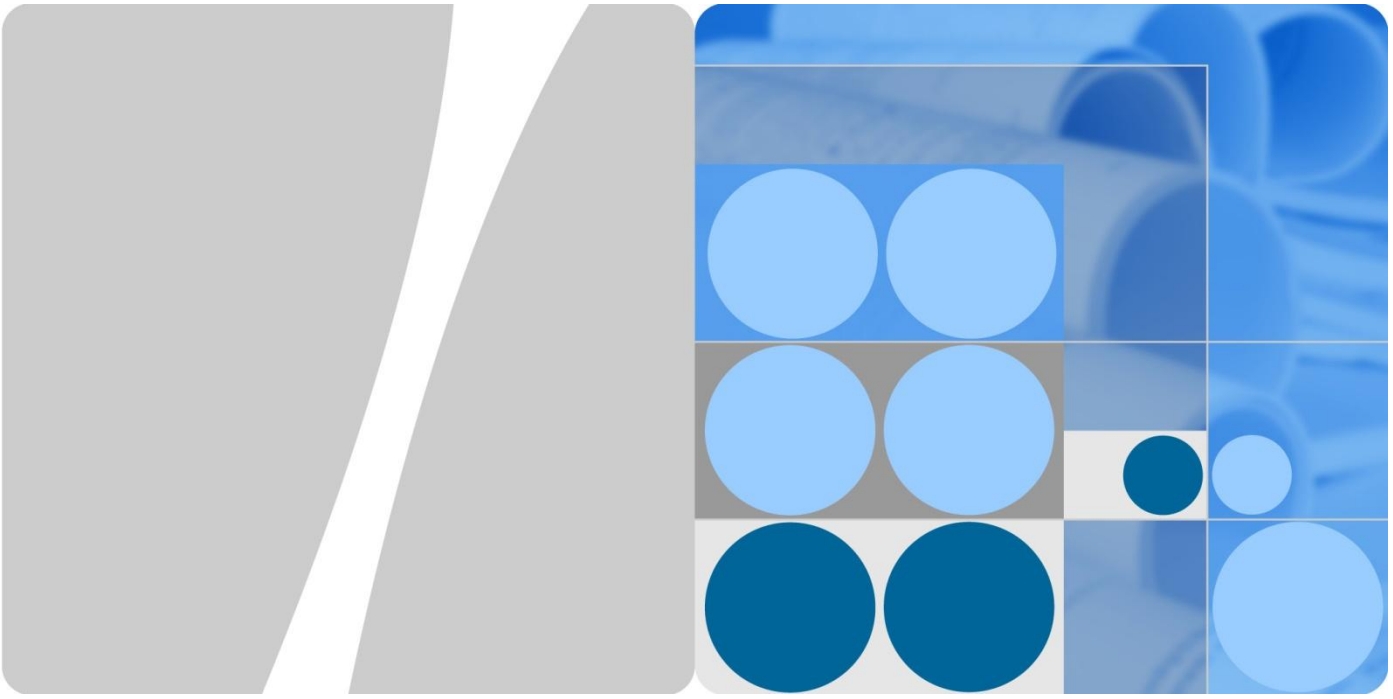

Huawei OceanStor S2200T V100R005 Technical White Paper

Issue V1.1

Date 2014-08

HUAWEI TECHNOLOGIES CO., LTD.

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

1 Executive Summary

This document begins by providing an insight into the development trend of storage technologies. Then based on user requirements, this document describes in detail the functions and highlights of Huawei OceanStor S2200T storage products in four aspects including excellent performance, proven reliability, massive scalability, and energy saving. This document introduces the unique features of S2200T series and its values for users.

2 Introduction

With the development of the information technology (IT), IT systems maintain an ever-increasing amount of data, and their architectures become ever more complicated due to the heterogeneous networks, distributed departments, and large-scale complicated service systems. Device management and maintenance are facing enormous challenges.

Authoritative surveys indicate that nowadays IT O&M is increasingly complicated. The O&M cost accounts for 72% of the total IT investment budget, and only 28% of the total investment is used to add new facilities.

More problems are naturally brought out, such as, complicated network environment, performance bottlenecks on a single storage system, difficulty in achieving a balance between costs and performance, and intense requirements for mixed storage. How to resolve these problems becomes the top priority in the IT industry.

The OceanStor S2200T is designed particularly to provide high reliability, high availability (HA), high scalability, and easy management. It fully meets customer requirements, and helps them get the most out of their initial investments.

3 Solution

With the development of the IT industry, storage gradually grows to be an independent category. Technology progresses bring the development of some industries, such as storage.

Technology development brings new concepts and standards that change the development trends of industries. According to experts, the storage industry is developing towards the following aspects:

- **A new level of storage performance due to the new standards of the PCI-E bus**

  The data transfer rate of the mainstream PCI-E 2.0 doubles that of the PCI-E from 2.5 Gbit/s to 5.0 Gbit/s. The new PCI-E 2.0 standard expands the bus bandwidth for storage systems.

- **Widespread application of Serial Attached SCSI (SAS) technologies**

  SAS adopts the serial and interconnected architecture, allowing enterprise users to flexibly expand and manage their storage systems. The evolution of the SAS storage interface from the 3 Gbit/s standard to the 6 Gbit/s standard enables the SAS storage interface to provide faster data transfer rate and higher reliability.

- **Solid State Disks (SSDs) bring new features for storage.**

  As the cost of the integrated circuit decreases, the SSD draws more attentions in the storage industry. The SSD has its own advantages:

  A. The data transfer rate is fast because the access speed of the NAND flash is 30 to 40 times as that of the traditional mechanical disk.

  B. The SSD has good shockproof capability because the flash chip is used to replace mechanical parts. In this way, the SSD can normally work even under fast movement.

  C. Power is saved because the IOPS power consumption of the SSD is much lower than that of the mechanical disk.

  In one word, as more and more SSDs are added into the storage system, the advantages become obvious.

- **Diversified interface types improve storage scalability.**

The advent of 8 Gbit/s FC and 10 Gbit/s ethernet brings diversified storage interface cards. The enterprise-level 2.5 inch disk brings chances for disk update in the storage industry.

- **Broadened Internet by Internet Protocol Version 6 (IPv6)**

  Internet Protocol Version 4 (IPv4) is the first widely used Internet protocol that laid the foundation for the current Internet technologies. The Internet flourished in the past decade, and there has been a great demand for IP addresses. As a result, IPv4 addresses are running out at present. This severely restrains the application and development of the Internet.

  IPv6 is the next-generation Internet protocol that is developed by Internet Engineering Task Force (IETF) to replace the current Internet protocol (IPv4). The length of an IPv6 address is 128 bits, increasing the number of IP addresses by 2^{96} times. It is said that if IPv6 is widely used, every sand in the world can have its IP address.

- **Green storage**

  Green storage aims to improve storage utilization ratio, reduce construction and maintenance costs, and raise the energy efficiency of all network storage devices. The global financial crisis since 2009 and the current tense situation of energy resources force most enterprises to consider the possibility of constructing a green data platform.

Based on the development trend of the storage industry and a comprehensive understanding of user needs, Huawei launches its OceanStor™ S2200T series unified storage products. The S2200T is a next-generation enterprise entry-level storage arrays oriented to the low-end markets. It has powerful hardware specifications and incorporates various advanced technologies, including the full-redundancy design, hot-swappable I/O modules, and diversified data protection technologies. It costs only the price of entry-level products and meets the ever-increasing and complicated requirements. Its target applications include databases, digital media, Internet operation, centralized storage, backup, disaster recovery, and data migration. These outshone features effectively protect customers' service security and continuity. Figure 3-1 and Figure 3-2 shows the appearance of the S2200T.

Figure 3-1 Appearance of the S2200T 2 U 12-disk controller enclosure

Figure 3-2 Appearance of the S2200T 2 U 24-disk controller enclosure

In addition, the S2200T has unique features on high performance, high reliability, powerful scalability, and energy savings.

High Performance

- **New-generation high-speed bus and high-speed I/O channels**
  The S2200T adopts the new-generation PCI-E 2.0 bus technology and the SAS 2.0 high-speed I/O channel technology, and provides a maximum of 32 Gbit/s bidirectional internal exchange bandwidth, which helps to improve the overall performance of the storage system.

- **Multi-core CPU groups and various memory capacities**
  - Multi-core CPU group: enhances the processing capability of the storage system.
  - Various memory capacities: The S2200T has two memory capacities (4 GB/8GB for dual controllers), and they can be flexibly selected as required.

- **SmartCache （Only supported by product of 8GB cache）**

  The concept of the SmartCache technology is to use one or more SSDs to form a SmartCache resource pool. The hotspot data blocks that the host frequently accesses are migrated from the mechanical disk to the SmartCache resource pool according to the real-time statistics on the access frequency of data blocks. The SmartCache technology makes use of the fast processing capability of SSDs to improve the read performance and access efficiency of the host. 0 shows the working mechanism of the SmartCache technology.

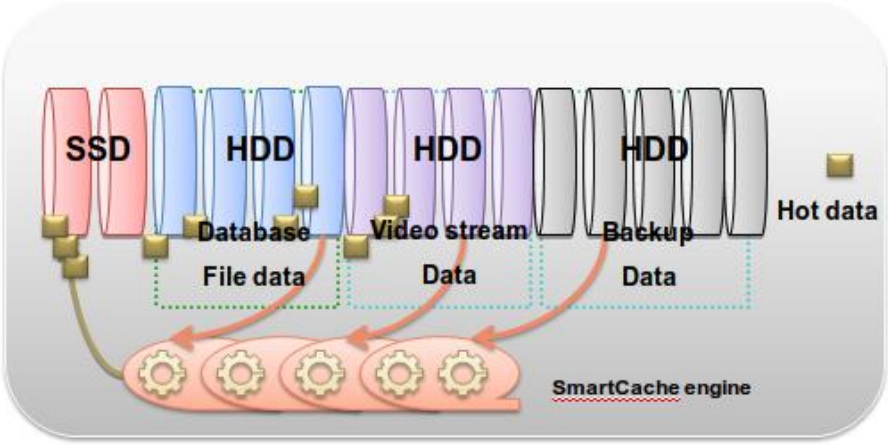

Figure 3-3 SmartCache working mechanism

The SmartCache technology features high performance, high availability, and powerful scalability.

- High performance: The test result indicates that the SmartCache technology improves the random read performance by several times for the applications of Web server and file server.
- High availability: The SmartCache resource pool consists of multiple SSDs; therefore, the failure of one SSD does not affect the functions of the SmartCache.
- Powerful scalability: The SmartCache supports adding SSDs online. In addition, the installation and uninstallation of the SmartCache do not affect the running of services.

- **Optimized application awareness**

  The S2200T provides application-based awareness, which greatly improves the overall performance of the system in terms of virtual machine and database application.

  - SQL Server: Based on the optimized application awareness function, the S2200T triples its speed in managing and dispatching virtualized SQL databases.
  - Oracle: The S2200T provides system reliability by optimizing database migration and upgrade, and reduces system downtime by 90% as it shortens the time needed for development, testing, and maintenance by using the snapshot and full LUN copy functions.

- **Application-oriented cache intelligent prefetch**

  The S2200T provides the application-oriented cache intelligent prefetch technology that can automatically identify how sequential the current I/Os are, enable or disable the cache prefetch function depending on the service pattern, and automatically set a optimal prefetch length for a specific application scenario. The cache prefetch technology of the S2200T series greatly improves read performance and reduces the access frequency of hard disks, and therefore extends the service life of hard disks.

- **VAAI technology: acceleration of virtual machine performance**

  Because of development of the server virtualization technology, more customers use virtual machines to replace traditional servers when they deploy their service systems. Virtual machines dramatically slash customers' investments and set up a sound basis for subsequent service system consolidation, service data migration, and data backup. In the virtualization server family, the VMware stands out. According to the relevant statistics, the VMware takes up more than 60% of the market share till now and keeps the increasing tendency.

  Compared with traditional servers, virtual machines also have a disadvantage, that is, most I/O operations on a virtual machine are processed by software. Therefore, a large number of resources such as CPUs, memory, and network bandwidth are required to process I/O operations. This disadvantage is obvious with the wide application of virtual machines. To solve the problem, the hardware acceleration concept is put forward. A dedicated VMware vStorage APIs for Array Integration (VAAI) plug-in interacts with a disk array. The disk array authenticated on compatibility takes over the I/O operations on virtual machines, improving the overall performance.

  Currently, the S2200T storage system has passed the certification on compatibility with the VMware and passed the test of interaction with the VMware VAAI

plug-in. The performance is satisfactory. Interacted disk array and VMware can implement the following functions:

- Zeroing function: The most common operation on a virtual machine is to clear virtual disks. This job occupies large part of DMA buffer and CPU period, and consumes shared resources of the virtual machine. When a S2200T series disk array interacts with the VMware, data block zeroing operations on the virtual machine are transferred to the disk array. Based on the powerful CPU performance of the disk array, clearing operations are quickly completed. This function reduces the I/O operations between the disk array and the ESX server by more than 10 times and accelerates the virtual disk initialization, improving the overall performance of the service system.
- Replication function: When the data on virtual disks is migrated or cloned, the virtual machine copies a large amount of data blocks. It takes a few hours to clone xx GB files. That consumes many host resources and occupies the network bandwidth for a long time, deteriorating the overall system performance. When a S2200T series disk array interacts with the VMware, the data block replication operation on the virtual machine is transferred to the disk array. The array hardware controls the replication operation within seconds and reduces the CPU load of the ESX server. In this way, the host software can make full use of virtual machine resources. This function reduces the I/O operations between the disk array and the ESX server by more than 10 times, improves the Storage vMotion operation speed, and simplify the deployment of virtual machines.
- Assisted locking function: In a cluster, to ensure the data consistency on virtual machines, the locking mechanism is employed to control resource occupation upon concurrent access. Traditionally, when an ESX server accesses a LUN, the system locks the LUN. Therefore, write I/O operations on other ESX servers are suspended and the overall write performance deteriorates a lot. In addition, a series of commands need to be executed to obtain and release the lock. The entire process prolongs I/O latency. When a S2200T series disk array interacts with the VMware, virtual machines access data blocks rather than LUNs. That improves concurrent write performance, shortens write I/O latency, enables more virtual machines to access a single LUN, and further improves the overall performance of the service system.

- Dynamic load balance between dual controllers

  The two controllers of the S2200T work in active-active mode. They concurrently process I/O requests sent by application servers, balancing the load between them. This avoids the situation where one controller is overloaded whereas the other is idle. Therefore, the active-active mode reduces the load of single controllers and helps the system to efficiently use system resources, improving system efficiency and performance.

- **Link data compression**

  The link data compression technology enables the S2200T array in the primary site (primary array) and the S2200T in the disaster recovery site (DR array) to check for each other's compression/decompression capability before a remote replication or remote LUN copy. If both of the two S2200T arrays obtain the compression/decompression capability, the primary array compresses data in its compression module, and then the iSCSI initiator transfers the compressed data to the DR array. The DR array receives the data and decompresses it in its

decompression module, and then stores the decompressed data onto a specific location.

In this way, the link data compression technology improves remote data transfer speed and capability. But note that this technology is only applicable to remote replication and remote LUN copy based on IP replication links.

High reliability

- **BBU and data coffer**

  To protect the system against external power supply failures, the S2200T storage system is equipped with a backup battery unit (BBU) and a data coffer. The BBU features small size, low cost, redundancy, and hot swapping. In case of an external power failure, the BBUs supply power to the controllers and data coffer simultaneously so that the cached data is securely flushed to hard disks. This ensures data integrity and reliability.

- **Disk precopy**

  Storage systems can use RAID mechanism to improve data availability. The reliability of RAID mechanism is based on that of hard disks. After a hard disk has been operating for a long time, there is a growing disk failure possibility. In addition, if a storage system uses the hard disks that are purchased in a batch, the failure of any hard disk means a potential failure of all hard disks. RAID mechanism allows only a limited number, not all of hard disks to fail simultaneously. If potential disk failures are not detected and handled in a timely manner, service systems may break down unexpectedly. Even if RAID groups can be reconstructed, the reconstruction takes a time and inevitably affects overall system performance.

  To prevent or reduce the impact of hard disk failures on the storage system, the S2200T adopts the disk pre-copy technology. The disk pre-copy technology enables the system to obtain first-hand information about disk status by using the disk prediction technology. The information is usually obtained by reading the self-monitoring analysis and reporting technology (S.M.A.R.T) properties. The precopy algorithm determines the operating status of hard disks and computes their latent failure rates. If a hard disk is considered under a high risk, its data is replicated to a hot spare disk. The replication process is preferentially performed in a system idle time to avoid interrupting ongoing services. Disk precopy eliminates or minimizes RAID reconstruction duration, and lowers the ratio of another disk failure during RAID reconstruction. Disk precopy enhances the security of Huawei storage systems and ensures the service continuity.

- **Bad sector repair**

  Authoritative surveys show that hard disks account mostly for storage system failures. The S2200T detects bad sectors by determining I/O responses. Once a bad sector is detected, the data on that sector is recovered through a RAID parity calculation, and is returned to the application server. Then the data is redirected and written to the free space. All read I/Os destined at the bad sector are redirected to the new location of the data. Bad sector repair lowers the disk failure rate, prolonging the service life of hard disks and

- **Multiple data protection mechanisms**

The S2200T provides the most comprehensive advanced data protection capability in the industry to meet the needs of backup, disaster recovery, and data migration in various service application.

- HyperMirror: maintains data duplicates at two or more sites to prevent data loss caused by disasters.
  The S2200T supports the remote replication for both file systems and block data. Users can select their remote disaster recovery method based on their actual requirements. The remote replication of the S2200T includes two mainstream replication technologies: synchronous remote replication (HyperMirror/S) and asynchronous remote replication (HyperMirror/A).
- Snapshot: The snapshot of the S2200T series (HyperImage) supports both file-system-level and block-level snapshots. It can generate a point-in-time consistent mirror for a source file system or source LUN, offering a rapid consistent data duplicate of the source file system or source LUN without interrupting services. The duplicate is available immediately. Data reads and writes on the duplicate have no impact on the source data. Therefore, the snapshot function greatly helps a user with online backup, data analysis, and application testing that were difficult to implement.
- LUN copy: The LUN copy of the S2200T (HyperCopy) includes full LUN copy and incremental LUN copy, supports FC and IP links, internal and inter-array data copying, and LUN copy between heterogeneous storage arrays. It meets various data backup requirements and can also be used in data migration services.

Flexible scalability

- **Flexible upgrade**

  The S2200T storage system supports the performance and capacity of the S2200T can expand linearly and easily.

- **TurboModule technology**

  The TurboModule technology includes hot swappable modules, flexible combination of host and expansion I/O modules, and high-density I/O modules and interfaces.

  - **Hot swapping of modules**: The S2200T supports full-redundancy hardware design. The redundant parts, including controllers, power supplies, fans, and hard disks, are all hot-swappable. Figure3-4 shows the hot-swappable modules of the S2200T.

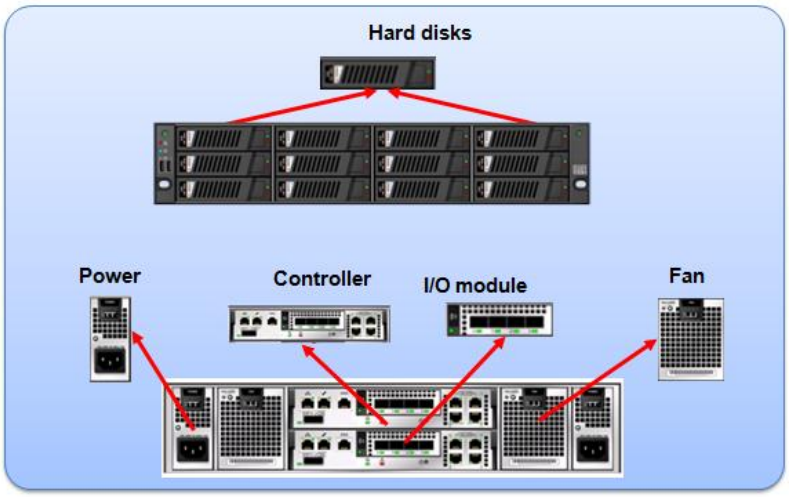

Figure 3-4 Hot-swappable modules of the S2200T

The hot swapping of I/O modules is the most extraordinary design of the S2200T in aspect of high scalability. Owing to this design, the S2200T can online add its I/O modules as service data explodes. This saves the cost and labor of adding switches. Furthermore, once an I/O module is faulty, it can be online replaced without service interruption, and the system reliability and service continuity are greatly protected.

- Flexible combination of host and expansion I/O modules: The S2200T supports 8 Gbit/s FC, 1 GE and 10 Gbps TOE ports (Only supported by product of 8GB cache); therefore, users can decide on the combination of host and expansion I/O modules depending on the service type.
- High-density design of I/O modules and ports: The S2200T supports both onboard I/O modules and online hot-swappable I/O modules, providing a maximum of 20 host ports and 2 expansion ports. This high-density design and flexible scalability greatly save the purchase and maintenance costs.

- **Multiple disk types**

  The S2200T supports 3.5-inch and 2.5-inch disks. SSD, SAS, NL SAS and SATA disks for on-demand configuration.

- **Two specifications of the disk enclosure**

  The S2200T supports 2U and 4U disk enclosures. Both types of disk enclosure adopt the design of high density slots (24 slots/enclosure). In average, 1U space houses up to 12 2.5-inch hard disks. Figure3-5 shows the disk enclosure of the S2200T.

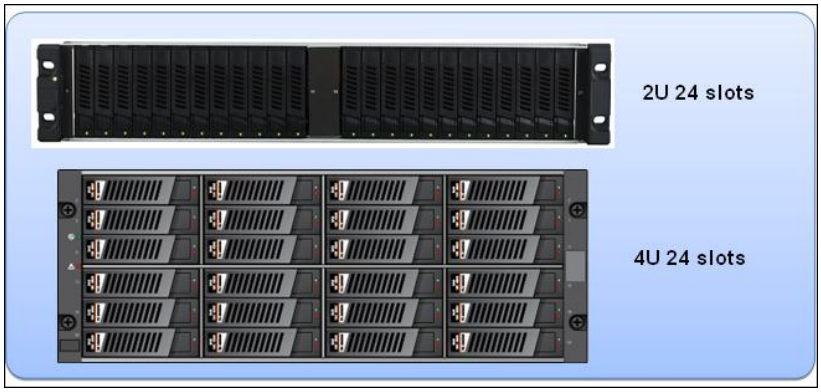

Figure3-5 Disk enclosure of the S2200T

- **HyperThin technology**

  The S2200T storage system provides a thin provisioning function, HyperThin (Only supported by product of 8GB cache). This function enables users to flexibly plan host capacity and assign virtual storage space according to specific service requirements. HyperThin also implements smooth capacity expansion without service interruption to meet the increase of data amount. The expansion process is transparent to upper-layer applications.

  The HyperThin technology reduces investment and O&M costs for customers and simplifies service system management.

- **Dynamic RAID group expansion**

  The S2200T supports the dynamic RAID group expansion technology. With this technology, users can add hard disks to a RAID group online. The expansion process is transparent to upper-layer applications. After expansion, the RAID group settings such as stripe depth and LUN configuration remain unchanged. In addition, RAID group expansion incorporates the disk defragmentation function, which frees disk space to improve disk utilization, keeps hard disks running in optimal read and write status, and prolongs their service life. Dynamic RAID group expansion applies to the scenarios that allow no downtime and require flexible capacity expansion.

- **Supporting the next-generation Internet protocol — IPv6**

  IPv6 is the next-generation Internet protocol, the advent of which follows the exhaust of limited address space defined by IPv4 due to the rapid development of the Internet. Insufficient address space will inevitably hinder the further development of the Internet. To expand the address space, IPv6 is supposed to define new address space. The length of an IPv6 address is 128 bits, increasing the number of IP addresses by 2^{96} times. IPv6 can provide almost unlimited IP addresses to users.

  The S2200T already supports IPv6. Users' service system can be upgraded from IPv4 to IPv6 smoothly without the need to replace storage arrays, and no compatibility issue is involved in the upgrade process, fully protecting users' initial investment.

Low Power Consumption

- **Energy saving**

The S2200T series green design consists of several energy-saving measures, including, intelligent CPU frequency control, delicate fan speed control, and intelligent disk spin-down.

- **Intelligent CPU frequency control**: The intelligent CPU frequency control technology of the S2200T can dynamically adjust the CPU frequency according to CPU utilization to reduce the overall system power consumption. This technology increases CPU frequency once the CPU utilization is above 80%, and smoothly reduces CPU frequency once the CPU utilization is below 20%, with a certain latency to maintain the stability of system performance. Meanwhile, this technology ensures sufficient voltage for the CPU during the intelligent frequency adjustment process. Therefore, the adjustment process has no impact on the stability of the service system.
- **Delicate fan speed control**: A temperature sensor is deployed at each heat-sensitive component of the S2200T series to monitor its temperature and smoothly adjust fan speed accordingly. The S2200T is designed to provide 16-increment fan speed control. Each increment is 2℃. This design effectively improves heat dissipation, reduces the noise that the device produces and the dust that enters the device. Therefore, this design prolongs the service life of the fans while reducing the power consumption of the overall storage system.
- Intelligent disk spin-down technology: Of all storage components, hard disks consume most of the power. The hard disks that run continuously for a long time consume much power. The intelligent disk spin-down technology of the S2200T spins down idle hard disks in some near-line or off-line applications depending on the service load. In this way, unnecessary power consumption is reduced. While this technology does not alter the working mode of hard disks frequently, it shortens the time of rotating and prolongs the service life of hard disks.

Huawei also provides the integrated storage management software OceanStor™ ISM (ISM) that centrally manages the OceanStor series storage devices. As a user-friendly Graphical User Interface (GUI) –based management tool, ISM helps customers with wizard service configuration, one-click upgrade, and alarm reporting.

In addition to its user-friendly OAM GUI, the ISM also features quick deployment, easy management, standard interfaces, and secure operation.

- **User-friendly management interface**
  - **User-friendly GUI:** The ISM provides a Windows-like GUI to suit the needs of the majority of users. Figure 3-6 shows the ISM GUI.

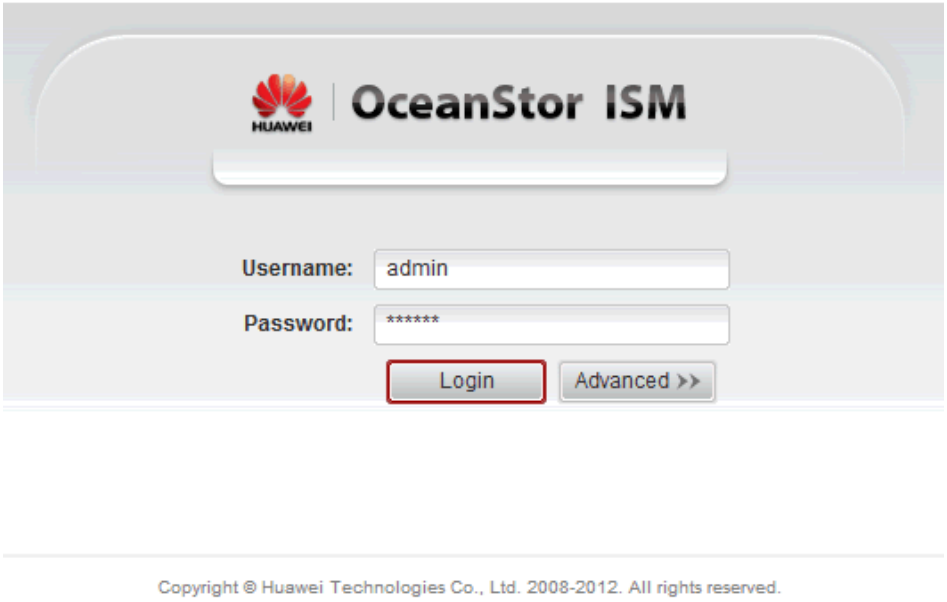

Figure 3-6 ISM GUI

- Diverse alarm management mechanisms: The ISM enables the real-time viewing and statistics collection of events and alarms. The alarms can be sent by indictor, SMS, and email.

- **Quick deployment**
  - **Wizard-based service configuration:** The ISM provides various wizards to facilitate service configuration. These wizards guide users through all deployment processes and provide typical configurations as default options help users initialize their storage devices efficiently.
  - Batched deployment: The ISM provides the batched discovery and configurations of devices, reducing repeated configurations and service configuration duration.
  - Configuration synchronization: The ISM provides configuration synchronization among different devices, for example, alarm SMS and email configurations. This improves the configuration efficiency of multiple devices.
  - One-click upgrade: One-click upgrade is supported on the ISM. You can upgrade all components and display the upgrade progress and upgrade component details. This simplifies the operation difficulties and improves the maintenance efficiency.

- **Easy management**
  - **Intuitive views**: The ISM provides an intuitive and real view of the device to show the current running status of the device. The device manager on the ISM provides various management functions to show the progress of all tasks being performed on the device.

- Information summary: The ISM can collect statistics on hardware status, storage resources, LUN mappings, and major alarms. Statistics export, user-defined statistical items, and capacity precaution are supported. The statistics intuitively display the device status and help decision making.

- **Standard interfaces**
  - **Unified management:** The ISM manages Huawei storage devices (such as SAN devices and NAS devices) and virtual gateways on the same GUI. Consistent operation interface and style are easy for users to learn and use.
  - Report to upper-layer NMS: The ISM can interwork with mainstream NMS, for example, Huawei I2000 and Symantec CCS.

- **Operation security**
  - Hierarchical management of operations: The ISM defines four operation levels based on the impacts of all configuration operations. The impact-based anti-misoperation design for these levels minimizes the possibility of misoperations caused by human errors. Figure 3-7 shows the ISM hazard levels.

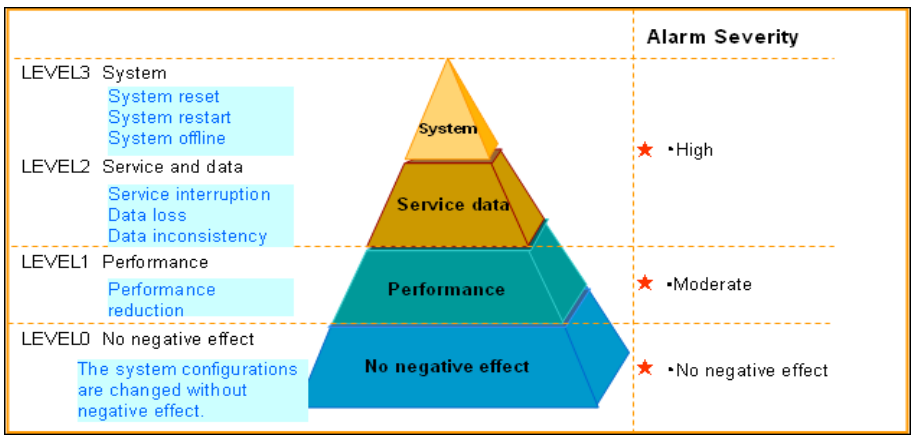

Figure 3-7 Definitions of ISM hazard levels

4 Conclution

Huawei is dedicated to providing high quality storage products and user-friendly services for customers. Guiding by this concept, the S2200T fully meets customers' requirements, and improves customers' values to the maximum in terms of functions, performance, and energy savings.

5 Acronyms and Abbreviations

Table 6-1 Acronyms and abbreviations of the S2200T

Acronyms and Abbreviations	Full Spelling
FC	Fibre Channel
LUN	Logical Unit Number
RAID	Redundant Arrays of Independent Disks
S.M.A.R.T	Self Monitoring Analysis And Reporting Technology
SAS	Serial Attached SCSI
SATA	Serial Advanced Technology Attachment
SCSI	Small Computer System Interface
BBU	Backup Battery Unit
OLTP	On-Line Transaction Processing
OLAP	On-Line Analytical Processing
MIS	Management Information System
ERP	Enterprise Resource Planning
MDN	Mobile Directory Number

