河北移动邮件系统负载均衡方案

设计目标

邮件系统在当今社会，作为一种基础通讯平台，已经不单纯是简单收发邮件，而是成为互联网各种应用的核心，承载越来越重要的应用。中国人口众多，上网人数逐年剧增，使用邮件也日益频繁。邮件系统从建成的那一天起，就面临升级的压力。如何设计一个结构良好的大容量邮件系统，对于系统稳定性、可靠性，对于日后的升级维护有着至关重要的作用。

春笛公司作为一个在邮件领域默默耕耘的公司，以小用户量的企业邮件系统立身，最终希望打造一个坚固的、高度可扩展的、容易管理维护的大容量分布式邮件系统。在设计时，我们主要考虑如下方面：

1．底层坚固、高度稳定。

为保证系统的稳定可靠，需要在硬件、操作系统、核心 MTA、应用层在内的每一个环节都稳定可靠才行。硬件通常选取知名品牌服务器，稳定性、可靠性都有保障，差别不大。

操作系统选择 Linux 或者 FreeBSD, 针对邮件系统的特点，内核需要特殊调整：如打开文件数（open files）、stack size 、max user processes 等。除了操作系统核心外，系统只加载必须的软件，屏蔽一切不不要的服务端口。

在操作系统之上，处理 smtp、pop3 请求的 MTA 的稳定性、效率也至关重要。当今世界上使用比较多的是 qmail 和 postfix, 都有分布世界各地的大批用户。相比较而言，qmail 有着更好的模块化设计、更好的安全性，更高的投递效率、更可靠的队列设计。Postfix 优势在于和 sendmail 有着很好的兼容性，部署容易，集成程度比较高，也是一个非常不错的 MTA 软件。

应用层我们选取 Apache+tomcat。Apache 久负盛名、久经考验，tomcat 背后有 SUN 支持，最重要的是 tomcat5 支持应用层负载均衡（Load Balance）。另外，java 作为一种面向对象的编程语言，最能体现软件工程思想，有一系列的 UML 设计工具、集成开发环境、应用服务器可以选择。很多学校也开设 JAVA 课程，以后 JAVA 会像 c 语言那样普及，变成程序员必备的技能之一。邮件系统应用层会根据用户的反馈增加信的增值服务品种，如果基于 java 开发，很容找到相关人才。这样缩短开发周期、节省开发成本、降低维护难度。Google 很多服务是基于 java 开发的。当然，java 也有执行效率低的缺点，但随着硬件速度的提升，单只程序运行速度的劣势很容易被良好的设计模式优势取代。

2．容易管理、维护。

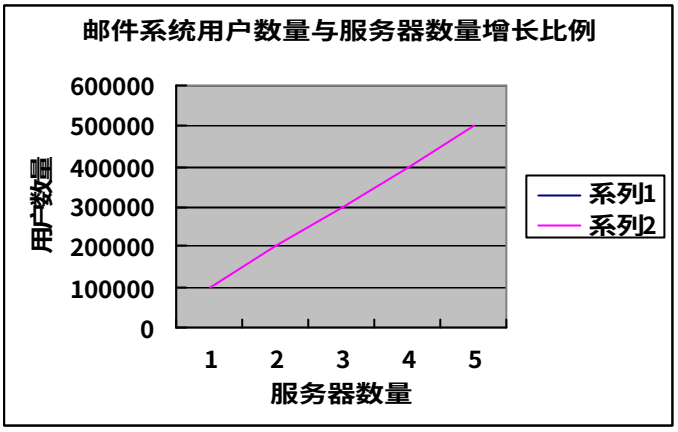
系统结构复杂之后，必然会带来管理维护上的麻烦。我们的设计思想是希望通过统一的一个管理控制界面，让系统管理员对每台服务器的运行状况、负载情况、流量了如指掌；同时通过一个管控界面，可以远程控制服务器启动、关闭，对服务进程远程启动、停止，对流量调配。当出现异常时，系统自动通过短信通知到管理人员的手机上。

3．增加设备简单、有效。

系统扩展性是衡量系统设计好坏的一个重要指标。好的系统能够通过简单添置硬件、软件做少许配置即可满足需要。我们设计的是让用户数量的增长和邮件服务器数量呈线性关系。由于 PC 服务器的硬件成本比较低，以一台高性能 PC 服务器 3 万元人民币计算，如果

作为 SMTP 服务器可以支撑 15 万用户，作为 POP 服务器可以支撑 30 万用户，作为存储服务器，可以支撑 3 万用户。平均每用户为 0.2 元、0.1 元、1 元。

在保证系统稳定可靠的同时，能够最大地降低成本。降低成本有 2 个途径，一是最大限度利用硬件，二是避免使用高端的存储备份设备、负载均衡设备、四层交换机等。



4．具有多级权限管理，支持个人用户、企业用户、运营商等。

大容量邮件系统，不仅仅是个人用户，还有企业用户。不同用户群体的需求是不同的。针对不同用户群，提供不同的服务套餐，无疑是市场营销重要手段。而这，需要有技术上做保障才行。金笛邮件应用层采用 java 技术，无疑是体现用户需求的最佳技术手段。

作为运营商，要有丰富权限管理体系，使邮件系统的每一个管理环节严紧、严密。

5．高度模块化，模块之间最小耦合。

模块化程度的高低，可以体现的系统的成熟度。充分利用开源宝库中的免费资源，将优秀的开源项目经过改良、优化，会搭建出优秀的系统。不赞成完全从底层开发。

模块之间通过标准接口通讯，耦合程度尽可能小，这样，即使出问题也不会影响全局。最重要的是，系统扩展性大大增强。如可以随时将病毒扫描过滤模块升级，或者将垃圾邮件过滤模块升级，其它模块无需做任何改动。

6．提供与其它系统直接的接口：如计费平台、短信彩信平台、防毒网关等。

作为邮件运营，计费模块很重要。良好的设计可以为灵活的计费提供原始数据。根据这些计费数据，可以制定灵活的促销手段。

随着彩信的普及，邮件系统与彩信会逐渐融合。其它的扩展平台，如防病毒网关、反垃圾网关、反黄网关都可以灵活对接。

7．应用层二次开发、部署简单方便。

邮件系统的生命力来自客户的需求，只有不断满足客户需求，推陈出新，与时俱进，才会不断有新的用户加入。根据用户需求进行二次开发，这是必不可少的。二次开发必须简单，方便。金笛邮件通过统一的二次开发接口 Jindi-Mailet 实现服务端应用的扩展。

邮件系统架构上的演化和优劣比较

大容量邮件系统按照存储方式不同，大致可以分为 2 类：

1．统一存储

邮件队列和邮件数据集中放在存储设备上。前端 smtp 服务器多台，POP 服务器多台，随机选择一台 smtp 服务器或者 POP 服务器（这个过程一般采用 DNS 轮询方式完成）。选定某台服务器后，与该服务器建立连接，通过认证系统确认用户身份后，发送或者接收邮件数据。因为是统一存储，数据都是放在磁盘阵列上，通过 NFS 方式挂在每台服务器上。不论是通过哪台功能服务器，都可以完成邮件收发。

用户对 Webmail 的请求和 smtp/pop3 类似，也是通过多台机器随机选取实现负载分布。

2．分布式存储

邮件数据分布在每一台服务器上，每台服务器都提供完整的邮件服务：smtp/pop3/imap/webmail/数据库。用户请求过来之后，首先查询目录服务器，验证用户身份，然后找到对应的服务器，建立连接，收取或者发送邮件。

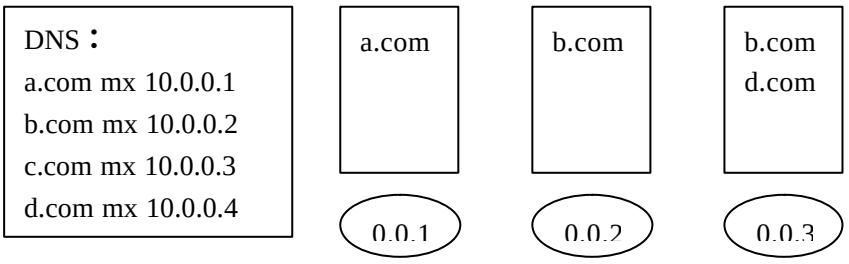
这种方式最大的优点是用户数量和邮件服务器数量可以保持线性增长。每台服务器支持 10 万邮箱，当一台服务器用户已达这个数字时，启用另外一台。可以通过管理程序在不同服务器直接迁移用户。动态管理服务器。

从早期 UNIX 几十用户的简单邮件系统，到现在几百万、上千万邮件系统，中间经历很多变化。不妨把这些梳理一下，比较邮件系统各种技术的优劣。

1．一机一域、一机多域（虚拟域）

一机一域代表用户是企业用户。一台服务器作为邮件服务器。一机多域象新网、万网，给很多企业提供邮箱。

所有的服务，如 smtp,pop3,imap 都在一台服务器上，对方邮件服务器通过查询 DNS 即可唯一锁定收件方服务 IP，直接投递过去。

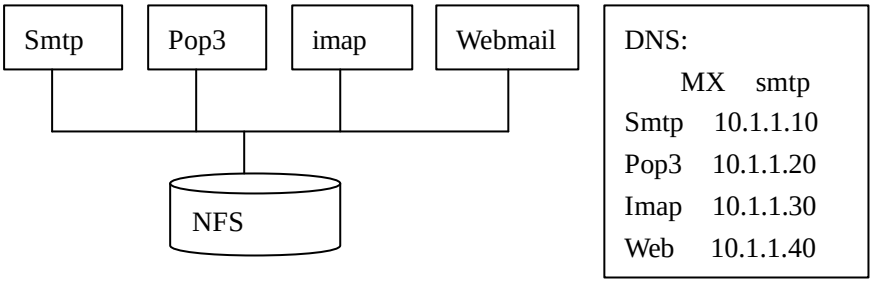


这种单机邮件系统，由于受 cpu、内存、硬盘的制约，用户最多一般不超过 10 万用户。考虑可靠性，一般采用 HA，将用户数据存放磁盘阵列上，正常只有一台服务器工作，异常时自动切换到另外一台。

对于一般的企业用户，这种单机邮件系统已经够用。但对于大型邮件系统，这种单机系统显然无法满足。

2．多机一域，功能分割，集中存储。

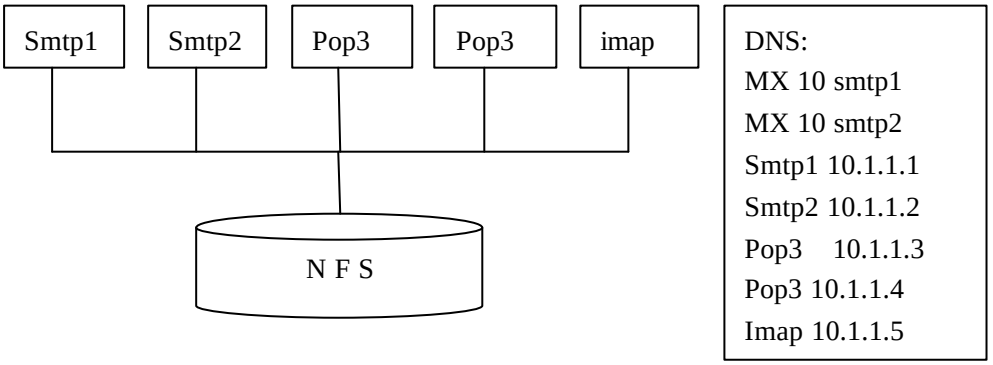
邮件服务器 smtp、pop3、imap、webmail **服务分布在不同的服务器上，通过** NFS **统一访问同一存储区域。**



这种做法，实现一个简单功能分布，使每一台服务器功能单一，几台服务器各司其职，处理能力大大增强。这种简单的功能分散，实现起来比较简单，只要在不同的服务器其上部署相关服务即可，将存储服务器通过 NFS mount **到每一个服务器上，然后在** DNS **中做相关配置。**

当用户增大到一定数量，并发量会比较大，单台 smtp **可以到** 256 **并发量，最多不会超过** 1000 **并发量。因此** smtp **很容易成为瓶颈。**

3．多机一域，负载均衡，集中存储。



当邮件用户并发量大时，smtp **很容易成为瓶颈，在系统中出现瓶颈的地方可以多增加几台服务器，然后更新** DNS**，通过** DNS **查询解析不同的** smtp **服务器地址，可以使请求平均分配到每一台服务器上。**

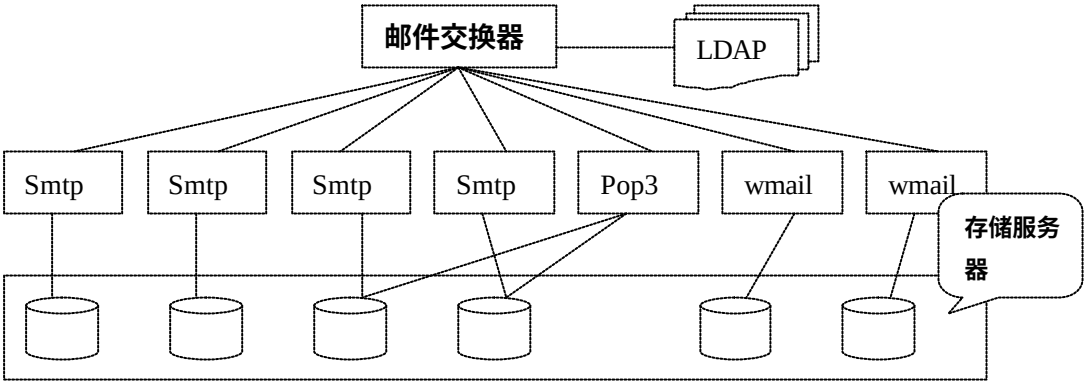
现在许多大容量邮件系统都脱胎于此，特点是部署容易，简单易行。缺点是当访问量增大后，数据访问全部通过 NFS**，也很容易出现磁盘** IO **瓶颈。**

4．多机一域，功能分布，存储分布。

该系统前端(front-end)**是一个邮件交换器和目录服务器** LDAP**，后端是多台独立的存储服务器，通过管理中心调度，将存储服务器通过** NFS **挂在相应的** smtp **服务器或者** pop **服务上。**

在目录服务器上，会保存用户名、密码、smtpserver、popserver、storeserver **等信息，当用户请求过来是，先查询** LDAP**，如果是收信，找到该用户对应的** smtp, **通过直接路由方式连接** smtpserver**，发信；如果是收信，找到对应的** pop,**连接** popserver **下载邮件。**

该架构解决了服务器处理瓶颈、数据存储 IO 瓶颈。缺点是主要的任务分发通过邮件交换器完成。一但邮件交换器出问题，整个系统都将无法运行。



5．多机一域，双层负载均衡，存储分布。

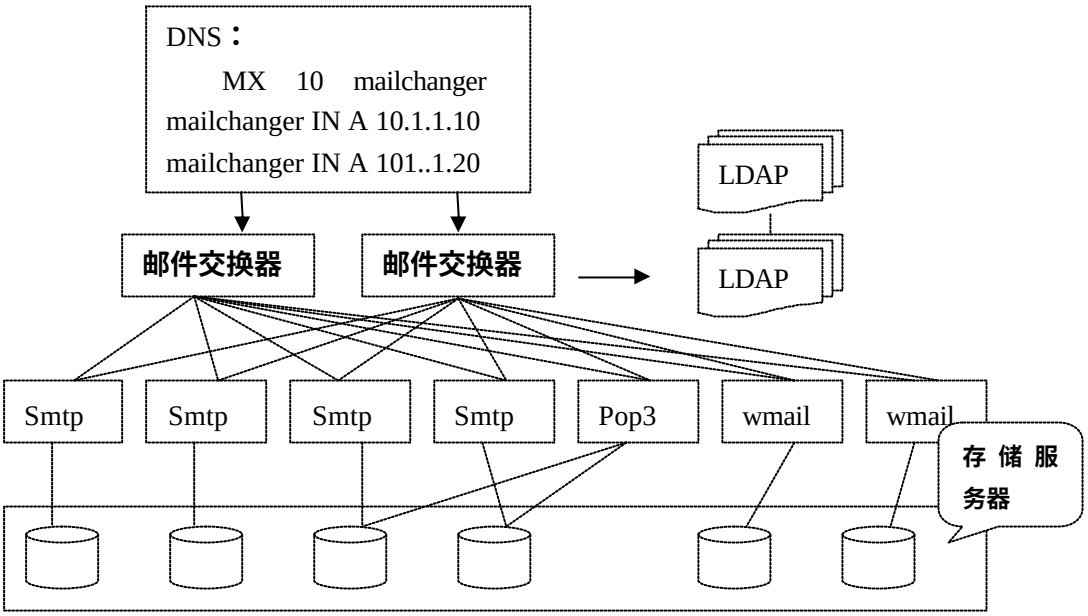
为避免出现单点故障，将 4 改良，增加多台邮件交换器和 LDAP 服务器。邮件交换器通过 DNS 轮询实现负载均衡，LDAP 服务器做成双机热备，任何一台有故障，另一台接替。存储服务器通过 DRBD 实现两两镜像，避免出现存储故障。

这样一套系统，可以支持千万用户级。以 2000 万用户为例，4000 并发量测算，按照处理能力：

Smtp:15 万用户/台，pop：30 万/台

需要：smtp:133 台，pop：66 台，共计约 199 台 PC 服务器。

这个方案的优点是没有瓶颈，可以无限扩充，缺点是需要很多存储服务器，资源上有些浪费。



6．多机一域，邮件功能服务器。

这种方案将以邮件服务器为单位，形成邮件服务器阵列。每台邮件服务器具有完整的邮件服务功能：smtp/pop3/imap/webmail**等。用户认证信息集中存放于**LDAP**服务器，通过**LDAP**查找用户所在邮件服务器。**

系统管理通过集中管控界面可以控制用户数据在服务器之间进行移动、复制、删除。邮件服务器进行两两互备。

设置独立的过滤服务器，主要过滤病毒、垃圾邮件和黄色图片。过滤服务器与邮件服务器之间通过 socket **通讯，所有处理在内存完成，大大加快了处理速度。**

该系统有如下特点：

（1） **这种方案系统造价比较低。以每台服务** 15 **万用户计算，**100 **台服务器可以承载** 1500 **万用户。**100 **台服务器大约** 200 **万－**300 **万人民币左右。**

（2） **结构简单，布置方便，容易维护。**

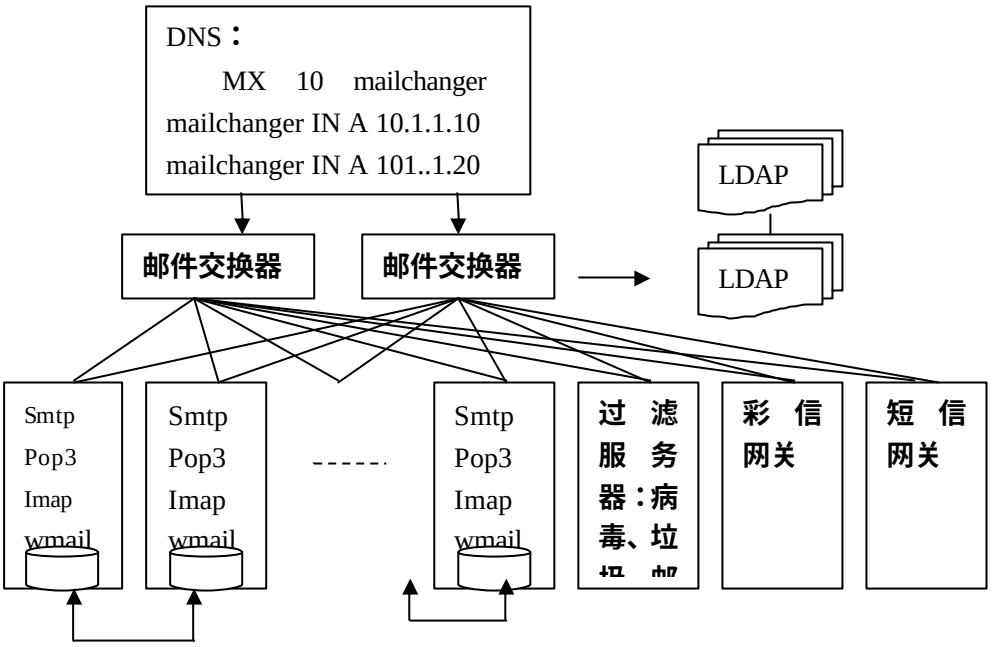
（3） **用户容量与服务器数量保持线性增长，扩容只需要增加服务器即可。**

（4） **帐户在服务器直接动态可调，可迁移。**

（5） **没有瓶颈，应付突发事件的能力强。**

（6） **小面积的故障，不会影响整理系统。**

基于以上优点，我们推荐优先考虑此方案。



功能设计

功能上我们遵循如下设计理念：

1．**集中管控。**

管理一个大型邮件系统，就象管理一台服务器那样简单。在单一管理界面下，通过 web **浏览器即可监控任何环节，完成大部分操作。**

2．**自动化、智能化。**

通过自动装置，如 SMS **监控系统流量、内存、进程、磁盘空间等，设定报警阀值，超过阀值就会自动报警，防范于未然。**

3．**最简化、最优化。**

用简单办法完成简单的事情。简单意味着效率、稳定、可靠。

1．主要功能：

基本功能：	Smtp,pop3,imap,webmail
增值模块：	SMS，MMS，NETFile，Photo，largeFile, Schedule,AntiSPAM，Anti－Virus, Anti-pronography, …
可选模块：	**语音邮件、视频邮件、邮件传真服务**

2．功能清单：

系统管理员	**域管理员**	**普通用户**
1．**统计、审计功能：**	**域用户管理**	**认证登录** WebMail
2．**监控功能：**	**邮件广播**	**用户** session **信息初始化**

3．服务器管理：远程开机、关机、启动服务、停止服务	域级过滤规则设定	显示首页
4．域管理：添加、变更、转移、删除	域级黑名单	保存图标位置
5．过滤服务器管理	用户邮箱大小、附件大小、邮件大小控制	邮件索引显示
6．DNS 管理	功能服务定制：sms/防病毒等。	邮件移动
		邮件排序
		显示页码
		回复信件
		转发信件
		删除信件
		永久删除信件
		显示信件原文
		显示邮件内容
		加入到地址本
		加入到拒收列表
		保存邮件
		pop 取信
		文件夹显示
		文件夹删除
		文件夹添加
		文件夹重命名
		查找邮件
		发邮件页面显示
		立即发送邮件
		定时发送邮件
		保存草稿
		发送短信邮件
		添加附件
		删除附件
		个人地址本显示
		团体地址本显示
		添加个人地址本
		删除个人地址本
		修改个人地址本
		添加团体地址本
		删除团体地址本
		修改团体地址本
		地址本排序
		导出地址本
		导入地址本

		显示自动转发
		修改自动转发
		显示自动回复
		修改自动回复
		显示签名档
		删除签名档
		增加签名档
		修改签名档
		显示 pop 收信设置
		pop 收信设置修改
		pop 收信设置增加
		pop 收信设置删除
		修改密码
		显示参数设置
		参数设置修改
		显示多风格设置
		多风格设置修改
		显示反垃圾级别设置
		反垃圾级别设置
		显示杀毒状态
		修改杀毒状态
		显示过滤设置
		邮件过滤设置添加
		邮件过滤设置修改
		邮件过滤设置删除
		显示拒收设置
		拒收设置修改
		外挂一次认证
		session 维护进程

3．后端统计

统计类别	统计项目	描述
系统资源	负载情况	5 分钟采样一次，自动绘制统计图
	CPU 使用情况	User,system,nice and idel
	内存	Totoal，used，free
	交换分区	Used ,free,cached,shared
	硬盘	Total，used,free, inode
	网络状况	收到字节数，发送字节数
邮件总量统计	邮件数量，总流量	统计每域、每用户收发邮件总量和字节数。

投递失败统计	统计投递失败的记录	按照每域、每用户统计
隔离邮件统计	邮件感染病毒，或者认定为垃圾邮件的	按照每域、每用户统计
短信统计	短信发送数统计	按照每域、每用户统计

4．后端服务器管控功能

项	功能	描述
服务器节点远程管理	添加	向集群中添加一台服务器
	除去	从集群中拿掉一台服务器
	重启	远程重启服务器
	关机	远程关机
服务远程管理	启动	远程启动某服务器服务进程
	停止	远程停止服务器进程
监控、报警	设定监控项目，过载报警	发送 SMS 到管理员手机
负载均衡管理	LVS，RR-DNS, LDAP	负载均衡远程调控
Tomcat 集群管理	添加节点、删除、更新	管理 webmail 服务器
Session 服务器管理	集中管理用户 session	