



解决方案

将繁琐的容灾工作交给我们
你来喝杯咖啡……

Solutio

ISCSI 环境下，集群（应用虚拟化）的实现

存储区域网络（SAN）的发展，使得更多的应用系统，可以有效的进行存储集中，更加多的手段来提升数据的安全性。存储虚拟化、服务器虚拟化，可以引申出应用的虚拟化。‘应用虚拟化’的实现是通过服务器集群来实现的，对于应用虚拟化这个概念，可以从我们的其它技术文章中看到详细描述。而 SAN 环境构架，曾经是 FC SAN 的专利，现在随着 ISCSI 技术的蓬勃发展，一种基于 IP 的 SAN 构架，正在迅速发展壮大起来。这篇文章我们从 ISCSI 环境构架讨论，探讨在 ISCSI 环境下，集群（应用虚拟化）的实现。

ISCSI 技术

首先介绍一下 ISCSI 技术

ISCSI 技术于 2003 年 2 月 11 日，由 IETF (Internet Engineering Task Force, 互联网工程任务组) 正式通过; 它由 IBM、思科共同发起，是一种基于网络的数据存储技术，具有硬件成本低廉，操作简单，扩充性强，传输速度快等特点。

ISCSI (互联网小型计算机系统接口) 是一种在 Internet 协议网络上，特别是以太网上进行数据块传输的标准。它是由 Cisco 和 IBM 两家发起的，并且得到了 IP 存储技术拥护者的大力支持。是一个供硬件设备使用的可以在 IP 协议上层运行的 SCSI 指令集。简单地说，ISCSI 可以实现在 IP 网络上运行 SCSI 协议，使其能够在诸如高速千兆以太网上进行路由选择。

ISCSI 对于大多数用户，比 FC SAN 更具有竞争力

ISCSI 技术和产品的出现，使得用户构架一个基于 IP 的 SAN 环境变得更加容易，由于目前市场上基于 ISCSI 的存储产品越来越多，通过这些产品构架一个 SAN 存储环境相对构架一个 FC SAN 环境整体投入相当的低。这是由于 FC SAN 需要专用的存储光纤交换机、HBA 卡、FC 存储设备等，这些设备目前还是相当的昂贵，对于一般用户来讲，这个投入可能不是很值得。当然 FC SAN 环境速度相当快，对于一些大规模数据库应用、并发度高的用户环境，这种配置是必须的。而大多数应用环境，对存储处理能力的要求并不是想象中的那么吓人，往往连用户自己都不知道自己需要什么样的处理能力的存储。实际上，ISCSI 环境下的性能瓶颈主要在网络上，因为 ISCSI 是通过以太网传输存储数据的。当然，在 ISCSI 存储设备端和服务器端存在封包和解包过程，这些会占用一些系统资源，如果使用专用的 ISCSI 卡，这种处理都在 ISCSI 卡上完成，不会占用系统资源。只有我们直接使用普通千兆以太网卡来连接到 IP SAN 环境中，才需要更多的系统资源。而随着网络的传输能力越来越强，服务器处理能力越来越强，这些都不是主要问题了。我们完全可以在相当低的预算范围内，甚者仅仅是以前的一个 DAS (直接连接存储，比如 SCSI 接口磁盘阵列柜) 设备的投入，就可以完成一个 IP SAN 环境的构建。

下表可以了解一下，通过两种方式，构建一个 SAN 环境所需要的设备列表：

设备	FC SAN	IP SAN
磁盘阵列存储	外部 FC 接口 (2GB/4GB)	外部以太网接口 (千兆)
磁盘模块	SATA/SAS/SCSI/FC	SATA/SAS/SCSI
交换机	2GB/4GB 存储交换机	千兆网络交换机
HBA 卡	2GB/4GB 光纤 HBA 卡	ISCSI 卡或千兆以太网卡
线缆	存储光纤用线缆	网络线缆

ISCSI 环境下，构建一个集群

IP SAN 的构建，使得我们的用户可以将网络中的各个服务器，上的数据存放在 ISCSI 存储设备上，并且能够像 FC SAN 一样，在 ISCSI 磁盘阵列上划分的分区 (LUN)，这些分区能够被其它连接在 IP SAN 中的服务器所共享，这样就使得我们构架一个高可用性集群环境成为可能。

我们知道，在 IP SAN 环境下，我们可以不管服务器存放在何处、ISCSI 存储设备存放在何处，只要网络链路能够通达就可以加入到这个 SAN 环境中。这样我们就在这个 IP SAN 环境中实现了存储虚拟化。那么怎么实现集群呢，这里我们讨论高可用性集群的实现。

高可用性集群解决 SAN 环境中各个应用系统的高可靠性，简单的说，就是任何服务器节点的宕机，不会导致运行在该服务器节点的应用系统停止工作，它会根据预定义的备援规则进行转移，这个转移包括了应用系统 (服务、进程等)、网络资源 (IP)、存储资源 (卷)。这些资源通常被捆绑在一起，我们称为任务包 (Package)，它在逻辑上独立于服务器主机 (服务器节点，Node)，这样实际上是把处理单元 (服务器节点) 和应用系统 (任务包) 逻辑上完全分开，而任务包可以在任何可以运行的服务器节点上运行，对于客户端访问者来讲，不知道数据是存放在哪里，也不知道自己使用的应用环境是运行在哪台服务器上，实现了应用环境的高可用。下面我们以 LanderCluster 为例，来阐述一下这个环境的实现过程。

LanderCluster 是一种集群管理软件，能够在 IP SAN 环境下，构建一个多种工作方式服务器互备环境，这种互备方式，可以是多台服务器相互备份、一台备份多台、多台备份多台等。而我们仅仅需要考虑的是集群系统的规划，也就是我们前面提到的备援规则定义，这是高可用集群系统定义的核心。主要要考虑的包括应用系统的兼容性、服务器处理能力等。

应用系统兼容性主要是指两个或多个应用系统是否可以在同一台服务器环境中运行，比如两个同实例名的 ORACLE 数据库是不能在同一台服务器上同时运行的，那么必须在备援规则上予以规划，使它们在任何时刻不能向同一台服务器上切换。

服务器处理能力主要是要分析是否两个或多个应用如果同时运行在一台服务器上，该服务器是否会因为过载而死机。通常情况是服务器不会死机，但会变得运行迟缓，这种情况发生，说明备援规则规划存在问题。

我们可以看出在服务器节点越多的情况下，备援规则会越复杂，但整个系统的规划会越发灵活，可靠性会越高。因为不同于双机环境，每台服务器应用只能是另外一台做备份服务器，在集群环境下，理论上的其它服务器，只要满足条件，都可以是备援服务器。

用 LanderCluster 构建这样集群

由于集群还是比较复杂，大多数用户和存储爱好者对它的认识，认识还不够深刻，不能够有效的和实际环境联系起来，也不知道如何下手来实现这样的环境。我们为相当多的用户构建了这样的集群环境，那么就结合我们的实践经验来分享一下这个概念。下面以大多数政府用户的典型环境进行设计分析，来理解前面提到的概念。

很多的政府用户典型的应用环境包括：业务数据库服务器（可能是 ORACLE、SQL Server、MySQL 等）、应用服务器（中间件服务器）、办公自动化服务器（OA 服务器，Notes 等）、文件服务器、财务服务器等，这些服务器往往都是单独运行的，个别环境会在核心的数据库服务器做个双机环境，确保关键部分高可用，而实际上，现在的用户系统已经是一个复杂的多机协同提供服务的环境了，数据库服务器是可靠的，那么应用服务器如果宕机，整个系统环境同样是不可靠的，会因为其中的某台服务器宕机而停止服务，这样环境的整体可靠性，取决于整个环境中各个服务器、应用的可靠性。因此，真正能有效提高整体可靠性的手段，就是采用集群技术。

这些服务器都可以整合在一个 IP SAN 环境中，通过 LanderCluster 实现高可用。我们来分析这个环境，这些服务器环境的应用数据都存放在 ISCSI 存储设备的相关分区上，ISCSI 环境规划参考下表：

注：下表仅仅是参考环境，取环境中的几个应用系统进行示范，所有定义内容均是参考数据

服务器	数据库服务器	应用服务器	文件服务器	财务服务器
主机名	DBserver	WEBserver	FILEserver	FINserver
应用类型(例)	ORACLE	WEBlogic		MySQL
ISCSI 分区 (LUN)	1	2	3	4
卷定义 (Windows 环境, Linux 是文件系统)	D:	E:	F:	G:
IP 规划	192.168.8.1	192.168.8.2	192.168.8.3	192.168.8.4
操作系统	Windows/Linux	Windows/Linux	Windows/Linux	Windows/Linux

在这样一个环境下，我们需要定义备援规则，首先看这几台服务器的应用系统：ORACLE、MySQL、文件服务、WebLogic 之间没有冲突，也就是说这些应用都可以同时运行在任何一台服务器上，这是最完美的状况，可以达到最高的可靠性定义。我们假定服务器处理能力都能够满足处理能力的要求。

那么，备援规则的定义可以如下表：

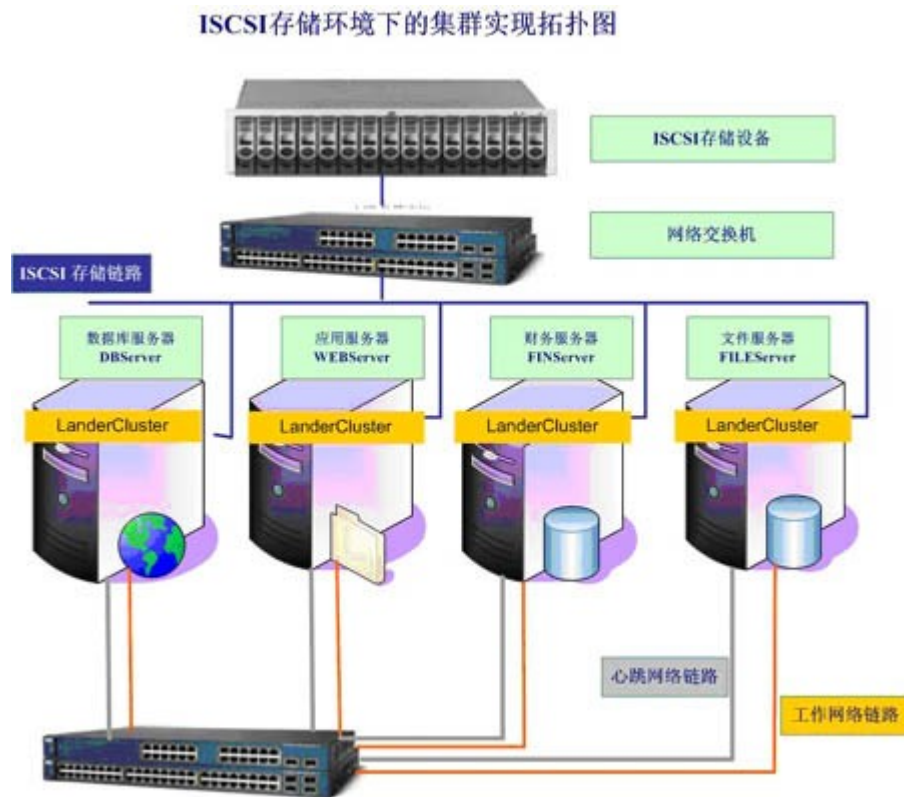
服务器	第一备援	第二备援	第三备援
DBServer	FINServer	FILEServer	WEBServer
WEBServer	FINServer	FINServer	DBServer
FILEServer	FINServer	DBServer	WEBServer
FINServer	DBServer	WEBServer	FILEServer

这是一种完美状态的环境规划，在实际运行环境中，每个服务器节点的备援理论上达到 2 个，就已经具有非常高的可靠性了，上面的定义包含了各种极端情况都能保证系统不停止提供服务，比如任意 3 台服务器

宕机，那台好的服务器也能提供 4 个应用的服务。

简单的以数据库服务器为例说明工作方式，在正常运行时，四台服务器分别运行自己的应用。某个时刻数据库服务器宕机，则 LanderCluster 会在 10 秒钟之内发现，立刻检索备援规则表（存放在每个节点机的内存中），发现第一备援机是财务服务器（FINServer），在确定财务服务器是完好的情况下，将数据库服务器上的资源包，整体切换到财务服务器上运行（包括 IP、ISCSI 分区、ORACLE 服务等），切换过程不超过 50 秒。如果 LanderCluster 侦测发现财务服务器是坏的，则会检查第二备援节点状态，如果第二节点文件服务器（FILEServer）是好的，则切换到文件服务器上工作，依次类推。这样我们知道，集群环境的规划相当的重要。

环境拓扑如下图：



这样我们就实现了在 ISCSI 环境下的集群。