1. 数据存储系统方案

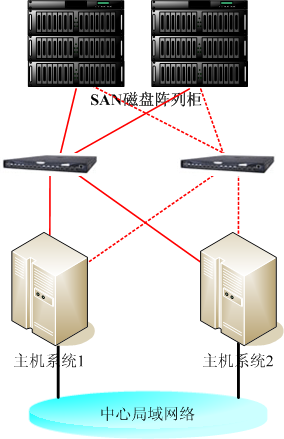
## SAN架构示意图

此次同济大学数据存储系统的方案，不仅仅是对容量的简单增加，更多地是为了全面提高整个系统的可用性，可恢复性，可扩展性，性能等全面技术指标，建立更加完善的存储架构，为今后的长期发展打下基础。

同济大学原有存储设备EMC CX300一台，考虑到CX300不支持镜像技术，以及灾备系统的建设，考虑将现有CX300磁盘阵列柜作为镜像阵列柜，在原有系统的基础上增加以下设备：

* EMC CLARiiON CX500磁盘阵列 2台；
* 主机上的光纤卡（HBA） 若干；
* 光纤连接线 若干。

改造后的存储架构示意图如下图：



## SAN的结构设计

本方案的结构设计依据高可用性，高性能，高可扩展性和高可管理性几方面的要求进行设计。

* **高可用性**

本方案在SAN网络连接方面全部采用双冗余设计，任何一条网络连接出现故障，都可以自动由另一条线路直接接管，从用户端根本感觉不到连接故障带来的影响。甚至于，假如在这个网络中的交换机整个出现故障，导致所有连接的中断，仍可以由另一台交换机接管起所有的访问。在这样的网络设计下，设备连接之间不存在任何单点故障，无论是故障和事故性的硬件失败，还是人工干预的系统维护，都不会中断应用的运行，可用性得到了保证。

针对备份管理系统同样提供高可用性。当光纤链路出现故障时，系统管理员可以简单地将磁带库连接手动转移到另一条链路上，从而达到对磁带库的高可用性。

* **高性能**

方案采用了同济大学已有的高性能的光纤交换机DS-16B3光纤交换机连接主机和存储设备。DS-16B3可以提供每个端口2Gbit/sec的高速度，相对其它竞争产品具有明显的优势。本方案中磁盘阵列子系统与2台交换机分别使用2条光纤连接，最多支持4条，全部线路双工，使对磁盘阵列的总访问速度可达8Gbit/sec。

* **可扩展性**

本方案采用的光纤交换机DS-16B3每台提供16个端口。这样，将来如果需要增加其它设备，可以在无需增加网络设备的前提下就直接接入SAN，易于扩展。

* **可管理性**

本方案的一切网络设备，均严格按照标准的SAN架构进行设计，将需要增加SAN管理工具，均可直接进行实施。

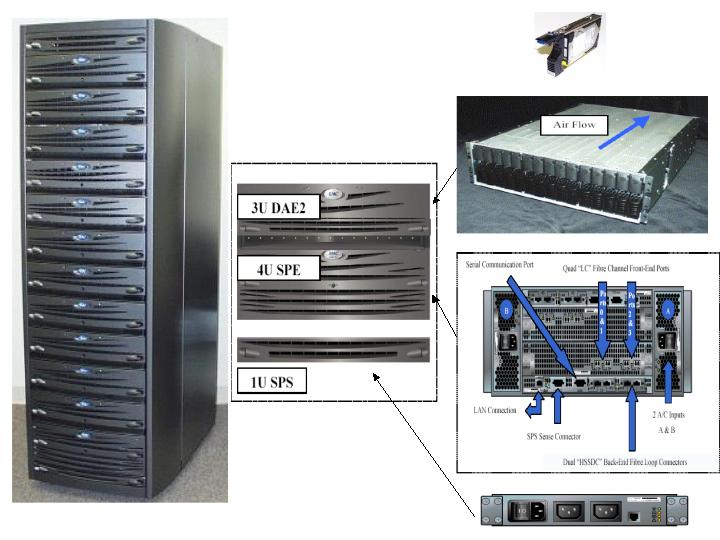
## EMC CX500配置

|  |  |  |  |
| --- | --- | --- | --- |
| **产品号** | **描述** |  | **数量** |
| CX500 | CX500 FC Array W 4GB MEM | CX500主机(主) | 2 |
| RACK-40U-C | 40U COMMON RACK | 标准机柜 | 2 |
| CX-2G10-146 | FC 146GB 10K 520B 2GB | 146G 10000转光纤硬盘+动态备盘 | 30 |
| PW40U-C-IEC3 | 40U-C PWR CORD - IEC309 | 机柜电源 | 2 |
| C-PH-CABLE | SRL CABLE, PHONE HOME SW | 远程连接MODEM连接电缆 | 2 |
| NAVAGT-SUNKIT | NAVI AGENT SUN MEDIA | SUN服务器连接软件介质 | 6 |
| MVS500-KIT | CX500 MIRRORVIEW S MEDIA | 镜像冗灾软件介质 | 2 |
| NAV-DPKIT | NAVI DEPARTMENTAL MEDIA | CX500管理软件介质 | 2 |
| 40UC-SECURE2 | CLRN 40UC ANTI-MOVE KIT | 远程连接MODEM | 2 |
| CX500-KIT | CX500 DOCS & RTU | 文档 | 2 |
| SYMMOD-CHN | MODEM CHINA | MODEM | 2 |
| UTIL-SUN | SUN CPCI S/W UTILITIES | SUN服务器连接软件 | 6 |
| PP-SUN-KIT | POWERPATH FOR SOLARIS KIT | 动态均衡软件介质 | 2 |
| PP-SUN-WG | PPATH SUN WGR | 动态均衡软件 | 6 |
| MVS500 | CX500 MIRRORVIEW SYNC | 镜像冗灾软件 | 2 |
| NAV500-DP | CX500 NAVI DEPARTMENTAL | CX500管理软件 | 2 |

## EMC CX500简介

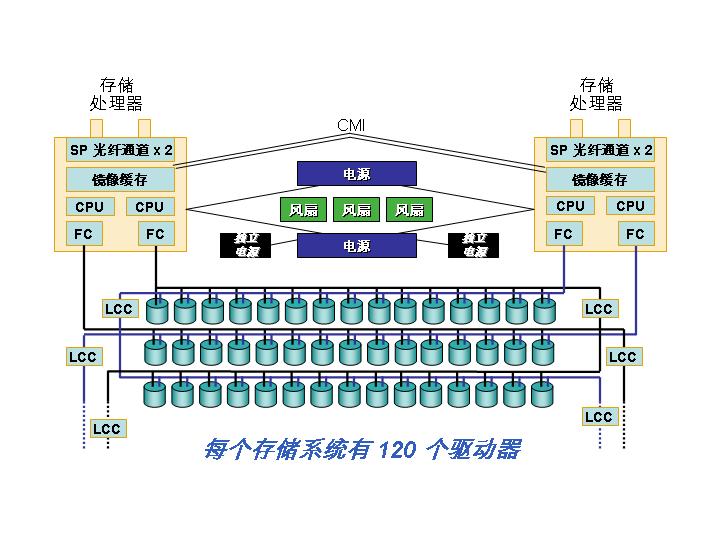
### EMCCX500产品介绍

1. **总体概览**

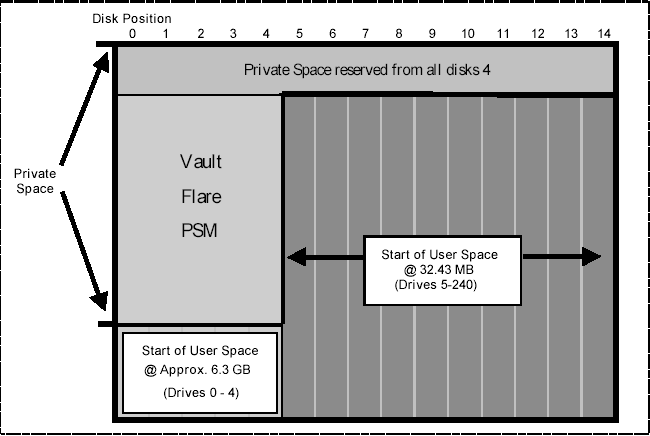


CX500在继承以往的体系结构的基础上利用了新的技术设计SP(StorageProcessor)。CX500数据存储系统目前每个SP有2个前端光纤端口、2个后端光纤端口、4GB的cache以及1个以太网端口用于系统管理。CX500提高了性能、数据管理、数据安全性、灾难恢复能力。目前CX500支持：同步远程镜像（MirrorView）；某一时间点的数据镜像（Snapshot）；不间断的软件升级（nondisruptiveupgradeprocess）；真正的同／异构计算环境。

从下图中可以看出，CX500的体系结构完全无单点故障。



1. **CX500的数据区**



SPE没有内置磁盘。为PSMLUN，bootdirectoryservices，FLAREdatabasedrives而设置的”保留区”位于第一个DAE。这个DAE同时和SPE中的SP和SPS连接。“FLAREdatabasedrives”被置于磁盘0，1，2。SPA的”bootinformation”被置于磁盘0和2，SPB的”bootinformation”被置于磁盘1和3。”PSMLUN”(1024MB)被置于磁盘0，1，2。这是一个3元镜像组。”PSMLUN”之后是”vault”区域。这是掉电后Cache内容的保留地。

1. **CX500特性**

CX500是新的一款企业级存储旗舰产品。她的设计要点在于智能化存储管理，通过多层次的软件，CX500能提供多种功能包括：LUN屏蔽、远程镜像、快照等。这使得CLARiiON成为市场上同档产品中功能最强大者。

CX500主要特性如下：

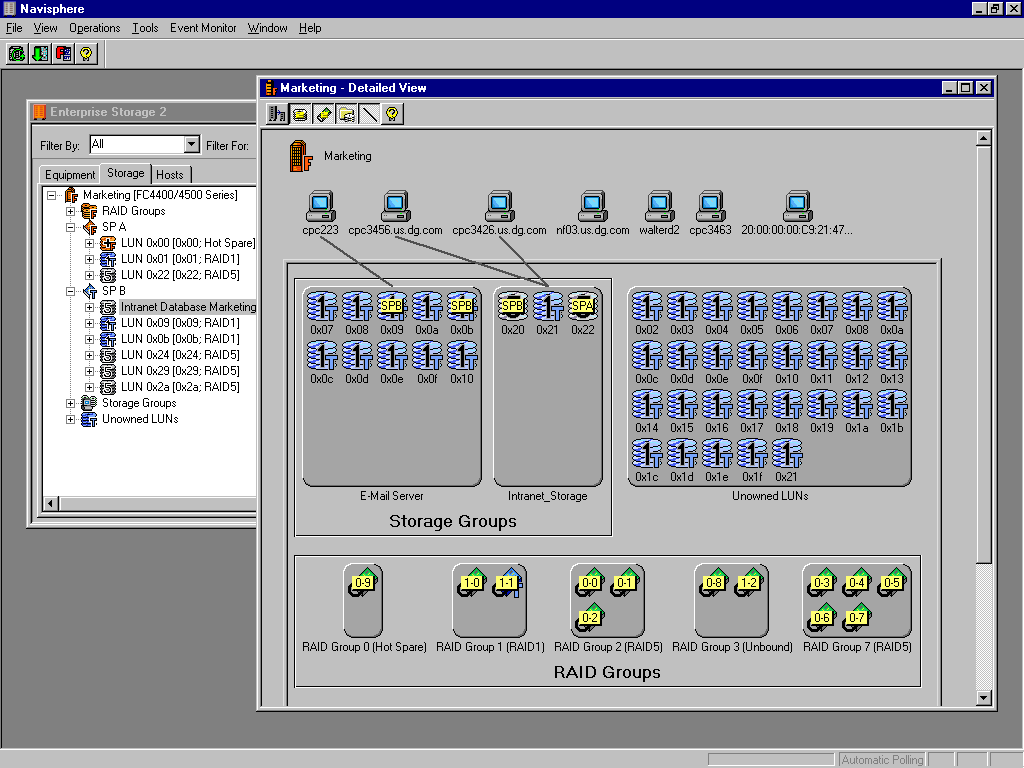
* 2个1.6GHz微处理器内置在SP中
* 2个200MB/s前端光纤口（每个SP），阵列总共4个前端光纤口
* 2个200MB/s后端光纤口（每个SP），阵列总共4个后端光纤口
* 2个200MB/sCMI通道（SP-to-SP通信）
* 每个SP2GBCache，阵列总共4GBCahce
* 完全2Gb光纤通道技术，不论是前端SwitchedFabric还是后端光纤环路
* 真正的同／异构多主机支持
* 支持FC-AL和FC-SW
* 最小5块硬盘的配置
* 支持的RAID级别：0,1,1/0,3,5和单个硬盘包括globalhotsparing
* 不间断的软件升级（NDU）
* 提供服务端口：1个115K波特率的串口
* 内置以太网端口提供Navisphere®Manager的连接
* 最初版本支持36/73GB10Krpm和36/73GB15Krpm硬盘,40U机柜最多可容纳165块磁盘,单个阵列可容纳120块磁盘
* 电源　　XPE:47-63Hz,520VA,510W;DAE:47-63Hz,400VA,392W;40U机柜:单相,47-63Hz,220-240VA
* 性能：120K+IOPS，760MB/s带宽

### Navisphere存储管理软件

强大的EMCNavisphere存储管理软件是一种适宜的工具，可帮助客户有效地配置、监控和调整其CLARiiON系统。Navisphere企业应用程序，如管理器、监控器、ATF（应用通明容错）、分析器、整合器、事件监控器和CLI等，使得管理人员能够很好地监控管理安装于网络上的所有CLARiiON系统。

信息保护和数据资产控制是阵列管理的一个重要组成部分。Navisphere提供了一系列平衡的性能以满足现今操作环境的方便管理和安全性要求。Navisphere与EMC控制中心整合在一起。通过ECC图形用户接口，客户可点击下载整套Navisphere软件产品。ECC控制中心和Navisphere一起通过遍及世界的中央单元，向整个企业存储网络提供了无与伦比的存储管理。Navisphere软件通过提供全面广泛的商务存储管理，将CLARiiON系统的功能扩展了。它还简化了所有CLARiiON系统的管理，提高了系统管理人员的工作效率，并减少了存储管理费用。

Navisphere存储管理软件向存储管理者提供了一套独一无二的工具，它是为跨带有CLARiiON系统的通用型异机种服务器平台进行常见的“观感”式管理而特别开发的。例如，用于管理Sun服务器磁盘阵列的指令同样可用于管理IBM或HP服务器上的磁盘阵列。这种结构大大减少了管理费用，因为随高度熟练技术人员工作效率的提高，它可减少对培训管理人员的需求。



1. **Navisphere管理器(NavasphereManager)**

Navisphere管理器为安装于整个公司的CLARiiON系统的配置和管理提供了方便的途径。管理器有一个WindowsNT型用户接口，它具有常见的分层树状结构，因而客户无须进一步培训就可使用。所有阵列都可在单个窗口中看到和管理。

avisphere允许客户对其视窗的优先级进行选择：1）硬件中心设备窗口，它显示了硬件配置“地图”，并能用于迅速识别硬件故障，或者是2）存储中心窗口，它显示了存储器各逻辑部件以及它们间相互关系。

Navisphere管理器可自动执行多个任务，如生成RAID群和固件下载，从而节省了时间。Navisphere管理器可支持大量阵列，使它们具有极佳的响应时间。

1. **Navisphere指令行接口（CLI）**

NavisphereCLI提供了一系列文本指令，它们可在非图形环境或远程单元环境下运行。NavisphereCLI用于向阵列发出指令、请求阵列状态、记录指令和结果输出，以及用作一种故障确定工具。它还可以脚本方式用于连接LUN和收集信息，而不会造成性能冲突。NavisphereCLI可通过接口进程在与存储系统相连接的主机上运行，其中接口进程的功能是作为一个传输器。

### PowerPath 软件简介

EMC公司的软件PowerPath是安装在UNIX，NT主机上的强有力的通道管理软件。

1. **PowerPath功能简介**
   * PowerPath的卓越功能包括多路存取，负载均衡，通道故障隔离，故障切换和恢复。
   * 通过多条通道（多达32条）并发存取，全面提升I/O吞吐率，更快地完成更多的任务。
   * 业界独特的自动负载均衡算法智能地管理多条I/O通道的流量，极大地提高了系统的效率和I/O的吞吐率。
   * 出色的动态配置功能减轻了系统管理负担。
   * 通过自动故障监测、通道切换、把I/O队列分布到预定的替换数据通道上，保持正常的应用程序运行和信息流。
   * 允许动态恢复已修复好的部件。
   * 全面提高UNIX和WindowsNT及高可用集群环境的可用性和灵活性。
2. **发掘企业服务器潜力，增加信息可用性**

EMCPowerPath是一种驻留在服务器上增强服务器性能和信息可用性的软件解决方案。EMCPowerPath把多条I/O通道的能力、自动负载均衡和通道切换功能集成在一个综合的软件包中，用在与EMC企业存储系统CLARiiON相连的HP(HP－UX)、Sun(Solaris)、IBM(AIX)、DEC(DigitalUNIX)及WindowsNT平台上。PowerPath还支持高可用的集群（Cluster）环境。PowerPath运行在与EMCCLARiiON系列存储系统相关联的环境下。PowerPath软件驻留在开放服务器平台上，支持FWDSCSI，UltraSCSI以及光纤通道接口。EMCPowerPath与CLARiiON的结合，使您能够利用以EMC企业存储为中心的业界领先的信息存储系统的所有优势以及最高的可用性和最高的性能来实现您的IT战略。

1. **多条通道和负载平衡，使您快速工作，抓住更多商机**

使用PowerPath，不仅可以做到多条通道共同分担I/O负载，而且可以做到自动负载均衡以保证服务器在尽可能最有效的状态下运行。多通道允许两条或更多条通道（多达32条）同时用来作读/写操作。通过自动地并且均匀地把数据分布在所有可用的通道上，大大地提高了系统的性能。EMCPowerPath是业界唯一的提供这种多条通道负载均衡能力的软件。正是这种能力使得在更短的时间内可以做更多的工作，在相同的时间内可以有更多的客户被服务，更多的应用在运行，更多有利的商业机会被利用。

PowerPath负载均衡特性保证不会发生一条通道非常繁重而其它通道带宽没用被充分利用从而引起I/O瓶颈这种事情。当一条或多条通道变得比其它通道更忙时，PowerPath就会把繁忙通道的I/O流量转移到其它通道上，进一步加强了已经非常高效的多通道特性的吞吐率。

1. **自动通道切换使您的业务蒸蒸日上**

PowerPath的自动通道切换和动态恢复特性使得当某一条通道出现故障时，系统允许数据立刻自动的并且不中断的从一条或多条替换通道被访问。这就消除了由于一块控制卡、电缆或通道控制卡失败造成的应用程序中断的可能性。当发生通道切换时，所有已在这条通道队列里的和后续来的I/O请求都会自动的并且是不中断的排到替换通道队列里。关键应用无中断的继续运行，业务保持正常活动。

1. **增强对高可用集群系统的支持**

PowerPath特别适用于集群环境，可以阻止昂贵的操作中断。PowerPath通道切换功能避免了节点的切换，即是在主节点出现某一通道断开的情况下也保持应用的无中断操作。当前支持的集群软件见下列表。

* **SANcopy**

SANCopy是一种CLARiiON软件应用程序，安装在CLARiiON系统上。客户可以使用SANCopy同时移动信息，而不用考虑主机操作系统或应用程序。这对于内容分发，亦即将应用程序或支持应用程序数据移动到分布式环境以提高性能来说，是十分重要的。目前，SANCopy运行于CLARiiONCX700、CX500、FC4700、CX400和CX600存储系统上。它相当于一个通过SAN（或LAN/WAN）基础结构在系统间移动数据的工具，从而不再需要占用宝贵的服务器CPU周期和LAN带宽（假如通过SAN）。

SANCopy可移动和拷贝数据的范围包括：单个CLARiiON系统内、CLARiiON系统之间、CLARiiON和Symmetrix系统之间，甚至其他供应商的阵列之间（EMC.com上的EMC支持列表列出了当前支持的所有阵列）。重要的是，所有SANCopy操作的管理与所有CLARiiON系统管理都是通过同一个图形用户界面执行的，即NavisphereManagementSuite。或者，对于例行操作过程，您也可以通过编写NavisphereCLI脚本来安排数据移动会话。

* **Snapview**

Snapview是某个时间点的数据镜像，可用来实施Non-disruptiveBackup、建立测试环境、数据仓库等多种功能。其原理如下图，在同一台阵列中，对生产数据卷做一份某时间点的“快照”，这份“快照”被mount到其它的主机用于非实时的应用。

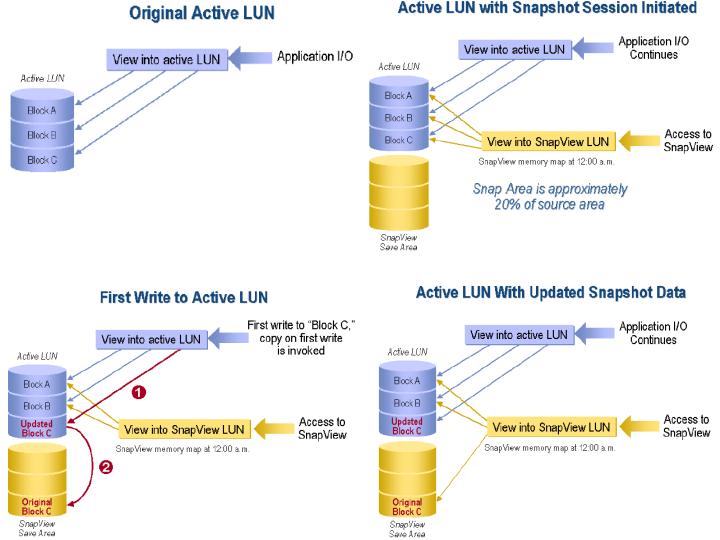
技术说明--“基于指针”的快照

1．初始状态：应用直接访问“源”数据

2．12:00时建立“快照”，应用可以开始访问“快照”。

3．建立“快照”后发生对“源”数据的第一次写操作，原始“源”数据被转而保存到“SaveArea”，数据块被刷新。

4．应用仍然可以访问12:00时的系统数据，12:00之后被覆盖的数据并没有丢失，而是转而保存到了“SaveArea”。发生数据访问时，访问被自动定向到原始的卷或“SaveArea”。



* **Mirrorview**

Mirrorview是远程数据镜像，容灾解决方案。在不同地点的两台阵列之间实现生产数据的复制和生产系统的切换。MirrorView软件对CX500,CX700阵列全面兼容。MirrorView和SnapView的结合使用就可以营建非常典型且强大的“MissionCritical”应用环境。

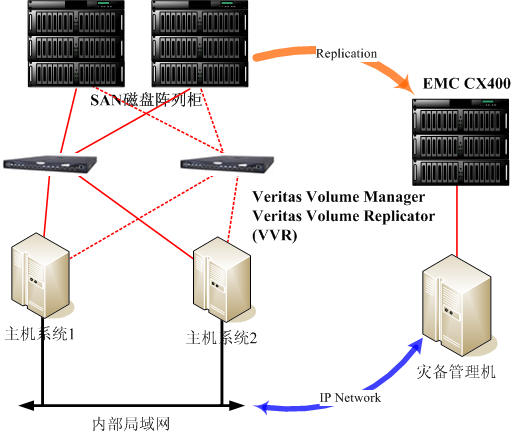
1. 远程数据备份方案

高可用系统虽然能够做到数据中心本地系统的高可用，但有一个问题，如果一些不可抗拒的因素如火灾、雷击等把整个中心的一些关键系统破坏了，还是会导致应用系统的长时间中断。而对于同济大学的校园信息化系统，数据的丢失是绝不允许的，因此，有必要建设一个具有容灾功能的备份中心，当数据中心发生故障时，可由备份中心接管部分或所有的业务处理。

考虑到在发生不可抗拒的灾难时，较短的停机（数量级为天）或系统的不可运行是允许的，以及一个具有实时应用切换的灾备中心建设的投资是相当巨大的，建议同济大学的灾备中心只是较简单的数据级的灾备，先期建立数据中心与灾备中心的数据同步传输系统，完成数据的灾备。

## 备份系统结构

整个系统的结构大致如下图所示：



系统的运作过程如下：

1. 正常情况下，数据中心和灾备中心的系统均处于运行状态，但业务处理系统只在数据中心运行；同时，业务系统对数据的任何修改，会实时同步地复制到备份中心
2. 当数据中心的某些部件发生故障，如进程出错、内存损坏等情况发生时，产生故障的机器上的应用系统会自动地快速切换到数据中心的其他机器，整个系统正常运行
3. 当灾难发生，导致数据中心整个系统瘫痪时，按照预定的规则在灾备中心重新建立应急系统
4. 数据中心的计算机网络系统修复后，VVR可将灾备份中心的当前数据复制回数据中心，然后将应用系统从灾备中心切换回数据中心，灾备中心的系统重新回到备份状态。

## 远程数据同步复制

远程数据同步复制的实现包括两个部分：有足够带宽的网络连接；优秀的数据复制管理软件。

对于网络连接，如果距离不超过40公里，建议采用光纤连接，如果连接距离超过40公里，建议采用广域网专线连接。

对于数据复制管理软件，建议采用VERITAS的Volume Replicator (VVR)。通过广域网远程连接，VVR可向远程备份系统同步进行逻辑卷复制，确保系统数据的高度可用。VVR采用可靠的连接和监听协议，保证远程备份站点与本地逻辑卷数据的一致性，为用户关键应用的灾难备份和恢复提供了有效的手段。

VVR具有下述特点：

* 支持广域网节点间数据的同步和异步复制
* 支持多点到多点的复制，一份数据可同时复制到32个节点，多个节点的数据也可同时复制到一点进行集中
* 容忍网络延迟：在同步模式下，若网络发生堵塞，可自动切换到异步模式，当网络恢复后，再重新同步

## 系统配置

容灾方案配置：

软件：

Veritas Volume Replicator (VVR) 同济大学已购

硬件：

EMC CX300磁盘阵列柜

灾备管理机（PC服务器）