

云计算的负载均衡 机制研究

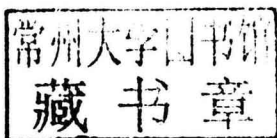
Yunjisuan de Fuzai Junheng
Jizhi Yanjiu

刘琨 著

中国农业大学出版社
CHINA AGRICULTURAL UNIVERSITY PRESS

云计算的负载均衡 机制研究

刘 琨 著



中国农业大学出版社
· 北京 ·

内 容 简 介

本书提出 Hadoop 的云存储负载均衡的优化策略、多目标优化的云存储负载均衡模型、基于动态副本的负载均衡策略、基于虚拟机迁移的资源调度负载均衡策略,以实现云计算系统的存储及资源调度的负载均衡。Hadoop 的云存储负载均衡的优化策略对超负载机架进行优先处理;多目标优化的云存储负载均衡模型综合多种因素计算负载进行数据迁移;基于动态副本的负载均衡策略根据文件访问热度决策副本的创建及删除;基于虚拟机迁移的资源调度负载均衡策略根据节点的负载进行虚拟机资源的调度。

本书可供从事云计算及相关领域研究的学者使用,也可对对负载均衡相关研究感兴趣的学者提供帮助。

图书在版编目(CIP)数据

云计算的负载均衡机制研究/刘琨著. —北京:中国农业大学出版社,2018.2

ISBN 978-7-5655-1988-8

I. ①云… II. ①刘… III. ①云计算-研究 IV. ①TP393.027

中国版本图书馆 CIP 数据核字(2018)第 020048 号

书 名 云计算的负载均衡机制研究

作 者 刘 琨 著

策划编辑 童 云

责任编辑 韩元凤

封面设计 郑 川

出版发行 中国农业大学出版社

社 址 北京市海淀区圆明园西路 2 号

邮政编码 100193

电 话 发行部 010-62818525,8625

读者服务部 010-62732336

编辑部 010-62732617,2618

出版部 010-62733440

网 址 <http://www.caupress.cn>

E-mail cbsszs@cau.edu.cn

经 销 新华书店

刷 涿州市星河印刷有限公司

版 次 2018 年 4 月第 1 版 2018 年 4 月第 1 次印刷

规 格 880×1 230 32 开本 7.25 印张 140 千字

定 价 35.00 元

图书如有质量问题本社发行部负责调换

前 言

随着互联网技术的迅猛发展,云计算这种新兴的商业模式应运而生,它是并行计算、网格计算、虚拟化、分布式计算、网络存储、负载均衡等技术融合发展的产物。云计算技术的出现,将原本用户端的工作放在云端执行,云数据中心承担着复杂忙碌的工作。

对于云数据存储,云数据中心涉及成千上万台服务器和网络设备,这些节点分布不均、节点的配置存在差异、资源访问热度不同,用户的需求多样、实时、复杂,造成云数据中心的数据存储不均衡。例如一些节点存储了大量数据而另一些节点负载较轻;一些节点存储的文件访问热度高,需要频繁地应对用户的访问请求,而另一些节点非常空闲等。数据存储的不均衡将影响系统的性能、降低系统的响应时间,更严重的会引起节点的宕机。对于云资源调度,云中的节点的异构性及用户需求的多样性、不确定性,导致某些节点负载重,非常忙碌;相反,另一些节点负载轻,非常轻松,将影响整个系统的性能及资源利用率。

因此必须解决云存储及云资源调度的负载均衡问题。良好的资源调度负载均衡策略能有效地避免网络负载分布不均、数据流量拥挤、响应时间长等问题,提高应用的执

行效率。

本书针对云计算负载均衡问题提出了解决方案,包括基于 Hadoop 的云存储负载均衡的优化策略、多目标优化的云存储负载均衡模型、基于动态副本的负载均衡策略、基于虚拟机迁移的资源调度负载均衡策略,以实现云计算系统的存储及资源调度的负载均衡。Hadoop 的云存储负载均衡的优化策略对超负载机架进行优先处理;多目标优化的云存储负载均衡模型综合多种因素计算负载进行数据迁移;基于动态副本的负载均衡策略根据文件访问热度决策副本的创建及删除;基于虚拟机迁移的资源调度负载均衡策略根据节点的负载进行虚拟机资源的调度。

本书共分 9 章,其中:

第 1 章为绪论,分析了目前数据中心负载不均衡产生的原因以及负载不均衡带来的问题;对云数据存储的负载均衡、云资源调度的负载均衡的意义进行了详细的阐述;介绍了云计算、云存储、云数据存储负载均衡及云资源调度负载均衡的国内外研究现状;最后简要介绍了主要研究内容。

第 2 章介绍了与负载均衡相关的技术。主要包括负载均衡的分类、常用的负载均衡算法及评价方法、负载的度量方法、数据存储的负载均衡以及资源调度的负载均衡问题。

第 3 章介绍了云计算及云存储的相关原理及技术,包括云计算、云存储的概念;GFS、HDFS 的原理;云存储中的副本技术。

第 4 章针对云计算资源调度的负载均衡问题,提出了

基于虚拟机迁移的负载均衡策略。该策略中详细阐述了负载均衡的框架模块；使用一次平滑指数算法进行负载预测，避免瞬时峰值引起的不必要的虚拟机迁移；使用信息熵算法确定影响因素的权值，客观地计算节点的负载值，使得负载均衡更加合理化。

第5章针对数据存储的负载均衡问题，提出了基于动态副本技术的负载均衡策略，这是一种通过动态调整副本数量、动态创建副本、动态删除副本解决负载均衡的策略，在确定副本负载值时综合了文件热度、CPU性能、带宽、副本一致性维护成本等因素。

第6章针对数据存储的负载均衡问题，提出了基于多目标优化的负载均衡模型，根据文件访问时间、大小、访问热度、节点CPU处理能力、带宽、节点内存大小等各个因素综合确定负载值。

第7章针对数据存储的负载均衡问题，对于Hadoop的HDFS及数据存储负载均衡算法进行了描述，针对该算法存在的超负载节点不能优先及时处理的问题提出了优先处理超负载机架的优化策略。

第8章针对数据存储的负载均衡问题，对于Hadoop负载均衡算法存在的超负载节点不能优先及时处理的问题提出了队列排序的优化策略。

第9章为全文总结与展望，并指出了下一步研究方向。

刘 琨

2017年12月



CONTENTS

目 录

第 1 章 绪 论 /1

- 1.1 研究背景和意义 /1
- 1.2 云计算及云存储 /5
 - 1.2.1 云计算概述 /5
 - 1.2.2 云存储概述 /9
- 1.3 研究现状 /12
 - 1.3.1 云计算的研究现状 /12
 - 1.3.2 云存储的研究现状 /13
 - 1.3.3 云数据存储负载均衡的研究现状 /15
 - 1.3.4 云资源调度负载均衡的研究现状 /17
- 1.4 主要研究内容 /18

第 2 章 负载均衡技术 /21

- 2.1 负载均衡概述 /21
- 2.2 负载均衡的意义 /22

- 2.3 负载均衡算法的分类 /23
 - 2.3.1 动态负载均衡算法和静态负载均衡算法 /23
 - 2.3.2 集中式负载均衡算法和分布式负载均衡算法 /25
 - 2.3.3 预测法和实测法 /26
 - 2.3.4 接收者和发送者启动策略 /26
 - 2.3.5 其他分类 /27
- 2.4 常用的负载均衡算法 /28
 - 2.4.1 轮询算法 /28
 - 2.4.2 加权算法 /29
 - 2.4.3 加权轮询均衡算法 /29
 - 2.4.4 最少连接数与加权最少连接数算法 /30
 - 2.4.5 最低缺失算法 /31
 - 2.4.6 最快响应算法 /31
 - 2.4.7 随机法 /31
- 2.5 负载均衡算法的评价 /32
- 2.6 负载的度量 /33
- 2.7 本章小结 /35

第3章 云计算及云存储 /36

- 3.1 云计算 /36
 - 3.1.1 云计算概述 /36
 - 3.1.2 云计算的技术特点 /40
 - 3.1.3 云计算关键技术 /41
 - 3.1.4 云计算的优势及不足 /43



- 3.1.5 现有开源云计算平台介绍 /46
 - 3.2 云存储 /50
 - 3.2.1 云存储概述 /50
 - 3.2.2 国外云存储的研究进展 /52
 - 3.2.3 国内产业界云存储方案 /54
 - 3.2.4 云存储的研究热点 /55
 - 3.3 GFS /57
 - 3.3.1 系统架构 /57
 - 3.3.2 工作流程 /59
 - 3.3.3 容错 /60
 - 3.4 HDFS /61
 - 3.4.1 HDFS 架构 /61
 - 3.4.2 数据访问方式 /62
 - 3.4.3 副本 /63
 - 3.4.4 文件的删除和恢复 /66
 - 3.4.5 读写文件的过程 /67
 - 3.5 云存储中的副本技术 /70
 - 3.5.1 副本技术概述 /70
 - 3.5.2 GFS 及 HDFS 中的副本技术 /73
 - 3.6 本章小结 /75
- 第4章 资源调度负载均衡策略——基于虚拟机迁移的策略 /76**

-
- 4.1 引言 /77
 - 4.2 研究现状 /79



- 4.3 虚拟机动态迁移的相关技术 /80
 - 4.3.1 虚拟机动态迁移机制 /80
 - 4.3.2 虚拟机动态迁移过程 /84
 - 4.3.3 两种动态迁移方法 /86
- 4.4 基于虚拟机迁移策略的相关定义 /89
- 4.5 基于虚拟机迁移的资源调度负载均衡策略的详细描述 /92
 - 4.5.1 负载均衡策略的架构 /92
 - 4.5.2 采集模块的功能 /93
 - 4.5.3 监测模块的功能 /94
 - 4.5.4 预测模块的功能 /97
 - 4.5.5 源机选择模块的功能 /99
 - 4.5.6 目标机选择模块的功能 /103
- 4.6 实验分析 /105
 - 4.6.1 实验环境配置 /105
 - 4.6.2 预测模块的实验分析 /106
 - 4.6.3 源机选择模块的实验分析 /109
 - 4.6.4 目标机选择模块的实验分析 /114
- 4.7 本章小结 /118

第5章 数据存储负载均衡策略——基于动态副本技术 /120

- 5.1 引言 /121
 - 5.1.1 副本数量 /122
 - 5.1.2 副本创建 /122

- 5.1.3 副本删除 /123
- 5.1.4 副本一致性 /123
- 5.2 研究现状 /123
- 5.3 基于文件热度的动态副本策略的原理 /125
 - 5.3.1 文件热度的定义 /125
 - 5.3.2 副本的选择及副本个数的确定 /126
 - 5.3.3 副本位置的确定 /129
 - 5.3.4 副本删除的策略 /131
- 5.4 基于文件热度的动态副本策略的算法描述 /133
 - 5.4.1 增加副本的算法描述 /133
 - 5.4.2 删除副本的算法描述 /134
- 5.5 实验分析 /135
 - 5.5.1 副本个数 /137
 - 5.5.2 文件的读取时间 /138
 - 5.5.3 增加副本后节点的文件访问量的标准差 /140
 - 5.5.4 节点的存储空间 /141
 - 5.5.5 删除副本后的文件访问时间 /143
- 5.6 本章小结 /144

第6章 数据存储负载均衡策略——多目标优化的模型 /146

- 6.1 引言 /147
- 6.2 研究现状 /149
- 6.3 多目标优化的云存储负载均衡模型的建立思想 /151



- 6.3.1 评价指标的选取 /151
- 6.3.2 负载值的计算 /152
- 6.3.3 相关定义 /153
- 6.4 多目标优化的云存储负载均衡模型的算法描述 /156
- 6.5 实验分析 /158
- 6.6 本章小结 /164

第7章 数据存储负载均衡策略——Hadoop的超负载机架优化策略 /165

- 7.1 研究现状 /166
- 7.2 Hadoop数据存储负载均衡的原理 /167
 - 7.2.1 Balancer程序 /167
 - 7.2.2 Hadoop数据存储负载均衡算法的详细描述 /168
 - 7.2.3 负载迁移时块移动的规则 /171
- 7.3 超负载机架优先处理的策略 /171
 - 7.3.1 Hadoop数据存储负载均衡算法的问题描述 /171
 - 7.3.2 相关定义 /172
 - 7.3.3 超负载机架优先处理策略的主要思想 /174
 - 7.3.4 超负载机架优先处理策略的算法描述 /176
- 7.4 实验分析 /178
- 7.5 本章小结 /181

第 8 章 数据存储负载均衡策略——Hadoop 的队列排序 优化策略 /183

- 8.1 Hadoop 的 Balancer 负载均衡算法的思想及问题 /183
- 8.2 队列排序的优化策略 /184
 - 8.2.1 相关定义 /184
 - 8.2.2 队列排序优化策略的主要思想 /184
 - 8.2.3 队列排序优化策略的算法描述 /185
- 8.3 实验分析 /186
- 8.4 本章小结 /189

第 9 章 总结与展望 /191

- 9.1 总结 /191
- 9.2 进一步工作 /195

参考文献 /197



第 1 章 绪 论

1.1 研究背景和意义

随着科学技术的飞速发展,人们对计算能力的要求越来越高,单独的系统已不能满足人们对计算能力的需求。于是,网络协作、充分利用闲散资源成了最初的解决方法——网格计算(Grid Computing)。网格计算通过整合网络中的大量闲散资源,在动态的、自治的、异构的环境中协调资源共享、解决大规模的具有挑战性问题。但是网格计算遇到了很多难以突破的问题,致使网格计算的思想虽好,但是商业应用却受到了极大的限制。

—网络上的信息量逐渐增大,越来越多的人使用互联网来获取信息、购物和娱乐,网络数据量和用户请求数出现了爆发式的增长,海量数据需要进行存储,同时高性能计算的发展,企业内部信息不断增加,外部需求快速增多。对服务器计算和处理能力提出了更高的要求,使它们在合

理接受客户端请求的基础上又要在最短的时间内做出应答响应,以提高用户体验度。根据易观国际提供的数据,淘宝网目前每天的活跃数据量超过 50 TB,有 4 亿条产品信息和 2 亿多注册用户在上面活动,每天数据访问量超过 4 000 万。谷歌每天需处理 9 100 万次查询,一年平均要在数据库保存 33 万亿条用户的查询记录。如此巨大的数据量和访问请求,迫切需要具备快速响应能力、高可用性、高扩展性、易于管理的服务器来提高网络吞吐量和及时响应用户请求的能力。对于服务提供商来说,现有的基础服务设备已不能满足客户的需求,为了改变这种情况,大部分公司首先考虑到的是购买更多处理能力更强的服务器,显然这样会耗费公司大量的费用,同时后期维护也需要大量资金,对于资金欠缺的中小型企业来说,这是很大的负担。但是这种选择是治标不治本,设备升级远远跟不上用户访问量增加的速度,而且每次硬件升级都会面临数据的迁移及可靠性、稳定性方面的风险。另一种解决方案是将多台服务器连接成一个整体,通过网络和分布式技术共享资源以实现用户请求和计算任务的并行高效处理。

在这样的背景下,随着互联网技术的迅猛发展和计算机技术的不断进步,更大规模、更新的互联网应用发展迅速,使得更多的用户通过互联网共享各种资源,由此产生了一种新的商业和计算模式——云计算。在这种新兴的商业模式下,用户能够按需使用其中的计算资源及服务,且不受时间限制,扩展性强。与传统网络相比,云计算的出现改变了网络的服务模式,也改变了人们使用网络和计



算机的方式,还是下一代网络运用的新技术。从用户的角度来说,用户购买云计算服务,云计算服务提供商就能按需动态地提供给用户相应的服务,保证用户 SLA;从云服务提供商角度而言,云计算是将存储资源、计算资源及软件服务通过互联网提供给用户的一种计算模式,实现用户与计算资源的管理相分离。因此用户不必去关心各种硬件及软件资源,通过仅在需要资源时申请获取资源并为所需的资源付费。

云计算的服务模型大致包括“用户端”“云端”及连接两者的传输介质。“用户端”指用户接入云端的终端设备,比如电脑、手机或其他终端设备;“云端”指的是提供计算或存储能力的基础设施中心、平台或应用服务器等及云数据中心,提供的服务类型包括基础设施、平台和应用等。云计算的思想就是要把服务处理及存储尽量集中到“云端”,“用户端”瘦身成一个简单的终端或浏览器,因此对于“云端”的设计尤为关键。

云计算系统服务的实现主要依靠云数据中心完成,由于云计算技术的发展,对云数据中心的要求越来越复杂。云数据中心主要由数量巨大的服务器和网络设备组成,这些网络设备和服务器的异构性强,用户的需求复杂,要求高质量的服务,要求更合理的动态资源管理,因此对云数据中心提出了更高的要求。但实际上云数据中心目前存在着效率低、成本高、能耗高等问题。据资料显示,在我国,云数据中心的众多服务器均处在空闲状态,其中只有10%左右的服务器被充分利用,处于空闲状态的服务器也

会有很大的功耗,相当于满载服务器的60%,云数据中心相当一部分能耗都被浪费了。

随着云数据中心的运行,新节点不断加入到云中,旧节点不断从云中删除,节点的动态变化造成系统数据负载的不均衡,必须对这些节点进行负载均衡^[1]。目前云数据中心的资源管理大都采用静态管理方式,无法适用于网络的动态变化,将造成资源分配的不合理;同时各个服务器上的数据存储也存在着不均衡的问题。这种资源、数据存储负载不均衡的问题将导致云资源的浪费,影响云数据中心的效率。因此,必须设计合理的云资源调度算法和云数据存储算法来解决云数据中心负载不均衡的问题,从而提升资源利用率,这是云计算研究领域的一个关键问题。一个好的负载均衡策略能够使得负载分布更加均衡,有效地避免数据流量的拥挤,缩短响应时间,提高执行效率,降低能耗。

负载均衡技术主要任务是在分布式环境下协调由网络连接的一组自治计算节点,使它们能够均衡地并行处理请求,充分挖掘各个计算节点的计算能力,以减少系统对请求的平均响应时间,从而提供服务质量。云计算的不断发展对负载均衡技术提出了更高的要求,负载均衡是云计算中资源管理、资源调度、数据存储的关键问题,它是云计算中亟待解决的问题之一。主要原因在于^[2]:

(1)负载均衡技术能够提升硬件的处理能力、减少硬件的投入。随着云计算中用户需求的不断增加,需要的资源越来越多,但云数据中心的硬件不能无限增加,必须对这些硬件设施进行有效的管理和利用。负载均衡技术能