

学校编码: 10384

分类号_____密级_____

学号: X2010230020

UDC _____

厦 门 大 学

工 程 硕 士 学 位 论 文

CDN 流量监控服务端的设计与实现

Design and Implementation of CDN Flow Monitor System

卢 钊

指 导 教 师: 邱 明 助 理 教 授

专 业 名 称: 软 件 工 程

论 文 提 交 日 期: 2 0 1 3 年 月

论 文 答 辩 日 期: 2 0 1 3 年 月

学 位 授 予 日 期: 2 0 1 3 年 月

指 导 教 师: _____

答 辩 委 员 会 主 席: _____

2013 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

随着宽带网络接入的普及以及个人计算机性能的不不断提高,促进了用户对网络服务质量提出更高的要求。用户希望能够更快更好得通过 Web 浏览器从 Web 服务器上获得网页,内容分发网络(Content Delivery Network, CDN)应运而生。所谓 CDN,是指一种利用避开互联网上有可能影响数据传输速度和稳定性的瓶颈和环节,从而使内容传输更快、更稳定的架构。而现在,CDN 不仅在提供静态 Web 上得到了广泛应用,而且在音频视频这些流媒体服务方面也显现了强劲的发展潜力。

不管是 Web 访问加速,还是流媒体加速都需要使用大量的 CDN 流量。所谓 CDN 流量,是特指 CDN 架构对外提供服务产生的流量,即 CDN 制作的网站镜像提供给用户使用的流量。CDN 流量是否稳定,不仅关系到网站的访问效果,更是影响到 CDN 厂商的直接利润。当流量出现异常的突增或突跌,都一定程度地反映出 CDN 服务出现问题,这时候需要及时通知运维人员来介入以解决问题,从而减少用户的流失,将损失减少到最低。

论文描述了一个面向 CDN 流量的监控平台,从处于最底层的 Squid 服务器的流量开始监控,结合 CDN 上层规划的负载均衡架构,统计出实际对外提供服务的流量情况,供产品人员决策流量调度方案,减少资源浪费,降低成本,更有效地起到 CDN 的作用。同时,用户能够分析历史流量走势,能否发现哪些机房长期提供的服务不稳定,哪些机房分发的内容访问率较低,哪些机房长期流量跑高,一直处于饱和状态。通过定期分析和观察,可调整内容分发方案,使 CDN 系统更趋于稳定和高效。由于 CDN 服务的对象有若干种,其流量报警阈值和走势均具有各自的特点,因此需要提供相对灵活的报警配置方案,才能减少不正常的报警或者遗漏报警。基于上述背景需求,本文致力于研究 CDN 平台下的流量监控系统,目的是更快更准确的发现异常并及时通知用户,帮助企业节省成本并提高 CDN 平台的稳定性。

关键词: CDN; 监控; 流量

Abstract

Nowadays the internet is more and more popular than any other before, and people require improving the speed of network. Consumers wish to get the web page as fast as they can, which push the Content Delivery Network into the internet. CDN is short of Content Delivery Network, which is used to bypass the node which slow the speed of content delivery, and keep the content delivery more stabilize. By now, CDN is not only used to speed up the visiting of static web page, but also considered to have a lot of potentiality on audio and video.

Not only the visit of web page, but also streaming media will cost a lot of CDN flow. The CDN flow, which is product by the CDN company, is used to supply service for the CDN architecture. The stability of CDN flow is relay to the visit effect of web site and the income of CDN company. When the flow raise up or drop down suddenly, the CDN service is usually abnormal at the same time. However when this happen, it must be important to notice the operation staff to deal with and reduce the loss.

This system provide a platform which is based on CDN flow, and it collects the basement data of Squid server which combine with the CDN load balancing to statistics the used flow. The result will be consulted by production staff to reduce the waste of resource, and poll down the prime cost, and raise the effect of CDN. At the same time, with the analysis of the flow data, people can tell which node is not stable enough, and which node's visiting rate is lower than others, and which node's flow is always running above the high line. Thanks to all above, people can adjust the delivery method to keep the CDN system more stable and effective. Also the monitor system provide agility alarm method which can make the alarm information more exactly. The dissertation is about a monitor system base on CDN flow, and it's purpose is to find out the problem of the CDN system, and to notice the user in time, saving the cost and boost up the constancy of the CDN platform.

Key words: CDN; Monitor; Flow

目 录

第一章 绪论	1
1.1 课题研究的背景	1
1.2 课题研究的意义	2
1.3 国内外发展现状	3
1.4 课题研究的目标及其内容	7
1.4.1 研究目标.....	7
1.4.2 研究内容.....	7
1.5 论文组织结构	7
第二章 相关技术和理论介绍	9
2.1 Java 技术介绍	9
2.2 SSH 技术简介	11
2.2.1 Spring.....	11
2.2.2 Struts	13
2.2.3 Hibernate.....	15
2.3 数据传输格式 JSON	16
2.4 JQuery EasyUI 简介	17
2.5 MySQL.....	18
2.6 iBATIS.....	20
2.7 本章小结	22
第三章 系统的需求分析与概要设计	23
3.1 需求分析	23
3.1.1 功能需求分析.....	23
3.1.2 非功能性需求.....	23
3.1.3 定义需求用例图.....	24
3.2 系统总体设计	24
3.2.1 系统架构.....	24

3.2.2 基本环境.....	27
3.2.3 数据库设计.....	27
3.2.4 功能设计.....	33
3.3 本章小结	35
第四章 系统的详细设计	36
4.1 核心层架构设计	36
4.1.1 runRule 程序介绍.....	36
4.1.2 SendMailThread 的介绍.....	37
4.1.3 SendMailThread 任务执行流程.....	38
4.2 报警查看模块设计	39
4.2.1 报警查看.....	40
4.2.2 报警查询算法.....	42
4.2.3 搜索应用服务和 Cache 组.....	43
4.2.4 结果分页的实现.....	43
4.3 邮件管理模块设计	44
4.3.1 发送规则管理.....	45
4.3.2 邮件模版管理.....	46
4.3.3 富文本的二次开发.....	47
4.3.4 模版预览功能.....	47
4.3.5 定时执行规则.....	48
4.3.6 多线程发邮件.....	49
4.4 本章小结	50
第五章 系统的实现与运行效果	51
5.1 系统环境的部署	51
5.2 主页面的实现	51
5.3 邮件管理运行效果	52
5.3.1 邮件发送管理运行效果.....	52
5.3.2 邮件模版管理运行效果.....	55

5.4 统计管理运行效果	55
5.4.1 流量趋势图运行效果.....	55
5.4.2 报警统计运行效果.....	57
5.5 规则管理运行效果	59
5.5.1 查询规则运行效果.....	59
5.5.2 修改规则运行效果.....	60
5.6 本章小结	60
第六章 总结与展望	61
6.1 论文总结	61
6.2 工作展望	61
参考文献	63
致谢.....	65

Contents

Chapter1 Introduction	1
1.1 The Background of the Study	1
1.2 The Meaning of the Study	2
1.3 Development Status at Home and Abroad.....	3
1.4 The Purpose and Contents of the Study.....	7
1.4.1 Purpose.....	7
1.4.2 Contents	7
1.5 Structure of Dissertation	7
Chapter2 Techniques and Theories Related.....	9
2.1 Java Technology	9
2.2 Techniques on the SSH	11
2.2.1 Spring.....	11
2.2.2 Struts	13
2.2.3 Hibernate.....	15
2.3 Data Transmission Format JSON	16
2.4 JQuery EasyUI Description	17
2.5 MySQL.....	18
2.6 iBATIS.....	20
2.7 Summary.....	22
Chapter3 Requirements Analysis and Overall Design of System	23
3.1 Requirements Analysis	23
3.1.1 Functionsl Requirements Analysis.....	23
3.1.2 Non-functional Requirements Analysis	23
3.1.3 Use Case.....	24
3.2 Overall Design	24
3.2.1 System Structure	24
3.2.2 Development Environment	27

3.2.3 Database Design.....	37
3.2.4 Function Design.....	34
3.3 Summary.....	35
Chapter4 Detailed Design of System.....	36
4.1 Core Architecture Design	36
4.1.1 runRule Program Description	36
4.1.2 SendMailThread Description	37
4.1.3 SendMailThread Task Process	38
4.2 Query Alarm Module Design	39
4.2.1 View Alarm	40
4.2.2 List Algorithm.....	42
4.2.3 Search Service And Cache Group.....	43
4.2.4 Result Paging	43
4.3 Mail Manage Module Design	44
4.3.1 Send Rule Manage.....	45
4.3.2 Mail Template Manage	46
4.3.3 Rich Text Development	47
4.3.4 Preview Mail Template	47
4.3.5 Execution Rule Timing	48
4.3.6 Multithreading Send Mail.....	49
4.4 Summary.....	50
Chapter5 Implementation and Running Results of System	51
5.1 Deployment Environment	51
5.2 Implementation of the Main Page	51
5.3 Mail Manage Running Results	52
5.3.1 Mail Send Manage Running Results.....	52
5.3.2 Mail Template Manage Running Results.....	55
5.4 Statistics Manage Running Results	55

5.4.1 Flow Trend Graph Running Results	55
5.4.2 Alarm Statistics Running Results.....	57
5.5 Rule Manage Running Results	59
5.5.1 Query Rule Running Results	59
5.5.2 Modify Rule Running Results	60
5.6 Summary.....	60
Chapter6 Conclusions and Prospect	61
6.1 Conclusions.....	61
6.2 Prospect.....	61
References	63
Acknowledgements	65

第一章 绪论

1.1 课题研究的背景

内容分发网络（Content Delivery Network, CDN）的基本思路是尽可能避开互联网上有可能影响数据传输速度和稳定性的瓶颈和环节，使内容传输更快、更稳定^[1]。通过在网络各处放置节点服务器所构成的在现有的互联网基础之上的一层智能虚拟网络，CDN 系统能够实时地根据网络流量和各节点的连接、负载状况以及到用户的距离和响应时间等综合信息将用户的请求重新导向离用户最近的服务节点上。其目的是使用户可就近取得所需内容，解决 Internet 网络拥挤的状况，提高用户访问网站的响应速度^[2]。CDN 平台如此重要，而我们实现的流量监控就是面向 CDN 平台。

所谓 CDN 流量，即 CDN 系统在网络各处放置的节点服务器传输的数据量，是判断 CDN 系统是否稳定的一项重要指标^[3]。当前的互联网技术蓬勃发展，覆盖面日益扩大，网民数量与日剧增，CDN 平台的服务器对外提供流量服务，一旦出现问题，流量出现大幅下跌，如果没有及时发现并修复，将给网站提供商和终端用户带来严重的损失。同时，CDN 平台的服务能力也能通过当前的流量来衡量。当服务器的流量远没有达到规划值时，可调整服务策略，分配更多的任务给该服务器。当服务器的流量将近达到上限值时，可修改配置，把流量分流到给有更大服务能力的服务器。

论文描述了一个面向 CDN 平台的流量监控系统，能帮助 CDN 服务提供商监控他们的流量服务和调度流量的使用，确保系统的稳定，并且在出现问题时，能够及时的将告警信息送达负责人。用户能够查看当前的报警，也能查询历史报警，便于找出并替换经常出问题的机器，进而提高了平台的服务效果。由于各地区、各类型的 CDN 服务其流量趋势图表现各异，对异常的敏感度也不一样，因此该系统也提供了高灵活度的流量报警规则管理功能。同时，由于 CDN 服务类型的多样性，报警通知的频率、内容也需要根据实际情况配置^[4]。基于上述背景，本文致力于研究 CDN 平台下的流量监控系统，探索更及时的异常发现方案和更

多维度的监控手段，充分挖掘服务器的潜在资源，真正做到帮助企业节省成本和提供更优质的服务，并让企业随时掌握其 CDN 资源的使用情况。

1.2 课题研究的意义

当今互联网上各大网站服务提供商大部分都有向 CDN 厂商购买 CDN 加速服务来保证网站的顺畅访问，更有甚者自身购置大量服务器来自建 CDN 服务^[5]。可见 CDN 一种网络加速方案已日趋人心，并随着互联网上内容的多样化而日趋规模扩大。如：淘宝、百度、腾讯、酷 6、网易等网站，已经非常多，相信以后还会有更多。当 CDN 平台达到一定程度的规模，单靠人工检查服务器的负载、硬盘等常规检测已无法满足系统稳定性的需求，且效率低下。因此，通过系统来监控服务器的各项指标是否健康，并能够决策异常的影响范围，确保最大的异常能够首先被通知到位和处理。监控系统不仅成为了提供 CDN 服务的必需品，而且已经彻底改变了 CDN 提供商对平台的监控方式^[6]。另一方面，由于系统的自动化，部分决策由系统自动产生，将极大的减少人工参与度，并减少人工判断可能带来的失误，从而提高运维部门的工作效率。

CDN 流量可以说是“一个服务器集群的流量汇总”，因为它们本质上跟单台服务器上网卡的流量一样，但是通过采用负载均衡技术，往往提供服务的不单单是一台机器，而是把几十台，上百台服务器当作一个整体对外提供服务^[7]。因此，CDN 流量的监控不仅要把每台的流量进行监控，更重要的是要对整个服务集群做分析，当整个服务集群的流量发生异常，能够及时报警出来。设计和开发对服务器集群的流量监控能够更真实的发现终端用户的访问异常。

CDN 平台的监控在从基于机器指标（负载、磁盘容量、CPU 使用率、内存使用率）的监控向基于流量监控的发展过程中势必会出现许多难题，如：如何讲一台台零散服务器的流量整合到一起成为一个集群的流量，如何区分哪些流量的变化是异常，如何控制报警邮件的发送使其不会大量重复的发导致淹没了最需要关注的报警，如何提高报警算法的效率，如何才能灵活根据时间来决定发送的对象及邮件内容的格式，如何用更直观易懂的方式去展现流量的走势^[8]。这些都值得去探索其实现策略、算法和技术支持。显然，研究 CDN 流量监控系统中出现

的难题并解决之具有重要的意义。

1.3 国内外发展现状

CDN 是以现有互联网为基础的专业化综合应用平台,它由分布在网络边缘的应用服务器群组成,通过应用层的内容智能路由,自动保证用户实现就近访问,为网站提供高效、稳定、安全的内容和应用的分布服务。

举个例子来说,当某用户访问网站时,网站会利用全球负载均衡技术,将用户的访问指向到距离用户最近的正常工作的缓存服务器上,直接响应用户的请求。当用户访问已经使用了 CDN 服务的网站时,其解析过程与传统解析方式的最大区别就在于网站的授权域名服务器不是以传统的轮询方式来响应本地 DNS 的解析请求,而是充分考虑用户发起请求的地点和当时网络的情况,来决定把用户的请求定向到离用户最近同时负载相对较轻的节点服务器上。通过用户定位算法和服务器健康检测算法综合后的数据,可以将用户的请求就近定向到分布在网络边缘的服务器上,保证用户的访问能得到更及时可靠的响应。由于大量的用户访问都由分布在网络边缘的 CDN 节点服务器直接响应了,这就不仅提高了用户的访问质量,同时有效地降低了源服务器的负载压力^[9]。

当前 CDN 的关键技术主要有内容路由技术、内容分发技术、内容存储技术、内容管理技术等。

1、内容路由技术

内容路由技术的作用是将用户请求导向整个 CDN 网络中的最佳节点。最佳节点的选择可以运用多种评估策略来实现,例如用户与服务器距离最近、服务器负载最轻等。这是 CDN 的核心,它决定了整个 CDN 的效率和性能。内容路由技术可以通过多种方法实现,包括 DNS、应用层重定向、传输层重定向等。

2、内容分发技术

内容分发技术指将内容从源到复制缓存到 CDN 边缘服务器的过程。从实现上看,有两种主流的分发技术: PUSH 和 PULL。PUSH 是一种主动分发的技术。通常, PUSH 由内容管理系统发起,将内容从源或者中心媒体资源库分发到各边缘的 Cache 节点。分发的协议可以采用 HTTP、FTP、RSTP 等。对于 PUSH 分

发需要考虑的主要问题是分发策略,即在什么时候分发什么内容,可以根据用户访问的统计信息,以及预定义的内容分发规则来确定。PULL 是一种被动的分发技术, PULL 分发通常由用户请求驱动。当用户请求的内容在本地的边缘 Cache 上不存在时, Cache 启动 PULL 方法从内容源或者其他 CDN 节点实时获取内容。在 PULL 方式下,内容的分发是按需的。在内容分发的过程中,对于 Cache 设备而言,关键的是需要建立内容源 URL、内容发布的 URL、用户访问的 URL,以及内容在 Cache 中存储的位置之间的映射关系。

3、内容存储技术

对于 CDN 系统而言,需要考虑两个方面的内容存储问题。一个是内容源的存储,一个是内容在 Cache 节点中的存储。对存储性能的要求是吞吐量大、可靠性和稳定性。

4、内容管理技术

本地内容管理主要针对一个 CDN 节点(由多个 CDN Cache 设备和一个本地负载均衡设备构成)进行,即决定 CDN 节点中应该存储何种资源以及如何进行节点内容更新。本地内容管理的主要目标是提高内容服务的效率,提高本地节点的存储利用率。通过本地内容管理,可以在 CDN 节点实现基于内容感知的调度,通过内容感知的调度,可以避免将用户重定向到没有该内容的 Cache 设备上,从而提高负载均衡的效率。通过本地内容管理还可以有效地实现在 CDN 节点内容的存储共享,提高存储空间的利用率^[10]。

为了提高访问服务质量,设计 CDN 系统,应该综合考虑以下要求:

1、安全可靠:为保证将来的业务应用,网络必须具有高可靠性。可以在采用硬件备份、冗余等可靠性技术的基础上,采用相关的软件技术,提供较强的管理机制、控制手段、事故监控和网络安全保密等技术措施,提高网络系统的安全可靠性。

2、灵活性和可扩展性:网络系统是一个不断发展的系统,所以它必须具有良好的扩展性。能够根据将来信息化的不断深入发展的需要,方便地扩展网络覆盖范围、扩大网络容量和提高网络各层次节点的功能。

3、开放性:具备与多种协议计算机通信网络互连互通的特性,确保网络系统

基础设施的作用可以充分发挥。

4、可管理性：能够实现监控、监测整个网络的运行状况，合理分配网络资源、动态配置网络负载，可以迅速确定网络故障等。

根据当前 CDN 的关键技术和实现原则可知，整个 CDN 监控体系中，流量监控是整个监控的基础和重点^[11]。流量检测主要目的是通过对网络数据进行实时连续的采集监测网络流量，对获得的流量数据进行统计计算，从而得到网络主要成分的性能指标。CDN 管理员根据流量数据就可以对 CDN 主要成分进行性能分析管理，发现性能变化趋势，并分析出影响 CDN 性能的因素及问题所在。此外，在 CDN 流量异常的情况下，通过扩展的流量检测报警系统还可以向管理人员报警，及时发现故障加以处理。

由此可见，对于一个有效的 CDN 监控系统来说，功能的实现都或多或少的依赖于流量信息的获取^[12]。因此 CDN 流量信息的采集可以说是 CDN 监控系统得以实现的核心基石。它的应用可以在一定程度上检测到入侵攻击，可以有效地帮助管理人员进行 CDN 性能管理，并利用报警机制协助 CDN 管理人员采取对应的安全策略与防护措施，从而减少入侵攻击所造成的损失。

我国 CDN 流量监控行业起步较晚，虽然随着技术人员的开拓创新和外部技术的引入，CDN 流量监控行业已取得了质的发展，但系统的稳定性普遍较低、各系统差异较小，定制化程度低，大部分仅能满足基础需求。目前的流量监控技术在中国 CDN 行业中的应用和发展仍存在诸多不可抗力因素。具体表现在：

1、海量数据处理瓶颈。在设备总数与日俱增的情况下，每分钟采集到的流量数据需要先进行汇总再持久化到数据库。这一过程需要消耗大量的磁盘 IO 和传输连接数。当系统的消耗达到一定程度，会出现连接不上，数据堆积，库表锁死等故障，并且恶性循环，最终导致系统崩溃；

2、网络传输慢。网络传输技术整体水平不高，骨干网调整的现象频频出现。一旦出现网络传输慢或断网的情况，采集的数据上报时将出现堆积，一旦网络恢复，堆积的数据爆发式涌入系统，系统崩溃的概率大大提升；

3、监控体系不完善。业界存在一些系统，虽然是以流量监控为主，但是没有结合实际 CDN 的架构来监控，没有层次区分，导致监控出来的结论不具有参

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库