

# 从系统运维角度看NSTL十年发展历程

□ 郝春云 毕军荣 / 中国科学技术信息研究所 北京 100038

**摘要:** NSTL网络服务系统的茁壮成长,凝聚了NSTL人十年奋斗智慧和勇气,体现了国家发展历程,伴随系统变化体现在运行维护工作的诸多变化,也印证了NSTL十年发展、壮大的过程和结晶,引出了前景的伟大和运维工作的艰巨。

**关键词:** NSTL网络服务系统,系统运维,发展历程

**DOI:** 10.3772/j.issn.1673-2286.2010.10.008

## 1 NSTL网络服务系统发展三阶段

NSTL网络服务系统于2000年12月26日正式开通运行,10年来,NSTL网络服务系统不断发展,数据量大幅增长,系统功能不断扩展和完善,用户数量稳定增长,系统服务量连年攀升,在科技界文献信息服务领域赢得了良好的口碑。

回顾10年的发展历程,从2000年至今,网络服务系统经历了三次大的升级改造:2002年二期系统建设、2005年二期系统改版、2007-2009年三期系统建设。服务模式由单一中心站转为以中心站为核心,东西南北镜像站为辅助,多方位服务站为节点的立体服务网模式,经过多年的运行服务,显示其独特的风格,适应不同区域各类用户的需求,是中国近期较为成功的范例。不断探求更适应我国科技文献服务体系和文献保障体系建设,是NSTL重要的发展理念,长期的研究与实践也证实了这种思路的正确性。

### 1.1 一期系统

为实现NSTL网络服务系统建

设目标——“分布加工数据,集中建库,集中检索,分布服务”,2000年一期系统选用典型的分布式系统架构。主要功能是二次文献查询和全文提供。

系统采用C/S结构,由一个中心站点和8个分中心站点组成,中心站点是由Web服务器、GTS应用层网管、Web Session管理器、HTML模板系统、用户管理应用服务器、文献检索应用服务器、原文订购应用服务器、帐务管理子系统等组成。Oracle用于账务等业务管理和流程管理,Trip全文数据库系统用于二次文献集中检索服务,前端采用Java Servlet作为表现层,Web服务器使用Apache,ISMQ中间件为各业务组件提供连接与集成。中心站点和分中心站点通过Internet网络互联,并通过ISMQ实现相互通信。分中心站点接收中心站点传递来的全文订购请求信息,处理后直接向用户提供全文,实现分布式全文文献服务。

系统构建在2台Sun/E3500(双CPU 400MHZ)对称式高可靠性系统平台上,存储系统选用Sun A5100,集中存贮二次文献数据约200G。

### 1.2 二期系统

为适应服务内容和数据不断扩展的需求,2002年对一期系统进行了根本性的改变。

二期系统采用B/S三层结构。系统主要功能在保留一期的文献检索与全文提供外,增加了网络版全文数据库、期刊分类目次浏览、专题信息服务、联机公共目录查询、专家咨询等服务功能。在系统实现上更注重人性化设计和个性化服务,如开通SDI功能便于资源推送服务,设置自助中心方便用户自我管理,开通网上支付功能方便用户交费等。用户类型也从单一普通用户扩展到集团用户和VIP用户。2005年二期改版进一步强化了资源揭示,提高网站的规范性设计,对系统进行了充分的优化,流程处理更趋合理,使系统综合服务能力明显提高,为后续开展各类服务奠定了良好的基础,如西部用户优惠政策的实现、试用卡和充值卡的启用、代查代借功能的实现等。

系统选用B/S结构建立一个中心站系统,二次文献数据和业务数据全部存储在中心系统,规避了一期系统中心和分中心之间业务数据的同步问题。Trip用于文献数据、

用户数据和业务管理信息等全部数据的管理和检索服务；Weblogic为业务中间件和Web服务器，应用系统与Trip数据库的连接通过自行开发连接池实现。

系统选用两台ALPHA ES45构建动态冗余体系，提供高可靠性服务，存储系统采用FC-SAN方式，数据存储于MA8000阵列上（最初为500G，2004年扩容为1T）。

随着国内用户需求的变化，2003年以后，服务系统调整为多层结构，服务模式由绝对集中变为相对分布。在中心站之下，建设多个镜像系统，共享一套服务体系，在中心站和镜像站之下，建设多个服务站，共享其数据和服务。2003年3月开始筹建NSTL镜像站系统，2003年11月兰州镜像站建成开通服务，2006年武汉服务站建成开通服务，不仅扩大了服务范围，还缓解了当时西部地区因网络等基础条件不足而影响科技文献信息快速获取的现状。

### 1.3 三期系统

随着服务内容和服务理念的不断变化，为解决子系统互不开放、自成体系、缺乏整体服务，以及软、硬平台存在影响效率和稳定等诸多问题，在借鉴二期系统良好功能和服务理念的基础上，2008年开始研究三期系统建设并进行了自主研发。

三期系统仍采用B/S三层结构，遵循J2EE标准，实行组件模式，使系统更具开放性。与二期相比，在系统架构和工作流管理上作了大幅调整。增加了期刊浏览、数字资源发现等协助用户提高信息发现能力的各种功能，在数据管理与利用方面增加了知识库及仓储的应

用管理，以数据仓储为中心，形成了结构化的数字资源加工、处理、发布与同步体系，也为资源的长期保存奠定了基础。

系统选用2台前端交换机和8台服务器构建访问负载均衡、统一WEB服务门户、应用服务器群和数据库群的三层服务体系，并通过各类接口连接镜像站、服务站以及第三方服务系统，形成多层服务体系。系统依前后台业务划分区域和分布部署，在优化服务、提高业务效率和加强数据和系统安全防护能力上有很大改善。系统采用Oracle进行用户管理、业务流程管理、信息统计以及仓储数据管理，Trip用于二次文献检索服务，Weblogic为业务中间件和Web服务器。存储系统采用FC SAN和IP SAN两种方式：FC SAN用于文献数据存储以提高检索性能；IP SAN具有扩充性好的特点，用于其他数据包括仓储数据存储。

## 2 NSTL系统服务指标的变化

自2000年系统开通以来，各类服务统计结果的逐年变化客观地反映出NSTL的发展。

### 2.1 用户数量持续增长

由于NSTL服务系统是一个开放式系统，并且一直免费为Internet用户提供二次文献检索浏览而无需注册，因此访问网站获取信息服务的用户数量非常大，可在下面的检索访问量和二次文献浏览量中体现。从2001年到2009年底，注册用户和缴费用户数也逐年增长，见表1。

从表中数据可以看出，经过了最初的快速增长，NSTL的交费用户数每年都保持一个相对稳定的增长量。服务站的建立和万方接口的加入，在各自服务区域自成体系建立用户群，共享中心站文献服务，但不在中心站用户统计之内，因此从某种意义上分流了一部分NSTL中心网络服务系统的用户，但实际上是扩大了服务范围和用户群。

### 2.2 系统服务量变化

表1 NSTL网络服务系统历年用户增长情况

年份	新增交费用户数	累计交费用户数	新增注册用户数	累计注册用户数
2001	3125	3125		
2002	4025	7150	7146	7146
2003	1967	9117	34048	41194
2004	1790	10907	27742	68936
2005	2182	13089	36047	104983
2006	1865	14954	50252	155235
2007	2343	17297	54462	210137
2008	2502	19799	55353	266143
2009	2601	22400	46589	311950

注：此统计只包括NSTL中心站和镜像站的用户

NSTL的系统服务量主要可以通过几个指标衡量,包括全文服务量、检索访问量、文献浏览量等。表2、表3、表4详细给出了历年的服务量变化情况。

表2 NSTL网络服务系统历年全文服务量

年份	全文服务量
2001	25026
2002	122068
2003	113448
2004	109575
2005	125924
2006	152157
2007	259516
2008	337285
2009	366225
总计	1611224

表3 NSTL网络服务系统历年检索访问量

年份	检索访问次数
2001	4735120
2002	11471201
2003	16668200
2004	19199451
2005	28256223
2006	41678335
2007	60462816
2008	61368572
2009	48057442
总计	291897360

从上述统计可以看出,相对于用户增长,检索访问的增长速度非常快,全文服务的增长则介于两者之间,文献浏览的情况非常随机。这里存在来自网络上恶意下载二次

表4 NSTL网络服务系统二次文献浏览量

年份	浏览记录数
2005	4271695
2006	2442922
2007	2324367
2008	6398658
2009	8643131

文献数据的情况。

### 2.3 系统数据量变化

NSTL网络服务系统在开通之初,本地数据库总量为177万条记录,包括8种文献类型(中文会议、中文学位论文、中文期刊、中文图书、西文会议、西文期刊、西文科技报告、西文科技图书)。经过10年的发展,本地数据量已逾7000万条,文献类型42种,另外开拓了500种网络版期刊全文数据库、3000余种核心期刊引文数据约4800万条、1122种期刊的回溯数据约350万条等,提供服务的数据总量超过一亿。

表5 NSTL网络服务系统本地数据库增长情况

年份	数据增长量	数据总量
2001	4395209	6168924
2002	5776277	11945201
2003	15284525	27229726
2004	2704357	29934083
2005	4931906	34865989
2006	5103160	39969149
2007	12917606	52886755
2008	7982490	60869245
2009	8924245	69793490

## 2.4 系统功能和服务模式的发展变化

NSTL网络服务系统开通之初仅提供文献检索和全文提供服务,在历年的发展,服务功能逐渐增加,服务模式越来越多样化,其独特的服务模式逐渐被国内同行认可和采用。表6简要描述了NSTL历年系统功能和服务的发展变化。

## 3 NSTL系统运维工作的变化

NSTL网络服务系统的技术支持工作在NSTL网络服务系统开通前半年就开始了,十年发展不断对运维技术和管理提出新的要求。

### 3.1 系统发展带来工作量的持续增长

无论是数据的大量增长,功能的不断增加,子系统、镜像站、服务站的陆续设立等等无一不带来运维工作量的巨幅增加。运维对象从最初的1个中心系统和9个分中心系统,发展到1个中心系统、8个镜像系统、十几个服务站系统、17个子系统以及其他相关的NAS备份系统和OA系统等,并且各应用系统比早期的系统复杂得多。因此,运维队伍也从最初的2、3人身兼数职,既负责系统、网络维护、数据库维护,也面向用户承担用户服务、宣传培训等工作,演变为专业分解,用户服务工作移出,各业务定岗定责,A、B角配合,引入外围专业服务,形成相对稳定的运维队伍。

### 3.2 异构平台带来的维护多样性

表6 NSTL网络服务系统功能和服务模式发展变化

年份	新增功能或服务	备注
2000	一期系统上线：文献检索与全文提供。	12月26日开通服务,177万条数据
2001	新增Science中国镜像站。	2005年终止
2002	开通英国MANEY公司和英国皇家学会全文期刊系统中国镜像站点；二期上线：新增目次浏览、目录查询、预印本、参考咨询、网上支付、SDI、后台系统管理（如：在线提交数据、在线数据装载等）。	2005年镜像站转移至清华
2003	开通兰州、成都镜像站；新增七国两组织专利、中外标准、外文学位论文、机械工业科学技术文献、计量基准等数据库；新增集团用户功能；开通西部用户优惠服务。	12月25日二期上线
2004	新增昆明、西安、哈尔滨镜像站；开通热门门户系统；新增小语种数据（日文期刊、俄文期刊）、全球科技投入要览；新增中国标准、国外标准、计量检定规程的订购功能、目次到二次文献的链接功能、科技信息资源指南；西部用户优惠方式由双倍加费改为半价计费；新增NAS备份系统。	
2005	新增南京镜像站；二期系统改版上线，提供更人性化的服务；新增试用卡发卡功能、全文文献、分类导航和机构导航、专利订购功能、全球科技人才要览、追加多馆藏数据功能；新增方正图书系统；开通800一站式用户服务。	
2006	新增郑州、杭州镜像站；开通武汉服务站；新增万方（中国科研机构、中国科技信息机构、中国高等院校）数据库。	
2007	新增代查代借功能；新增天津、重庆、南宁、广州、宁波、青岛、乌鲁木齐、西宁、苏州服务站；新增文集汇编数据库；新增专业平台数据抽取服务。	
2008	新增合肥、深圳、长春、泰达服务站；调整页面数据传输模式，提高数据安全。	
2009	新增服务站手工同步功能。	
2010	三期系统上线，新增文献浏览导航功能、全文自助下载功能、集团用户自我管理功能、知识库功能、第三方接口、数据仓储等。	

由于系统建设是在需求不断变化下逐步扩大起来的，从2002年的二期系统开始，数据类型和功能扩展很快，新增了很多功能、子系统、镜像站和服务站系统，这些系统的运行平台及系统架构并不统一，从硬件环境、操作系统、数据库系统到应用软件各不相同，对运维技术提出更高的要求。主要的

硬件平台包括HP ALPHA服务器、SUN 服务器、HP9000服务器、HP PC服务器、IBM PC服务器、Dell服务器，以及相应的阵列系统；存储系统有NAS，也有SAN。操作系统有SOLARIS、TRU64、LINUX、HP-UX和Windows的各种版本，Web服务器产品包括Weblogic、Tomcat及Apache、IIS等，数据库产品涉及

Trip、Oracle、MySQL、SQL Server等。这些系统相对独立，种类繁多，标准规范不统一，各系统接口不一致，带来维护工作的高成本，对运维技术和管理是一个相当大的挑战。

### 3.3 数据质量和服务质量对运维工作的高要求

随着NSTL的不断发展，对数据质量和服务质量提出更高的要求。为保证数据质量，中心发布了NSTL数据加工规范，建立了质检小组，开发了不同级别的专用质检软件，从规范、技术和组织管理等不同角度、各个流程，加强数据质量的控制。最后一道防线是数据入库前的质检，因此数据加载过程由最初先装实验库后入正式库变为：下载数据到本机质检—入实验库—入正式库—追加多馆藏数据—生成统计报表等。在保证数据质量的前提下，要检查目次加工与二次文献数据加工的时差，判定数据的时效性是否符合要求。另外，二期、三期系统升级均要对文献数据、用户数据等进行大规模的清洗、整合、迁移，前提是要对其数据进行完整的抽取，而不影响原有系统的正常服务。

全文服务的时效性要求由最初的2个工作日，变为48小时、1个工作日、24小时，这对系统正常运行时间的要求更高了，运维工作必须保证系统7\*24小时不间断服务。

### 3.4 服务器老化带来运维问题

与功能越来越多、数据量越来越大、用户访问连年攀升的情况相

比,前期系统建设的一些服务器,包括镜像站,设备早已逐渐进入故障期,系统和数据库的版本更新不能优良匹配,加大运行风险,同时空余磁盘空间也逐渐变小,数据工作区紧张,系统环境变差。这类问题此起彼伏,为满足运行要求,需加强系统监控,提高监控技术手段,才能及时排除故障。

### 3.5 其他需求对运维工作的影响

近年来,为配合NSTL开展的各项研究课题对NSTL网络服务系统各类数据的需求,专门开发程序对文献数据、其他数据和日志信息等数据进行定题抽取或统计,因系统结构所限,大量抽取工作对系统运行有影响,要避免对正常服务产生不良影响,就加大了该项工作的难度。

## 4 系统运维工作的展望

### 4.1 运维对象的升级减轻了性能维护的压力

三期系统上线以后,硬件环境得到很大提升,更合理的系统架构

提升了系统稳定性,负载均衡,应用并行部署提高了系统并发处理能力,新系统在稳定性方面的特性能相对减轻部分运维工作量。同时,大磁盘阵列的使用,避免每日进行日志备份和数据转移等工作。

### 4.2 新的运维工具将增强运维便利

三期系统将针对实际应用开发一系列后台运维管理工具,包括服务监控、系统调优、数据备份和各类数据抽取统计等内容,将极大方便运维人员的工作,使得运维工作更加科学合理。

### 4.3 运维服务对象更多

三期系统变封闭的服务为开放的服务,增加多种开放接口,支持第三方机构调用NSTL资源和服务,支持特殊机构的文献需求服务,因此将来运维工作的对象除了普通用户,NSTL各服务单位、各镜像站、服务站、各课题组等,还将包括各类接口用户。

### 4.4 加强运维队伍建设

随着三期系统上线服务,面

对差异较大的系统,截然不同的工作流程,对现有维护人员是一项新的挑战,对整个维护队伍的建设也提出了新要求。因此,需要调整理念,建立正确的组织管理机制,重新梳理业务流程,比如二期数据装载变为数据收割、数据发布、数据自动加载等流程,要针对各环节统筹考虑现有人员工作分工,并尽快引进新的运维人员,以补充仓储系统、知识库系统、回溯系统和引文系统等新增内容的运行维护工作。同时,必须加强和完善各项运维工作规范的制定和落实,保障运维工作有条不紊地开展。

现有运维人员在原有知识结构和工作经验的基础上,接手新系统的运维工作难度较大,需要接受多方位不同层面的培训,包括Oracle、SQL Server等关系型数据库的使用维护、AIX操作系统的使用维护、存储阵列的使用维护等,同时也需要尽快完成对现有应用系统的全面熟悉和使用。通过各种学习和实践,提高运维人员的技术水平,才能使运维工作对NSTL网络服务系统的正常服务提供保障。

“未雨绸缪、呕心沥血、患得患失”,正是十年运维工作的一个写照。

#### 作者简介

郝春云(1973-),毕业于天津大学计算机系,中国科学技术信息研究所信息技术支持中心高级工程师,硕士生导师,研究方向为信息技术、数字图书馆技术。通讯地址:北京市复兴路15号中国科学技术信息研究所信息技术支持中心100038。E-mail:chyhao@istic.ac.cn

毕军荣(1956-),毕业于长沙工学院计算机系,中国科学技术信息研究所信息技术支持中心高级工程师,研究方向为信息技术、数字图书馆技术。通讯地址同上。E-mail:bjr@istic.ac.cn

#### Review on the 10-year Development of NSTL Focusing on the Network System Administration

Hao Chunyun, Bi Junrong / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: With the combination of 10-years' wisdom and courage from everyone working on NSTL, NSTL network service system got a thriving development, which reflects the country's development process. With the network system changing, the system maintenance experienced a lot of changes, which is fully described in the article.

Keywords: NSTL network service system, System administration, Development history

(收稿日期: 2010-08-30)