

## · 专题研究 ·

## 试论流媒体网站的建设\*

张文勇 伍 艺

(湖南图书馆 湖南长沙 410011)

**〔摘要〕** 随着 Internet 带宽的不断增加,多媒体技术在互联网的应用越来越广泛。多媒体信息与传统的文字、文件信息在处理上有很大的区别,多媒体信息一般以流的方式在网上传输。本文将讨论如何建设一个流畅、稳定、安全的流媒体网站。

**〔关键词〕** 流媒体 流媒体服务器 负载均衡

**〔分类号〕** G250.76

多媒体信息由于所占空间比较大,因此一般不能下载后再进行播放,而是以流的方式边下载边播放,这就是所谓的流媒体技术。流技术简单来说就是把连续的影像和声音信息经过压缩处理后放在网站服务器上,让用户一边下载,一边观看、收听,不需要等整个压缩文件下载到自己的机器上后才可以观看的网络传输技术。流技术先在用户端的电脑上创建一个缓冲区,于播放前预先下载一段资料作为缓冲,当网络实际连线速度小于播放所耗用资料的速度时,播放程序就会取用这一小段缓冲区内的资料,以避免播放的中断,也使得播放品质得以维持。有人形象地将流媒体比作淋浴,水是边流边洗,而不是等浴缸里放满了水以后再洗。与单纯的下载方式相比,这种对多媒体文件边下载边播放的流式传输方式不仅使启动延时大幅度地缩短,而且对系统缓存容量的需求也大大降低,是目前 Internet 常用的视频服务平台应用技术。本文将从流媒体网络的建设、流媒体文件格式、流媒体服务器的构建来讨论如何建设一个流畅、稳定、安全的流媒体网站。最后简单介绍一下湖南图书馆的流媒体网站系统。

## 1 网络的建设

流媒体 WEB 网站的建立首先需要建立一个局域网,局域网的模式有很多: FDDI、ATM、快速以太网等,其中最流行、最常用的是快速以太网。快速以太网的建立技术十分成熟,不再赘叙。在局域网的基础上选择一个合适的服务器添加一个 WEB 服务器软件就可以建立一个 Internet,如采用 Windows NT 系列网络操作系统 + IIS、采用 UNIX、LINUX 网络操作系统 + Apache。若有一条专线与 Internet 相连,那么这个服务器就是 Internet 中的一个网站。

另外,流媒体传输网络是适合多媒体传输协议甚至是实时传输协议的网路。流媒体在因特网上的传输必然涉及到网络传输协议,这是制约流媒体性能的最重要的因素。为了保证对网络拥塞、时延和抖动极其敏感的流媒体业务在面向无连接的 IP 网络中的服务质量,必须采用合适的协议,其中包括 Internet 本身的多媒体传输协议及一些实时流式传输协议等。

## 2 流媒体文件格式

流媒体文件是将一个资料(动画、影音等)分段传送,用户不必等待整个内容传送完毕就可以观看到即时、连续的内容,甚至可以随时暂停、快进、快倒,由于不同公司开发的文件格式不同,传送的方式也有差别,以下是目前流行的几种流媒体文件格式。

### (1) Real Networks 公司的 RM 视频影像格式和 RA 音频格式

RA 格式是 Real Networks 公司开发的一种新型流式音频 Real Audio 文件格式, RM 格式则是流式视频 Real Video 文件格式,主要用

来在低速率的网络上实时传输活动视频影像,可以根据网络数据传输速率的不同而采用不同的压缩比,在数据传输过程中边下载边播放视频影像,从而实现影像数据的实时传送和播放,客户端通过 Real Player 播放器进行播放。

### (2) Microsoft 公司的 ASF 格式

微软公司也推出了一整套包括流媒体制作(Media Tools)、发布(Media Server)和播放软件(Media Player)的信息流式播放方案 Microsoft Media Technology。ASF 是一种包含音频、视频、图像以及控制命令、脚本等多媒体信息在内的数据格式,通过分成一个个的网络数据包在 Internet 上传输,实现流式多媒体内容发布。ASF 支持任意的压缩/解压缩编码方式,并可以使用任何一种底层网络传输协议,具有很大的灵活性。微软公司将 Microsoft Media 技术捆绑在 Windows 中,具有方便、先进、集成、低费用等特点,此外,ASF 还将被用作 Windows 版本中多媒体内容的标准文件格式,这无疑将对 Internet 特别是对流式技术的应用和发展产生重大影响。

### (3) Apple 公司的 QT 格式

Quick Time Movie 的 QT 格式是 Apple 公司开发的一种音频、视频文件格式,用于保存视频和音频信息,具有先进的视频和音频功能,它几乎支持所有主流的个人计算机平台和各种格式的静态图像文件、视频和动画格式,是创建 3D 动画、实时效果、虚拟现实、A/V 和其他数字流媒体的重要基础。此外, MPEG-4 被认为是对抗微软向媒体市场进军的一种技术, MPEG-4 将很大的数字文件压缩成较小的文件,以便在互联网上进行传输,并提供交互功能,使影视节目具有。

要保证流媒体文件在播放时能流畅、稳定,就必须选择合适的流媒体文件格式和压缩技术。现在流行的视频压缩格式是采用 Helix Producer 压缩的支持动态码流技术的 Real 格式和采用 DIVX、XVID、MPEG-4 等技术压缩的 Windows Media 格式。

## 3 流媒体服务器的构建

流媒体系统中的媒体服务器用于存放和控制流媒体的数据。随着流媒体规模的扩大,流媒体服务器的性能成为制约流媒体服务扩展能力的重要因素。流媒体服务器性能的关键指标是流输出能力和能同时支持的并发请求数量。影响流媒体服务器性能的因素很多,包括 CPU 能力、I/O 总线、存储带宽等。

从技术上讲,要提供稳定的服务,在构建流媒体服务器时需要做到以下几点:

第一、采用专业的磁盘阵列。视频节目库容量大,更新快,读取

\* 本文系湖南省社会科学重点基金项目“湖南农村文化信息服务创新与保障机制研究”(项目编号:06ZD08)研究成果之一。

频繁,对数据吞吐带宽要求高,因此采用带有电源备份、RAID 技术,支持热插拔的专业磁盘阵列是有必要的。

第二、重视磁盘阵列的规划和维护。磁盘阵列的容量都为 TB 级,一般具有 2TB-5TB 的磁盘空间,可以提供 5000 小时以上的节目内容。建议将整个空间划分为 2-3 个或更多分区,每个分区 1TB 左右。先用一个分区,满了再用第二个分区,在整个磁盘空间使用率超过 70% 时,就必须对最早使用的分区进行文件删除整理,同时进行磁盘碎片整理。对一个 TB 级空间的硬盘进行碎片整理是非常耗时和占用 CPU 资源的工作,建议采用专业的软件安排在网络访问量小的时候进行,比如凌晨 1、2 点。

第三、在网站架构上,要实现用户数据、视频资料描述和视频内容分开存储。这包含两个层面,一是在软件上用户数据和视频资料描述单独存放在数据库中,在数据库中记录视频资料的物理存放地址,用户访问时系统会自动从数据库中读取相关资料并动态生成连接地址。二是在服务器构建上,网站服务器和点播服务器分离,网站服务器用来提供视频内容介绍,用户认证,计费等工作,而实际提供视频内容服务则由点播服务器来完成。点播服务器可以有数台,用来提供不同的内容或者实现负载均衡。对于大型网站,建立单独的数据库服务器也是值得考虑的。

第四、做好数据库的备份。用户数据和视频资料数据库最好每天做一次增量备份,备份的数据要存放在不同位置的服务器上。

第五、实现点播服务器的负载均衡。单台服务器的容量和性能毕竟是有限的,服务器集群调度策略可以将多个服务器并行使用来为更多的用户服务。集群系统通过客户端和多个流服务器之间的调度与网管服务器实现。在负载均衡应用中,集中式调度是现在使用较多的一种集群管理模式,系统主要由三部分组成:调度器,位于服务集群的最前端,根据服务集群的负载情况将到来的服务请求转发到后台真实服务器上;服务器池,位于调度器的后端,由一组为用户提供具体服务的 Server 组成,对用户不可见,用户通过调度器与服务器建立 C/S 关系;存储池,后台服务器的一个共享存储空间,使得集群中的各台服务器能够提供相同的内容服务。

客户端将用户的请求先传送到调度器,调度器则通过网络管理协议对各节点进行监控,动态获取各节点的状态信息,随时对各节点的性能进行评估,根据当前各节点的负载情况,通过一定的智能算法选择转发至最合适的服务器,从而达到集群中各个服务器间的动态负载均衡。这种方法通过增加新的服务结点大大扩充了服务器支持的并发流数量,克服了单台服务器处理能力的局限性,使得整个流媒体服务的性能和容量得到了无限制的提升。

对于用户来说,整个集群服务器是透明的,也就是不存在每个单独服务器,因此必须实现 IP 隐藏,使得用户无法观察到每个单独服务器的 IP,这样可以更好地保证整个集群的负载均衡,因为这样集群的负载是由调度器来统一执行,用户无法自己指定,而且也可以更好地保证系统的安全性。

调度服务器上的网络管理平台通过特定的网管协议来实现各个服务器之间的协调管理,可以选用 SNMP 作为管理协议,并针对流媒体服务的特点开发了一套流媒体服务集群管理系统。一个 SNMP 网络管理系统从逻辑上可抽象为四个部分:被管代理,网络管理工作站,网络管理协议,网络管理信息库。网管工作站负责整个网络的管理,监控被管设备状态从而保证网络中的各个设备正常运行。被管代理是驻留在被管理设备上的一个进程,负责收集组织被管设备状态信息,响应管理端访问管理信息库的请求,另外当系统出现问题的時候可以主动向管理端发出通告信息。MIB 是管理信息的组织形式,呈树形结构。管理协议则定义了网管工作站和被管设备之间信息交换的接口,双方以约定的格式进行消息传递。

#### 4 流媒体存储结构

传统存储的体系结构有两种:集中式和分布式。集中存储可实现服务的负载均衡,由于流媒体服务器间的数据都是共享且统一的,当发生热点繁忙时,所有流媒体服务器都可为其提供服务,分减压力,而不像分布式的存储会出现热点繁忙,没有热点内容的存储出现空闲这种不均匀情况。但集中结构存在容易产生 I/O 瓶颈、容量扩展性差、性能不可扩展、容易发生单点故障等严重缺陷。而分布式结构则在每台服务器上提供文件共享服务,由应用层来实现媒体资源数据在各个服务器集群之间的迁移,从而比较好地解决了集中存储的 I/O 瓶颈问题,但分布式的存储没有负载均衡,利用率相对较低,重复数据大量存在,且份数多,备份、恢复、远程容灾比集中存储实现成本高。

传统的集中和分布都存在不同的问题,随着技术的不断进步,产生了一种新的存储结构“分布式的集中存储”。分布式集中存储相对于传统存储系统来说拥有如下优点:采用统一的全局命名空间,支持文件级共享,且采用分布式存储结构,能实现高聚合 I/O 带宽及跨 Linux 平台和 Windows 平台的文件共享,还能够线性扩展 I/O 带宽,拥有良好的系统负载平稳性,并能够动态扩展存储容量,实现成本可控。该架构能较好地满足网络视频对存储系统的需求。

下面就以湖南图书馆的视频点播系统为例,来介绍流媒体网站的建设。

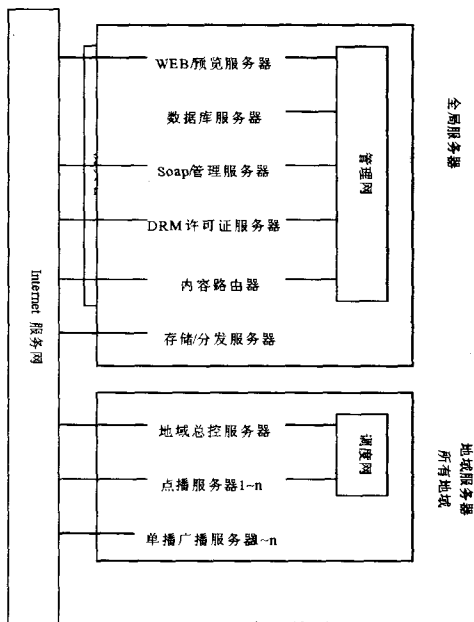


图 1 系统总体结构图

湖南图书馆于 2007 年启动了全省的流媒体视频点播系统项目,并建立了以湖南图书馆为省中心,浏阳市图书馆、宁乡县图书馆、长沙县图书馆、望城县图书馆、衡山县图书馆为分中心的流媒体视频服务器试验系统。

该系统流媒体视频服务器由多个高性能的节点机组成,这些节点机通过内联网相互连接。系统中包括高性能的 Gigabit 以太网和连接存储设备的光纤通道等通用网络系统,系统具有强大 I/O 能力和良好的并发处理能力。为了支持大量的流媒体操作,采用了两级存储结构。采用分布式或集中分布式存储结构,即每个节点机独立内置 1-6 个热插拔 SCSI 节目硬盘阵列,由各节点所有硬盘构成大容量存储器。也可将节点分成多组,每个节点内不配置节目硬盘,每组节点通过 SCSI 接口或 FC 光纤通道共享一个外置 RAID 磁盘阵列,多组节点之间实现分布式存储方式,这种方式对多节点大规模流媒体服务系统而言,可减少节目硬盘的数量。采用内联网结构实现节点之间的互连,用于节点之间的节目调度和动态负载均衡。该流

媒体系统中节点机配置高性能的 Intel 处理器,系统总线可达到 400MHz。两路交叉访问的高速 DDR 内存,容量为 1GB,可扩展至 16GB。节点机内置 1~6 个高速 SCSI 接口的视频节目硬盘,双通道 Ultra160 SCSI 控制器。节点机提供 5 个以上 PCI 扩展槽。这些扩展槽将用于增加节点机的功能,提高系统性能。节点机可以支持连接多种标准 PCI 接口的适配器卡,包括 Gigabit Ethernet, 10/100BaseT, MPEG2 编码器, MPEG2 解码器, FC 卡, SCSI 设备等。

湖南图书馆流媒体服务器采用分布式网络结构,整个系统统一认证、统一计费。全局服务器放在湖南图书馆中心机房,而地域服务器分布在中心机房和各个边缘地区(长沙县、望城县、宁乡县、浏阳市和衡山县)。全局服务器和地域服务器之间通过外网连接和通信(图 1)。

全局服务器中数据库服务器、WEB 服务器、Soap 服务器、域控制服务器、预览服务器,是整个系统的管理控制中心。数据库服务器、WEB 服务器、Soap 服务器均采用了群集配置的方式,防火墙加强系统的安全性。存储服务器既可提供流媒体服务同时是整个系统流媒体文件的存储仓库,是所有节目的备份。各分中心 VOD 节点机之间通过交换机组成高速内联网,实现流媒体内容的动态负载均衡(图 2)。

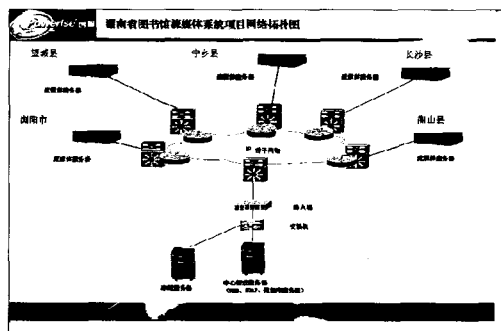


图 2 系统网络拓扑结构图

#### 参考文献:

1. Steve Mack. 流媒体宝典. 北京:电子工业出版社,2003
2. 廖勇,周德松. 流媒体技术入门与提高. 北京:国防工业出版社,2005
3. 钟玉琢,向哲,沈洪. 流媒体和视频服务器. 北京:清华大学出版社,2003

### On the Construction of Media Stream Website

Zhang Wenyong Wu Yi (Hunan Library)

[Abstract] With the continuous increase of the bandwidth of Internet, the multi-media technique is more and more extensive in the Internet application. The multi-media information and traditional writing, the document informations have the very big differentiation on the processing, the multi-media information is generally to deliver on the net by flow. This paper discussed how to construct a safe, free and stable flow medium website.

[作者简介] 张文勇,男,湖南图书馆自动化发展部系统管理员,网络工程师;伍艺,女,湖南图书馆副馆长,副研究馆员。

(上接 82 页)

表 6 图书馆学核心期刊 2005 年发文被 2006 年引用的相对 H 指数

排序	刊名	H 指数	发文量	相对指数
1	大学图书馆学报	6	129	0.0465
2	中国图书馆学报	5	137	0.0365
3	图书馆	7	215	0.0326
4	图书情报知识	4	180	0.0222
5	图书馆工作与研究	3	203	0.0148
6	图书馆杂志	4	340	0.0118
7	图书与情报	2	172	0.0116
8	现代图书情报技术	3	280	0.0107
9	图书馆理论与实践	3	339	0.0088
10	图书馆论坛	4	512	0.0078
11	图书馆工作	3	444	0.0068

由表 5.6 可知,《图书馆》在两年的相对 H 指数排序中均排名第 3,与表 2 的影响因子排序是一样的,说明该刊在图书馆学专业已具有相当的学术地位,其学术影响力之大是显而易见的。期刊的指数和相对 H 指数是对总被引频次、影响因子等传统计量指标的很好补充和发展,也是对期刊整体评价与论文评价相结合的一种探索、实践和创新,目前已得到国内外文献情报界专家和同行的广泛关注、重视和认可。

#### 3 结论

通过从载文被引篇数、被引频次、影响因子、即年指标、H 指数以及相对 H 指数等多个方面对《图书馆》在我国图书馆

学核心期刊的排序分析,结果表明,10 年来《图书馆》在编辑人员的辛勤劳动和广大作者的积极支持下,已发展成为我国图书馆学研究领域重要的信息源,近年来其办刊质量和水平有显著提高,已形成自己独特的办刊风格和特色,拥有一批读者和作者群,该刊文献计量学的各项指标在该学科领域的排序均列前茅,其学术影响力不断得到提升,受到图书馆界的瞩目,成为图书馆学领域专业人员关注的重要学术刊物之一。在调查中发现该刊的自引率较高,说明同一主题的论文具有一定的研究深度和较高水平,但要注意抓住本学科研究中近期、中期的热点问题,以及交叉学科的热点问题,提高期刊的他引率,进一步扩大其在学术领域的知名度和影响力。

(来稿时间:2007 年 10 月)

#### 参考文献:

1. 戴龙基,蔡蓉华. 中文核心期刊要目总览:2006 年版. 北京:北京大学出版社,2004
2. 何小清.《图书馆》及图书馆学期刊学术水平与影响评价——图书馆学期刊 1998 年被引排序与分析. 图书馆,2001(6):23-7
3. 程刚,邹志仁.《大学图书馆学报》被引分析与研究. 大学图书馆学报,2001(2):81-4
4. 曹海燕,陈青.《图书馆》计量与评价. 图书馆,2004(1):69-72
5. 姜春林等. H 指数和 G 指数——期刊学术影响力评价的新指标. 图书情报工作,2006,50(12):63-5,104

### Citation Analysis of Library by CSSCI in 2006

Yang Zhi (Library of Hunan Normal University)

[Abstract] The database CSSCI 2006 is used for evaluating our country library science core journal, and analysing "Library" in bibliometric.

[作者简介] 杨志,女,湖南师范大学图书馆馆员,发表论文 10 余篇。