restored and powered on, it serves as a standby SIR engine automatically.

If HA configurations of SIR engines are used, every SIR engine can access not only its own storage unit but also the storage units of other SIR engines through physical channels. In this way, each SIR engine can access its own storage unit no matter on which physical SIR device the SIR software application runs. In addition, all SIR engines are connected to the VTL engines through physical channels. In this way, the SIR engine can read the original backup data from the VTL storage space for deduplication, and the VTL engine can read the data blocks from the SIR storage space for recovery. In general, all SIR engines adopt the same physical connections, thus supporting failover between active and standby nodes.

In addition, the SIR software module also supports high performance. If the VTL6000 adopts the HA configuration of three (2+1) SIR engines, two active SIR engines are used to perform deduplication. The SIR engines form an HP cluster, thus providing higher deduplication performance more efficiently.

The VTL6000 also supports configuring one or two SIR engines. In this case, the HA feature of the SIR engine(s) is unavailable. The VTL6000 configuring two SIR engines provides a high performance (HP) cluster.

3.3 Energy Conservation

According to the statistics of International Energy Agency (IEA) in 2008, the global energy consumption rose by 73% from 1973 to 2006. Energy prices have kept rising since the 1970s, resulting in increasing cost pressure to the industrial and manufacturing sections. The society pays more attention than ever to energy conservation. In a society that values energy conservation and environmental protection, more and more users look for energy-saving features of storage products. More and more institutes and companies take energy conservation as an essential requirement when purchasing storage products. To some extent, the energy-saving feature determine the success of a product.

3.3.1 Energy Conservation Through Deduplication Technologies

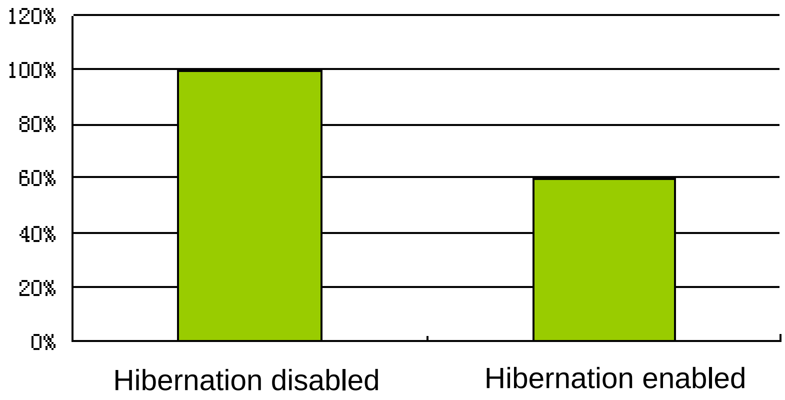
Deduplication technologies save substantial storage investments and management cost for users and reduce energy consumption of storage systems significantly. By means of deduplication technologies, users can store more data on smaller storage. On the basis of the same storage density, smaller storage means fewer storage devices and lower energy consumption. The following example shows how the VTL6000 improves energy conservation by using deduplication technologies.

Assume that a full backup involves 10 TB of data and that the daily incremental data is 50 GB. The user requires daily incremental backup, weekly full backup, and backup data storage for six months. The backup system needs to store 25 full backup data sets and 155 incremental backup data sets on the basis of 30 days per month. To meet the user requirement, a storage device that does not support deduplication needs to provide a storage capacity of 362.75 TB (10 x 25 + (25 x 24 x 0.05 x 7)/2 + 155 x 0.05). The VTL6000 with deduplication function needs to provide a storage capacity of 40 TB only (20 TB for the VTL storage space and 20 TB for the SIR storage space. The SIR storage space available for data is 18.95 TB (10 + 179 x 0.05)). To store the data in 1-TB SATA disks, a storage device that does not support deduplication requires about 400 disks, while the VTL6000 requires about 50 disks only (including the disks for RAID5 and hot backup). The ratio between the disks required by the two systems is about 8:1. In addition, the storage devices that do not support deduplication requires more enclosures than the VTL6000 does. In this case, the power consumed by the storage units of the VTL6000 is only one tenth of that consumed by the storage device without the deduplication function. Considering that the power consumed by the storage units of the VTL6000 accounts for 40% of the whole-system consumption of the VTL6000, it is safe to say that the VTL6000 saves about 36% of the whole-system consumption by using deduplication technologies.

3.3.2 Energy Conservation Through Hibernation Technologies

In addition to deduplication, hibernation can also reduce the energy consumption of storage systems substantially. If a disk receives no access request for a long time, hibernation is invoked to set the disk into a hibernated or power-off state, thus saving energy and prolonging the service life of the disk. In a storage device that supports hibernation, all idle disks are in the hibernated state, and only the read or written disks are in the normal working state. Once a hibernated disk is written or read, the disk enters the normal working state. A disk that is in the normal working state enters the hibernated state of different levels according to the relevant conditions after the read-write operation is completed. Therefore, the hibernation technology is applicable to mass near-line storage devices and hierarchical storage devices that are not accessed frequently and require low instant data availability. The hibernation technology is applicable especially to the data backup and archive devices used for data recovery (for example, VTL products). All disks in the VTL that supports hibernation remain in the working state for several hours of the backup window everyday in extreme circumstances, and are in the hibernated state for the rest of the day.

Figure 3-7 Energy-saving performance of hibernation of VTL6000 storage units



The VTL6000 supports hibernation. The disks in the VTL6000 storage unit are divided into several Redundant Array of Independent Disks (RAID) groups. The user can configure different hibernation policies for different RAID groups (for example, invoking hibernation at a preset time or staying hibernated for a certain period of time). After the hibernation policies are configured for the RAID groups, the idle disks in the groups may enter the hibernated or power-off state according to the hibernation policies. Once the backup software (such as NetBackup) sends read-write requests to the LUNs in the hibernated RAID groups, the hibernated disks in the groups are powered on automatically and are ready for read-write operations. The hibernation technology can save 40% energy from the storage units of the VTL6000. Considering that the energy consumed by the storage units of VTL6000 accounts for 40% of the whole-system consumption of VTL6000, it is safe to say that VTL6000 can save about 16% of the whole-system consumption by means of hibernation technology.

4 Promotion and Experience

The VTL6000 supports VTL and SIR engine clusters, deduplication, and hibernation. It provides excellent backup performance with a backup speed of 1400 MB/s and a raw storage capacity of nearly 200 TB. The VTL6000 can satisfy the demands of medium- and high-end users for HP, high-capacity, energy-saving, and HA VTL products.

This chapter describes two typical application scenarios (the VTL backup system and the remote backup system) and analyzes the solutions and customer benefits of THE VTL6000. For information on other application scenarios (such as hierarchical backup and remote disaster recovery), solutions, and customer benefits, see the *Technical White Paper for the VTL3500*.

4.1 Typical Scenarios and Customer Benefits

4.1.1 VTL Backup System

Application Scenarios

The VTL backup system is applicable to the following scenarios:

- Setting up a new backup system.

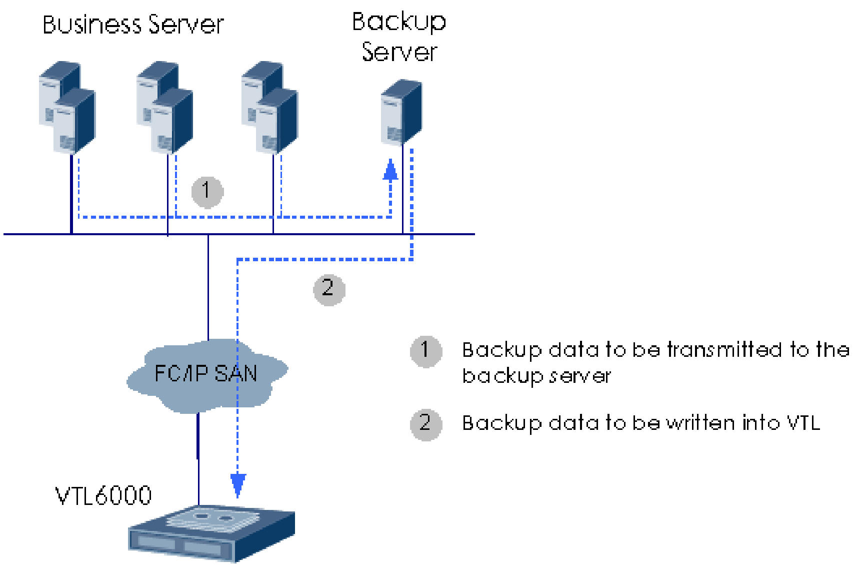
  A customer needs to purchase backup devices because no existing backup system is available or the existing backup system needs to be reconstructed.
- Replacing the existing physical tape library with the VTL6000.

  A customer needs to replace the existing backup system based on physical tape library by a new backup device to improve the performance, reliability, and management.

Solution

Figure 4-1 shows the networking of the VTL6000.

Figure 4-1 VTL backup system



The VTL600 is connected to the backup server through an FC SAN or an IP SAN.

Customer Benefits

The VTL backup system can provide customers with the following benefits:

- High performance that meets customer requirements for backup windows

  If only one VTL engine is configured, the tested backup rate of the VTL6000 reaches 1250 MB/s and and 36TB of data can be backed up in eight hours. If two VTL engines are configured, the tested backup rate of the VTL6000 reaches 2500MBMB/s and 72TB of data can be backed up in eight hours. For customers that need to back up less than 72TB data at a time, the VTL6000 can meet the customers' requirements for backup windows perfectly.

- High capacity that meets customerrequirements for storage capacity

  If two VTL engines are configured, the VTL6000 supports a raw capacity up to 384 TB (320TB available for data storage), and a capacity up to 152TB for deduplicated data if the deduplication function is enabled. The available capacity of 320 TB consists of the available capacity of the VTL storage space (160 TB), the SIR index disk (8TB), and the SIR data disk (152 TB). If the deduplication ratio is 20:1, the VTL6000 can store 3.0 PB of backup data, thus satisfying the data backup requirements of mid-rang and high-end users.

- Deduplication that saves energy consumption and storage investments

  The deduplication and hibernation functions supported by the VTL6000 help to reduce the storage disks needed for the backup system significantly and save energy consumption and storage investment greatly. The deduplication ratio and energy-saving performance are related closely to the type of backup data and the backup policies. As described in section 3.3 "Energy Conservation," the VTL6000 saves over 50% energy from the backup devices for customers, compared with the VTL products that do not support deduplication or hibernation.

- HA cluster configuration that meets customer requirements for high availability

  If two HA VTL engines are configured, the two VTL engines work independently. If either VTL engine fails, the other VTL engine takes over the work of the failed engine automatically. The failover process takes about four minutes, and then the

VTL6000 continues to work. In this way, the VTL6000 meets customers' requirements for high availability of backup systems. If three HA SIR engines (2+1) are configured, the standby SIR engine automatically takes over the work of the failed engine when one SIR engine fails to support the deduplication process. The failover process takes about four minutes, and then the VTL6000 continues to work. In this way, the VTL6000 satisfies customers' requirements for high availability of backup systems.

4.1.2 Remote Backup System

Application Scenario

The remote backup system is applicable to the following scenario:

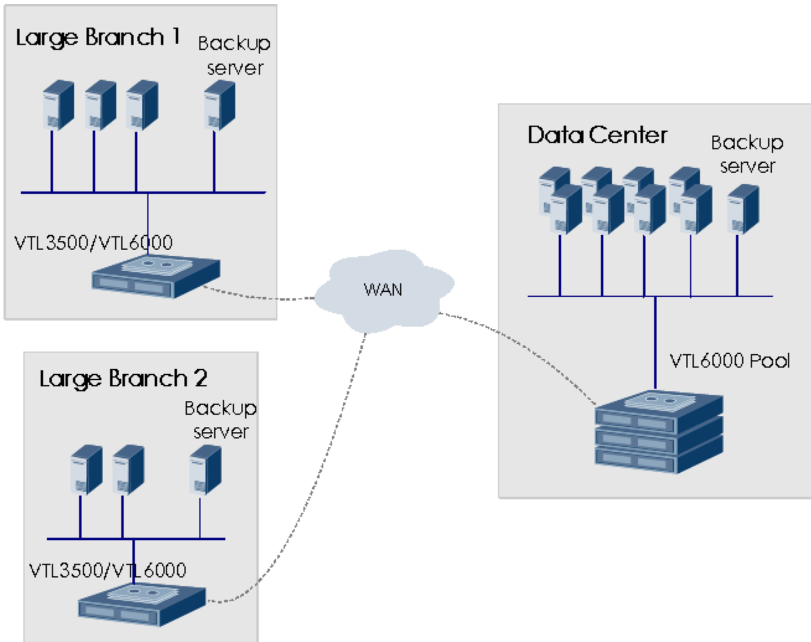
- Insufficient bandwidth resources in customers' networks.

  In addition to the data center, multiple remote branch nodes in the customer's network require data backup. The customer has no backup system in its network, or the available bandwidth resources are insufficient between the branch nodes and the data center in the existing backup system. The data backup performance of the branch nodes are too poor to satisfy the customer's backup window requirements.

Solution

Figure 4-2 shows the networking of a remote backup system.

Figure 4-2 Remote backup system



- Deploy the VTL3500 with deduplication function (see the product description of the VTl3500) or the VTL6000 All In One (see he product description of the VTL6000) for the branch nodes. Deploy the VTL6000 Multi-Nodes with deduplication function (see the production description of VTL6000) for the data center.

- Data of the branch nodes is backed up to the local VTL3500 or VTL6000, and then replicated to the VTL6000 at the data center through remote replication based on deduplication. Data of the data center is directly backed up to the local VTL6000.

Customer Benefits

The remote backup system can provide customers with the following benefits:

- Local VTL backup that satisfies customers' backup window requirements for branch nodes

  Data of the branch nodes is directly backed up to the local VTL3500 or VTL6000. For more information on the customer benefits of the VTL backup system, see section 4.1 The VTL supports high performance (The backup rate of the VTL3500 is 400 MB/s and that of the VTL6000 All in One is 1250 MB/s). Therefore, it takes less than 5 hours to back up as much as 10 TB data for branch nodes, thus meeting the customers' backup window requirements for branch nodes.

- Deduplication-based remote replication that greatly reduces the required network bandwidth and saves investments in network bandwidth

  The local VTL first deduplicates the backup data of the branch nodes, and then replicates the deduplicated data to the VTL at the data center through the WAN. Only the data blocks that are different from those in the data center are replicated. This greatly saves replication bandwidth, reduces the required bandwidth, and saves investment in network bandwidth for customers.

- Global deduplication that further reduces storage investments for customers

  The VTL6000 support global deduplication. After deduplication, only the data blocks that are different from those in the data center are replicated. The global deduplication eliminates the duplicate data among the VTLs of different branch nodes and that between the VTL of each branch node and the VLT of the data center. Compared with the deduplication function of the VTL backup system described in section 4.1.1 , the global deduplication achieves higher efficiency and further reduces storage investments for customers.

5 Conclusion

With deduplication, hibernation, and HA cluster configuration of VTL and SIR engines, the VTL6000 is an ideal solution for medium- and high-end users to solve such problems as the poor performance, insufficient capacity, high energy consumption, and low availability of backup systems.

- For customers that need to back up less than 72 TB data at a time, the VTL6000 meets their requirements for backup windows perfectly.
- If the deduplication ratio is 20:1, the VTL6000 can store3.0PB of backup data, thus satisfying the data backup requirements of mid-rang and high-end users.
- By means of deduplication and hibernation, the VTL6000 hepls to save over 50% energy cost for customers.
- The VTL6000 supports HA cluster configurations of VTL and SIR engines. The failover takes about four minutes. The HA configurations satisfy customers' requirements for high availability of backup systems.

6 Acronyms and Abbreviations

Acronym/Abbreviation	Expansion
VTL	virtual tape library
SATA	serial advanced technology attachment
HA	high availability
SIR	single instance repository
FC	fibre channel
IPMI	intelligent power management interface
LUN	logical unit number
RAID	redundant array of independent disks