Linux

集群技术研究

彭丽艳/著

云南大学出版社
YUNNAN UNIVERSITY PRESS

Linux 集群技术研究

彭丽艳　著

云南大学出版社
YUNNAN UNIVERSITY PRESS

图书在版编目（CIP）数据

Linux集群技术研究 / 彭丽艳著. -- 昆明 : 云南大学出版社, 2019

ISBN 978-7-5482-3611-5

Ⅰ. ①L… Ⅱ. ①彭… Ⅲ. ①Linux操作系统 Ⅳ. ①TP316.85

中国版本图书馆CIP数据核字(2019)第008773号

策划编辑：王翌沣
责任编辑：王翌沣
封面设计：周　凡

Linux集群技术研究
彭丽艳　著

出版发行：云南大学出版社
印　　装：昆明理煜印务有限公司
开　　本：880mm×1230mm　1/32
印　　张：7.5
字　　数：150千字
版　　次：2019年1月第1版
印　　次：2019年1月第1次印刷
书　　号：ISBN978-7-5482-3611-5
定　　价：39.8元
社　　址：昆明市一二一大街182号
（云南大学东陆校区英华园内）
邮　　编：650091
电　　话：（0871）65033244　65031071
E-mail：market@ynup.com

若发现本书有印装质量问题，请与印厂联系调换，联系电话：0871-64167045。

作者简介

彭丽艳，生于1980年，四川雅安人，大学本科学历，讲师职称。毕业于电子科技大学计算机学院，现任职于四川托普信息技术职业学院计算机系，主要研究方向为计算机网络技术。曾发表《虚拟存储产生容灾问题的数据备份关键技术分析》《浅谈数字证书在网络安全中的应用》《职业技术学院中计算机网络课程教学方法探讨》等数篇论文，并于2015年参加四川省信息化教学课堂大赛，获得三等奖。

前　言

Linux 是当今发展最为迅速、关注度非常高的操作系统之一。随着人们对 Linux 服务器依赖的加深，用 Linux 集群技术构建网络服务器就成为未来网络服务器发展的一个方向。用集群技术构建网络服务器的基本思路，就是把原先独立的服务器通过网络技术连接起来，作为一个整体（集群系统）对外提供服务，并且要把到达的服务请求分配到集群中的各台服务器上，让它们均衡地分摊负载，缩短访问的响应时间。构建集群系统的核心问题是实现服务器间的负载均衡，它直接关系到集群系统的系统性能和可扩展性、可用性。

计算机系统被广泛应用于社会生活的方方面面，为人们提供各种及时可靠的服务及信息。不论是在日常生活中，还是在通信、金融、物流等关键性行业，我们都需要通过计算机系统对信息进行处理。这不仅使社会信息化程度不断增加，更进一步提高了整个社会的运转效率，实现了社会生产资源的合理利用。然而随着对计算机系统依赖性的不断增加，服务器能否持续可靠地运行成为一个尤为突出的问题，如何能最大限度地提高系统可用性就显得至关重要。互联网用户数和网络流量的迅速增长，对网络服务器的可扩展性和可用性提出了更高的要求。传统的单服务器模式已经不能应对不断增长的负载。高性能服务器集群系统将成为实现高可扩展性、高可用性网络服务的有效体系。随着互联网公司集群规模的扩大，服务器数量越来越多，服务器之间的关联关系日渐复杂，安全问题带来的风险也在不断增加。安全问题涵盖的内容包括产品设计、开发、测试、运维、基础设施（IDC、内网、外网、办公网）等各个方面，其中 IDC 线上服务器的权限是一个很重要的环节，需要进行合理、规范、统一的登录权限验证、管理。

当今计算机技术已进入以网络为中心的计算时代。由于客户/服务器模型的简单性、易管理性和易维护性，因此，客户/服务器计算模式在网上被大量采用。在20世纪90年代中期，万维网出现并以其简单操作方式将图文并茂的网上信息带给普通大众，Web也正在从一种内容发送机制成为一种服务平台，大量的服务和应用（如新闻服务、网上银行、电子商务等）都是围绕着Web平台进行。这促进了Internet用户剧烈增长和Internet流量爆炸式增长。现在Web服务中越来越多地使用CGI、动态主页等CPU密集型应用，这对服务器的性能有较高要求。未来的网络服务会提供更丰富的内容、更好的交互性、更高的安全性等，需要服务器具有更强的CPU和I/O处理能力。例如，通过HTTPS取一个静态页面需要的处理性能比通过HTTPS高一个数量级，HTTPS正在被电子商务站点广为使用。所以，网络流量并不能说明全部问题，要考虑到应用本身的发展，也需要越来越强的处理性能。

目前，Linux操作系统的应用日趋成熟，Linux在很多计算机领域都占有一定的市场份额。基于Linux构建Web、Mail、FTP等应用服务器，已成为当今的主流，本书对Linux集群技术在应用历程中的一系列问题进行了系统的剖析，并提出了部分具有建设性的建议和意见，对于Linux集群技术的开展与推进具有非常重要的现实意义和理论价值。由于时间、水平有限，书中难免有疏漏之处，恳请广大读者批评指正。

目　录

第一章　Linux 集群技术研究及应用

第一节　浅析 Linux 集群技术　02

第二节　Linux 集群技术在校园网中的应用　14

第三节　Linux 集群文件系统 GFS 的研究应用　23

第四节　一种基于 Linux 集群技术的负载均衡方法　28

第二章　一种高可用性 Linux 集群管理系统

第一节　集群管理工具 KUSU　58

第二节　高可用性集群管理系统的需求分析　68

第三章　Linux 集群系统任务调度策略

第一节　目前集群技术的研究情况　82

第二节　Linux Virtual Server 集群　87

第四章　Linux 集群运维平台用户权限管理及日志审计系统实现

第一节　Linux 集群运维平台设计中的关键问题及解决方案　96

第二节　Linux 集群运维平台权限管理模块的系统需求　121

第三节　Linux 集群运维平台日志审计系统的系统需求　125

第五章　大规模 Linux 集群部署系统的研究及实现

第一节　Linux 集群系统部署的方法　132

第二节　一种基于 Script 和镜像的混合部署方法　140
第三节　Linux 集群部署系统的实现　153

第六章　基于 Linux 和 MPI 的集群并行系统的研究与实现

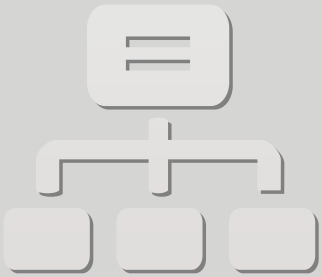
第一节　集群体系结构及集群进程间通信　166
第二节　并行计算理论　183
第三节　Linux 技术知识　191

第七章　基于 Linux 集群的可伸缩电子邮件系统的研究

第一节　服务器集群技术发展状况　198
第二节　Linux 集群的体系结构　208
第三节　集群电子邮件系统的体系结构　219

结　语　223

参考文献　224



第一章

Linux 集群技术研究及应用

集群技术（Cluster 技术）就是将多台服务器用集群软件连接在一起，组成一个高度透明的大型服务器群的计算机系统，作为一个整体为客户端提供服务，客户端能共享网络上的所有资源，如数据或应用软件等，当集群系统内某一台服务器出现故障时，其备援服务器便立即接管该故障服务器的应用服务，继续为前端的用户提供服务。从客户端来看，集群中的所有服务器是一个系统，就像一台大型的计算机系统，其上运行着客户端需要的应用服务。同时客户端的用户并不关心其应用 Server 运行在哪台服务器上，只关心其应用 Server 能否连续工作。集群系统能够保证用户的业务是连续的并且具有持续可用的特性，即具有 7×24 小时的可用性，在一年之内可达 99.99％可用性时，这样的集群系统称为高可用性的集群系统。

第一节　浅析 Linux 集群技术

随着现代社会网络应用的逐步深入，越来越多的服务器采用 Linux 操作系统，提供邮件、Web、文件存储、数据库等服务。现在已有非常多的公司在企业内部网中利用 Linux 服务器提供这些服务。由于人们对 Linux 服务器依赖的加深，对其可靠性、负载能力和成本也倍加关注，Linux 集群技术应运而生。其以低廉的成本、高效的性能很好地满足了人们的需要。

一、Linux 集群技术简介

Linux 集群是最近几年最为主要的一种 HPC 硬件，集群(Cluster)就是一组 MPP 的集合。集群中的处理器通常被称为节点，

它具有自己的 CPU、内存、操作系统、I/O 子系统，并且可以与其他节点进行通信。目前市场上出现了很专业的集群设备，但有很多地方都使用常见的工作站运行 Linux 和其他开放源码软件来充当集群中的节点。与专业的硬件设备相比，使用 Linux 工作站和相应的开放源码软件来实现集群功能的最大优势在于成本低，且企业可根据自身的需求进行二次开发。

（一）集群技术的用途

我们知道，每台服务器所能承载的连接和负载量都是有限的，为了使服务器能够承载更大的负载，我们一般采用对称多处理(Symmetric Multi-Processor,SMP)技术来提升服务器的整体性能。但是，SMP 服务器的可扩展能力有限，显然不能满足高可伸缩、高可用网络服务中的负载处理能力不断增长的需求。随着负载处理能力不断增长，会导致服务器不断升级。这种服务器升级有下列不足：一是升级过程烦琐，机器切换会使服务暂时中断，并造成原有计算资源的浪费；二是越高端的服务器，所花费的代价越大；三是 SMP 服务器是单一故障点（Single Point of Failure），一旦该服务器或应用软件失效，就会导致整个服务的中断。而通过高性能网络或局域网互联的服务器集群则可以解决上述问题，并能实现高可伸缩、高可用网络服务。与提升单个服务器配置方法相比，服务器集群技术具有以下典型的优点：

1. 性能

网络服务的工作负载通常是大量相互独立的任务，通过一组服务器分而治之，可以获得很高的整体性能。

2. 性能 / 价格比

组成集群系统的 PC 服务器或 RISC 服务器和标准网络设备因为大规模生产降低了成本，价格低，具有最高的性能 / 价格比。若整体性能随着节点数的增长而接近线性增加，该系统的性能 /

价格比接近于 PC 服务器。所以，这种松耦合结构比紧耦合的多处理器系统具有更高的性能 / 价格比。

3. 可伸缩性

集群系统中的节点数目可以增长到几千，乃至上万个，其伸缩性远超过单台超级计算机。

4. 高可用性

集群技术在硬件和软件上都有冗余，通过检测软硬件的故障，将其屏蔽，由存活节点提供服务，以实现高可用性。

（二）Linux 集群类型

专家按照集群侧重点的不同，把 Linux 集群分为以下三类：

1. 高可用性集群

最简单的高可用性集群有两个节点：一个节点是活动的，另一个节点是备用的。备用节点会一直对活动节点进行监视，一旦活动节点出现故障，备用节点就会接管它的工作，这样就能使关键的系统能够持续工作。

2. 负载均衡集群

负载均衡集群通常会在非常繁忙的 Web 站点上采用，它们由多个节点来承担相同站点的工作，每个获取 Web 页面的新请求都被动态路由到一个负载较低的节点上。动态路由是指路由器能够自动地建立自己的路由表，并且能够根据实际情况的变化适时地进行调整。

3. 高性能集群

高性能集群用来运行那些对时间敏感的并行程序，它们对科学社区意义特殊。高性能集群通常会运行一些模拟程序和其他对 CPU 非常敏感的程序，这些程序在普通的硬件上运行需要花费大量的时间。

高可用性集群、负载均衡集群及高性能集群三者的工作原理

不同，适用于不同类型的服务。通常负载均衡集群适用于提供静态数据的服务，如 HTTP 服务；而高可用性集群既适用于提供静态数据的服务，又适用于提供动态数据的服务，如数据库等。高可用性集群之所以能适用于提供动态数据的服务，是由于节点共享同一存储介质，如 RAID BOX。也就是说，在高可用性集群内，每种服务的用户数据只有一份，存储在共用存储设备上，在任一时刻只有一个节点能读写这份数据。

二、Linux 集群的实现方案

根据不同的应用需求，Linux 集群有三种类别的集群方案：高可用性集群方案、负载均衡集群方案、超级计算集群方案。

（一）高可用性集群方案

高可用性集群系统中不同的服务器承担不同的任务，当其中一个服务器发生故障时，系统根据设定的条件，将发生故障的服务器的任务转移给另外一台服务器，对最终用户来说，并没有反映出系统故障，系统的可用性将得到提高。

（二）负载均衡集群方案

1. 负载均衡集群的架构

在负载均衡的解决方案中，若干台服务器做同样的工作。这样一来，以前由一台服务器来做的工作被分配给多个服务器来做，因而整个系统的处理能力得以提高。

（1）负载均衡器（Load Balancer）

Load Balancer 是整个集群系统的前端，负责把客户的请求通过特定的调度算法转发到 Real Server 上。Backup 是备份 Load Balancer，当 Load Balancer 不可用时接替它，成为实际的 Load Balancer。Load Balancer 通过 Idirectord 监测各 Real Server 的健康状况，在 Real Server 不可用时把它从群中剔除，恢复时重新加入。

（2）真实服务器组（Server Array）

Server Array 是一组运行实际应用服务的机器，如 Web、Mail、FTP、DNS、Media 等。在实际应用中，Load Balancer 和 Backup 也可以兼任 Real Server 的职责。

（3）共享存储（Shared Storage）

Shared Storage 为所有 Real Server 提供共享存储空间和一致的数据内容。

2. 负载均衡集群的算法

目前，Linux 平台下的集群软件有很多，其中 Linux Virtual Server（LVS）最为流行。LVS 安装在负载均衡器上，使用虚拟 IP 地址对外服务，当接收到客户请求后，根据特定的调度算法将客户请求转发到选择的真实服务器。LVS 支持的调度算法如下。

（1）轮叫调度（RR）

调度器通过“轮叫”调度算法将外部请求按顺序轮流分配到集群中的真实服务器上。它均等地对待每一台服务器，而不管服务器上实际的链接数和系统负载。

（2）可调加权轮叫调度（WRR）

调度器通过“加权轮叫”调度算法根据真实服务器的不同处理能力来调度访问请求。这样可以保证处理能力强的服务器处理更多的访问数据。调度器可以自动问询真实服务器的负载情况，并动态地调整其权值。

（3）最小连接（LC）

调度器通过“最小连接”调度算法，动态地将网络请求调度到已建立的链接数最小的服务器上。如果集群系统的真实服务器具有相近的系统性能，采用“最小链接”调度算法可以较好地均衡负载。

（4）加权最小连接数（WLC）

在集群系统中的服务器性能差异较大的情况下，调度器采用“加权最小连接”调度算法优化负载均衡性能，具有较高权值的服务器将承受较大比例的活动连接负载。调度器可以自动问询真实服务器的负载情况，并动态地调整其权值。

（5）基于局部性的最小连接调度（LBLC）

基于局部性的“最小连接”调度算法是针对目标 IP 地址的负载均衡，目前主要用于 Cache 集群系统。该算法根据请求的目标 IP 地址找出该目标 IP 地址最近使用的服务器，若该服务器是可用的且没有超载，将请求发送到该服务器；若该服务器不存在，或者该服务器超载且有服务器处于一半的工作负载，则用“最小连接”的原则选出一个可用的服务器，将请求发送到该服务器。

（6）带复制的基于局部性最小连接调度（LBLCR）

带复制的基于局部性“最小连接”调度算法也是针对目标 IP 地址的负载均衡，目前主要用于 Cache 集群系统。它与 LBLC 算法的不同之处是，它要维护从一个目标 IP 地址到一组服务器的映射，而 LBLC 算法维护从一个目标 IP 地址到一台服务器的映射。该算法根据请求的目标 IP 地址找出该目标 IP 地址对应的服务器组，按“最小连接”原则从服务器组中选出一台服务器，若服务器没有超载，将请求发送到该服务器；若服务器超载，则按“最小连接”原则从这个集群中选出一台服务器，将该服务器加入服务器组，将请求发送到该服务器。同时，当该服务器组有一段时间没有被修改，将最忙的服务器从服务器组中删除，以降低复制的程度。

（7）目标地址散列（DH）

目标地址散列调度算法，是以请求的目标 IP 地址作为散列键（Hash Key）从静态分配的散列表找出对应的服务器，若该服

务器是可用的且未超载，将请求发送到该服务器，否则返回空。

（8）源地址散列（SH）

源地址散列调度算法是以请求的源 IP 地址作为散列键从静态分配的散列表找出对应的服务器。若该服务器是可用的且未超载，将请求发送到该服务器，否则返回空。

（三）超级计算集群方案

超级计算集群是并行计算的基础，以解决复杂的科学问题。它可以使集群系统通过高速链接来链接一组单处理器或双处理器微机，并且在公共消息传递层上进行通信以运行并行应用程序。因此，所谓廉价 Linux 超级计算机，实际是一个 Linux 计算机集群，其处理能力与真正的超级计算机相当，但 Linux 集群系统的价格比超级计算机便宜很多。支持 Linux 集群的软件及系统有 En Fuzion、Beowulf 等。Beowulf 软件的主要功能包括分发任务到各个计算节点、监测任务运行的情况、监测故障及故障恢复、控制任务队列等。

集群是在理论和实际应用上都有重大意义的技术，目前倍受各大 IT 厂商的关注，但现有集群系统在某些方面仍存在不少问题，在性能与可用性方面还不是十分理想，因此，人们仍在寻找更好的方法以使集群系统具有更高的性能与可用性，而 Linux 由于它的开源特性，必然会有更多的有志之士为它出谋划策。

三、防火墙及其集群相关技术浅析

传统单一的防火墙已经限制网络宽带的增长，极大地制约了网络的实际应用，同时也降低了网络性能及其可扩展性。主要原因有：随着网络视频、IPTV 和 P2P 业务的广泛应用，导致互联网络流量高速增长，防火墙网络带宽不够；防火墙所能处理的连接数目有限，无法处理过多用户的通信需求；防火墙身兼认证、

访问控制、完整性检查等多项任务，处理能力有限。单一的防火墙不仅存在性能问题，还存在单点故障的瓶颈问题。为了解决这种问题，目前大多数防火墙厂商采用了双机热备的方式。双机热备系统由两台防火墙及双机热备软件组成，它采用主从工作方式，即一台防火墙节点处于活动工作状态，称为活动防火墙；另一台防火墙节点为备用机，处于热等待监控状态，但是并没有改变防火墙的单点接入方式。

（一）防火墙技术

防火墙是一种网络安全产品，是在两个网络之间强制实施访问控制策略的一个系统或一组系统。它应具有的功能大致包括：拒绝未经授权的用户访问网络；阻止未经授权的用户存取敏感数据；允许合法用户不受妨碍地访问网络资源。

防火墙经历了由简单到复杂、功能不断丰富和性能不断增强的过程，其采用的技术大致分为几种类型。①包过滤技术：析进出内部网络的所有数据包，按照一定的安全策略（包过滤规则）决定数据包是否能被允许通过。包过滤规则是以所收到的数据包包头信息为基础，如正数据包源地址、正数据包目的地址、封装协议类型（TCP、UDP、ICMP 等）、TCP/UDP 源端口号、TCP/UDP 目的端口号、ICMP 报文类型等。当一个数据包满足过滤规则，则允许该数据包通过，否则拒绝通过。②代理技术：代理技术与包过滤技术完全不同，包过滤是在网络层拦截所有的数据包，而代理技术是针对每一个特定应用都有一个程序。代理的原理是在应用层实现防火墙功能，即彻底隔断通信两端的直接通信，所有的通信都必须经过应用层代理转发。③状态检测技术：状态检测防火墙是由 Checkpoint 公司率先提出的，又称为动态包过滤防火墙。状态检测技术对于新建的应用链接，首先检查预先设置的安全规则，允许符合规则的链接通过，然后在内存中记录下该链