



《Linux 系统运维之系统架构》

UNIXHOT 运维社区

<http://www.unixhot.com>

版权信息:

Copyright (c) 2010 Zhao Shundong. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

使用说明:

1. 为保证本文的完整性和可用性，本文遵循 GFDL 协议。
2. 可以在 <http://www.unixhot.com/pdf/cluster.pdf> 找到本文的最新版本。
3. 本文仅供参考使用，不承担任何因文档错误而造成的任何损失。
4. 有任何问题可以在 UnixHot 运维社区讨论交流。
5. 有相关问题或业务合作。请邮件至 admin@unixhot.com。

相关文档:

- | | |
|-----------------------------|---|
| 1. 《Linux 系统运维之系统架构》 | http://www.unixhot.com/pdf/cluster.pdf |
| 2. 《Linux 系统运维之系统管理》 | http://www.unixhot.com/pdf/admin.pdf |
| 3. 《Linux 系统运维之运维监控》 | http://www.unixhot.com/pdf/monitor.pdf |
| 4. 《Linux 系统运维之性能优化》 | http://www.unixhot.com/pdf/tuning.pdf |
| 5. 《Linux 系统运维之运维规范》 | http://www.unixhot.com/pdf/guifan.pdf |
| 6. 《Linux 系统运维之 MySQL DBA》 | http://www.unixhot.com/pdf/mysql.pdf |
| 7. 《Linux 系统运维之 Oracle DBA》 | http://www.unixhot.com/pdf/oracle.pdf |

修订历史记录

[illegible]

内容简介

本文通过生产应用实例，从运维工程师的角度对目前流行的 Web 架构做了实验性的讲解，该系列的文档属于手册类型，可以直接在生产环境部署运行。本文讲解的方案都是在互联网公司应用比较成熟，也比较通用的方案。如 LAMP、LNMP、LVS Keepalived、Apache+Tomcat 负载均衡和 Nginx+Tomcat 负载均衡等。

目录 (Contents)

第 1 章 系统架构概述

- 1.1 Web 应用架构
- 1.2 JSP 网站概述
- 1.3 PHP 网站概述
- 1.4 什么是集群
- 1.5 集群的主要类型

第 2 章 LAMP 应用

- 2.1 Apache 服务器简介
- 2.2 Apache MPM 原理和设置
- 2.3 源码安装 LAMP

第 3 章 LNMP 应用

- 3.1 Nginx 安装
- 3.2 MySQL 客户端安装
- 3.3 PHP 扩展模块安装
- 3.4 PHP FastCGI 模式安装
- 3.5 启动测试

第 4 章 集群中的文件共享

4.1 DAS、NAS 和 SAN

4.2 inotify+rsync 触发式同步数据

4.3 Sersync 部署

第 5 章 负载均衡中的 Session 解决

5.1 负载均衡中 Session 的问题

5.2 负载均衡中 Session 的解决方法

5.3 Nginx 做负载均衡 Session 解决

5.4 Apache Tomcat 负载均衡 Session 解决

第 6 章 Apache Tomcat 负载均衡

第 7 章 Nginx Tomcat 负载均衡

第 8 章 LB 负载均衡集群

2.1 LVS 简介

2.2 LVS-NAT 方式部署

2.3 LVS-DR 方式部署

第 9 章 LVS Keepalived 集群

5.1 LVS Keepalived 集群简介

5.2 部署 LVS Keepalived

5.3 LVS 配置

5.4 Keepalived 配置

5.5 LVS Keepalived 集群测试

第 1 章 系统架构概述

1.1 Web 应用架构

作为跨世纪青年，我们处于云计算和 Web 2.0 的时代，搜索、视频、SNS、微博等 WEB 应用扑面而来。根据美国知名 IT 产业分析机构 IDC 的白皮书表述：以 Blog、Wiki、Tagging 和 SNS 四类应用为代表的 Web2.0 趋势在中国互联网市场中引起了广泛的反响。在加上目前琳琅满目的电子商务网站，团购网站。太多太多的问题等待着有人去解决。不同的 Web 脚本语言，不同的架构让很多架构师和运维工程师不知到如何入手。

1.2 JSP 网站概述

1.2.1 相关名词解释

1> J2EE 、J2SE 、J2ME 三者的区别

J2EE 是 Java 2 enterprise edition 是 Java 的一种企业版用于企业级的应用服务开发

J2SE 是 Java 2 standard edition 是 Java 的标准版,用于标准的应用开发

J2ME 是 Java 2 Micro Edition 是 Java 的微型版,常用于手机上的开发

J2EE, J2SE, J2ME 是 java 针对不同的的使用来提供不同的服务，也就是提供不同类型的类库。

针对企业网应用的 J2EE (Java 2 Enterprise Edition)、针对普通 PC 应用的 J2SE (Java 2 Standard Edition) 和针对嵌入式设备及消费类电器的 J2ME (Java 2 Micro Edition) 三个版本

2> JDK、JRE、SDK 三者的区别

JDK Java 开发工具包, Java Development Kit 里面有运行环境 JRE 和开发时所需要的 Java 类库，以及一些编译调试运行的程序，如 java.exe, javac.exe, javaw.exe 等。

JRE Java 运行环境 Java Runtime Environment. 这个仅仅只是 Java 运行的环境，没有 Java 无法运行，一般 JRE 被包括在 JDK 中，也可以单独装一个独立的 JRE。

SDK 是一些公司针对某一项技术为软件开发人员制作的一套辅助开发或者减少开发周期的工具. 比方你用的 Eclipse 就是 Java 的 SDK, 它全称就是 Eclipse SDK.

Java 世界中，只有独一无二的一套 JDK。J2ME, J2EE 都是在这套 JDK 基础上的扩展。

1.2.2 JSP 应用服务器

运行 JSP 应用的中间件主要有 Tomcat、JBOSS、Weblogic、WebSphere。

1.3 PHP 网站概述

在近几年的编程语言排名上，PHP 始终在第 4 名左右徘徊。PHP 站点到处可见，而且很多我们每天都登陆的

1.3.1 PHP 常用开源框架

1>ThinkPHP（本土优秀的开源 PHP 框架） <http://thinkphp.cn/>

2>Zen-Cart（对外贸易的顶级开源框架） <http://zen-cart.com/>

3>Zend Framework（做 P H P 的都知道） <http://framework.zend.com/>

4>Smarty（学习 MVC 都用过） <http://www.smarty.net/>

1.4 什么是集群

将多台同构或异构的计算机连接起来协同完成特定的任务就构成了集群系统。

1.5 集群的主要类型

1.5.1 HA（High Availability）

高可用性集群的主要功能就是提供不间断的服务。有许多应用程序都必须一天二十四小时地不停运转，如所有的 web 服务器、工业控制器、ATM、远程通讯转接器、医学与军事监测仪以及股票处理机等。对这些应用程序而言，暂时的停机都会导致数据的丢失和灾难性的后果。

HA 集群通过特殊软件把独立的系统连接起来，组成一个能够提供故障切换功能的集群，HA 集群可以保证在多种故障中，关键服务的可用性、可靠性及数据完整性，HA 集群主要用于文件服务、WEB 服务，数据库服务等关键应用中。

HA 集群的开源项目：Heartbeat 详情见官方网站：<http://www.linux-ha.org>

1.5.2 LB (Load Balancing)

负载均衡集群，在 LB 服务器上使用专门的路由算法，将数据包分散到多个真实服务器中进行处理，从而达到网络服务均衡负载的作用。

LB 集群的开源项目：LVS 详情见官方网站：<http://www.linuxvirtualserver.org>

1.5.3 HPC (High performance Computing)

DC (Distributed Computing)

PC (Parallel Computing)

高性能集群通过将多台机器连接起来同时处理复杂的计算问题。模拟星球附近的磁场、预测龙卷风的出现、定位石油资源的储藏地等情况都需要对大量的数据进行处理。传统的处理方法是使用超级计算机来完成计算工作，但是超级计算机的价格比较昂贵，而且可用性和可扩展性不够强，因此集群成为了高性能计算领域瞩目的焦点。

分布式高性能计算 (DC) OpenMosix OpenSSI

并行式高性能计算 (PC) Beowulf

第 2 章 LAMP 应用

2.1 Apache 服务器简介

2.2 Apache MPM 原理和设置

在我们编译安装 Apache 之前，要考虑的是让 Apache 在什么样的模式下运行，因为从 Apache 2.0 就加入了 MPM (Multi-Processing Modules, 多道处理模块)。

(摘自网络，很经典)

Apache 2.0 在性能上的改善最吸引人。在支持 POSIX 线程的 Unix 系统上，Apache 可以通过不同的 MPM 运行在一种多进程与多线程相混合的模式下，增强部分配置的可扩充性能。相比于 Apache 1.3, 2.0 版本做了大量的优化来提升处理能力和可伸缩性，并且大多数改进在默认状态下即可生效。但是在编译和运行时刻，2.0 也有许多可以显著提高性能的选择。本文不想叙述那些以功能换取速

度的指令，如 HostnameLookups 等，而只是说明在 2.0 中影响性能的最核心特性：MPM（Multi-Processing Modules，多道处理模块）的基本工作原理和配置指令。

毫不夸张地说，MPM 的引入是 Apache 2.0 最重要的变化。大家知道，Apache 是基于模块化的设计，而 Apache 2.0 更扩展了模块化设计到 Web 服务器的最基本功能。服务器装载了一种多道处理模块，负责绑定本机网络端口、接受请求，并调度子进程来处理请求。扩展模块化设计有两个重要好处：

- ◆ Apache 可以更简洁、有效地支持多种操作系统；
- ◆ 服务器可以按站点的特殊需要进行定制。

在用户级，MPM 看起来和其它 Apache 模块非常类似。主要区别是在任意时刻只能有一种 MPM 被装载到服务器中。

2.2.1 prefork 的工作原理及配置

如果不用“--with-mpm”显式指定某种 MPM，prefork 就是 Unix 平台上缺省的 MPM。它所采用的预派生子进程方式也是 Apache 1.3 中采用的模式。prefork 本身并没有使用到线程，2.0 版使用它是为了与 1.3 版保持兼容性；另一方面，prefork 用单独的子进程来处理不同的请求，进程之间是彼此独立的，这也使其成为最稳定的 MPM 之一。

若使用 prefork，在 make 编译和 make install 安装后，使用“httpd -l”来确定当前使用的 MPM，应该会看到 prefork.c（如果看到 worker.c 说明使用的是 worker MPM，依此类推）。再查看缺省生成的 httpd.conf 配置文件，里面包含如下配置段：<IfModule prefork.c>

```
StartServers          5
MinSpareServers       5
MaxSpareServers       10
MaxClients             150
MaxRequestsPerChild   0

</IfModule>
```

prefork 的工作原理是，控制进程在最初建立“StartServers”个子进程后，为了满足 MinSpareServers 设置的需创建一个进程，等待一秒钟，继续创建两个，再等待一秒钟，继续创建四个……如此按指数级增加创建的进程数，最多达到每秒 32 个，直到满足 MinSpareServers 设置的值为止。这就是预派生（prefork）的由来。这种模式可以不必在请求到来时再产生新的进程，

从而减小了系统开销 以增加性能。

MaxSpareServers 设置了最大的空闲进程数，如果空闲进程数大于这个值，Apache 会自动 kill 掉一些多余进程。这个值不要设 得过大，但如果设的值比 MinSpareServers 小，Apache 会自动把其调整为 MinSpareServers+1。如果站点负载较大，可考虑 同时加大 MinSpareServers 和 MaxSpareServers。

MaxRequestsPerChild 设置的是每个子进程可处理的请求数。每个子进程在处理了 “MaxRequestsPerChild” 个 请求后将自动销毁。0 意味着无限，即子进程永不销毁。虽然缺省设为 0 可以使每个子进程处理更多的请求，但如果设成非零值也有两点重要的好处：

- ◆ 可防止意外的内存泄漏；
- ◆ 在服务器负载下降的时候会自动减少子进程数。

因此，可根据服务器的负载来调整这个值。笔者认为 10000 左右比较合适。

MaxClients 是这些指令中最为重要的一个，设定的是 Apache 可以同时处理的请求，是对 Apache 性能影响最大的参数。其缺省值 150 是远远不够的，如果请求总数已达到这个值（可通过 `ps -ef|grep http|wc -l` 来确认），那么后面的请求就要排队，直到某个已处理请求完毕。这就是系统资源还剩下很多而 HTTP 访问却很慢的主要原因。系统管理员可以根据硬件配置 和负载情况来动态调整这个值。虽然理论上这个值越大，可以处理的请求就越多，但 Apache 默认的限制不能大于 256。如果把这个值设为大于 256，那么 Apache 将无法起动。事实上，256 对于负载稍重的站点也是不够的。在 Apache 1.3 中，这是个硬限制。如果要加大这个值，必须在 “configure” 前手工修改的源代码树下的 `src/include/httpd.h` 中查找 256，就会发现 “`#define HARD_SERVER_LIMIT 256`” 这行。把 256 改为要增大的值（如 4000），然后重新编译 Apache 即可。在 Apache 2.0 中新加入了 `ServerLimit` 指令，使得无须重编译 Apache 就可以加大 MaxClients。下面是笔者的 `prefork` 配置 段：

```
<IfModule prefork.c>

    StartServers      10
    MinSpareServers   10
    MaxSpareServers   15
    ServerLimit       2000
    MaxClients        1000
    MaxRequestsPerChild 10000

</IfModule>
```

上述配置中，ServerLimit 的最大值是 20000，对于大多数站点已经足够。如果一定要再加大这个数值，对位于源代码树下 server/mpm/prefork/prefork.c 中以下两行做相应修改即可：

```
#define DEFAULT_SERVER_LIMIT 256#define MAX_SERVER_LIMIT 20000
```

2.2.2 worker 的工作原理及配置

相对于 prefork，worker 是 2.0 版中全新的支持多线程和多进程混合模型的 MPM。由于使用线程来处理，所以可以处理相对海量的请求，而系统资源的开销要小于基于进程的服务器。但是，worker 也使用了多进程，每个进程又生成多个线程，以获得基于进程服务器的稳定性。这种 MPM 的工作方式将是 Apache 2.0 的发展趋势。

在 configure -with-mpm=worker 后，进行 make 编译、make install 安装。在缺省生成的 httpd.conf 中有以下配置段：<IfModule worker.c>

```
StartServers          2
MaxClients            150
MinSpareThreads       25
MaxSpareThreads       75
ThreadsPerChild       25
MaxRequestsPerChild   0

</IfModule>
```

worker 的工作原理是，由主控制进程生成“StartServers”个子进程，每个子进程中包含固定的 ThreadsPerChild 线程数，各个线程独立地处理请求。同样，为了不在请求到来时再生成线程，MinSpareThreads 和 MaxSpareThreads 设置了最少和最多的空闲线程数；而 MaxClients 设置了所有子进程中的线程总数。如果现有子进程中的线程总数不能满足负载，控制进程将派生新的子进程。

MinSpareThreads 和 MaxSpareThreads 的最大缺省值分别是 75 和 250。这两个参数对 Apache 的性能影响并不大，可以按照实际情况相应调节。

ThreadsPerChild 是 worker MPM 中与性能相关最密切的指令。ThreadsPerChild 的最大缺省值是 64，如果负载较大，64 也是不够的。这时要显式使用 ThreadLimit 指令，它的最大缺省值是 20000。上述两个值位于源码树 server/mpm/worker/worker.c 中的以下两行：#define DEFAULT_THREAD_LIMIT 64

```
#define MAX_THREAD_LIMIT 20000
```

这两行对应着 `ThreadsPerChild` 和 `ThreadLimit` 的限制数。最好在 `configure` 之前就把 64 改成所希望的值。注意，不要把这两个值设得太高，超过系统的处理能力，从而因 Apache 不起动使系统很不稳定。

Worker 模式下所能同时处理的请求总数是由子进程总数乘以 `ThreadsPerChild` 值决定的，应该大于等于 `MaxClients`。如果负载很大，现有的子进程数不能满足时，控制进程会派生新的子进程。默认最大的子进程总数是 16，加大时也需要显式声明 `ServerLimit`（最大值是 20000）。这两个值位于源码树 `server/mpm/worker/worker.c` 中的以下两行：`#define DEFAULT_SERVER_LIMIT 16`

```
#define MAX_SERVER_LIMIT 20000
```

需要注意的是，如果显式声明了 `ServerLimit`，那么它乘以 `ThreadsPerChild` 的值必须大于等于 `MaxClients`，而且 `MaxClients` 必须是 `ThreadsPerChild` 的整数倍，否则 Apache 将会自动调节到一个相应值（可能是个非期望值）。下面是笔者的 `worker` 配置段：`<IfModule worker.c>`

```
StartServers      3
MaxClients        2000
ServerLimit       25
MinSpareThreads   50
MaxSpareThreads   200
ThreadLimit       200
ThreadsPerChild   100
MaxRequestsPerChild 0
</IfModule>
```

通过上面的叙述，可以了解到 Apache 2.0 中 `prefork` 和 `worker` 这两个重要 MPM 的工作原理，并可根据实际情况来配置 Apache 相关的核心参数，以获得最大的性能和稳定性。

2.3 源码安装 LAMP

```
[root@Apache-Server src]# chmod +x httpd-2.2.13.tar.gz
[root@Apache-Server src]# tar zxvf httpd-2.2.13.tar.gz
[root@Apache-Server src]# cd httpd-2.2.13
[root@Apache-Server httpd-2.2.13]# ./configure --prefix=/vol1/apache \
--with-mpm=prefork \
--enable-so --enable-modules="all" \
--enable-mods-shared="all"
[root@Apache-Server httpd-2.2.13]# make && make install
[root@Apache-Server httpd-2.2.13]# cp support/apachectl /etc/init.d/httpd
[root@Apache-Server httpd-2.2.13]# chmod +x /etc/init.d/httpd
```

第3章 LNMP 应用

3.1 Nginx 安装

3.1.1 安装 PCRE 软件包

pcre (Perl Compatible Regular Expressions) 包括 perl 兼容的正规表达式库.这些在执行正规表达式模式匹配时用与 Perl 5 同样的语法和语义是很有用的.

```
[root@Web-Server ~]# cd /usr/local/src/
[root@Web-Server src]# wget
http://cdnetworks-kr-1.dl.sourceforge.net/project/pcre/pcre/8.10/pcre-8.10.tar.gz
[root@Web-Server src]# tar zxvf pcre-8.10.tar.gz
[root@Web-Server src]# cd pcre-8.10
[root@Web-Server pcre-8.10]# ./configure
[root@Web-Server pcre-8.10]# make && make install
```

3.1.2 安装 Nginx 软件包。

```
[root@Web-Server ~]# cd /usr/local/src
[root@Web-Server src]# wget http://nginx.org/download/nginx-0.8.54.tar.gz
[root@Web-Server src]# tar zxvf nginx-0.8.54.tar.gz
```

```
[root@Web-Server src]# cd nginx-0.8.54
```

如果你了解该版本的一些信息，请查看 CHANGES 文件。

```
[root@Web-Server nginx-0.8.54]# useradd -s /sbin/nologin -M www
```

```
[root@Web-Server nginx-0.8.54]# ./configure --prefix=/usr/local/nginx \
```

```
--user=www --group=www \
```

```
--with-http_stub_status_module --with-file-aio \
```

```
--with-http_ssl_module \
```

```
[root@Web-Server nginx-0.8.54]# make && make install
```

3.1.3 Nginx 配置文件和启动

```
[root@Web-Server ~]# /usr/local/nginx/sbin/nginx -t
```

the configuration file /usr/local/nginx/conf/nginx.conf syntax is ok

configuration file /usr/local/nginx/conf/nginx.conf test is successful

```
[root@Web-Server ~]# /usr/local/nginx/sbin/nginx （启动）
```

```
[root@Web-Server ~]# /usr/local/nginx/sbin/nginx -s reload （重新加载）
```

3.2 MySQL 客户端安装

```
[root@Web-Server ~]# cd /usr/local/src
```

```
[root@Web-Server src]# wget http://mysql.cdpa.nsysu.edu.tw/Downloads/MySQL-5.1/mysql-5.1.56.tar.gz
```

```
[root@Web-Server src]# tar zxvf mysql-5.1.56.tar.gz
```

```
[root@Web-Server src]# cd mysql-5.1.56
```

```
[root@Web-Server mysql-5.1.56]# ./configure --prefix=/usr/local/mysql \
```

```
--localstatedir=/data/mysql --enable-asm \
```

```
--with-client-ldflags=-all-static --with-mysqld-ldflags=-all-static \
```

```
--with-pthread --enable-static --with-big-tables --without-ndb-debug \
```

```
--with-charset=utf8 --with-extra-charsets=all \
```

```
--without-debug --enable-thread-safe-client --enable-local-infile --with-plugins=max
```

```
[root@Web-Server mysql-5.1.56]# make && make install
```

```
[root@Web-Server mysql-5.1.56]# groupadd mysql
```

```
[root@Web-Server mysql-5.1.56]# useradd -s /sbin/nologin -M -g mysql mysql
```

```
[root@Web-Server mysql-5.1.56]# chown -R root:mysql /usr/local/mysql/
```

3.3 PHP 扩展模块安装

```
[root@web-node1 ~]# cd /usr/local/src
```

```
[root@web-node1 src]# wget http://ftp.gnu.org/pub/gnu/libiconv/libiconv-1.13.1.tar.gz
```

如银联在线的网上支付接口就需要 mcrypt 和 bcmath 和 curl 三个 PHP 扩展库的支持,

```
[root@web-node1 src]# wget
```

```
http://cdnetworks-kr-2.dl.sourceforge.net/project/mcrypt/Libmcrypt/2.5.8/libmcrypt-2.5.8.tar.gz
```

```
[root@web-node1 src]# wget http://ftp.gnu.org/pub/gnu/libiconv/libiconv-1.13.1.tar.gz
```

```
[root@web-node1 src]# wget
```

```
http://cdnetworks-kr-1.dl.sourceforge.net/project/mcrypt/MCrypt/2.6.8/mcrypt-2.6.8.tar.gz
```

```
[root@web-node1 src]# wget
```

```
http://cdnetworks-kr-1.dl.sourceforge.net/project/mhash/mhash/0.9.9.9/mhash-0.9.9.9.tar.gz
```

```
[root@web-node1 src]# tar zxvf libiconv-1.13.1.tar.gz
```

```
[root@web-node1 src]# cd libiconv-1.13.1
```

```
[root@web-node1 libiconv-1.13.1]# ./configure --prefix=/usr/local
```

```
[root@web-node1 libiconv-1.13.1]# make && make install
```

```
[root@web-node1 src]# tar zxvf libmcrypt-2.5.8.tar.gz
```

```
[root@web-node1 src]# cd libmcrypt-2.5.8
```

```
[root@web-node1 libmcrypt-2.5.8]# ./configure && make && make install
```

```
[root@web-node1 libmcrypt-2.5.8]# ldconfig
```

```
[root@web-node1 libmcrypt-2.5.8]# cd libltdl/
```

```
[root@web-node1 libltdl]# ./configure --enable-ltdl-install
```

```
[root@web-node1 libltdl]# make && make install
```

```
[root@web-node1 src]# cd mhash-0.9.9.9
```

```
[root@web-node1 mhash-0.9.9.9]# ./configure && make && make install

[root@web-node1 lib]# ln -s /usr/local/lib/libmcrypt.la /usr/lib/libmcrypt.la
[root@web-node1 lib]# ln -s /usr/local/lib/libmcrypt.so /usr/lib/libmcrypt.so
[root@web-node1 lib]# ln -s /usr/local/lib/libmcrypt.so.4 /usr/lib/libmcrypt.so.4
[root@web-node1 lib]# ln -s /usr/local/lib/libmcrypt.so.4.4.8 /usr/lib/libmcrypt.so.4.4.8
[root@web-node1 lib]# ln -s /usr/local/lib/libmhash.a /usr/lib/libmhash.a
[root@web-node1 lib]# ln -s /usr/local/lib/libmhash.la /usr/lib/libmhash.la
[root@web-node1 lib]# ln -s /usr/local/lib/libmhash.so /usr/lib/libmhash.so
[root@web-node1 lib]# ln -s /usr/local/lib/libmhash.so.2 /usr/lib/libmhash.so.2
[root@web-node1 lib]# ln -s /usr/local/lib/libmhash.so.2.0.1 /usr/lib/libmhash.so.2.0.1
[root@web-node1 lib]# ln -s /usr/local/bin/libmcrypt-config /usr/bin/libmcrypt-config

[root@web-node1 src]# tar zxvf mcrypt-2.6.8.tar.gz
[root@web-node1 src]# cd mcrypt-2.6.8
[root@web-node1 mcrypt-2.6.8]# ldconfig
[root@web-node1 mcrypt-2.6.8]# ./configure && make && make install
```

3.4 PHP FastCGI 模式安装

```
[root@Web-Server src]# wget http://cn.php.net/distributions/php-5.3.5.tar.gz
[root@Web-Server src]# tar zxvf php-5.3.5.tar.gz
[root@Web-Server src]# cd php-5.3.5
./configure --prefix=/usr/local/php \
--with-config-file-path=/usr/local/php/etc \
--with-mysql=/usr/local/mysql \
--with-mysqli=/usr/local/mysql/bin/mysqli_config \
--with-freetype-dir --with-jpeg-dir --with-png-dir --with-iconv-dir=/usr/local \
--with-zlib --with-libxml-dir=/usr --enable-xml \
```



```
--with-curl --with-curlwrappers --enable-sqlite-utf8 \  
--with-pdo-mysql=/usr/local/mysql \  
--enable-bcmath --enable-shmop --enable-sysvsem \  
--enable-inline-optimization \  
--enable-mbregex --with-openssl --enable-pcntl \  
--enable-fpm --with-fpm-user=www --with-fpm-group=www \  
--enable-mbstring --with-gd --enable-gd-native-ttf \  
--enable-sockets --with-xmlrpc --enable-zip --enable-soap \  
--disable-debug --enable-safe-mode --with-mcrypt --with-mhash
```

提示：如果编译出错，请根据错误提示安装相应的软件包支持，可以让你更熟悉 PHP 的运行细节，如果不想那么麻烦，那只好使用下面的 yum 安装了，再次建议您自己解决问题：

```
yum -y install gcc gcc-c++ autoconf libjpeg libjpeg-devel libpng libpng-devel freetype  
freetype-devel libxml2 libxml2-devel zlib zlib-devel glibc glibc-devel glib2 glib2-devel  
bzip2 bzip2-devel ncurses ncurses-devel curl curl-devel e2fsprogs e2fsprogs-devel krb5  
krb5-devel libidn libidn-devel openssl openssl-devel openldap openldap-devel nss_ldap  
openldap-clients openldap-servers  
[root@Web-Server php-5.3.5]# make && make install  
[root@Web-Server php-5.3.5]# cp php.ini-production /usr/local/php/etc/php.ini  
[root@Web-Server php-5.3.5]# cd /usr/local/php/etc/  
[root@Web-Server etc]# cp php-fpm.conf.default php-fpm.conf
```

3.5 启动测试

```
[root@Web-Server ~]# /usr/local/nginx/sbin/nginx  
[root@Web-Server ~]# /usr/local/php/sbin/php-fpm
```

第 4 章 集群中的文件共享

在第 1 章我们已经了解集群的基本技术和原理，在第 2 章和第 3 章中熟悉了目前应用比较广泛的 LAMP 和 LNMP 应用。在进入集群的架构之前，我们需要先考虑一下，在集群中可能存在的问题，

并提前了解解决方案，然后才能继续进行。如集群节点中文件如何共享，Session 如何同步等。

4.1 DAS、NAS 和 SAN

目前企业存储应用的体系结构主要有 DAS、NAS 和 SAN 三种模式。从企业发展和 Web 从单点到集群的发展就需要在 DAS、NAS 和 SAN 三种存储方案进行不同的选择。具体三者之间的丝丝连连读者可以通过网络搜索，我推荐 IT168 这篇经典文章：

<http://publish.it168.com/2004/0819/20040819005701.shtml>。

4.2 inotify+rsync 触发式同步数据

在这里主要讲解使用文件同步的方式，进行集群中文件的异步复制操作，rsync 当然是首选的目标，但是 rsync 在同步文件时需要分析目录中所有文件的更新标记，每次都需要全部扫描，对于文件很多，或者目录级别很多，即使没有文件需要同步，rsync 的检查时间也会很长。

对于 2.6.13 以后的 Linux 内核，已经加入了 inotify 模块，它是内核支持的一种，可以在后台监控文件系统的改变，可以在任何文件的修改时间发生变化后，发出通知。所以我们使用 inotify 加 rsync 的方式实现实时的文件同步，我们通常称之为触发式同步。

4.2.1 inotify 简介

4.2.2 inotify 可以监视的文件系统事件包括：

IN_ACCESS，即文件被访问

IN_MODIFY，文件被 write

IN_ATTRIB，文件属性被修改，如 chmod、chown、touch 等

IN_CLOSE_WRITE，可写文件被 close

IN_CLOSE_NOWRITE，不可写文件被 close

IN_OPEN，文件被 open

IN_MOVED_FROM，文件被移走,如 mv

IN_MOVED_TO, 文件被移来, 如 mv、cp

IN_CREATE, 创建新文件

IN_DELETE, 文件被删除, 如 rm

IN_DELETE_SELF, 自删除, 即一个可执行文件在执行时删除自己

IN_MOVE_SELF, 自移动, 即一个可执行文件在执行时移动自己

IN_UNMOUNT, 宿主文件系统被 umount

IN_CLOSE, 文件被关闭, 等同于(IN_CLOSE_WRITE | IN_CLOSE_NOWRITE)

IN_MOVE, 文件被移动, 等同于(IN_MOVED_FROM | IN_MOVED_TO)

4.2.3 内核是否支持

```
[root@web-node1 ~]# uname -r
```

2.6.18-194.el5 （查看你的内核版本）

```
[root@web-node1 ~]# ls /proc/sys/fs/inotify/ （如果该虚拟目录存在说明你的内核已支持）
```

max_queued_events max_user_instances max_user_watches

4.3 Sersync 部署

当然你可以使用 inotify 脚本和 rsync 开始进行分发和同步, 但是, 我们有更好的选择。这个选择就是 sersync, sersync 是使用 C++编写的, 它可以使用多线程进行同步, 尤其在同步较大文件时, 能够保证多个服务器实时保持同步状态。更过优势请访问项目首页: <http://code.google.com/p/sersync/>。

4.3.1 rsync 安装

需要在所有目标服务器中部署 rsync, 并启动 rsync。

```
[root@web-node2 ~]# cd /usr/local/src
```

```
[root@web-node2 src]# wget http://rsync.samba.org/ftp/rsync/src/rsync-3.0.8.tar.gz
```

```
[root@web-node2 src]# tar zxvf rsync-3.0.8
```

```
[root@web-node2 src]# cd rsync-3.0.8
```

```
[root@web-node2 rsync-3.0.8]# ./configure --prefix=/usr/local/rsync && make && make install
```

```
[root@web-node2 ~]# mkdir /etc/rsyncd/
```

```
[root@web-node2 ~]# cat /etc/rsyncd/rsyncd.conf
```

```
uid=root
gid=root
max connections=36000
use chroot=no
log file=/var/log/rsyncd.log
pid file=/var/run/rsyncd.pid
lock file=/var/run/rsyncd.lock
```

```
[shop]
path=/data/shop
comment = shop.unixhot.com
ignore errors = yes
read only = no
hosts allow = 172.16.1.0/24
hosts deny = *
```

在所有目标服务器上开启 rsync 监听。

```
[root@web-node2 ~]# /usr/local/rsync/bin/rsync --daemon --config=/etc/rsyncd/rsyncd.conf
```

4.3.2 sersync 安装

```
[root@web-node1 src]# wget
http://sersync.googlecode.com/files/sersync2.5\_64bit\_binary\_stable\_final.tar.gz
[root@web-node1 src]# tar zxvf sersync2.5_64bit_binary_stable_final.tar.gz
[root@web-node1 src]# mv GNU-Linux-x86/ /etc/sersync
```

4.4.3 sersync 配置

```
[root@web-node1 ~]# cat /etc/sersync/confxml.xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<head version="2.5">
    <host hostip="localhost" port="8008"></host>
```

```
<debug start="false"/>
<fileSystem xfs="false"/>
<filter start="false">
  <exclude expression="(.*).svn"></exclude>
  <exclude expression="(.*).gz"></exclude>
  <exclude expression="^info/*"></exclude>
  <exclude expression="^static/*"></exclude>
</filter>
<inotify>
  <delete start="true"/>
  <createFolder start="true"/>
  <createFile start="false"/>
  <closeWrite start="true"/>
  <moveFrom start="true"/>
  <moveTo start="true"/>
  <attrib start="false"/>
  <modify start="false"/>
</inotify>

<sersync>
  <localpath watch="/data/shop">
    <remote ip="172.16.1.12" name="shop"/>
    <remote ip="172.16.1.14" name="shop"/>
  </localpath>
<rsync>
  <commonParams params="-artuz"/>
  <auth start="false" users="root" passwordfile="/etc/rsync.pas"/>
  <userDefinedPort start="false" port="874"/><!-- port=874 -->
  <timeout start="false" time="100"/><!-- timeout=100 -->
  <ssh start="false"/>
```

```

</rsync>

<failLog path="/etc/sersync/rsync_fail_log.sh" timeToExecute="60"/><!--default every 60mins
execute once-->

<crontab start="false" schedule="600"><!--600mins-->

    <crontabfilter start="false">

        <exclude expression="*.php"></exclude>

        <exclude expression="info/*"></exclude>

    </crontabfilter>

</crontab>

<plugin start="false" name="command"/>

</sersync>

<plugin name="command">

<param prefix="/bin/sh" suffix="" ignoreError="true"/>    <!--prefix /opt/tongbu/mmm.sh suffix-->

<filter start="false">

    <include expression="(.)\.php"/>

    <include expression="(.)\.sh"/>

</filter>

</plugin>

<plugin name="socket">

<localpath watch="/opt/tongbu">

    <deshost ip="192.168.138.20" port="8009"/>

</localpath>

</plugin>

<plugin name="refreshCDN">

<localpath watch="/data0/htdocs/cms.xoyo.com/site/">

    <cdninfo domainname="ccms.chinacache.com" port="80" username="xxxx" passwd="xxxx"/>

    <sendurl base="http://pic.xoyo.com/cms"/>

    <regexurl regex="false" match="cms.xoyo.com/site([a-zA-Z0-9]*).xoyo.com/images"/>

```

```
</localpath>  
</plugin>  
</head>
```

4.4.4 sersync 启动

```
[root@web-node1 sersync]# ./sersync2 -h
```

set the system param

```
execute: echo 50000000 > /proc/sys/fs/inotify/max_user_watches
```

```
execute: echo 327679 > /proc/sys/fs/inotify/max_queued_events
```

parse the command param

参数-d:启用守护进程模式

参数-r:在监控前，将监控目录与远程主机用 rsync 命令推送一遍

参数-n: 指定开启守护线程的数量，默认为 10 个

参数-o:指定配置文件，默认使用 confxml.xml 文件

参数-m:单独启用其他模块，使用 -m refreshCDN 开启刷新 CDN 模块

参数-m:单独启用其他模块，使用 -m socket 开启 socket 模块

参数-m:单独启用其他模块，使用 -m http 开启 http 模块

不加-m 参数，则默认执行同步程序

```
[root@web-node1 ~]# /etc/sersync/sersync2 -r -d （启动服务，如果测试有问题，可以打开 debug 查看）
```

第 5 章 负载均衡中的 Session 解决

注意：对于 cookie 和 session 的介绍，和他们之间的恩恩怨怨，还请查询有关文档。

5.1 负载均衡中 Session 的问题

无论是 PHP 还是 Java，只要使用服务器保存 Session，在做负载均衡时都需要考虑 Session 的问题。从用户端来解释，就是当一个用户第一次访问通过后端的一台服务器登录，当用户再次发送请求时，被分发器分发到后端不同的服务器中，由于这台服务器没有用户的登录信息，所以导致用户

需要重新登录。这对用户来说是不可忍受的。所以，在实施负载均衡的时候，我们必须考虑 Session 的问题。

5.2 负载均衡中 Session 的解决方法

针对 Session 的处理，我们一般有以下几种方法：

- 1.将 Session 写入客户端浏览器的 Cookies 里面。
- 2.Session 保持，保证每个客户端固定访问后端的同一台应用节点服务器。
- 3.Session 复制，将每个应用服务器中的 Session 信息复制到其它服务器节点上。
- 4.Session 共享，将 Session 统一存放，程序在同一个地方读取。

5.3 Nginx 做负载均衡 Session 解决

对于 Nginx 可以选作用 Session 保持的方法实行负载均衡，nginx 的 upstream 目前支持5种方式的分配方式，期中有两种比较通用的 Session 解决方法，ip_hash 和 url_hash。

5.3.1 ip_hash

每个请求按访问 ip 的 hash 结果分配，这样每个访客固定访问一个后端服务器，达到了 Session 保持的方法。

例：

```
upstream bakend {  
    ip_hash;  
    server 192.168.0.11:80;  
    server 192.168.0.12:80;  
}
```

5.3.2 url_hash

按访问 url 的 hash 结果来分配请求，使每个 url 定向到同一个后端服务器，后端服务器为缓存时比较有效。

例：在 upstream 中加入 hash 语句，server 语句中不能写入 weight 等其他的参数，hash_method 是使用的 hash 算法

```
upstream backend {  
    server squid1:3128;  
    server squid2:3128;  
    hash $request_uri;  
    hash_method crc32;  
}
```

5.4 Apache Tomcat 负载均衡 Session 解决

第 6 章 Apache Tomcat 负载均衡

6.1 Apache Tomcat 负载均衡简介

6.2 Apache Tomcat 简介

6.2.1 集群架构图

6.2.2 集群环境

集群节点	节点主机名	节点 IP 地址	节点说明
Apache-Server	Apache-Server	192.168.140.136	Apache 和 JK 分发器
Tomcat-Node1	Tomcat-Node1	192.168.140.137	Tomcat 节点 1
Tomcat-Node2	Tomcat-Node2	192.168.140.139	Tomcat 节点 2

第 7 章 Nginx Tomcat 负载均衡

Nginx Tomcat 负载均衡已经在越来越多的方案中脱颖而出，在高并发的情况下表现的要远远超过 Apache 和 Tomcat 负载均衡的方案，Nginx 使用 upstream 把用户请求分发到后端的应用服务器上，而且还支持“权重”的分发方式，主要的修改 Nginx 的配置文件完成。

7.1 Nginx 做前端分发器

Nginx 的安装配置见以上章节，这里主要给出配置文件。

```
user    website website;

worker_processes  4;


error_log  logs/error.log;

pid        logs/nginx.pid;

worker_rlimit_nofile 65535;


events {
    use epoll;
    worker_connections  10240;
}


http {
    include        mime.types;
    default_type  application/octet-stream;

    server_names_hash_bucket_size 128;

    client_header_buffer_size 32k;
    large_client_header_buffers 4 32k;
    client_max_body_size 8m;

    sendfile on;
```

```
tcp_nopush      on;
keepalive_timeout 60;
tcp_nodelay on;
```

```
gzip on;
gzip_min_length 1k;
gzip_buffers     4 16k;
gzip_http_version 1.0;
gzip_comp_level 2;
gzip_types       text/plain application/x-javascript text/css application/xml;
gzip_vary on;
```

```
server_tokens off;
```

```
upstream web #设置 web 集群池
{
ip_hash; #
server 192.168.0.141:8080;
server 192.168.0.142:8080;
server 192.168.0.143:8080;
server 192.168.0.144:8080;
server 192.168.0.145:8080;
server 192.168.0.146:8080;

}
```

```
upstream wap #设置 wap 集群池
{
ip_hash;
```

```
server 192.168.0.151:8080;  
server 192.168.0.152:8080;  
server 192.168.0.153:8080;  
server 192.168.0.154:8080;  
server 192.168.0.155:8080;  
server 192.168.0.156:8080;
```

```
}
```

```
server {  
    listen      80;  
    server_name www.***.com;  
  
    location / {  
        root    html;  
        index   index.html index.htm;  
        proxy_redirect off;  
        proxy_set_header Host $host;  
        proxy_set_header X-Real-IP $remote_addr;  
        proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;  
        proxy_pass http://web; #注意设置在这里  
    }  
  
    error_page   500 502 503 504   /50x.html;  
    location = /50x.html {  
        root    html;  
    }  
}
```

```
server {  
    listen      80;  
    server_name  wap.***.com;  
  
    location / {  
        root    html;  
        index   index.html index.htm;  
        proxy_redirect off;  
        proxy_set_header Host $host;  
        proxy_set_header X-Real-IP $remote_addr;  
        proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;  
        proxy_pass http://wap; #注意：设置在这里  
    }  
    error_page   500 502 503 504   /50x.html;  
    location = /50x.html {  
        root    html;  
    }  
}  
}
```

7.2 Nginx Upstream 介绍

在负载均衡的 Session 解决中，已经介绍了 Nginx Upstream 支持的几种分配方法，还有一些 upstream 的配置如下：

每个设备的状态设置为：

- 1.down 表示单前的 server 暂时不参与负载
- 2.weight 默认为 1.weight 越大，负载的权重就越大。
- 3.max_fails ： 允许请求失败的次数默认为 1.当超过最大次数时，返回 proxy_next_upstream 模块定义的错误

4.fail_timeout:max_fails 次失败后，暂停的时间。

5.backup: 其它所有的非 backup 机器 down 或者忙的时候，请求 backup 机器。所以这台机器压力会最轻。

nginx 支持同时设置多组的负载均衡，用来给不用的 server 来使用。

client_body_in_file_only 设置为 On 可以讲 client post 过来的数据记录到文件中用来做 debug

client_body_temp_path 设置记录文件的目录 可以设置最多 3 层目录

location 对 URL 进行匹配.可以进行重定向或者进行新的代理 负载均衡

第 8 章 LB 负载均衡集群

LB 负载均衡集群，最简单的方法就是通过 DNS 中设置多个 A 记录，来实现 DNS 轮询，来达到负载均衡的效果，当然这个是最简单，配置维护也最简单的方案，在很多场景中都可以见到。

还有一种比较有名的就是 LVS 负载均衡了，这一章主要通过讲解 LVS 来部署一个企业级的负载均衡集群方案。

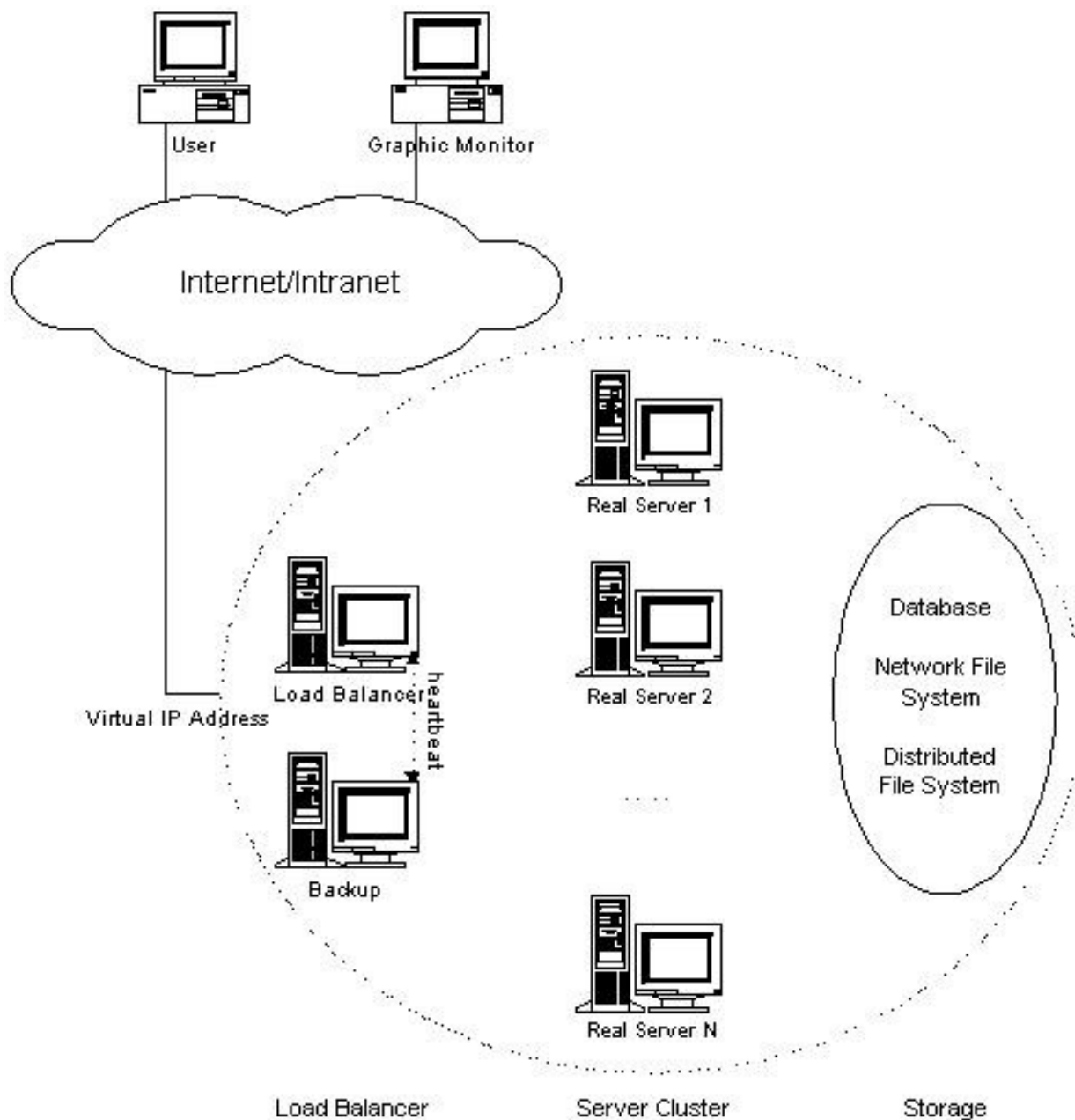
8.1 LVS 简介

LVS 是 Linux Virtual Server 的缩写，是由章文嵩博士的基于 Linux 内核的负载均衡技术。建议读者看本文档之前熟读 LVS 官方手册。

8.1.1 LVS 体系结构

LVS 建筑于实际的服务器集群之上，用户看不到提供服务的多台实际服务器，而只能看见一台作为负载均衡器的服务器。实际的服务器通过高速局域网或地理上分散的广域网连接。实际服务器的前端是一台负载均衡器，他将用户的请求调度到实际服务器上完成，这样看起来好像所有服务都是通过虚拟服务器来完成的。Linux 虚拟服务器能够提供良好的可升级性、可靠性和可用性。用户

可以透明地增加或减少一个节点，可以对实际服务器进行监测，如果发现有节点失败就重新配置系统。



8.1.2 LVS 调度算法

LVS 提供了十种调度算法：

可以在这里查看：

```
[root@Web-node ~]# ls /lib/modules/2.6.18-164.el5/kernel/net/ipv4/ipvs/
```

```
ip_vs_dh.ko  ip_vs.ko      ip_vs_lblcr.ko  ip_vs_nq.ko  ip_vs_sed.ko  ip_vs_wlc.ko
ip_vs_ftp.ko ip_vs_lblc.ko  ip_vs_lc.ko     ip_vs_rr.ko  ip_vs_sh.ko   ip_vs_wrr.ko
```

1. 轮叫 (Round Robin 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连接数和系统负载。

2. 加权轮叫 (Weighted Round Robin 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流量。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. 最少链接 (Least Connections LC)

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

4. 加权最少链接 (Weighted Least Connections 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“加权最少链接”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. 基于局部性的最少链接 (Locality-Based Least Connections LBLC)

“基于局部性的最少链接”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“最少链接”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少链接 (Locality-Based Least Connections with Replication LBLCR)

“带复制的基于局部性最少链接”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，将该服务器加入到服务器组中，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 (Destination Hashing DH)

“目标地址散列”调度算法根据请求的目标 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 (Source Hashing SH)

“源地址散列”调度算法根据请求的源 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。

9. 最短期望延迟 (Shortest Expected Delay Scheduling SED)

分配一个接踵而来的请求以最短的期望的延迟方式到服务器。

10. 最小队列调度 (Never Queue Scheduling NQ)

分配一个接踵而来的请求到一台空闲的服务器，此服务器不一定是最快的那台，如果所有服务器都是繁忙的，它采取最短的期望延迟分配请求。

8.1.3 LVS 负载均衡方法

LVS 提供了三种 IP 级的负载平衡方法：

Virtual Server via NAT 、Virtual Server via IP Tunneling、Virtual Server via Direct Routing。

Virtual Server via NAT 方法使用了报文双向重写的方法， Virtual Server via IP Tunneling 采用的是报文单向重写的策略， Virtual Server via Direct Routing 采用的是报文转发策略，这些策略将在以后的文章中详细描述。

8.1.4 常用术语介绍

DGW	公网 IP 地址的默认网关
VIP	客户端访问的公网 IP 地址，Director 的虚拟 IP 地址
DIP	Director 的真实 IP 地址
RIP	真实机的 IP 地址
CIP	客户端 IP 地址

8.1.5 相关资源

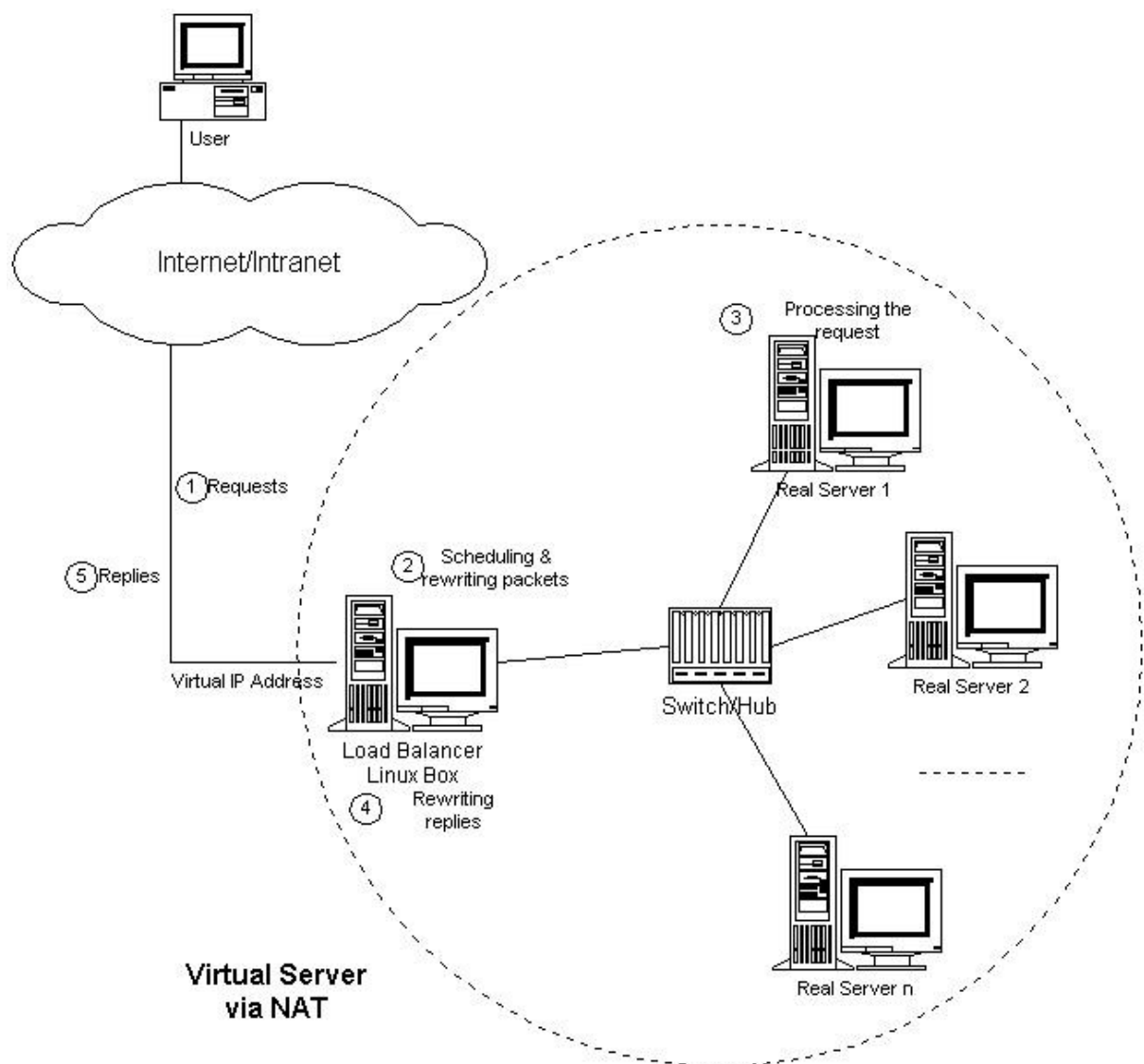
LVS 官方网站: <http://www.linuxvirtualserver.org/>

LVS 官方中文文档: <http://www.linuxvirtualserver.org/zh/index.html>

8.2 LVS-NAT 方式部署

8.2.1 LVS-NAT 方式体系结构

客户通过 Virtual IP Address (虚拟服务的 IP 地址) 访问网络服务时, 请求报文到达调度器, 调度器根据连接调度算法从一组真实服务器中选出一台服务器, 将报文的目标地址 Virtual IP Address 改写成选定服务器的地址, 报文的目标端口改写成选定服务器的相应端口, 最后将修改后的报文发送给选出的服务器。同时, 调度器在连接 Hash 表中记录这个连接, 当这个连接的下一个报文到达时, 从连接 Hash 表中可以得到原选定服务器的地址和端口, 进行同样的改写操作, 并将报文传给原选定的服务器。当来自真实服务器的响应报文经过调度器时, 调度器将报文的源地址和源端口改为 Virtual IP Address 和相应的端口, 再把报文发给用户。我们在连接上引入一个状态机, 不同的报文会使得连接处于不同的状态, 不同的状态有不同的超时值。在 TCP 连接中, 根据标准的 TCP 有限状态机进行状态迁移。在 UDP 中, 我们只设置一个 UDP 状态。不同状态的超时值是可以设置的, 在缺省情况下, SYN 状态的超时为1分钟, ESTABLISHED 状态的超时为15分钟, FIN 状态的超时为1分钟; UDP 状态的超时为5分钟。当连接终止或超时, 调度器将这个连接从连接 Hash 表中删除。这样, 客户所看到的只是在 Virtual IP Address 上提供的服务, 而服务器集群的结构对用户是透明的。



8.2.2 部署前的准备工作

1> 服务器规划

IP 地址	主机名	描述
192.168.130.130	Web-Master	Director 分发器 (VIP)
192.168.140.132	Web-Master	Director 分发器 (DIP)
192.168.140.133	Web-node1	Real Server Web 节点 1
192.168.140.134	Web-node2	Real Server Web 节点 2

192.168.130.131	Web-Client	测试客户端
-----------------	------------	-------

2 > 设置主机名解析和 ssh 验证

```
[root@Web-node ~]# vim /etc/hosts

# Do not remove the following line, or various programs
# that require network functionality will fail.
127.0.0.1      localhost.localdomain localhost
::1           localhost6.localdomain6 localhost6

192.168.140.132 Web-node
192.168.140.133 Web-node1
192.168.140.134 Web-node2

[root@Web-node ~]# ssh-keygen
[root@Web-node ~]# cp .ssh/id_rsa.pub .ssh/authorized_keys
[root@Web-node1 ~]# mkdir .ssh
[root@Web-node2 ~]# mkdir .ssh

[root@Web-node ~]# scp .ssh/authorized_keys Web-node1:/root/.ssh
[root@Web-node ~]# scp .ssh/authorized_keys Web-node2:/root/.ssh
[root@Web-node ~]# scp /etc/hosts Web-node1:/etc
[root@Web-node ~]# scp /etc/hosts Web-node2:/etc
```

8.2.3 在 Real Server 上的部署

1> Web-node1:

```
[root@Web-node1 ~]# mount /dev/cdrom /mnt
[root@Web-node1 ~]# rpm -ivh /mnt/Server/httpd-2.2.3-31.el5.i386.rpm
[root@Web-node1 ~]# echo Web-node1 > /var/www/html/index.html
[root@Web-node1 ~]# /etc/init.d/httpd start
[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
DEVICE=eth0
```

```
BOOTPROTO=static
```

```
IPADDR=192.168.140.133
```

```
NETMASK=255.255.255.0
```

```
GATEWAY=192.168.140.132 注意：默认网关设置为 DIP
```

```
ONBOOT=yes
```

```
HWADDR=00:0c:29:7c:cf:ba
```

```
[root@Web-node1 ~]# /etc/init.d/network restart
```

2> Web-node2

```
[root@Web-node2 ~]# mount /dev/cdrom /mnt
```

```
[root@Web-node2 ~]# rpm -ivh /mnt/Server/httpd-2.2.3-31.el5.i386.rpm
```

```
[root@Web-node2 ~]# echo Web-node2 > /var/www/html/index.html
```

```
[root@Web-node2 ~]# /etc/init.d/httpd start
```

```
[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
```

```
DEVICE=eth0
```

```
BOOTPROTO=static
```

```
IPADDR=192.168.140.134
```

```
NETMASK=255.255.255.0
```

```
GATEWAY=192.168.140.132 注意：默认网关设置为 DIP
```

```
ONBOOT=yes
```

```
HWADDR=00:0c:29:5d:2d:90
```

```
[root@Web-node2 ~]# /etc/init.d/network restart
```

请用浏览器访问两个 Web 几点，保证服务是正常运行。

8.2.4 在 Director 上的部署

1> 打开 IP_Forward

```
[root@Web-node ~]# vi /etc/sysctl.conf
```

```
net.ipv4.ip_forward = 1
```

```
[root@Web-node ~]# sysctl -p
```

2> 绑定 DIP 和 VIP

```
[root@Web-node ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
DEVICE=eth0
BOOTPROTO=static
IPADDR=192.168.140.132
NETMASK=255.255.255.0
ONBOOT=yes
HWADDR=00:0c:29:17:39:9c
```

3> 安装 ipvsadm 软件包

```
[root@Web-node ~]# rpm -ivh /mnt/Cluster/ipvsadm-1.24-10.i386.rpm
```

4> 设置 ipvsadm

```
[root@Web-node ~]# modprobe iptable_nat
[root@Web-node ~]# ipvsadm -A -t 192.168.130.130:80 -s rr
[root@Web-node ~]# ipvsadm -a -t 192.168.130.130:80 -r 192.168.140.133 -m
[root@Web-node ~]# ipvsadm -a -t 192.168.130.130:80 -r 192.168.140.134 -m
[root@Web-node ~]# service ipvsadm save
Saving IPVS table to /etc/sysconfig/ipvsadm:          [ OK ]
[root@Web-node ~]# chkconfig ipvsadm on
```

8.2.5 LVS -NAT 方式集群测试

1> 手动效果测试

```
[root@Web-Client ~]# elinks http://192.168.130.130 发现访问的是 Web-node1
[root@Web-Client ~]# elinks http://192.168.130.130 发现访问的是 Web-node2
```

2> 压力负载测试

在这里使用 Apache 自带的 ab 测试工具。

```
[root@Web-node ~]# ab -n 100 -c 100 http://192.168.140.133/
This is ApacheBench, Version 2.0.40-dev <$Revision: 1.146 $> apache-2.0
Copyright 1996 Adam Twiss, Zeus Technology Ltd, http://www.zeustech.net/
Copyright 2006 The Apache Software Foundation, http://www.apache.org/
```

Benchmarking 192.168.140.133 (be patient).....done

```
Server Software:      Apache/2.2.3      #服务器平台和版本
Server Hostname:      192.168.140.133  #服务器主机名
Server Port:          80                #服务器端口号

Document Path:        /                #测试的页面
Document Length:      10 bytes          #测试的页面大小

Concurrency Level:     100              #并发数
Time taken for tests:   0.91990 seconds #整个测试持续时间
Complete requests:     100              #完成的请求数
Failed requests:       0                #失败的请求数
Write errors:          0
Total transferred:     27200 bytes       #整个测试场景的网络传输量
HTML transferred:      1000 bytes        #整个测试场景的 HTML 内容传输量 (10X100)
Requests per second:   1087.07 [#/sec] (mean) #平均每秒处理的事务数
Time per request:      91.990 [ms] (mean)  #平均事物响应时间
Time per request:      0.920 [ms] (mean, across all concurrent requests) #每个请求运行
                        的平均时间
Transfer rate:         282.64 [Kbytes/sec] received
```

#平均每秒网络上的流量，可以帮助排除是否存在网络流量过大导致响应时间延长的问题

Connection Times (ms)

	min	mean[+/-sd]	median	max
Connect:	1	19 9.3	21	33
Processing:	9	32 14.2	32	65
Waiting:	8	31 14.1	31	56
Total:	10	52 22.9	54	90

Percentage of the requests served within a certain time (ms)

50%	54
66%	66
75%	71
80%	75
90%	83
95%	87
98%	89
99%	90
100%	90 (longest request)

测试图表

类 型	请求和并发数	每个请求平均处理时间
单个节点 Web-node1	100 个请求，每个请求 100 并发	0.961 [ms]
单个节点 Web-node2	100 个请求，每个请求 100 并发	0.950 [ms]
LVS 负载均衡	200 个请求，每个请求 200 并发	0.819

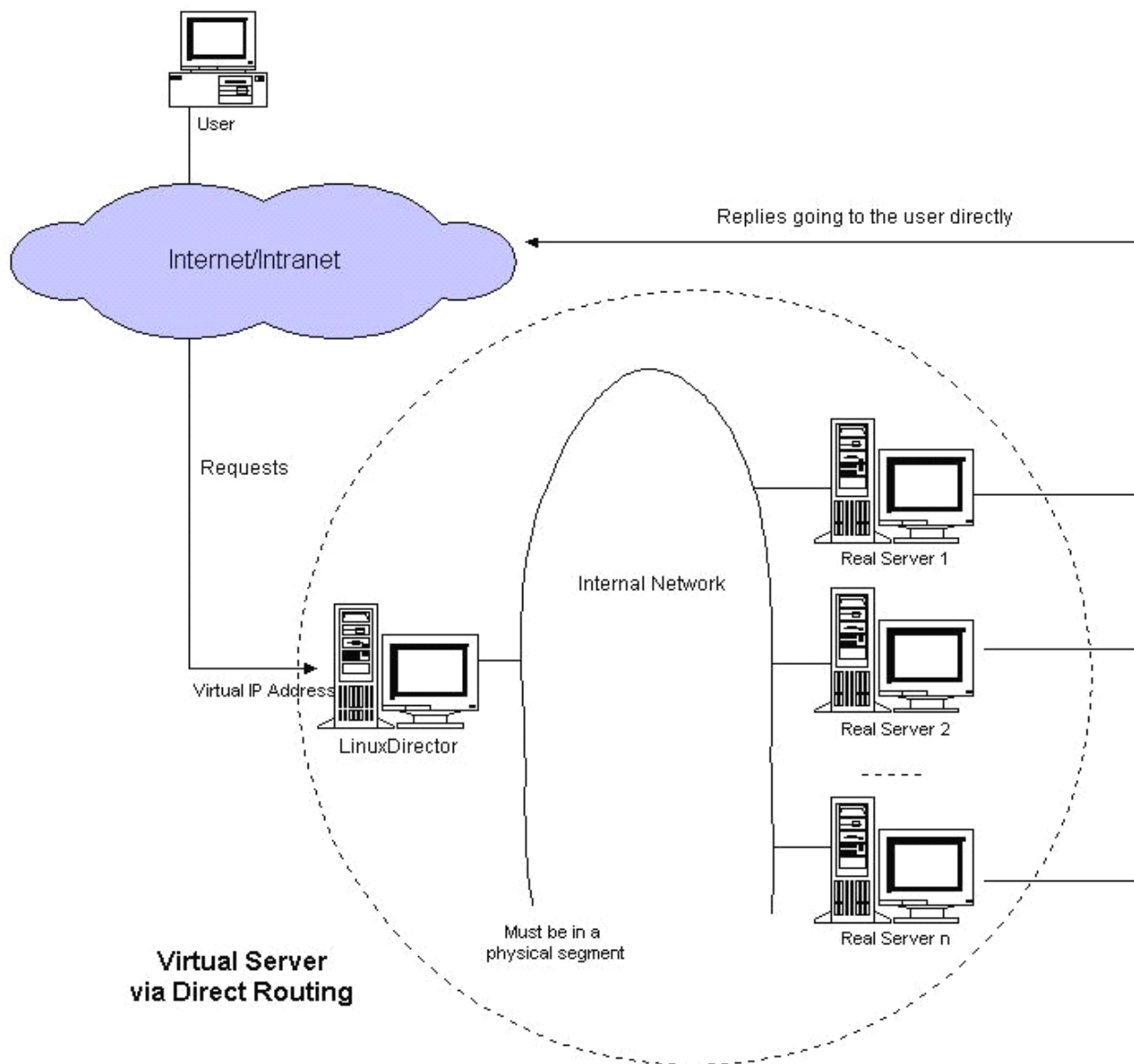
8.3 LVS-DR 方式部署

8.3.1 LVS-DR 方式体系结构

在 LVS-DR 中，调度器根据各个服务器的负载情况，动态地选择一台服务器，不修改也不封装 IP 报文，而是将数据帧的 MAC 地址改为选出服务器的 MAC 地址，再将修改后的数据帧在与服务器组的局域网上传送。因为数据帧的 MAC 地址是选出的服务器，所以服务器肯定可以收到这个数据帧，从中可以获得该 IP 报文。当服务器发现报文的目标地址 VIP 是在本地的网络设备上，服务器处理这个报文，然后根据路由表将响应报文直接返回给客户。

在 LVS-DR 中，根据缺省的 TCP/IP 协议栈处理，请求报文的目标地址为 VIP，响应报文的源地

址肯定也为 VIP，所以响应报文不需要作任何修改，可以直接返回给客户，客户认为得到正常的服务，而不会知道是哪一台服务器处理的。



8.3.2 部署前的准备工作

1> 服务器规划

IP 地址	主机名	描述
192.168.140.128	Web-node	Director 分发器 (VIP)

192.168.140.132	Web-node	Director 分发器 (DIP)
192.168.140.133	Web-node1	Real Server Web 节点 1
192.168.140.134	Web-node2	Real Server Web 节点 2
192.168.140.130	Web-Client	测试客户端

在这里仅需要修改上个实验的 Direcoor 分发器的 IP 地址即可，使用 192.168.140.0 这个网段模拟公网地址。

2> 清除上个实验的配置

```
[root@Web-node ~]# echo "" > /etc/sysconfig/ipvsadm
```

```
[root@Web-node ~]# service ipvsadm restart
```

8.3.3 在 Real Server 上的部署

1> Web-node1

第一步：

```
[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
```

修改 GATEWAY 为 DGW: 192.168.140.1

```
[root@Web-node1 ~]# /etc/init.d/network restart
```

第二步：

LVS-DR 方式中的 Real Server 需要忽略 ARP 解析，在这里用脚本实现

```
[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/arp.sh
```

```
#!/bin/bash
```

```
#=====
```

```
# $Name:      RealServer.sh
```

```
# $Revision:   1.0
```

```
# $Function:   Config realserver lo and apply noarp
```

```
# $Author:     Shundong Zhao
```

```
# $organization: UnixHot
```

```
# $Create Date: 2010-08-10
```

```
#=====
```

```
WEB_VIP=192.168.140.140
```

```
. /etc/rc.d/init.d/functions
```

```
case "$1" in
```

```
start)
```

```
    ifconfig lo:0 $SNS_VIP netmask 255.255.255.255 broadcast $WEB_VIP
```

```
    /sbin/route add -host $WEB_VIP dev lo:0
```

```
    echo "1" >/proc/sys/net/ipv4/conf/lo/arp_ignore
```

```
    echo "2" >/proc/sys/net/ipv4/conf/lo/arp_announce
```

```
    echo "1" >/proc/sys/net/ipv4/conf/all/arp_ignore
```

```
    echo "2" >/proc/sys/net/ipv4/conf/all/arp_announce
```

```
    sysctl -p >/dev/null 2>&1
```

```
    echo "RealServer Start OK"
```

```
;;
```

```
stop)
```

```
    ifconfig lo:0 down
```

```
    route del $WEB_VIP >/dev/null 2>&1
```

```
    echo "0" >/proc/sys/net/ipv4/conf/lo/arp_ignore
```

```
    echo "0" >/proc/sys/net/ipv4/conf/lo/arp_announce
```

```
    echo "0" >/proc/sys/net/ipv4/conf/all/arp_ignore
```

```
    echo "0" >/proc/sys/net/ipv4/conf/all/arp_announce
```

```
    echo "RealServer Stopped"
```

```
;;
```

```
*)
```

```
    echo "Usage: $0 {start|stop}"
```

```
    exit 1
```

```
esac
```

```
exit 0[root@Web-node1 ~]# bash /etc/sysconfig/network-scripts/arp.sh
[root@Web-node1 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
          inet addr:192.168.140.128  Mask:255.255.255.255
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
```

2> Web-node2

第一步:

```
[root@Web-node2 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
```

修改 GATEWAY 为 DGW: 192.168.140.1

```
[root@Web-node2 ~]# /etc/init.d/network restart
```

第二步:

```
[root@Web-node1 ~]#
scp /etc/sysconfig/network-scripts/arp.sh Web-node2:/etc/sysconfig/network-scripts/
[root@Web-node2 ~]# bash /etc/sysconfig/network-scripts/arp.sh
[root@Web-node2 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
          inet addr:192.168.140.128  Mask:255.255.255.255
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
```

8.3.4 在 Director 上的部署

1> 绑定 VIP

```
[root@Web-node ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0:1
DEVICE=eth0:1
BOOTPROTO=static
IPADDR=192.168.140.128
NETMASK=255.255.255.0
ONBOOT=yes
HWADDR=00:0c:29:17:39:9c
```

2> 配置 ipvsadm

```
[root@Web-node ~]# ipvsadm -A -t 192.168.140.128:80 -s wrr -p 3600
[root@Web-node ~]# ipvsadm -a -t 192.168.140.128:80 -r 192.168.140.133:80 -g -w 10
[root@Web-node ~]# ipvsadm -a -t 192.168.140.128:80 -r 192.168.140.134:80 -g -w 20
[root@Web-node ~]# service ipvsadm save
```

8.3.5 LVS-DR 方式集群测试

请读者用 ab 测试，模拟环境测试的性能指标可能区别不太明显（略）

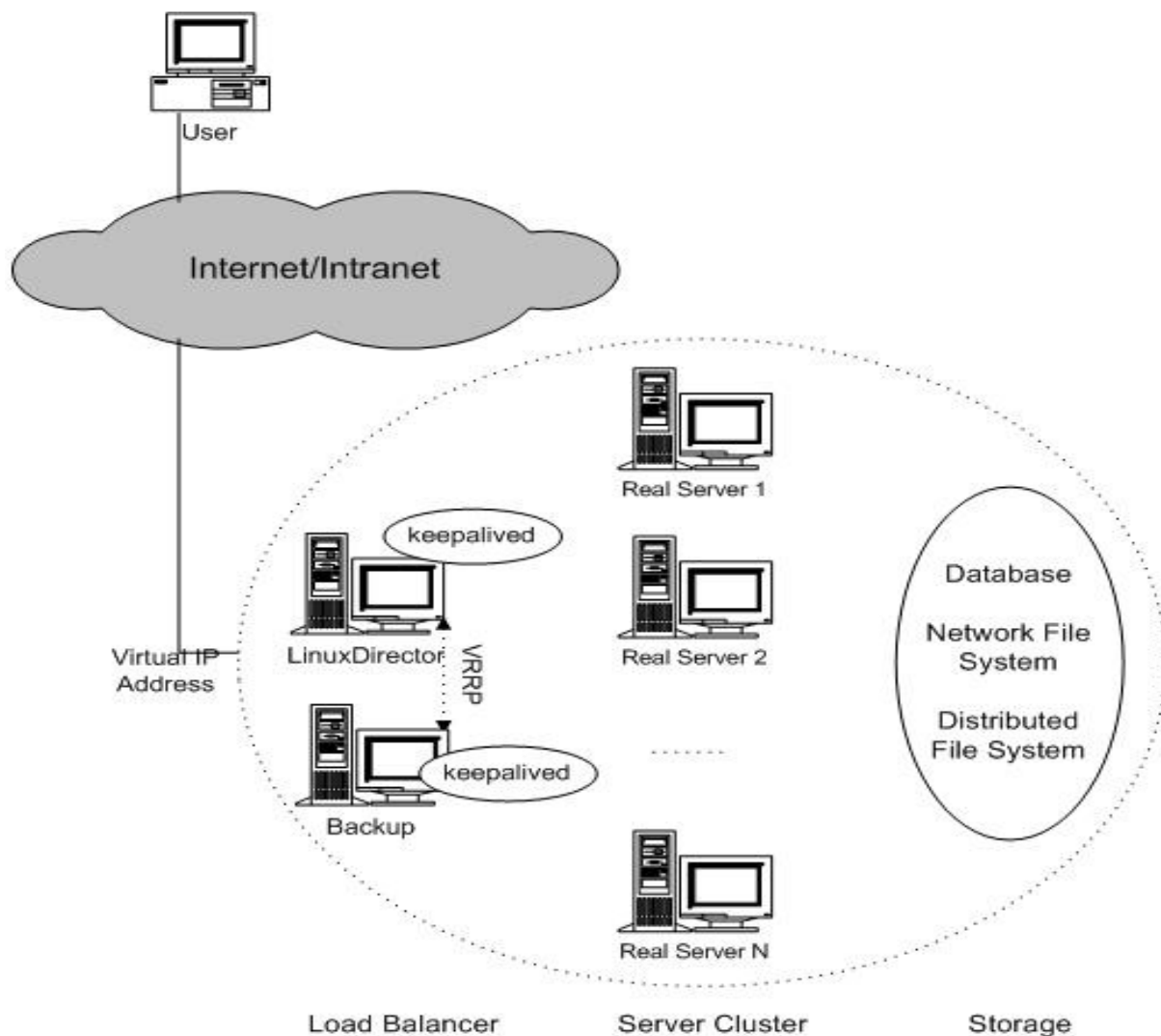
第 9 章 LVS Keepalived 集群

9 LVS Keepalived 方案

9.1.1 LVS Keepalived 方案简介

LVS Keepalived 方案是和 LVS Heartbeat 一样的高可用加负载均衡方案，Keepalived 于 Heartbeat 相比在部署和管理方面都大大简化了。

9.1.2 集群架构图



9.1.3 集群环境列表

IP 地址	主机名	说明
192.168.140.140		VIP
192.168.140.141	Web-Director	LVS Director
192.168.140.143	Web-Backup	Keepalived Backup
192.168.140.137	Web-node1	Web 节点
192.168.140.139	Web-node2	Web 节点

9.1.4 其它注意事项

- 1> LVS Keepalived 方案 ipvsadm 的管理有 keepalived 负责。
- 2> LVS 后端节点的监控检查也是 keepalived 进行。
- 3> 注意内核版本和 ipvsadm 版本的对应。
- 4> 注意 ip_vs 模块是否载入内核。

9.2 部署 LVS Keepalived

9.2.1 检查内核版本和模块

```
[root@Web-Director ~]# uname -r
2.6.18-164.el5
[root@Web-Director ~]# modprobe ip_vs
[root@Web-Director ~]# lsmod | grep ip_vs
ip_vs                121473  0
```

9.2.2 下载软件包

```
[root@Web-Director ~]# ln -s /usr/src/kernels/2.6.18-164.el5-x86_64/ /usr/src/linux
[root@Web-Director ~]# cd /usr/local/src
[root@Web-Director src]# wget
http://www.linuxvirtualserver.org/software/kernel-2.6/ipvsadm-1.24.tar.gz
[root@Web-Director src]# wget
http://www.keepalived.org/software/keepalived-1.1.20.tar.gz
```

9.2.3 安装 ipvsadm

注意：Web-Director 和 Web-Backup 安装相同

```
[root@Web-Director src]# tar zxvf ipvsadm-1.24.tar.gz
```

```
[root@Web-Director src]# cd ipvsadm-1.24
```

有关安装的相关信息，可以查看该目录下的 README 文件。

```
[root@Web-Director ipvsadm-1.24]# make && make install
```

安装完毕后会生成以下文件：

```
/sbin/ipvsadm
/sbin/ipvsadm-save
/sbin/ipvsadm-restore
/usr/man/man8/ipvsadm.8
/usr/man/man8/ipvsadm-save.8
/usr/man/man8/ipvsadm-restore.8
/etc/rc.d/init.d/ipvsadm
```

9.2.4 安装 Keepalived

注意：Web-Director 和 Web-Backup 安装相同

```
[root@Web-Director src]# tar zxvf keepalived-1.1.20.tar.gz
```

```
[root@Web-Director src]# cd keepalived-1.1.20
```

```
[root@Web-Director keepalived-1.1.20]# ./configure --prefix=/usr/local/keepalived
```

Keepalived configuration

```
-----
Keepalived version      : 1.1.20
Compiler                : gcc
Compiler flags          : -g -O2
Extra Lib               : -lpopt -lssl -lcrypto
Use IPVS Framework      : Yes
IPVS sync daemon support : Yes
```



```
Use VRRP Framework      : Yes
```

```
Use Debug flags         : No
```

如果出现以上输出,说明可以正常编译安装。

```
[root@Web-Director keepalived-1.1.20]# make && make install
```

```
[root@Web-Director ~]# cp /usr/local/keepalived/etc/rc.d/init.d/keepalived  
/etc/rc.d/init.d/
```

```
[root@Web-Director ~]# cp /usr/local/keepalived/etc/sysconfig/keepalived /etc/sysconfig/
```

```
[root@Web-Director ~]# /usr/local/keepalived/sbin/keepalived --help
```

如果你执行了 keepalived 命令,你会发现它默认去/etc/keepalived/keepalived.conf 找配置文件,所以我们要把配置文件创建到此处方便启动,当然,你也可以自己制定配置文件位置。

```
[root@Web-Director ~]# mkdir /etc/keepalived
```

```
[root@Web-Director ~]# cp /usr/local/keepalived/etc/keepalived/keepalived.conf  
/etc/keepalived/
```

```
[root@Web-Director ~]# cp /usr/local/keepalived/sbin/keepalived /usr/sbin/
```

9.3 LVS 配置

9.3.1 LVS Director 配置

LVS 配置是由 keepalived 依靠/etc/keepalived/keepalived.conf 来进行配置的,也就是说我们不需要手动配置 ipvsadm 程序,但是我们做为 LVS DR 模式运行,不能忘了对真实机的配置,对 RealServer 的配置参考“第 2 章 2.3.3 小节”。

9.3.2 RealServer 配置

```
[root@Web-node1 ~]# /sbin/realserver.sh start
```

```
RealServer Start OK
```

```
[root@Web-node2 ~]# /sbin/realserver.sh start
```

```
RealServer Start OK
```

在 RealServer 执行完 realserver.sh 后,一定要检查:

```
[root@Web-node1 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
          inet addr:192.168.140.140  Mask:255.255.255.255
          UP LOOPBACK RUNNING  MTU:16436  Metric:1

[root@Web-node2 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
          inet addr:192.168.140.140  Mask:255.255.255.255
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
```

9.4 Keepalived 配置

这是 LVS Keepalived 集群方案的核心，主要集中在/etc/keepalived/keepalived 文件上。

Keepalived 有两种状态，MASTER 和 BACKUP，配置大致相似，但也有不同的地方，在 MASTER 的配置时，会用红色将，MASTER 和 BACKUP 不同的地方标出。

9.4.1 Web-Director 配置

下面是一个配置好的 keepalived.conf 文件，对于每行用注释说明。

```
[root@Web-Director ~]# cat /etc/keepalived/keepalived.conf
! Configuration File for keepalived

global_defs {
                                #全局定义块

    notification_email {
        admin@unixhot.com      #邮件通知模块，定义通知邮件地址。
    }

    notification_email_from localhost #
    smtp_server 127.0.0.1         #定义 SMTP Server
    smtp_connect_timeout 30       #SMTP 链接超时时间
    router_id LVS_MASTER          #运行 keepalived 机器的一个标识，注意：Web-Backup 应该不同，修改为 LVS_BACKUP
}
```

```
vrrp_instance VI_1 {  
    state MASTER                #实例状态，注意：Web-Backup 应该修改为 BACKUP  
    interface eth0              #指定对外服务的网卡。  
    virtual_router_id 51  
    priority 101                #优先级，注意：Web-Backup 应该修改比这个值小  
    advert_int 1  
    authentication {  
        auth_type PASS  
        auth_pass 1111  
    }  
    virtual_ipaddress {  
        192.168.140.140  
    }  
}  
  
virtual_server 192.168.140.140 80 {  
    delay_loop 6                #健康检查的时间间隔  
    lb_algo wrr                 #负载均衡的调度算法  
    lb_kind DR                  #负载均衡转发模式  
    nat_mask 255.255.255.0  
    persistence_timeout 50      #会话保持时间，针对动态网站  
    protocol TCP                #转发的协议  
  
    real_server 192.168.140.137 80 {        #真实机的设置  
        weight 1                 #真实机的权重，在带有加权调度的调度算法中  
    }                                     #TCP 健康检查  
    TCP_CHECK {  
        connect_timeout 10  
        nb_get_retry 3  
        delay_before_retry 3  
        connect_port 80
```

```
    }  
}  
  
real_server 192.168.140.139 80 {  
    weight 1  
    TCP_CHECK {  
        connect_timeout 10  
        nb_get_retry 3  
        delay_before_retry 3  
        connect_port 80  
    }  
}
```

9.4.2 BACKUP 端配置

BACKUP 端配置和 MASTER 几乎一样，可以直接用 scp 从 MASTER 端复制一份过来，做以下修改即可：

```
router_id LVS_MASTER  
state MASTER  
priority 101
```

9.5 LVS Keepalived 方案测试

9.5.1 启动 keepalived 服务

```
[root@Web-Director ~]# /etc/init.d/keepalived start  
[root@Web-Backup ~]# /etc/init.d/keepalived start
```

9.5.2 VIP 切换测试

```
[root@Web-Director ~]# ip ad li eth0
```

```
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast qlen 1000
    link/ether 00:0c:29:57:04:db brd ff:ff:ff:ff:ff:ff
    inet 192.168.140.141/24 brd 192.168.140.255 scope global eth0
    inet 192.168.140.140/32 scope global eth0
    inet6 fe80::20c:29ff:fe57:4db/64 scope link
        valid_lft forever preferred_lft forever
```

可以发现，默认的 VIP 是绑定在优先级比较高的这台 MASTER 上，也就是 Web-Director 服务器。

```
[root@Web-Backup ~]# ip ad li eth0
```

```
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast qlen 1000
    link/ether 00:0c:29:45:5e:0e brd ff:ff:ff:ff:ff:ff
    inet 192.168.140.143/24 brd 192.168.140.255 scope global eth0
    inet6 fe80::20c:29ff:fe45:5e0e/64 scope link
        valid_lft forever preferred_lft forever
```

你可以手动来停掉 Web-Director 上的 keepalived 服务，VIP 就会自动切到 Web-Backup 上。

```
[root@cache-node1 ~]# cd /usr/local/src
```

```
[root@cache-node1 src]# wget http://www.squid-cache.org/Versions/v2/2.7/squid-2.7.STABLE9.tar.gz
```

```
[root@cache-node1 src]# tar zxvf squid-2.7.STABLE9.tar.gz
```

```
[root@cache-node1 src]# cd squid-2.7.STABLE9
```

实验答疑: <http://www.unixhot.com>

<http://www.bosshot.com>

附录: GFDL 协议

(注: 可以参考本协议的中文翻译版本 <http://www.thebigfly.com/gnu/FDLv1.3/>)

GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc. <<http://fsf.org/>>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels)

generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other

implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or

state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other

copyright notices.

- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard. You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to

the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License. However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been

terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of

another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (C) YEAR YOUR NAME.

Permission is granted to copy, distribute and/or modify this document

under the terms of the GNU Free Documentation License, Version 1.3

or any later version published by the Free Software Foundation;

with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU

Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the

Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

实验答疑: <http://www.unixhot.com>

<http://www.bosshot.com>